SURROGATE LANGUAGES AND THE GRAMMAR OF LANGUAGE-BASED MUSIC

EDITED BY: Yoad Winter and Laura McPherson
PUBLISHED IN: Frontiers in Communication and Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-716-0
DOI 10.3389/978-2-88974-716-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SURROGATE LANGUAGES AND THE GRAMMAR OF LANGUAGE-BASED MUSIC

Topic Editors:
Yoad Winter, Utrecht University, Netherlands
Laura McPherson, Dartmouth College, United States

Citation: Winter, Y., McPherson, L., eds. (2022). Surrogate Languages and the Grammar of Language-Based Music. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-716-0

Table of Contents

04 ***Editorial: Surrogate Languages and the Grammar of Language-Based Music***
Laura McPherson and Yoad Winter

07 ***Categoricity, Variation, and Gradience in Sambla Balafon Segmental Encoding***
Laura McPherson

18 ***Perception of Nigerian Dùndún Talking Drum Performances as Speech-Like vs. Music-Like: The Role of Familiarity and Acoustic Cues***
Cecilia Durojaye, Lauren Fink, Tina Roeske, Melanie Wald-Fuhrmann and Pauline Larrouy-Maestri

31 ***Linguistic Surrogacy With Minimal Semantics Among the Dagomba of Ghana***
Fusheini Hudu

38 ***Igbo Speech Surrogacy: Preliminary Findings Based on the Oja Flute***
Aaron Carter-Ényì, Nnaemeka C. Amadi, Quintina Carter-Ényì, Charles Chukwudozie, Jude Nwankwo and Ebruphiyor Omodoro

44 ***Feeling the Beat in an African Tone Language: Rhythmic Mapping Between Language and Music***
Kathryn H. Franich and Ange B. Lendja Ngnemzué

68 ***Rhythm-Speech Correlations in a Corpus of Senegalese Drum Language***
Sofiya Ros

74 ***When Music Speaks: An Acoustic Study of the Speech Surrogacy of the Nigerian Dùndún Talking Drum***
Cecilia Durojaye, Kristina L. Knowles, K. Jakob Patten, Mordecai J. Garcia and Michael K. McBeath

91 ***Systems of Communication: Aspects of Culture and Structure in Speech Surrogates***
Lucas James

106 ***Singing the Individual: Name Tunes in Oyda and Yopno***
Azeb Amha, James Slotta and Hannah S. Sarvasy

113 ***The Language of Gángan, A Yorùbá Talking Drum***
Samuel Kayode Akinbo

127 ***A Flute, Musical Bows and Bamboo Clarinets that "Speak" in the Amazon Rainforest; Speech and Music in the Gavião Language of Rondônia***
Julien Meyer and Denny Moore

145 ***The Speech Surrogacy Systems of the Yoruba Dùndún and Bàtá Drums. On the Interface Between Organology and Phonology***
Mariano González and Olupemi Oludare

164 ***A Preliminary Account of the Northern Toussian Balafon Surrogate Language***
Anthony Struthers-Young

frontiers in Communication

EDITORIAL
published: 25 February 2022
doi: 10.3389/fcomm.2022.838286

Editorial: Surrogate Languages and the Grammar of Language-Based Music

Laura McPherson[1]* *and Yoad Winter*[2]*

[1] Dartmouth College, Hanover, NH, United States, [2] Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands

Keywords: linguistics, music, language, linguistic theory, phonology, syntax

Editorial on the Research Topic

Surrogate Languages and the Grammar of Language-Based Music

Most cultures have ways of keeping a close relation between language and musical expression. In many parts of the world, these two forms of expression can directly coincide in musical surrogate languages: systems in which musical instruments mimic or indirectly encode concrete linguistic expressions. Drums, wind instruments, and xylophones are commonly used for speech surrogates across different cultures, and these systems often demonstrate closer contact between language and music than is familiar in Western traditions. The importance of this "transfer" for linguistics was observed as early as work by Sapir (1933). However, while the documentation of speech surrogates has been of continuous interest to anthropologists and musicologists (Sebeok and Umiker-Sebeok, 1976), it is only recently that theoretical linguists have started to focus on speech surrogacy as a major area of study that interacts with these adjacent disciplines. Recent results reveal implications of whistled and instrumental speech surrogates for phonetics (Meyer, 2008), phonology (McPherson, 2018; Seifart et al., 2018), and syntax (Winter, 2014). This linguistic interest in speech surrogacy appears at a time of renewed theoretical interest in the linguistic analysis of music (Katz and Pesetsky, 2011; Schlenker, 2017) and dance (Patel-Grosz et al., 2018; Charnavel, 2019) also in circumstances that do not involve speech surrogacy.

The aim of this volume is to highlight state of the art work on speech surrogacy from theoretical linguistic perspectives. It represents the first explicit collection on speech surrogacy since the seminal volumes of Sebeok and Umiker-Sebeok in the 1970s, and the first ever collection to focus on the connections between these systems and theoretical linguistics.

James presents a broad overview of musical surrogate languages as systems of communication. He surveys decades of interdisciplinary literature to show how a complex interplay of discursive context, linguistic structure, and cultural constraints shape the form and usage of speech surrogates.

Four articles in the collection focus on Yorùbá drumming, perhaps one of the most famous cases of "talking drums." Akinbo explores the surrogate phonology of the *gángan*, part of a family of tension drums known in Yorùbá as *dùndún*. Based on primary data collected with five drummers, he shows that the *gángan* encodes the phonetic realizations of both lexical and grammatical tone. Syllable structure likewise plays a role in the system, with the drum membrane only optionally struck for onsetless (V) syllables. Durojaye, Knowles, et al. likewise study the acoustic properties of the *dùndún* including fundamental frequency, duration, and intensity. In addition to comparing these properties in *dùndún* "speech mode" to spoken Yorùbá, they also compare "drum singing" to sung Yorùbá, and find that the *dùndún* is capable of quite precise mimicry across modes. While the papers by Akinbo, and Durojaye, Knowles et al. focus on

Edited and reviewed by:
Mila Vulchanova,
Norwegian University of Science and Technology, Norway

****Correspondence:***
Laura McPherson
laura.e.mcpherson@dartmouth.edu
Yoad Winter
y.winter@uu.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *23 December 2021*
Accepted: *26 January 2022*
Published: *25 February 2022*

Citation:
McPherson L and Winter Y (2022) Editorial: Surrogate Languages and the Grammar of Language-Based Music. Front. Commun. 7:838286. doi: 10.3389/fcomm.2022.838286

production, the work by Durojaye, Fink et al. turns to questions of perception. In their experiment, 107 participants listened to samples of *dùndún* in both speech mode and musical mode and were asked to classify each sample as either speech-like or music-like; around half of the participants were familiar with the *dùndún*, and 28 participants were speakers of Yorùbá. On average, all listeners were able to distinguish between speech mode and musical mode, though accuracy was higher amongst experienced participants. Finally, the paper by González and Oludare looks at Yorùbá talking drums beyond the *dùndún* by comparing this surrogate tradition to the *bàtá* drum ensemble, which differs considerably in its organology from the tension drum. While the *dùndún* surrogate uses the fundamental frequency to mimic Yorùbá tone and intensity to encode limited vowel contrasts, the *bàtá* system is based in sophisticated timbral differences. Their study highlights the importance of considering instrumental constraints in the analysis of musical surrogates.

Two other papers, by Ros and Hudu, investigate drum surrogate traditions outside of Nigeria. Ros investigates linguistic encoding in the *sabar* drumming of Wolof in Senegal. Unlike Yorùbá, Wolof is not a tonal language, which raises questions of how linguistic structure is represented in drummed form. Ros' analysis of a corpus of rote and improvised phrases shows that different drum strokes are correlated with vowel quality, including height (like Yorùbá or whistled speech of non-tonal languages) as well as backness. Hudu focuses on surrogate traditions of the Dagomba of Ghana, among which are talking drums, bells, and fiddles. What is interesting in this case is that all of these speech surrogates have been borrowed from other ethnolinguistic groups, and as such, surrogate speech is not in Dagbani but in other languages like Akan and Hausa, which the musicians do not typically speak. Hudu discusses the acquisition and processing of these non-native surrogate languages and highlights the important role that they play in Dagomba culture.

Beyond talking drums, the two papers by Struthers-Young and McPherson cover phonetic and phonological encoding in balafon (resonator xylophone) surrogate languages of Burkina Faso. Struthers-Young provides a first account of the Northern Toussian balafon surrogate, which he shows to encode the language's three tones at a post-lexical level (including downstep). In addition, the system also encodes some elements of syllable structure. McPherson digs into syllable structure encoding in the related Sambla balafon surrogate tradition. She demonstrates that while some aspects are categorically encoded by the balafon, others show free variation between encoding strategies. Beyond these conscious encoding choices, acoustic analysis of duration shows a very tight relationship between notes on the balafon and vowel-to-vowel intervals in spoken Seenku, suggesting that surrogate timing is controlled by perceptual centers of syllables in the musician's inner voice.

The paper by Carter-Ényì et al. analyses bidirectional mapping (speech-to-music and music-to-speech) in the Ìgbò *ọ̀jà* flute surrogate of Nigeria. Carter-Ényì et al. use contour analysis to demonstrate the close connection between the *ọ̀jà* phrases and phonological tone in Ìgbò. By contrast, phonetic effects like declination were not found. A pilot identification task with one listener showed that most phrases could be correctly—and quickly—identified. Misidentifications were tied to dialectal differences, showing the importance of considering dialect in studies of musical surrogate speech.

Meyer and Moore also focuses on a flute surrogate tradition, this time among the Gavião of the Amazon. The *kotiráp* flute is one of three instrumental traditions said to be able to "sing," alongside a pair of mouth bows (*iridnáp*) and a trio of bamboo clarinets (*totoráp*). The authors show how these instruments encode the tone of Gavião songs rather than spoken language directly and represent a little-studied case of "surrogate song," which may show different patterns of phonological encoding from surrogate speech.

While most of the papers in the collection focus on instrumental surrogates, language-based music is not limited to this surrogacy. Franich and Lendja Ngnemzué presents an analysis of text-setting in Medʉmba, a Grassfields Bantu language of Cameroon. Rhythmic textsetting in African languages is understudied compared to tonal textsetting, since the metrical structure of these languages is often ill-understood and the polyrhythmic nature of many African musical traditions poses a challenge for identifying "the beat." The authors find correspondences between Medʉmba linguistic structure and musical rhythm that provide another source of evidence for stem-initial prominence in the language.

Finally, Amha et al. provides the first crosslinguistic documentation of "name tunes," a tradition wherein individuals are identified and addressed using either sung or whistled unique musical phrases. This rare system has developed independently amongst Oyda speakers in Ethiopia and Yopno speakers in Papua New Guinea and represents a unique way in which music can play a linguistic and communicative function, without directly encoding linguistic structure of the spoken language.

The wide range of topics covered in this special collection points to both the multifaceted nature of the language-music connection as well as to the broad distribution of language-based musical traditions worldwide. We hope this collection will spark further interest amongst linguists and language researchers and be the first of many exploring connections between linguistic theory and language-based music.

AUTHOR CONTRIBUTIONS

LM and YW are responsible for writing and editing this editorial.

FUNDING

This Research Topic has been supported by the National Science Foundation under grant BCS-1943896 (LM) and by a Free Competition grant of the Dutch Research Council (NWO) Project 360-89-060, When language has a beat (YW).

REFERENCES

Charnavel, I. (2019). Steps towards a universal grammar of dance: local grouping structure in basic human movement perception. *Front. Psychol.* 10, 1364. doi: 10.3389/fpsyg.2019.01364

Katz, J., and Pesetsky, D. (2011). *The Identity Thesis for Music and Language*. MIT. Available online at: https://ling.auf.net/lingbuzz/000959

McPherson, L. (2018). The Talking Balafon of the Sambla. *Anthropol. Linguist.* 60, 255–294. doi: 10.1353/anl.2019.0006

Meyer, J. (2008). Typology and acoustic strategies of whistled languages: phonetic comparison and perceptual cues of whistled vowels. *J. Int. Phon. Assoc.* 38, 69–94. doi: 10.1017/S0025100308003277

Patel-Grosz, P., Grosz, P. G., Kelkar, T., and Jensenius, A. R. (2018). "Coreference and disjoint reference in the semantics of narrative dance," in *Proceedings of Sinn und Bedeutung 22, Vol. 2*, eds U. Sauerland, and S. Solt (Berlin: Leibniz-Centre General Linguistics (ZAS)), 199–216. doi: 10.21248/zaspil.61.2018.492

Sapir, E. (1933). "Language," in *Encyclopaedia of Social Sciences, Vol. 9*, eds E. R. A. Seligman, and A. Johnson (London: MacMillan & Co. Ltd), 155–169.

Schlenker, P. (2017). Outline of music semantics. *Music Percept.* 35, 3–37. doi: 10.1525/mp.2017.35.1.3

Sebeok, T. A., and Umiker-Sebeok, D. J. (1976). *Speech Surrogates: Drum and Whistle Systems*. The Hague: Mouton. doi: 10.1515/9783110804416

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *R. Soc. Open Sci.* 5, 170354. doi: 10.1098/rsos.170354

Winter, Y. (2014). On the grammar of a Senegalese drum language. *Language* 90, 644–668. doi: 10.1353/lan.2014.0061

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Communication

ORIGINAL RESEARCH
published: 17 May 2021
doi: 10.3389/fcomm.2021.652635

Categoricity, Variation, and Gradience in Sambla Balafon Segmental Encoding

Laura McPherson *

Department of Linguistics, Dartmouth College, Hanover, NH, United States

This paper analyzes the musical surrogate encoding of Seenku (Mande, Burkina Faso) syllable structure on the balafon, a resonator xylophone used by the Sambla ethnicity. The elements of syllable structure that are encoded include vowel length, sesquisyllabicity, diphthongs, and nasal codas. Certain elements, like vowel length and sesquisyllabicity, involve categorical encoding through conscious rules of surrogate speech, while others, like diphthongs and nasal codas, vary between being treated as simple or complex. Beyond these categorical encodings, subtler aspects of rhythmic structure find their way into the speech surrogate through durational differences; these include duration differences from phonemic distinctions like vowel length in addition to subphonemic differences due to phrasal position. I argue that these subconscious durational differences arise from a "phonetic filter", which mediates between the musician's inner voice and their non-verbal behavior. Specifically, syllables encoded on the balafon may be timed according to the perceptual center (p-center) of natural spoken rhythm, pointing to a degree of phonetic detail in a musician's inner speech.

Keywords: phonology, phonetics, speech surrogate, rhythm, Burkina Faso, Mande

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco, Sevilla University, Spain

Reviewed by:
Lei He, University of Zurich, Switzerland
Gertraud Fenk-Oczlon, University of Klagenfurt, Austria

****Correspondence:***
Laura McPherson
Laura.e.mcpherson@dartmouth.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *12 January 2021*
Accepted: *26 April 2021*
Published: *17 May 2021*

Citation:
McPherson L (2021) Categoricity, Variation, and Gradience in Sambla Balafon Segmental Encoding. Front. Commun. 6:652635. doi: 10.3389/fcomm.2021.652635

INTRODUCTION

For linguists studying musical surrogate languages, the main question of interest is which linguistic features get encoded into musical form and how? Many older studies of surrogate languages, written largely by anthropologists or ethnomusicologists, contain general statements on the connection between spoken and musical language (e.g. "drums encode tone and speech rhythm") with no further explanation. This leaves us to wonder, which tones are encoded, lexical or grammatical? Which aspects of rhythm—only the phonemic ones, such as vowel length or syllable shape, or also more subtle rhythmic effects?

Through careful comparison of spoken and surrogated language, we can unveil these patterns of encoding. However, I suspect that in many cases we can draw a distinction between conscious elements of encoding—let's call these the grammatical rules of the speech surrogate—and subconscious patterns of encoding. Determining which linguistic elements are consciously encoded and which are subconscious stands to shed light on a speaker's metalinguistic knowledge and the division between phonology and phonetics.

This paper explores these questions by looking at segmental encoding strategies on the Sambla balafon, a West African resonator xylophone used among the Seenku-speaking population of southwestern Burkina Faso. Previous work (McPherson 2018) on the surrogate system focused largely on the encoding of tone, showing a distinction between lexical/morphological and postlexical tone, with the former encoded and the latter ignored by musicians. In this paper, I take a closer look at segmental encoding, specifically syllable structure (as individual consonant and vowel phonemes are not distinguished in the speech surrogate). I show that musicians' conscious rules of surrogate encoding marks a categorical distinction between simplex and complex syllables, with complex defined as CVː (long vowels) and C(ə)CV (sesquisyllables; Matisoff 1990; Pittayaporn and Enfield, 2015). Two other syllable shapes, namely CVN (nasal codas) and CVV (diphthongs) are variably treated as complex. In addition to these conscious encoding strategies, syllable structure and speech rhythm more broadly also influence the Sambla balafon surrogate in subtler subconscious durational differences that closely mirror spoken language.

This range of encoding behavior—categorical, variable, gradient—can shed light on the phonetic and phonological structure of the spoken language, and what speakers know about that structure. First, the variable encoding of diphthongs is shown to be the result of conscious disambiguation on the part of the musician, demonstrating a recognition of the distinction between monophthongs and diphthongs when such a contrast is needed. The variation with nasal codas, on the other hand, is argued to arise from their weak phonological representation in the language itself (McPherson 2020a; McPherson 2020b; McPherson 2020c), with more "complex" encodings correlated with environments in which they surface as true coda consonants, and more "simplex" encodings correlated with environments in which they are subsumed into the following onset or the preceding vowel. Finally, the presence of gradience in the system suggests that surrogate encoding passes through a subconscious "phonetic filter", in which surrogate speech is mediated by the musician's inner voice. Specifically, syllable timing on the balafon appears to be tied to the perceptual center (p-center) of spoken syllables (Morton et al., 1976; Ryan 2014), despite the fact that musicians are not speaking as they play. This in turn points to a relatively high degree of phonetic detail in inner speech (Corley et al., 2011; Oppenheim 2013; Martin et al., 2018, *inter alia*), demonstrating yet another way in which musical surrogate languages can serve as a crucial source of data for understanding the human language faculty.

The paper is structured as follows. In *Background on Sambla Music and the Seenku Language*, I provide background information on Sambla music and the spoken language, Seenku. *Encoding Strategies* turns to the different encoding strategies, covering categorical (*Categoricity*), variable (*Variability*), and gradient (*Gradience*) encoding in turn. *Discussion* discusses the results in light of the phonetic filter and p-centers, and *Conclusion* concludes.

BACKGROUND ON SAMBLA MUSIC AND THE SEENKU LANGUAGE

The Sambla people are a Mande ethnicity with a population of less than 20,000 living in southwestern Burkina Faso. The name "Sambla" (also spelled Sembla) is an exonym, but is used by the people to refer to themselves when speaking French or Jula. I will be referring to the spoken language by its endonym, Seenku (IPA [sɛ̃́ː-kû]), though it is likewise referred to as Sambla or Sembla in the literature.

In this section, I will briefly lay out the pertinent background information on Sambla balafon music and Seenku phonology, which will set the stage for an analysis of segmental encoding.

Sambla Balafon

Arguably the most important instrument in the Sambla musical tradition is the balafon. The term "balafon" is applied to a range of resonator xylophones found throughout West Africa, where tuned gourds are hung beneath each wooden key to amplify the sound. The balafon is played at all major cultural events, including weddings, funerals, communal work parties, and religious festivals.

Balafon playing is a hereditary profession, passed from father to son in just a small number of families belonging to the griot caste. Unlike some balafon traditions, such as the Senufo balafon orchestras (Zemp and Soro 2010), Sambla balafon music involves just a single instrument at any event. However, three people play at the same time: a simple middle part sets the tempo, while a more complicated bass part creates a layered polyrhythmic backdrop to the "soloist" on the treble. It is this soloist who practices speech surrogacy, through both melodic lines that stand for proverb-rich lyrics of songs (the "sung mode" of surrogacy) and improvised surrogate utterances that communicate directly with the audience or other musicians (the "speech mode" of surrogacy). This paper focuses solely on speech mode, which, as the name suggests, displays a much tighter connection to the spoken language.

The Sambla balafon is tuned to a pentatonic scale, though the exact notes/frequencies vary instrument to instrument; as musicians put it, just like people, each instrument has its own voice. The intervals between the notes, however, remain fixed. The names of the notes, along with their closest corresponding Western scale degrees, are shown in **Table 1**.

As this table shows, the Sambla scale is rather unusual, with the presence of both the minor and major third. However, the minor third, the "fetish balafon key", is only rarely used. It is reserved for spiritual uses and typically does not figure in the speech surrogate, with the exception of one lexical ideogram expression corresponding to "yes" or "that's correct", in which the player alternates between ♭3 and 5. It should be noted that the spatial relations of "under" and "above" in the key names refer to physical location, since low notes have larger gourds beneath them and are hence "higher" from the ground than high notes; in other words, the keys of the balafon slope downwards as one moves from the bass to the treble.

For more information on the Sambla balafon, including its history, construction, and social uses, see Strand (2009) and McPherson (2018).

Seenku Phonology

Seenku is a member of the Samogo group of Western Mande languages. It is phonologically rich, with complex consonantal, vocalic, and tonal inventories. Balancing this complexity is a largely monosyllabic vocabulary, though syllable structure itself

TABLE 1 | Notes and scale degrees of the Sambla balafon.

Western scale degree	Seenku note name	Abbreviation	Gloss
1	*bâ̰a̰-ɲȁ / sərȁ-kȕa*	B / S	"balafon mother" / (no translation)
♭3	*ĵio-bȁ̰a̰-dȅn*	J	"fetish balafon key"
3	*bâ̰a̰-ɲȁ-gṵ̏-nɔ̰̏n*	Bg	"the one under the balafon mother"
5	*tərɔ́n-tərɔ́n*	T	(no translation)
6	*sərà-kùa-kɔ̰̀n-nɔ̰̀n*	Sk	"the one above the *sərȁ-kȕa*"

TABLE 2 | Seenku consonant inventory.

	Bilabial	Alveolar	Palatal	Velar	Labiovelar
Plosive	p b	t d	c ɟ <j>	k g	kp gb
Nasal	m	n	ɲ	ŋ	ŋm
Affricate		ts dz			
Fricative	f	s			
Approximant		l	j <y>		w

can be complex. For more detail on Seenku's phonological system, see McPherson (2020a), McPherson (2020b).

Segmental Inventories

The consonant inventory is shown in **Table 2**, where any orthographic deviations from IPA are shown in angled brackets.

As this table shows, Seenku contrasts five places of articulation, including labiovelars. Unusually for Mande, it displays voiced and voiceless alveolar affricates. There is no contrast between [l] and [ɾ], with the realization depending upon the context.

The oral vowel inventory contains either eight or nine vowel qualities, depending upon the speaker; this inventory results from an [ATR] contrast among high and mid vowels, though for some speakers, I have seen no evidence of high [-ATR] /ɪ/. Thus, the vowel inventory can be summarized as /i (ɪ) e ɛ a ɔ o ʊ u/. These oral vowels can be either phonemically short or long.

Seenku also has phonemic nasal vowels, though the inventory is more constrained, with just a five-way distinction: /ĩ ɛ̃ ã ɔ̃ ũ/; these nasal vowels may also be short or long.

In addition to these monophthongs, Seenku displays an impressive array of diphthongs, including (but not limited to) /ia iɛ ie ua uo uɔ ɔɛ oe/, etc. A striking fact of Seenku phonology is that diphthongs can also be either short or long, due to the non-moraic nature of the diphthong-initial vocalic element (McPherson, 2020a; McPherson, 2020b).

Tonal Inventory

Seenku has four contrastive tone levels, which I call Superhigh (S), High (H), Low (L), and Extralow (X), marked with double acute (a̋), acute (á), grave (à) and double grave (ȁ) diacritics, respectively. The four tone levels are distinguished solely by f0, with no significant role played by vowel length or phonation. A minimal set for level tones is shown in (1):

(1) a. sí̋ 'tree sp.'
b. sí 'reciprocal'
c. sì 'first son (birth order name)'
d. sȉ 'water jar'

These four level tones can combine to create a vast array of two- and three-tone contours. The most common lexical contours are HX (â), LS (ǎ), and HS (áa̋); contours created by grammatical tone and clitic elision include SX (ä), XH (ȁá), SH (a̋á), and HL (áà). Three-tone contours include XHX (e.g. *gɔ̏ɔ̂n* "sorrel") and LSX (e.g. *nàä* "come (perfect)").

Syllable and Word Structure

Most Seenku vocabulary is monosyllabic, but there is a wide range of possible syllable shapes, including sesquisyllabic words (Matisoff 1990; Pittayaporn and Enfield, 2015), i.e. a short minor syllable (Cə) followed by a full syllable; it may also be possible to analyze these as syllables with a complex onset, broken up by an epenthetic or even excrescent vowel (McPherson 2020a). Seenku syllable structure is schematized in (2):

(2) (C)(əC)V(V)(:)(N)

The only obligatory element is the syllable nucleus, V (though in the case of the 1sg *ń*, it can also be a nasal). Only a small number of pronouns (e.g. 3sg *ȁ,* 2pl *í,* etc.) are vowel-initial (indeed, consist of only V). All other vocabulary must have an onset consonant (e.g. *kâ* "griot", *mí̋* "1pl" etc.). Syllables can be either mono- or sesquisyllabic, with the addition of əC after the initial onset consonant (e.g. *məní̋* "woman", *səgȁ* "sheep", etc.). The syllable nucleus can be either a monophthong, as seen so far, or a diphthong (e.g. *kű̋ɛ* "others", *səgȕa* "stack (v.)", etc.). Both monophthongs and diphthongs can be either short or long, with long vowels indicated here by doubling the vowel (e.g. *kâa* "fight", *kȕaa* "farm (intr.)", etc.). Finally, the only permissible coda is a nasal whose place of articulation is non-contrastive. In fact, in most cases, it is realized as either late nasalization of the preceding vowel (contrastive with phonemically nasal vowels) or as nasalization of a following sonorant. Only in rare instances is it realized fully as a nasal stop; for more details, see McPherson (2020c). The repercussions of this weak phonological element for the balafon surrogate language will be discussed below.

Any of these syllable elements may co-occur, with lexical frequency decreasing as complexity increases (McPherson 2020a). In other words, simple CV syllables like *sȍ* 'horse' are vastly more frequent than complex syllables like CəCVVːN, of which we find just a single instance *təgǔaan* 'tree sp. (*Carapa procera*)'.

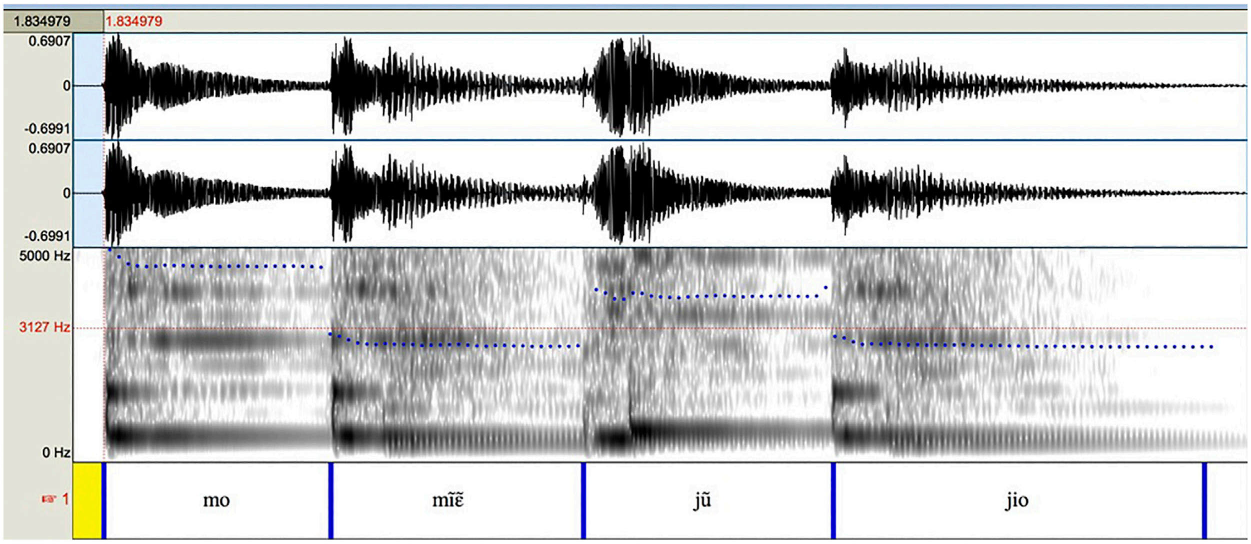


FIGURE 1 | Praat TextGrid of balafon phrase *mó mîɛ̰ jṵ ĵio* “I saw a long snake”.

ENCODING STRATEGIES

With this musical and phonological background in place, we can now turn to their confluence in the balafon surrogate language.

The findings in this paper are drawn from a corpus of data collected by the author since 2014. The majority of the data is elicited, though the corpus also includes some naturally occurring phrases from field recordings. All phrases are in the balafon speech mode—the surrogate mode most closely emulating speech—rather than musical or sung mode; for more details, see McPherson (2018) and McPherson and James (forthcoming). In total, the corpus contains 1259 syllables, only slightly higher than the number of words given the largely monosyllabic nature of Seenku vocabulary. Data were recorded with four musicians, all from the same Diabate family of griots from Toronsso, Burkina Faso, though the majority of the data are from a single musician, Mamadou Diabate.

In order to determine the encoding strategies, each syllable in the corpus was coded for a number of binary linguistic features related to syllable structure and tone: Sesquisyllabicity, Diphthong, Long vowel, Coda nasal, Contour tone. In addition, the syllables were coded for the binary musical feature Flam, that is, whether the syllable was played with a flam (two strikes) or a single strike.

The total duration in milliseconds of each syllable played on the balafon was also included. Measurements were made by creating TextGrids in Praat (Boersma and Weenink 2017), with boundaries placed at the attack of the first strike of each word (or the only strike for single-strike syllables). The interbeat duration between the two strikes of a flam is not measured but rather the two are considered as a single unit for the purposes of measurement, since the two strikes together encode aspects of syllable structure as a whole. **Figure 1** shows one such example, where both *mîɛ̰* “snake” and *jṵ* “long” are played with a flam (visible in the spectrogram) which are treated as single interbeat durations. The final word of the phrase, here *ĵio* “saw”, is marked in the TextGrid but its measurement is not included in the database since it is not followed by another beat and as such, no explicit endpoint exists for measuring its duration. For audio recordings of this and all other examples in the paper, see the **Supplementary Materials**.

Position in the line was also coded by numbering each syllable in a line and dividing that number by the total number of syllables in the line. Some syllables were excluded from the final analysis since their original line membership had been lost and hence their position could not be calculated, and all line-final syllables were removed since their duration could not be measured. This left 1053 syllables for the durational analysis in *Gradience*.

This paper focuses on the encoding of segmental aspects of Seenku rather than tone. Briefly, tone is encoded in the notes of the balafon, with the four notes of the scale excluding the “fetish balafon key” (♭3) corresponding to the four phonemic tone levels; the highest tone, S, is generally pegged to the center of the musical mode and the remaining three tone levels—H, L, X—are played subsequently on each lower note. For more details of tonal encoding, see McPherson (2018).

Consonant and vowel qualities are not encoded in the surrogate language, i.e. there is no way to encode the difference between /a/ and /i/ or between /m/ and /t/. As such, segmental encoding consists of the encoding of the different aspects of syllable structure described in *Syllable and Word Structure*, namely vowel length, diphthongs, sesquisyllabicity, and nasal codas. When we look at the data corpus, we find three patterns of encoding: categorical encoding, variable encoding, and gradient encoding. In the subsections below, I will describe each pattern in turn and the elements of syllable structure for which they are employed.

Categoricity

The conscious rules of balafon segmental encoding involve a single formal distinction: The word can be played with a single strike, or with a flam (two strikes in quick succession, with the first significantly shorter than the second, like a grace note). This choice is made categorically for long vowels and for sesquisyllabic words, both of which are exceptionlessly played with a flam (resulting in potential ambiguity). For instance, **Figure 2** shows a short phrase that could encode either *bɛ̏ɛ kərɛ̂* 'male pig/boar' or *səgə̏ bâ* 'male sheep/ram' (among other things). Both words are played with a flam; the first involves two strikes on the same note, corresponding to the level-toned long vowel or sesquisyllable, while in the second, the strikes are on two different notes to encode the HX contour tone. In the case of *bâ,* this brings up the important point that if a word has a contour tone, it will always be encoded as a flam, regardless of syllable structure.

It would be considered ungrammatical on the balafon to play the first word of this phrase with just a single strike, and the corpus contains no instances in which a sesquisyllable or long vowel is played with a single strike.

Note that this is a binary opposition: Long vowels result in two strikes, and so do sesquisyllables, but a syllable type with both complexities like CəCVː still falls into the two-strike category. Likewise, the presence of any one complexity (like sesquisyllabicity) will put the syllable into the two-strike category. That is, even though CəCV has a short vowel, it is encoded with two strikes due to its sesquisyllabic nature.

Variability

While long vowels and sesquisyllabicity are exceptionlessly encoded with a flam, other elements of syllable structure are subject to variation in their encoding. Specifically, diphthongs and nasal codas are variably encoded either with one strike or two; I present the data below and suggest reasons why these elements may behave differently from those we saw in the last section.

Diphthongs

As stated in *Syllable and Word Structure*, the syllable nucleus in Seenku may be occupied by either a monophthong or a diphthong. All diphthongs are either rising sonority (higher to lower vowels) or back to front at the same vowel height. The first vocalic (**V̲**V) element is non-moraic, acting like a glide, leaving the second (V**V̲**) to carry contrastive vowel length. In addition,

	S (8)				
	Sk (6)				
	T (5)				
	Bg (3)				
	J (b3)				
➢	B (1)				
	Words1	bɛ̏ɛ		kərɛ̂	
	Words2	səgə̏		bâ	

FIGURE 2 | Balafon encoding of "male pig" or "male sheep". Arrow by B(âa-ɲä) indicates the center of the mode.

diphthongs may be either oral or nasal, though vowel nasality plays no role in the balafon surrogate language.

In total, the corpus contains just 66 instances of diphthongs. In order to isolate the effect of the diphthong on surrogate encoding, we must restrict our analysis to only level-toned syllables of the shape CVV. Contour tones, long vowels, or sesquisyllables will all result in categorical flam encoding independently of the diphthong. For the time being, I will also exclude diphthongs carrying a nasal coda, which as we will see below also can trigger a flam. This leaves 41 CVV syllables in the corpus. Of these, 9 are played with a flam and 32 with a single strike (22% flam). However, while we have 41 individual tokens, these represent just 18 distinct words, with many words showing multiple repetitions. Out of these 18 unique words, only 3 are responsible for the flams. These three words are shown in (3):

(3) kùa 'farm/cultivate (tr.)'
sòe 'horses'
sùɛ 'three'

None showed free variation; in other words, they were uniformly encoded with flams. Thus, it is possible that what we find is not free variation in diphthong encoding but rather lexical variation, with certain words encoded with a single strike and others with a flam. However, in this data set, multiple repetitions were given by the musician back-to-back, and so there may be a priming effect, such that once a decision is made to play a word with a flam, the same encoding will be used in subsequent repetitions. These phrases would need to be elicited on another day, and also ideally with different musicians to test whether there is any inter-player variability (which I reiterate is unattested for either long vowels or sesquisyllables).

Why should diphthongs be subject to variable encoding when long vowels and sesquisyllables are not? I hypothesize that there is competition between two contrasts that musicians may try to maintain: the contrast between a monophthong and a diphthong, but perhaps more importantly, the contrast between short and long diphthongs. From a vowel length standpoint, CVV is a simplex syllable (while CVVː is complex); but at the same time, CVV is still more complex than CV. Overall, musicians appear to err on the side of treating CVV as simplex. But as the examples in (3) show, this is not always the case.

We can ask whether there is anything special about the words in (3) that may lead them to be played with a flam rather than a single strike. For *kùa* 'farm (tr.)' (underlyingly *kȕa,* but with tone raising in its phrasal context), it isn't clear. It is possible that there is interference from the intransitive *kȕaa,* which would be played with a flam, though it seems unlikely to me that musicians would confuse this grammatical vowel length contrast. For *sòe* 'horses' and *sùɛ* 'three', though, a functional explanation may be available.

First, *sòe* 'horses' is the plural form of *sȍ* 'horse'. In isolation, the two are distinguished both by the diphthong and by tone raising in the plural (McPherson 2017). In the phrasal context elicited on the balafon, however, a process of tone sandhi raises the X-tone of *sȍ* 'horse' to the same L-tone as the plural. These two phrases are shown in (4):

(4) a. mó nǎ sò sa̰n
1SG PROSP horse buy.IRR
'I will buy a horse'
b. mó nǎ sòe sa̰n
1SG PROSP horse.PL buy.IRR
'I will buy horses'

When I elicited the phrase 'I will buy horses', the consultant first offered the singular phrase in (4a), in which *sò* 'horse' was represented with a single strike, as expected due to its level-toned CV nature. In playing the plural form immediately afterward, the musician chose to encode *sòe* as a flam, likely to disambiguate the two phrases. It is an open question whether a musician would think to disambiguate in this way if the singular and plural were not immediately juxtaposed

A similar explanation may be at play for *sùɛ* 'three' (lexically *sȕɛ,* but once again raised to L tone in the phrasal context). Played on a single strike, *sùɛ* would be confusable with *nɔ̀* 'five' (lexically L-toned), a relatively common numeral played on the balafon thanks to the monetary system, and as such, the musician may have chosen to explicitly encode the diphthong with a flam. Nevertheless, this flam encoding would then make *sùɛ* 'three' confusable with other single digit numerals like *nàa* 'four' (also lexically X-toned *nȁa*). My impression is that *nȁa* is less common on the balafon and hence would be less salient as a potential confound to a musician. More data are required to test this functional hypothesis, such as the numeral 'three' in a context without tone raising so that *sȕɛ* and *nɔ̀* would be tonally distinct.

Musicians are keenly aware of ambiguity in the signal, which they have reported to me in other instances. Curiously, sometimes they do not offer disambiguation by flam. For instance, one consultant told me that saying *bȉ̃ fűɛ* 'twenty goats' on the balafon would be unlikely, because it would be confused for *bȉ̃ fȉ̃* 'two goats'; *fűɛ* 'twenty' was not played with a flam. Then again, it may be unlikely to speak of twenty goats in general (who would offer to slaughter so many?), so less thought was put into how to make the signal clear, whereas 'three' vs. 'five' is a more realistic difference to need to make.

Nasal Codas

The other variable aspect of syllable structure is the nasal coda, which will be transcribed with <n>, though its realization and place of articulation are variable and non-contrastive. In the spoken language, this is a weak phonological element whose realization depends heavily on context (segmental, phrasal) but which is also subject to free variation. In phrase-final position, it is variably realized as late nasalization of the preceding vowel, with or without weak lingual articulation (palatal for front vowels, velar or uvular for back vowels, but not achieving closure), or it goes unrealized. If followed by a sonorant /l/ or /w/, it will be realized on the sonorant, yielding either nasal stops [n] and [m] or nasalized sonorants [l̃] and [w̃]. Before nasals, it goes unrealized. Only before obstruents is it variably realized as a nasal stop. For more details, see McPherson (2020a, 2020b 2020c).

Given the variable nature in the spoken language, it is unsurprising that it should also receive variable treatment in the balafon surrogate language. The corpus contains 109 instances of nasal codas, but once again we must focus only on those cases where another syllable element cannot be responsible for its balafon encoding. Of the 109 nasal codas, 70 are found on level-toned CVN syllables, without the influence of long vowels, contour tones, sesquisyllables, or diphthongs. 7 out of 70, or 10%, are played with flams. Another word, *wɛ̋n* 'money', varies in its spoken pronunciation between *wɛ̋n* and *wɛ̋ɛ(n)*; it is a very common word on the balafon, since asking for money is one of the surrogate language's main functions, and it is uniformly played with a flam. Since it cannot be determined whether this is the result of the nasal coda, vowel length, or both, I exclude it from consideration here.

Unlike in the case of diphthongs, we do see some free variation in the corpus for nasal codas. For instance, the word *dân* '10,000' appears four times in the same recording, three of which are played with flams and one of which is not. Similarly, the word 'sauce', which appears as both *nàn* and *nân* depending on the phrasal context, appears eight times, twice with a flam and otherwise as a single strike. These two flam realizations are found in the phrase in (5), but here too we see free variation, with one repetition of the phrase showing just a single strike for *nàn*:

(5) gɔ̀ɔ́ǹ nàn kɔ̀rɔ̀ sḭ̌ mó tɛ̏
sorrel sauce desire be 1SG GEN
'I want sorrel sauce'

In contrast, the phrasal variants in (6a) and (6b) never see the word *nán* played with a flam:

(6) a. mó nǎ ȁ nán lɛ̂ kʊ́ nɔ̏
1SG PROSP DEF sauce DEM D.DEF eat.IRR
'I will eat that (aforementioned) sauce'
b. mó nǎ ȁ nán lɛ̂ ȁ nɔ̏
1SG PROSP DEF sauce DEM DEF eat.IRR
'I will eat that sauce'

While there are just six tokens of these phrases (three each) and thus we should take any patterns with some degree of caution, the fact that these cases are never played with a flam while those in (5) are may find a linguistic explanation: In (5), the nasal coda on *nàn* appears before an obstruent and is pronounced as a velar nasal [nàŋ kɔ̀rɔ̀], whereas in (6), the nasal appears before a sonorant and is realized entirely on this following segment [ná nɛ̂]. In other words, the surface realization in (6) renders *nán* a simple CV.

It is by no means the case that every coda nasal before an obstruent will be realized with a flam—in fact, the majority are not; out of nine repetitions of the phrase *ȁ sḭ̌ sȁn kȁ* 's/he is in God's hands', none were produced with a flam. But at the same time, all seven instances of nasal coda flams are found before an obstruent, suggesting that surface realization does have a role to play.

Thinking about why nasal codas should be variably rather than categorically encoded, the first explanation relates to its weak representation. Given its variable pronunciation even in the spoken language, it may be that nasal codas are not fully activated phonological representations, in the sense of

Gradient Symbolic Representation (Rosen 2016; Smolensky and Goldrick 2016; Zimmermann 2019). As such, when musicians come to represent speech on the balafon, they may not be activated enough to be consciously registered as contributing to syllabic complexity. Another possible explanation is that in the case of nasal codas, we have a mismatch between the timing of syllable complexity (after the nucleus) and the timing of surrogate syllable complexity (at the beginning of the word, with a flam). This stands in contrast to diphthongs and sesquisyllables, both of which find their "weaker" or "faster" element at the left edge. A larger corpus will reveal whether the higher rates of flam encoding in contexts where the nasal coda is realized as a true coda are statistically significant.

We could also ask whether there are any musical or phrasal effects on variation, such as the length of a line or the presence of other flams in the vicinity of the target syllable. For instance, we might predict that in longer phrases, musicians may counter this complexity by using fewer flams (cf. Fenk-Oczlon et al., 2009). As far as the current data show, there appears to be no effect of phrase length; both flam and single strike realizations are found in phrases of varying lengths and in both the first and second half of the phrase. However, as with the linguistic explanations above, the current dataset is too small to draw any firm conclusions. With more data, we could also test whether the presence of other flams in the immediate vicinity either encourages or inhibits a flam encoding. I leave these questions to future research.

Before leaving the topic of variable encoding, I want to briefly point out that it may be the case that words that contain two variably encoded elements—that is, a diphthong followed by a nasal coda—are more likely to be encoded as a flam. The dataset only contains one such word, *sóen* 'one', but in 8 out of 9 tokens, it is encoded with two strikes. If we think about variable flam encoding as a probability, then it makes sense that two factors that result in such an encoding would increase its odds. Since we are dealing with just a single type, though, itself a high frequency word in the natural surrogate language, we should not draw any strong conclusions about the interaction of these two probabilities. Like the word *w̃ɛ̃n* 'money' discussed above, musicians may have a lexically specific representation of this word as a flam regardless of the productive rules of surrogate encoding. Further cases of level-toned CVVN words must be sought to test the rest of flam encoding with these two variable elements of syllable structure.

Gradience

Thus far, all encoding we have discussed has been what could be considered conscious rules of the balafon surrogate language, a reduction of contrastive elements of syllable structure to a binary choice between a single strike and a flam on the balafon. Musicians even report rampant neutralization due to these rules of encoding, such as the example shown in **Figure 2**.

Nevertheless, while the notes of the instrument are fixed, and the number of strikes (single vs. flam) is likewise a categorical feature, duration is gradient. Specifically, the timing between strikes of the balafon (interbeat durations, Seifart et al., 2018) is not isochronous and varies depending upon the word that is encoded. These gradient distinctions appear to be subconscious, since at the level of conscious encoding musicians report neutralizations where duration measurements suggest a distinction.

As noted at the beginning of **Section 4**, each syllable in the corpus was coded for a variety of binary factors, in addition to its phrasal position and its duration in milliseconds. To investigate the influences on gradient duration, a linear mixed effects regression analysis was carried out in R using the lme4 package (Bates et al., 2015), with duration in milliseconds as the dependent variable. Independent variables included the position of the word in a phrase, whether the nucleus of the syllable was a diphthong or a monophthong, whether the nucleus vowel was phonemically short or long, whether the syllable was sesquisyllabic or not, whether the syllable had a coda or not, whether the syllable's tone was a contour or level, and finally whether the syllable was played on the balafon with a flam or a single strike. The interaction between phonemic vowel length and the presence of a coda was also included in the model. Random intercepts for player and word were included. The package lmerTest (Kuznetsova et al., 2017) was used to test the statistical significance of the results.

First, there is a significant interaction between phonemic vowel length and codas ($\beta = -52.3 \pm 21.4$, $t(930) = -2.4$, $p = 0.02$). When the syllable has no coda, the difference in duration between long and short vowels is 77 ms (Long: 302±54 ms, Short: 225±69 ms). However, when the syllable does have a coda, the difference is much smaller, at only 12 ms (Long: 284±38 ms, Short: 272±77 ms). This result captures the fact that segments tend to be compressed and shorter in more complex (and especially closed) syllables (e.g. Fowler 1983; Maddieson 1985; Clements and Hertz 1996). As main effects, both Vowel length and Coda were significant ($\beta = 25.0 \pm 7.9$, $t(935) = 3.1$, $p < 0.01$ for Vowel length and $\beta = 39.5 \pm 6.9$, $t(940) = 5.7$, $p < 0.001$ for Coda).

The model found a main effect of position ($\beta = 100.2 \pm 11.3$, $t(430) = 8.8$, $p < 0.001$); syllables closer to the end of the phrase have longer durations. There is also a main effect of diphthong ($\beta = 17.9 \pm 7.8$, $t(937) = 2.3$, $p = 0.02$); syllables with a diphthong nucleus are significantly longer than those with a monophthong. Sesquisyllabicity was also significant ($\beta = 17.1 \pm 8.6$, $t(939) = 2.0$, $p = 0.05$); sesquisyllabic syllables had significantly longer durations than simple C(V)V syllables. Finally, syllables played with a flam were significantly longer than those played with a single strike ($\beta = 52.2 \pm 8.3$, $t(939) = 6.3$, $p < 0.001$). The factor of contour tone was not significant ($p = 0.38$).

Figure 3 plots duration of syllables in the surrogate language by position. This gradient effect differs from the others (vowel length, sesquisyllabicity, etc.) in that it does not arise from a phonemic contrast in the underlying words being encoded. Instead, we may attribute this phrase-final lengthening to one (or both) of two sources: 1) the phonetic realization of spoken phrases, which would be subject to cross-linguistic patterns of phrase-final lengthening (Klatt 1975; Scott 1982, *i.a.*), or 2) the same effect of phrase-final lengthening attested in musical performance and perception (Todd 1985; Palmer 1989, *i.a.*). In other words, gradient durational differences in the balafon surrogate language are controlled by more than just contrastive categories; we may also see the effects of low-level phonetic differences, lying beneath the level of a musician's

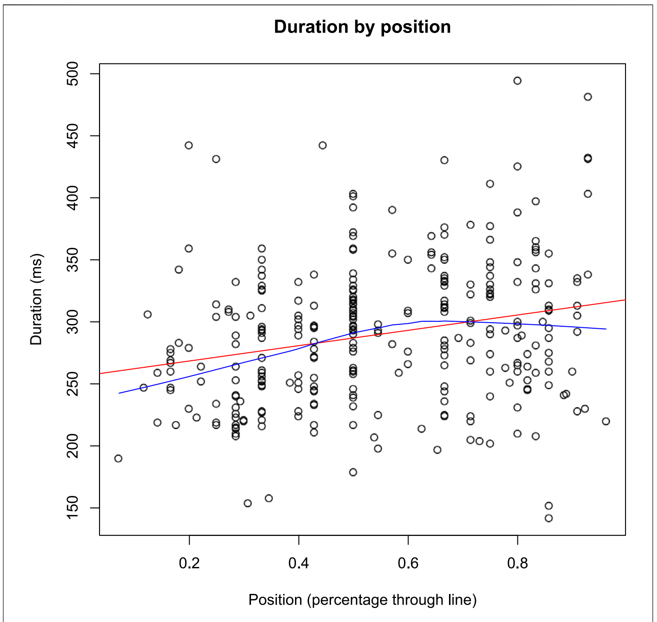


FIGURE 3 | Duration plotted by phrasal position (syllable #/total number of syllables in the line).

consciousness. I return to this point in the next section and offer some thoughts on how to conceptualize these different influences on musical surrogate speech.

DISCUSSION

In this paper, I have shown that encoding of syllable structure in the Sambla balafon surrogate language is more complex than first meets the eye. In terms of conscious encoding, i.e. the explicit grammatical rules of the surrogate language, there is just a binary contrast of a single strike vs. a flam. The flam can be employed categorically—for contour tones (not discussed here), vowel length, or sesquisyllabicity—or it can be employed variably, with diphthongs and coda nasals vacillating between a simple and complex syllabic treatment. Beneath these conscious encodings, we also find a range of subconscious, gradient durational effects, ranging from gradient encoding of phonemic contrasts like vowel length to phonetic or musical phrasing effects. Musicians appear to be unaware of these durational contrasts, at least consciously. Future work could investigate whether they play any role in comprehension, e.g. in a forced choice task asking musicians to choose between two interpretations of a supposedly ambiguous balafon phrase.

How can we account for these different levels of surrogate speech encoding? I suggest that the encoding of surrogate speech passes through a "phonetic filter", formalized in the schema shown in **Figure 4**.[1]

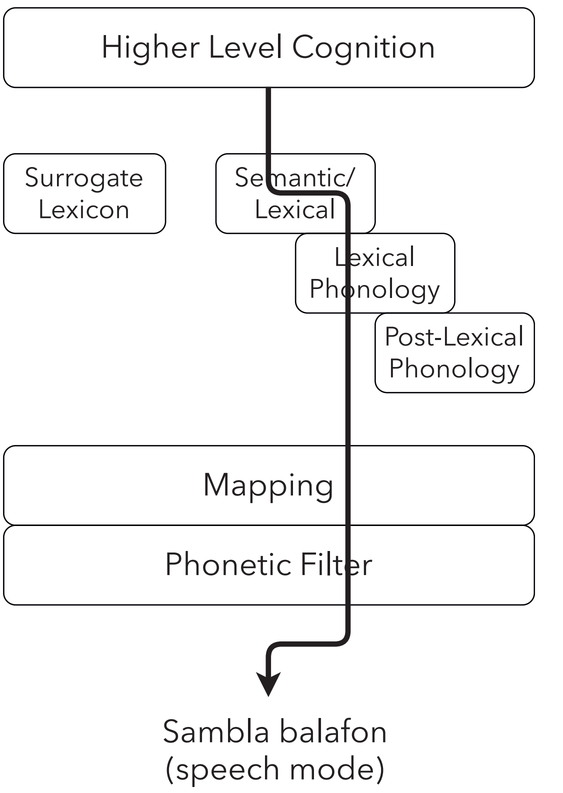


FIGURE 4 | Schema of Sambla balafon surrogate language encoding.

The idea of what a musician wants to play first comes from "higher level cognition", and its linguistic content is selected from the "semantic/lexical" component.[2] These words are then passed through the lexical phonology, where they receive their phonological form. In the Sambla balafon surrogate language, it is the underlying lexical and morphological level of tone that is encoded, not surface tone resulting from post-lexical tone rules (McPherson 2018), and so the encoding does not pass through post-lexical phonology. However, given the possible influence of the surface realization of nasal codas (for instance), the model may require more nuance, separating tonal from segmental phonology; in other surrogate traditions where the post-lexical output is uniformly encoded, the arrow would pass through that component. The output of the phonological component(s) is then passed through the "mapping" component, containing the conscious rules of surrogate encoding. It is here that the notes of the balafon and the choice between a single strike and a flam are selected by the musician. Finally, on the way to being played, the surrogate encoding passes through the phonetic filter where subconscious phonetic elements of speech can influence the surrogate realization, depending upon instrumental constraints. On a variably pitched instrument, this could be

[1]Schema credit: Lucas James.

[2]In a lexical ideogram (Stern 1957) surrogate system, with arbitrary musical signs standing in for words or phrases, there would be a separate step of going to the "surrogate lexicon".

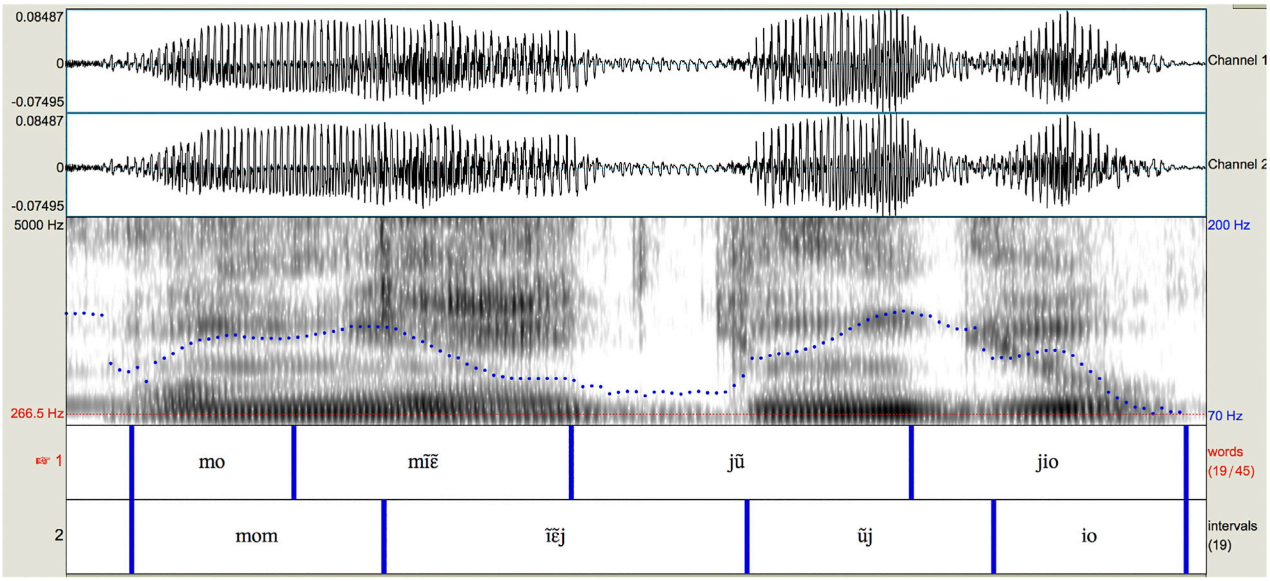


FIGURE 5 | Syllable **(top tier)** and interval **(bottom tier)** timing of spoken phrase *mó mḭ̂ɛ̰̃ jṵ̌ ȷ̂io* "I saw a long snake".

subtle phonetic aspects of tonal realization (see Akinbo 2019 for one such possibility on the Yorùbá *dùndún*). In the case of the balafon, the influence can be seen in micro-timings.

I propose that these phonetic effects exert influence over surrogate speech as musicians think about the words as they play, which controls the timing of their strokes. This suggests that the timing of surrogate speech should largely mirror the timing of spoken language, a finding reported for the Bora *manguaré* slit log drumming tradition (Seifart et al., 2018). In the case of Bora, the authors determined that beat timing tracked vowel-to-vowel intervals rather than syllables.

To test this explanation behind the phonetic filter for the Sambla balafon language, I ran a pilot study comparing the relative timing of words in natural speech and surrogate speech, spoken and played by the same individual. For natural speech, measurements were taken for both syllables and vowel-to-vowel intervals, as shown in **Figure 5**, the spoken phrase corresponding to the balafon phrase in **Figure 1**.

Final syllables were excluded due to the inability to measure an interbeat duration for their balafon correspondents. To time-normalize the balafon and spoken phrases, the total duration of all beats/syllables, excluding the final, was calculated, then each individual duration measure was divided by this total to indicate its individual percentage of the overall phrase. 125 words/syllables were compared, each with three corresponding measures: 1. Interbeat balafon duration; 2. Syllable duration; 3. Vowel-to-vowel interval duration. The deviation between the interbeat duration and each of the syllable and interval durations was calculated to determine which measurement provided a closer match to the balafon timing. The average absolute deviation for syllables was over twice as high as that for intervals, 0.055 (5.5% off) vs. 0.0268 (2.68% off). Visual inspection of plotted phrases,

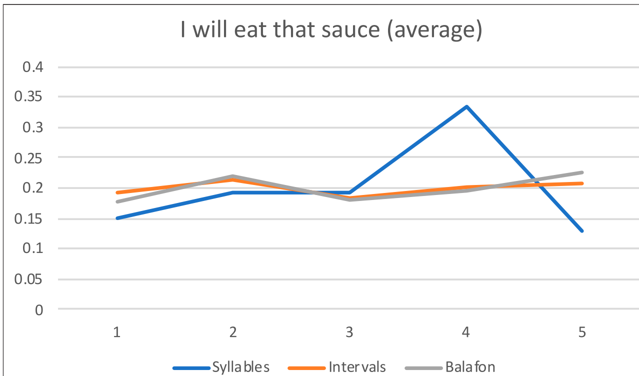


FIGURE 6 | Relative timing of first five words of *mó nǎ kʊ́ nǎ̀n bɛ́ nɔ̀̃* "I will eat that sauce", averaged across three spoken and played iterations.

such as that shown in **Figure 6**, corroborates these results, with the interval data much more closely tracking the relative balafon interbeat durations.

In all likelihood, vowel-to-vowel intervals are simply an approximation of the perceptual center, or p-center, of the syllable, which is aligned closely with the left edge of the rime but which may be pulled leftwards into the onset by consonant clusters or otherwise long onsets (Morton et al., 1976; Ryan 2014). We see a suggestion of this effect in the pilot data: Sesquisyllables and the consonant /f/ are both longer onset events in Seenku, and all of their instances in the data show an interval duration shorter than the balafon duration; if it were the p-center that controlled the timing, the left edge of the interval would be anticipated, rendering the whole measurement longer and bringing it closer to that of the balafon. Seifart et al. (2018) make a similar nod to the p-center but do not explore the hypothesis further.

What is interesting about the phonetic filter on the Sambla balafon is that it comes solely from mental speech—the musician is not speaking while playing, and yet precise relative timings of words in spoken language appear to control the timing of instrumental rhythm. This suggests a fair amount of phonetic detail in either the lexicon, the inner voice, or both. Even though the oral articulators are not being used, their influence is felt on this musical modality.

Most studies of the p-center focus on speaking syllables aloud to an isochronous beat, in other words, a beat controlling the timing of spoken syllables. The study of musical surrogate languages like the Sambla balafon offer the opportunity to flip the script and study the effect of natural non-isochronous syllable timing on a beat. In future work, it would be interesting to compare interbeat duration on the balafon when the musician is speaking aloud vs. when they are simply thinking the words in their head. This could help us tease apart exactly how much phonetic detail is present in the inner voice.

CONCLUSION

At first glance, the Sambla balafon surrogate language is like most African surrogate languages, based largely on tone. The rich syllable structure of the spoken language, Seenku, is ostensibly collapsed to a binary distinction between "simple" and "complex". But a deeper look reveals that musicians are sensitive to a host of phonological and phonetic factors that inform their playing. Long vowels and sesquisyllables are both invariably treated as complex syllables, while diphthongs and nasal codas vary in their encoding. The variation seen with diphthongs seems to reflect to the tension between the complex nucleus and the pressure to maintain a short vs. long contrast even within diphthongs. Variation for nasal codas, on the other hand, may relate to their weak phonological representation and variation in surface form in the spoken language. At a subconscious level, the mental timing of speech influences the timing of words on the balafon, suggesting gradient distinctions between words that musicians report to be neutralized.

These details of encoding demonstrate the various ways in which musical surrogate languages act as a unique window onto phonological and phonetic structure of the spoken language. Differing treatment of phonemic contrasts can provide a source of evidence for differing phonological representation or for competition in the phonological component. The presence of phonetic detail in the surrogate signal raises questions about how it gets there: Could this provide evidence for an exemplar model of phonology, with mental storage of detailed tokens of speech (Pierrehumbert 2002; Johnson 2006)? Does thinking of the phrases activate the motor neurons for articulation (Tian and Poeppel 2010; Pickering and Garrod 2013), without reaching the threshold of producing the gestures but still replicating the timing patterns of external speech?

Further detailed studies of musical surrogate languages will allow us to better understand the deep connections between the human language faculty and these unique modalities of communication.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Center for the Protection of Human Subjects, Dartmouth College. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This material is based on work supported by the National Science Foundation-National Endowment for the Humanities Documenting Endangered Languages Program under National Science Foundation Award No. BCS-1664335.

ACKNOWLEDGMENTS

I would first like thank Mamadou Diabate and the other members of the Diabate family for sharing their tradition and expertise with me. I would like to thank Kevin Ryan and Lucas James for helpful conversations in the preparation and revision of this work. Finally, I am deeply grateful to Rolando Coto Solano for his help with statistical analysis. Any remaining errors are my own.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.652635/full#supplementary-material

REFERENCES

Akinbo, S. (2019). Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis. *Linguistique et Langues Africaines* 1, 11.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *J. Stat. Soft.* 67, 1. doi:10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (2017). Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.28, Available at: http://www.praat.org/ (Accessed January 2, 2017).

Clements, G. N., and Hertz, S. R. (1996). An Integrated Model of Phonetic Representation in Grammar. *Working Pap. Cornell Phonetics Lab.* 11, 34–116.

Corley, M., Brocklehurst, P. H., and Moat, H. S. (2011). Error Biases in Inner and Overt Speech: Evidence from Tongue Twisters. *J. Exp. Psychol. Learn. Mem. Cogn.* 37 (1), 162–175. doi:10.1037/a0021321

Fenk-Oczlon, G., Fenk, A., and August, F. (2009). Some Parallels between Language and Music from a Cognitive and Evolutionary Perspective. *Musicae Scientiae*. 13 (2 Suppl. l), 201–226. doi:10.1177/1029864909013002101

Fowler, C. A. (1983). Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet. *J. Exp. Psychol. Gen.* 112 (3), 386–412. doi:10.1037/0096-3445.112.3.386

Johnson, K. (2006). Resonance in an Exemplar-Based Lexicon: The Emergence of Social Identity and Phonology. *J. Phonetics* 34 (4), 485–499. doi:10.1016/j.wocn.2005.08.004

Klatt, D. H. (1975). Vowel Lengthening Is Syntactically Determined in a Connected Discourse. *J. Phonetics* 3 (3), 129–140. doi:10.1016/s0095-4470(19)31360-9

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* 82, 1–26. doi:10.18637/jss.v082.i13

Maddieson, I. (1985). Phonetic Cues to Syllabification. in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Editor Fromkin (New York, NY: Academic Press), 203–221. doi:10.18637/jss.v082.i13

Martin, S., Iturrate, I., Millán, J. d. R., Knight, R. T., and Pasley, B. N. (2018). Decoding Inner Speech Using Electrocorticography: Progress and Challenges toward a Speech Prosthesis. *Front. Neurosci.* 12. 422. doi:10.3389/fnins.2018.00422

Matisoff, J. (1990). Bulging Monosyllables: Areal Tendencies in Southeast Asian Diachrony. *Proc. 16th Annu. Meet. Berkeley Linguistics Soc.*, 543–559. doi:10.3765/bls.v16i0.1680

McPherson, L. (2017). Multiple Feature Affixation in Seenku Plural Formation. *Morphology* 27 (2), 217–252.

McPherson, L. (2018). The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. *Anthropological Linguistics* 60 (3), 255–294.

McPherson, L. (2020a). A Grammar of Seenku. *MGL 83*. Berlin/Boston: De Gruyter Mouton

McPherson, L. (2020b). Illustration of Seenku. *J. Int. Phonetic Assoc.* 50 (2), 220–239.

McPherson, L. (2020c). On Latent Nasals in Samogo. *Mandenkan* 63, 68–88.

McPherson, L., and James, L. (2020). "Artistic Adaptation of Seenku Tone: Musical Surrogates vs. Vocal Music,"in. Selected Proceedings of the 50th Annual Conference on African Linguistics. Editor Akinlabi (Berlin, Germany: Language Science Press).

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual Centers (P-Centers). *Psychol. Rev.* 83, 405–408.

Oppenheim, G. M. (2013). Inner Speech as a Forward Model. *Behav. Brain Sci.* 36 (4), 369–370. doi:10.1017/S0140525X12002798

Palmer, C. (1989). Mapping Musical Thought to Musical Performance. *J. Exp. Psychol. Hum. perception Perform.* 15 (2), 331. doi:10.1037//0096-1523.15.2.331

Pickering, M. J., and Garrod., S. (2013). An Integrated Theory of Language Production and Comprehension. *Behav. Brain Sci.* 36, 329–392. doi:10.1017/S0140525X12001495

Pierrehumbert, J. (2002). *Word-specific Phonetics*. Laboratory Phonology 7

Pittayaporn, P. (2015). "Typologizing Sesquisyllabicity," in *Languages of Mainland Southeast Asia: The State of the Art*. Editors Enfield and Comrie (Berlin/Boston: De Gruyter Mouton), 500–528.

Rosen, E. (2016). "Predicting the Unpredictable: Capturing the Apparent Semi-regularity of Rendaku Voicing in Japanese through Harmonic Grammar," in. Proceedings of BLS. Editors Clem (Berkeley, CA: Berkeley Linguistic Society), 42, 235–249.

Ryan, K. (2014). Onsets Contribute to Syllable Weight: Statistical Evidence from Stress and Meter. *Language* 90 (2), 309–341.

Scott, D. (1982). Duration as a Cue to the Perception of a Phrase Boundary. *J. ASA* 71, 996. doi:10.1121/1.387581

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing Language to Rhythm: Amazonian Bora Drummed Language Exploits Speech Rhythm for Long-Distance Communication. *R. Soc. Open Sci.* 5, 170–354. doi:10.1098/rsos.170354

Smolensky, P., and Goldrick, M. (2016). *Gradient Symbolic Representations in Grammar: The Case of French Liaison, Ms*. Johns Hopkins Univeristy and Northwestern University. Available at: http://roa.rutgers.edu/content/article/files/1552_smolensky_1.pdf (Accessed September 15, 2019).

Stern, T. (1957). Drum and Whistle "Languages": an Analysis of Speech Surrogates. *Am. Anthropologist* 59, 487–506.

Strand, J. (2009). The Sambla Xylophone: Tradition And Identity in Burkina Faso. PhD Dissertation. Middletown (CN): Wesleyan University.

Tian, X., and Poeppel, D. (2010). Mental Imagery of Speech and Movement Implicates the Dynamics of Internal Forward Models. *Front. Psychol.* 1, 166. doi:10.3389/fpsyg.2010.00166

Todd, N. (1985). A Model of Expressive Timing in Tonal Music. *Music Perception*. 3 (1), 33–57.

Zemp, H., and Soro, S. (2010). Talking Balafons. *Afr. Music.* 8 (4), 7–24.

Zimmermann, E. (2019). Gradient Symbolic Representations and the Typology of Ghost Segments. Proceedings of the Annual Meeting on Phonology, Vol. 7.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 20 May 2021
doi: 10.3389/fpsyg.2021.652673

Perception of Nigerian Dùndún Talking Drum Performances as Speech-Like vs. Music-Like: The Role of Familiarity and Acoustic Cues

Cecilia Durojaye[1,2†], Lauren Fink[1,3†], Tina Roeske[1], Melanie Wald-Fuhrmann[1,3] and Pauline Larrouy-Maestri[3,4]*

[1] Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany, [2] Department of Psychology, Arizona State University, Tempe, AZ, United States, [3] Max Planck-NYU, Center for Language, Music, and Emotion, Frankfurt am Main, Germany, [4] Neuroscience Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

OPEN ACCESS

Edited by:
Laura McPherson,
Dartmouth College, United States

Reviewed by:
Elizabeth Margulis,
Princeton University, United States
Aaron Carter-Enyi,
Morehouse College, United States

***Correspondence:**
Cecilia Durojaye
cdurojay@asu.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: 12 January 2021
Accepted: 21 April 2021
Published: 20 May 2021

Citation:
Durojaye C, Fink L, Roeske T, Wald-Fuhrmann M and Larrouy-Maestri P (2021) Perception of Nigerian Dùndún Talking Drum Performances as Speech-Like vs. Music-Like: The Role of Familiarity and Acoustic Cues. Front. Psychol. 12:652673. doi: 10.3389/fpsyg.2021.652673

It seems trivial to identify sound sequences as music or speech, particularly when the sequences come from different sound sources, such as an orchestra and a human voice. Can we also easily distinguish these categories when the sequence comes from the same sound source? On the basis of which acoustic features? We investigated these questions by examining listeners' classification of sound sequences performed by an instrument intertwining both speech and music: the dùndún talking drum. The dùndún is commonly used in south-west Nigeria as a musical instrument but is also perfectly fit for linguistic usage in what has been described as speech surrogates in Africa. One hundred seven participants from diverse geographical locations (15 different mother tongues represented) took part in an online experiment. Fifty-one participants reported being familiar with the dùndún talking drum, 55% of those being speakers of Yorùbá. During the experiment, participants listened to 30 dùndún samples of about 7s long, performed either as music or Yorùbá speech surrogate (n = 15 each) by a professional musician, and were asked to classify each sample as music or speech-like. The classification task revealed the ability of the listeners to identify the samples as intended by the performer, particularly when they were familiar with the dùndún, though even unfamiliar participants performed above chance. A logistic regression predicting participants' classification of the samples from several acoustic features confirmed the perceptual relevance of intensity, pitch, timbre, and timing measures and their interaction with listener familiarity. In all, this study provides empirical evidence supporting the discriminating role of acoustic features and the modulatory role of familiarity in teasing apart speech and music.

Keywords: speech surrogacy, Yorùbá, categorization, amplitude modulation spectrum, pitch, intensity, timbre, timing

INTRODUCTION

When we turn on the radio it seems trivial to determine whether what we are hearing is music or someone speaking. Sound sequences can generally be described in terms of pitch, timbre, and timing (e.g., Kraus et al., 2009); acoustical differences have been shown between sequences associated with music and language domains. For instance, speech typically comprises many gliding tones and more variation in pitch trajectory than (Western) music (Patel, 2008), with a temporal modulation spectrum peaking around 5 Hz, approximating the syllable rate (Ding et al., 2017). On the other hand, music is characterized by discrete pitches sustained for longer durations (Zatorre and Baum, 2012), and a temporal modulation spectrum peaking around 2 Hz, approximating the average beat rate (Ding et al., 2017). However, note that such studies often examine material that has different sound sources, such as the human voice vs. musical instruments, which might enhance the distinctive characteristics of the sequences associated with each of the two categories.

When coming from the same source (e.g., the vocal instrument), music and language categories can show a certain overlap. For instance, certain types of speech are considered more musical than others (e.g., child-directed speech, rhymes, poetry), while certain types of vocal music are considered speech-like (e.g., rap). To better understand the ambiguity of vocal stimuli, Merrill and Larrouy-Maestri (2017) presented several versions of Arnold Schoenberg's Pierrot lunaire—a piece notable for its use of sprechstimme or "speech-song" (Stadlen, 1981)—to vocal experts and found a large variety in the description of the material, from very spoken-like to very sung-like. Interestingly, the same exact material can be interpreted as either song or speech. A phenomenon called the speech-to-song illusion has been reported by Deutsch et al. (2008, 2011): A spoken sentence can be perceived as sung after several repetitions. In other words, the same acoustic contents can be perceived as belonging to two distinct categories, which makes it difficult to identify clear boundaries between speech and music, even in a culture in which the idea of contrasting categories is widely accepted (see Brown, 2000 for a gradual view of music-language). Further, it has been shown that categorization, in the context of the speech-to-song illusion task, changes when the material is difficult to pronounce (Margulis et al., 2015) or to understand (Jaisin et al., 2016), which supports the role of listeners' familiarity or prior knowledge in stimulus categorization (Vanden Bosch der Nederlanden et al., 2015) and suggests a potential downside of using vocal stimuli, which are highly familiar to listeners.

In the present study, we focus on the perceptual categorization of non-vocal material into speech-like vs. music-like, as well as the role of familiarity in shaping these categories. Familiarity/culture effects on the perception of speech and music are well-described (Palmer and Krumhansl, 1987; Morrison and Demorest, 2009; Perrachione et al., 2011; Bregman and Creel, 2014; Sharma et al., 2020). With respect to timbre, familiar sound sources are recognized more quickly than unfamiliar ones (Siedenburg and McAdams, 2016). In the time domain, both rhythm (Hannon et al., 2012) and meter perception (Kalender et al., 2013) are altered as a function of listener familiarity with a stimulus type. Regarding pitch, Weidema et al. (2016) showed that the same contours are perceived differently depending on the context in which they are embedded, with better discrimination in music than speech context. These studies highlight that the role of specific acoustic features in shaping perceptual categorization may, in part, be driven by previous exposure.

Acoustically, sound sequences from music and language domains can be defined in several ways, with summary statistics (i.e., mean), as reported earlier, but also in terms of changes over time. In the case of pitch, for example, speech and music are rarely monotonous. In speech, the intonation or pitch direction (e.g., utterances interpreted as statements or questions, Bolinger, 1986; Ladd, 2008), the pitch accent (e.g., Ladd et al., 1999), or the prosodic patterns (e.g., Bänziger and Scherer, 2005; Kraljic and Brennan, 2005; Dilley and McAuley, 2008), play a considerable role in carrying paralinguistic information, such as the emotional state (Banse and Scherer, 1996) or the intention (Hellbernd and Sammler, 2016) of the speaker. Additionally, pitch contrasts and thus changes of pitch over time additionally carry lexical information in tonal languages (Yip, 2002; Carter-Ényì and Carter-Ényì, 2016). In music, changes of pitch over time also provide crucial information that allow listeners to recognize, evaluate, and enjoy a musical performance. There exist different musical (cultural) systems that define pitch movements, with specific scales and rules (Krumhansl, 1979; Lerdahl and Jackendoff, 1983; Cross, 2001; Ringer, 2002; Thompson, 2013), but some features such as the presence of small intervals or descending melodies are present in different cultures (e.g., Huron, 2001) and have been described as statistical universals for music (Savage et al., 2015). Besides the relevance of pitch changes over time (in addition to mean pitch), a large range of literature in the music and language domains supports the perceptual relevance of changes in other dimensions such as timbre, intensity, or duration. It is thus important to explore acoustic features that take into account dynamics (rather than means) not only with regard to pitch, but also more broadly in the dimensions of timing, timbre, and intensity.

In this study, we further examine speech-music specificities by employing an instrument intertwining language and music: the dùndún talking drum. The dùndún is commonly used in south-west Nigeria as a musical instrument. The dùndún of the Yorùbá is played by people of all ages, though mostly men, and in a variety of sacred and secular cultural contexts (Durojaye, 2020). It is used to play musical rhythms without semantic information but also to communicate announcements, warnings, prayer, jokes, proverbs, or poetry (Sotunsa, 2009). While a dùndún ensemble consists of drums of varying sizes and functions, we focus here on the ìyá ìlù dùndún—the lead drum in the ensemble—usually performing the role of "talker" during performance, imitating Yorùbá, which is a tonal language, in what has been described as speech surrogacy (Durojaye et al., in review; McPherson, 2018). Villepastour, in her analysis of bàtá drums—a very close relative of the dùndún—argues for the interdependence of speech tone and music and highlights the significance of relative pitch and rhythm in the surrogacy system of the dùndún (Villepastour, 2010, 2014).

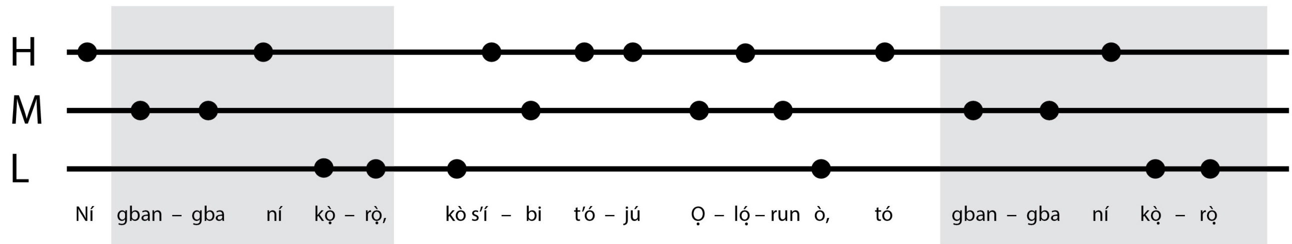


FIGURE 1 | Transcription of a Yorùbá proverb played on the dùndún (stimulus 6S in our corpus). The dots placed on the three horizontal lines represent a schematic depiction of the pitch changes (high, medium, or low) as the phrase written underneath would be spoken in the Yorùbá language. The phrase translates to "Public or private, there is no place the God cannot see, public or private." Though the phrase would typically be spoken without the repetition at the end, when played on the dùndún the proverb is elaborated by repetition of the first phrase (gray shaded areas), as repetition is often a means to provide context and remove ambiguity in meaning (Stern, 1957; Arewa and Adekola, 1980).

The Yorùbá language uses three relative tone levels: Low (grave accent), Middle (usually left unmarked), and High (acute accent). The tone levels are vital to distinguish the meaning of words (Carter-Ényì and Carter-Ényì, 2016). Like the language, the drum also consistently employs three relative tone levels. The dùndún is a waisted (hourglass shaped), double-headed membrane drum, with gut or leather cords securing the skins around the wooden frame of the drum. The cords are manipulated with one hand, while the other strikes the top membrane with a curved stick. Pressure on the cords changes the pitch of the drum, allowing for a full octave range and effects like glissandi (Blades, 1992; Euba, 1990). For the drum to produce the lowest pitch, minimal pressure is applied on the cords; the more the pressure, the higher the frequency. Thus, the drum can manipulate tone levels and contours, as in Yorùbá language. This imitation is confirmed by recent acoustical analyses of mono or disyllabic words performed on the drum which demonstrated that the three Yorùbá tones (Low, Middle, and High) are produced on a global level with three measurably different fundamental frequencies (Akinbo, 2019). The technique of representing syllables can take many forms such as (1) using one drum stroke for each syllable (as for a single tone level and vowel elisions); (2) many strokes for one syllable; (3) one drum stroke for two or more syllables; (4) one drum stroke for a syllable with many speech tone levels as would be the case for some glides, or assimilations (see also, Euba, 1990; Villepastour, 2010, for bàtá drums). A transcription of dùndún "talking" is provided as an example in **Figure 1**.

In the experiment outlined below, we seek to identify acoustic features associated with the perception of speech-like vs. music-like dùndún performances and the potential role of familiarity in such classification. To do so, we first compared speech-like vs. music-like dùndún performances with regard to different acoustic features related to pitch, intensity, timbre, and timing. Second, we presented the same samples to familiar and unfamiliar listeners and examined their ability to classify the performances as intended by the performer, as well as their confidence in the classification. Finally, a statistical model was created to quantify the role of listener familiarity and acoustic features of dùndún performance, in predicting listeners' perception of dùndún as speech-like vs. music-like.

MATERIALS AND METHODS

The experimental procedure was in accordance with guidelines ethically approved by the Ethics Council of the Max Planck Society. Participants provided informed consent before proceeding with the study.

Participants

One hundred and seven participants (36 self-reported as females, 71 as males, aged from 18 to 75 years old, M = 39.22, SD = 15.06) were recruited via the research participant database of the Max Planck Institute for Empirical Aesthetics and via personal contacts. From various geographical locations (15 different mother tongues were represented), fifty-one participants reported being familiar with the dùndún talking drum (i.e., they knew about the dùndún prior to the survey). Of these 51 participants, 28 (55%) were speakers of Yorùbá. Participants declared to have normal hearing ability and reported various musical training levels. Participants received no financial compensation.

Material

Thirty-six dùndún samples were created from performances by one professional dùndún musician from Ibadan, South-West Nigeria. The performer (male) is a native Yorùbá and fluent English speaker with more than 25 years of experience playing the drum. Performances were recorded at a local music studio with a SHURE SM57 dynamic microphone directed at the face of the drum, at a 3-inch distance, sampling at 44.1 kHz. Note that clicks and environmental noise can be heard in some recordings.

Half of the performances were music-like material consisting of Yorùbá àlùjó rhythms (literally dance drumming); the other half were speech-like material, composed of Yorùbá proverbs and oríkì (poetry). The performer was first asked to use the drum to say different phrases ("talk"), after which he was asked to "drum" (the equivalent of music). All instructions were given in the Yorùbá language. Note that in Yorùbá dùndún performances, when drummers say they "talk" with the drum, they refer to the performance of oríkì, proverbs, or the signal mode of the drum. Similarly, when they talk about "drumming" or "playing music" (for those who use the English term), they are making a reference to àlùjó. These categories were confirmed by the performer after the task. Also note that dance rhythms, proverbs, and poetry are used for any occasion (e.g., weddings, burials, religious events). In the current performances, the "talking" contents covered various themes, such as a praise to a deity, prayers, vilification, and admonition.

To confirm that the performances clearly represented the category of speech or music, all recordings were presented to three independent professional dùndún drummers located in Nigeria and South Africa. The judges were asked to categorize the performances according to whether they represented speech or music. Like the performer, they used the terms àlùjó, oríkì, owe (proverbs), in their responses. The 30 samples on which the judges unanimously agreed to represent speech (n = 15) or music (n = 15) were selected. The duration of the samples ranged from 5 to 10 s (M = 7.37 s, SD = 1.1 s). All recordings can be accessed at: https://edmond.mpdl.mpg.de/imeji/collection/ovmWl7rLtIiGSv1v.

Procedure

The task was implemented as an online experiment in Unipark Enterprise Feedback Suite (QuestBack GmbH, Cologne, Germany). Prior to the classification task, a brief presentation of the dùndún was given (origin, uses, description, picture) without sound examples. The aim was to provide a basic knowledge for those who reported being unfamiliar with the drum or the potential use of this instrument as speech surrogates. To determine participants' familiarity with the dùndún, we asked if they knew about the dùndún prior to the survey. Participants were instructed to listen to each excerpt and to indicate whether it was best described as "speech-like" or "music-like." For each excerpt, the forced choice identification was followed by a confidence rating on a 4 point-scale (1 = not confident, 4 = very confident). The order of stimuli and response pattern (speech-like button as the first or the second option) was randomized for each participant.

Acoustic Analyses

The analysis of acoustic features was carried out in MATLAB 2018b (The MathWorks, Inc., Natick, Massachusetts, United States).

Segmentation

Segmentation of single notes was performed semi-automatically on each recording's amplitude envelope, using an adaptive threshold. First, amplitude envelopes were slightly smoothed with a Hodrick-Prescott (HP) filter (coefficient = 50). Then, the adaptive segmentation threshold was created by applying a stronger HP filter to the amplitude envelope (coefficient = #10^7). The difference between the slightly smoothed amplitude envelope and the adaptive filter provided robust segmentation in most recordings. We visually inspected the segmented waveforms and sonograms while listening to the audio to validate segmentation. In cases where the automated segmentation had failed, we manually outlined note onsets with custom-written MATLAB code.

Acoustic Measures of Interest

First, we computed the amplitude modulation spectra of the recordings, following the procedure of Ding et al. (2017), with MATLAB code kindly provided by Nai Ding. In brief, we extracted the sound envelope in narrow frequency bands (corresponding to frequency bands used by the human cochlea), then, following a re-scaling procedure, we calculated the root mean square of the Discrete Fourier Transform of each frequency band and binned over frequencies. High frequencies in the amplitude modulation spectrum correspond to fast modulations of intensity, low frequencies to slow modulations (for details, see Ding et al., 2017). We then calculated the peak frequency in the spectrum (i.e., the frequency exhibiting the greatest amplitude modulation), for each recording. To further analyze differences in timing, we also computed the inter-onset interval (IOI, in ms) between notes, the two-interval ratio: interval1/(interval1+interval2), and the pulse clarity. The first measure corresponds to the timing at the note level and the second to the change in timing between consecutive time intervals. A short interval preceding a long interval would result in a ratio < 0.5, a short interval *following* a longer one has a ratio > 0.5, and an isochronous rhythm of two similar intervals has a ratio of 0.5. Pulse clarity, a measure that estimates the temporal regularity of events in the signal (Lartillot et al., 2008), was calculated using all recommended default parameters of the *mirpulseclarity* function from the music information retrieval toolbox v.1.7.2 (i.e., using a frame length of 5 s, a hop factor of 10%, and the maximum value of the autocorrelation curve to define clarity). Pulse clarity ranges between 0 (no clear pulse) and 1 (perfectly regular pulse).

Besides these time-related measures, we selected various features typically used to describe auditory signals: pitch, intensity, and timbre measures. At the *note level* and *between consecutive notes*, for each stimulus, we calculated: pitch height, intensity, and Wiener entropy (timbre, the maximum value is a pure sine tone). Scaling was performed within recording, on the millisecond-wise acoustic features between the 0.5 and 99.5th percentiles (instead of between minimum and maximum) to exclude outliers. Amplitude envelope and Wiener entropy were extracted from the audio in 10 ms time windows and steps of 1 ms using the MATLAB package Sound Analysis for MATLAB (by Sigal Saar). The pitch function from MATLAB Audio Toolbox (The MathWorks Inc., 2020) was used for pitch extraction. For pitch, intensity, and timbre measures (as well as IOI and ratio), we also calculated the probability densities for each group of stimuli (music-like and speech-like). Additionally, we computed the average of each feature, across each stimulus, resulting in the following final measures: mean pitch, mean intensity, and mean timbre; as well as average of absolute differences between consecutive tones (leading to mean intensity change, mean timbre change, and mean pitch change measures). Note that scaled values (0–1) were used to compute the measures capturing changes between consecutive notes.

Behavioral Analyses

Participants' Classification of Stimuli

Participants' responses on the task can be summarized using a 2 × 2 contingency table, or confusion matrix M:

$$M = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

where we arbitrarily define music as positive and speech as negative, such that true positives (TP) represent music intended by the performer and classified by the listener, while true negatives (TN) represent speech intended by the performer and similarly classified by listener. TPs (music) and TNs (speech) represent correct classifications. False positives (FP) designate instances of intended speech perceived as music, while false negatives (FN) indicate cases of intended music perceived as speech. Perfect classification is thus defined as 0 FPs and FNs: $\begin{bmatrix} n^{+} & 0 \\ 0 & n^{-} \end{bmatrix}$. Accuracy on the task is defined as the number of correct classifications, divided by the total number of observations:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

However, accuracy has been shown to be biased if the classes are unbalanced (e.g., if there would be more instances of perceived speech than music). Therefore, as widely used and recommended, we computed Matthews Correlation Coefficient (MCC):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

MCC is robust to unbalanced datasets, and has been shown to be a more reliable measure than accuracy (Chicco and Jurman, 2020). It is a measure of classification performance across all classes that takes into account the size of each class. MCC ranges between -1 and 1, with 0 indicating chance performance, 1 perfect performance, and -1 perfect misclassification. As is common practice with MCC, in the case of 0 in the denominator, we set the denominator to 1.

Statistical Analyses

Statistical analyses were conducted in R Core Team (2013). To compare acoustic measures between speech- and music-like dùndún recordings, we performed independent, two-tailed *t*-tests. For measures relative to changes between consecutive notes, absolute (rather than signed) differences were used. At an alpha of 0.05, the threshold *p*-value after correcting for multiple comparisons (10 *t*-tests) was $p = 0.005$. Cohen's d is used to report effect sizes and was calculated in R using the cohen.d function in the *effsize* package (Torchiano, 2020).

To predict participants' perception of the stimuli as music- or speech-like, a generalized linear mixed effects logistic regression model was fit via maximum likelihood using the Laplace approximation method, with bound optimization by quadratic approximation, implemented using the glmer function from the *lme4* package (Bates et al., 2015). Acoustic variables were centered and scaled before being entered into the model. Multicollinearity was checked using variance inflation factors (VIFs). Some predictors had a VIF > 5 and, therefore, required removal from the model. Specifically, inter-onset-interval (VIF = 5.90) and amplitude difference between adjacent notes (VIF = 8.27) were removed. Correlations between means of all acoustic features are provided in **Supplementary Figure 1**. Note that inter-onset interval has a high correlation with intensity [$r_{(28)} = 0.63$, $p < 0.001$] and intensity difference [$r_{(28)} = -0.82$, $p < 0.001$], and that intensity and intensity difference have a high correlation with each other [$r_{(28)} = -0.71$, $p < 0.001$].

RESULTS

Acoustic Properties of Dùndún Performances

We first investigated whether the amplitude modulation spectrum (AMS) systematically differs between speech-like and music-like dùndún stimuli, as the AMS has previously been shown to distinguish between speech (around 4–6 Hz) and different kinds of Western music (around 2 Hz). As illustrated in **Figure 2A** and confirmed via *t*-test, however, the AMS of the two types of stimuli did not significantly differ [$t_{(27.93)} = 1.60$, $p = 0.120$, $d = 0.59$]; both peaked around 5 Hz, which corresponds to the previously established amplitude modulation rate typical of *speech* (Ding et al., 2017; **Figure 2A**, bottom panel). In terms of pulse clarity, we find a significant difference between the two stimulus categories, with a greater pulse clarity (i.e., greater temporal regularity) in the music-like stimuli, $t_{(26.14)} = 4.03$, $p < 0.001$, $d = 1.47$ (**Figure 2B**).

In addition to the AMS measure, we examined four types of features (see section "Materials and Methods" for description of all measures) that could capture differences between music- and speech-like stimuli: intensity, pitch, timbre, and timing. **Figure 3** illustrates each measure at the note level (mean for each stimulus and their distribution). As reported in **Table 1**, we observed significant differences for intensity and timing, with higher intensity level (**Figure 3A**) and longer IOI (**Figure 3D**) in the speech-like recordings. **Figure 4** illustrates each measure at the consecutive note level (mean of differences between consecutive notes for each stimulus and their distribution). Besides being louder, consecutive notes also varied less in intensity in the speech-like stimuli (**Figure 4A**), in line with the distribution depicted in **Figure 3A**, with a narrower range for the speech-like stimuli. Also, we observed that near-isochrony is very common in both speech- and music-like excerpts (**Figure 4D**). However, in music-like ones, intervals tend to speed up (the second interval in an isochronous pair being a little shorter, moving the near-isochronous peak slightly right) while slowing down in speech-like ones (near-isochronous peak moved slightly left). Some very high and very low ratios (due to a short interval next to a long one) become apparent in the wider spread of speech-like data, in the top right scatter plot, and in the peaks of the probability density plot, marked by arrowheads.

Listeners' Classification of Dùndún Performances

As can be seen in **Figure 5A**, participants clearly separated the stimuli into two distinct speech vs. music categories that largely aligned with the intention of the performer. We

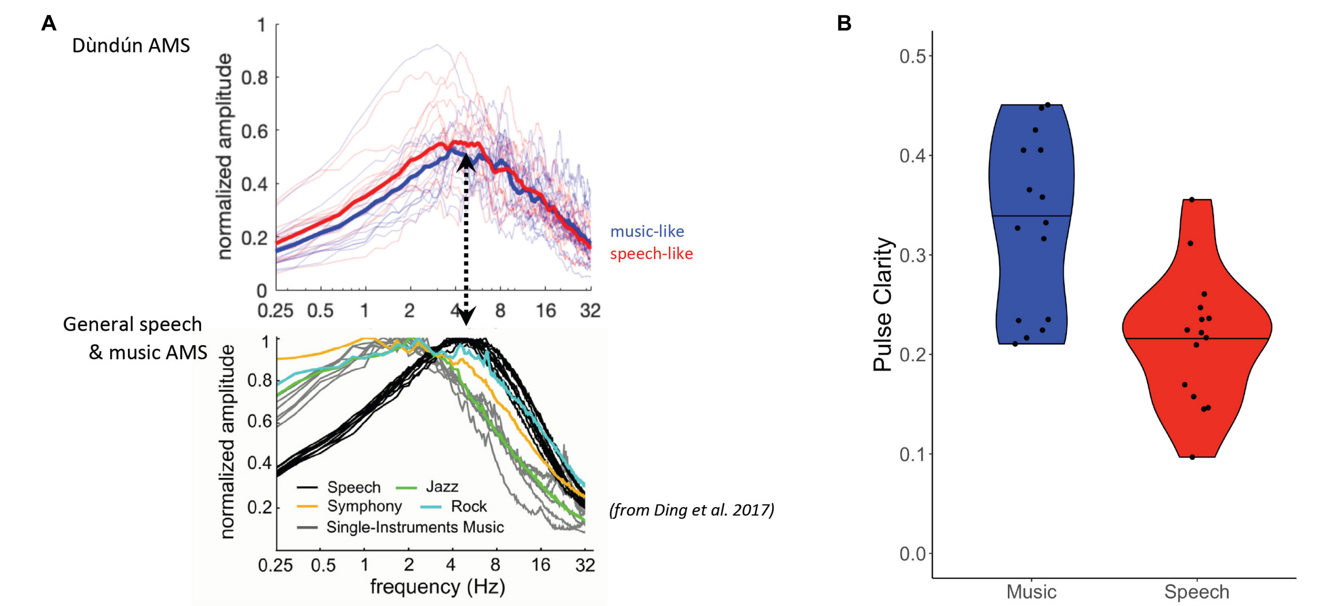


FIGURE 2 | (A) Amplitude modulation spectra of the dùndún performances (top panel). Thin lines are for AMS of individual recordings, thick lines represent the mean for 15 music (blue) and 15 speech-like (red) stimuli. Bottom panel shows AMS for speech and music corpora from Ding et al. (2017) for comparison. Dashed arrow illustrates that dùndún recordings peak at higher modulation rates than most Western music, namely rates that are typical for speech (black line) but also occur in rock music (teal line). The lower panel in A is reprinted from Neuroscience and Biobehavioral Reviews, 81, Nai Ding, Aniruddh D. Patel, Lin Chen, Henry Butler, Cheng Luo, David Poeppel, Temporal modulations in speech and music, 7, Copyright Elsevier Ltd. (2017), with permission from Elsevier. **(B)** Violin plots representing the pulse clarity of the music (blue) and speech-like (red) stimuli. The black horizontal line indicates the median of both distributions. Black dots represent pulse clarity for individual stimuli.

observed that only four participants categorized every sample as music (solid blue rows near the bottom of the plot), whereas the large majority showed few confusions. Twelve participants exhibited perfect classification (top rows). In the figure, within the speech and music categories, stimuli (columns) are sorted by the number of errors made per stimulus (i.e., the left-most column, stimulus 13M, was least often confused for speech, while 3M was most often confused for speech). Within the speech category, 13S was most clearly perceived as speech, while 5S was most often confused for music. Readers can access all stimuli online (link in section "Materials and Methods").

TABLE 1 | Output of the independent *t*-tests carried out for the four types of acoustic features (Intensity, Pitch, Timbre, and Timing) between music- and speech-like stimuli.

	Intensity	Pitch	Timbre	Timing
Mean for note level	$t_{(18.18)} = -5.37$	$t_{(20.61)} = 2.06$	$t_{(18.84)} = -1.38$	$t_{(16.18)} = -8.53$
	$p < 0.001$	$p = 0.052$	$p = 0.185$	$p < 0.001$
	$d = -1.96$	$d = 0.75$	$d = -0.50$	$d = -3.12$
Mean change for consecutive notes	$t_{(26.32)} = 8.95$	$t_{(17.53)} = -0.75$	$t_{(27.60)} = 4.09$	$t_{(15.74)} = 2.17$
	$p < 0.001$	$p = 0.465$	$p < 0.001$	$p = 0.046$
	$d = 3.27$	$d = -0.27)$	$d = 1.49$	$d = 0.79$

*For the four types of comparisons, measures consisted of the mean of all notes (**Figure 2**) and the mean of the absolute difference between all consecutive notes (**Figure 3**). Note that the threshold p-value (for an alpha of 0.05) after correcting for multiple comparisons was p = 0.005. Thicker frames highlight the measures significantly differing after correction.*

A confusion matrix for perceived vs. intended music and speech-like performances is plotted in **Figure 5B**. Overall, the average accuracy of participants on the task was 66%. The average rate people perceived speech when the performance was intended to be music was 12%, while the average rate at which people perceived music when it was intended to be speech was 29%. Collectively, these latter two rates indicate that participants were more likely to perceive speech as music than music as speech. The illustration of confidence ratings (underlying histograms in **Figure 5B**, with ratings from 1 to 4) showed similar patterns, with moderately high confidence even in the case of false classifications. Note, however, that the listeners who were unfamiliar (gray) with the dùndún seem to be least confident when they perceive the stimulus to be speech (right column in confusion matrix). Confidence means for unfamiliar (gray) and familiar (gold) participants are displayed in the lower left and right corners of each quadrant, respectively.

Given the imbalance in perceiving speech vs. music, and the statistical properties outlined in the methods, our main metric of interest for participants' classifications of the stimuli was the Matthews Correlation Coefficient (MCC). An MCC of 1 indicates perfect classification, 0 chance, and -1 perfect misclassification. Participants' average MCC was 0.61 (±0.33).

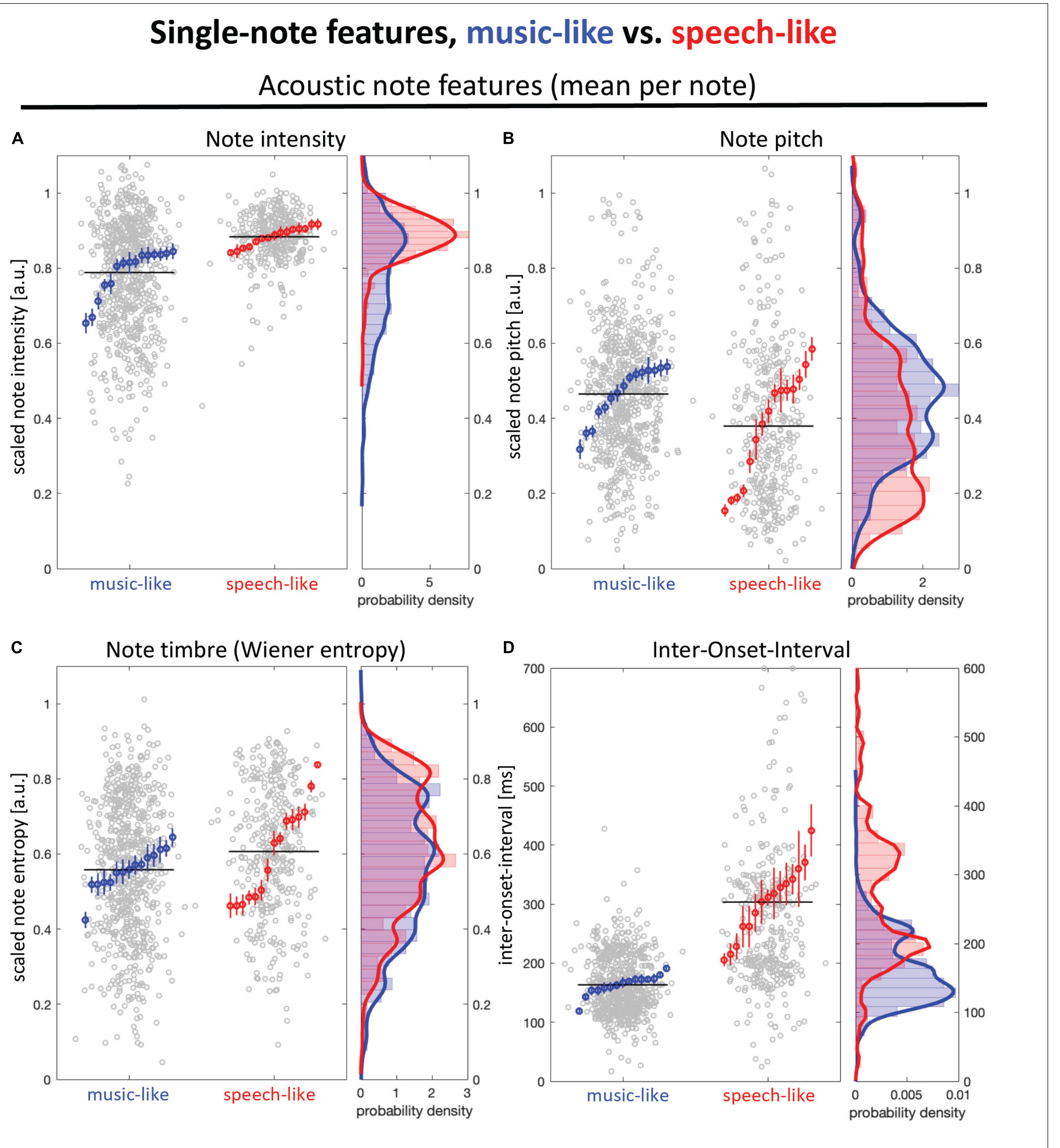


FIGURE 3 | Acoustic measures computed at the note level for music-like (blue) and speech-like (red) dùndún performances. **(A)** Mean intensity, **(B)** Mean pitch, and **(C)** Mean timbre (Wiener entropy) **(D)** Mean inter onset intervals (IOI). Left panels show mean note values for 15 speech- and music-like pieces each in gray, with mean and SEM per piece in color. Right panels show the respective probability densities.

Participants who were familiar[1] with the dùndún exhibited a significantly higher MCC, compared to those who were unfamiliar with the dùndún [Welch's independent two-tailed, t-test: $t_{(93.08)} = 6.12$, $p < 0.001$, Cohen's $d = 1.16$, ΔMCC = 0.33]; see **Figure 5C**; N.B. each participants' familiarity (gold = familiar; gray = unfamiliar) is plotted in the left color bar of **Figure 5A**). Nonetheless, familiarity is not required to perform the task, as unfamiliar participants still exhibited an average MCC

[1]We ran an additional t-test comparing the MCC of the subset of familiar participants who did not speak Yorùbá to those unfamiliar and still found a large, significant effect of familiarity, $t_{(58.74)} = 3.18$, $p = 0.002$, $d = 1.16$, ΔMCC = 0.211.

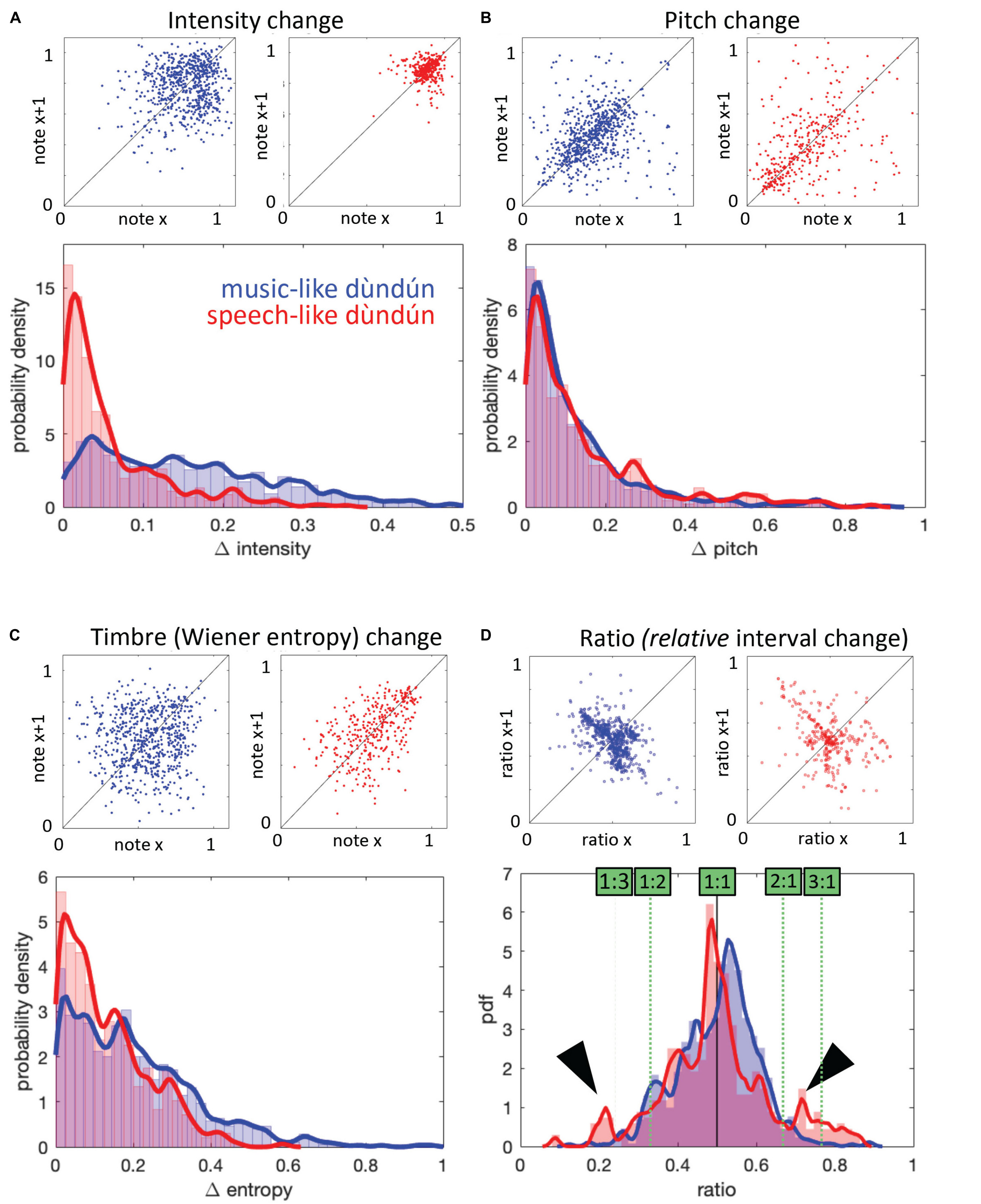


FIGURE 4 | Acoustic measures computed at the level of consecutive notes for music-like (blue) and speech-like (red) dùndún performances. All measures are scaled from 0 to 1. **(A)** Mean intensity change: The top panel shows, for each pair of consecutive notes, the intensity of the pair's first note on the *x*-axis, and of the second note on the *y*-axis. The diagonal line indicates where note pairs with two similar note intensities would fall. The bottom panel shows the probability density of intensity differences. **(B,C)** are the Mean pitch change and Mean timbre change, illustrated following the same logic as **(A)**. **(D)** Similar illustration for change in IOIs, as measured by ratio = interval1 / (interval1+interval2). In the probability density plot (bottom), green boxes and dashed lines mark interval ratios in terms of "interval1: interval2," an alternative way to describe interval ratios, commonly used in music theory.

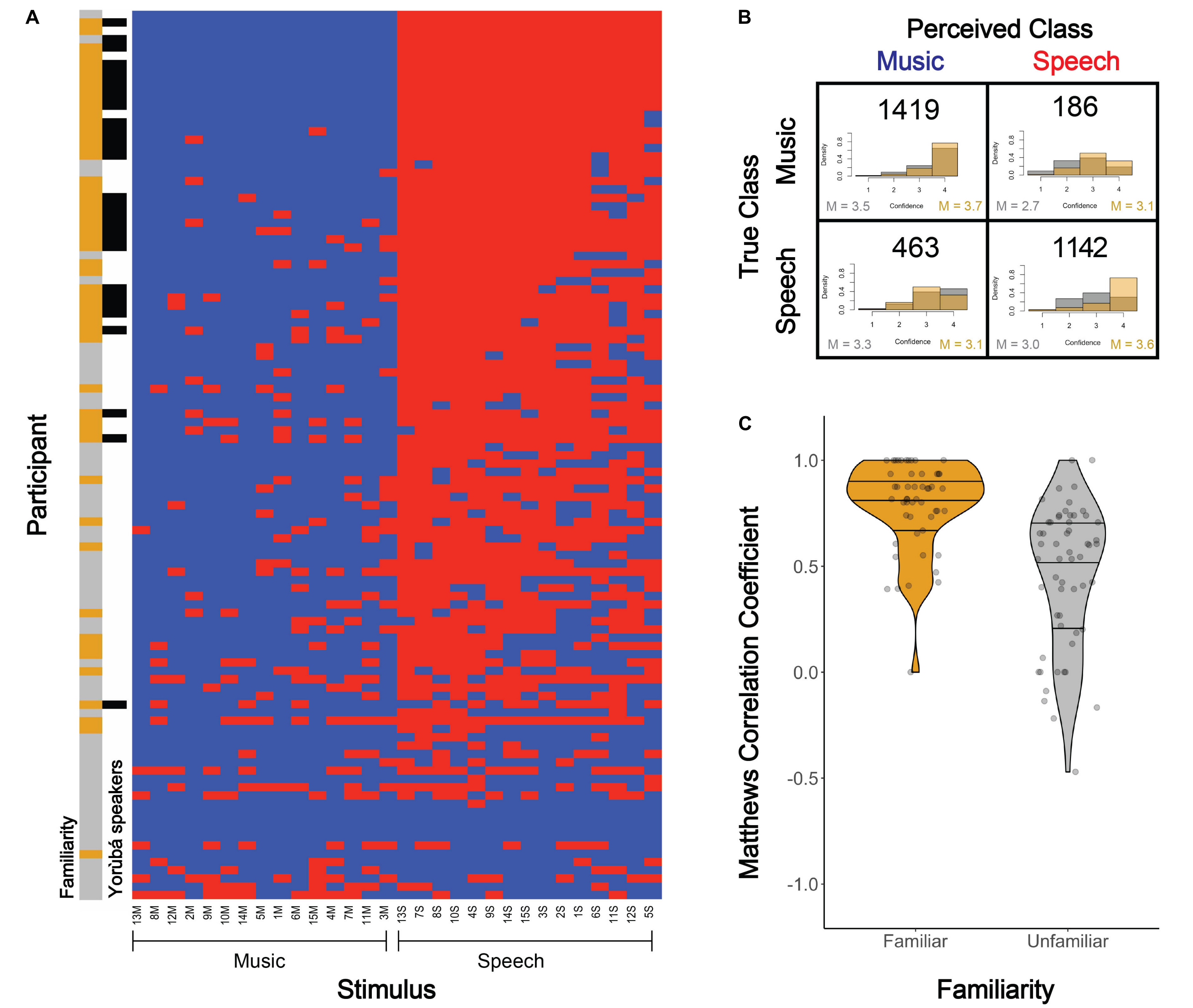


FIGURE 5 | Participants' classification of music vs. speech-like dùndún performances. **(A)** Each participant's (vertical axis) judgment of each stimulus (horizontal axis) as music (blue) or speech (red). Participant performance is sorted descending from least to most errors. Columns are sorted from left to right, within each category (speech and music), from least to most errors per stimulus. The color bar to the left of the plot indicates whether each participant was familiar (golden) or unfamiliar (gray) with the dùndún and whether they spoke Yorùbá (black) or not (white). **(B)** Confusion matrix for perceived vs. intended stimulus classes, with histograms of participants' confidence ratings (1–4) for each response type (i.e., quadrant), grouped by familiarity (gold = familiar; gray = unfamiliar). Confidence rating densities were computed within each response type (quadrant) for each familiarity group separately. Means for unfamiliar (gray) and familiar (gold) groups are displayed in the lower left and right corners of each quadrant, respectively. **(C)** Violin plots and underlying data points indicating the Matthews Correlation Coefficient (MCC) for each participant, separated according to those who were familiar with the dùndún (golden) vs. unfamiliar (gray). The bottom and top horizontal black lines in each distribution represent the 25th and 75th percentiles, the middle line represents the median. An MCC of 1 indicates perfect classification; 0 represents chance, and -1 perfect misclassification.

of 0.46, well above chance (0), $t_{(55)} = 10.10$, $p < 0.001$, $d = 1.35$.

Predictors of Listeners' Perception of Speech vs. Music

In an effort to understand which acoustic features were most relevant in participants' perception of the dùndún excerpts as music vs. speech-like, we built a linear mixed effects logistic regression model. The binary dependent variable was participants' perception (speech = 0, music = 1). On the stimulus level, fixed effects included all measures reported in **Figures 2–4**, except intensity difference and inter-onset-interval, which had to be excluded due to high correlation with intensity and each other (see section "Materials and Methods"). Since we observed an effect of familiarity on the classification performance index, with better classification for listeners who were familiar with the dùndún, we included familiarity as a fixed effect and in

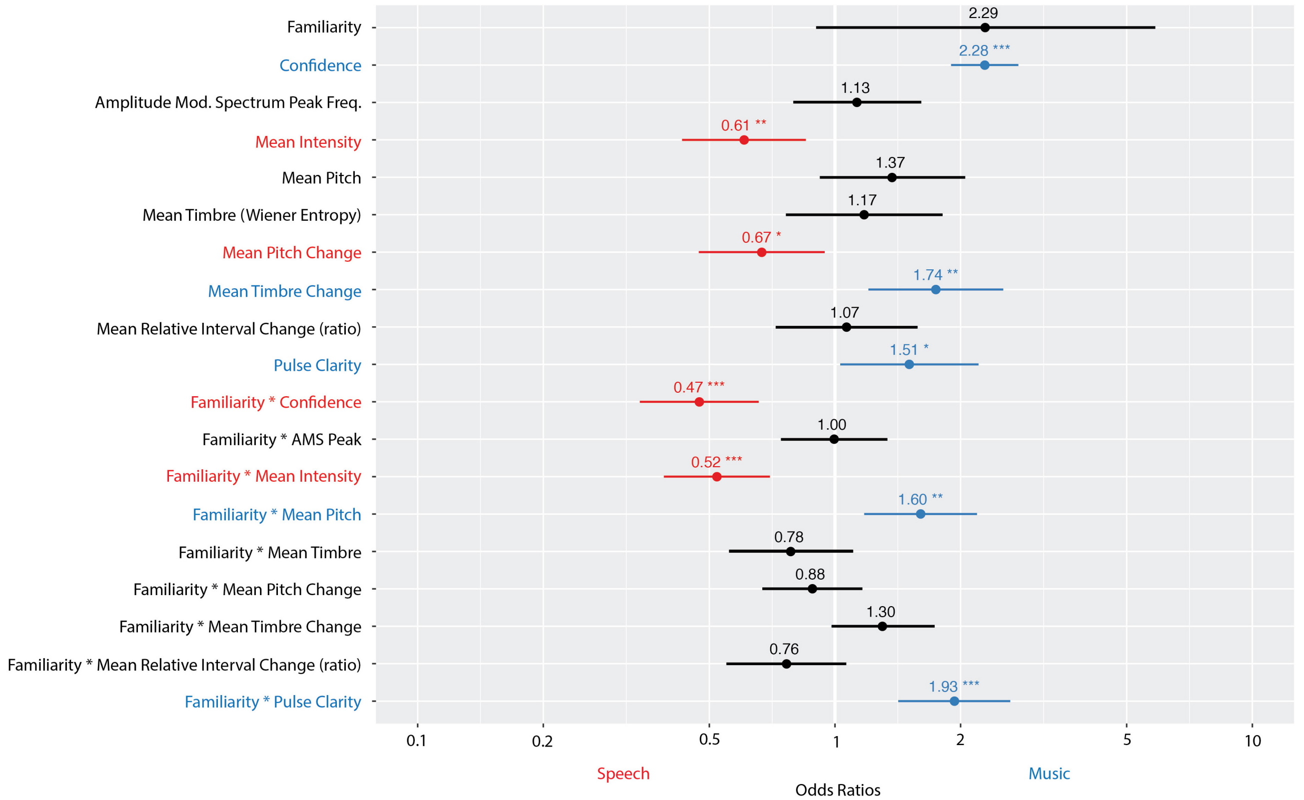


FIGURE 6 | Odds ratios, with confidence interval (CI), for each fixed effect in a logistic regression model predicting participants' perception of stimuli as music-like (1) or speech-like (0). The vertical line at 1 indicates no effect (i.e., any fixed effect predictor whose odds ratio CI overlaps 1 does not significantly predict participants' perception). Fixed effects with an odds ratio <1 (red) indicate that a high value on that feature leads to the perception of speech. Odds ratios >1 (blue) indicate that a high value on that feature leads to the perception of music. The significance of each fixed effect is indicated with stars (***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$). Number of observations: 3,210. Familiarity was a binary predictor (0 = no; 1 = yes); confidence ranged from 0 to 3; all other variables were continuous.

interaction with all acoustic measures. Confidence ratings were also entered as fixed effects. Random intercepts were included for participants and stimuli.

Figure 6 shows the odds ratios and confidence intervals for each fixed effect in the model. Fixed effects with an odds ratio <1 (red) indicate that a high value on that feature leads to the perception of speech. Odds ratios >1 (blue) indicate that a high value on that feature leads to the perception of music. Overall, the model had a prediction accuracy of 85% and an MCC of 0.70. The model explained a significant proportion of variance in the data, with a marginal R^2 of 0.46 (amount of variance explained by fixed effects alone) and a conditional R^2 of 0.65 (amount of variance explained by fixed and random effects).

Greater pulse clarity predicted perception of music. At the note level, greater mean intensity predicted speech. In terms of changes between notes, greater changes in pitch predicted speech, while greater changes in timbre predicted music. However, mean intensity, pitch and pulse clarity all interacted with participants' familiarity (same direction of the overall effects but enhanced magnitude). Additionally, confidence predicted the categorization "music-like," and interacted with familiarity, such that those familiar and confident were more likely to classify a given stimulus as speech, while those not familiar and confident were more likely to classify the stimulus as music. These confidence/familiarity results are in line with the general trends presented in **Figure 5B**.

DISCUSSION

In this study, we used an instrument capable of speech surrogacy to explore the boundaries between speech and music. While several instruments such as trumpets (Kaminski, 2008), flutes (Moore and Meyer, 2014), xylophones (Strand, 2009; Zemp and Soro, 2010; McPherson, 2018), and whistling (Stern, 1957) can produce speech surrogates, we focussed here on the dùndún talking drum to examine listeners' perception of music and speech and, more specifically, the role of acoustic features in distinguishing these two categories. To do so, we recorded expert dùndún performances, which have the advantage of being able to create both musical and speech-like stimuli without requiring the human voice (highly familiar to all listeners). Acoustic measures relative to pitch, timbre, intensity, and timing were used to describe the stimuli and we measured listeners' ability to classify each performance into two pre-defined categories intended by the performer: speech- or music-like.

Participants could accurately classify the dùndún performances in the category intended by the performer, with an unsurprising bias toward the music-like category, given that drums are more commonly associated with music than speech. Listeners who were familiar with the instrument seem to have a clearer representation of what is grouped into speech or music categories, as visible by their better classification performance. Such results could be driven by the fact that slightly over half of the familiar participants also reported speaking Yorùbá, though the familiarity effect remained even when Yorùbá-speaking participants were removed from the analysis. In any case, if familiarity (broadly defined) or Yorùbá fluency sharpens the categories, it is not required to perform the task, as shown by the above chance level MCC and relatively high confidence for participants who were not familiar with the instrument. Such results suggest the relevance of commonly recognized acoustic cues that shape each category but become fine-tuned through repeated exposure.

In the current study, we asked broadly about listeners' familiarity and thus are not able to disentangle what exactly underlies the familiarity effect. Teasing apart perceptual vs. cultural vs. linguistic familiarity might all be of interest in future research. Additionally, manipulation of familiarity, via priming or additional explicit information about the dùndún, might allow one to quantify the amount and type of previous exposure that affects the classification. Further, exploration of listeners' perception of dùndún performances with less constrained answer types might reveal other categories that could include both music and speech-like performances or less strong boundaries between these two culturally shaped categories (Brown, 2000).

The model proposed to examine the predictors of participants' classification revealed that participants relied on several features. Pulse clarity, mean note intensity, and mean timbre change between consecutive notes were significant predictors. Additionally, familiarity seemed to amplify the effects of pulse clarity and intensity in predicting music vs. speech, respectively. These perceptual results are in line with those we find to distinguish acoustically between the two different stimulus categories. However, we also observed that features which did not significantly differ between the speech and music recordings nonetheless contributed significantly to the perceptual model. Specifically, mean pitch did not differ acoustically between the two types of stimuli performed by the musician, though it interacted with familiarity in predicting the perception of music. Similarly, mean pitch change between consecutive notes did not significantly differ acoustically but was associated with the perception of the stimuli as speech. In the future, the nature of the music and speech categories (that are slightly modulated by the familiarity of the listeners) could be clarified by means of psychophysical experiments that parametrically manipulate the relevant acoustic features reported here.

It is interesting to note the considerably faster amplitude modulations of the dùndún performances, compared to the AMS previously established for music (around 2 Hz, Ding et al., 2017) or to the perceived rate in natural sounds (Roeske et al., 2020). In the current study, peaks stand around 5 Hz for both music and speech-like dùndún. Though some musical styles have been shown to also peak around 5 Hz (e.g., rock), this rate is consistent with the modulation rate of speech (Ding et al., 2017). However, while Ding et al. (2017) analyzed a variety of Indo-European languages (American and British English, French, German, Swedish, Dutch, Danish, Norwegian; exception: Chinese), they did not include any Niger-Congo languages, such as Yorùbá, and only included Western types of music, which limits the generalizability of their findings. Here we extend Ding et al.'s AMS analysis to non-Western stimuli (dùndún) and show that its peak closely resembles that of speech. Future work might extend the AMS analyses to spoken Yorùbá and compare with that of the dùndún to better understand the findings about the temporal aspect reported here. Note also that, while mean IOI was significantly different between the two types of stimuli, it unfortunately could not be included into our statistical model (like the mean intensity change measure), as it had high correlation with other features (**Supplementary Figure 1**). Thus, it could well be the case that participants are using IOI (as well as intensity changes) in their classification. This issue could be addressed by systematically manipulating IOI, as well as mean intensity and intensity change, to tease apart their perceptual relevance.

Regarding the dynamic aspect of timing, the present study focused mainly on consecutive notes or intervals, though we did include a measure of pulse clarity. Our pulse clarity metric was related to maxima in the autocorrelation function of our stimuli (i.e., periodic repetitions), but our measure set could also be extended to perceived beat and/or meter (e.g., Tomic and Janata, 2008; London et al., 2017) or the detection of repeated patterns (e.g., via recurrence quantification analysis Fukino et al., 2016 or multi-fractal analysis Roeske et al., 2018). The analysis/evaluation of longer stimuli would allow for application of a more extensive set of timing measures to investigate their role in speech vs. music distinction. Also, though our stimulus set is novel in that it consists of speech-like and music-like performances on the dùndún from the same performer, it is also limited in scope. Future studies might consider developing larger corpora with more examples of speech-like and music-like material from multiple performers. In addition, future research might also more closely consider the relationship between measures like IOI, perceived beat and meter, and AMS. Though IOI, AMS, and perceived pulse / meter all have previously reported preferred temporal ranges, which broadly seem to align with each other (e.g., Fraisse, 1963; Parncutt, 1994; Farbood et al., 2013; Gotham, 2015; Ding et al., 2017), it is likely that IOI, AMS, and perceived pulse / meter do not form a trivial and/or consistent relationship to one another across all timescales.

Previous studies have suggested that surrogate languages or language-based music, such as talking drums, may enhance our understanding of music and language (Patel, 2008; Winter, 2014). In the present study, such ecologically valid material provided the unique opportunity to compare stimuli coming from the same sound source (and performer) while representing different conceptual domains, which paves the way to a more in-depth understanding of speech/music differences/similarities. Altogether, our findings confirm the relevance of acoustic features relative to intensity, pitch, timbre, and timing in distinguishing speech and music, as well as the role of culture and/or exposure in defining such categories.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the links below: https://edmond.mpdl.mpg.de/imeji/collection/ovmWl7rLtIiGSv1v, https://github.com/lkfink/Dundun.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Council of the Max Planck Society. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CD, MW-F, and PL-M designed the study. CD collected the data. TR conducted acoustic analyses. LF conducted perceptual analyses and mixed modeling. LF and PL-M drafted the manuscript. All authors edited and approved the manuscript.

ACKNOWLEDGMENTS

We thank David Poeppel for helpful discussion and feedback on analyses. We thank the performer for providing the stimuli and the dùndún experts who assisted in classifying the stimuli for final use.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.652673/full#supplementary-material

REFERENCES

Akinbo, S. (2019). Representation of Yorùbá tones by a talking drum: an acoustic analysis. *Linguist. Lang. Afr.* 5, 11–23.

Arewa, O., and Adekola, N. (1980). Redundancy principles of statistical communications as applied to Yorùbá talking-drum. *Anthropos* 75, 185–202.

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614

Bänziger, T., and Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Commun.* 46, 252–267. doi: 10.1016/j.specom.2005.02.016

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Blades, J. (1992). *Percussion Instruments and Their History*. Wesport, CT: Bold Strummer Limited.

Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford, CA: Stanford: Stanford University Press.

Bregman, M. R., and Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition* 130, 85–95. doi: 10.1016/j.cognition.2013.09.010

Brown, S. (2000). "The 'musilanguage' model of music evolution," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 271–300.

Carter-Ényì, A., and Carter-Ényì, Q. (2016). "Perception of syntagmatic tone intervals in Ìgbò and Yorùbá," in *Proceedings of the Tonal Aspects of Languages 2016*, Buffalo, NY, 107–110. doi: 10.21437/TAL.2016-23

Chicco, D., and Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21:6. doi: 10.1186/s12864-019-6413-7

Cross, I. (2001). Music, cognition, culture, and evolution. *Ann. N. Y. Acad. Sci.* 930, 28–42. doi: 10.1111/j.1749-6632.2001.tb05723.x

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252. doi: 10.1121/1.3562174

Deutsch, D., Lapidis, R., and Henthorn, T. (2008). The speech-to-song illusion. *J. Acoust. Soc. Am.* 124:2471. doi: 10.1121/1.4808987

Dilley, L. C., and McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *J. Mem. Lang.* 59, 294–311. doi: 10.1016/j.jml.2008.06.006

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 18, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Durojaye, C. (2020). The dùndún talking drum of the Yorùbá community in south-west Nigeria. *J. Arts Hum.* 9, 11–19. doi: 10.18533/journal.v9i7.1902

Durojaye, C., Knowles, K., Jakob Patten, K., Garcia, M., and McBeath, M. (in review). *When Music Speaks: An Acoustic Study of the Speech Surrogacy of the Nigerian Dùndún Talking Drum*.

Euba, A. (1990). *Yorùbá Drumming: The Dùndún Tradition*. Bayreuth: Bayreuth African Studies.

Farbood, M. M., Marcus, G., and Poeppel, D. (2013). Temporal dynamics and the identification of musical key. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 911–918. doi: 10.1037/a0031087

Fraisse, P. (1963). *The Psychology of Time*. Oxford: Harper & Row.

Fukino, M., Hirata, Y., and Aihara, K. (2016). Coarse-graining time series data: recurrence plot of recurrence plots and its application for music. *Chaos* 26:023116. doi: 10.1063/1.4941371

Gotham, M. (2015). Attractor tempos for metrical structures. *J. Math. Mus.* 9, 23–44. doi: 10.1080/17459737.2014.980343

Hannon, E. E., Soley, G., and Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: a cross-cultural comparison of American and Turkish listeners. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 543–548. doi: 10.1037/a0027225

Hellbernd, N., and Sammler, D. (2016). Prosody conveys speaker's intentions: acoustic cues for speech act perception. *J. Mem. Lang.* 88, 70–86. doi: 10.1016/j.jml.2016.01.001

Huron, D. (2001). Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Percept.* 19, 1–64. doi: 10.1525/mp.2001.19.1.1

Jaisin, K., Suphanchaimat, R., Figueroa Cnadia, M., and Warren, J. (2016). The speech-to-song illusion is reduced in speakers of tonal (vs. non-tonal) languages. *Front. Psychol.* 7:662. doi: 10.3389/fpsyg.2016.00662

Kalender, B., Trehub, S. E., and Schellenberg, E. G. (2013). Cross-cultural differences in meter perception. *Psychol. Res.* 77, 196–203. doi: 10.1007/s00426-012-0427-y

Kaminski, J. (2008). Surrogate speech of the Asante ivory trumpeters of Ghana. *Ybook. Trad. Music* 40, 117–135.

Kraljic, T., and Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: for the speaker or for the addressee? *Cogn. Psychol.* 50, 194–231. doi: 10.1016/j.cogpsych.2004.08.002

Kraus, N., Skoe, E., Parbery-Clark, A., and Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre and timing. *Ann. N. Y. Acad. Sci.* 1169, 543–557. 10.1111/j.1749-6632.2009.04549.x

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cogn. Psychol.* 11, 346–374. 10.1016/0010-0285(79)90016-1

Ladd, D. R., Faulkner, D., and Faulkner, H. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *J. Acoust. Soc. Am.* 106:1543. doi: 10.1121/1.427151

Ladd, R. (2008). *Intonational Phonology*, 2nd Edn. Cambridge: Cambridge University Press.

Lartillot, O., Eerola, T., Toiviainen, P., and Fornari, J. (2008). "Multi-feature modeling of pulse clarity: design, validation, and optimization," in *Proceedings of the 9th International Conference on Music Information Retrieval*, eds J. P. Bello, E. Chew, and D. Turnbull (Philadelphia, PA: Drexel University), 521–526.

Lerdahl, F., and Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Percept.* 1, 229–252. doi: 10.2307/40285257

London, J., Polak, R., and Jacoby, N. (2017). Rhythm histograms and musical meter: a corpus study of Malian percussion music. *Psychon. Bull. Rev.* 24, 474–480. doi: 10.3758/s13423-016-1093-7

Margulis, E., Simchy-Gross, R., and Black, J. (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Front. Psychol.* 6:48. doi: 10.3389/fpsyg.2015.00048

McPherson, L. (2018). The talking Balafon of the Sambla: grammatical principles and documentary implications. *Anthropol. Linguist.* 60, 255–294. doi: 10.1353/anl.2019.0006

Merrill, J., and Larrouy-Maestri, P. (2017). Vocal features of song and speech: insights from Schoenberg's pierrot lunaire. *Front. Psychol.* 8:1108. doi: 10.3389/fpsyg.2017.01108

Moore, D., and Meyer, J. (2014). The study of tone and related phenomena in an Amazonian tone language: Gavião of Rondônia. *Lang. Doc. Conserv.* 8, 613–636.

Morrison, S. J., and Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Prog. Brain Res.* 178, 67–77. doi: 10.1016/S0079-6123(09)17805-6

Palmer, C., and Krumhansl, C. L. (1987). Pitch and temporal contributions to musical phrase perception: effects of harmony, performance timing, and familiarity. *Percept. Psychophys.* 41, 505–518. doi: 10.3758/BF03210485

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* 11, 409–464. doi: 10.2307/40285633

Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

Perrachione, T. K., Del Tufo, S. N., and Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science* 333:595. doi: 10.1126/science.1207327

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ringer, A. L. (2002). "Melody: definition and origins," in *The New Grove Dictionary of Music Online*, ed. L. Macy (London: Macmillan Online Publishing).

Roeske, T., Kelty-Stephen, D., and Wallot, S. (2018). Multifractal analysis reveals music-like dynamic structure in songbird rhythms. *Sci. Rep.* 8:4570.

Roeske, T., Larrouy-Maestri, P., Sakamoto, Y., and Poeppel, D. (2020). Listening to birdsong reveals basic features of rate perception and aesthetic judgements. *Proc. R. Soc. B.* 287:20193010. doi: 10.1098/rspb.2019.3010

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci.* 112, 8987–8992. doi: 10.1073/pnas.1414495112

Sharma, N., Krishnamohan, V., Ganapathy, S., Gangopadhayay, A., and Fink, L. (2020). Acoustic and linguistic features influence talker change detection. *J. Acoust. Soc. Am. Express Lett.* 147:5. doi: 10.1121/10.0002462

Siedenburg, K., and McAdams, S. (2016). The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory* 25, 550–564. doi: 10.1080/09658211.2016.1197945

Sotunsa, M. (2009). *Yorùbá Drum Poetry*. London: Stillwatersstudios.

Stadlen, P. (1981). Schoenberg's Speech-Song. *Music Lett.* 62, 1–11. doi: 10.1093/ml/62.1.1

Stern, T. (1957). Drum and whistle "languages": an analysis of speech surrogates. *Am. Anthropol.* 59, 487–506. doi: 10.1525/aa.1957.59.3.02a00070

Strand, J. (2009). *The Sambla Xylophone: Tradition and Identity in Burkina Faso*. Unpublished Ph.D. thesis. Middletown, CT: Wesleyan University.

The MathWorks Inc. (2020). *Audio Toolbox Toolbox*. Natick, MA: The MathWorks Inc.

Thompson, W. F. (2013). "Intervals and scales," in *The Psychology of Music*, 3rd Edn. ed. D. Deutsch (London: Elsevier), 107–140. doi: 10.1016/B978-0-12-381460-9.00004-3

Tomic, S. T., and Janata, P. (2008). Beyond the beat: modeling metric structure in music and performance. *J. Acoust. Soc. Am.* 124, 4024–4041. doi: 10.1121/1.3006382

Torchiano, M. (2020). *effsize: Efficient Effect Size Computation. R Package Version 0.8.1*. doi: 10.5281/zenodo.1480624

Vanden Bosch der Nederlanden, C. M., Hannon, E. E., and Snyder, J. S. (2015). Finding the music of speech: musical knowledge influences pitch processing in speech. *Cognition* 143, 135–140. doi: 10.1016/j.cognition.2015.06.015

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bata Drum*. Farnham: Ashgate.

Villepastour, A. (2014). Talking tones and singing speech among the YoruÌbaì of Southwest Nigeria. *Jahrb. Phonogrammarchivs Österr. Akad. Wiss.* 44, 29–47.

Weidema, J., Roncaglia-Denissen, M. P., and Honing, H. (2016). Top-down modulation on the perception and categorization of identical pitch contours in speech and music. *Front. Psychol.* 23:1647. doi: 10.3389/fpsyg.2016.00817

Winter, Y. (2014). On the grammar of a Senegalese drum language. *Language* 90, 644–668. doi: 10.1353/lan.2014.0061

Yip, M. (2002). *Tone. Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press.

Zatorre, R. J., and Baum, S. R. (2012). Musical melody and speech intonation: singing a different tune. *PLoS Biol.* 10:e1001372. doi: 10.1371/journal.pbio.1001372

Zemp, H., and Soro, S. (2010). Talking Balafons. *Afr. Music* 8, 7–24.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

BRIEF RESEARCH REPORT
published: 26 May 2021
doi: 10.3389/fcomm.2021.649416

Linguistic Surrogacy With Minimal Semantics Among the Dagomba of Ghana

Fusheini Hudu[1,2]*

[1]*Department of Linguistics, University of Ghana, Accra, Ghana,* [2]*Department of Gur-Gonja Education, University of Education, Winneba, Ajumako, Ghana*

This paper discusses critical questions on the processing of non-native surrogate languages of the Dagbamba (Dagomba) of Ghana. The Dagbamba use the fiddle, talking drum and double bell to encode speech in Hausa, Akan and other languages they do not speak. Fiddling and talking drums are integral to their festivals, funerals, the installation of chiefs and other cultural events. These instruments are used to entertain, praise, and send messages ranging from daybreak notifications to mobilizing people for war. The surrogate language they produce is a specialized language, interpreted mainly by people deeply rooted in their culture. It indicates nobility and statesmanship. While the performers and their patrons do not understand Akan or Hausa, they process and communicate with Akan and Hausa surrogate languages. The maintenance of the languages of performance is part of the practitioners' desire to preserve the cultural heritage of Dagbamba. This raises questions about the acquisition of these surrogate languages, the level of accuracy of production and comprehension, the role of music in the processing and the implications of these for linguistic theory. These questions are discussed on the basis of data from recorded interviews of talking drummers and fiddlers. The overarching goal is to highlight the gaps in our understanding of language processing that surface in the study of surrogate language, when processing takes place with a poverty of grammatical content.

Keywords: Dagomba, talking drum, fiddle, double bell, language processing, semantic processing

OPEN ACCESS

Edited by:
Laura McPherson,
Dartmouth College, United States

Reviewed by:
Abdulai Salifu Asuro,
Madina Institute of Science and Technology, Ghana
Anne Storch,
University of Cologne, Germany

Correspondence:
Fusheini Hudu
fahudu@ug.edu.gh

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *04 January 2021*
Accepted: *11 May 2021*
Published: *26 May 2021*

Citation:
Hudu F (2021) Linguistic Surrogacy With Minimal Semantics Among the Dagomba of Ghana. Front. Commun. 6:649416. doi: 10.3389/fcomm.2021.649416

INTRODUCTION

This brief research report concerns the processing of non-native surrogate language among the Dagbamba (Dagomba) of Ghana. Surrogate instruments of the Dagbamba are of two types. Some encode speech in Dagbani, the native language of the Dagbamba; others encode speech in non-native languages. The surrogate language encoded by both of these instruments is specialized for people deeply rooted in the culture and traditions of the Dagbamba. The use of non-native surrogate language marks statesmanship and indicates nobility. While the performers and their patrons do not understand the languages encoded by the instruments, they use them to convey messages that are of critical importance to the performance of rituals and the projection of their culture, among other uses.

This raises questions about the acquisition/learning or processing of these surrogate languages by the performers and patrons. For instance, how do the fiddlers and drummers learn to play the instruments and produce the appellations of dozens of people over several decades in languages they neither speak nor understand? How do the patrons retrieve the intended messages? How does the lack of understanding of the foreign languages affect the accuracy of production and perception of drummed and fiddled speech? How similar is the performers' processing of these languages to second language users' processing? In the processing of a surrogate language produced by some instruments such as the fiddle, music intermediates between the encoder and encoded message on the one hand, and the patrons and what they perceive on the other. The music is strictly neither the intended message produced nor the target of processing, raising further questions about the role of music in the interpretation of the message. At the heart of all these is the implication of processing for linguistic theory. While linguistic theory typically assumes that the target of language processing is its literal meaning, this cannot be the case here, as the practitioners do not necessarily know that meaning. The paper sheds light on these issues and provides a further perspective on language processing. It shows that contrary to widespread belief, there are cases of language use where semantic processing of literal meaning is not the goal.

The Centrality of Meaning in Surrogate Language

In surrogate speech, meaning is encoded directly with musical instruments playing the role of articulatory organs. The nature of the semantic encoding and the level of perfection of surrogate instruments are subject to debate. For some, surrogate instruments do exactly what the speech organs do and communicate (nearly) as much meaning as verbal articulation of speech. Nicholls (1993), for instance, recounts a description of a surrogate instrument in a 1591 publication by a Portuguese explorer named Duarte Lopez to the Kingdom of Kongo as an instrument by means of which users could communicate their thoughts in the same way others would, with words. Finnegan (2012) argues that the signals of a drum language directly represent the words themselves and can even be considered as literature. Chernoff (1980) (cited in Mukuna, 1987) also argues that drum language is not a code but the actual language of the people.

The opposite view sees the output of surrogate instruments as a mere abstraction of the meaning intended by users, and aimed at a privileged few (e.g., Nketia, 1971; Locke and Agbeli, 1981; Carter, 1984; Mukuna, 1987; Neeley, 1996). The direct acoustic output is solely prosodic, unlike speech produced by humans, which has both segmental and prosodic units of speech. Surrogate instruments also encode only some tonal melodies of tone languages, not all, as noted by Nketia (1971). The sequential order of these tonal melodies is also disrupted in instrumental encoding. Mukuna emphasises the fact that surrogate speech is often neither meant for everyday communication nor for the comprehension of the average speaker of the language. For this reason, special attention is often paid to training and preservation of the production and perception of surrogate speech, unlike conventional speech.

The viewpoints presented in these studies raise significant questions about how many of the components of the grammar of these languages are available to the speakers when the languages are encoded instrumentally. They leave no doubt regarding the difficulties and complexities speakers of a language are faced with in the learning, production and comprehension of surrogate speech. All these are within the context of the psycholinguistics of one's native speech that is encoded by surrogate instruments. The report presented in this paper relates to the learning and processing of surrogate speech by speakers who do not speak the languages of the instrument and have no access to any of the components of the grammar of the surrogate language in their mental lexicon.

Fiddling and Talking Drums of Dagbon

DjeDje (2008) lists twenty-five potential surrogate instruments the Dagbamba use, although many of them are used mainly for music, and only marginally serve the purpose of encoding speech. Notable surrogate instruments are the double-membrane hourglass drum, known as *luŋa,* the single-membrane goblet-shaped talking drum (*timpani*), the one-stringed fiddle (*goonje*), the double bell (*dawule*), and the transverse wooden flute (*kikaa*). The *luŋa* is typically drummed in an ensemble to dancers. In isolation or even in a group, it is also used to encode speech. *Luŋa* (plural = lunsi) also refers specifically to the person who drums the hourglass drum. Similarly, *goonje* (plural = *goonjenima*) refers both to the fiddle and the fiddler. The person who plays the *timpani* is known as the Akarima (plural = Akarimanima). The instruments of interest here are the *goonje, timpani*, *dawule* and *kikaa*. These are used to encode speech in Hausa, Akan, Gourmantché and other non-native languages.

The fiddle and talking drum have been part of the culture of Dagbon since the 1700s (see DjeDje 2008 on the fiddle). Oral tradition also has it that the *timpani* was introduced into Dagbon during the reign of Naa Gariba. This was confirmed by two of the drummers interviewed for this study (see *Sources of Data*). Naa Gariba assumed the throne of Dagbon as Yaa Naa in 1740 (Staniland, 1975). The use of the *timpani* and *dawule* for encoding surrogate speech was borrowed from the Asantes of southern Ghana. The names *timpani*, *dawule* and *Akarima* are all loans from Akan *atumpan, dawuro ɔkyerɛma* respectively. The shape, size, and other physical features of the *timpani* and *dawule* are identical to the Akan *atumpan* and *dawuro*. The use of the *timpani* by the Dagbamba mimics its use by the Asantes, with hardly any adaptations. The language encoded by the drum is Akan, the native language of the Asantes (Akans); and messages encoded by the talking drum can also be encoded by the dawule or trumpet, as is the case among the Akan. However, the practice in Dagbon is that, the person who bears the title Akarima only plays the *timpani* and *dawule*. The trumpet is blown by a different person, and considered below the rank of the Akarima.

The *goonje*, (from Hausa *goge*) is used to encode speech in languages (typically Hausa) that the typical fiddler of today and his patrons do not understand. DjeDje (2008) has a detailed comparison between the Dagomba *goonje* and the Hausa *goge*. The ancestors of present-day fiddlers came into Dagbon from Gurmaland, in modern day Burkina Faso. Many believe they were

Hausa who first settled in Grumaland before coming into Dagbon; others believe they were of Gurma ethnicity (see extensive discussion on the different views in DjeDje 2008: 189–196).

All these instruments along with others that are native to the Dagbamba are primarily (palace) court instruments. They play significant roles in defining the life in the palace and the culture of the Dagbamba, including festivals, the installation of chiefs, the performance of funerals and virtually every activity that is characterized by Dagbon culture.

Sources of Data

Six Akarimanima in Yendi, (the capital of the Dagbon Kingdom and seat of the Yaa Naa, the King of Dagbon), and near-by communities were interviewed separately. They include two septuagenarians, Wumbei Dawuni, who performs at the palace of the Yaa Naa, and Wumbei Kwame, who had retired from drumming. Another retired Akarima was Abdulai Yakubu, in his sixties. Two others, Natogma Neindow (the Akarima of the chief of Zohe, a suburb of Yendi), and Akarima Awolu (the Akarima of the Paramount chief of Miɛŋ) were in their fifties. The youngest was Zakaria Alhassan in his mid-twenties, who was the Akarima of Kuga, a suburb of Yendi. None of them could speak or understand any dialect of Akan, and none appeared to know of any Akarima anywhere in Dagbon who could speak Akan. Except for Zakaria, the rest had practiced drumming for periods ranging from 20 to 50 years.

Most of the data on fiddling came from DjeDje (2008), a product of decades of research on fiddling in Dagbon and other ethnic groups in West Africa. I also interacted with Alhassan Sulemana, a native of Yendi and an accomplished fiddler of exceptional talent and decades of experience in fiddling. He was also among the fiddlers DjeDje interviewed extensively for her research and quoted copiously in DjeDje (2008). These sources were complemented by my intuitions as a Dagbana who was born and raised in Yendi. I only needed expert confirmation or perspectives on some of the issues I knew and experienced.

THE ROLE OF THE FIDDLE AND TALKING DRUM IN DAGBON

The Akarima and Goonje are a part of everyday life at the palace. They notify, announce events, narrate history, and send appellations on selected days, at funerals, festivals, installation of chiefs, and in the company of the chief during his royal trips. The Akarima visits the palace at dusk on Sunday, Thursday, and the eve of the day of a festival to remind the chief that the next day is Monday, Friday, or a festival respectively. He returns at dawn the following day – Monday, Friday, or the day of the festival – to notify the chief that it is daybreak. He then drums the chief's praise name and those of his ancestors, dating back several centuries. He is followed by the Goonjenima who go into the inner court of the chief to wake him up with their performance. The Goonjenima also fiddle the chief's praises and those of his relatives and ancestors. In the case of festivals, the Akarima returns after sunrise to notify the sub-chiefs, elders and courtiers that the day has begun and that they must be at the palace. He also welcomes visitors with eulogies.

Events that the Akarima announces include the exit of the chief from the palace, the enskinment of a new chief and his identity, and the death of a chief and notable personalities and their identities. The Goonjenima are the next to take over and perform these functions. The Akarima also summons the sub-chiefs, courtiers, and notable personalities to the chief's palace, especially during emergencies such as war. In the performance of all these functions, he recounts the praise names of the individuals concerned and those of their ancestors. This is preceded or followed by drumming the relevant phrases or sentences. For instance, in announcing death, the Akarima repeats the Akan phrase *Damirfa due* "rest in peace" and similar ones that signal that death has occurred. He then drums the praise names of the deceased or that of their father or the youngest ancestor if the deceased has no praise name, to provide an indication of the house or clan within which the death has occurred. When summoning people to the palace, the phrase *bra ntem* "come in haste" is used and preceded or followed by the praise name of each of the individuals being summoned.

The Goonje entertains the chief and ordinary people during events and cheers on the chief and warriors during wars. According to Alhaji Sulemana, historically, the fiddle was primarily a war instrument, as the very first invitation of fiddlers into Dagbon by the reigning Yaa Naa was primarily to help him fight off an attack on the Kingdom. It has been part of war contingents over the centuries. During wars, they invoke the sense of patriotism in the chief and warriors, remind them of the bravery and exploits of their forbears, and urge them on to emulate them. Alhaji Sulemana further adds that the melodies typically had the opposite effects of mesmerizing their enemies, who were not familiar with them. Akarima Abdulai notes that the Akarima also performs this role in times of war. With the rarity of wars in modern times, the primary duty of the fiddlers has shifted to one of entertainment for the chief and non-royals at every event that brings the Dagbamba together.

The eulogies that are drummed or fiddled are typically proverbs and wise sayings that reflect the philosophy of the people, teach valuable lessons in life, and sometimes challenge the listener to live up to certain standards (see also Salifu 2020). The Goonje melodies are useful "in socializing and controlling the behavior of individuals in the society" (DjeDje, 2008: 214). According to Mahama, one of the fiddlers interviewed by DjeDje and cited in DjeDje (2008), when the Goonje was first introduced into Dagbon, the chiefs had a greater interest in the educational value of the messages in the songs than the melody of the instruments. For all Dagbamba, Mahama says that Goonje is a source of therapy in times of misery, it promotes chieftaincy, adds value to the customs and traditions and improves on the lives of the people with educative messages. Bokor (2014) presents a similar argument on the value of drum language in Africa in general, arguing forcefully that beyond their use for entertainment and rituals, the drums are used for rhetorical purposes to influence social behavior, generate awareness, and call on people to act in specific ways for the good of the society.

When asked whether they sometimes drum in Dagbani, Akarima Wumbei Dawuni's immediate response was "no, there is no Dagbani in it". However, when asked whether they can drum a Dagbani praise name someone chooses for himself, he responded in the affirmative, as did all the other Akarimanima. They mentioned Dagbani and Hausa as languages they often drummed in. Wumbei Dawuni's statement that they do not drum in Dagbani reflects the tradition and norm. Praise names drummed by the Akarima are in Akan, those drummed by the *lunsi* are in Dagbani. Chiefs rarely select a Dagbani phrase or a phrase in any other language beside Akan as an Akarima praise name. The choice of a Dagbani praise name for Akarima will normally be made by a non-royal, or royals who are not chiefs. From my interaction with the Akarimanima, it became apparent that drumming in any other language besides Akan is a recent innovation limited to cases where the demand is made by the one being eulogised.

Similarly, the message of the Goonje is typically in a foreign language, predominantly Hausa, and Gourmantché, although very few fiddlers of today speak any of these languages. Quoting Alhassan Sulemana, DjeDje notes that all but two praise songs honoring Dagbon Kings between 1786 and 1968 (a total of 15 out of 17) are in Hausa. As is the case of the Akarima, modern innovation has seen the incorporation of other languages, including Dagbani, Arabic and English, into the texts (see DjeDje, 2008). In some cases, words of these languages may be code-switched into the text; in others the entire song may be in any of these languages. DjeDje noted that an analysis of about 100 Goonje songs she collected between 1972 and 1974 revealed that more than half were either entirely in Hausa or mixed with Hausa.

THE COGNITIVE ASPECTS OF DAGBAMBA FIDDLED AND DRUMMED SPEECH

How do the Dagbamba, who do not understand Hausa or Akan, process Goonje and Akarima speech in these languages? This section discusses this question.

The Acquisition of Fiddled and Drummed Speech

Almost everyone in Dagbon who understands the praises of the Goonje and the Akarima learns it through explicit instruction. For the average Dagbana, the task is perceiving the tune, knowing its meaning and the personality being praised, and being able to repeat it. For many, knowing the meaning and repeating it are of lesser importance. For Goonje praise songs, because they also double as music for entertainment, some patrons are more focused on enjoying the music and singing it than processing it for meaning. The performers' task is to learn the praise, its meaning, the person praised, as well as the art of producing it instrumentally. Learning to drum or fiddle is the least of their rather daunting tasks, which include the rote memorization of proverbs in a foreign language along with the history and genealogy of hundreds of people over several centuries. The Goonje has an added task of verbal production, as they are required to sing what they fiddle. For these reasons, training an Akarima or Goonje typically starts at a very young age. Five of the six Akarimanima interviewed and some of the fiddlers interviewed by DjeDje indicated that they started learning before the age of 10. The training begins with explicit instruction at home on the praises of the chiefs and other nobility of Dagbon using makeshift drums and mini-sized fiddles. The young ones also accompany their fathers and observe them play their instruments. While by their sides, they are taught the meaning of the praise names that are played, and the people being praised. In addition to practising at home, they also practise on the job and their mistakes are corrected by their elder relatives.

The ultimate sources of the praise names differ for people of different categories. According to Wumbei Kwame, historically, the chiefs got their praise names from their Asante friends. In contemporary times, the chief could ask his Akarima to get him a praise name by giving him an idea of the meaning he wants or leaving it open for the Akarima to decide. The Akarima would then suggest a name after consulting people who understand Akan, for the approval of the chief. Sometimes, the Akarima of a new chief can take this step without a prompting from the chief. Five of the Akarimanima interviewed were unanimous that getting a praise name is the responsibility of the chief, who also decides whether to involve his Akarima. The same rule applies to princes and the noble men and women of Dagbon, who wish to have a name they can be praised with. Until a chief or any other personality gets a praise name for themselves, they are praised with the praise name of their father or youngest ancestor. According to Akarima Awolu, some chiefs who admire their father's praise name may choose to adopt it rather than get one for themselves.

The Production of Fiddled and Drummed Speech

During the interviews, the Akarimanima were requested to orally produce and translate praise names of some chiefs. Their articulation of some of the words was poor, not unexpected of people who do not speak Akan. Although they also knew the meanings of the praises, some of their responses were rough interpretations of these praises, not actual translations. This is not surprising because, as Akarima Awolu put it, "our understanding of the meaning is restricted to what we were told by our fathers and grandfathers". For many of the praise names of the chiefs that came centuries ago, narration of the same meaning to generations of Akarimanima over the centuries evidently led to distortion in meaning.

The most significant observation was the difficulties some of them had with the oral production of the praise names when asked to do so. Retired Wumbei Kwame gave up, saying he needed to first drum them before he could remember how to articulate them orally. He had no difficulty drumming any praise name, but he encountered difficulties verbalizing them. In their profession, their hands are more articulate than their tongues. Akarima Awolu, Abdulai and Zakaria were explicit in saying that in the performance of their work, “it is our hands that hear and speak, not our ears or tongues”.

The Akarima are aware of the similarity between their work and the work of the Goonje. Akarima Abdulai stated, without any prompting, that “we are like the Goonje. It is our hands that hear, not our ears”. The reference to the hands as the perceiver and producer of speech is obviously limited to the drumming of praise names, some of which they do not understand. In summoning people to the palace, announcing events etc., they obviously understand what they drum. In this regard, Awolu was emphatic that “the message of the Akarima comes from the heart but articulated with the hand”.

All six drummers indicated that they could play any praise name they perceived from another Akarima, including those of foreign chiefs. They also admitted that they would not be able to orally say every praise name they drummed. All of them also indicated that they could perceive errors in drummed praise names and correct them with the drum, but they could not always perceive such errors when the praise names are uttered orally. Retired Wumbei Kwame likened his ability to correct errors in other people’s drumming to the ability of a Muslim reader of the Qur’an to correct mistakes in someone else’s recitation, even though the person correcting would not understand the meaning of what he reads. This comparison is of relevance because of the dominance of Islam in Dagbon, where the overwhelming majority practise it, and the religion was integrated into the culture since its introduction into the Kingdom by Yaa Naa Zangina in the 17th Century (Staniland 1975). While all Dagbamba Muslims read the Qur’an, only an insignificant minority who learn the Arabic language understand it. For the rest, Arabic is a liturgical language for fulfilling their spiritual needs.

Dawuni and Awolu recalled being praised by Asantes when they performed together, and the amazement of the Asantes when they got to know that their Dagbamba colleagues could play the instruments with perfection but needed a translator to communicate with them in Akan. Nevertheless, all six Akarimanima agreed that knowledge of Akan enhances the work of the Akarima. Dawuni and Wumbei referenced late Akarimanima who excelled by virtue of their knowledge of Akan.

Unlike the Akarima, the Goonje cannot concentrate solely on the instrumental production, as singing the praise name is part of his tasks. In his interview with DjeDje, Sulemana notes that both the text and melody are of interest to the Goonje and their patrons. However, the accuracy of pronunciation depends on the Goonje’s level of competence in Hausa or how meticulous he was in learning the text. The Goonje who neither understands Hausa nor takes his time to learn the text ends up mispronouncing the text during singing. The result is that, the melody is performed accurately while the meaning is lost in the song.

The Processing of Fiddled and Drummed Speech by Patrons

The patrons of drummed and fiddled speech are the chiefs, people of royal lineage and the commoners. Unlike the preceding discussion on the producers, the difficulty with comprehension of surrogate speech by the patrons is not unique to the Dagbamba. Mukuna (1987) notes that among the Luba in Zaire (Democratic Republic of Congo), proverbs transmitted instrumentally are only understood by trained receivers; but the same proverbs are easily understood by all speakers when spoken. Herzog (1934) is reported by Kaminski (2008) to have noted the poor intelligibility of speech produced by the horn among the Jabo of Liberia. He described it as essentially as technical jargon that is difficult for the average speaker’s comprehension. In his study of the drummed speech by the Ewondo people of Cameroon, Neeley (1996) estimates that only five to ten percent of the 400 residents understand most of the message of the drum, except the commonly drummed phrases, which were understood by a greater percentage. The fact that the transmitted messages of the Akarima and Goonje are in foreign languages not comprehensible to the receivers implies a much poorer level of comprehension. DjeDje describes it as an irony that the importance of the message of the Goonje is not fully appreciated because most Dagbamba do not understand the literal meaning of the lyrics.

Dagbamba chiefs, court officials, people of royal lineage and people with great interest in Dagbon culture fairly understand announcements, summons and the praise names of the kings and notable chiefs who have ruled Dagbon over the centuries. Beyond these, the average Dagbana would have to inquire the meaning of the praise name from the Akarima. From my experience, which was confirmed by the Akarimanima interviewed, it is common for an Akarima to drum the praises of someone of royal lineage and not get any response from them because they do not understand the meaning and do not even know that it is their father or grandfather who is being praised. For this reason, the Akarima also doubles as a teacher of praise names of notable people in Dagbon, as they would call such a person and teach them the praise name and the ancestor who bears it.

The Dagbamba compensate for the lack of understanding of the literal meaning of the message of Goonje by dwelling on the beauty of the lyrics, which could move them to dance. For those with interest in the message but lack enough competence in the foreign language, an alternative, noted by DjeDje, is to link the songs with a symbol or proverb that identifies the individual. DjeDje illustrates this with two praise songs called *Nantoh* (a small poisonous reptile) and *bawuna* (bush cow). They are titles of praise songs for Yaa Naa Yakuba I and Yaa Naa Abdulai I, respectively, and have come to symbolize these Kings in Goonje songs. Their mention in a Goonje song is enough to inform the listener that these Kings are being praised and suffices for the detailed semantic content they miss.

Why Surrogate Languages in a Foreign Tongue?

The low level of semantic and other grammatical content detailed here on the Goonje songs and Akarima speech raises a question on what the motivations of the Dagbamba are for sticking to these surrogates amid several alternatives, including those provided by the Lunsi, and the over a dozen other surrogates. Why are the transmitters not trained to gain enough competence in the respective foreign languages of transmission? What accounts for the people's adherence to the tradition for three centuries? These questions apply more to the Akarima because, in addition to the foreign language content, the Akarima, unlike the Goonje, provides no entertainment to compensate for the lack of comprehension.

The answers to these questions lie in the role of history in the Dagbamba's understanding of culture and the central role of chieftaincy in the definition of history. For the Dagbamba, history is royalty. Virtually every aspect of their history revolves around royalty. The important events in Dagbon history cannot to be told without the role of the chiefs, especially the Yaa Naa, either in facilitating or combatting them. This ranges from events that define the territorial boundaries and cultural practices of Dagbon such as wars, conquests, festivals, religion, and spirituality, to forces of nature like drought and famine. Thus, Dagbon culture is what the chiefs experienced, the actions they took, what they rejected, and what they endorsed over the centuries. Behind every cultural activity that defines the tradition of Dagbon, which is typically linked to a clan, there is a Yaa Naa who played a role in its integration into the everyday life of the Dagbamba. In addition to the warrior clan (*Sapashininima*) (of whom the Akarimanima are part) and the Goonje Clan, we have clans for butchers (*Nakohinima*), blacksmiths (*Machelnima*), drummers (*Lunsi*)[1], barbers (*Wanzama*), the fetish priests (*Tindaannima*), the Islamic priests (*Afanima*) and others. These clans all got incorporated into Dagbon at different periods in its history. Some of these clans, like the Goonje and Akarima are of foreign origin. Unlike the Goonje and Akarima, the use of their heritage languages does not characterize the performance of their distinctive clan activity.

To the Dagbamba, all these clans, by their distinctive activities, are of equal significance in defining the culture and everyday life of the citizens of the Kingdom. Although the Lunsi play their instruments in Dagbani, their performances are not viewed to be of greater significance or benefit than those of the Akarima and the Goonje. As cultural activities that were licensed by a Yaa Naa centuries ago, they are part of what defines royalty, nobility, and statesmanship in Dagbon. The foreign language component contributes to their uniqueness and significance. Any royal, statesman or person of noble decent is required to accept and respect them, regardless of how much of the produced surrogate speech they understand. The maintenance of the languages of performance is part of the intense desire to preserve the cultural heritage of Dagbon.

GRAMMAR IN GOONJE AND AKARIMA PERFORMANCE

In mainstream linguistic theory, the grammar of a language includes the phonetics, phonology, morphology, syntax, and semantics as core components. Language processing takes place at all these levels as well as the interactions between them and at the level of discourse and pragmatics. Cultural and sociolinguistic factors also affect the processing of meaning and interpretation of linguistic units. When viewed under the lenses of these components of grammar and factors influencing language processing, the surrogate speech of the Goonje and Akarima is clearly defective. The only component that is present for all users is the phonetics. The best response supporting this conclusion comes from Akarima Zakaria, who said "...for some of the praises of the ancient kings, we only learn them from our fathers by how they sound, we do not know what they mean, neither can we articulate them".

Knowledge of the phonology and semantics might develop from frequent processing of these praises by speakers who know a lot of them. However, this can only be confirmed after a detailed study. For most users whose understanding of the praises is limited to the few that are of interest to them, the semantics is minimally restricted to the few they understand. The other core components of the grammar – morphology, syntax – can only be present for those who understand the foreign language and are able to use that understanding to parse these praises. Beyond these core components of grammar, what appears to be present is a bit of pragmatics and sociolinguistics. The processing of these praise songs is primarily for the significance it holds within their culture and tradition.

CONCLUSIONS

Meaning is important in the use of surrogate speech in Dagbon. However, meaning and other core components of the grammar do not entirely define surrogate language use. The encoding of meaning and other aspects of the grammar is secondary to the preservation of cultural heritage, the definition of nobility and statesmanship and, to a lesser extent, entertainment. This conclusion partly applies to the use of surrogate instruments to encode speech in native languages. The main difference between Dagbani and these languages is the level of comprehension on the part of the producers. For the average user of these languages with little or no interest in the deeper cultural significance of these instruments, there is little difference between the Dagbamba and the Akan, Hausa etc. This challenges the potency of psycholinguistic theories of language acquisition and processing, especially those relating to second language, to the analysis of surrogate speech in a foreign language. Considering the substantial role of surrogate speech in the lives of the peoples of Africa, the application of models of second language acquisition and processing to surrogate language processing promises to enrich our understanding of language processing in general.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation.

[1]The *Lunsi* Clan is different from the clan of the Akarimanima, who are part of the warrior Clan, though both the Lunsi and the Akarimania are drummers.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Ghana Ethics Committee for the Humanities. The patients/participants provided their oral informed consent to participate in this study. Oral informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The research benefited from research funds awarded to the School of Languages, University of Ghana, by Andrew Mellon Foundation.

ACKNOWLEDGMENTS

My sincere gratitude to Karim Iddrisu for his help in conducting the fieldwork and doing the recordings; and to Alhaji Alhassan Sulemana (AKA Alhaji Goonje) and all the Akarimanima for the time and valuable information they provided me. Many thanks to Prof. Kofi Agyekum for reading and commenting on an initial draft of the paper. All errors are mine.

REFERENCES

Bokor, M. J. K. (2014). When the Drum Speaks. *Rhetorica: A J. Hist. Rhetoric* 32 (2), 165–194. doi:10.1525/rh.2014.32.2.165

Carter, W. G. (1984). *Asante Music in Old and New Juaben*. dissertation. Los Angeles, (CA) University of California.

DjeDje, J. C. (2008). *Fiddling in West Africa: Touching the Spirit in Fulbe, Hausa and Dagbamba Cultures*. Bloomington: Indiana University Press.

Finnegan, R. (2012). *Oral Literature in Africa*. Cambridge, UK: Open Book Publishers. doi:10.11647/OBP.0025

Herzog, G. (1934). Speech-Melody and Primitive Music. *Musical Q.* XX (4), 452–466. doi:10.1093/mq/xx.4.452

Kaminski, J. S. (2008). Surrogate Speech of the Asante Ivory Trumpeters of Ghana. *Yearb. Traditional Music* 40, 117–135. doi:10.2307/S0740155800012133

Locke, D., and Agbeli, G. (1981). Drum Language in Adzogbo. *The Black Perspective in Music* 9 (1), 25–50. doi:10.2307/1214103

Mukuna, K. (1987). Function of Musical Instruments in Surrogate Languages in Africa: A Clarification. *Africa: Revista do centro estudos Africanos* 10, 3–8.

Neeley, P. (1996). Drummed Transactions: Calling the Church in Cameroon. *Anthropological Linguistics* 38 (4), 683–717.

Nicholls, R. (1993). The Language of the Drums: Africa's Traditional Broadcast Technologies. *The World A. Chronicle. our changing era* 8 (4), 244–255.

Nketia, J. H. K. (1971). "Surrogate Languages of Africa" in *Linguistics in Sub-sahara Africa*, Editor. Sebeok, T. A. The Hague: Mouton, 669–732. doi:10.1515/9783111562520

Salifu, A. A. (2020). Dagbamba and Akan Praise Poetry Acts as a Lucid Narrative Platform for Historical and Political Events. *South Afr. J. Afr. Languages* 40 (1), 11–18. doi:10.1080/02572117.2020.1733819

Staniland, M. (1975). *The Lions of Dagbon: Political Change in Northern Ghana*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511759543 CrossRef Full Text

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 14 June 2021
doi: 10.3389/fpsyg.2021.653068

Igbo Speech Surrogacy: Preliminary Findings Based on the Oja Flute

Aaron Carter-Ényì[1], Nnaemeka C. Amadi[2], Quintina Carter-Ényì[3], Charles Chukwudozie[2], Jude Nwankwo[2] and Ebruphiyor Omodoro[2]*

[1] Morehouse College, Atlanta, GA, United States, [2] University of Nigeria Nsukka, Nsukka, Enugu, Nigeria, [3] University of Georgia, Athens, GA, United States

This research report presents analyses of recordings from the Ìgbò culture of southeastern Nigeria of an *ọ̀jà* flute player, a female speaker, and a male speaker. After a prepared performance, the participants completed two tasks: (1) mapping speech to flute playing and (2) identifying phrases played on the flute. Contour analysis is applied to annotated recordings to study the mapping of speech tone and rhythm from voice to instrument in parallel utterances by the three participants (male, female, and flute). Response time between the flute playing and spoken phrase identification indicates each prompt's relative clarity. Using a limited but not predetermined inventory of related praise epithets, participants successfully converted speech to music and music to speech. In the conversion of speech to music, we found that declination was not part of the mapping, indicating it is a phonetic artifact of speech and does not carry a functional load. In identifying surrogate phrases played on the flute (music to speech), we found that dialectical variation caused some misidentification because idioms known in one area of the Igbo dialect cluster are not necessarily known throughout the region. However, *ọ̀ọ̀jà* speech surrogacy is found throughout the region. Possibilities and predictions for further research are presented.

Keywords: speech surrogacy, tone language processing, speech rhythms, music analysis, contour theory, flute acoustics, Nigeria, Igbo

OPEN ACCESS

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Gerald Nweya,
University of Ibadan, Nigeria
Maduabuchi Agbo,
University of Benin, Nigeria

****Correspondence:***
Aaron Carter-Ényì
aaron.carterenyi@morehouse.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *13 January 2021*
Accepted: *18 May 2021*
Published: *14 June 2021*

Citation:
Carter-Ényì A, Amadi NC, Carter-Ényì Q, Chukwudozie C, Nwankwo J and Omodoro E (2021) Igbo Speech Surrogacy: Preliminary Findings Based on the Oja Flute. Front. Psychol. 12:653068. doi: 10.3389/fpsyg.2021.653068

INTRODUCTION

We present preliminary findings from a computer-assisted study of Ìgbò *ọ̀jà* speech surrogacy based on a 30-min participant-observation session recorded at the University of Nigeria Nsukka on November 2, 2020. In the session, an *ọ̀jà* flute player, a female speaker, and a male speaker gave prepared performances. Then, the performance participants were asked by the researchers to complete two tasks:

(1) Mapping of speech to flute playing: the male speaker spoke a phrase praising the woman, the woman repeated, and then the flute played it. The participants repeated the process 16 times. The number of repetitions was not specified when the task was described.
(2) The recognition of phrases played on the flute: the flute player played a common phrase for the male speaker to identify. After a primer, this task was performed 18 times.

The Ìgbò language is spoken in southeastern and southern Nigeria, primarily in the states of Abia, Anambra, Ebonyi, Enugu, and Imo. It belongs to the Benue-Congo branch of the Niger-Congo family. Ìgbò is a two-tone language with downstep. Because of the relatively small number of tone levels–two levels, while many in Nigeria have three or more levels–Maddieson (2013) classifies it as a simple tone system in the World Atlas of Language Structures. However, previous research on Ìgbò and Yorùbá suggests that the functional load of tone in Ìgbò may be higher than that of Yorùbá. A comparison of two widely available dictionaries [Williamson's (1972) Ìgbò Dictionary and the University of Ibadan's Yorùbá Dictionary] revealed that 60% of disyllable entries formed minimal pairs in Ìgbò. In contrast, only 48% of disyllable entries in Yorùbá formed minimal pairs (Carter-Enyi, 2016).

Ọ̀jà is a small wooden high-pitched flute, approximately seven inches (18 cm) in length, indigenous to the Ìgbò people (Nwachukwu, 1997). Its usage is vast, but Lo-Bamijoko (1987) notes that it is used "more for chanting than for singing." Lo-Bamijoko (1987) defines chanting as an "extended form of speaking," more commonly known as speech surrogacy. The *Ọ̀jà* is played for the Ígweē (traditional ruler), notable chiefs or influential people in the community for entertainment, praise-singing, or relaxation. It may also take on a more important role during life-cycle celebrations such as naming ceremonies or marrying a new wife. The *ọ̀jà* may be seen as the soul of Ìgbò cultural music. The instrument is used to sing laments for the dead. In Ìgbò myths, the *ọ̀jà* is believed to possess spiritual power capable of even raising the dead. It is played for the *mmoọnwu* (masquerade representing spirit manifestation) as a morale booster during public displays. In recent times, composers use the instrument for soundtracks of Nollywood movies. It is sometimes described as "the oil with which Ìgbò music is eaten." The sound energizes the weak and calls up the very aged to jump up in strength as they dance to its calls. In summary, *ọ̀jà* is a musical instrument of immense cultural significance among the Ìgbò people. In recent years, Christian Onyeji (2006, 2016) of the University of Nigeria Nsukka has advocated for the *ọ̀jà* and other Igbo instruments as mediums for art music composition. Nwachukwu (1997) is a detailed acoustic and organological study of the instrument.

Like Lo-Bamijoko, we assume that the "chanting" (language-based) mode of the *ọ̀jà* is primary to the instrument's performance practice. Our aim was to collect data on the chanting mode, which would serve for further explorations of this under-studied genre. We recorded a participant-observation session to examine the mapping of speech to flute and flute to speech. This brief research report summarizes the findings from the computer-assisted analyses of these recordings.

MATERIALS AND METHODS

CC is a retired lecturer in African Studies at the University of Nigeria Nsukka. He is known in the community by nicknames, including "Akionu" and "Member." Smartphone videos of his dancing are popular on social media[1]. He hosts a weekly radio show on the university radio station (Lion FM) on Ìgbò culture entirely in the Ìgbò language.

Mr. Chukwudozie approached Aaron Carter-Ényì about recording an ọ̀jà performance for the Africana Digital Ethnography Project (ADEPt, radar.auctr.edu/adept). Three videos are available on YouTube from the recording session: (1) Igbu Ọ̀jà (Playing the Flute): "Igwe O, Igwe[2];" (2) Ìgbò Woman praised by the Ọ̀jà flute[3]; and (3) Ọ̀jà Phrase Identification[4]. In the recorded performance, Bartholomew Ogbu (*ọ̀jà* player), Chinyelum Ewelum, and Mr. Chukwudozie demonstrated *ọ̀jà* praise-singing for an Ìgwē (king) and Lọ̀lọ̀ (queen), observed in videos 1 and 2, respectively. The prepared performance included the flute "speaking," playing common praise phrases on the flute by replicating the pitch and rhythm contours of speech.

After the group completed their prepared performance, they were asked to perform specific tasks. Although the prepared performance demonstrated the *ọ̀jà's* capacity for surrogacy, articulating words such as "Ígwē" (king), the researchers deemed it necessary to have a closer comparison of equivalent phrases. Because the performance participants were aware of the *ọ̀jà's* capacity for surrogacy, they could respond to specific tasks involving the transfer of speech to music and music to speech. The two tasks were (1) mapping speech to flute playing and (2) identifying phrases played on the flute. We describe the performance processes in detail with analysis in the following sections.

For Videos 1 and 2, CC handwrote the transcriptions and translations of video excerpts. NA and Ugonna Okonkwo entered his written text as timed-text captions in YouTube Studio. Quintina Carter-Ényì completed all stages of language annotation for Video 3. The timed-text Igbo and English captions completed in YouTube Studio were then downloaded in the sub-rip title format (.srt) and imported into ELAN. However, this report focuses on the pitch and time domain, not segmental phones (phonemes). Aaron Carter-Ényì made additional annotations in Celemony's Melodyne Editor (see **Figure 1**), which encodes pitch and timing information in MIDI format (where C4 = 60, C5 = 72, etc.). MIDI data (.mid) is interoperable with many software from MATLAB to Logic Pro. The Melodyne annotation was the final stage of annotation.

RESULTS: TASK 1: CROSS-DOMAIN MAPPING OF SPEECH TO MUSIC

In discussion with the participants, we agreed that the "king" (Chukwudozie) would speak a praise epithet (e.g., "Ńné múrú-ọ̀rà" meaning "mother of a multitude"), the "queen" (Ewelum) would repeat it, and finally, the flutist (Ogbu) would play it. Everyone knew that the phrases would praise women, especially mothers. Every phrase belonged to a standard inventory of praise

[1] https://twitter.com/Gidi_Traffic/status/1237310343266938880
[2] http://hdl.handle.net/20.500.12322/adept.ibo:0013
[3] http://hdl.handle.net/20.500.12322/adept.ibo:0014
[4] http://hdl.handle.net/20.500.12322/adept.ibo:0015

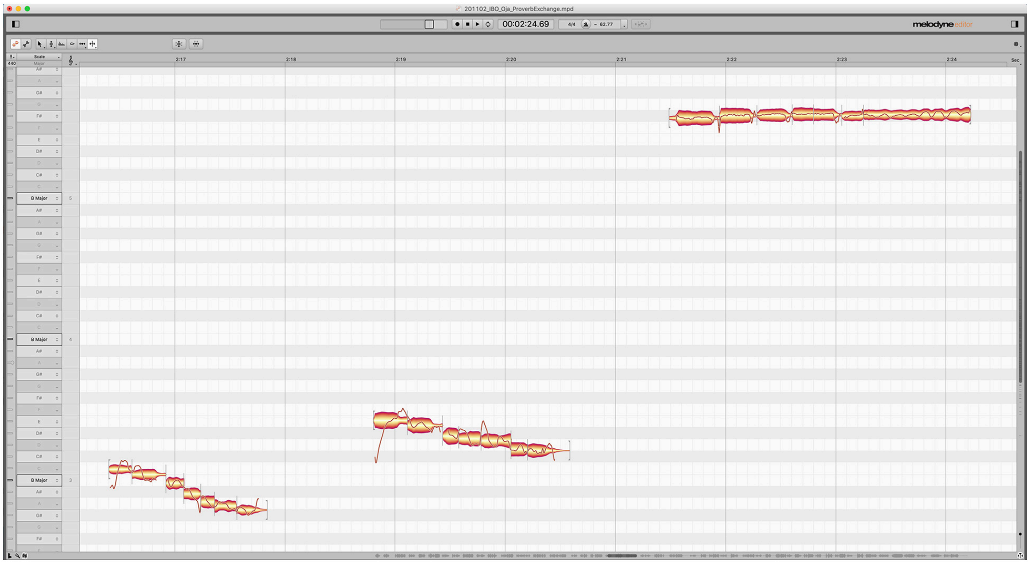


FIGURE 1 | Task 1–Phrase 7 in Melodyne with male voice **(left)**, female voice **(middle)**, and *ọ̀jà* **(right)**.

epithets (see Barber, 1991). The participants did not rehearse the phrases or their order. This task is available on YouTube as "Ìgbò Woman praised by the ọ̀jà flute (see footnote text 3)."

The video starts with an *ọ̀jà* introduction (timecode 00:00) and is followed by vocalizations by Chukwudozie portraying the Ìgwē (timecode 01:02). The praise sayings begin at timecode 01:20. All 16 of the phrases are in celebration of motherhood. Women in Ìgbòland are celebrated and praised because they are seen as the pillars of the home. The woman's ability to manage the household and her husband's wealth is all captured in the praises. The phrases highlight the woman's qualities and characteristics, including the woman's ability to bear, breastfeed, and raise a child. They also refer to the physical attributes of women, such as beauty and shapeliness.

We analyzed data from the first task to evaluate the similarity of the pitch and rhythm content between the three versions (male, female, and flute) for each of the 16 phrases. Specifically, an implementation of musical contour theory was applied to computer-assisted melodic transcriptions produced using Melodyne software. A script written in MATLAB calculated Quinn's (1997) contour similarity to assess the similarity of the male speaker's speech and the *ọ̀jà* interpretation (script attached, "Frontiers.m"). This analysis addresses the mapping of speech tone and rhythm to *ọ̀jà* playing through comparing combinatorial matrices of pairwise pitch height comparisons (or segment durations in seconds in the case of rhythm).

In **Figure 2**, at the left, each pitch height (60 = C4/Middle C) is compared to every other pitch height in the melodic segment. In this case, there are seven "notes." The "melody" of the male voice speaking the phrase starts at 60 (C4) and gradually descends to 56 (Ab3). Quinn's method (1997) codes binary pairwise comparisons as "1" for greater than and "0" for less than or equal. We compared the notes at the top of the columns to the notes along the left side. This yields self-comparisons (0 for equal) along the central diagonal from the top left to bottom right. The middle matrix is for the *ọ̀jà* interpretation. All of the notes are at the same pitch height (90 or F#6). All of the pairwise comparisons are "0" because all of the notes are equal. The rightmost matrix measures similarity between the speech matrix and music matrix. In this case, "1" indicates a match, and "0" indicates a non-match for each respective cell. Out of the 16 phrases recorded, annotated, and analyzed, Phrase 7 (see **Figures 1**, **2**) had the lowest pitch contour similarity between speech and music, 61.2%. Notably, the phrase is on a single tone "Ónyé áká ghárá-ghárá" (all high tone level). This result suggests that declination is not necessarily part of the mapping from speech to music. The contrast in phrase declination between speech and music can be observed aurally by listening closely to Phrase 7 in the audio of the YouTube video linked above (timecode 02:16) and visually by examining the Melodyne transcription in **Figure 1**.

Although flute playing is studied here, not singing, this exclusion of declination effects is not limited to surrogacy. Chanted or sung realization of tones may also avoid declination (Carter-Enyi, 2016). The single-tone phrase; none of the other phrases consisted of a single tone. 11 out of 16 of the phrases had high pitch contour similarity, with above 80% of pairwise comparisons matching (see the second to last column, "Pitch," in **Table 1**). The difference between the phrase declination characteristic of speech and the stable pitch height characteristic of music is not so pronounced in a phrase where the speech tone is more varied and the contour more complex.

We also applied contour similarity metrics to rhythm, specifically duration in seconds (the last column, "Rhythm," of **Table 1**). **Figure 3** shows the application to segment durations in hundredths of a second. Similarly, the binaries are coded as "1" when the column duration is greater than the row duration, or "0" if it is equal to or less than for the first two matrices (speech on the left and *ọ̀jà* in the middle). On the right, we compare the

	60	60	59	58	57	57	56
60	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0
59	1	1	0	0	0	0	0
58	1	1	1	0	0	0	0
57	1	1	1	1	0	0	0
57	1	1	1	1	0	0	0
56	1	1	1	1	1	1	0

	90	90	90	90	90	90	90
90	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0

1	1	1	1	1	1	1
1	1	1	1	1	1	1
0	0	1	1	1	1	1
0	0	0	1	1	1	1
0	0	0	0	1	1	1
0	0	0	0	1	1	1
0	0	0	0	0	0	1

FIGURE 2 | Contour matrices for the pitch of Task 1–Phrase 7 with male speech demonstration **(left)**, *ọ̀jà* interpretation **(middle)**, and similarity matrix **(right)**.

TABLE 1 | Results for Task 1 (speech to music).

#	Time	Ìgbò transcription	English gloss	Segments	Pitch	Rhythm
1	1:19	Nìné múrú-ọ̀rà	Mother of a multitude	6	0.917	0.611
2	1:27	Álá nàzù ńwā	Breast that feeds the babies	6	0.917	0.389
3	1:35	Óchìé díkē-ńnēm	My strong mother	5	0.880	1.000
4	1:44	Éléléb'úkwù è gbù'éwū	The waist that deserves to be celebrated with the killing of a goat	7	0.918	0.959
5	1:54	Ọ̀zùlù éké, zùọ̀ Óriè, zùọ̀ Áfọ̀, zúọ̀ Ǹkwò	She sells on Eke, Orie, Afor, and Nkwo market days	10	0.740	0.800
6	2:07	Ńné ńnọ̀ọ̀má	Mother, good mother	5	0.880	0.840
7	2:16	Ónyé áká ghárá-ghárá	A resourceful person	7	0.612	0.796
8	2:25	Ọ̀mụ̀rụ̀ Órìe, mụ́ọ́ Àfọ̀, mụ́ọ́ Ǹkwọ́	She gave birth to Orie, Afor and Nkwo	8	0.766	0.563
9	2:37	Ọ̀dị̀ dí yé ḿmā	Her husband's delight	6	0.639	0.278
10	2:44	Ódózī àkụ̀ dí yá	The manager of her husband's wealth	7	0.898	0.592
11	2:53	Ùgòó dí yá	The glory of her husband	5	0.760	0.520
12	3:01	Ézē nwáānyì	Queen mother	4	0.875	0.750
13	3:08	Ọ̀chílụ̀ ọ́zụ́ọ́	Trainer of all	6	0.889	0.611
14	3:16	Ágbàrà k'ìibeè yà	A woman greater than other women	6	0.944	0.778
15	3:24	Ńné ọ̀mááàá!	A virtuous woman	3	1.000	1.000
16	3:33	Iì gàdị́ ńdụ́ rúé m̀gbè ébiÌghèbì	You will live forever and ever	10	0.880	0.540

	0.31	0.24	0.16	0.33	0.73
0.31	0	0	0	1	1
0.24	1	0	0	1	1
0.16	1	1	0	1	1
0.33	0	0	0	0	1
0.73	0	0	0	0	0

	0.32	0.28	0.23	0.40	1.30
0.32	0	0	0	1	1
0.28	1	0	0	1	1
0.23	1	1	0	1	1
0.40	0	0	0	0	1
1.30	0	0	0	0	0

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

FIGURE 3 | Contour matrices for the rhythm of Task 1–Phrase 3 with male speech demonstration **(left)**, *ọ̀jà* interpretation **(middle)**, and similarity matrix **(right)**.

first two matrices, yielding all "1" values because all entries in the first two matrices match each other, yielding 100% similarity.

RESULTS: TASK 2: PHRASE-LEVEL IDENTIFICATION OF Ọ̀JÀ SURROGACY

Task 2 is available on YouTube as "Ọ̀jà Phrase Identification (see footnote text 4)." The second task assesses the intelligibility of speech surrogacy on *òòjaÌ*. Response time between the end of the flute phrase and the phrase's identification by the speaker (Chukwudozie) indicates the ease with which the speaker can identify each specific phrase. The response time is measured as the time between the end of the *ọ̀jà* phrase and the start of the speaker's identification. In general, the speaker quickly identified the phrase demonstrated by the *ọ̀jà*. Chukwudozie correctly identified 14 out of 18 phrases with a mean response time of 0.5 s. The *ọ̀jà* player had to prime the topic twice (the first two phrases) before the Ígwē (Chukwudozie) could start identifying. Only three phrases were misidentified. Of particular note is "Óbòdò dìkē o!" which means "Strong city," when the *ọ̀jà* player intended "Peace be with you" consistent with utterances correctly identified later in the task. Also, dialect seems to be a factor because "Déèjé nù o!" (which is a common phrase in Enugu state) required some prompting from the *ọ̀jà* player (from Enugu state) for the Iìgwē (from Anambra state) to recognize it.

DISCUSSION

While the Yorùbá *dùndún* (talking drum) is the most iconic speech surrogate in Nigeria, perhaps West Africa, speech surrogates are found in many other Niger-Congo cultures, notably the Ìgbò. Awareness of speech surrogacy is embedded in Ìgbò culture and practiced on several indigenous instruments, notably the *ọ̀jà* (small wooden flute) and *ùfiè* (large log drum). Our preliminary study of the *ọ̀jà* suggests that the mapping from speech to music is more easily accomplished than the mapping from music to speech. While we cannot generalize this finding based on one small group of participants, this outcome is logical because there is a loss of information in the mapping from speech to music, namely the segmental phonemes. The recognition of the musical phrases as speech requires the reconstruction of missing information from an inventory of known idioms. It is likely that the speech phrases come from a limited inventory and may need to be associated with musical mapping through experience. They may not be recognizable to fluent speakers without significant cultural experience with surrogacy and ọ̀jà surrogacy specifically.

We found that there is considerable precision in the representation of both pitch and rhythm in Ìgbò *ọ̀jà* surrogacy, similar to Seifart et al.'s (2018) study of Amazonian Bora drumming. Tonal stability across Igbo dialects was first proposed by Emananjo (1978). According to Clark (1990), variation in segmental phonemes (such as/r/and/l/) is common between dialects, but tonemes are usually consistent in analogous phrases. Toneme consistency made it possible for the *ọ̀jà* artist (from Enugu) to communicate on his instrument with a man from Anambra state. However, when the flute spoke a common phrase in the Enugu dialect, Chukwudozie (the respondent) did not readily identify it. Observation and analysis of Task 2 suggest some characteristics of Ìgbò *ọ̀jà* speech surrogacy. Toneme consistency across dialects makes it possible for *ọ̀jà* speech surrogacy (which represents the pitch and rhythm of speech) to be communicative across dialects. Because it is idioms that are usually "spoken" by instruments, unless the idiom is known across dialects, the pitch and rhythm pattern will not be familiar.

Regarding Task 1, we observe that speech declination is not part of the mapping from speech to music. This result is consistent with observations of singing (Carter-Enyi, 2016) and suggests that declination is purely an effect of production, which does not seem to affect intelligibility. This cumulative evidence supports the position that declination is not a phonological aspect of language even in a "terraced" tone language such as Ìgbò.

Based on 6 min of recordings, these preliminary findings provide a basis for future research predictions. However, much more work must be done to determine the extent to which these observations may be generalized to other instruments and even different cultures. If we conducted a more extensive study with multiple participants responding to Task 2 (identification of surrogate phrases), we predict that participants will most readily identify stereotyped phrases. Likely, single words out of context cannot be identified. Even phrases without a larger context (e.g., a topic such as praise of a woman or king) are hard to identify.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the Africana Digital Ethnography Project (ADEPt) collection of the Repository of AUC Digital collections, Archives and Research: https://radar.auctr.edu/adept.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Nigeria Nsukka, International Directorate. The participants provided their written informed consent to participate in the audiovisual recordings collected for this study. Written informed consent was obtained from the individuals for the publication of the audiovisual recordings referenced by this article.

AUTHOR CONTRIBUTIONS

ACÉ conducted the participant research, recorded and edited the videos, did the musical transcription, completed the data analysis, and drafted the article. QCÉ completed transcription and translation of one of the two videos analyzed, and also interpreted the data qualitatively as well as discussing extensively with the lead author. JN conducted ongoing speech surrogate research with the lead author and contributed text on both Igbo culture and the Oja flute in Igbo culture. NA prepared captions and the description for one of the videos which became part of the basis for a brief section of the report. EO assisted the lead author in recording the session and communicating with the participants. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Endowment for the Humanities, United States (Grant No. PW-264175-19) for research participants.

ACKNOWLEDGMENTS

CC (king), Chinyelum Ewelum (queen), Bartholomew Ogbu (*ọ̀jà*) were the performers. ACÉ and EO were the videographers. NA, QCÉ, CC, and Ugonna Okonkwo were the annotation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.653068/full#supplementary-material

REFERENCES

Barber, K. (1991). *I Could Speak Until Tomorrow: Oriki, Women and The Past in a Yoruba Town*, Vol. Vol. 7. Edinburgh: Edinburgh University Press.

Carter-Enyi, A. (2016). *Contour Levels: An Abstraction of Pitch Space based on African Tone Systems*. Ph.D. thesis. Ohio: Ohio State University.

Clark, M. M. (1990). *The Tonal System of Igbo. Dordrecht, Holland*. Providence, RI: Foris Publications.

Emananjo, E. N. (1978). *Elements of Modern Igbo Grammar*. Ibadan: Oxford University Press.

Lo-Bamijoko, J. N. (1987). Classification of Igbo musical instruments, Nigeria. *Afr. Music J. Int. Libr. Afr. Music* 6, 19–41. doi: 10.21504/amj.v6i4.1259

Maddieson, I. (2013). "Tone," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Nwachukwu, M. A. (1997). On the resonant frequencies of the oja. *Nigerian J. Technol.* 18, 1–21.

Omojola, B. (1995). *Nigerian Art Music: With an Introduction Study of Ghanaian Art Music*, Vol. 47. Nairobi: Institut Français de recherche en Afrique.

Onyeji, C. (2006). "Oja (Igbo wooden flute): an introduction to the playing technique and performance," in *Centering on African Practice in Musical Arts Education*, ed. M. Mans (Cape Town: African Minds). doi: 10.1163/9789004425842_002

Onyeji, C. (2016). "Chapter five: playing technique and contemporary compositions for the Oja (Wooden Flute)," in *Music and Social Dynamics in Nigeria*, ed. B. Omojola (New York: Peter Lang). doi: 10.1017/cbo9780511552427.004

Quinn, I. (1997). Fuzzy extensions to the theory of contour. *Music Theory Spectr.* 19, 232–263. doi: 10.2307/745755

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: amazonian Bora drummed language exploits speech rhythm for long-distance communication. *R. Soc. Open Sci.* 5:170354. doi: 10.1098/rsos.170354

Williamson, K. (1972). *Igbo English Dictionary*. Benin City: Ethiope Publishing Corporation.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 16 June 2021
doi: 10.3389/fcomm.2021.653747

Feeling the Beat in an African Tone Language: Rhythmic Mapping Between Language and Music

Kathryn H. Franich[1]* *and Ange B. Lendja Ngnemzué*[2]

[1]University of Delaware, Newark, DE, United States, [2]Paris 8 University, Paris, France

Text-setting patterns in music have served as a key data source in the development of theories of prosody and rhythm in stress-based languages, but have been explored less from a rhythmic perspective in the realm of tone languages. African tone languages have been especially under-studied in terms of rhythmic patterns in text-setting, likely in large part due to the ill-understood status of metrical structure and prosodic prominence asymmetries in many of these languages. Here, we explore how language is mapped to rhythmic structure in traditional folksongs sung in Medʉmba, a Grassfields Bantu language spoken in Cameroon. We show that, despite complex and varying rhythmic structures within and across songs, correspondences emerge between musical rhythm and linguistic structure at the level of stem position, tone, and prosodic structure. Our results reinforce the notion that metrical prominence asymmetries are present in African tone languages, and that they play an important coordinative role in music and movement.

Keywords: African tone languages, music, rhythm, prosody, metrical structure

OPEN ACCESS

Edited by:
Laura McPherson,
Dartmouth College, United States

Reviewed by:
Lev Blumenfeld,
Carleton University, Canada
Rosalía Rodríguez-Vázquez,
University of Vigo, Spain

Correspondence:
Kathryn H. Franich
kfranich@udel.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *15 January 2021*
Accepted: *21 April 2021*
Published: *16 June 2021*

Citation:
Franich KH and Lendja Ngnemzué AB (2021) Feeling the Beat in an African Tone Language: Rhythmic Mapping Between Language and Music. Front. Commun. 6:653747. doi: 10.3389/fcomm.2021.653747

INTRODUCTION

Music has historically served as an important, if somewhat overlooked, source of information on the prosodic organization of language. In particular, patterns of *text-setting*—or the ways in which speakers of a language opt to align text to music—have played an important role in shaping analyses of metrical structure in language (Liberman 1975), syllable weight effects (Vance 1987), and syntactic/prosodic phrasing (Lehrdal and Jackendoff, 1983; Lerdahl 2001; Dell and Halle 2009; Turpin and Laughren 2013). Indeed, some aspects of linguistic theory itself are deeply inspired by aspects of music theory, including the characterization of syllable stress as a manifestation of "prominence" interpreted through a hierarchically organized system of alternating strong and weak beats (Liberman 1975; Liberman & Prince 1977; Kiparsky, 1979; Prince 1983; Kiparsky, 2016). An example from English which illustrates the link between word-level metrical structure and musical structure is given in **Figure 1**. Many native English speakers will intuit that the two words *gérbil* and *giráffe*, which differ crucially in terms of their stress placement (initial syllable vs. final syllable), require different rhythmic structures when incorporated into the song, as demonstrated by the two different rhythmic patterns in the second measure of the song. In this particular rendition, the singer's objective is to place the stressed syllable in either word on an odd-numbered main beat of the

Abbreviations: 1, 1st person; 2, 2nd person; 3, 3rd person; ASSO, associative; CERT, certitude; CL, noun class; COP, copula; COMP, complementizer; DISC, discourse particle; EXCL, exclamative; FOC, focus; H, high tone; ITER, iterative; L, low tone; NEG, negation; OBJ, object; PL, plural; PROX, proximal; REDUP, reduplicant; SG, singular; SUBJ, subjunctive; Q, question particle; VOC, vocable.

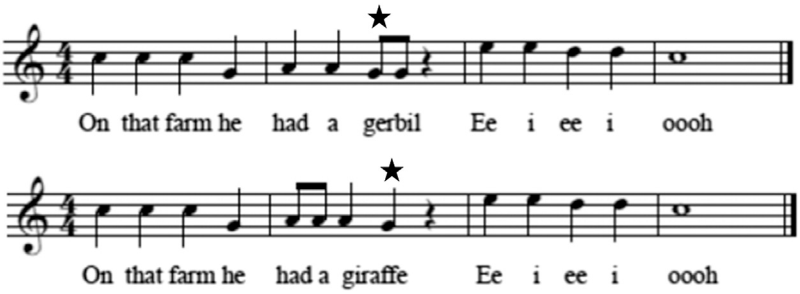


FIGURE 1 | Rhythmic strategies corresponding to *gerbil* and *giraffe* for the song Old MacDonald.

4/4 timed measure, which corresponds with the sense of where the "strong" beat in the musical line is located. Note that readers may intuit that there are other possible rhythmic strategies for the song beyond what is represented in **Figure 1**; the key observation is simply that lexical stress placement predicts the two words should be arranged differently within the rhythmic line; in either case, the stressed syllable of the target animal word is placed on the third beat of the second measure (marked with a star above it).

While the stringency of the constraints governing the mapping of stressed syllables to specific musical beats varies across languages and musical genres, stress nonetheless appears to play an important role in text-setting across a range of languages and musical traditions. For example, evidence from Spanish children's songs presented by Morgan and Janda (1989) demonstrates a tendency for musical downbeats (i.e., the first beat of each measure) to be matched with stressed syllables. They also demonstrate that mismatches between stress and beat strength are tolerated in cases where conflicts arise between a word's syllable count and its stress pattern; in other words, the musical rhythmic structure of a song line is less likely to be altered to accommodate stress patterns, as is commonly done in English songs (e.g., in **Figure 1** for *Old MacDonald*). Similarly, Temperley and Temperley (2012) show that French early 20th century folk songs display quite close alignment between linguistic metrical structure and beat strength, including the regular cooccurrence of musically strong beats with word-final syllables, which are stressed in French (Schane 1968), and the occurrence of metrically weak function words in musically weak positions. Despite this trend, they demonstrate that French songs display a greater degree of mismatch between stress and musical beat strength than English songs. The importance of stress in text-setting can also be found for languages outside of the Indo-European family. For example, Fitzgerald (1998) finds that stress placement is highly regulated in Tohono O'odham traditional songs, with stress obligatorily occurring at the beginning of every song line and prohibited from occurring at line ends. These text-setting rules appear to be quite strict, with various grammatical operations, including vacuous reduplication, employed in the songs in order for the rules to be satisfied.

Cross-linguistic variation in prosody and speech timing appear to be at play in determining some of the observed variation in the strictness of mapping between stress and musical beat structure. For example, Spanish and English have been argued to differ in their speech timing patterns, with Spanish exhibiting more even timing between consecutive syllables (*syllable timing*), and English exhibiting variable timing between syllables (primarily due to unstressed syllable reduction), but more even timing between consecutive stressed syllables (*stress timing*) (Pike 1945, Abercrombie 1967; Grabe and Low 2002; Ramus 2002)[1]. While English songs are found to condense and stretch syllable durations to occupy different musical beat ratios in order for stress to align with strong musical beats, mirroring the process of foot-internal shortening found more generally in English and other stress-timed languages (Fowler 1977; Kim and Cole 2006), this practice is not common in songs in Spanish and other syllable-timed languages[2] (Huron and Ollen 2003; Patel and Daniele 2003).

Not all deviations from stress-beat alignment are attributable to cross-linguistic or cross-dialectal variation in prosody, however. For example, Nancarrow (2010) describes a creative device in Lardil *burdal* songs in which repeated lines will be produced first in their default form with alignment between stressed syllables and musical beat structure, and then again with stress shifted to normally unstressed positions within the text. Temperley (1999) describes a similar phenomenon of *syncopation* in English-based rock music, which involves the characteristic 'misplacement' of stressed syllables onto musically weak beats. Liberman (2007) provides an alternative interpretation of syncopation as involving an underlying *polyrhythm*, such that stressed syllables, rather than being paired with weak beats, are paired with beats which are actually strong if interpreted relative to a competing rhythmic line which is in opposition to the song's primary rhythmic pulse. An example which Liberman draws on is from the Afro-Cuban "habanera rhythm" (also commonly known among music theorists as the *clave* rhythm), which is shown to involve a 3 + 3+2 beat pattern (contrary to the "square" 4 + 4 pattern found across many musical traditions, including in European classical and folkloric music). Indeed, Mead (2007) shows that

[1]Note that many varieties of English show patterns more consistent with syllable-timing (Mesthrie & Bhatt 2008).

[2]Note that Spanish does display some patterns of rhythmic adjustment, as in cases of synalepha (metrically driven vowel elision), which frequently take place within sequences of function words (Espinosa 1924).

stress placement in music featuring this rhythm is driven both by attraction to typically strong beats within a 4 + 4 meter, as well as (to a lesser extent) by the placement of rhythmically accented beats within the habanera pattern.

Thus, stress-beat alignment strategies vary as a function of language and as a function of rhythmic structure within a given musical tradition. In spite of this variability, the role of stress itself in guiding text-setting practices is robust across a wide variety of languages and contexts. Given the importance of stress in determining rhythmic alignment in text-setting, an interesting question concerns whether languages which lack clear evidence of stress—as is the case for many lexical tone languages—also show evidence for rhythmic constraints on text-setting, and if so, how these constraints can be characterized. While correspondences between tones and musical melody have been investigated in some depth in a variety of tone languages, including Cantonese, (Chan 1987; Wong and Diehl 2002), Dinka (Ladd and Kirby 2020), Fe'Fe' (Proto 2016), Hausa (Richards 1972), Mandarin (Chan 1987; Wee, 2005), Shona (Schellenberg, 2009; Schellenberg, 2012), Thai (Ketkaew and Pittayaporn 2014), Tommo So (McPherson and Ryan 2018), Vietnamese (Kirby and Ladd 2016), and Zulu (Rycroft 1959), rhythmic considerations in these languages have received less attention. Interestingly, in some of these languages, tone and rhythm appear to interact in terms of influencing text-setting constraints. In Mandarin Chinese, for example, various studies have suggested that tones occurring on metrically prominent syllables[3] (sometimes termed "prosodic heads") are more resistant to tone sandhi phenomena, whereas non-prominent syllables are more flexible in this regard. In a text-setting study of Mandarin songs, Wee (2015) finds that musical beat structure has a similar influence on tonal melody: while there is a close correspondence between musical melody and phonemic tone value on rhythmically strong beats in the music (a phenomenon Wee refers to as "tonal integrity"), rhythmically weak beats are more tolerant of melody-tone mismatches. Similar findings have been reported for so-called "pitch-accent" languages, such as Japanese (Cho 2017).

Investigations into the role of linguistic rhythm in text-setting have been even sparser among African tone languages[4]. There are likely various reasons for this gap. First off, there has been considerable debate over the existence of stress in many of these languages (see Downing (2004), Downing (2010) for comprehensive overviews); indeed, some researchers have gone so far as to question whether it is even necessary to posit metrical structure in such languages at all (Odden 1999; Hyman, 2014; Hyman, 2015). While several studies have identified possible correlates of metrical structure in African languages beyond canonical stress cues, such as positional restrictions on consonant and vowel contrasts (Akinlabi and Urua 2003), vowel harmony patterns Pearce (2006, 2007), tone spreading patterns (Downing 1990; Leben 2002, 2003; Weidman and Rose 2006; Green 2015), and other morphophonological patterns (Dimmendaal 2012; Green 2015), it has been unclear until recently whether any of these patterns can be linked with rhythmic prominence similar to that found in stress languages[5]. Another reason why rhythmic text-setting studies on African music may be uncommon is that the rhythmic structure of music in many African cultures—including those of broad swaths of West and Central Africa—is characterized as polyrhythmic, meaning that several rhythms may be at play in a given piece of music at one time. For researchers unfamiliar with such rhythms, the task of locating a central "beat" on which to base an analysis may prove challenging.

Given this gap in the literature on rhythmic text-setting in African music, the present work seeks to provide an analysis of text-setting patterns in one particular language and musical context, that of Medʉmba folk songs from the West Region of Cameroon. Medʉmba is an interesting language to examine due to the fact that it shares with several other African tone languages positional restrictions on vowel and consonant contrasts which are characteristic of metrical prominence alternations; recently, (Franich, 2017; Franich, 2018; Franich, 2021) has shown that these asymmetries are linked with rhythmic timing, similar to stressed syllables in languages which have stress. We ultimately show that there are a number of restrictions on how syllables in Medʉmba can be rhythmically mapped to song which depend on such factors as position of a syllable within the stem, tone, and morpho-prosodic status.

Another aspect of the present work which will be of theoretical interest is the fact that Medʉmba folksongs bear the hallmark polyrhythmic structure of West and Central African music, allowing us to investigate how distinct rhythms within the music align with linguistic structure. Our analysis reveals that mapping between metrically prominent syllables and strong beats is flexible across different songs, and even within songs, depending on rhythmic mode. Nonetheless, mapping between linguistic structure and beat structure is far from random: we show that certain syllable types pattern together consistently within a rhythmic mode, whether they are being treated as "prominence-attracting" or not. We also show that some apparent deviations in the mapping of linguistic prominences to strong beats can be understood to constitute an alternative mapping of these elements to competing rhythmic patterns which are reinforced by the drummers within the ensemble (the master drummer, in particular).

[3]While there has been considerable debate as to the existence of stress in Mandarin, a growing body of research supports the idea that metrical prominence asymmetries do exist, even if not marked phonetically by stress (Chao 1948; Yip 1980; Duanmu 1990; Duanmu, 1996; Duanmu, 2004; Moore 1993).

[4]Notable exceptions to this generalization include Schuh (2010) and Schuh (2014) which investigate rhythmic factors such as syllable weight in text-setting patterns in Hausa and other Chadic languages.

[5]Some research within music theory has suggested a role for syllable stress in text-setting in African music (e.g. Agawu 1995; Temperley 2000), but these studies tend not provide a clear linguistic definition for stress.

FIGURE 2 | Photos of the metallophone ("complex ndŭnlàm") and the idiophone ("simple ndŭnlàm") **(left)** and nə̀cà' tin shakers, tómŋkà' standing drum, and the mɛ̂nfámə́ sitting drum **(right)**.

TABLE 1 | Consonant and vowel distributions by stem position and affix type.

Consonants				
Stem-initial (48)	Stem-medial (7)	Stem-final (7)	Prefix (1)	Suffix (1)
^{m}ʙ, ^{m}b, ^{m}b^{w}, ^{n}t, ^{n}d, ^{ɲ}c, ^{ɲ}c^{w}, ^{ɲ}ɟ, ^{ɲ}ɟ^{w}, ^{ŋ}k, ^{ŋ}g, ^{m}v, ^{n}z, ^{n}z^{w}, ^{n}ts, ^{n}dz, ^{n}ʧ, ^{n}ʤ, ɲ^{w}, ŋ^{w}, ʙ, b, b^{w}, t, d, t^{h}, c, c^{w}, k, k^{h}, k^{w}, m, n, ɲ, ŋ, f, v, s, s^{w}, z, ʒ, ɣ, ts, dz, ʧ, ʤ, j, l	b, ʔ, l, ɣ, m, n, ŋ	p, t, k, ʔ, m, n, ŋ	n	d
Vowels				
Stem-initial (11)	Stem-medial	Stem-final (1)	Prefixes (1)	Suffixes (1)
i, u, ɪ, ʉ, e, o, ɛ, ɔ, ə, a, ɑ	N/A	ə	ə	ə

The paper proceeds as follows: in §2, we provide an overview of the two Medʉmba folkloric songs investigated in the study. In §3 we provide details of some aspects of phonological and morphological structure in Medʉmba which are most likely to be relevant to text-setting behavior. We then describe our text-setting methodology in §4, and provide results and discussion of this analysis in §5 and §6.

DATA SET

Our data set consists of two Medʉmba folksongs sung in the context of a *ntánlá'* community meeting in the Lafeng quarter of Bangoulap village, located southwest of the city of Bangangte.

The *ntánlá'* meeting is held every 8 days per the Bamileke calendar. *Ntánlá'* derives from *ntánə́* meaning "market" and *lá'* "village", literally translating to "village market day", or the day when people rest from manual labor and attend the market to buy and sell goods. The meeting serves primarily as a site for community governance. *Ntánlá'* meetings also serve an important function of bringing community members together during major events, such as after the death of a member, or in preparation for a member's wedding or move to a new location. In our case, the occasion was the return of the second author to his village after being away, and a visit from the first author as a guest. Under all of these special circumstances, music and dance typically take place in combination with monetary transactions meant to fund funeral services, medical expenses, or moving expense (see Jo-Keeling 2011 for details on the link between banking and music-making in Bamileke culture).

The specific style of music to be described comes from the *Làm* folk music tradition, which translates to "iron" or "forge", as it is historically the music of blacksmiths in the region (see Lendja 2016 for a deep history of ironwork in the Bamileke region). *Làm* music is played by a percussion ensemble which is made up of various members of the community; different members may be present and playing instruments on any given meeting day. Many of the instruments played within the ensemble, including the *ndŭnlàm* metallophone and idiophone which keep the pulse of the music, are made of iron. The ensemble is led by a master drummer, who is a fixture of the ensemble. Master drummers are typically identified at a young age within the community and are trained throughout their lives to play the rhythms of local songs. The ensemble recorded for the present work included the metallophone and idiophone, both referred to as *ndŭnlàm* (elsewhere we will refer to the metallophone as the "complex *ndŭnlàm*" and the idiophone as the "simple *ndŭnlàm*"), the *nə̀cà'* tin shakers, the *tómŋkà'* standing drum, and the *mɛ̂nfámə́* sitting drum, played by the master drummer (**Figure 2**).

Historically, *Làm* is a music born out of the activities of the forge, to perpetuate in the context of an ensemble the noise and dialogue of the anvils and other tools emanating from local workshops around the village all day long. Importantly, *Làm* music was used as a tool for advertising the products of the forge, such as hoes, dibbles, and other farming equipment, which were on display on tables next to the musicians. The music was used to draw interested customers, and to entice community members to dance in order to create an even more alluring spectacle for potential buyers. Nowadays, *Làm* music is typically accompanied by a circular dance which is performed by making a round around the musicians who play their different instruments in the center of the performance space. Songs consist of a call and response pattern, with the master drummer calling out to members of the community—often using traditional names of community members—and the community responding in kind. *Làm* music and dance events are some of the most egalitarian of such events in Bamileke culture, given that they are sung and danced without a mask (a feature of certain, more sacred dance rituals), and that they welcome all singers, dancers and instrumentalists regardless of gender or social rank. Just like other folkloric music of the Bamileke people, such as *Késsoú*, *Máŋgàmbéu*, and *Nètchà*, where singers display their talents in a free and creative style, *Làm* events provide the master drummer and other community members the opportunity to express themselves and for the master drummer to speak to the community through song. Through his song, the master drummer greets the community, and also tells the history of the village through proverbs and through listing community members' traditional names and their relationships to one another. The present work will focus specifically on the rhythmic structure of these speech-based songs, how they map to the rhythms played by the instruments in the *Làm* ensemble, and, most importantly, how these musical rhythmic structures interact with aspects of Medʉmba linguistic structure.

PHONOLOGICAL AND MORPHOLOGICAL STRUCTURE IN MEDɄMBA

There are a variety of ways in which linguistic structure has been found to interact with rhythmic properties in music, several of which will be considered in the present section in the context of Medʉmba *Làm* music. As mentioned previously, metrically prominent syllables (whether indexed phonetically by stress, or not) tend to co-occur with musically strong beats across languages. First, we outline various segment-level patterns in Medʉmba which provide evidence for metrical foot structure, which we hypothesize may interact systematically with musical rhythmic structure. We then discuss tonal patterns, which, in addition to interacting with melodic patterns in music, have been shown to influence rhythmic patterns, as well. Finally, we consider evidence from syllable structure and tone in the context of possible syllable weight distinctions, which are also know to play into text-setting patterns in various languages.

Segmental Structure

Medʉmba is one of dozens of Bamileke languages classified within the Eastern Grassfields subgroup of Grassfields Bantu. While evidence suggests these languages are descended from Proto Bantu, Medʉmba and other Grassfields languages look quite different from many well-studied Eastern and Southern Bantu languages in that they have more isolating morphology and have lost many segmental affixes typical of Bantu, some of which persist instead in the form of floating tone morphemes (Voorhoeve 1971; Hyman 2003). Medʉmba patterns with other Grassfields Bantu languages, as well as other Bantu languages from the Northwest regions (those located in Guthrie zones A and B; Guthrie 1948) and several non-Bantu languages of West and Central Africa in exhibiting positional prominence effects, such that stem-initial syllables bear a greater number of consonantal and vocalic contrasts than do non-initial and non-stem syllables (see Hyman et al., 2019 and references therein). In Medʉmba, in stem-initial position and in monosyllabic stems, 48 consonants and 11 vowels can appear (**Table 1**)[6]. In non-initial position, the number of contrasts is reduced to 7 consonants and one vowel ([ə])[7]. Medʉmba displays few segmental affixes, but those that do exist exclusively contain the vowel [ə]. Additional positional restrictions on tone are discussed in Franich, 2021.

Non-compound native stems in Medʉmba are for the most part either monosyllabic (N)CV or (N)CVC or disyllabic (N)CVCV. As described in Franich (2021), in disyllabic forms, distributional asymmetries of consonants shown in **Table 1** derive in part from a lenition process, as demonstrated in examples like (1,2), which targets consonants occurring in foot-medial position[8]. Words such as *^{m}b^{w}áɣə́* "fire" and *^{m}bálə́* "hill" are realized as disyllabic in isolation or phrase-finally (1a-b), and as monosyllabic phrase-initially or phrase-internally (1c-d). As seen in (1a), the velar stop /k/ is realized as [ɣ] word-internally due to spirantization (as well as voicing). A similar pattern is found for the consonant /d/, which lateralizes to [l] in the same environments where /k/ undergoes spirantization (1b).

(1) Spirantization of /k/ and lateralization of /d/

a.	^{m}b^{w}á**ɣ**ə́	'Fire'	/^{m}b^{w}á**k**ə́/[9]
	sá**ɣ**ə́	'Sauce'	/sá**k**ə́/
b.	^{m}bá**l**ə́	'Hill'	/^{m}bá**d**ə́/
	^{m}vɛ́**l**ə́	'Brother	/^{m}vɛ́**d**ə́/

[6]This list includes most of the same consonants and vowels described by Voorhoeve (1965), Voorhoeve (1976), with some departures. For example, while Voorhoeve argued for a contrast between /k/ and /g/ in the language, we find no evidence that these are distinct phonemes in the Bangangte/Bangoulap dialects. Furthermore, root-internally, Voorhoeve's /mf/ is always produced as /mv/ in the dialects examined here; we therefore transcribe them as such. Aspirated consonants are contrastive in loanwords only. Finally, the vowel inventory is updated in places to reflect more recent acoustic analyses of the Bangangte dialect by Olson and Meynadier (2015). Vowels analyzed as diphthongs by Voorhoeve (1965) are also excluded from the present discussion.

[7]Note that in stem-final position, plosive consonants become devoiced.

[8]While Danis (2011) analyzes lenition as conditioned by the prosodic word, Franich (2017), Franich (2021) provides arguments for why a foot-based analysis is more appropriate.

[9]Voorhoeve (1965) treats the underlying medial consonants in examples like (1–3) as /g/ and /d/, rather than /k/ and /d/ (thereby allowing for a more unified treatment of lenition as targeting consonants which are [+voice]). As mentioned in footnote 6, we find no evidence synchronically for a contrast between /k/ and /g/, and velar stops are realized by default as voiceless (with voiced [g] predictable in contexts where it occurs, such as within some prenasalized sequences). We therefore opt to treat the underlying form of the medial consonant in the forms in (1–3) as /k/.

(2) Foot structures for Medʉmba nouns and verbs

a.	$_{Ft}$(^{m}b^{w}áɣə́)$_{Ft}$	b.	$_{Ft}$(^{m}bálə́)$_{Ft}$
	$_{Ft}$(káɣə́)$_{Ft}$		$_{Ft}$(tɛ́lə́)$_{Ft}$

As demonstrated in Franich (2017, 2018, 2021), stem-initial syllables in Medʉmba display evidence of rhythmic prominence in speech timing, while stem-final syllables, prefixes, and suffixes display evidence of rhythmic weakness. Assuming that foot/stem-initial syllables behave as rhythmically prominent in music (as they have been found to in speech production), we might expect there to be a systematic relationship between these syllables and musically strong beats.

Pronominal enclitics, which also trigger lenition (if vowel-initial) but which realize the full range of vowel contrasts in the language (3), behave similarly to stem-initial syllables in showing rhythmic prominence in speech timing, though they may be undergoing a process of prosodic weakening.

(3) Lenition triggered by pronominal enclitics

a.	mbwáɣ=ám	mbwáɣ=ú
	'My fire'	'Your fire'
	sáɣ=ám	sáɣ=ú
	'My sauce'	'Your sauce'
b.	mbál=ám	mbál=ú
	'My hill'	'Your hill'
	mvɛ́l=ám	mvɛ́l=ú
	'My brother'	'Your brother'

Assuming that pronominal enclitics behave similarly to stem/foot-initial syllables in their rhythmic timing in music, we expect similar co-occurrence patterns to hold for both stem/foot-initial syllables and pronominal enclitic syllables in text-setting.

Tonal Structure

Phonemically, Medʉmba has only a binary tonal contrast between high and low tones. Falling and rising tones can occur in many contexts, though they are analyzable, as in most other African tone languages, as sequences of level tones (Leben 1971; Goldsmith 1976; Clements and Goldsmith 1984). For example, verb stems carry one of two tone melodies, H or LH, and can be either monosyllabic or disyllabic. As can be seen in (4), monosyllabic verbs can host LH contours, while the LH melody is distributed as a sequence of two level tones on a disyllabic verb. Phrase-internally, disyllabic verbs are reduced to monosyllabic (N)CVC structure. Reduced verb forms with a LH melody are also realized with a contour tone (4c). For disyllabic verbs with a H melody, reduced forms are realized with a single high tone (4d).

(4) Mono- and disyllabic verbs with H and LH tone melodies

a.	zí "sleep"	zínə́ "walk"[10]	(phrase-final form)
	ʒʉ́ "eat"	ʒʉmə́ "be dry"	(phrase-final form)
b.	bǎ "be ripe"	bɑ̀ɣə́ "split"	(phrase-final form)
	sǒ "press"	sɔ̀ŋə́ "throw"	(phrase-final form)
c.	zín "walk" (phrase-internal form)		
	ʒʉm "be dry" (phrase-internal form)		
d.	bǎk "split" (phrase-internal form)		
	sɔ̌ŋ "throw" (phrase-internal form)		

Falling and rising tones can occur as a result of various other processes, including docking of a floating tone morpheme to a word which is already specified for a tone, as in some instances of the associative construction (5).

(5) Tonal morpheme docking in the associative construction (Voorhoeve 1971)

a.	bàm	x́	mɛ́n	→	bǎm ꜜmɛ́n	'Sack of the child'
	Sack	ASSO	Child			
b.	mɛ́n	x̀	mɛ́n	→	mɛ̂n mɛ́n	'Child of the child'
	Child	ASSO	Child			

Contours may also form where two morphemes merge and a vowel is elided but its tone remains, as happens with a sequence of a complementizer and a pronoun, or a copula and an object focus marker (6).

(6) Tonal contours resulting from merging/vowel elision

a.	zə̀	+	á	→	zǎ
	COMP		1SG.FOC		COMP.1SG
b.	bə́	+	à	→	bâ
	COP		FOC		COP.FOC

Verbs also undergo a process of tonal overwrite in relative clause, *ex-situ wh*-questions, and in the subjunctive mood, such that all verbs are realized with a falling contour in those contexts, even if underlyingly high toned (7) (Voorhoeve 1976; Keupdjio 2020). Various other morphemes, such as some tense markers, also bear contours which appear to have arisen through historical processes of vowel loss and tone docking.

(7) Falling tone overwrite in relative clauses and *wh*-constructions

a.	bǎ "be ripe" (main clause form)	bâ "be ripe" (relat./*wh*-/subjunctive form)
b.	zí "sleep" (main clause form)	zî "sleep" (relat./*wh*-/subjunctive form)

One potentially interesting question concerns whether contours formed from different sources—for example, those arising from a word's lexical tone alone vs. those formed

[10]It is standardly assumed, in order to avoid a stem-internal Obligatory Contour Principle violation (Leben 1971), that the high tone in disyllabic examples such as those in (4a) is a single tone linked to both stem syllables (Hyman and Tadadjeu 1976).

through morphological processes—show different patterns in text-setting in Medʉmba. For example, McPherson and Ryan (2018) show that melodic text-setting requirements are more stringent in Tommo So for lexical tone vs. grammatical tone. One could imagine a similar type of dichotomy when it comes to rhythmic constraints; we will explore this possibility in the current data set.

Morphology: The Associative Construction

Another aspect of linguistic structure which will be important to consider in our analysis pertains to morphologically complex forms, including compounds and possessive forms, all of which are formed in Medʉmba using the associative construction, as was seen in (5). This description also extends to traditional names, which typically involve this construction. Though the associative construction can be used to form verbs and adjectives, we limit ourselves here to a description of noun-based constructions, which were the most common within our song dataset.

In the associative construction, the noun class of the head noun in the construction (the leftmost noun) determines the tone of the associative morpheme (H or L) and this tone then docks to the head noun to mark the construction (Voorhoeve 1971). Generally speaking, if the associative tone is opposite that of the head noun, it forms a contour (5a,b); if the two tones match, they merge into a single level tone (8a)[11]. As can be seen in (8b-e), aside from possessive relations, a variety of other relationships between nouns can be conveyed through the associative construction, and several words can be combined to form multiple embedded associative relations. Some of the names of instruments to be described in the present study are also formed through the associative construction (8b).

(8) Examples of the associative construction

a.	mvə̀n chief	x̀ ASSO.CLI[12]	mɛ́n child	→	mvə̀n mɛ́n	“Chief of the child”
b.	ndùn drum	x́ ASSO.CLII	làm Iron	→	ndǔnlàm	“Metallophone”
c.	ndà house	x́ ASSO.CLV	ncà Word	→	ndǎncà	“Palace of justice”
d.	ndǎncà pal. of just	x́ ASSO.CLV	Bangangté Bangangté		→	ndǎncǎ Bangangté “Bangangté palace of justice”
e.	nyə̀m creature	x̀ ASSO.CLI	nkù PL.leg	x̀ nkuə̀ ASSO four	→	nyə̀mnkùnkuə̀ “Quadruped”

The associative construction is commonly used in the construction of traditional names, including those referred to as ndàp, which are assigned by generation to children through the mother's side (though women also receive a separate ndàp from the father's side). Some of these names are semantically transparent, such as *ʙ^{w}ŏndà*, comprised of *ʙ^{w}ŏ* “good” and *ndà* “gift,” literally translating to “good gift,” or *mântú'kámə́*, comprised of *má* “mother”, *ntú'* “calabash/vessel”, and *kámə́* “power/nobility”, which translates directly to “mother of the vessel of power,” a reference to the noble status of individuals bearing this particular ndàp (for further details on the history of ndàp names and nobility in Medʉmba culture, see Mkammi 2009).

Other names are less transparent, but bear evidence of being derived from similar constructions. As will be seen in §5, a feature of many Medʉmba folkloric songs is a series of greetings of individuals in the community by their traditional names.

Unlike disyllabic stem syllables, associative forms each contain two or more noun stems, each of which is hypothesized to contain a metrical head. It is possible, however, that there exists internal metrical structure to the associative construction, similar to what is found with compound stress in languages like English, in which the syntactic head and its complement receive distinct levels of prominence. In our analysis, therefore, we will pay attention to which members of an associative construction occur in which positions within the rhythmic structure of songs.

Syllable Weight

One outstanding question related to phonological structure in Medʉmba concerns whether the language distinguishes between syllables of different weights. There is no clear evidence of a vowel length distinction in the language, nor is there evidence for distinct patterning between CV and CVC syllables that might suggest that coda consonants contribute to syllable weight. However, there are some aspects of phonological patterning which suggest that weight could play a role. One piece of evidence concerns positional restrictions on tone. Specifically, Franich, 2021 shows that foot-initial syllables, which typically align with the left edge of the stem, are uniquely able to bear contour tones, whereas non-initial syllables are prohibited from bearing contours. As was seen in (4), native monosyllabic stems of various shapes can host contour tones. Since Medʉmba features few native polysyllabic words bearing contour tones which are unambiguously monomorphemic, English loanwords provide a source of evidence that contours are limited to stem-initial position. As

[11]However, see Voorhoeve (1971) for exceptions to this generalization.

[12]Noun class numbering is based on Voorhoeve (1968).

seen in (9), whereas monosyllabic words and syllables bearing primary stress in English are normally borrowed with a high tone, those with nasal codas are typically borrowed into Medʉmba with a falling contour.

(9) English loanwords in Medʉmba; CVN syllables borrowed with falling contours

a.	tʰí	[tʰi]	"Tea/coffee"
b.	kát	[kʰa:d]	"Card"
c.	há.mə̀	[ˈhæ.mə]	"Hammer"
d.	sú.ndì	[ˈsʌn.ˌdej]	"Sunday"
e.	bà.ná.nà	[bə.ˈna.nə]	"Banana"
f.	pîn	[pɪn]	"Pin"
g.	tâm	[tajm]	"Time"
h.	ⁿgûm.nə̀	[ˈgʌv.nə]	"Governor"
i.	sîŋ.lì	[ˈsɪŋ.glət]	"Singlet"

However, HL contours are only permitted on such syllables when they occur stem-initially: if a disyllabic word with a second syllable of CVN shape is borrowed, for example, an epenthetic vowel is inserted after the nasal (either [i] or [ə], depending on the place of articulation of the preceding consonant), and the HL contour is distributed as separate H and L tones across the final two syllables (10). Note that words of the shape CVCVN are licit in the language, as evidenced by forms like *bə̀lɔ̀ŋ* "potato," indicating that tone, and not segmental phonotactics, must be motivating epenthesis in the forms in (10).

(10) HL contours limited to stem/foot-initial position

	Medʉmba loanword		English source IPA	English translation
a.	dósínì	(*dósîn)	[ˈdʌzən]	"Dozen"
b.	flébánə̀	(*flébân)	[ˈfɹaj ˌpan]	"Fry pan"

Franich (2021) shows that this pattern can be accounted for by analyzing these trisyllabic forms as involving a disyllabic foot aligned to the left edge of the stem (11). From this perspective, vowel epenthesis can then be seen as a strategy for avoiding a contour tone on the weak syllable of the foot.

(11) Foot-based analysis of trisyllabic loans

a.	dósínì	$_{FT}$(dósí)$_{FT}$ nì
b.	flébánə̀	$_{FT}$(flébá)$_{FT}$ nə̀

Given that many languages show a correspondence between weight and metrical prominence, one way to account for this fact would be to posit that foot-initial syllables in Medʉmba are (or can be) heavy, while non-initial syllables are light. This treatment would be in line with analyses of African tone languages which posit the mora as the tone-bearing unit—where two moras are present, a contour tone may occur; where only one mora is present, only a level tone may occur (Hyman 1985). It is also notable that syllables bearing contour tones in Medʉmba have been found to be significantly longer in duration—up to double in length—than those bearing level tones (Franich, 2014)[13].

This link between contour tones, duration, and syllable weight is open to other interpretations, however: Zhang (2002, 2004) provides evidence that contour tone licensing need not make reference to moras, and that positional restrictions on contour tone licensing may be better captured through positional markedness constraints which prohibit contour tones from surfacing in positions which are articulatorily and/or perceptually ill-suited to bear contour tones (i.e., syllables with shorter/less sonorous rhymes which cannot facilitate efficient production and perception of contour tones). Importantly, there is no evidence in Medʉmba that syllables in stem/foot-initial position are generally longer in duration than in other positions, and contour tones in the language are allowed on all syllable types, including both CV and CVC syllables with both voiced and voiceless coda consonants (the latter generally forming a poor phonetic host for a contour tone, as reflected in Zhang, 2002 cross-linguistic typology).

Zhang's approach assumes that phonetic properties of syllables have a direct influence on their tone-bearing abilities. The account therefore dispenses with the traditional mora-based representational account of contour tone licensing. Another alternative to Zhang's proposal is found in Gordon (1999), Gordon (2001), who recognizes a role for phonetic properties such as sonority and duration in influencing contour tone licensing, but still argues for a representational account. Specifically, Gordon proposes an enriched typology of weight distinctions which can make reference to the kinds of phonetic properties Zhang considers, but which are formalized using skeletal slots to represent timing units. In the case of a language like Medʉmba, where contours are allowed on syllables of all shapes but where vowels are clearly longer in the presence of a contour tone, an analysis might involve a constraint that contour tones only be realized on vowels with two timing slots.

The different ways in which representational theories and phonetically based theories like Zhang's conceive of the relationship between duration and contour tone licensing could have interesting implications for the present musical analysis. Assuming there is a principled relationship between timing in language and music for Medʉmba, as has been found for other languages (Patel and Daniele 2003), this relationship may or may not reflect weight-based distinctions. Contour toned

[13]This pattern would mimic the behavior of a typologically rare foot type, the "uneven trochee" (Hayes 1995): while iambic feet typically display durational asymmetries across initial/nonhead and final/head syllables, trochees do not show such durational asymmetries, instead favoring other types of phonetic prominence asymmetries, such as asymmetries in loudness. Uneven trochees are not completely unheard of in the realm of African languages, however: Green (2015), for example, documents variable cases of uneven CVV.CV trochees in Bambara, a Mande language, which he argues are undergoing a change toward becoming more typical CV.CV trochees.

syllables of different shapes vary widely in their rhyme durations in Medʉmba, with CVC syllables typically realized with much shorter durations than CV syllables (Franich, 2014). We might expect, then, that note durations of contour-toned syllables in text-setting will be quite variable, too, if no reference is being made to a more abstract, uniform weight distinction between contoured and level-toned syllables. On the other hand, if syllable weight, and not just phonetic duration, is being relied upon as a guide for text-setting, we might expect more uniform note durations for contoured syllables, regardless of their shape.

To review, Medʉmba does not display robust evidence for a syllable weight distinction, as in languages with clear minimal pairs for vowel length. However, contour-toned syllables in Medʉmba show both longer duration than level-toned syllables, and also distributional restrictions which prohibit them from occurring in prosodically weak positions. Both of these phenomena are consistent with the treatment of contour-toned syllables as phonologically heavy, though other explanations are possible. Examining text-setting patterns of note length for contoured vs. level-toned syllables may provide us a window into this complicated area of the grammar. If it is the case that contour-toned syllables are heavier than level-toned syllables in the language, we might predict that they will occupy greater note lengths than corresponding level-toned syllables in text-setting, and that the mapping between note length and tone will be consistent and based on relatively uniform abstract weight distinctions, rather than variable as a function of phonetic duration alone.

Summary

To summarize, we will pay specific attention to the position of syllables within a stem/foot, the tone of syllables, and their morphological status in our text-setting analysis. Taking these areas as a starting point, several hypotheses can be made. First, we hypothesize that stem/foot-initial syllables may behave differently from non-initial syllables and segmental affixes in being drawn with greater frequency to rhythmically strong beats. We also hypothesize that note length may be determined by aspects of tone and morpho-prosodic structure, with contour toned syllables most likely to occupy the greatest note lengths, and stem-final/affix syllables most likely to occupy the shortest note lengths, being less prosodically prominent. An open question concerns whether finer-grained note length distinctions will be observed between contour-toned syllables of different shapes. It is also possible that contour tones formed through different means, e.g., those which represent a word's lexical tone vs. those which are formed through the concatenation of a floating "grammatical" tone or from tonal overwrite, might display distinctive rhythmic properties. If, for example, grammatical tones are generally less faithfully realized in text-setting as found in McPherson and Ryan (2018), floating grammatical tones may not contribute as reliably to a syllable's rhythmic timing within songs. Finally, we will pay attention to whether word stems within the associative construction consistently behave as separate prosodic units, or whether they may form a single unit under any circumstances.

METHODOLOGICAL APPROACH

A challenge in evaluating the degree to which rhythmic structures in the Medʉmba language and music align is that we must decide on a strategy for representing musical rhythms in a way that will facilitate such a comparison. Medʉmba folkloric music of the kind described here is not traditionally written down; thus, there is no notational system indigenous to the culture that would facilitate such a comparison. Rather, transmission of music is experience-based: newer drummers learning to play the music do so by playing it alongside more experienced drummers, and community members learn dance steps by dancing alongside their elder family members.

There are a variety of ways in which music theorists have approached the question of notating African music, many of which dispense with notions of Western staff-based-notation such as the time signature. Chernoff (1991), for example, notates rhythms as continuous lines of alternating symbols reflecting the sounding of an instrument or the feel of a metrical pulse. Others, such as Anku (2000), opt to represent African rhythms in terms of cycles of sets of beats, each set having one regulative beat which need not be associated with any type of audible "accent" or note onset. These approaches essentially dispense with treating musical beat structure as hierarchically organized, an implicit feature of Western staff-based notation, where time-signatures reflect the division of pulses into beats and measures (Temperley 2000). However, there is reason to believe that music from West and Central Africa is—or at least can be—hierarchically organized. Toussaint (2013), for example, uses an empirical approach based on the probabilistic positioning of note onsets across beats in several different African rhythms[14], showing that note onsets (drum beat attacks, chord changes, etc.) are more likely to occur in certain beat positions than others, and that multiple, discrete levels of beat strength can be seen to emerge from these positional restrictions on note onsets. This is in spite of the fact that the specific pattern of metrical organization differs between the examined African songs, where metrical patterns tend to be arranged in a 3-3-2 beat pattern, and other styles of music, such as German folk songs, where beats are arranged in 4-4-4 grouping patterns. Importantly, Toussaint points out that these implied patterns of beat strength in African songs, unlike in much Western folk and classical music, need not occur with explicit acoustic accenting (as through increased duration or loudness). This latter point accords with Agawu's analysis of Ghanaian music as having clear rhythmic structure

[14]This work does not mention explicitly which regions of Africa are represented in the analysis, though patterns described are consistent with Central and West African rhythms.

FIGURE 3 | Staff transcription of Song 1.

FIGURE 4 | Staff transcription of Song 2.

(which guides, for example, the coordination of dance feet) without having a rigid alternation of accented and non-accented beats.

Agawu (1995) argues that the use of Western musical notation, while not without potential problems, is an appropriate choice, particularly as this approach can facilitate comparison of African musical rhythms with those from other musical traditions (pp. 185–195). In the present work, we follow music theorists such as Agawu in adopting staff-based notation, which allows us not only to observe clearly how different instruments and voices within the ensemble interact, but also provides us with some level of precision in notating note durations, which will form an important part of our text-setting analysis[15]. Agawu stresses that the use of such a staff-based approach for representing rhythm does not necessarily entail the acceptance of a fixed hierarchical beat structure, in which there is a sense of a single beat within a measure being consistently treated as metrically "strong" to the exclusion of other beats. Rather, there is "…a multiplicity of competing accents, which are always held in check by a simple, regular background" (p. 191). This brings up the additional issue of how to determine what the appropriate grouping pattern is for something like meter/time signature within this music. Here, we follow Agawu, Anku (2000), Toussaint (2015) and others in using repeating rhythmic motifs among instruments to establish rhythmic grouping patterns, as well as Kubik (1983) in exploring cues beyond the music itself, incorporating information about the timing of dance steps and body movements to guide our thinking on where the pulse of the music is optimally felt. Note that melodies are encoded only approximately within staff-based notation; since the primary focus of the present work is on rhythmic structure, no formal analyses of musical key was conducted on these songs. Finally, for ease of interpretation, we also include grid-based notations of song parts to facilitate the visualization of alignment of main beats and beat fractions with linguistic structure. Note that some slight deviations exist between the staff- and grid-based representations since the 12-beat subdivisions in the grid make it more constrained in the rhythmic patterns it can capture.

ANALYSIS

We start by characterizing some of the basic rhythmic elements of the song structures, which are transcribed in **Figures 3**, **4** (red timestamps allow the reader match the transcription to the accompanying sound recordings). In both songs, the simple *ndŭnlàm* (idiophone) and the *nəcà'* (tin shakers) double one another in playing a consistent pulse throughout each song. These parts define the basic beat of the song, and we follow Agawu (1995), Anku (2000), Toussaint (2015) in interpreting the attacks of these instruments as falling on the whole numbered beats of the measure throughout the song. We note, too, that dancers' arm movements and footsteps are aligned with these beats: specifically, dancers' arms move back and forth on every hit of the idiophone and shake of the shaker (**Figure 5**), and dancers take a step forward on every other attack of these two instruments (what we have indicated as the odd-numbered beats of the song measures) (**Figure 6**).

As indicated in **Figures 3**, **4**, assuming a 12/8 compound meter, these instruments sound four times per measure, at equal intervals. Layered on top of these instruments is the sound of the complex *ndŭnlàm* (iron metallophone), which realizes a triplet eighth note pattern for every beat of the idiophone and *nə̀cà'* (except for the final beat of each measure of Song 2). Over this rhythmic backdrop, the *tómŋkà'* (standing drum) and *mɛ́nfámə́* (sitting drum played by the master drummer) add their voices, introducing syncopation at various points with respect to the rhythmic line played by the idiophone and shaker. In Song 1, for example, the standing drum sounds an extra-long note beginning on the second triplet eighth note of the second beat, while the master drummer highlights the second triplet eighth note of the third beat by not playing on the first one. In Song 2, the standing drummer highlights the second triplet eighth note of the second and fourth beats in a similar manner, while the sitting drummer highlights the second triplet eighth notes of the third and fourth beats[16]. The frequent occurrence of these drum beats on the *second* note of the triplet pattern can often give the feel of a different subdivision of the beat structure altogether, in which measures are broken down into three groups of four eighth notes, rather than four groups of three; this represents a classic duple vs. triple polyrhythmic opposition found in many genres of West and Central African music (Temperley 2000).

As will be described below, several of the predicted patterns described in §3 are substantiated in Songs 1 and 2, including the preferred timing of stem-initial syllables on main beats within the song, and timing of stem-final syllables, affixes, and functional elements on beat fractions. Interestingly, this pattern is found to be completely reversed within one particular motif found in Song 2, but prominent and non-prominent syllables nonetheless continue to pattern together within that section. We also find that syllables occurring outside of main beats frequently occur on beat fractions which are accentuated by the sitting drum, further emphasizing a polyrhythmic pattern. In terms of tone, we find that both contoured and level-toned monosyllabic stems tend to occupy the same note length of two triplet eighth notes (though some exceptions apply in the case of compounds),

[15]As one editor points out, a shortcoming of using staff-based notation is that its implied hierarchical nature may obscure some microvariation in note timing which might arise if, for example, two or more voices are not perfectly rhythmically aligned. We acknowledge this shortcoming, and the potential for some oversimplification of timing in our staff-based representations. Of course, an ideal investigation of rhythmic timing would involve time-locked individual recordings of each voice in the ensemble, which we do not have access to at present; we would like to undertake this type of investigation at a future time.

[16]Note that there is further variation in the rhythms that the drummers play from verse to verse, but the patterns transcribed here represent recurring patterns throughout the songs.

FIGURE 5 | Dancers' swing arms back and forth for very attack of the idiophone and shaker.

while individual syllables in disyllabic roots each occupy a single triplet eighth note; affixes and many function words also tend to occupy shorter note lengths. Patterns of phrase-final lengthening are also reflected in note durations at the ends of musical phrases. We now provide more detailed, line-by-line analysis of each song.

Song 1: Line-by-Line Analysis

Focusing now on the sung portions of Song 1, we hear the master drummer and chorus trading lines throughout both songs, in a call-and-response pattern. The master drummer initiates the song before the chorus joins in. Song 1 begins with the master drummer repeating a line to the chorus, one time as a solo, the second with the accompaniment of the drummers. The second of these two repetitions is the first line that appears in the staff-based representation of Song 1 (starting at :14 in the sound file of that song). This first line is presented in (12), with separate lines for surface phonetic pattern, underlying phonological structure, gloss, and translation, and in a grid formulation in **Figure 7**. Toneless syllables receive their tone either by default (as in the 3sg pronoun *a*) or through tone spreading (as in the final syllable of verb stem cóbə). Floating tone morphemes are indicated with a H or L in the gloss line.

the song. The line is the same each time, and always initiated on the second triplet of the third beat of the measure, and continues for another measure and a half, with the final syllable *b*ʷ*ó* (the result of merging *b*ʷ*ə̀* "there" and *ò*, a marker of certitude/evidential) initiating the final measure of the line on the first beat.

Within this line, we can notice a few interesting patterns. Looking at the positions within the rhythmic line where individual syllables fall, the main beats (numbered in **Figure 7**), where they are articulated, are occupied exclusively by stem-initial syllables. For example, the fourth beat of the initial syllable is initiated with the stem-initial syllable in *cóbə* "speak," the first beat of the second measure is occupied by *bwə̀* "there" (combined with the certitude marker) and the first beat of the third measure is occupied by *tà* "far away" (combined with the H associative marker). It is interesting to note that main beats 1 and 4 are always articulated in the opening lines/chorus, while beats 2 and 3 are left empty; indeed, Beats 1 and 4 seem to show great similarities in their patterning throughout the song. The significance of the first and fourth beats in a 12/8 rhythm has been observed in other African musical traditions (Vetter 1996).

Looking now to syllables which do not occur on whole-numbered beats, the picture is more mixed. Among the elements occurring in these rhythmic positions are the stem-final syllable in *cóbə* "speak," the third person singular pronoun *a*, and the two syllables in *Băbá*,

(12) Master drummer's opening line/chorus

à	cóbə́	Bʷə̀	à	cóbə́	tǎ	Băbá	à	cóbə́	Bʷò
a	cóbə	Bʷə̀	a	cóbə	tà.H	Băbá	a	cóbə	Bʷə̀+ò
3SG	speak	there	3SG	speak	far.ASSO	Băbá	3SG	speak	there + CERT

"It speaks from over there, it speaks from the Băbá (river) (it's clear) it speaks from there."

As we have analyzed Song 1 to involve a 12/8 time signature, grids are comprised of 4-beat measures with each beat divided into triplets, for a total of 12 pulses per measure. Thus, each individual pulse represents a triplet eighth note unit within the staff notation. In addition to labeling the whole numbered beats of the measure within the grids with numbers 1–4, we have also labeled those beat fractions where the two drums (abbreviated St and Si for "standing" and "sitting," respectively) provide articulations.

The opening line, with the same rhythm, will become the refrain of the chorus after the master drummer begins singing the verses of the name of the river being sung about. Alignment of syllables by morpheme type and tone for the chorus is summarized as proportions in **Figures 8**, **9**. As can be seen in **Figure 8**, the majority of elements occurring on beat fractions appear to be stem-final and functional elements, with the exception of the two syllables in *Băbá*. Interestingly, several of the syllables occurring outside of whole numbered beats (including the second syllable of *Băbá*) occur on beat fractions which are articulated by one of the two drummers (see also **Figure 7**). In terms of tone, following from the fact that stem-final syllables are more common on beat fractions, we

FIGURE 6 | Dancers' synchronized forward steps occur on every other attack of the idiophone and shaker.

DRUM-ACCENTED FRACTIONS					St			Si							
WHOLE-NUMBERED BEATS	1			2			3			4			1		
LYRICS								à		có	bɔ́				
DRUM-ACCENTED FRACTIONS					St			Si							
WHOLE-NUMBERED BEATS	1			2			3			4			1		
LYRICS	bwɔ̀							à		có	bɔ́				
DRUM-ACCENTED FRACTIONS					St			Si							
WHOLE-NUMBERED BEATS	1			2			3			4			1		
LYRICS	tǎ		bɑ̌		bɑ́			à		có	bɔ́		bwò		

FIGURE 7 | Grid representation of master drummer's opening lines and chorus. Numbers 1–4 represent whole-numbered beats on which the idiophone/shaker sound and dancers move. Individual cells represent beats on which the complex ndǔnlàm/metallophone sounds. St/Si represent beat fractions on which the Standing or Sitting drum sounds. Gray filled cells indicate beats on which the voice is sounding.

see that these beats, where are articulated, are largely associated with toneless syllables. Beat 1 is mixed between L and LH toned syllables, while Beat 4 is consistently H. The master drummer's accented beat fractions on the sitting drum always occur on low tone syllables, while the standing drum accents beat fractions consistently on high tones; this pattern in fact parallels the tuning of the two drums, as the larger *mɛ̂nfámə́* sitting drum sounds at a lower pitch than the tall, thin *tómŋkàʼ* standing drum[17].

Another aspect of this initial line that is important to examine is the relative note durations attributed to each syllable in the line. In **Figure 3**, while the disyllabic verb root *cóbə* receives one eight note of the triplet pattern on each syllable every time it is uttered[18], monosyllabic roots such as *tà* (combined with a floating associative morpheme) and *bwə̀* (combined with the certitude marker) and the two syllables in *Bǎbá* receive double this note value. The third person singular pronoun *α* also consistently occupies a single eighth note value. This is consistent with the idea that disyllabic stems such as *cóbə* possess a similar overall weight to monosyllabic stems, both of which are overall heavier than the pronoun. However, we note that two of these monosyllabic stems are combined with additional morphemes, and two bear contour tones, two facts which may also contribute to greater note length for those syllables. We note, as well, that words occurring in final position of a musical line were consistently lengthened more than in any other position, mimicking a linguistic process of phrase-final lengthening (Turk and White 1999). This provides us with evidence of alignment not only in terms of metrical structure, but also in terms of prosodic grouping in the songs (Dell and Halle 2009).

Shifting our attention now to the master drummer's verses, which are improvised each time, we see much more rhythmic variety across different verses. This pattern mirrors findings from McPherson and Ryan (2018) on improvised vs. rote lyrics described for Tommo So folk music. The master

[17]Thank you to one of the editors for pointing out this potential link.

[18]We note that there is some variation in note length assigned to the final syllable of *cóbə* "speak" as it seems this syllable is sometimes matched with a single eighth note, and sometimes a longer quarter note. This variation is not represented in the staffs or grids provided. As we discuss in §6, outside of a musical context, there would typically be a focus marker *à* that would be uttered between the verb and the word *jà* "where" or the word *bwə̀* "there." Thus, it seems the singer is alternating between two forms, one the simple verb form, and the other with the addition of the focus marker, which corresponds with additional note length.

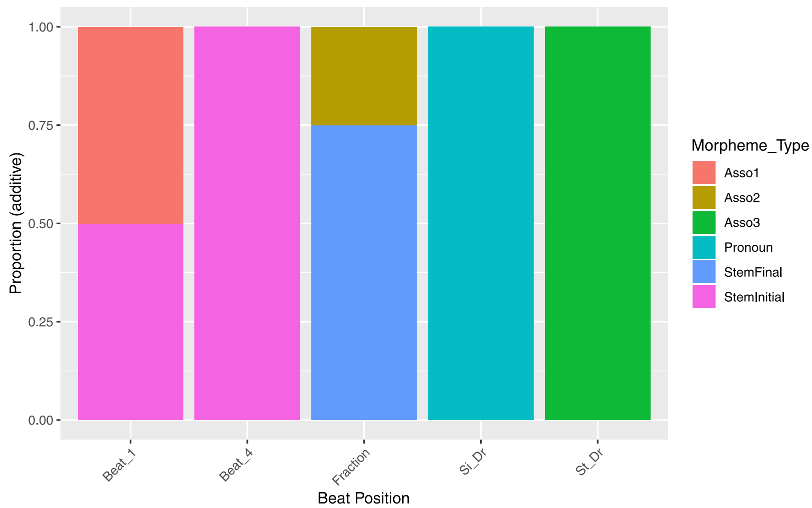


FIGURE 8 | Alignment of syllables by morpheme type in the opening lines/chorus of Song 1 represented as proportions of occurrences.

drummer's (henceforth MD) verses each consist of essentially one line each, which can range in length from 9 to 11 syllables. Here, we analyze ten lines of his part. Each of these lines involves the MD singing directly to the community, often calling out to different members using their *ndàp* or another social name they may have. Examples of three of these lines are given in (13–15). Note that for some particles listed, such as the exclamative particle *mə*, it is unclear whether the particle bears a low tone underlyingly or whether it may be toneless and surfacing with a default low tone.

(13) Line 1, master drummer verse, Song 1

wò	mə̀	ndŭn-làm	cóbə́	jà	ò	jà
wo	mə̀	ndùn.H-làm	cóbə	jà	o	jà
VOC	EXCL	ndùn.ASSO-làm	Speak	Where	VOC	Where

"Where does the ndŭnlàm speak from?!"

(14) Line 2, master drummer verse, Song 1

wó	mə̀	wʉ̆	və́.ꜜnsí	ú	ják	jù	lú	à
wo	mə̀	wʉ̀.H	və́.nsí	u.H	ják	jù	lú	à
VOC	EXCL	2SG.ASSO	sit.down	2SG.SUBJ	pass. afternoon	2SG.OBJ	PROX	Q

"You who are sitting down, how are you this afternoon?!"

(15) Line 3, master drummer verse, Song 1

é	kə̀-lɛ̂n	tâ-kwá-ŋkə̀'	mvɛ́t	Bʷŏ-ndà
e	kə̀-lɛ̀n.HL	tá.L-kwá.H-ŋkə̀'	mvɛ́t.H	Bʷò.H-ndà
VOC	NEG-know.SUBJ	tá.ASSO-kwá.ASSO-ŋkə̀'	brother.ASSO	Bʷò.ASSO-ndà

"Tâ-kwá-ŋkə̀' is the brother of Bʷŏ-ndà!"

Lines 1 and 3 in examples (13) and (15) above represent one of two forms that the MD's lines standardly take, incorporating a total of 9 syllables—we refer to this as "Song 1, Motif 1," or S1M1, for short. Six out of the ten lines analyzed followed the structure of S1M1. As can be seen in **Figure 10** (where lines have been ordered by motif), though the word structures vary across lines 1 and 3, there is overall considerable parallelism across lines in terms of rhythmic distribution (i.e., beat position at which words are initiated within the rhythmic line) and note duration. So, for example, the prefix of the exclamative *kə̀-lɛ̂n* (literally "not know", comprised of a negative prefix and the root *lɛ̂n* "know" in its subjunctive form) in (15), which occurs in measure 16 in **Figure 3**, occurs on the same beat and occupies a similar note length to the monosyllabic exclamative particle *mə̀* in (13) occurring in measure 8. Similarly, the words that follow these two syllables—the first syllable *ndŭn* in *ndŭn-làm* in (13) and the root *lɛ̂n* in (15)—occur on the same beat and occupy the same note lengths, both longer than their preceding syllables (two eight notes to the preceding syllables' single eight notes). It is interesting that both of these syllables bear contour tones, albeit with different tone values (LH vs. HL) and formed through two different processes, the first through concatenation of a floating morpheme, and the other through a process of tonal overwrite.

Another observation concerns the note durations of the syllables in the word *cóbə* "speak", which are each the same duration as the previously mentioned exclamative particle and prefix, such that the entire verb together comprises the same duration (two eighth notes) as one of the two contoured syllables previously discussed. A closer look at text-setting of other morphemes indicates, however, that level-toned monosyllabic forms can also occupy notes of two triplet eight notes in length, as is the case for the low-toned word *jà* "where" and the high-toned word *və́* "sit." This would seem to indicate that such level-toned CV words possess a different musical status than the exclamative particle and prefix, consistent either with a difference in prominence or syllable weight between the two types of level CV syllables. Finally, we

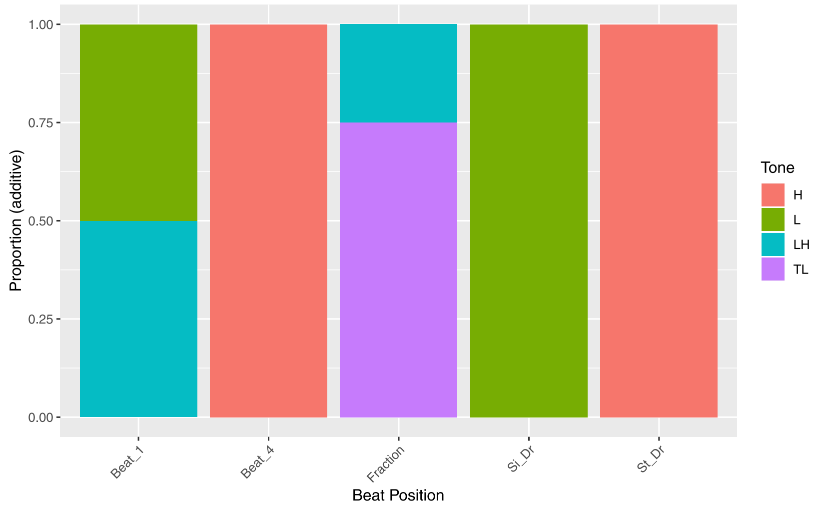


FIGURE 9 | Alignment of syllables by tone in the opening lines/chorus of Song 1 represented as proportions of occurrences.

see that certain compound forms are treated similarly in terms of their note lengths to disyllabic stems like *cóbə*, such as the last two syllables of the ndàp *tâkwáŋkə̀*. Here, the final two syllables of the traditional name are assigned the same note duration when combined as the first, contoured syllable of the name.

In contrast to examples (13) and (15), Line 2 presented in example (14) has an additional syllable, and a somewhat different rhythmic structure from the other two lines—we refer to this as Song 1, Motif 2 (abbreviated as S1M2; four of the ten lines analyzed followed this pattern). Lines like this are partially lexicalized and always feature a measure with several tied pairs of triplet eighth notes and second measure containing some variant of the greeting *ú ják jù lú à* (roughly, "How are you this afternoon?"). This last phrase is condensed into a little over half of a measure, with the first four syllables receiving single triplet eighth note lengths, and the final question particle receiving double the duration. Here, even the verb root is reduced in duration from what would have been expected in Lines 1 and 3. In contrast, in the first part of the line, the exclamative particle *mə̀* is realized with two triplet eighth notes, as opposed to the single eighth not it was associated to in lines 1 and 3. Thus, there appears to be flexibility in terms of note length for morphemes depending on whether they occur in a line of S1M1 or of S1M2.

Turning now to beat alignment patterns within the MD's lines, we analyze the morphological and tonal composition of each main beat, broken down by motif in the master drummer's singing for Song 1, in **Figures 11**, **12**. We note first that S1M2 involved no syllables articulated on either Beat 4 or on the beat fraction articulated by the sitting drum played by the master drummer. Apart from this, the specific morphological makeup of each syllable is quite varied across the two motifs: for example, while Beat 1 in S1M1 is made up nearly entirely of stem-initial syllables, Beat 1 in S1M2 is primarily made up of pronouns, with Beat 2 containing some stem-initial syllables, as well. Both motifs show a higher proportion of high tones on main beats than on beat fractions (including those fractions which are articulated by the drummers). Interestingly, contour tones appear to be entirely absent on main beats in S1M1, being more common on beat fractions (in spite of the similar tone patterns and morphological structures of contours found across motifs). In S1M2, contours occur on Beats 1 and 3 as well as on beat fractions (including the one articulated by the standing drum), but not on Beat 2. Within both motifs, we see considerably more variability in morpheme type occurring at beat fractions, as opposed to on whole beats. Certain syllable types, including stem-final syllables, prefixes, vocables, exclamative particles, question particles, and other discourse particles only occurred on beat fractions. Interestingly, the morphological makeup of the sitting drummer's accented beat fraction is both more consistent and more likely to contain stem-initial syllables (either monosyllabic stems, or as a part of an associative construction) than other beat fractions.

Song 2: Line-by-Line Analysis

Moving now to Song 2, we will only discuss the master drummer's song patterns, since the regular chorus in this song is composed entirely of vocables. Later in the song, the chorus repeats two verses identically to the master drummer—this pattern will be described below in the context of Song 2, Motif 2 (S2M2) of the master drummer's song. Transcriptions, glosses, and translations for four of the ten lines analyzed for this song are given in (16–19). Similar to findings for Song 1, we see in

Line 1, master drummer verse (Motif 1 / S1M1)												
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**				
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS		wò			mə	ndǔn		làm		có	bə́	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**				
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	jà		o		jà							
Line 3, master drummer verse (Motif 1 / S1M1)												
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**				
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS			é		kə	lɛ́n		tâ		kwá	ŋkə'	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**				
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	mvɛ́t		ʙʷǒ		ndà							
Line 2, master drummer verse (Motif 2 / S1M2)												
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**				
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS			wó		mə		wǔ		və́		nsí	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**				
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	ú	ják		jù	lú	à						

FIGURE 10 | Grid representation of three lines of master drummer's verses, Song 1. Numbers 1—4 represent whole-numbered beats on which the idiophone/shaker sound and dancers move. Individual cells represent beats on which the complex ndǔnlàm/metallophone sounds. St/Si represent beat fractions on which the Standing or Sitting drum sounds. Gray filled cells indicate beats on which the voice is sounding.

Figure 4 a pattern in which note lengths of monosyllabic stem syllables and contoured particles, such as *jú'*, *mɛ̂n*, *bâ*, and *jà* are longer (usually 2 triplet eighth notes), while pronouns such as *wʉ́*, and *jù* and affixes such as the iterative suffix *-də́* and the infinitival prefix *nə̀-* tend to occupy shorter note lengths, usually one triplet eighth note. There are, however, some exceptions to this. For example, the first syllable in the traditional name *ŋgú'sàm* (occurring in measure 4), which appears to derive from a possessed nominal construction (*sàm* being one class of possessive enclitic pronoun), is only one triplet eighth note in length. Interestingly, *sàm* bears the same note length—the equivalent of two eighth note triplets—as the preceding contoured word *mɛ̂n* "child" (modified with a high associative marker). This does not seem to be an isolated occurrence, as the pronoun enclitic, *cám* (occurring in the following measure), bears the same note duration of two eighth note pulses, which is equal to that of the monosyllabic noun it modifies, *fɛ́t* "brother." Thus, it seems that possessive enclitics deviate from other pronouns in having note lengths more similar to noun stems. It is interesting to note that these pronominal enclitics tend to have a CVC shape, whereas subject pronouns more often have a V or CV shape. This could reflect a phonological weight distinction among pronouns of different types. However, note that stems of both CV and CVC shapes, regardless of whether they bear a level or contour tone, tend to pattern similarly in note length, suggesting that different syllable shapes does not universally map to different phonological weights, at least in a way that is observable through these songs.

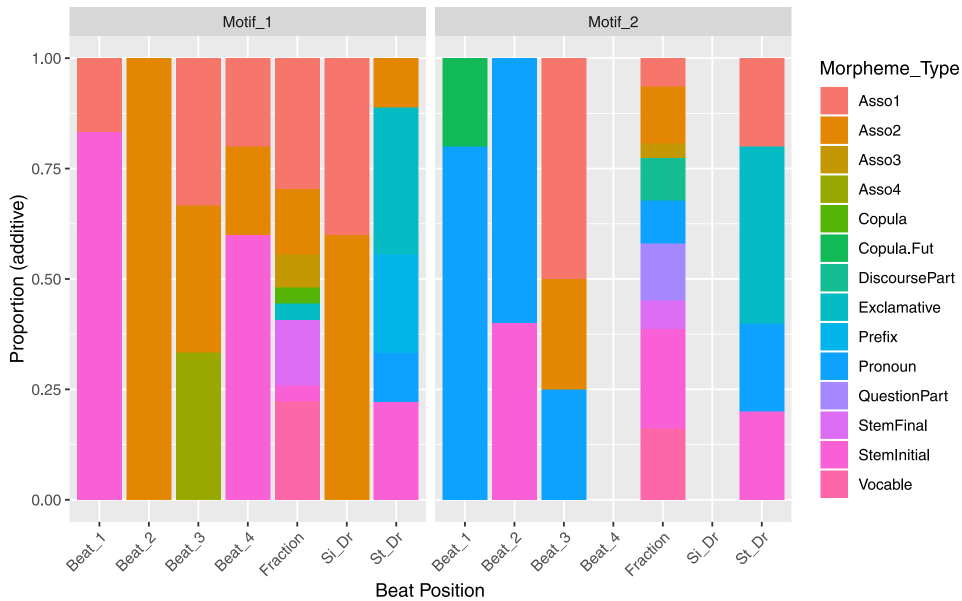


FIGURE 11 | Alignment of syllables by morpheme type and motif in master drummer's verses of Song 1 represented as proportions of occurrences.

(16) Line 1, master drummer verse, Song 2

wó	mɛ̂n	wʉ́	bâ	jù	dí	ó	ŋkwá	làm
wo	mɛ́n.L	wʉ́	bə́.à	jù	dí	o	ŋkwá	làm
VOC	child.ASSO	Who	COP + FOC	2SG.OBJ	DISC	DISC	Song	làm

"Whose child represents the song you choose?"

(17) Line 2, master drummer verse, Song 1

wó	mɛ̂n	ŋgú'sàm	bâ	jà	dí	fɛ́t	cám
wo	mɛ́n.L	ŋgú'sàm	bə́.à	jà	dí	fɛ́t =	cám
VOC	child.ASSO	ŋgú'sàm	COP + FOC	Where	DISC	brother	1SG.POSS

"My brother, where is the child of ŋgú'sàm?"

(18) Line 3, master drummer verse, Song 1

ù	jú'-də́	ŋkwá	nə̀-jú'-də́	BʷǒBʷò
u	jú'-də́	ŋkwá	nə̀-jú'-də́	BʷǒBʷò
2SG	groove.ITER	Dance	INF-groove-ITER	beautiful.REDUP

"Dance sensually, beautiful ones."

(19) Line 4, master drummer verse, Song 1

ò	báɣə-də́	ŋkwá	nə̀- báɣə-də́	BʷǒBʷò
o	báɣə-də́	ŋkwá	nə̀- báɣə-də́	BʷǒBʷò
2SG	bend.ITER	Dance	INF-bend-ITER	beautiful.REDUP

"Bend and dance, beautiful ones."

Examining now the beat alignment of syllables in Song 2, we break the song down again into two different motifs, based on lexical content and rhythmic patterning. Song 2, Motif 1 (abbreviated S2M1), represented by Lines 1 and 2 as presented in (16) and (17), typically consists of 9 syllables, while Song 2, Motif 2 (abbreviated S2M2), represented by Lines 3 and 4 as presented in (18) and (19), consists of 9–10 syllables. Seven out of the ten lines analyzed followed the pattern of S2M1, and three followed the pattern of S2M2. Once again, there is strong (but not perfect) rhythmic parallelism within S2M1, despite lexical and morphological variation. Lexical content varies considerably across motifs, as can be seen by comparing examples (16–17) with examples (18–19). In particular, the verb forms found in (18–19), which are a fixture of this motif, include both prefixes and suffixes; though such morphological elements do occur within some lines of S2M1, they occur less consistently, as evidenced by their absence in (16–17). S2M1 also contains a greater number of stem-initial syllables on average per line than S2M2. As can be seen in **Figure 13**, many of the same beats are occupied across both motifs, though there are some differences, such as the presence of a syllable on the standing drummer's accented beat fractions.

Figures 14, **15** show more striking differences between the two motifs by breaking down beat positions by morpheme type and tone. Whereas stem-initial syllables are favored on Beat 1 in S2M1, Beat 1 is comprised entirely of prefixes in S2M2. Likewise, Beats 2 and 4 in S2M2 are made up entirely of suffixes, while affixes in S2M1 occur exclusively on beat fractions, which are also primarily occupied by stem-final syllables, question particles, and other types of discourse particles. Though the greater scarcity of stem-initial syllables within S2M2 may explain some of this difference in patterning, it is striking that stem-initial syllables, where they do occur in S2M2, occur exclusively on beat fractions, and never on

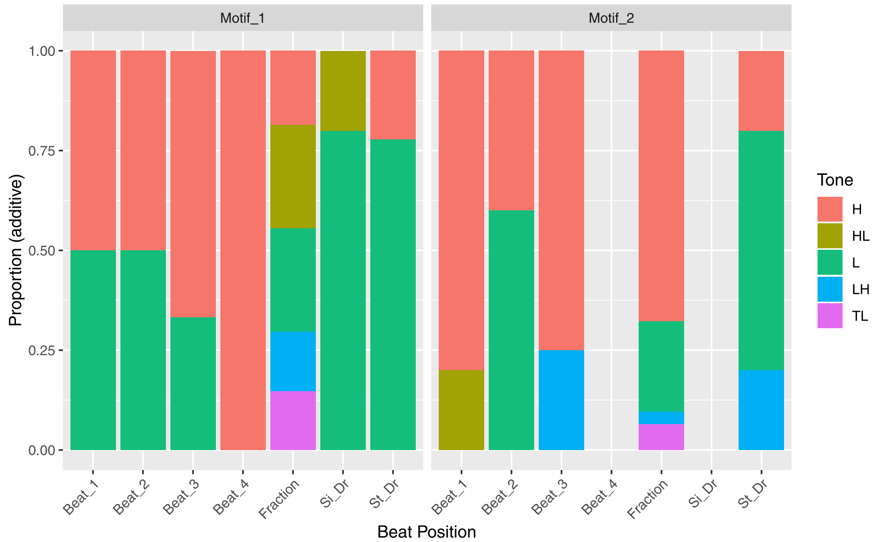


FIGURE 12 | Alignment of syllables by tone and motif in master drummer's verses of Song 1 represented as proportions of occurrences.

main beats. The greatest proportion of stem-initial syllables are in fact found on the beats articulated by the sitting drum/master drum, who sounds his drum only in the presence of stem-initial syllables. As with Song 1, beat fractions are more variable across both motifs in Song 2 than main beats in terms of the types of morphemes they host.

In terms of tone, we see much more variability overall in S2M1 than in S2M2 in terms of which tones occur. In both motifs, interestingly, low tones appear to be strongly preferred on Beat 1; they are also preferred on Beat 3 in S2M1. Outside of Beat 1, high tones are strongly preferred in all positions in S2M2. Contour tones are once again seen to be just as common, if not more common, on fractions of beats as on whole numbered beats in Song 2.

DISCUSSION

Our analysis illustrates the creative flexibility of rhythm in Medʉmba *Làm* folk songs while also highlighting some consistent patterns in the mapping between music and language. To begin with, rhythmic variation is highly dependent on the voices and parts being considered: at the level of the percussion ensemble, the drummers clearly have more flexibility to deviate from the underlying pulse of the music than do the other instruments. Likewise, as the storyteller and creative center piece within the ensemble, the master drummer has far more rhythmic flexibility in sung portions of the music compared with the chorus who responds to him. However, rhythmic "deviations" in both drumming and singing are systematic: though patterns vary measure by measure, there are recurring intervals at which the two drums sound (and conspicuously do not sound) throughout the songs. In particular, regular occurrence of drumbeats on the second note within triplet eighth note groupings in **Figures 3**, **4** can be interpreted as an alternate rhythmic line which is felt in 3, rather than 2, which is highlighted by the drummers. Such a pattern of conflicting duple and triple rhythmic lines overlaid on one another is consistent with a polyrhythmic analysis of this music, a pattern which is often found in the music of Central and West Africa, more generally (Temperley 2000). We can conclude from all of this that text-setting in Medʉmba *Làm* folk songs is governed by multiple underlying rhythms; we unpack this observation a bit more below.

Rhythmic Alignment and Mismatch Between Language and Song

In spite of the considerable rhythmic variation noted within and among the songs in our dataset, we find substantial alignment between song structure and linguistic structure in the sung parts of both the chorus and the master drummer voices. In the first song, both the chorus refrain and S1M1 of the master drummer's verses reflect a preference for stem-initial syllables (whether occurring within isolated stems or as a part of an associative construction) to be placed on whole-numbered beats within the staff-notated lines. Meanwhile, stem-final syllables and other prosodically weak particles are relegated to beat fractions. Indeed, the prevailing pattern is one in which whole-numbered beats within the song are either occupied by stem-initial syllables or are completely empty; rarely are they found to host a prosodically weak element.

S1M2 breaks with this pattern somewhat in that pronoun—including many subject pronouns, which consistently receive shorter note lengths—are found in abundance on main beats. Pronouns also feature heavily into main beats in S2M1, where they also occur in parallel positions with stem-initial syllables on several beats. Even within rhythmic motifs where pronouns are given main beat status, however, we find that other types of function words,

	Line 1, master drummer verse (Motif 1 / S2M1)											
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS					wó			mɛ̂n		wʉ́	bâ	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	jù	dí	ó			ŋkwá	làm					
	Line 2, master drummer verse (Motif 1 / S2M1)											
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS					wó			mɛ̂n		ŋgú'	sàm	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	bâ		jà	dí		fɛ́t	cám					
	Line 3, master drummer verse (Motif 2 / S2M2)											
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS							ù	jú'		də́	ŋkwá	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	nə̀	jú'		də́	ʙʷɔ̌		ʙʷó					
	Line 4, master drummer verse (Motif 2 / S2M2)											
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS						ò		bá	ɣə́	də́	ŋkwá	
DRUM-ACCENTED FRACTIONS					**St.D**			**Si.D**			**St.D/ Si.D**	
WHOLE-NUMBERED BEATS	**1**			**2**			**3**			**4**		
LYRICS	nə̀	bá	ɣə́	də́	ʙʷɔ̌		ʙʷó					

FIGURE 13 | Grid representation of three lines of master drummer's verses, Song 2. Numbers 1–4 represent whole-numbered beats on which the idiophone/shaker sound and dancers move. Individual cells represent beats on which the complex ndǔnlàm/metallophone sounds. St/Si represent beat fractions on which the Standing or Sitting drum sounds. Gray filled cells indicate beats on which the voice is sounding.

including discourse particles, as well as stem-final syllables, remain confined to beat fractions instead of main beats. These observations are consistent with a view whereby pronouns occupy a kind of middle-ground prosodically, neither consistently weak, nor consistently strong. On the subject of pronouns, we also found that certain types of pronouns—namely possessive pronominal enclitics—are consistently produced with longer note duration, and can also occur at musically strong positions. This is consistent with earlier work demonstrating rhythmic prominence of these enclitic syllables within non-musical speech production tasks (Franich, 2017; Franich, 2018; Franich, 2021).

Interestingly, the overall pattern of alignment between stem-initial syllables and whole-numbered beats is completely reversed in the context of S2M2: here, affixes reign supreme in filling main beat positions, while stem-initial syllables are consistently located on beat fractions. The systematicity with which these two types of syllables—affixes, on the one hand, and stem-initial syllables, on the other—are swapped in terms of their beat positions across motifs within Song 2 is a

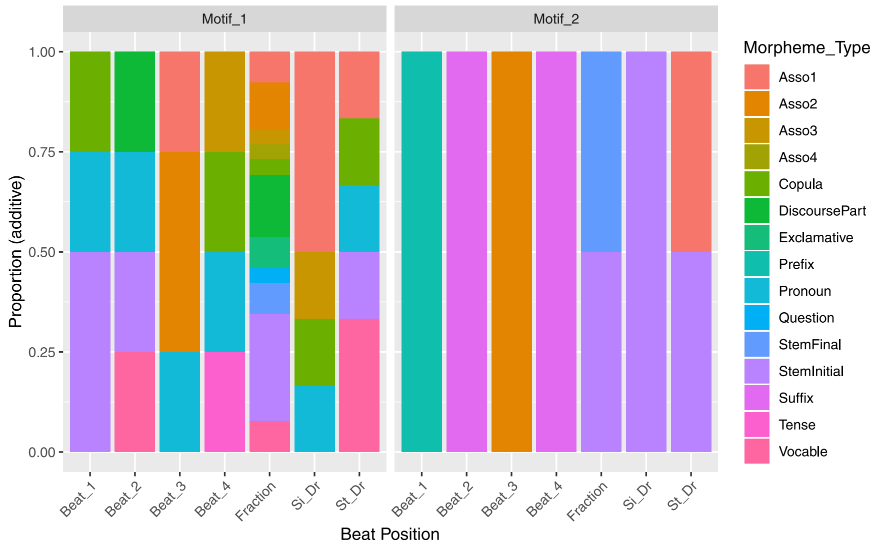


FIGURE 14 | Alignment of syllables by morpheme type and motif in master drummer's verses of Song 2 represented as proportions of occurrences.

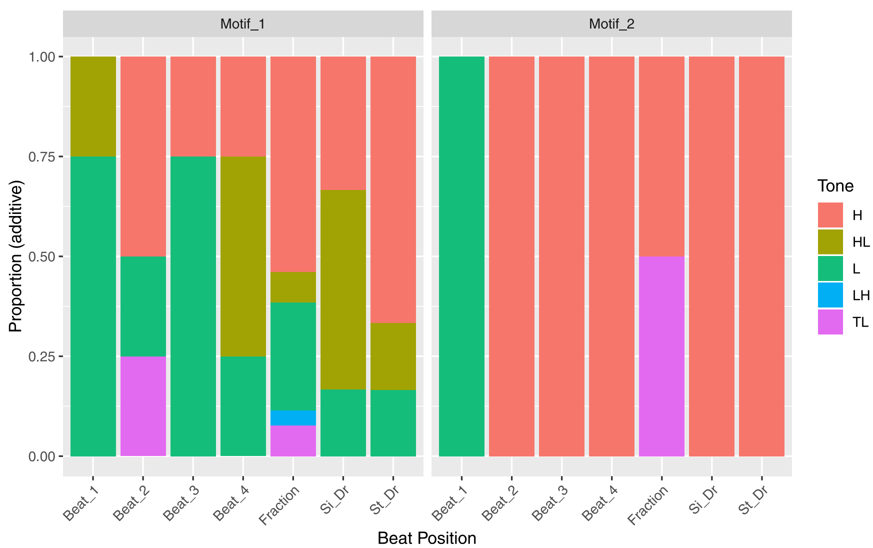


FIGURE 15 | Alignment of syllables by tone and motif in master drummer's verses of Song 2 represented as proportions of occurrences.

testament to the internal consistency of each syllable type regarding their rhythmic role. S2M2 is somewhat of an outliner across all of the rhythmic structures investigated, suggesting that the exchange of rhythmic roles between affixes and stem-initial syllables reflects a subversion of the usual link between the primary rhythmic pulse of the song and linguistically prominent stem-initial syllables. In this way, this pattern somewhat resembles what was described for Lardil *burdal* songs in §1 (Nancarrow 2010). Whether this pattern is idiosyncratic in Song 2 as presented here, or representative of a more broadly employed creative device in text-setting, is unclear; data from additional songs and genres would be useful in answering this question.

Another interesting source of consistency we find in the mapping of musical rhythm to language structure is in the syllables on which the master drummer accents beat fractions outside of the main beats on the *mə̂nfámə́* sitting drum. Across the three different motifs (S1M1, S2M1, and S2M2), the morphological makeup of these beats is typically one of the most consistent of any beat position, and considerably more consistent than what is found for the accented beat fractions of the *tómŋkàʼ* standing drum. The morphology corresponding to the accented fractions of the *mɛ̂nfámə́* is also similar to what is found on main beats, in that stem-initial syllables and pronouns tend to be favored in these positions when those types of syllables are overall aligning with main beats within a particular motif, and affixes are preferred when those syllables are instead populating the main beats within a motif. Indeed, it is as if these beat fractions are behaving themselves as main beats, which falls out naturally if we assume that they *are*, in fact, main beats, when contextualized within a competing triple rhythm overlaid on the duple rhythm being reinforced by the idiophone/shaker and dancers. The consistency of the morphological makeup on the master drummer's articulated beats also makes sense if we consider the difficult rhythmic task of the master drummer, who is essentially reinforcing *both* rhythms at once through his singing and his playing: this is likely more easily done if what the hands and the voice are doing is rhythmically and structurally consistent and somewhat predictable.

Tone, Syllable Structure, and Text-Setting

In terms of tone, we again found considerable variation in text-setting patterns. In most rhythmic motifs, high tones tended to be preferred on main beats, with low tones more common on beat fractions, and toneless syllables occurring exclusively on beat fractions. A major exception to this is in Beat 1 of Song 2 for both motifs, where low tones were more prevalent (they were also more prevalent on Beat 3 within the first motif). Interestingly, contour tones, where they did occur, tended to be as common—if not more so—on beat fractions as on main beats. An exception to this is in S1M2, where contours occurred to some degree on both Beats 1 and 3. Overall, however, there does not seem to be any strong rhythmic preference when it comes to the positioning of contour tones within the rhythmic line.

Turning now to note durations, we found that contoured syllables did consistently occur with longer note lengths (usually two eighth notes/one quarter note) in the data set. We observed no clear differences in alignment or duration between words bearing contour tones arising from their own underlying lexical tone (e.g., *ʙwǒ* "beautiful"), those derived from concatenation of a floating tone morpheme (e.g., *mɛ̂n* "child.asso"), or those derived from a tonal overwrite process (e.g., *lɛ̂n* "know.subj"); rhythmic differences between LH rising and HL falling contours were also not apparent in our data set. There was also no consistent difference between contour-toned syllables of different syllable shapes, as might have been predicted if musical note lengths were based on pure durational differences across syllables. This suggests that contour tones were given a uniform treatment in terms of note length, apparently reflective of a more abstract mapping between linguistic structure and musical beat structure.

Importantly, it was not only contour-toned syllables which bore relatively longer note lengths in the songs: monosyllabic stems of all syllable shapes also tended to bear similar note lengths of two eighth notes/one quarter note. Thus, it seems tone is not the primary driver of increased note length. Note lengths for disyllabic stems also tended to constitute a total of two eighth notes, with a single eighth note assigned to each syllable. Given that stems are thought to align with feet in this language (see §3.1), assuming a moraic account of syllable weight, this pattern could reflect the very common cross-linguistic constraint requiring that feet minimally contain two moras (Prince and Smolensky 1993). However, it is also possible that this pattern reflects a phonetically based pattern of foot-internal shortening, rather than a weight-based pattern (Fowler 1977; Kim and Cole 2006).

Another area of the grammar in which syllables are consistently mapped to shorter note lengths is among functional item such as discourse particles and affixes. It is well-known that function words in the world's languages tend to be prosodically weaker than lexical words; thus, this pattern may simply be a reflection of relative syllable prominence. We note, though, that stem-initial syllables, despite bearing greater rhythmic prominence, are only realized with greater note length in monosyllabic words, and not within disyllabic stems; thus, there is no perfectly straightforward mapping between syllable prominence and note length in our dataset. This gives us further reason to consider a weight-based analysis in which functional items are both prosodically weak and light in terms of their syllable weight. We conclude that there is still no clear answer from the present work about whether Medʉmba does contrast syllables in terms of weight, but it appears a purely phonetic explanation of the mapping between syllables and note lengths falls short of capturing some of the generalizations that speakers/singers are making when enacting this mapping.

Finally, we note that there are several cases in which mismatches occur between musical beat strength and note length in the songs presented here. For example, while contour-toned syllables and monosyllabic stems regularly occur with longer note lengths, they need not always be positioned on "strong" musical beats; conversely, though subject pronouns tend to bear shorter note lengths, they also consistently occur on main beats within the rhythmic line. This is in spite of our earlier observation that stem-initial syllables and monosyllabic stems (which also tend to take up longer note lengths) are generally preferred on main beats in most rhythmic motifs, with stem-final syllables and affixes (which bear shorter note lengths) occurring outside of these beats. The relationship between note length and beat strength is therefore not absolute, and is almost certainly modulated by other factors, such as overall position within the sentence/phrase.

Song-Based Deviations From Linguistic Structure

Finally, we note that grammatical structure within songs can deviate in some interesting ways from what is expected in everyday speech. For example, unlike in regular speech, stem-final syllables are able to be overtly produced phrase-medially in these songs. One example of this is the initial line/chorus of Song 1, in which the word *cóbə* "speak" is produced with two syllables, where it would normally be produced without the final vowel. Also missing from this construction as it would usually occur is the focus marker *à* which would normally occur between the verb *cób* 'speak' and the following directional *ʙwə̀* "there" (i.e. *à cób **à** ʙwə̀*). It is interesting to note that the verb consistently occurs with an eighth note rest after it in Song 1, almost as if rhythmic space is being held for the focus marker, even though it isn't being produced. Another place where a stem-final syllable is produced phrase-internally is in Song 2, where the group sings *ó báɣə́-də́ ŋkwá* ("o, bend and dance"), instead of *ò bá**k**-də́ ŋkwá*, as would be expected in regular speech. These patterns may reflect an older form of the language where stem-final syllables were produced regularly in all contexts, before being reduced in phrase-internal position; however, the historical evolution of these syllables is as yet unclear. It is also possible that these patterns simply reflect an idiosyncrasy of song structure which deviates from the spoken language grammar.

CONCLUSION

The present work has provided description and analysis of mapping patterns between language and music in Medʉmba folkloric songs from the *Làm* blacksmith tradition. Our findings provide support for earlier work suggesting that syllables in the language differ in terms of their prosodic strength: as predicted, stem-initial syllables in the language, which are analyzed as metrically prominent, tend to be preferred on musically prominent beats, while non-initial and non-stem syllables, generally speaking, are preferred on less prominent beats. However, we have also highlighted various aspects of rhythmic flexibility which suggests that mapping between linguistic strength and musical strength is not absolute in these songs. Nonetheless, even where mismatches between musical structure and linguistic structure did emerge, they were principled, in that syllables of distinct morpho-prosodic status tended to pattern together in either being attracted to, or repelled by, musically strong beats within any given rhythmic motif. Our results thus highlight the fact that rhythmic variability of a particular morpho-prosodic structure within song does not indicate its lack of rhythmic status, but rather a flexibility in the mapping constraints between language and music. Variability in mapping constraints can only be ascertained through careful analyses of distinct rhythmic modes within and across songs. Finally, it should be noted that the present work represents a rich, but small dataset on Medʉmba speech-based song. Undoubtedly, a wider survey of musical genres from the region will reveal additional patterns and complexities which can increase our understanding of musical and linguistic grammars, and mappings between the two.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Delaware Institutional Review Board. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

KF conceived of the study and oversaw all data collection, data processing, analyses, and manuscript preparation. ALN assisted with data collection and contributed to data interpretation and manuscript writing.

FUNDING

This work was supported by National Science Foundation Linguistics Program Grant No. BCS-2018003 (PI: KF). The National Science Foundation does not necessarily endorse the ideas and claims in this paper. All errors are our own.

ACKNOWLEDGMENTS

We thank Paul Hansen and Michael Lewis for their generous help with music transcription and discussions about African rhythm, as well as Kofi Agawu, Tim Broscious, and Gene Koshinski for additional discussion and feedback. As the authors we take responsibility for any and all errors. We are deeply grateful to Guy-Flaubert Kuitchou for coordinating the music for the *ntánlá'* and for his perspective as a master drummer. We also thank Emmanuel Kuitcheu, Ernest Kwanga, Tagni Amos Datcheu, Magni Madeleine Tchouta and the many members of the Lafeng community who participated in the *ntánlá'* musical ensemble and contributed to this study. We thank His Majesty Mbiandou Yogang Bernard, Chief of the Lafeng District, and His Majesty Yonkeu Kuika Jean, Paramount Chief of the Bangoulap community, for their support. The music and lyrics presented in this text and accompanying recordings are the intellectual property of the Lafeng *ntánlá'* association. Finally, we thank audiences at Boston University, Keio University, the International Christian University, Japan, and New York University for helpful feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.653747/full#supplementary-material

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago, IL: Aldine-Atherton.

Agawu, V. K. (1995). *African Rhythm: A Northern Ewe Perspective*. Cambridge University Press.

Akinlabi, A., and Urua, E. E. (2003). Foot Structure in the Ibibio Verb. *J. Afr. Languages Linguistics* 24 (2), 119–160. doi:10.1515/jall.2003.006

Anku, W. (2000). Circles and Time: A Theory of Structural Organization of Rhythm in African Music. *J. Soc. Music Theor.* 6, 1.

Chan, M. K. M. (1987). Tone and Melody Interaction in Cantonese and Mandarin Songs. *UCLA Working Pap. Phonetics* 68, 132–169.

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press. doi:10.4159/harvard.9780674337145

Chernoff, J. M. (1991). The Rhythmic Medium in African Music. *New Literary Hist.* 22 (4), 1093. doi:10.2307/469080

Cho, S.-H. (2017). Text Alignment in Japanese Children's Song. *Penn Working Pap. Linguistics* 23, 1. Accessible online at http://repository.upenn.edu/pwpl/vol23/iss1/5/.

Clements, G. N., and Goldsmith, J. (1984). *Autosegmental Studies in Bantu Tone*. Dordrecht, Netherlands: Foris.

Danis, N. (2011). "Downstep and Contour Formation in Medumba: A Prosodic Account," in Proceedings of the 42nd Annual Conference on African Linguistics.

Dell, F., and Halle, J. (2009). "Comparing Musical Textsetting in French and in English Songs," in *Towards a Typology of Poetic Forms*. Editors Jean-Louis, Aroui, and Andy. Arleo (Amsterdam: John Benjamins), 63–78. doi:10.1075/lfab.2.03del

Dimmendaal, G. (2012). Metrical Structures: A Neglected Property of Nilotic (And Other African Language Families). *Stud. Nilotic Linguistics* 5, 1–26.

Downing, L. (2010). "Accent in African Languages," in *A Survey of Word Accentual Patterns in the Languages of the World*. Editors R. W. N. Goedemans, H. G. van der Hulst, and E. A. van Zanten (Berlin, Germany: Mouton de Gruyter), 381–427.

Downing, L. J. (1990). Local and Metrical Tone Shift in Nguni. *Stud. Afr. Linguistics* 21, 261–317.

Downing, L. J. (2004). What African Languages Tell Us about Accent Typology. *Zaspil* 37, 101–136. doi:10.21248/zaspil.37.2004.247

Duanmu, S. (1990). *A Formal Study of Syllable, Tone, Stress and Domain in Chinese languages*. Cambridge (MA): MIT. [dissertation].

Duanmu, S. (1996). Pre-juncture Lengthening and Foot Binarity. *Stud. Linguistic Sci.* 26, 95–115.

Duanmu, S. (2004). Tone and Non-tone Languages: An Alternative to Language Typology and Parameters. *Lang. linguistics* 5 (3), 891–924.

Espinosa, A. M. (1924). *Synalepha and Syneresis in Modern Spanish*. Washington: Hispania, 299–309. doi:10.2307/331182

Fitzgerald, C. M. (1998). The Meter of Tohono O'odham Songs. *Int. J. Am. Linguistics* 64 (1), 1–36. doi:10.1086/466345

Fowler, C. A. (1977). *Timing Control in Speech Production*. Bloomington: Indiana University, Linguistics Club.

Franich, K. (2014). Contour tones and prosodic structure in Medʉmba, in *Proceedings of the 40th Annual Meeting of the Berkeley Linguistic Society*. Editor H. Lueng, 102–124.

Franich, K. (2021). *Metrical Prominence Asymmetries in Medʉmba, a Grassfields Bantu Language.* Language 92 (2).

Franich, K. (2017). *The Interaction of Prominence, Rhythm, and Tone in Medʉmba*. Chicago (IL): University of Chicago. [dissertation].

Franich, K. (2018). Tonal and Morphophonological Effects on the Location of Perceptual Centers (P-centers): Evidence from a Bantu Language. *J. Phonetics* 67, 21–33. doi:10.1016/j.wocn.2017.11.001

Goldsmith, J. (1976). *Autosegmental Phonology*. Cambridge (MA): MIT. [dissertation].

Gordon, M. (2001). A Typology of Contour Tone Restrictions. *Stud. Lang.* 25, 405–444. doi:10.1075/sl.25.3.03gor

Gordon, M. (1999). *Syllable Weight: Phonetics, Phonology, and Typology*. Los Angeles, CA: UCLA. [PhD Thesis].

Grabe, E., and Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. *Pap. Lab. Phonology* 7.

Green, C. R. (2015). The Foot Domain in Bambara. *Language* 91 (1), e1–e26. doi:10.1353/lan.2015.0009

Guthrie, M. (1948). *The Classification of the Bantu Languages*. Oxford University Press.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: University of Chicago Press.

Huron, D., and Ollen, J. (2003). Agogic Contrast in French and English Themes: Further Support for Patel and Daniele (2003). *Music Perception* 21 (2), 267–271. doi:10.1525/mp.2003.21.2.267

Hyman, Larry. (1985). *A Theory of Phonological Weight*. Dordrecht: Foris. doi:10.1515/9783110854794

Hyman, L. M. (2014). "Do all Languages Have Word Accent?," in *Word Stress: Theoretical and Typological Issues*. Editor H. van der Hulst (Cambridge University Press), 56–82.

Hyman, L. M. (2015). *Positional Prominence vs. Word Accent: Is There a Difference?* Berkeley, CA: UC Berkeley Phonology Lab Annual Report, 86–99.

Hyman, L. M. (2003). Why Describe African Languages? 4th World Congress on African Linguistics/34th Annual Conference on African Linguistics. Rutgers University.

Hyman, L. M., Rolle, N., Sande, H., Chen, E., Jenks, P., Lionnet, F., et al. (2019). "Niger-Congo Linguistic Features and Typology," in *The Cambridge Handbook of African Linguistics and A History of African Linguistics*. Editor E. Wolff (Cambridge University Press).

Hyman, L. M., and Tadadjeu, M. (1976). Floating Tones in Mbam-Nkam. *South. Calif. Occas. Pap. Linguistics* 3, 57–111.

Jo-Keeling, S. R. (2011). *'Musicking Tradition in Place: Participation, Values, and Banks in Bamiléké Territory'*. Ann Arbor (MI): University of Michigan. [dissertation].

Ketkaew, C., and Pittayaporn, P. (2014). Mapping between Lexical Tones and Musical Notes in Thai Pop Songs. Proceedings of the 28th Pacific Asia Conference on Language. *Inf. Comput. (Paclic)* 28, 160–169.

Keupdjio, H. S. (2020). *The Syntax of A'-dependencies in Bamileke Medumba*. Vancouver(BC): University of British Columbia. [dissertation].

Kim, H., and Cole, J. (2006). "*Evidence for Rhythm Shortening in American English as Conditioned by Prosodic Phrase Structure*," To appear in the Proceeding of 42nd annual Meeting of the Chicago Linguistic Society.

Kiparsky, P. (1979). Metrical Structure Assignment Is Cyclic. *Linguistic Inq.* 10 (3), 421–441.

Kiparsky, P. (2016). *Stress, Meter, and Text-Setting. The Handbook of Historical Phonology*. Oxford, UK: Oxford University Press.

Kirby, J., and Ladd, D. R. (2016). "Tone-melody Correspondence in Vietnamese Popular Song," in Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL-2016), 48–51.

Kubik, G. (1983). ""Transkription Afrikanischer Musik Vom Stummfilm" [Transcripton of African Music from Silent Film]," in *Musik in Afrika*. Editor Artur. Simon (Berlin: Museum für Völkerkunde), 202–216.

Ladd, D. R., and Kirby, J. (2020). "Tone–Melody Matching in tone-Language Singing," in *The Oxford Handbook of Language Prosody*. Editors C. Gussenhoven and A. Chen (Oxford University Press), 676–687.

Leben, W. (2002). Tonal Feet and the Adaptation of English Borrowings into Hausa. *Stud. Afr. Linguistics* 25, 139–154.

Leben, W. (2003). "Tonal Feet as Tonal Domains," in *Trends in African Linguistics 5: Linguistic Typology and Representations of African Languages*. Editor J. Mugane (Trenton, NJ: Africa World Press), 129–138.

Leben, W. R. (1971). *Suprasegmental Phonology*. Cambridge (MA): MIT. [dissertation].

Lehrdal, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lendja, N. A. B. (2016). "La territorialisation/dérritorialisation du corps du roi: Bamiléke-Bangoulap (Cameroun)," in *Les territoires du sacré: images, discours, pratiques*. Editors R. Ziavoula, A. Kouvouama, and P. Yengo, 73–111.

Lerdahl, F. (2001). The Sounds of Poetry Viewed as Music. *Ann. N. Y. Acad. Sci.* 930, 337–354. doi:10.1111/j.1749-6632.2001.tb05743.x

Liberman, M. (2007). *Rock Syncopation: Stress Shifts or Polyrhythms?* LanguageLog (blog), published. Retrieved from http://itre.cis.upenn.edu/~myl/languagelog/archives/005154.html.

Liberman, M. (1975). *The Intonational System of English*. Cambridge (MA): MIT. [dissertation].

Liberman, M., and Prince, A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inq.* 8, 249–336.

McPherson, L., and Ryan, K. M. (2018). Tone-tune Association in Tommo So (Dogon) Folk Songs. *Language* 94, 119–156. doi:10.1353/lan.2018.0003

Mead, R. (2007). *Text-to-tune Alignment in the Music of La Charanga Habanera*. Palo Alto (CA): Stanford University. [undergraduate thesis].

Mesthrie, R., and Bhatt, R. M. (2008). *World Englishes*. Cambridge： Cambridge University Press. doi:10.1017/cbo9780511791321

Mkammi, J. (2009). *Les noms et les titres de noblesse dans le Nde*. Bangangte, Cameroon: Comité de Langue pour l'Etude et la Production des Œuvres Bamiléké-Medumba.

Moore, C. B. (1993). Phonetic Observations on Tone and Stress Levels in Mandarin Classifiers. Proceedings of the 5th North American Conference on Chinese Linguistics 29.

Morgan, T. A., and Janda, R. D. (1989). "Musically-conditioned Stress Shift in Spanish Revisited: Empirical Verification and Nonlinear Analysis," in *Studies in Romance Linguistics, Selected Proceedings from the XVII Linguistic Symposium on Romance Languages*. Editors C. J. Kirschner and A. DeCesaris (New Jersey: Rutgers University).

Nancarrow, C. (2010). What's that Song about?: Interaction of Form and Meaning in LardilBurdalSongs. *Aust. J. Linguistics* 30 (1), 81–92. doi:10.1080/07268600903134046

Odden, D. (1999). "Typological Issues in Tone and Stress in Bantu," in Proceedings of the Symposium on Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics. Editor Shigeki Kaji (Tokyo: ILCAA), 187–215.

Olson, K., and Meynadier, Y. (2015). "On Medumba Bilabial Trills and Vowels," in Proceedings of the 18th International Congress of Phonetic Sciences (Glasgow, Scotland: The Scottish Consortium for ICPhS 2015).

Patel, A. D., and Daniele, J. R. (2003). Stress-Timed vs. Syllable-Timed Music? A Comment on Huron and Ollen (2003). *Music Perception* 21, 273–276. doi:10.1525/mp.2003.21.2.273

Pearce, M. (2007). "Iambicity in Kera," in Proceedings of the 37th Annual Conference on African Linguistics. Editors D. L. Payne and J. Pena (Somerville, MA: Cascadilla Proceedings Project), 66–76.

Pearce, M. (2006). The Interaction between Metrical Structure and Tone in Kera. *Phonology* 23 (2), 259–286. doi:10.1017/s095267570600090x

Pike, K. (1945). *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Prince, A. (1983). Relating to the Grid. *Linguistic Inq.* 14, 19–100.

Prince, A., and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*.

Proto, T. (2016). "Methods of Analysis for Tonal Text-Setting. The Case Study of Fe'Fe' Bamileke," in Proceedings of the 5th International Symposium on Tonal Aspects of Language (Buffalo, New York). doi:10.21437/tal.2016-35

Ramus, F. (2002). Language Discrimination by Newborns: Teasing Apart Phonotactic, Rhythmic, and Intonational Cues. *Annu. Rev. Lang. Acquisition* 2. doi:10.1075/arla.2.05ram

Richards, P. (1972). A Quantitative Analysis of the Relationship between Language Tone and Melody in a Hausa Song. *Afr. Lang. Stud.* 13, 137–161.

Rycroft, D. (1959). African Music in Johannesburg: African and Non-african Features. *J. Int. Folk Music Counc.* 11, 25–30. doi:10.2307/834851

Schane, S. (1968). *French Phonology and Morphology*. Cambridge, MA: MIT Press.

Schellenberg, M. (2012). Does Language Determine Music in Tone Languages? *Ethnomusicology* 56 (2), 266–278. doi:10.5406/ethnomusicology.56.2.0266

Schellenberg, M. (2009). "Singing in a Tone Language: Shona," in Selected Proceedings of the 39th Annual Conference on African Linguistics: Linguistic Research and Languages in Africa. Editors A. Ojo and L. Moshi (Somerville, MA: Cascadilla Proceedings Project), 137–144.

Schuh, R. G. (2010). *The Form and Metrics of Ngizim Songs. Ms*. Los Angeles, CA: UCLA.

Schuh, R. G. (2014). Where Did Quantitative Metrics in Hausa and Other Chadic Songs Come from? Proceedings of the 1st International Conference on Endangered Languages. Kano: Japan, August 2014. doi:10.1628/978-3-16-153197-2

Temperley, D. (1999). *Syncopation in Rock: A Perceptual Perspective*. Cambridge University Press.

Temperley, D. (2000). Meter and Grouping in African Music: A View from Music Theory. *Ethnomusicology* 44 (1), 65. doi:10.2307/852655

Temperley, N., and Temperley, D. (2012). Stress-meter Alignment in French Vocal Music. *J. Acoust. Soc. Am.* 134 (1), 520–527. doi:10.1121/1.4807566

Toussaint, G. T. (2015). Quantifying Musical Meter: How Similar Are African and Western Rhythm? *Anal. Approaches World Music* 4, 1.

Toussaint, G. T. (2013). *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?* Taylor & Francis Group: CRC Press.

Turk, A. E., and White, L. (1999). Structural Influences on Accentual Lengthening in English. *J. Phonetics* 27, 171–206. doi:10.1006/jpho.1999.0093

Turpin, M., and Laughren, M. (2013). Edge Effects in WarlpiriYawulyuSongs: Resyllabification, Epenthesis, Final Vowel Modification. *Aust. J. Linguistics* 33 (4), 399–425. doi:10.1080/07268602.2013.857569

Vance, T. (1987). *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Vetter, R. (1996). *Rhythms of Life, Songs of Wisdom: Akan Music from Ghana*. West Africa: Compact disc recording and 24-page accompanying booklet. Smithsonian/Folkways SFCD 40463.

Voorhoeve, J. (1976). *Contes Bamileke*. Museé Royal de L'Afrique Centrale.

Voorhoeve, J. (1965). The Structure of the Morpheme in Bamileke (Bangangté Dialect). *Lingua* 13, 319–334.

Voorhoeve, J. (1971). Tonology of the Bamileke Noun. *J. Afr. Languages* 2, 44–53.

Voorhoeve, J. (1968). Noun Classes in Bamileke. *Lingua* 21, 584–593. doi:10.1016/0024-3841(68)90077-6

Wee, L-H. (2015). Unraveling the Relation between Mandarin Tones and Musical Melody. *J. Chin. Linguistics* 35 (1), 128–144.

Weidman, S., and Rose, S. (2006). A Foot-Based Reanalysis of Edge-In Tonal Phenomena in Bambara. Paper Presented at West Coast Conference on Formal Linguistics 25. Seattle, WA: University of Washington. doi:10.21236/ada463728

Wong, P. C. M., and Diehl, R. L. (2002). How Can the Lyrics of a Song in a Tone Language Be Understood? *Psychol. Music* 30, 202–209. doi:10.1177/0305735602302006

Yip, M. (1980). *The Tonal Phonology of Chinese*. Cambridge (MA): MIT. [dissertation].

Zhang, J. (2002). *The Effects of Duration and Sonority on Contour Tone Distribution–Typological Survey and Formal Analysis*. Los Angeles, CA: UCLA. [PhD Thesis].

Zhang, J. (2004). "The Role of Contrast-specific and Language-specific Phonetics in Contour Tone Distribution," in *Phonetically Based Phonology*. Editors B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press), 157–190. doi:10.1017/cbo9780511486401.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 01 July 2021
doi: 10.3389/fcomm.2021.643683

Rhythm-Speech Correlations in a Corpus of Senegalese Drum Language

*Sofiya Ros**

Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands

In some African cultures, drumming is used for expressing linguistic meanings. Our research focuses on Senegalese musical traditions of encoding linguistic messages on the *sabar* drums. Senegalese drummers have the practice of playing drums in correlation to speech. We consider rhythms and their linguistic correlates as being part of a *Sabar drum language.* The long-term goal of this investigation is to establish the linguistic properties of the Sabar drum language. To this end, this work relies on two kinds of research materials collected from Senegalese drummers: *bàkks* (classical sabar phrases, not improvised on the spot) and sabar improvisations including their translation to Wolof. We study the regularities between Wolof units and sabar rhythms in the collected data. We tested the hypothesis of a syllable-level correspondence between Sabar and Wolof, assuming that each sabar stroke represents a syllable or a number of syllables in Wolof, where the nature of the correspondence depends on the phonetic or phonological properties of a vowel in a syllable. The analysis has shown that different drum strokes are more commonly associated with different types of vowels (front, central or back; open, mid-open/mid-closed or closed vowels).

Keywords: drum language, Wolof, Senegal, music, phonology, drums, sabar

OPEN ACCESS

Edited by:
Laura McPherson,
Dartmouth College, United States

Reviewed by:
Harry Van Der Hulst,
University of Connecticut, United States
Ka Omar,
University of Maryland, Baltimore County, United States

****Correspondence:***
Sofiya Ros
s.ros@uu.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *18 December 2020*
Accepted: *21 June 2021*
Published: *01 July 2021*

Citation:
Ros S (2021) Rhythm-Speech Correlations in a Corpus of Senegalese Drum Language. Front. Commun. 6:643683. doi: 10.3389/fcomm.2021.643683

INTRODUCTION

Speech surrogates using drums are present in Africa, South America, Asia and Oceania (Stern, 1957; Sebeok and Umiker-Sebeok, 1976). These are emulated speech systems, which are obtained by transforming spoken language into drum sounds (Seifart et al., 2018). Most languages of the Niger-Congo family, present in Africa, are tonal, meaning that differences in relative pitch trigger differences in lexical meaning and syntactic functions. Drummed speech has been described almost exclusively for tonal languages. For example, the Yoruba people of Nigeria play drums to mimic the spoken Yoruba language (Euba 1990; Villepastour, 2010) and in Ghana drumming was widely used among the Akan people to imitate the spoken Akan language (Nketia, 1963).

Our research focuses on Senegalese traditions of encoding linguistic messages on drums. Senegalese drummers follow the practice of playing drums in correlation to speech. These drummers belong to the social class of griots (Hale, 1998, Tang, 2007), and their most common drum is a single-headed drum known as *sabar*. In Senegal, sabar drumming appears in different sorts of events such as sport events, life-cycle ceremonies, political gatherings. Although nowadays the sabar drums are rarely used as a speech surrogate and their main function is to entertain the listener rather than to convey a linguistic message, the practice of playing the sabar still maintains a close connection to linguistic expressions (Winter, 2014). The formal correspondence of sabar rhythms with spoken language is different from other documented African drum languages. Unlike most other languages of

TABLE 1 | Sabar phonemes.

Code	Variants	Description
Hand strokes		
Gin	Gi, bin	Bass sound; the hand strikes the edge or the middle of the skin
Pin		The hand strikes the edge of the drum
Pax	Pa, ba, bax, *mbar, mbax*	The full palm strikes the whole skin
Stick strokes		
tan	Ta, sa, san, dan, ja	The stick strikes the centre of the drum and bounces off
tac	Tas, *tach*	The stick strikes the centre of the drum and bounces off, while the hand damps the edge of the drum
ce	Ña, ca, cek, cex, te	The stick strikes the drum and is left there
Hand + stick strokes		
rwan	Rwa	pin tan or pax tan
rwe	Rwex	Pax ce
drin		Tan gin

the Niger-Congo family, Wolof is not a tonal language and sabar rhythms do not mimic the pitch of word sounds.

We focus on linguistically meaningful sabar rhythms. This class of rhythms and their linguistic correlates is referred to here as *Sabar*. To examine possible regularities between Wolof and sabar rhythms, we carry out a case study on our collected dataset. Two general hypotheses are examined. According to the *word-level hypothesis*, each stroke in Sabar represents a class of words in Wolof that share some specific sound properties. According to the *syllable-level hypothesis*, each stroke in Sabar represents a class of syllables in Wolof that share some specific sound properties. Each of the hypotheses will be described in further detail below.

MATERIALS AND METHODS

Basic Phonemic Units

Playing sabar involves at least nine different drum strokes, which can be seen as the basic units of the genre. These strokes appear in longer sabar rhythmic phrases which can be correlated with spoken utterances in Wolof. Sabar sounds are produced by one hand and one stick. Both the stick and the hand are used for beating rhythms and for damping the sound (Winter, 2014). Sabar basic units, which we call Sabar phonemes, are produced by different combinations of applying the hand and the stick onto certain parts of the sabar head. Each sabar phoneme has a special oral correlate.

As documented in (Winter, 2014) Sabar phonemes can be divided into the following three classes:

- Hand strokes: produced by one hand, which may bounce or stop on the skin.
- Stick strokes: produced by the stick, where the hand may be used for damping the sound.
- Hand + stick strokes: sequences of hand strokes and stick strokes; these sequences are perceived as minimal rhythmic units.

Each of the Sabar phonemes has an oral code that griots use when referring to rhythms. The main Sabar phonemes and their oral codes are given in **Table 1**. The table is taken from (Winter, 2014) and with a few more variants for the oral codes (in *italic*).

The code for any given phoneme refers to the way the sound is perceived, not to the precise technique of its production, which may vary between different types of sabar drums (Tang, 2007). The codes can be called slightly differently, depending on personal choice of a drummer, speed of playing or the drum that is being used.

Materials

This work relies on research materials collected during previous expeditions to Senegal. Materials include improvised material as well as traditional *bakks*. These are traditional texts or phrases in Sabar, known to many griots and learnt by heart. Recordings start with the phrase in Wolof, which is followed by the corresponding rhythm. For each recording there is a transcription of both the text and the rhythm. To transcribe the rhythm, the sabar stroke coding system is used, without any detailed annotation of temporal relations between strokes or their acoustic properties.

402 recordings were made. Of them, six recordings were excluded due to the fact that the transcription of the sabar rhythm was missing. Of the remaining 396 pieces, 35 are bàkks and 361 are improvised texts. The average number of words and syllables per piece in the spoken language is 11 and 15, respectively. The average number of strokes per piece is 15. To gain more insight into sabar practices, fifty of the Wolof pieces were translated into English by a Wolof speaker.

The recordings were made in live sessions with the drummers. The work was conducted in the years 2018–2019 in Campement Nguekhohk, Senegal[1]. All recordings were made with griots of the same family, where 2–3 drummers were present in each session. The drummers were asked to come up with a traditional bakk or improvisation in Wolof, play the corresponding rhythm and utter the rhythm's oral codes.

[1]The data were collected by C.L.A. Bourdeau in the years 2018–2019 as part of the NWO-project no. 360-89-060

TABLE 2 | Frequencies of strokes.

Stroke	Frequency	Percentage (% of 5,454)
Gin	1966	36
Tan	1814	33.3
Rwan	689	12.6
Pax	428	7.8
Tac	317	5.8
Ce	182	3.3
Drin	32	0.6
Rwe	26	0.5

TABLE 3 | The most common word-stroke combinations.

Wolof word	Sabar stroke(s)	English translation	Count
La	Tan	You	66
Ci	Gin	At/in	63
Bu	Gin	Which	51
Ko	Tan	Her/him/it	47
Li	Gin	The	47

Example 1) illustrates an improvised text. The text in Wolof is followed by the codes of the corresponding drum strokes:

(1) xol bu baax
rwan gin tan

moo gën xol bu bon fuuf
rwan tan tan tan gin gin

(A beautiful heart
Is way better than an ugly heart)

Methods of Analysis

In order to examine possible regularities between Wolof units and sabar strokes in the collected dataset, two hypotheses were studied. According to the word-level hypothesis, each word in Wolof has a specific stroke or stroke sequence associated to it. According to the syllable-level hypothesis, each syllable in Wolof has a specific stroke or stroke sequence associated to it, where the nature of the correspondence depends on the phonetic or phonological properties of the vowel in a syllable.

For the word-level analysis all the texts were divided into pairs, the first element of each being the Wolof word and the second element being the corresponding stroke or combination of strokes. This resulted in 4,290 pairs.

To test the syllable-level hypothesis, Wolof words were syllabified (Ka, 1988). The texts were divided into pairs: a Wolof syllable and the corresponding stroke. In most of the cases the number of strokes and syllables per line was the same. Out of 5705 I excluded pairs where there was no correlation between the number of strokes and syllables (for example, a one-syllable word in Wolof that has two strokes correlated to it; or a longer word in Wolof that has one correlating stroke). In total there were 251 such cases, less than 5% of pairs in the dataset. This resulted in 5,454 pairs of syllables and strokes.

TABLE 4 | Example of restructured data.

Sabar	Wolof	Wolof_pos	Wolof_len	Wolof_open
Gin	a	Central	Short	Open
Drin	bes	Front	Short	Mid-open
Pax	Daan	Central	Long	Open
Tan	Dem	Front	Short	Mid-open

In **Table 2** the frequencies of eight different strokes in the data are presented. The stroke *pin* was not present throughout the whole data.

Word-Level Hypothesis

According to the word-level hypothesis each Wolof word is associated to a specific sabar stroke or combination of strokes. In **Table 3** the five most common word-stroke combinations in the dataset are presented. The fourth column shows the number of occurrences of each combination.

As seen in the table, the most common combinations involve one-syllable words. Therefore, for now we focused on the syllable-hypothesis, since testing the word-level hypothesis is not straightforward using the data available to us at the moment.

Syllable-Level Hypothesis

According to the syllable-level hypothesis, syllables in Wolof have specific drum strokes associated to them, where the nature of the correspondence depends on the phonetic or phonological properties of the vowel in a syllable: length, openness and front/central/back property of a vowel.

As documented in (Ward, 1939: 322, Nikiforova 1981: 15, Unseth 2009: 1, Ka 1993: 7), vowel-length is a significant element of Wolof (for instance, *xol* 'heart' vs. *xool* 'to look at'). Therefore, all the syllables were divided into two groups: syllables with short and long vowels. From the total of 5,454 analysed Wolof syllables, 4,375 (80%) contained short vowels and 1,079 (20%)—long vowels. The syllables were further divided into three groups based on the front/central/back property of the vowel. Wolof front vowels are [i], [e]; central — [a] [ə] (spelled "ë" here); back — [o] [u] [ɔ] (Ward, 1939: 321, Nikiforova, 1981: 15, Ka 1993: 4). From the total of 5,454 analysed Wolof syllables, 2,445 (45%) contained central vowels, 1,660 (30%) — front vowels, 1,349 (25%) — back vowels. Further, syllables were divided into three groups based on their openness: open, middle (mid-open/mid-closed), closed. Wolof open vowels are [a], middle — [e] [ə] [o], closed — [i] [u] (Ward, 1939: 321, Nikiforova, 1981: 15). From the total of 5,454 analysed Wolof syllables, 2,160 (40%) contained open vowels, 1835 (34%) — closed vowels, 1,459 (27%) — middle (mid-open/mid-closed) vowels.

In the paper the orthographic symbols suggested by the drummers are used. The orthographic symbols used in the paper are presented in the **Figure 1**.

In order to test the syllable-level hypothesis, the data was presented in a table where in each row there was a specific stroke, an associated syllable and the properties of a vowel in this stroke: length (short/long), position (front/central/back) and openness (open/middle/closed). **Table 4** is an example, where Wolof_pos

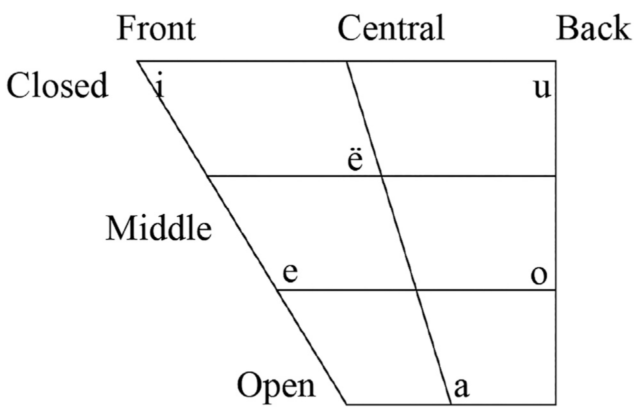


FIGURE 1 | Chart with the used symbols.

TABLE 5 | Cross-tabulation of stroke types and vowel lengths.

Stroke	Vowel length	
	Short	Long
Ce	148 (0.4)	34 (–0.4)
Drin	20 (–2.5)	12 (2.5)
Gin	**1731** **(10.9)**	**235** **(–10.9)**
Pax	355 (1.5)	73 (–1.5)
Rwan	**451** **(–10.4)**	**238** **(10.4)**
Rwe	21 (0.1)	5 (–0.1)
Tac	**232** **(–3.2)**	**85** **(3.2)**
Tan	1,417 (–2.8)	397 (2.8)

Absolute adjusted standardised residuals greater than three are highlighted in bold.

stands for the front/central/back property of the vowel in the Wolof syllable associated with the given stroke, Wolof_len—for the length and Wolof_open—for the openness of the syllable. This data was analysed using SPSS software.

RESULTS

Length

A chi-square test of independence was conducted between the stroke type and the length of the vowel in the corresponding syllable. All expected cell frequencies were greater than five. There was a statistically significant association between the stroke type and the length of the vowel in the corresponding syllable, $\chi2(7) = 193.94$, $p < 0.0005$. The association was weak, Cramer's $V = 0.189$ (Cohen 1988).

A cross-tabulation of stroke types and vowel lengths is presented in **Table 5**. Adjusted residuals appear in parentheses below the observed frequencies. A residual is the difference between the expected frequency and the observed frequency. The residuals are standardized so that they have an approximately standard normal distribution with the approximation improving at larger sample sizes. The adjusted standardized residual higher than 3 mark the cells that deviate significantly from independence (Agresti, 2007). Absolute adjusted standardised residuals greater than three are highlighted in bold.

TABLE 6 | Cross-tabulation of stroke types and vowel positions.

Stroke	Vowel position		
	Front	Central	Back
Ce	**81** **(4.2)**	81 (–0.1)	**20** **(–4.4)**
Drin	8 (–0.7)	17 (0.9)	7 (–0.4)
Gin	**837** **(14.6)**	**362** **(–29.5)**	**767** **(18.3)**
Pax	**61** **(–7.6)**	**300** **(10.9)**	**67** **(–4.5)**
Rwan	**113** **(–8.6)**	**397** **(7.2)**	179 (0.8)
Rwe	**21** **(5.6)**	4 (–3.0)	1 (–2.5)
Tac	**197** **(12.6)**	**95** **(–5.5)**	**25** **(–7.2)**
Tan	**342** **(–13.1)**	**1,189** **(21.7)**	**283** **(–11.0)**

Absolute adjusted standardised residuals greater than three are highlighted in bold.

Position

A chi-square test of independence was conducted between the stroke type and the front/central/back property of the vowel in the corresponding syllable. All expected cell frequencies were greater than five. There was a statistically significant association between the stroke type and the position of the vowel in the corresponding syllable, $\chi2(14) = 1,274.9$, $p < 0.0005$. The association was moderately strong, Cramer's $V = 0.342$.

A cross-tabulation of stroke types and vowel positions is presented in **Table 6**. Adjusted residuals appear in parentheses below the observed frequencies. Absolute adjusted standardised residuals greater than three are highlighted in bold.

Openness

A chi-square test of independence was conducted the between stroke type and the openness of the vowel in the corresponding syllable. All expected cell frequencies were greater than five. There was a statistically significant association between the stroke type and the openness of the vowel in the corresponding syllable, $\chi2(14) = 2,476$, $p < 0.0005$. The association was very strong, Cramer's $V = 0.476$.

A cross-tabulation of stroke types and vowel openness positions is presented in **Table 7**. Adjusted residuals appear in parentheses below the observed frequencies. Absolute adjusted standardised residuals greater than three are highlighted in bold.

The statistical analysis therefore shows that some regularities between sabar strokes and Wolof syllables exist, namely the following (only the results with absolute adjusted standardized residuals greater than three are presented):

TABLE 7 | Cross-tabulation of stroke types and vowel openness.

Stroke	Vowel position		
	Open	Mid-open/mid-closed	Close
Ce	71 (–0.2)	**81 (5.5)**	**30 (–5.0)**
Drin	9 (–1.3)	15 (2.6)	8 (–1.0)
Gin	**244 (–30.8)**	**295 (–14.7)**	**1,427 (45.7)**
Pax	**285 (11.9)**	**84 (–3.5)**	**59 (–9.1)**
Rwan	**371 (8.2)**	**257 (6.7)**	**61 (–3.2)**
Rwe	3 (–2.9)	**22 (6.7)**	**1 (–3.2)**
tac	74 (–6.1)	**206 (15.8)**	**37 (–8.5)**
Tan	**1,103 (22.6)**	499 (0.9)	**212 (–24.2)**

Absolute adjusted standardised residuals greater than three are highlighted in bold.

Strokes' preferences for vowel length (weak association):

gin—short
rwan—long

Strokes' preferences for vowel position (moderate association):

ce—front
gin—back (and front)
pax—central r
wan—central
tac—front
tan—central

Strokes' preferences for vowel openness (strong association):

ce—mid-open/mid-closed
gin—closed
pax—open
rwan—open
rwe—mid-open/mid-closed
tac—mid-open/mid-closed
tan—open

The association was weak for the strokes' preferences for vowel length, moderately strong for vowel position and very strong for vowel openness, therefore we suggest to take into account only the results for the vowel position and openness for now. These results suggest there is a Wolof-Sabar correspondence that depends on the phonetic and phonological properties of a vowel in a syllable.

DISCUSSION

This paper reports the first study that is meant to uncover statistical regularities between between units in the spoken language and strokes in the drum language. We used a dataset of sabar rhythms and their corresponding Wolof phrases. Two different hypotheses—the word-level hypothesis and the syllable-level hypothesis—were examined. While the data did not allow a detailed study of the word-level hypothesis, descriptive and inferential statistics were used in order to test the syllable-level hypothesis. Evidence for this hypothesis was found: the vowel position and the vowel openness affect the preference of an associated stroke with moderate and strong strength of association respectively. Such parameters as vowels openness and the front/back distinctions are the most salient and basic parameters for representing vowel systems and for this reason they are reflected in the Sabar phonology.[2]

It has to be metioned that Wolof also has ATR vowel harmony, meaning that Wolof vowels harmonise based upon the phonological feature [ATR], or advanced tongue root, a widespread phonological pattern in African languages (Casali, 2008; Ka 1993; Unseth, 2009; Van der Hulst, 2018). In the current analysis we do not include this feature, however, it might also be reflected in the drum phonology.

Some of the limitations of this work should be pointed out in order to outline the room they leave for further research. First, the work is based on a limited number of pieces collected from one family of the drummers in Senegal. This has probably led to the fact that it did not allow for a detailed study of the word-level hypothesis. Second, the inherent difficulty of working with Sabar should also be mentioned. Unlike other drum languages, which are based on tonal languages and therefore imitate the pitch levels of the spoken language, Wolof and the Sabar are not tonal, and working with them reguires different methods. This paper documents one attempt to develop such a method. Our study of the syllable-level hypothesis, while showing certain correlations, could not fully predict the relation between Wolof and Sabar. This suggests that there is much room for exploring further hypotheses about the relations between speech and rhythm in Sabar.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The present paper was ethically approved by the Ethics Assessment Committee of the Faculty of Humanities, FEtC reference number: 20-328-02.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

[2]With thanks to an anonymous reviewer for making this observation.

FUNDING

This research was supported by NWO grant no. 360-89-060, project name "When language has a beat: Senegalese drum language and linguistic theory".

ACKNOWLEDGMENTS

The author would like to thank Yoad Winter, Myrthe Hemker, Khady Postma, Corentin Bourdeau, Sophie Sabar, Art Mbaye, Ouzin Diop, Gana and Ndongo Diop and other members of Diop family in Senegal.

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. 2nd ed. Hoboken, NJ: Wiley.

Casali, R. F. (2008). ATR Harmony in African Languages. *Lang. Linguist. Compass* 2 (3), 496–549.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New York: Psychology Press.

Euba, A. (1990). *Yorùbá Drumming: The Dundun Tradition*. Bayreuth: Bayreuth African Studies.

Hale, T. A. (1998). *Griots and Griottes*. Bloomington: Indiana University Press.

Ka, O. (1993). *Wolof Phonology and Morphology*. Lanham, MD: University Press of America.

Ka, O. (1988). "Wolof Syllable Structure: Evidence from a Secret Code," in Procedings of the Eastern States Conference on Linguistics 5th; Philadelphia, PA, September 30-October 1988. (Washington, DC: ERIC Clearinghouse).

Nikiforova, L. A. (1981). *Yazyk Wolof [Wolof Language]*. Moscow: Nauka.

Nketia, J. H. K. (1963). *Drumming in Akan Communities of Ghana*. Edinburgh: Published on behalf of the University of Ghana by T Nelson.

T. A. Sebeok and D. J. Umiker-Sebeok (Editor) (1976). *Speech Surrogates: Drum and Whistle Systems*. The Hague, Netherlands: Mouton.

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing Language to Rhythm: Amazonian Bora Drummed Language Exploits Speech Rhythm for Long-Distance Communication. 5.4, 170354. doi:10.1098/rsos.170354

Stern, T. (1957). Drum and Whistle "Languages": An Analysis of Speech Surrogates. *Am. Anthropologist* 59, 487–506. doi:10.1525/aa.1957.59.3.02a00070

Tang, P. (2007). *Masters of the Sabar: Wolof Griot Percussionists of Senegal*. Philadelphia: Temple University Press.

Unseth, C. (2009). *Vowel Harmony in Wolof*. Australia, Occasional Papers in Applied Linguistics No. 7.

Van der Hulst, H. (2018). *Asymmetries in Vowel Harmony: A Representational Account*. Oxford, UK. Oxford University Press.

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code*. Surrey, UK. Ashgate.

Ward, I. C. (1939). A Short Phonetic Study of Wolof (Jolof): As Spoken in the Gamiba and in Senegal. *Africa* (3), 320–334. doi:10.2307/1155382

Winter, Y. (2014). On the Grammar of a Senegalese Drum Language. *Language* 90 (3), doi:10.1353/lan.2014.0061

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fcomm.2021.652690

When Music Speaks: An Acoustic Study of the Speech Surrogacy of the Nigerian Dùndún Talking Drum

Cecilia Durojaye[1,2]*, Kristina L. Knowles[3], K. Jakob Patten[4], Mordecai J. Garcia[3] and Michael K. McBeath[1,2]

[1]Psychology Department, Arizona State University, Tempe, AZ, United States, [2]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany, [3]School of Music, Dance and Theatre, Arizona State University, Tempe, AZ, United States, [4]College of Health Solutions, Arizona State University, Tempe, AZ, United States

Yorùbá dùndún drumming is an oral tradition which allows for manipulation of gliding pitch contours in ways that correspond to the differentiation of the Yorùbá linguistic tone levels. This feature enables the drum to be employed as both a musical instrument and a speech surrogate. In this study, we examined four modes of the dùndún talking drum, compared them to vocal singing and talking in the Yorùbá language, and analyzed the extent of microstructural overlap between these categories, making this study one of the first to examine the vocal surrogacy of the drum in song. We compared the fundamental frequency, timing pattern, and intensity contour of syllables from the same sample phrase recorded in the various communicative forms and we correlated each vocalization style with each of the corresponding drumming modes. We analyzed 30 spoken and sung verbal utterances and their corresponding drum and song excerpts collected from three native Yorùbá speakers and three professional dùndún drummers in Nigeria. The findings confirm that the dùndún can very accurately mimic microstructural acoustic temporal, fundamental frequency, and intensity characteristics of Yorùbá vocalization when doing so directly, and that this acoustic match systematically decreases for the drumming modes in which more musical context is specified. Our findings acoustically verify the distinction between four drumming mode categories and confirm their acoustical match to corresponding verbal modes. Understanding how musical and speech aspects interconnect in the dùndún talking drum clarifies acoustical properties that overlap between vocal utterances (speech and song) and corresponding imitations on the drum and verifies the potential functionality of speech surrogacy communications systems.

Keywords: talking drum, speech surrogate, music, acoustics, dundun, Nigeria, pitch, rhythm

OPEN ACCESS

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Maarten Mous,
Leiden University, Netherlands
Bruce Connell,
York University, Canada
Meagan Curtis,
SUNY Purchase, United States

Correspondence:
Cecilia Durojaye
cdurojay@asu.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *12 January 2021*
Accepted: *18 June 2021*
Published: *27 July 2021*

Citation:
Durojaye C, Knowles KL, Patten KJ, Garcia MJ and McBeath MK (2021) When Music Speaks: An Acoustic Study of the Speech Surrogacy of the Nigerian Dùndún Talking Drum. Front. Commun. 6:652690. doi: 10.3389/fcomm.2021.652690

INTRODUCTION

Yorùbá *dùndún* drumming is a musical-oral tradition wherein the characteristic of the drum as a variable-pitched membranophone allows the manipulation of intensity and pitch.[1] in ways that can mimic the tones and gliding contours of the Yorùbá, a tonal language, spoken in south-west Nigeria. This unique feature enables the dùndún drum (commonly referred to as the "talking drum") to be

[1]Following Villepastour (2014), we will use pitch to reference the fundamental frequency produced by the drum and the fundamental of a speech utterance, and "speech tone" to refer to the relative pitch of a spoken syllabus in Yorùbá.

FIGURE 1 | Dùndún percussionists. Drummers can play individually, but more commonly in ensembles.

employed as both a musical instrument and a speech surrogate. The dual function of the drum, its role in the Yorùbá social-cultural milieu, and the belief that it is the most eloquent of Nigerian talking drums (Euba 1990) have thus drawn the attention of linguists, (ethno) musicologists, and anthropologists with a focus on various aspects such as the structure of the drum ensemble (Akpabot 1975; Vidal 2012; Durojaye 2020), the social and religious functions of the drum (Adegbite 1988), principles of dùndún communication (Arewa and Adekola 1980), and more. The dùndún is played both individually and, more commonly, in ensembles, as illustrated in **Figure 1**.

Prominent characteristics of the dùndún include its speech and musical features. The aesthetic and stylistic attributes of drum poetry inescapably share literary and musical space when imitating Yorùbá oral literature (Sotunsa 2009). *Bàtá* drums, a very close relative of the dùndún, use drum strokes as a code that translate into Yorùbá language (Villepastour 2010). Dùndún drummers, however, draw elements from music and speech to communicate emotions on the drum (Durojaye 2019a). The historical and social background around the dùndún tradition, organization and training of drummers, and the three main modes of the dùndún drum are addressed in Akin Euba's monograph *Yorùbá Drumming* (1990), the most extensive work on the dùndún to date.

Of particular importance to the present study is the observation that three drum tones are consistently used to imitate the relative tones of the Yorùbá language. However, when drum speech and human speech are compared, linguistic representations on the drum tend to differ from the original speech utterances (Euba 1990). Acoustical analyses confirm this result, demonstrating that the three distinct tones (Low, Mid, and High) are produced on a global level with three measurably different fundamental frequencies (Connell and Ladd 1990; Akinbo 2019). In a word-level analysis, however, drummers often produce the three tones slightly differently in a way that mimics the spoken language. For instance, a high tone produced before a low tone rises and is significantly higher than one preceding a mid tone (Connell and Ladd 1990; Laniran and Clements 2003). Moreover, Akinbo (2019) noted that the number of drum strikes tend to be correlated with the number of syllables present in either mono- or disyllabic words, However, this research was conducted with single words rather than full speech phrases. Akinbo's research also focused solely on the speech mode of drumming and not the drum as a surrogate for song.

In contrast to the previous work, the precise specific acoustic characteristics and relationships of dùndún drumming modes were examined in two sister articles. Durojaye et al. (2021) provides a *macrostructural* analysis of the overall distributional matches between two dùndún modes for acoustical metrics of intensity, fundamental frequency (f_0), timing, and an entropy measure of timbre. This timbre metric analyzes the slope of the spectral envelope to gauge changes in speech features (Toh et al., 2005). In the present study, we examine the *microstructural* correlations between individual Yorùbá vocalizations at the syllable level and expand the acoustic continuum to four modes of dùndún drumming. Here, we analyze the microtiming patterns, fundamental frequency (f_0) and intensity contours of individual drummers playing the same auditory sequence as one that is vocalized and compare the correlations between the different modes of production. The ongoing acoustic features of f_0 and intensity are fundamentally important to the characterization of music and language (McBeath and Neuhoff 2002; Patten and McBeath 2020; Patten et al., 2019; Yu et al., 2021) and aid the intelligibility of the dùndún communication (Euba 1990; Akinbo 2019). As there are various drums in the general dùndún category, we refer to the *ìyáàlù* (mother of the drum) as the dùndún in this study (see Durojaye 2019b, for detailed description from an emic perspective). Gaining a full knowledge of the communicative mechanisms of the dùndún and its role in the language-music relation benefits from contributions with various perspectives, including acoustical analyses and comparisons of different modes of Yorùbá vocalization and dùndún drumming.

The Yorùbá language, which the drum imitates, uses three relative speech tones: Low, Mid, and High.[2] While the H and L tones are realized as rising and falling intonations, in comparison,

[2]The Yorùbá standard orthography is employed for transcriptions: H tone is marked with an acute accent (´), L tone with the grave (`) and the M tone is usually unmarked. Diacritics such as ẹ/ɛ/, ọ/ɔ/, ṣ/ʃ/, p/k͡p/, and gb/g͡b/are also used for Yorùbá words.

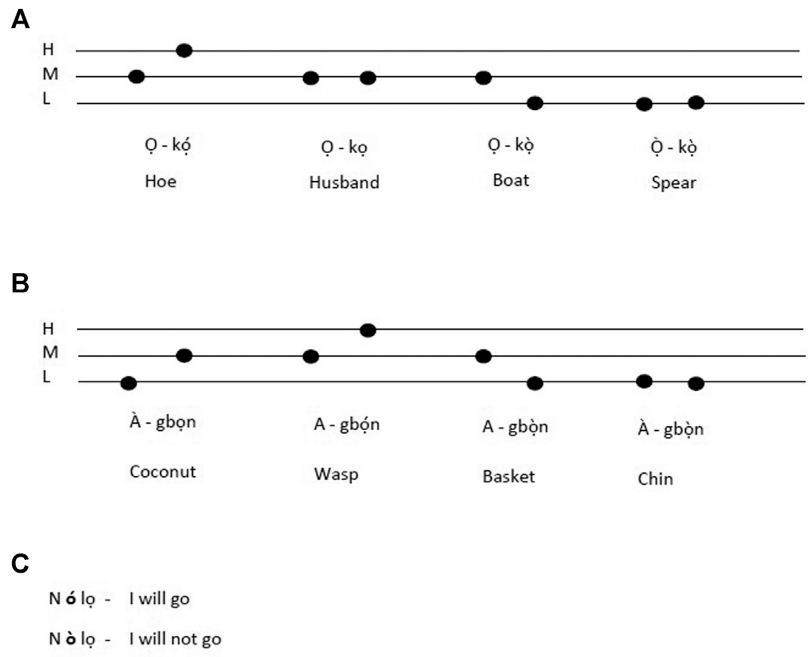


FIGURE 2 | Three examples exemplifying various standard f_0 glide patterns of the Yorùbá language.

the mid tone (M) is flat (its level remains the same and often represents the default tone of a speaker). In other words, the mid tone uses a relaxed laryngeal position, while the H and L tones necessitate muscular tension in the larynx to create a rising (H) or falling (L) tone. The relative speech tones are essential for word and sentence signification and even indicate opposing meaning, i.e., they are lexically and grammatically contrastive, as illustrated in the following examples shown in **Figure 2**.

Unlike monosyllables that have three tonal possibilities, disyllabic words have nine possible tonal patterns (HH, HM, HL, LH, LM, LL, MH, ML, and MM). Although one can employ all possible speech tone combinations in the language, this is, however, dependent on the particular utterance as a word meaning can be distorted if the wrong configurations are employed. (Arewa and Adekola 1980). For example, while different words can be produced from the word *agbọn,* only four tonal combinations are meaningful in the Yorùbá language.

Yorùbá language consists of twelve vowels (seven oral vowels and five nasal vowels), eighteen consonants, and a syllabic nasal/ n/.[3] Every syllable in Yorùbá language contains a vowel. The vowels constitute the Tone Bearing Unit (TBU) of the language and are also the essence of the tonal glides occurring in the language (Eme and Uba 2016). Glides, comprised of both rising and falling tones, occur when a syllable consists of more than one tone or when there are two adjacent vowels, as in *ọlọ́pàá* (the police) or *jọ̀ọ́* (please). Words ending in a vowel are often elided or assimilated when followed by words beginning with a vowel (Pulleyblank 2009). Since the majority of Yorùbá verbs end in a vowel and most nouns begin with a vowel, elision (the deletion of a vowel) is a common occurrence. For example, *gbé od*ó (lift a mortar) becomes *gbódó*. Elisions also occur when a preposition is followed by a noun as in *ní ilé* (at home) to *níle*. In vowel assimilation, the assimilated vowel becomes the same as the vowel to which it was added, as in *Mo rí i* (I saw him/her), or *mo pè é* (I called him/her) (Barber and Oyetade 1998). As a result, these glides, elisions, or assimilation of vowels have an effect on the actual speech tone and the syllable duration (Villepastour 2014). Also, unlike the nasalized vowels which do not impact the musical setting, the syllabic nasal, like vowels, adds to the tone length (Villepastour 2014, 31).

Speech perception relies on multiple acoustic features produced by the vocal apparatus. Fundamental frequency has been identified as the major component of tone perception in Yorùbá (Hombert 1976). Intensity of the sound pressure wave of speech (perceived as loudness), which is correlated with syllable accent and emotionality in some non-tonal languages (Sluijter and van Heuven et al., 1996; Laukka et al., 2011), has been identified as a lesser component of tone in tonal languages including Yorùbá and Cantonese (Ọdẹ́jọbí, 2007). Lastly, f_0 and intensity are not perceived independently of one another. Perception of pitch is often correlated with intensity (Neuhoff and McBeath 1996; McBeath and Neuhoff 2002). The reverse is also true in speech production; as speakers raise the intensity of their voice, the f_0 also rises (Scharine and McBeath 2019). Intercorrelations between f_0, intensity, and timbre have also been observed (Patten and McBeath 2020). As mentioned, Yorùbá speech tones are specifically dependent on modulation of the f_0, which is perceived as pitch. As dùndún drummers are attempting to replicate speech, a multiply determined construct,

[3]Yoruba vowels include seven oral vowels a/a/, e/e/, ẹ/ɛ/, I/i/, o/o/, ọ/ɔ/, u/u/; five nasal vowels an/ã/, ẹn/ɛ̃/, in/ĩ/, ọn/ɔ̃/, and un/ũ/.

TABLE 1 | Characteristics of various techniques, execution, and sound types produced on the dùndún drum.

Dùndún sounds	Stick/syllabification technique	Execution
L and M tones	Free stroke (loose wrist movement; stick bounds off the drum head after playing the tone. One drum stroke per syllable.	Moderate pressure on strings (M) Minimal or no tension (L)
H tone	Muted stroke (firm wrist movement; stick does not bounce off the drum head). Muted stroke is also usually used in the speech mode. One drum stroke per syllable.	Maximum pressure on tension strings
Glides (from L to H or vice versa)	One drum stroke for two tone levels/syllables	L–H: Gradual tightening of tension strings immediately after executing the lower tone and maintains the pressure for the glided tone H–L: Gradual release of pressure on tension strings
Vowel assimilation	One drum stroke for two tone levels/syllables	Tightening or releasing of tensioning strings according to the tone contour
Singing	As used for speech as highlighted above. Sometimes the use of hand and stick technique.	Vibration of the arm in addition to loosening and tightening of the tension strings

with an instrument that can most readily vary in f_0 and intensity, it is unclear which elements of the speech signal drummers rely on to create an effective speech surrogate.

Dùndún drummers reproduce the Yorùbá speech tones by loosening or tightening the tension strings (*ọsán*) surrounding the drum's resonator and connecting the two drumheads from one end of the drum to the other. Although there are different dialects spoken among the Yorùbá, the dùndún only imitates the Ọ̀yọ́ dialect (Akpabot 1986; Durojaye 2019b), believed to be the standard Yorùbá (Euba 1990; Villepastour 2014). The *iyáàlù dùndún* is carried with a shoulder strap which hangs from the player's (usually left) shoulder, wherein the combination of the left fingers, the wrist and hip bone are employed in the manipulation of the strings (*ọsán*). The other hand plays the drum using a stick known as *ọ̀pá* or *kọ̀ngọ́* (a curved stick shaped like lower case 't' without the cross). For the drum to produce the lowest pitch, minimal pressure is applied to the strings, and the more the pressure, the higher the frequency. Thus, the H speech tone is executed with the maximum pressure on the strings (**Table 1**). Given that the drum only captures vowels (the Yorùbá tones) and vowels occur in every syllable of the Yorùbá language, interpretation and analyses of drum rhythm and tone are usually in relation to syllables. The technique of representing syllables can take many forms such as 1) using one drum stroke for each syllable (as for a single tone level and vowel elisions); 2) many strokes for one syllable; 3) one drum stroke for two or more syllables and 4) one drum stroke for a syllable with many speech tone levels as would be the case for some glides, or assimilations (see also, Euba 1990).

Past research (for example, Euba 1990) has shown that the dùndún connects music and language through three modes. First, the "musical speech" mode in which strict or danceable rhythm is employed, which we refer to as *Drum–Dance Rhythm* (*D-DR*). Second, the "song form" where the drum is used to mimic vocal singing of a text, which we refer to as *Drum Singing* (*DS*). Third, "speech mode" in which the drum more strictly imitates speech and follows speech rhythm, which we divided into two sub-categories, referred to as *Drum Talking, Performative* (*DT-P*), and *Drum Talking, Direct* (*DT-D*). We added the subcategory of DT-D in order to test the drumming acoustical match with speech when drummers directly try to maximize their imitation of talking without musical constraints. The categories we compared are shown in **Table 2**. To our knowledge, no prior empirical studies have acoustically compared all of these drum modes to their corresponding vocally spoken or sung forms. The principal goal of this study is to determine the extent of microstructural acoustic representation of speech sounds on the drum, and second to distinguish and acoustically compare the four dùndún mode categories. Here we examine acoustic f_0, intensity, and microtiming patterns and test the microstructural relationship between the various modes of Yorùbá vocalization and dùndún drumming.

Our principal hypothesis, shown in **Table 2**, is that there are successively larger positive correlations between the patterns of (a) drum attack interonset intervals (i.e., duration from one attack to the next) and interonset intervals from one syllable to the next, and between the (b) drum and speech f_0, as the drum sequentially increases in extent of speech surrogacy from drum as song (DS) to direct talking (DT-D). Such a pattern validates the functional variance represented by the different drum modes. Our secondary hypothesis is that there is a progression toward musical rhythms across adjacent drumming modes (e.g., DT-D to DT-P to DS to DD-R) with the more musical modes featuring greater similarity in the average length of IOIs than more speech-like modes, which are more likely to feature non-isochronous rhythms. Together, the hypothesized pattern of findings would acoustically confirm that the dùndún drum spans the range of surrogacy communication systems between speech and music at a microstructural (syllable-by-syllable) level.

METHODS

Recording and Data Collection Procedure

Three Yorùbá vocal performers produced 10 examples each of both speech and singing, for a total of 60 vocalizations. Three independent professional drummers from three Yorùbá towns were randomly assigned 10 each of the previously recorded speech and singing examples and were asked to represent their assigned examples on the *dùndún.* Drummers performed 10 examples each of Drum-Dance Rhythm (D-DR), Drum Singing (DS), Drum Talking, Performative (DT-P), and Drum Talking, Direct (DT-D) for a total of 120 performances. Drum

TABLE 2 | Categories of speech and drumming that were compared.

Vocalizing	Drumming	
Vocal talking (*VT*)	Drum talking–Direct (*DT-D*)	↑Predicted larger vocal vs drum correlation
	Drum talking–Performative (*DT-P*)	
Vocal singing (*VS*)	Drum singing (*DS*)	
	Drum–Dance rhythm (*D-DR*)	

and vocal performances were matched such that there were 30 song/DS, 30 speech/DT-P, and 30 speech/DT-D pairs. D-DR performances did not have a matched vocal recording. All the performers were male and monolingual Yorùbá. Drummer A and C have approximately 35 years of experience of dùndún drumming, while drummer B has been playing for 28 years. The Yorùbá speakers also doubled as singers as they are well conversant with traditional Yorùbá songs.

Sample phrases were recorded at a professional studio in Ibadan, South-West Nigeria. After listening to each of the spoken utterances, the drummers were asked to replicate the same sequence in the corresponding drum language. The same procedure was followed for the songs and their drum representation. Each drummer, with their personal drums, reproduced ten samples each of the previously recorded spoken utterances and songs. For the song samples, performers were asked to choose from *dùndún* popular repertoire and were thus at liberty to perform similar pieces. For each of the speech samples, performers were asked to create two recordings, one of the drum as it is typically used to represent speech (DT-P) and one where the drummer tries to maximize their imitation of the speech example (DT-D). The recordings ranged between 5 and 10 s in length. The drum data were recorded in a soundproofed room with a SHURE SM57 dynamic microphone at a 3-inch distance from the drum. Spoken utterances and songs were recorded with Audio Technica AT 2035 cardioid condenser microphones. All recordings were made at a sampling rate of 44.1 KHz in WAV format and saved as separate files.[4] The recordings were then analyzed independently for timing, and f_0 and intensity information. In the initial analysis stage, recording errors resulting in missing data were discovered for two of the 180 recordings. As recordings were paired for vocal utterances and corresponding drum modes, we removed the data that was paired with the incomplete recordings, resulting in a total of five recordings omitted from analysis (one each of VT, DT-D, DT-P, VS, and DS), leaving a total of 175 recording samples in our final analysis.

Analysis of Timing in Speech and Dùndún Drumming

In order to allow for comparison of timing patterns between the drum excerpts and speech and song excerpts we measured interonset intervals (IOIs). An IOI is defined as the duration between successive event onsets, a standard measurement within studies on sensorimotor synchronization (Madison 2001; Repp 2005), expressive timing in music (Benadon 2006; Goldberg 2015; Ohriner 2018), and cognitive processes like free recall of items from memory (e.g., Rhodes and Turvey 2007; Patten et al., 2020). In keeping with standard methods for measuring IOIs in prerecorded music (Repp 1992; Ashley 2002), drum samples were uploaded into the sound processing program *Audacity* (Mazzoni 2021) and were then analyzed in two stages. In the first stage, the time point for each attack on the drum was marked in the recording. Since the physical onset of the sound envelope for a drum typically corresponds to the perceptual attack time, IOIs were calculated based on the physical onset of the sound which was verified in the sound file both visually and aurally. In the second stage analysis, onsets were coded for whether they were produced by a single strike on the head of the drum, or by utilizing rhythmic embellishments such as flams (short rapid note on the drum) or *àfikún* (additions), a pair of rapid 16th notes. Following musical conventions for both Western and Nigerian drumming, the second attack of a flam was marked as the onset for the drum stroke.

Excerpts containing verbal utterances (recordings of speech and song excerpts by performers) underwent a related process for analyzing timing information. Unlike drum attacks, the perception of which corresponds to the onset of the sound event, perceptual attack times in language are correlated with the onset of vowels, which are considered the syllable nucleus (Peterson and Lehiste 1960; Allen 1972; Greenberg et al., 2003). In order to calculate onsets of vowels in the speech (VT) and song (VS) samples, recordings were uploaded into Praat (Boersma and Weenink, 2021) and then parsed first for word boundaries, then for syllables, and finally vowels. Vowels were identified by the onset of characteristic vowel formants. Following onset identification, attention was given to the mid-point of the amplitude rise (approxmiately 50%) at the beginning of the vowel, which is broadly considered to be the "perceptual center," or p-center, of a vowel. (Pompino-Marschall 1991; Harsin 1997; Ohriner 2019). The p-center was used when calculating IOIs for speech and song excerpts.

Although Yorùbá is a language without diphthongs (Bamgbose 2000), assimilation (the approximation of two successive phonemes toward each other and away from their isolated pronunciation), and elision (the deletion of a phoneme) present similar challenges to phonetic segmentation. Some scholars have attributed an independent attack to each vowel in an assimilated or elided phrase; however, this refers to a descriptive, phonological end and not a perceptual one (Ola Orie and Pulleyblank 2002). The current paper relies on perceptual work regarding diphthongs to segment speech into

[4]Sample recordings provided in Results section.

syllabic intervals. While longer in duration, English diphthong vowels are not perceived as categorically different from their monophthong counterparts (Lehiste and Petersen 1961; Fox 1983). Diphthongs are also acoustically distinct from glides between two monophthong vowels in that diphthongs do not reach the formant range of their ending vowel portion but have a trajectory approaching it. This formant trajectory difference serves to differentiate a diphthong from a monophthong rather than cue a link between two vowels (Gay 1970). As such, diphthongs are typically treated as a single vowel in the speech timing analysis. In the current work, formant frequencies were used to identify vowel onsets and assimilations; clear vowel differences were marked as independent onsets, but formant signatures more indicative of a trajectory toward another vowel (a signature common to diphthongs and elisions) were treated as a single onset (Byrd 1992; Jun 2004). For example, the phrase "bi a" (if we) is two distinct words but is phonetically produced as [bjɑ]. As a result, when analyzing the timing of syllable onsets, elisions and assimilations were treated as a single onset. The duration of syllables, an analog of the drum IOI, was then calculated as the duration between successive vowels between syllables. We estimate that our method for identifying vowel onsets produces a degree of variability in the order of 5–10 ms, though this window of error may be marginally higher in the case of vowel onsets that are preceded by nasal or liquid consonants (e.g., [m] or [l]), as the vowel amplitude rise is not readily apparent. In either case, this variation is likely in the realm of an imperceptibly small change (Huggins 1972). Timing data for speech and song were then compared to their matched drum modes to determine correlations in timing profiles.

In order to test for the occurrence of a continuum of acoustic features spanning the four drumming modes, the difference between successive IOIs was computed and compared as a measure of rhythmicity over sample phrases. Our hypothesis here is that the average difference between IOIs increases linearly across modes between drumming as a purely musical instrument (featuring a greater incidence of isochronous rhythms and even rhythmic ratios) up to directly mimicking speech, which features largely non-isochronous rhythms and is most prone to longer breaks. Specifically, the average IOI differences of the four drumming modes are predicted to be ordered D-DR < DS < DT-P < DT-D. We also gathered data to perform a post-hoc descriptive analysis of the relationship between the IOIs of speech samples and those of their matched drumming modes, but we do not have an a priori hypothesis regarding any relationship differences between the modes.

Analysis of Fundamental Frequency (f_0) and Intensity

Fundamental frequency (f_0) and intensity were extracted from vocal talking (VT), vocal singing (VS), drum as song (DS), drum as performative speech (DT-P), and drum as direct speech (DT-D) samples with Praat (Boersma and Weenink 2021). The tempo of recorded performances differed both between vocal and drum phrases and between sub-phrases within a particular song. Acoustic information was extracted and compared on a sub-phrase basis, identified by long gaps of silence in both vocal and drum recordings. Occasionally, some quiet time-keeping rhythms that occurred between sub-phrases and had no f_0 information were omitted. These rhythms are discussed further in the microtiming analysis results. Tempo still differed in the resulting sub-phrases, so longer sample phrases were downsampled to match the shorter one. To demonstrate similarity between vocal and drum performances, f_0 and intensity correlations were calculated. However, pitch - the perception of the f_0 of speech - is not solely dependent on fundamental frequency; pitch is also influenced by intensity (e.g., Patten and McBeath 2020). Thus, correlations are not only computed between drum and vocal fundamental frequencies and drum and vocal intensities, but also across acoustic features. In other words, this analysis attempts to understand how dùndún speech surrogacy uses acoustic features other than just f_0 to convey pitch information. The hypothesis for this analysis is similar to that of the timing analysis, that correlations between all acoustic features will increase as there are fewer musical features (i.e., DS < DT-P < DT-D). Similarly, it is expected that f_0 and intensity will significantly correlate between drum and speech phrases and that cross correlations between features will also be significant.

RESULTS

Timing Analysis

The timing analysis consisted of three parts, each examining and comparing the interonset intervals (IOIs) of the various vocal and drumming modes. The initial, principal analysis tested temporal correlations of IOIs between corresponding vocal and drumming modes in order to confirm the various levels of instrumental speech surrogacy. The secondary analysis compared the average IOI rates for each of the drumming modes to confirm the distinctiveness of the drumming mode categories. The third analysis compared average IOI rates between corresponding vocal and drumming modes to explore overall systematic acoustic trends regarding vocal vs. drumming modes.

Table 3 summarizes the results of the principal temporal analysis comparing corresponding vocal vs. drumming modes across all samples. As shown in the third column, our principal hypotheses were supported, with significant correlations occurring between each of the corresponding vocal and drum modes, and systematically decreasing correlations when moving from drum mimicking direct talking (DT-D) through talking in performances (DT-P) and drum singing (DS). The fourth column is included to verify that the drumming IOIs of the different drumming modes, which are composed of different performances and should not show a relationship, do not correlate with each other. The only exception to this is the case of the DT-D and DT-P modes, which replicate the same speech sample and exhibit a small correlation of $r(28) = 0.24$.

The overall average correlation between the interonset intervals for Vocal Talking (VT) and Drum Talking, Direct (DT-D) was $r = 0.72$, a large correlation that was backed up by a significant one-sample t-test, $t(28) = 13.95$, $p < 0.001$ with a very large effect size of Cohen's $d = 2.59$. To put this in perspective, Cohen (1988) norms

TABLE 3 | Average timing correlations of syllabic occurrences between different modes of vocalizing and drumming on the same verbal-musical sequence. One-sample *t*-tests comparing correlation coefficients to a null set are included below average correlations.

Vocalizing	Drumming	Vocal and drum	Drum and drum
Vocal talking (*VT*)	Drum talking–Direct (*DT-D*)	$r_{VT,DTD} = 0.72^{**}$ $d = 2.59$	$r_{DTD,DTP} = 0.35^{**}$ $d = 1.00)$
	Drum talking–Performative (*DT-P*)	$r_{VT,DTP} = 0.37^{**}$ $d = 0.93$	—
Vocal singing (*VS*)	Drum singing (*DS*)	$r_{VS,DS} = 0.25^{**}$ $d = 0.75$	$r_{DS,DTD} = 0.07$ $r_{DS,DTP} = 0.10$
	Drum–Dance rhythm (*D-DR*)	—	$r_{DDR,DTD} = -0.05$ $r_{DDR,DTP} = 0.06$ $r_{DDR,DS} = 0.09$

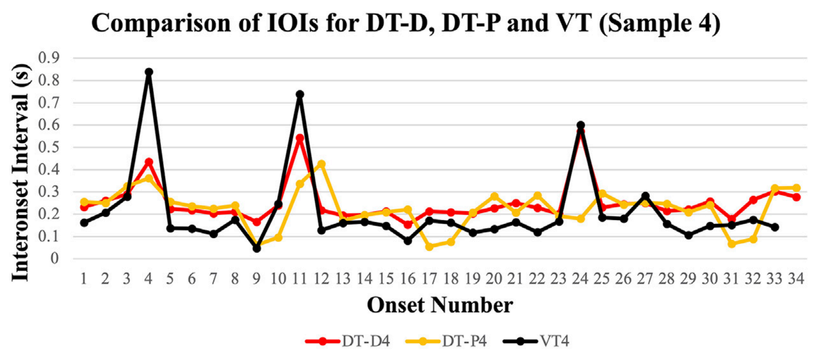


FIGURE 3 | Interonset intervals (IOIs) for a typical vocal talking (VT) sample and corresponding imitations of the speech sequence on the drum, both direct imitation (DT-D) and while musically performing the speech sequence (DT-P). Syllabic and drum onsets are presented as a monotonic temporal series on the *X*-axis, and the duration of each onset is indicated on the *Y*-axis. Thus, shorter IOIs correspond to lower values and longer durations correspond to higher values. Significant peaks represent gaps between sub-phrases. Note how the DT-D timing profile closely resembles VT, aligning robustly at the IOI peaks between sub-phrases, while the use of rhythmic embellishments in DT-P results in a somewhat less related timing profile than that of VT and DT-D (see sound samples **Supplementary Samples 1A–C**).

suggest a *d* over 0.8 is a large effect. This confirms that drummers are able to mimic the timing patterns of Yorùbá speech very robustly when tasked to do so directly without the need to add musical context. When playing in performance mode with some musical context, the average correlation between VT and DrumTalking, Performative (DT-P) drops to $r = 0.37$ while maintaining a significant one-sample *t*-test, $t(28) = 5.39$, $p < 0.001$, Cohen's $d = 0.93$. One possible reason for this lower correlation is the increased use of rhythmic embellishments in DT-P. However, removing flams and the first attack of every paired 16th note that makes up an àfikún does not have a significant effect on the overall correlation ($r_{\text{original}} = 0.37$, $r_{\text{corrected}} = 0.43$, an increase of only 4% variance explained). **Figure 3** illustrates the match between VT, DT-D, and DT-P for a typical representative sequence. Peaks in the graphs of duration onsets (IOIs) indicate breaks between sub-phrases. Note the relative alignment of IOI peaks between vocal and drumming modes showing alignment in phrasing. We also include a link to corresponding auditory recordings of this sequence to allow readers to experience the different levels of drumming IOI and vocal sequence synchronization.

Figure 4 illustrates the match between Vocal Singing (VS) and Drum Singing (DS), again with auditory recordings of the sequences. Here, the correlation between vocal and drumming modes drops further to $r = 0.25$, though still maintains a significant one-sample *t*-test, $t(28) = 4.02$, $p < 0.001$, Cohen's $d = 0.75$, a medium effect size. A contributing factor to the lower correlation between VS and DS is the increased use of rhythmic embellishments of some syllable onsets in the form of flams and àfikún (Euba 1990) along with the insertion of short rhythmic patterns that may serve a time-keeping function (Panteleoni 1972). Removing flams and the second onset from each pair of 16th notes that make up the àfikún ranges from a minor positive impact on correlations to a significant improvement on correlations for individual performers, but did not produce a significant change in the overall pattern of correlations ($r_{\text{original}} = 0.25$, $t(28) = 4.02$, $p = <0.001$, $d = 0.75$, $r_{\text{corrected}} = 0.40$, $t(28) = 5.53$, $p < 0.001$, $d = 2.59$, a 10% increase in variance accounted for).

Figure 5 illustrates the Drum-Dance Rhythm (D-DR) IOI pattern with corresponding auditory recording. Here, the sequence is not correlated with any particular vocal pattern so none are shown. This IOI pattern shows the most musical rhythmic characteristics, with repeating patterns of durations that relate via even ratios (e.g., 2:1) which follow the same patterns of musical beats and their subdivisions, alternating with series of periodic durations. As with the other samples, large peaks represent breaks between sub-phrases.

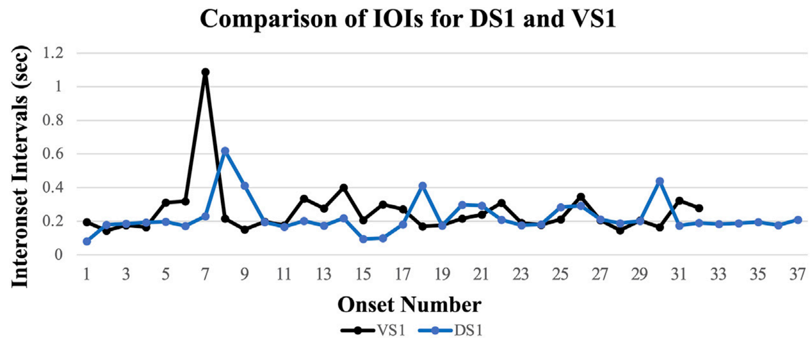


FIGURE 4 | Interonset intervals (IOIs) for a typical vocal singing sample (VS) and the corresponding imitation of the song on the drum (DS). Present in the graph of DS are rhythmic embellishments that shift the timing profiles of DS and VS out of alignment (see sound samples **Supplementary Samples 2A,B**).

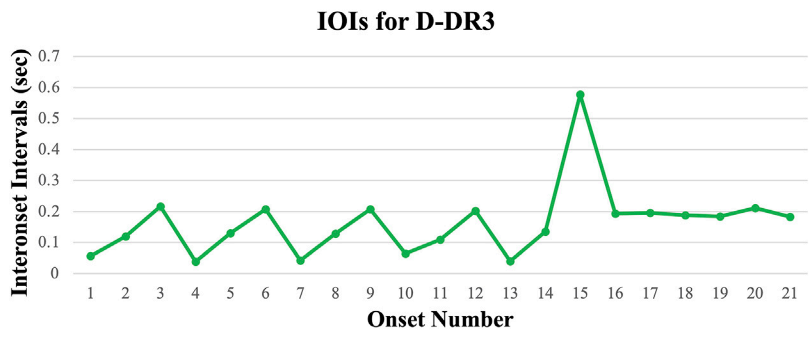


FIGURE 5 | Interonset intervals (IOIs) for a typical drumming pattern in the Drum Dance Rhythm condition (D-DR). The repeating pattern of durations present in the opening of the sample roughly follows a 4:2:1 ratio, which translates into durations of 200, 100, and 50 ms, respectively, with microtiming variation in each category. The sample ends with an isochronous series of durations (see sound **Supplementary Samples 3**).

A three-way, omnibus ANOVA comparing the effects of and interactions between drum and vocal correlations across mode (DT-D and VT, DT-P and VT, and DS and VS), performer, and correction for rhythmic embellishments revealed only a significant main effect of mode, $F(2,156) = 21.29$, $p < 0.001$, partial $\eta^2 = 0.21$, a medium–though verging on large–effect according to conventional norms (Cohen et al., 2003). There was a significant linear contrast between individual modes with VT and DT-D yielding the highest average correlation coefficients (average $r = 0.72$), VT and DT-P yielding a moderate amount (average $r = 0.40$), and VS and DS yielding the lowest average correlation coefficients (average $r = 0.33$), $F(1,171) = 31.57$, $p < 0.001$, $\eta^2 = 0.20$, a large effect according to Cohen (1988) norms. The discrepancy between the effect sizes of the omnibus test and the linear contrast is likely due to the difference in degrees of freedom. Either way, there is between a medium and large effect of mode. These analyses (for original data only) can be seen graphically in **Figure 6**.

In our secondary timing analysis, we compared the average difference between successive IOI times of the four drumming modes. Timing differences between IOIs differed as a function of drum mode type, as can be seen in **Figure 7**. As predicted, there was a significant linear trend indicating that differences between successive IOIs increased as drumming became more constrained to speechlike characteristics (e.g., from pure musical rhythm (D-DR) to direct speech imitation (DT-D)), $F(1,116) = 43.38$, $p < 0.001$, $\eta^2 = 0.27$, which is a large effect according to Cohen (1988) norms. **Figure 7A** illustrates four typical IOI patterns for comparison, and **Figure 7B** illustrates the linear trend of increasing IOI mean as the drumming mode has more speechlike qualities and fewer musical rhythms.

As a third and final timing analysis, we also performed a post-hoc comparison of timing data from speech excerpts and their correlated drum samples. Flams and àfikún differed across mode and drummer. A two-way ANOVA of flam rates indicated a significant main effect of drummer, but not mode or an interaction between the two, $F(2,80) = 5.44$, $p < 0.01$, partial $\eta^2 = 0.12$, a medium effect. Pairwise comparison revealed that this effect was driven by Drummer A's higher average use of flams ($M = 0.40$ flams per phrase) than either Drummer B ($M = 0$) or C ($M = 0.10$). A two-way ANOVA of àfikún rates, however, revealed a significant main effect of mode, drummer, and an interaction between the two. Drum/vocal mode yielded a difference, $F(2,80) = 9.10$, $p < 0.001$, partial $\eta^2 = 0.19$, a medium effect, that pairwise comparisons indicated was due to the significantly lower rate of àfikún in DT-D ($M = 0.17$ per

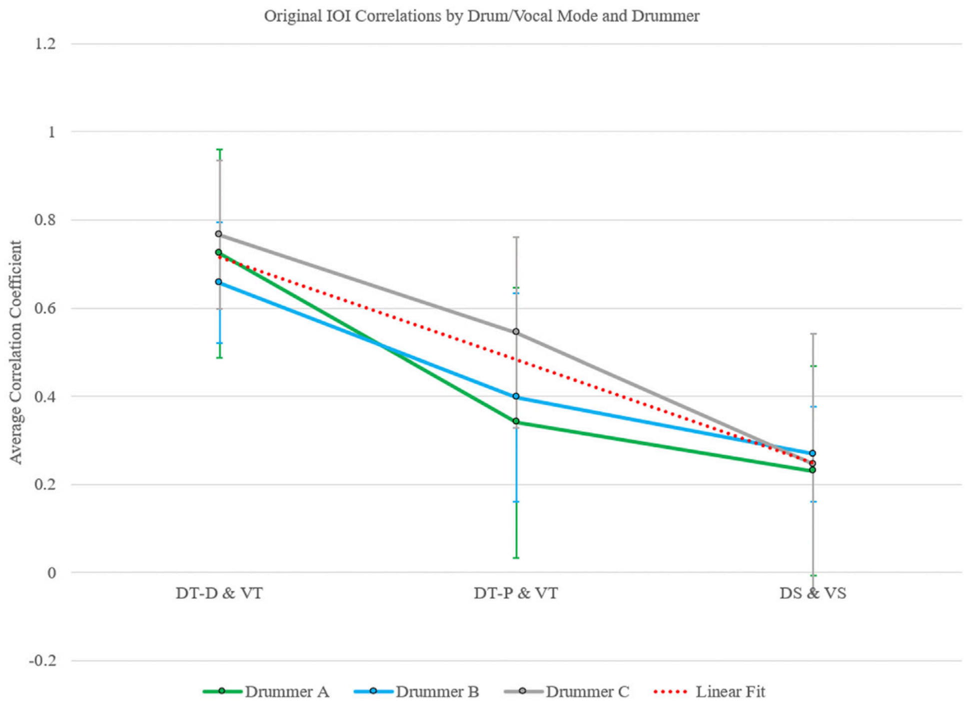


FIGURE 6 | Original IOI correlations between modes and drummers. The data closely follows the linear fit (dotted red line), showing a successive, incremental decrease in average correlation coefficient as the drum/vocal gains musical rhythms. As discussed later, this is likely due to the presence of increasing inter-phrase rhythms as modes become more musical. Error bars are ±2 SE.

phrase) compared to DT-P (M = 1.13) and DS (M = 1.09). Drummer yielded a difference, $F(2,80) = 12.67$, $p < 0.001$, $\eta^2 = 0.24$, a large effect, that pairwise comparisons indicated was due to Drummer C's higher usage (M = 1.52) compared to either Drummer A (M = 0.60) or Drummer B (M = 0.27). Finally, the correlation between mode and drummer was also significant, $F(4.80) = 3.39$, p <0 .05, $\eta^2 = 0.15$, a medium effect. According to pairwise comparisons, Drummer A rarely used àfikún at all, Drummer B more evenly recruited the technique across their modes of playing, and Drummer C relied most heavily on àfikún in DS.

Similarly, inter-phrase attacks (drum attacks without pitch information that appear to serve a time-keeping role) differed greatly among the recorded phrases. A two-way ANOVA comparing the amount of inter-phrase rhythms between drum/vocal modes and performers yielded a significant main effect of drummer, $F(2,80) = 16.06$, $p < 0.001$, $t(28) = 3.71$, $p < 0.001$, partial $\eta^2 = 0.28$, a large effect. Pairwise comparisons revealed that Drummer B drove this difference, $p < 0.001$ compared to both other drummers. There was also a significant main effect of drum/vocal mode, $F(2,80) = 26.32$, $p < 0.001$, partial $\eta^2 = 0.39$, a large effect. Pairwise comparisons revealed that this difference was driven by the DS and VS mode, $p < 0.001$ compared to both other modes. Finally, the interaction between drummer and mode was significant, $F(4,80) = 17.04$, $p < 0.001$, partial $\eta^2 = 0.46$, also a large effect. Drummer B produced, on average, 7.6 attacks per phrase in the DS mode, but zero inter-phrase attacks in both other modes. Drummer A produced, on average, 1.4 attacks in the DS mode, 0.4 in the DT-P mode, and zero in the DT-D mode. Drummer C produced no inter-phrase attacks in any mode. An example of these inter-phrase rhythms in the use of the drum as song surrogate is shown in **Figure 8**, with the inter-phrase rhythms resulting in more drum onsets than syllabic onsets in the corresponding song sample. It appears that in some instances the drummer is simultaneously recreating the song sample phrases and playing a pacemaker rhythm to keep time (Anku 1997).

Discussion of Timing Results

The results confirmed our hypotheses, with the highest correlations for timing occurring between DT-D and VT and getting progressively smaller as drum samples become increasingly more music-like, with the lowest correlation occurring between DS and VS. DT-D samples typically present a near-direct mapping of syllables to drum onsets (r = 0.72), following the same phrasing patterns with gaps between sub-phrases as seen in **Figure 3**. Most DT-D samples do feature between 1–3 additional attacks beyond the number of syllables present in the speech samples, though this is likely due to differences in the representation of elisions on the drum vs. in speech (Euba 1990). In contrast, DT-P samples typically incorporate a greater number of attacks, including the use of significantly more rhythmic embellishments such as flams and àfikún. The inclusion of these rhythmic embellishments varies across the samples collected, resulting in variances in correlation of the timing profiles between DT-P and VT. In most cases, the

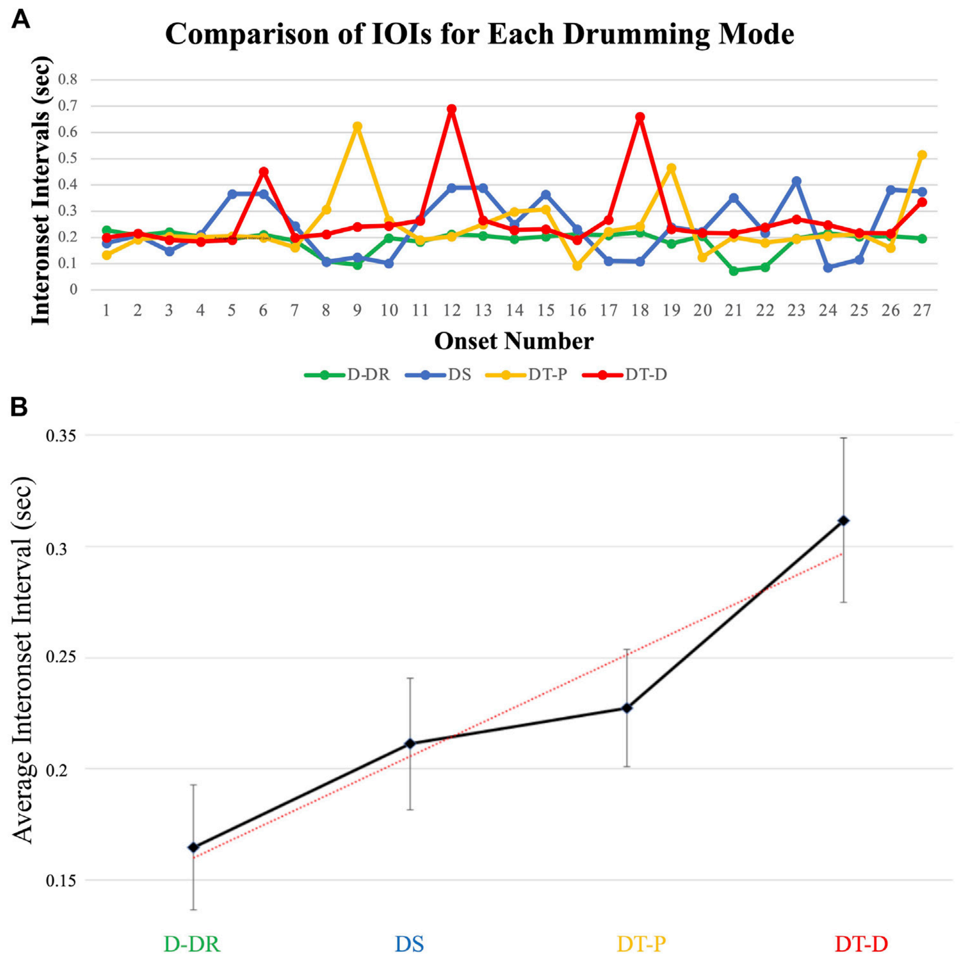


FIGURE 7 | A comparison of IOIs in seconds for representative samples of the four drumming modes. **(A)** Top graph illustrates the temporal IOI pattern for each sample (see sound samples **Supplementary Samples 4A–D**). **(B)** The bottom graph summarizes the average IOIs across all samples for each drumming mode (with error bars indicating ±2 SE). Note the increase in average duration length as samples move from musical rhythms (D-DR, sound sample **Supplementary Samples 4A**) towards the direct imitation of speech (DT-D, sound sample 4d) where more speechlike constraints are introduced to drumming rhythms. Dotted line illustrates a significant linear trend.

general phrase structure is preserved, but more attacks are added into each sample phrase, altering the timing profile while still retaining the same general shape. In keeping with this pattern of increasing musicality changing the drum signal, differences between IOIs linearly increased from D-DR to DS to DT-P to DT-D. Significant differences in IOIs, particularly those that are not in an even ratio with one another (e.g., 2:1) can be understood as either breaks or deviations from a prevailing rhythmic pattern, or as a representation of the more irregular timing patterns of speech. In essence, this result simply shows that DS more closely adheres to rhythmic synchrony than DT-D, which shows little evidence of rhythmic synchrony. The occurrence of inter-phrase rhythms follows a similar pattern.

Though the majority of the DT-P samples feature a greater, though not significantly greater, number of onsets than their DT-D and VT counterparts, the overall rhythmic profile is still significantly more irregular than what is seen in the samples when the drum is being used as a musical instrument (D-DR). Samples where the drum is imitating specific songs (DS) represents a middle ground between the rhythmic profiles of D-DR and DT-P samples, combining some aspects of the irregular timing patterns seen in the DT-P and DT-D samples with periods of isochrony or repeating rhythmic patterns reminiscent of the rhythmic regularity of the D-DR samples. Like the DT-P samples, some of the DS samples also use flams and àfikún, and these rhythmic embellishments can at times obscure the similarities in timing profile that do exist between the DS samples and their corresponding song excerpts (VS). Removing flams and the second attack from the paired 16th notes that make up the àfikún gestures does not significantly improve the overall correlations across all samples, though it does have a differential effect on individual correlations for some performers. Filtering these embellishments out of the data set does not address all of the differences between DS and VS samples that can be attributed to rhythmic variation. For example, **Figure 8** (top) shows the timing profiles for a DS sample that

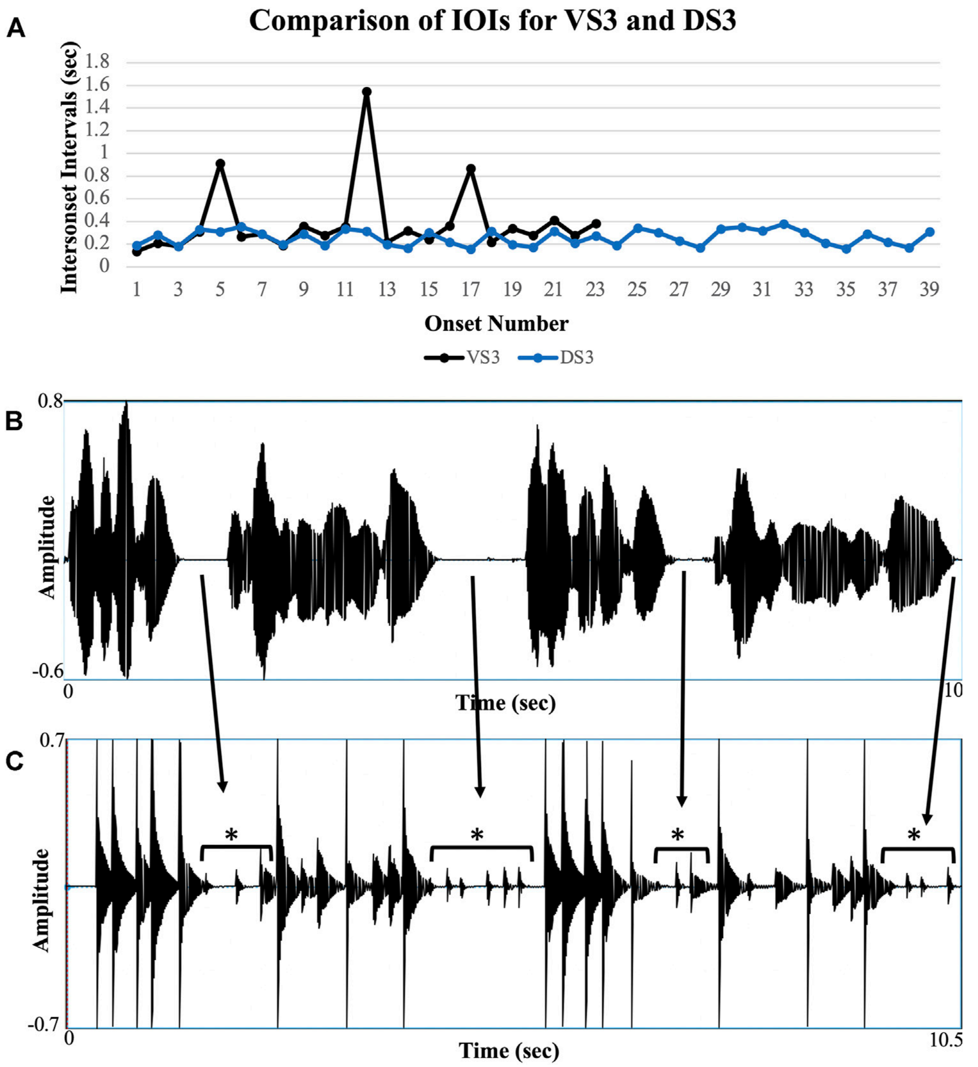


FIGURE 8 | Example of the temporal acoustic correspondence between a typical VS and DS sample phrase. Shown are sound pressure levels during a voice-drum sequence of about 10 s **(A)** IOI timing pattern of a matched set of VS and DS sample phrases (see sound samples **Supplementary Samples 5A,B**). **(B)** Temporal acoustic intensity graph for the VS sample. **(C)** Temporal acoustic intensity graph for the DS sample. Note how the two samples have approximately the same discrete onsets (syllables or drum strikes, respectively) and rate, but that the DS sample has more onset, particularly between sub-phrases consisting of quieter, repeating patterns of attacks that maintain a rhythmic musical time-keeping structure (exemplars marked with an asterisk).

uses neither flams nor àfikún. While similarities in the timing profiles between this DS sample and its paired VS sample are visually apparent, the timing profile of DS is altered by the use of inter-phrase rhythms that shift it away from the timing profile of VS. The overall regularity and repetitive nature of most of the inter-phrase rhythms is suggestive of a time-keeping function. It may be that the ìyáàlù dùndún is simultaneously representing the song sample phrases and providing a pacemaker rhythm to keep time (Locke 1982; Anku 1997). Differences in performers regarding the use of inter-phrase rhythms results in differing effects on correlations when removing rhythmic embellishments. Performer C uses no inter-phrase rhythms, so the removal of rhythmic embellishments from Performer C's DS samples significantly increases correlations between DS and VS ($r_{original}$ = 0.22, $r_{corrected}$ = 0.73), an increase from three of nine significant individual correlations to eight of 9, and an increase of 49% variance accounted for. In contrast, Performer B uses inter-phrase rhythms in all of his DS samples, with an average of 7.6 inter-phrase rhythm onsets per sample. Consequently, removing rhythmic embellishment has a negligible impact on correlations between DS and VS ($r_{original}$ = 0.26, $r_{corrected}$ = 0.27), with no changes in the significance of individual correlations. Performer A represents a middle ground between Performers C and B, using inter-phrase rhythms in some (but not all) samples, along with a greater use of rhythmic embellishments than Performer B, and so the removal of rhythmic embellishment from his DS samples has a minor positive effect on correlations between VS and DS ($r_{original}$ = 0.17, $r_{corrected}$ = 0.23), with an

TABLE 4 | Rhythm-matched fundamental frequency and intensity correlations between different modes of vocalizing and their matched drumming modes.

Vocalizing	Drumming	Vocal f_0 and drum f_0	Vocal intensity and drum intensity	Vocal f_0 and drum intensity	Vocal intensity and drum f_0
Vocal talking (*VT*)	Drum talking–Direct (*DT-D*)	$r_{VT,DTD} = 0.32^{**}$ $d = 1.46$	$r_{VT,DTD} = 0.13^{**}$ $d = 0.97$	$r_{VT,DTD} = 0.20^{**}$ $d = 1.27$	$r_{VT,DTD} = 0.05$ $d = 0.29$
	Drum talking–Performative (*DT-P*)	$r_{VT,DTP} = 0.25^{**}$ $d = 1.31$)	$r_{VT,DTP} = 0.08^{*}$ $d = 0.47$	$r_{VT,DTP} = 0.17^{**}$ $d = 0.87$	$r_{VT,DTP} = 0.04$ $d = 0.19$
Vocal singing (*VS*)	Drum Singing(*DS*)	$r_{VS,DS} = 0.58^{**}$ $d = 3.61$	$r_{VS,DS} = 0.29^{**}$ $d = 1.83$	$r_{VT,DS} = 0.27^{**}$ $d = 1.85$	$r_{VT,DS} = 0.16^{**}$ $d = 0.78$

increase from 0 of 10 significant individual correlations to three of 10.

Results of Fundamental Frequency (f_O) and Intensity Analysis

Table 4 summarizes the correlational results comparing the f_0 and intensity patterns between vocal and corresponding drumming modes. Fundamental frequency (f_0) of non-singing speech (VT) correlated significantly with the f_0 of drumming as direct speech (DT-D) and drumming as performative speech (DT-P), though the latter to a lesser extent. Twenty-three of twenty-nine individual samples (average $df = 248$) in which we compared f_0 of speech and DT-D produced significant correlations, with an average group correlation of $r = 0.32$. A single-sample t-test comparing individual correlations to a null set was significant, $t(28) = 7.88$, $p < 0.001$, Cohen's $d = 1.46$, a large effect according to Cohen (1988) norms. Similarly, twenty-four of twenty-nine individual samples (average $df = 236$) comparing the f_0 of speech and DT-P correlated significantly, producing an average group correlation of $r = 0.25$. A single-sample t-test of these correlations also was significant, $t(28) = 7.07$, $p < 0.001$, Cohen's $d = 1.31$, a large effect. DT-D produced significantly larger f_0 correlations with drum ($M = 0.32$) than DT-P ($M = 0.25$), paired $t(28) = 2.50$, $p < 0.05$, Cohen's $d = 0.46$, nearly a medium effect. Like the pattern found with the f_0 correlations, the correlation in intensity patterns between VT and DT-D was larger than that between VT and DT-P (with Cohen's d indicating a small effect size), however the difference did not reach statistical significance, $t(27) = 1.53$, $p = 0.14$, Cohen's $d = 0.29$.

Similarly, speech intensity was correlated with drum intensity for both DT-D and DT-P, albeit with smaller effect sizes. Sixteen of twenty-nine individual samples (average $df = 328$) correlated significantly, producing an average group correlation of $r = 0.13$, a small effect. A single-sample t-test yielded a significant result, $t(28) = 5.23$, $p < 0.001$, Cohen's $d = 0.97$, a large effect. Fourteen of twenty-eight individual correlations between speech and DT-P intensity were significant (average $df = 324$) correlated significantly, producing an average group correlation of $r = 0.08$, a null effect. A single-sample t-test indicated that these correlations significantly differed from a null set, $t(27) = 2.51$, $p < 0.05$, Cohen's $d = 0.47$, a small effect. These correlations are also presented in **Table 4**. Though these intensity correlations differ slightly in both coefficient and effect size, a paired t-test indicated no significant difference between DT-D and DT-P, $t(27) = 1.53$, $p = 0.14$, Cohen's $d = 0.29$, a small effect.

Because perception of f_0 and intensity are not one-to-one correlations, cross-feature correlations were also examined. Drum f_0 and speech intensity were correlated for neither DT-D (average $r = 0.05$, one-sample $t(28) = 1.55$, $p = 0.13$, Cohen's $d = 0.29$) nor DT-P (average $r = 0.04$, one-sample $t(28) = 1.0$, $p = 0.33$, Cohen's $d = 0.19$). Drum intensity and speech f_0, however, were both significantly correlated. The average correlation for DT-D intensity and speech f_0 ($r = 0.20$, one-sample $t(28) = 6.86$, $p < 0.001$, Cohen's $d = 1.27$, a large effect) was slightly larger than DT-P intensity and speech $f0$ ($r = 0.17$, one-sample $t(28) = 4.70$, $p < 0.001$, Cohen's $d = 0.87$), though not significantly so (paired $t(28) = 1.10$, $p = 0.28$, Cohen's $d = 0.20$). An exemplar showing the similar f_0 and intensity contours can be seen in **Figure 9**.

Vocal singing (VS) and the corresponding drum imitation (DS) were also significantly correlated in f_0. Twenty-nine of twenty-nine individual samples (average $df = 472$) correlated significantly, producing an average group correlation of $r = 0.58$. A single-sample t-test comparing individual correlations to a null set was significant, $t(28) = 19.42$, $p < 0.001$, Cohen's $d = 3.61$, a large effect. Similarly, the intensity of singing and the matched drum rhythm was significantly correlated. Twenty-eight of twenty-nine individual samples (average $df = 611$) correlated significantly, producing an average group correlation of $r = 0.29$. A single-sample t-test comparing individual correlations to a null set was significant, $t(28) = 9.85$, $p < 0.001$, Cohen's $d = 1.83$, a large effect. Cross-acoustic correlations were also significant. Twenty-two of twenty-nine drum f_0 and singing intensity correlations were significant, producing an average correlation of $r = 0.16$ (average $df = 472$), a small effect, and a significant one-sample t-test, $t(28) = 4.08$, $p < 0.001$, Cohen's $d = 0.76$, a medium effect. Twenty-eight of twenty-nine drum intensity and singing f_0 correlations were significant, producing an average correlation of $r = 0.27$ average $df = 566$), a nearly medium effect, and a significant one-sample t-test, $t(28) = 9.94$, $p < 0.001$, Cohen's $d = 1.85$, a large effect. These correlations are also summarized in **Table 4**.

Differences in average correlation across drum/vocal modes and the acoustic features correlated were examined in an omnibus ANOVA. There was a significant difference of average correlation between DT-D, DT-P, and DS, $F(2,335) = 36.75$, $p < 0.001$, $\eta^2 = 0.07$, a medium effect according to Cohen (1988) norms. Pairwise comparisons reveal that this effect is carried by the much larger average correlation for DS ($r = 0.33$) compared to DT-D ($r = 0.18$)

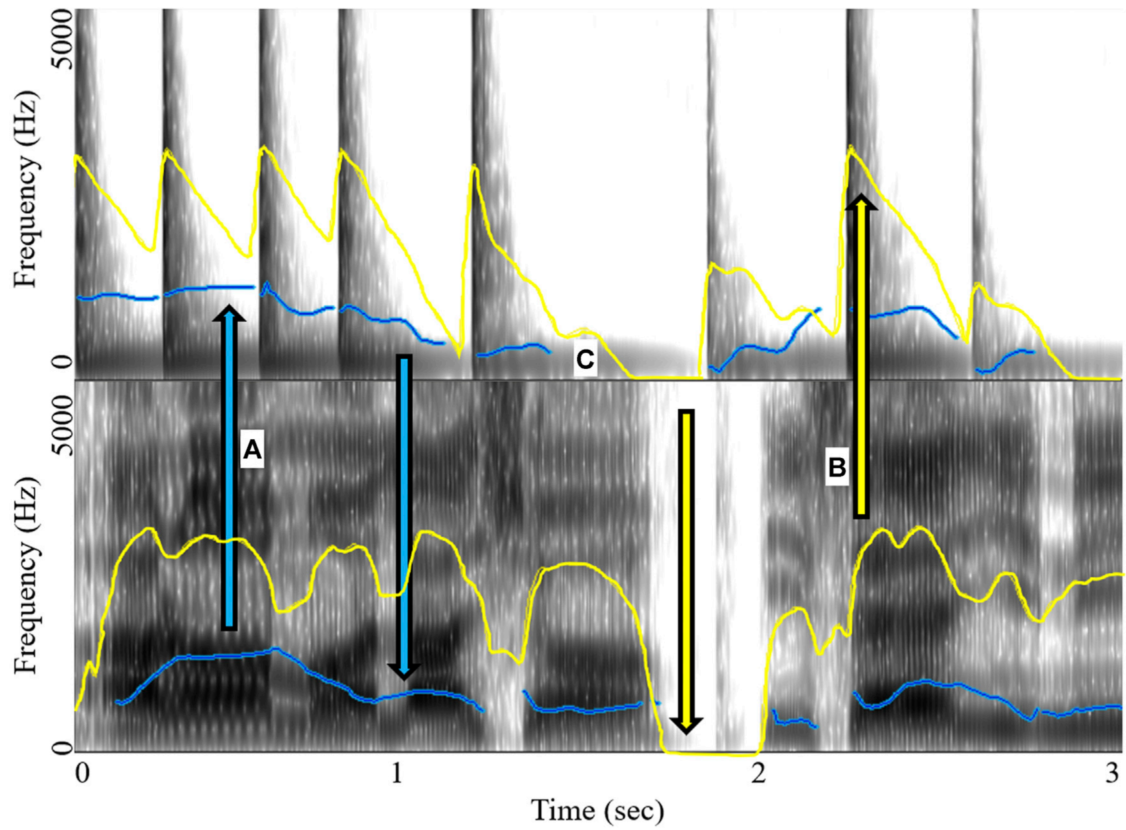


FIGURE 9 | Fundamental frequency (f_0) and intensity contours for drum and speech. This exemplar of the f_0 and intensity data (pulled from Praat for Drummer B's Song 5, which exhibited an f_0 correlation of $r(429) = 0.80$ and intensity $r(664) = 0.50$) shows the correlated movements between **(A)** drum and speech f_0 and **(B)** drum and speech intensity. This graphic also shows **(C)** gaps that appear in the data (most often in f_0), which are often caused by an actual lack of data at the moment or a dearth of data points that does not allow for an accurate estimate of the acoustic feature. Finally, the temporal asymmetry between samples is also evident. This is what necessitates the downsampling procedure to equalize the length of each sample phrase.

and DT-P (r = 0.13). There was also a significant effect of the acoustic features (f_0 or intensity) correlated with one another, $F(3,335) = 42.96$, $p < 0.001$, $\eta^2 = 0.13$, nearly a large effect. Pairwise comparisons show that this effect is due to the significantly higher f_0/f_0 correlations (r = 0.38) compared to others, and the significantly lower drum intensity/speech f_0 correlations (r = 0.08) compared to others. The intensity/intensity (r = 0.17) and drum f_0/speech intensity (r = 0.21) correlations are not statistically different from one another. There was also a statistically significant interaction between drum/voice mode and correlated acoustics, $F(6,335) = 2.86$, $p <0 .05$, $\eta^2 = 0.01$, a small effect, though very close to the benchmark minimum. This interaction is likely spurious as it seems to indicate no meaningful difference. These average correlations are depicted graphically in **Figure 10**. A post hoc one-way ANOVA to identify the nature of these differences investigated the standard deviation of f_0 differences across all modes independently (DT-D, DT-P, VT, DS, and VS), which revealed a significant effect, $F(4,140) = 11.91$, $p < 0.001$, $\eta^2 = 0.25$, a large effect. Pairwise comparisons revealed that this result was driven only by a significantly lower f_0 standard deviation for VT (M = 17.83) compared to DT-D (M = 23.50), DT-P (M = 27.37), DS (M = 26.00), and VS (M = 23.53).

Discussion of Fundamental Frequency (f_0) and Intensity Results

Overall, significantly higher correlations are observed Drum Singing and Vocal Singing (DS and VS), which is in line with neither our hypothesis nor the results of the timing analysis. This result is notable because this is the first study to show the relationship between dùndún drumming as song and vocal singing. The DS and VS correlations are larger for both f_0 and intensity as well as their cross correlations, which may be due, in part, to the more restricted range of the f_0 for Vocal Talking (VT). It is typical for the frequency range of singing to be larger than that of talking (Hacki 1996), so this is not an abnormal result. While singing, however, the variance in f_0 increases to match the exhibited f_0 range of the drum. DS and VS were the only correlates to show a significant average correlation coefficient for all acoustic feature correlations. This indicates that in addition to the similar f_0 range, drummers may be using different information from the VS phrases to produce their DS phrases

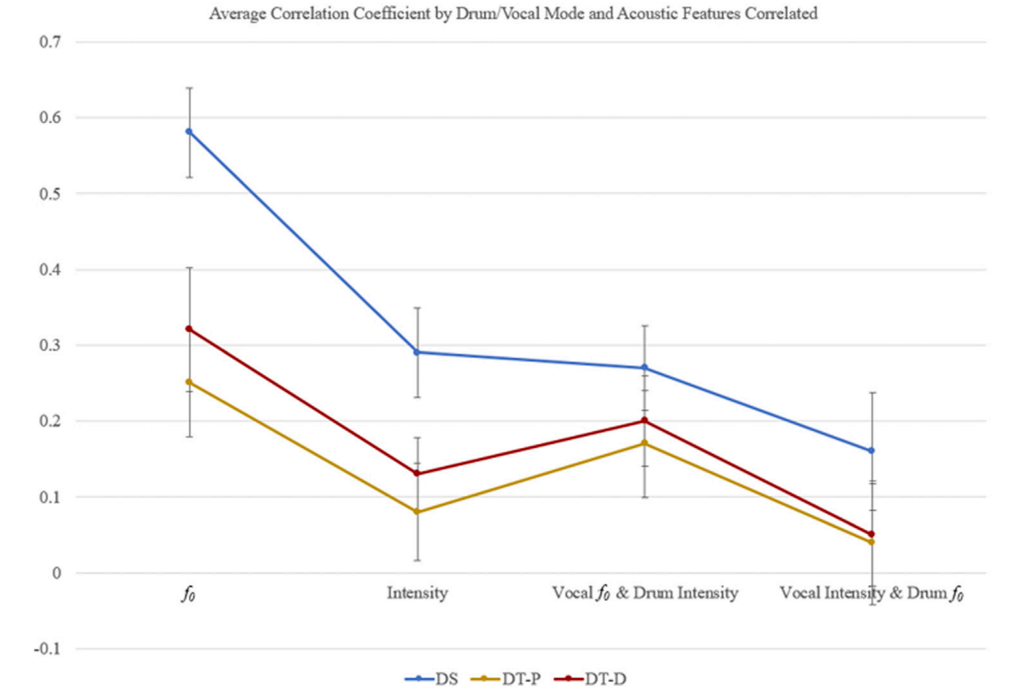


FIGURE 10 | Average correlation coefficients for all songs broken down by drum/vocal mode (DS, DT-P, and DT-D) and acoustic features used to obtain the correlation (drum f_0/vocal f_0, drum intensity/vocal intensity, drum intensity/vocal f_0, and drum f_0/vocal intensity). The significant difference of DS compared to both DT-P and DTD is apparent, as is the large effect of f_0/f_0 correlations, small-to-medium effect of intensity/intensity and drum intensity/vocal f_0, and the insignificant effect of drum f_0/vocal intensity. Error bars are ±2 SE.

than when listening to VT and performing either DT-D or DT-P. While the DT-D and VT and DT-P and VT correlations were not as high and did not produce a significant correlation between all acoustic features, the correlations were significant between vocal and drum f_0, vocal and drum intensity, and vocal f_0 and drum intensity.

DS and VS may also have exhibited stronger correlations due to differing perceptions of f_0 in speech and music. While f_0 is an important component of speech perception, it is not always directly perceived. Interactions with intensity and spectral centroid, as mentioned, can deviate pitch perception from frequency (things). Similar pitch changes are also perceived differently based on their location in speech (at the beginning or ending of a syllable) and their location relative to silence (House 1990, House 1995; Mertens 2004). Very quiet f_0s are also not often perceived when occluded by other speech sounds (Mertens 2013, Mertens 2014). While speech is characterized by formant frequencies that vary greatly between speech sounds, this variation is often less pronounced in singing–possibly in an effort to maintain consistent vocal quality for the duration of the sung vocalization (Bloothoofft and Plomp 1986). In the case of music, both vocal and instrumental songs have been found to exhibit similar correlations between changes in f_0 and intensity, so the current music correlations are not unexpected (Scharine and McBeath, 2019; Patten and McBeath, 2020). There are three levels of correlations for acoustic features. The highest correlation, across all modes, is between vocal f_0 and drum f_0. As Yorùbá is a tonal language and f_0 change captures most of the variance of changes between High, Mid, and Low tones (Ọdẹ́jọbí 2007), this result is not surprising. Across all modes, this correlation explains 14% of the variance and, for DS and VS, 34% of the variance. Correlations between vocal intensity and drum intensity, and those between vocal f_0 and drum intensity do not differ from one another significantly. As with the f_0/f_0 correlations, the intensity/intensity correlations are not surprising and simply confirm the dùndún directly imitates speech. The cross correlation between vocal f_0 and drum intensity, however, is surprising. This correlation may indicate that drummers are using intensity variations to perceptually change the pitch of the notes they play and further enhance the dynamics of the drum. While this correlation only explains 4% of the variance overall, it explains 7% for DS and VS. Lastly, the correlation between drum f_0 and vocal intensity was significantly lower than all other correlations. In fact, for all modes except DS and VS, it was significantly different from a null set of correlations. This indicates that, while some acoustic features inform others, the intensity changes in voice are not features used by dùndún drummers to inform the f_0 of their performances (Oyetade et al., 2003). While this correlation is significant for DS and VS and may indicate a difference in how DS is performed to imitate singing, the correlation is small according to Cohen (1988) norms and the significance of the one-sample t-test could be due to a large sample size.

GENERAL DISCUSSION

Yorùbá dùndún drumming is a classic example of speech surrogacy in which a musical instrument represents pitch and rhythmic characteristics of vocal utterances. The primary purpose of this research was to examine the speech-music relationship in the dùndún, thereby laying the groundwork for the understanding of the functioning of speech surrogate systems. Our goal was to determine the extent of acoustic representation of speech sounds on the drum, as well as to compare four dùndún performance modes. Here, we examine microtiming, f_0, and intensity patterns and provide a microstructural acoustic analysis that verifies the acoustic correspondence between talking and singing vocalization modes as well as their corresponding drumming modes. Analysis of 29 spoken and sung verbal utterances (n = 58) and corresponding drum modes (n = 87), along with samples of the drum performing as a musical instrument (n = 30) demonstrated that the four distinct drumming modes reflect a variety of imitation styles, ranging from purely rhythmic to direct speech imitation. Our microtiming findings confirmed our general hypotheses that comparisons between vocal and corresponding drumming modes are highest for when the drum is directly imitating speech (DT-D vs VT), next when the drum is performatively imitating speech (DT-P vs VT), and weakest when the drum is imitating song (DS vs VS). As anticipated, the purely rhythmic drumming mode (D-DR) differed significantly in comparison to other drumming modes, demonstrating that the dùndún can embody multiple distinct modes (e.g., rhythmic instrument, song surrogate, and speech surrogate). Correlations between f_0 and intensity, however, revealed a somewhat contradictory finding in that DS and VS yielded significantly better coefficients across all acoustic feature correlations, while DT-D and VT and DT-P and VT did not differ. It is possible this difference is, in part, due to the removal of inter-phrase rhythms from the f_0 and intensity analysis, a necessary change to equalize the duration of sub-phrases and accurately model correlations. The significantly greater number of rhythmic embellishments in DT-P and DS likely hindered IOI correlations but not those of acoustic features, as many did not produce f_0 information, were produced with low intensities, or occurred quickly enough that variations between correlates were negligible.

While previous studies have demonstrated frequency correlation between drum and word-level utterances (Akinbo 2019), we showed that this frequency correlation also exists at the phrase level. Novel to this study is our finding of significantly higher frequency correlations between the DS mode and song excerpts when compared to speech, demonstrating that the use of the drum as a speech surrogate can also be extended to the imitation of song as well. Previous research has often treated drum patterns as either representations of pure linguistic utterances (e.g., poetic phrases) or has treated those same drum patterns as musical patterns depending on the disciplinary focus of the researcher. Following Nketia (1963), most researchers discuss the signal, dance, and speech modes of the dùndún talking drum, with most acoustical analysis taking place on the speech mode of the drum (Akinbo 2019). To our knowledge, Euba (1990) is the only scholar to acknowledge the singing mode of the drum, and to date, no one has studied the acoustical features of this mode, making our study the first to do so. One reason scholars may have traditionally omitted the singing mode from discussion is due to difficulties in distinguishing between the speech and song modes of the dùndún drum (Euba 1990). However, our results show a consistent microstructural difference between speech (DT-D and DT-P) modes and singing modes (DS) on the drum, offering new evidence that can assist in acoustically differentiating between these modes in future studies. One of the differences we found between the speech and song drum modes was the inclusion of inter-phrase rhythms in several of the DS samples. These inter-phrase rhythms helped to communicate a clear and consistent pulse through the addition of repetitive rhythms between sub-phrases or in the use of one- or two-timed attacks. Research on West African drumming consistently cites the importance of a steady beat, often established and maintained via the bell pattern, a recurrent timeline played by one or more instruments in a drum ensemble (Locke 1982; Anku 1997; Oludare 2016). The inter-phrase rhythms in the DS samples may serve a similar purpose, providing a pacemaker rhythm that keeps time underneath the song sample phrases that the drum is performing. Indeed, it was found that the inter-phrase rhythm of performer B tends to follow the Long-short-short (L-S-S) pattern, known as "konkolo" rhythm among the Yoruba (Oludare 2016), and is also consistent with the accompaniment pattern in Segu Bamana drumming (Polak and London 2014). It is worth noting that the use of inter-phrase rhythms varied on a performer basis, with Performer B consistently using inter-phrase rhythms in all of his imitations of song on the drum, while Performer C used no inter-phrase rhythms. This variation between performers in our study parallels similar inter-performer differences shown by Polak and London (2014) in Malian drumming and also shows a characteristic element of dùndún performance in which drummers employ individual forms of expression while also maintaining a more general time-keeping function.

Also of note is our finding of significant cross-feature correlations between drum and vocal phrases. Correlations between the f_0 of drum and vocal phrases, and those between intensity, demonstrate that the drum can be used as a surrogate in both talking and singing modes, as mentioned. However, the significant correlation between vocal f_0 and drum intensity may indicate a perceptual correlation between f_0 and intensity in speech surrogacy drumming. Similarly, the use of drum intensity to indicate f_0 changes in speech suggests that drummers are collapsing the highly multidimensional qualities of speech into a few dimensions that can be captured by dùndún drumming.

Future research is needed to continue to deepen our understanding of the acoustical differences between the speech and singing modes of the ìyáàlù dùndún. While our study is the first to demonstrate clear microstructural acoustical differences between these modes, further study is warranted. Additionally, further examination of the different modes of the dùndún could compare them to modes of other percussion instruments in the talking drum family that do not seem to mimic speech as directly. Such analyses could complement our current findings, especially how other drumming modes are similar to or differ from the dùndún's speech and song surrogacy modes. Such a focus promises to reveal the deeper relations between speech and its imitation on musical instruments, but also aid further understanding of general percussive mechanisms of speech surrogates.

In conclusion, our findings confirm that the dùndún can very accurately mimic microstructural acoustic temporal characteristics of Yorùbá vocalization when doing so directly, and that this acoustic match systematically decreases as more musical context is specified. This pattern of imitation is maintained to a high degree for both speech and song in talking drum performances and is largely absent when used principally as a dance rhythm instrument. Frequency and intensity characteristics, on the other hand, are more closely matched for song than talking, which may be due to the constrained frequency range in the vocal talking phrases. Our findings acoustically verify the distinction between the four drumming mode categories DT-D, DT-P, DS, and D-DR, and their acoustical match to corresponding verbal modes. Understanding the process of music and speech interconnectivity in the dùndún talking drum helps clarify acoustical properties that overlap between these modes of communication and verifies the potential functionality of speech surrogacy communication systems.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CD, KK, and MM designed the study. CD collected the data. KK, JP, and MG analyzed the data. CD, KK, and JP wrote the article. MM contributed to writing. All authors reviewed, edited and approved the content of the article.

ACKNOWLEDGMENTS

Special thanks to Mitchell Ohriner, who offered advice on methods for analyzing timing of syllable onsets in speech, and to Robert Ghilduta for assistance. We also thank the performers and the ELYON studio staff for the time spent on recording the samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.652690/full#supplementary-material

REFERENCES

Adegbite, A. (1988). The Drum and its Role in Yoruba Religion. *J. Religion Afr.* 18, 15–26. doi:10.1163/157006688X00207

Akinbo, S. (2019). Representation of Yorùbá Tones by a Talking Drum: an Acoustic Analysis. *Ling. Lang. Afric.* 5, 11–23. doi:10.31234/osf.io/43gf6

Akpabot, S. (1975). The Talking Drums of Nigeria. *Afric. Mus.* 5, 36–40. doi:10.21504/amj.v5i4.1616

Akpabot, S. (1986). *Foundations of Nigerian Traditional Music*. Ibadan: Spectrum Books.

Allen, G. D. (1972). The Location of Rhythmic Stress Beats in English: an Experimental Study I. *Lang. Speech* 15, 72–100. doi:10.1177/002383097201500110

Anku, W. (1997). Principles of Rhythm Integration in African Drumming. *Black Music Res. J.* 17, 211–238. doi:10.2307/779369

Arewa, O., and Adekola, N. (1980). Redundancy Principles of Statistical Communications as Applied to Yorùbátalking-Drum. *Anthropos* 75, 185–202.

Ashley, R. (2002). Do[n't] Change a Hair for Me: The Art of Jazz Rubato. *Music Percept.* 19, 311–332. doi:10.1525/mp.2002.19.3.311

Bamgbose, A. (2000). *A Grammar of Yoruba*. Cambridge: Cambridge University Press, Vol. 5.

Barber, K., and Oyetade, A. (1998). *Yoruba Wuyi. Iwe Kinni (Book One): A Beginners' Course in Yoruba*. London: Hakuna Matata Press.

Benadon, F. (2006). Slicing the Beat: Jazz Eighth-Notes as Expressive Microrhythm. *Ethnomusicology* 50, 73–98.

Bloothofft, G., and Plomp, R. (1986). The Sound Level of the Singer's Formant in Professional Singing. *The J. Acoust. Soc. America* 79, 2028–2033. doi:10.1121/1.393211

Boersma, P., and Weenink, D. (2021). Praat: Doing Phonetics by Computer. Version 6.1.37. Available at: http://www.praat.org/(Accessed January 1, 2021).

Byrd, D. (1992). Perception of Assimilation in Consonants Clusters:A Gestural Model. *Phonetica* 49 (1), 1–24. doi:10.1159/000261900

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed.. Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd ed.. New York: Routledge.

Connell, B., and Ladd, D. R. (1990). Aspects of Pitch Realisation in Yoruba. *Phonology* 7, 1–29. doi:10.1017/s095267570000110x

Durojaye, C. (2019a). *Evoked Emotional Responses in the Performance Practices of Selected Yorùbá Dùndún Ensembles [dissertation]*. [South Africa]: University of Cape Town.

Durojaye, C. (2019b). Born a Musician: the Making of a Dundun Drummer Among the Yoruba People of Nigeria. *J. Arts. Hum.* 8, 43–55. doi:10.18533/journal.v8i1.1551

Durojaye, C. (2020). The Dùndún Talking Drum of the Yorùbá Community in South-West Nigeria. *J. Arts. Hum.* 9, 11–19. doi:10.18533/journal.v9i7.1902

Durojaye, C., Fink, L., Roeske, T., Wald-Fuhrmann, M., and Larrouy-Maestri, P.. (2021). Perception of Nigerian Talking Drum Performances as Speech-like vs Music-like: the Role of Familiarity and Acoustic Cues. Manuscript in Preparation.doi:10.4324/9781003175049

Eme, C. A., and Uba, E. D. (2016). A Contrastive Study of the Phonology of Igbo and Yoruba. *Ujah J. Art Hum.* 17, 65–84. doi:10.4314/ujah.v17i1.4

Euba, A. (1990). *Yorùbá Drumming: The Dùndún Tradition*. Bayreuth: Bayreuth African Studies.

Fox, R. A. (1983). Perceptual Structure of Monophthongs and Diphthongs in English. *Lang. Speech* 26, 21–60. doi:10.1177/002383098302600103

Gay, T. (1970). A Perceptual Study of American English Diphthongs. *Lang. Speech* 13, 65–88. doi:10.1177/002383097001300201

Goldberg, D. (2015). Timing Variations in Two Balkan Percussion Performances. *Emp. Mus. Rev.* 10, 305–328. doi:10.25148/lawrev.10.2.8

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal Properties of Spontaneous Speech-A Syllable-Centric Perspective. *J. Phonetics* 31, 465–485. doi:10.1016/j.wocn.2003.09.005

Hacki, T. (1996). Comparative Speaking, Shouting and Singing Voice Range Profile Measurement: Physiological and Pathological Aspects. *Logopedics Phoniatrics Vocology* 21, 123–129. doi:10.3109/14015439609098879

Harsin, C. A. (1997). Perceptual-center Modeling Is Affected by Including Acoustic Rate-Of-Change Modulations. *Perception & Psychophysics* 59 (2), 243–251. doi:10.3758/bf03211892

Hombert, J. M. (1976). Consonant Types, Vowel Height, and Tone in Yorùbá. *UCLA Working Pap. Phonetics* 33, 40–54.

House, D. (1990). *Tonal Perception in Speech*. Lund: Lund University Press.

House, D. (1995). "Perception of Prepausal Tonal Contours: Implication for Automatic Stylization of Intonation," in Proceedings of the 4th European Conference on Speech Communication and Technology, 4 (1), 949–952. Available at: https://www.speech.kth.se/prod/publications/files/3131.pdf.

Huggins, W. F. (1972). Just Noticeable Differences for Segment Duration in Natural Speech. *J. Acous. Soc. Am.* 51, 270. doi:10.1121/1.1912971

Jun, J. (2004). "Place Assimilation," in *Phonetically Based Phonology*. Editors B. Hayes, R. Kirchner, and D. Steriade (Cambridge: Cambridge University Press), 58–86. doi:10.1017/cbo9780511486401.003

Laniran, Y. O., and Clements, G. N. (2003). Downstep and High Raising: Interacting Factors in Yoruba Tone Production. *J. Phonetics* 31, 203–250. doi:10.1016/S0095-4470(02)00098-0

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., and Elenius, K. (2011). Expression of Affect in Spontaneous Speech: Acoustic Correlates and Automatic Detection of Irritation and Resignation. *Computer Speech Lang.* 25, 84–104. doi:10.1016/j.csl.2010.03.004

Lehiste, I., and Peterson, G. E. (1961). Transitions, Glides, and Diphthongs. *J. Acoust. Soc. America* 33, 268–277. doi:10.1121/1.1908638

Locke, D. (1982). Principles of Offbeat Timing and Cross-Rhythm in Southern Eve Dance Drumming. *Ethnomusicology* 26, 217–246. doi:10.2307/851524

Madison, G. (2001). Variability in Isochronous Tapping: Higher Order Dependencies as a Function of Intertap Interval. *J. Exp. Psychol. Hum. Perception Perform.* 27, 411–422. doi:10.1037/0096-1523.27.2.411

Mazzoni, D. (2013). AudacityI: Free Audio Editor and Recorder (Version 2.0.3) [Mac OS X]. Available at: https://audacityteam.org (Accessed January 1, 2021).

McBeath, M. K., and Neuhoff, J. G. (2002). The Doppler Effect Is Not what You Think it Is: Dramatic Pitch Change Due to Dynamic Intensity Change. *Psychon. Bull. Rev.* 9, 306–313. doi:10.3758/BF03196286

Mertens, P. (2004). "The Prosogram: Semi-automatic Transcription of Prosody Based on Tonal Perception Model" in Proceedings of Speech Prosody 2004, Nara. *Japan*, 23–26. doi:10.20396/joss.v4i2.15053

Mertens, P. (2013). "Automatic Labelling of Pitch Levels and Pitch Movements in Speech Corpora," in Proceedings of Tools and Resources for the Analysis of Speech Prosody 2013. Editors B. Bigi and D. Hirst (Aix-en-Provence: Laboratoire Parole et Language), 42–46. ISBN: 978-2-7466-6443-2.

Mertens, P. (2014). Polytonia: a System for the Automatic Transcription of Tonal Aspects in Speech Corpora. *J. Speech Sci.* 4, 17–57.

Neuhoff, J. G., and McBeath, M. K. (1996). The Doppler Illusion: the Influence of Dynamic Intensity Change on Perceived Pitch. *J. Exp. Psychol. Hum. Perception Perform.* 22, 970–985. doi:10.1037/0096-1523.22.4.970

Nketia, J. H. K. (1963). *Drumming in Akan Communities of Ghana*. Edinburgh: Thomas Nelson & Sons.

Ohriner, M. (2018). *"Expressive Timing,"* in *The Oxford Handbook of Critical Concepts in Music Theory*. Editors A. Rehding and R. Steven (New York: Oxford University Press), 369–396.

Ohriner, M. (2019). Lyric, Rhythm, and Non-alignment in the Second Verse of Kendrick Lamar's "Momma". *Mus. Theo. Online* 25, 1. doi:10.30535/mto.25.1.10

Ola Orie, O., and Pulleyblank, D. (2002). Yoruba Vowel Elision: Minimality Effects. *Nat. Lang. Ling. Theo.* 20, 101–156. doi:10.1023/a:1014266228375

Oludare, O. (2016). An Analysis of the Two Forms of the 'Kónkónkóló' Rhythm in Sakara Music. *J. Assoc. Nig. Mus.* 10, 186–196.

Oyetade, A., Hayward, K., and Watkins, J. (2003). "The Phonetic Interpretation of Register: Evidence from Yoruba," in *Phonetic Interpretation. Papers in Laboratory Phonology IV*. Editors J. Local, R. Ogden, and R. Temple (Cambridge: Cambridge University Press), 305–321.

Ọdéjọbí, Ọ. À. (2007). A Quantitative Model of Yorùbá Speech Intonation Using Stem-ML. *INFOCOMP J. Computer Sci.* 6 (3), 47–55.

Panteleoni, H. (1972). Three Principles of Timing in Anlo Dance Drumming. *Afr. Music* 5, 50–63. doi:10.21504/amj.v5i2.1419

Patten, K. J., Greer, K., Likens, A. D., Amazeen, E. L., and Amazeen, P. G. (2020). The Trajectory of Thought: Heavy-Tailed Distributions in Memory Foraging Promote Efficiency. *Mem. Cogn.* 48, 772–787. doi:10.3758/s13421-020-01015-7

Patten, K. J., and McBeath, M. K. (2020). "The Difference between Shrieks and Shrugs: Spectral Envelope Correlates with Changes in Pitch and Loudness," in Proceedings of the 2nd International Conference on Timbre (Timbre 2020), September 2020, 3–4. Thessaloniki (online), Greece

Patten, K. J., McBeath, M. K., and Baxter, L. C. (2019). Harmonicity: Behavioral and Neural Evidence for Functionality in Auditory Scene Analysis. *Aud. Percept. Cogn.* 1, 158–172. doi:10.1080/25742442.2019.1609307

Peterson, G. E., and Lehiste, I. (1960). Duration of Syllable Nuclei in English. *J. Acoust. Soc. America* 32, 693–703. doi:10.1121/1.1908183

Polak, R., and London, J. (2014). Timing and Meter in Mande Drumming from Mali. *Music Theor. Online* 20, 1-27. doi:10.30535/mto.20.1.1

Pompino-Marschall, B. (1991). The Syllable as a Unit and the So-Called P-center Effect. *Forschungberichte des Instituts für Phonetik und Sprachliche Kommunikationder Universitdt München (FIPMK)* 29, 65–123.

Pulleyblank, D. (2009). Vowel Deletion in Yoruba. *J. Afri. Lang. Ling.* 10, 117–136. doi:10.1515/jall.1988.10.2.117

Repp, B. H. (1992). Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann's "Träumerei". *J. Acoust. Soc. America* 92, 2546–2568. doi:10.1121/1.404425

Repp, B. H. (2005). Sensorimotor Synchronization: a Review of the Tapping Literature. *Psychon. Bull. Rev.* 12, 969–992. doi:10.3758/BF03206433

Rhodes, T., and Turvey, M. T. (2007). Human Memory Retrieval as Lévy Foraging. *Physica A: Stat. Mech. its Appl.* 385, 255–260. doi:10.1016/j.physa.2007.07.001

Scharine, A. A., and McBeath, M. K. (2019). Natural Regularity of Correlated Acoustic Frequency and Intensity in Music and Speech: Auditory Scene Analysis Mechanisms Account for Integrality of Pitch and Loudness. *Aud. Percept. Cogn.* 1, 205–228. doi:10.1080/25742442.2019.1600935

Sluijter, A. M. C., and van Heuven, V. J. (1996). "Acoustic Correlates of Linguistic Stress and Accent in Dutch and American English," in Proceedings of Fourth International Conference on Spoken Language Processing (USA: ICSLP '96, Philadelphia, PA), Vol. 2, 630–633.

Sotunsa, M. (2009). *Yorùbá Drum Poetry*. London: Stillwatersstudios.

Toh, A. M., Togneri, R., and Nordholm, S. (2005). Spectral Entropy as Speech Features for Speech Recognition. *Proc. PEECS* 1, 92.

Vidal, A. (2012). ""Traditional Music Instruments of the South-West Nigeria: Forms and Distribution," in *Selected Topics on Nigerian Music*. Editors F. Adedeji (Ile-Ife: Ife: Obafemi Awolowo University Press),", 43–53.

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bata Drum*. Burlington, VT: Ashgate.

Villepastour, A. (2014). "Talking Tones and Singing Speech Among the Yorùbá of Southwest Nigeria," in *Jahrbuch des Phonogrammarchivs der Österreichischen Akademie der Wissenschaften*, 29–47.

Yu, C. S.-P., McBeath, M. K., and Glenberg, A. M. (2021). The Gleam-Glum Effect: Phonemes Generically Carry Emotional Valence. *J. Exp. Psychol. Learn. Mem. Cogn.* 47. doi:10.1037/xlm0001017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 09 August 2021
doi: 10.3389/fcomm.2021.653268

Systems of Communication: Aspects of Culture and Structure in Speech Surrogates

Lucas James *

Undergraduate Program in Linguistics, Dartmouth College, Hanover, NH, United States

The practice of speech surrogacy is used for communication across many cultures. Previous work has historically engaged with the study of speech surrogates as part of anthropological or ethnomusicological inquiry; more recently, scholars have explored aspects of the formal relationship between spoken and surrogate linguistic structures. How speech surrogates function as systems of communication is not yet well understood. Based on evidence from an interdisciplinary corpus of documentation, characteristics of culture and discourse, as well as features of linguistic structure, are shown to play a role in fostering communicability in speech surrogates. Cultural constraints are linked to the development of a speech surrogate-mediated discourse within a community of practice, facilitating comprehension of the surrogate system. Moreover, specific structures including formulas, enphrasing, and framing devices are identified as common to various speech surrogate traditions, suggesting a common function as aids to communication. This analysis points to the need to investigate speech surrogates as linguistic systems within a discursive context.

Keywords: speech surrogate, communication, discourse, enphrasing, functional pressures

OPEN ACCESS

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Julien Meyer,
Centre National de la Recherche Scientifique (CNRS), France
Neil Coulter,
Dallas International University, United States

***Correspondence:**
Lucas James
david.l.james.21@dartmouth.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *14 January 2021*
Accepted: *29 July 2021*
Published: *09 August 2021*

Citation:
James L (2021) Systems of Communication: Aspects of Culture and Structure in Speech Surrogates. Front. Commun. 6:653268. doi: 10.3389/fcomm.2021.653268

INTRODUCTION

Speech surrogacy is (broadly) the practice of imitating verbal speech without the use of the larynx, often by means of whistling or through the use of musical instruments. The best-known examples include Silbo, the whistled Spanish of the Canary Islands, as well as "talking drums" throughout West Africa, but from the 19th century onwards hundreds of speech surrogates have been attested spanning every inhabited continent (James et al., 2021).

Scholarship on speech surrogates has been historically incidental to documentary work in anthropology or ethnomusicology. Bagemihl (1988), asking "But is it language?", noted that speech surrogates were at the time a marginal subject within linguistics. He concluded that "virtually all of the sources which I have examined . . . fall into the category of descriptive, nontheoretical studies . . . There is an obvious reluctance to delve into an area which smacks of the nonlinguistic" (26). Since many speech surrogates are integrated into larger cultural and musical traditions, a typical early approach was to analyze them as cultural performance or a "musical process" (Kaminski 2008), with a few notable exceptions (Stern 1957; Carrington 1949; Nketia (1971) 1976). This situation recalls past attitudes on writing systems in linguistics: "writing [being] clearly a cultural rather than biological endowment . . . seemed accordingly less interesting [to linguists] than spoken language" (Sampson 2015, 47)—though unlike writing, the origins of speech surrogacy are as yet unclear.

In recent years, speech surrogates have more commonly been taken seriously as a part of language, since researchers have recognized formal similarities between spoken and surrogate modalities.

Analyses suggest that speech surrogates can reproduce aspects of speech including acoustic and phonetic properties (e.g., Rialland, 2005; Meyer, 2008), speech rhythm (e.g., Seifart et al., 2018), phonemic structure (e.g., Villepastour, 2010; McPherson, 2018), and morphosyntactic processes (e.g., Winter 2014). Such contributions show how speech surrogates rely on practitioners' linguistic competence. This makes linguistics well-suited to the task of understanding speech surrogate structure. To adapt terminology from folkloristics: within cross-disciplinary speech surrogate studies, nonlinguistic perspectives may focus on *texts* and *context*, while linguistics provides the tools to study *texture,* the actual structures that make up surrogate speech (cf. Dundes, 1980, 20).

Advancements in the structural accounts of speech surrogates have not yet been matched with a flourishing of insight into functional questions: How are they used and understood as linguistic communication systems? How do they fit into the discursive lives of their listeners and practitioners? In-depth studies on speech surrogate communicability have been rare in linguistics, though some authors have explored aspects of perception and discourse (see *Previous Work on Speech Surrogate Communication*).

Surrogate language communication, embedded as it is within larger systems, may at first seem to stray from the domain of language proper. But Vigliocco et al. (2014) challenge the distinction between "*language proper*, i.e., language as a structured system amenable to linguistic analysis and *communication*, i.e., the broader context of language use, which includes the use of other channels of information" (1). They argue that "[t]he majority of language studies have been firmly focused on language proper, to the exclusion of context and multimodal expression that contribute to utterance and meaning construction" (1). This paper makes the case that the study of speech surrogates is relevant to linguistics not only because it illuminates the former category ("language as a structured system"), but also because it is part of the latter ("context and multimodal expression"). In other words: speech surrogates both tell us about language *and are language*.

This paper focuses on the relationship between form and function in speech surrogates. There is a significant body of evidence that linguistic structure is shaped by its communicative niche (Coupé et al., 2019). An analysis of communication in speech surrogate systems suggests the same functional pressures on spoken and signed languages also affect speech surrogates. Studying these functional pressures can help us distinguish modality effects from universal tendencies, bettering our understanding of language writ large.

Defining Speech Surrogacy discusses the definitions crucial for the typology of speech surrogate communication. *Previous Work on Speech Surrogate Communication* provides a review of existing work on perception and discourse in speech surrogates. *Challenges for Speech Surrogate Communication* presents the challenges to communicability that speech surrogates pose, and *Compensation Strategies in Speech Surrogate Communication* analyzes some attested strategies to counteract these challenges. *Discussion* presents a discussion and avenues for further research.

DEFINING SPEECH SURROGACY

What is a "speech surrogate"? There is no broad scholarly consensus on how to define the phenomenon, or even what to call it, and the definitions rarely have the same scope from author to author. I prefer "speech surrogate" only because it is in common usage, probably because of Stern (1957) and because it is in the title of Sebeok and Umiker-Sebeok (1976) collection of articles in two volumes, an oft-cited source on the subject.

For Sebeok and Umiker-Sebeok, "speech surrogates" are "specimens of one species of transmutation, one which is 1) a true substitutive system, 2) a first-order rather than a second-order system, and 3) in the acoustic modality" (XIX). "First-order" refers to a direct relationship between language and sign, rather than one mediated via another system (such as Morse code's reliance on the alphabet). Therefore, in other words, speech surrogacy is a practice of 1) systematically replacing utterances in 2) a spoken language with 3) other sounds. An alarm bell (say) does not count, as it violates 1) by having no systematic connection to language; Morse code does not count, as it violates 2) by referring to an alphabet rather than to speech itself; and alphabets in turn do not count, as they violate 3) for being silent. What is included is the practice of using communication through manipulating pitch, rhythm, and timbre in lieu of speech sounds, as typified by the "talking drum" and "whistled speech". Based on the predominance of those two forms, Sebeok and Umiker-Sebeok often refer specifically to "drum and whistle surrogates" in the text.

Sebeok and Umiker-Sebeok's definition contains a controversial assumption: how can one know if a system traces back to a spoken language? A number of articles included in the collection describe systems with no apparent connection to the phonology of their corresponding spoken languages. Instead, they arbitrarily signify words, phrases, or perhaps concepts; this is what Stern (1957) calls "lexical ideograph" systems, in contrast with phonologically-based "abridgment" systems (172). Sebeok and Umiker-Sebeok write:

> "In some drum and whistle systems, particularly those observed in Oceania and South America, the symbolic principle appears to be dominant over the iconic . . . More work is necessary on this sort of drum and whistle surrogate before it will be possible to discuss in detail its exact semiotic nature" (XIX).

I am not aware of any theoretical work that solves this puzzle. Some scholars, including (Nketia, 1971), exclude lexical ideograph systems entirely from the domain of surrogate speech. For the purposes of this analysis, I find it necessary to include them. As I describe in *Compensation Strategies in Speech Surrogate Communication*, the line between abridging and lexical ideograph systems is harder to draw than it appears. Moreover, many of these systems are used in functionally identical circumstances. So, for a full description of the speech surrogate communicative niche—that is, using implements or whistling as a genuine replacement for the speech act—I think it is important to account for these systems.

This leads to a second point of clarification: what is meant by "communication"? This paper cannot take a broad view of all the things that speech surrogates *mean*. For a not insignificant (though diminishing) population scattered around our planet, speech surrogates are woven into the fabric of life among their families and neighbors. Speech surrogates are meaningful insofar as they are shared; as such, they are never separate from the identity of the communities that practice them, often part of their music, spirituality, orality, and entertainment. Even for generations losing their grip on the practice, the sound of a surrogate language can signify a lot: shame at a perceived failure to uphold tradition (Coulter 2007), longing for the past or a homeland left behind (Poss 2005). This territory of meaning, linked to tradition and identity, is vast and unsuited for the analysis presented here.

Instead, this paper is focused on how people use speech surrogates to talk to (and about) each other. As meanings attributable to surrogate language go, this is an important one. Speech surrogates of all kinds are used to praise (Kaminski 2008), insult (Vercelli 2006), ridicule (Ames et al., 1971), or court others (Armstrong 1953; Catlin 1982; Dugast (1955)); to tell stories (Armstrong, 1953), to refer to people's names (Dugast, 1955) or their titles (Nketia, 1971; Kaminski, 2008), as well as the names of places (Nketia, 1971), clans (Seifart et al., 2018), and ancestors (Nketia, 1963); to ask for help (Burridge, 1959) or money (Strand, 2009); to address announcements to a whole community (Cloarec-Heiss, 1999) or to a specific person (Damm, 2003); to broadcast calls for celebration (Wojtylak, 2016), mourning (Lewis, 2018), and worship (Neeley, 1999); to coordinate hunting (Blench, 1987), warfare (Gourlay, 1982), agriculture (Wilken, 1979), and group performances (Stone, 1972); and so on. In other words, speech surrogates are often used the same way we all use language around other people.

This is true regardless of a system's structural properties. A Canarian whistler, a Bora *manguaré* drummer and an 'Are'are conch player use radically different methods to communicate, and these differences must have an impact on the way their messages are produced and understood. But all three practitioners have a common purpose: to expand the human communicative palette beyond its natural boundaries. A speech surrogate message might sound louder or travel farther than the human voice, or deliver a heightened register to a culturally meaningful text; it may represent the voice of authority, or simply enliven otherwise plain language. It is this quality—the expansion of communicative possibility—that represents the significance of speech surrogate traditions, and that attracts this investigation into their structure and function.

PREVIOUS WORK ON SPEECH SURROGATE COMMUNICATION

Speech surrogate documentation usually gives some indication of the system's use in communication; detailed theoretical accounts are much rarer. Background knowledge for this analysis comes from a thorough review of the sources listed in the Online Database of Speech Surrogates (ODSS) (James et al., 2021), which catalogues information on roughly 200 speech surrogates of over 100 language varieties. These include recent and very early literature alike, from this decade to the late 19th century. Of those, about 60 have descriptions contained in Sebeok and Umiker-Sebeok (1976). Another dozen or so are found in Meyer (2015), which provides a detailed overview of the typology of documented whistling systems, and is my primary source on that topic. I give preference to substantial works over brief descriptions, and avoid impressionistic comments for which I cannot elaborate on all relevant examples individually. All transcriptions are included verbatim.

I recognize the challenge of comparing studies from across more than a century of developments in linguistic theory. Though early transcriptions usually lack the theoretical grounding that would permit a retrospective phonological analysis, properties of syntax and discourse are often included. Moreover, some of the systems under discussion have now fallen from use and are thus out of the reach of further fieldwork. Many speech surrogates are extinct or endangered because of cultural and technological changes to communication around the world. Along with redoubling our efforts to document living systems, we need strategies to mine the existing literature.

One body of evidence on speech surrogate communication concerns the ability for listeners to understand abridging systems analytically; that is, how they associate individual sounds of surrogate speech with phonemes in their spoken language. Experimental work on this subject has been performed for several whistling systems: whistled Béarnaise in Aas, France (Busnel et al., 1962), whistled Turkish in Küskoy, Turkey (Busnel, 1970; Moles, 1970), and the Silbo Gomero of the Canary Islands (Rialland 2005). These are thoroughly reviewed by Meyer (2015), and I will not reiterate them here. All of these experiments tested the comprehension abilities of practitioners, finding cross-linguistically that whistled phrases could be identified at greater-than-chance rates, though with wide variations in accuracy depending on the type of utterance (from sentences and words to nonsense syllables).

In addition to phrase-identification tasks, several neurolinguistic studies have been performed in the past 2 decades, all on whistled speech processing. Carreiras et al. (2005) used fMRI technology to analyze neural activity in a processing task of Silbo Gomero. While listening to recorded whistle speech, skilled practitioners experienced activation in the left superior temporal gyrus and right-hemisphere superior–midtemporal region, cortical areas associated with speech processing. No such activation was recorded for control participants unfamiliar with Silbo Gomero. The authors posit that "the language-processing regions of the human brain can adapt to a surprisingly wide range of signalling forms" (31). The experimental group included only skilled whistlers rather than participants merely familiar with (but not practitioners) of the whistling system.

Güntürkün et al. (2015) and Villar González et al. (2020) used dichotic listening tasks to probe the localization of language processing in whistled Turkish and Silbo Gomero, respectively. Participants were tasked with identifying a whistled signal after listening to two simultaneous stimuli, one in each ear. Both papers conclude that the standard attested effect of left-hemisphere superiority in language processing tasks is not reproduced for whistled speech, but Meyer et al. (2019) identify methodological issues in the experiments that may have downplayed any effect of left-lateralization. Like Carreiras et al. (2005), both studies relied on proficient whistlers.

Poss (2012) performed a phrase-identification study of an instrumental speech surrogate. The study's participants were Hmong speakers who were knowledgeable listeners (but not necessarily skilled practitioners) of the *raj* speech surrogate, an aerophone abridgment system based on Hmong tone and consonant types. The experiment tasked participants with translating *raj* speech, with stimuli drawn from common phrases in the repertoire. Participants were highly successful at identifying *raj* phrases. Moreover, "incorrect" translations frequently had similar tone patterns to the expected value, suggesting that listeners were sensitive to the surrogate system's tone encoding rules; no such pattern was identified with the system's encoded consonant types. This indicates that Hmong listeners take the surrogate language's grammar into account when processing individual words, but only selectively. The author also notes that "there is evidence that some subjects responded in terms of phrases rather than individual words" (Poss 2012).

Each of these studies focused on a very granular level of speech surrogate comprehension: the linguistic processing aspect. A more global analysis—one that can help to relate speech surrogates structures to their cultural and discursive context—calls for other methods. Particularly useful in the literature have been interviews with practitioners alongside analyses of corpora of speech surrogate performance. One substantial source is the work of Paul Neeley (1994), Neeley (1996), Neeley (1999), which describes slit-log drumming in Mekomba, an Ewondo-speaking community of Cameroon. Neeley (1999) provides a book-length analysis of a single genre of drummed Ewondo performance: calls to Christian worship as performed by the community's then catechist Antoine Owono over a 4-month period in 1988. The book breaks down the abridging system, which uses pitch and rhythm on the slit-log drum (*nkul*) to reproduce the underlying tonal phonology as well as (variably) the structure of consonant clusters in the spoken language (64). It also presents an interdisciplinary set of discourse, textual and rhetorical analyses. Neeley characterizes the Mekomba drumming tradition as a fixed, conventional, and unidirectional transaction between drummer and audience, the latter being all Mekomba residents within earshot of the signal. The catechist gave a twice-weekly performance to convince residents to attend Christian church services; the exact verbal content of the drummed messages was not widely understood by residents, but the general meaning of the performance was universally understood.

Works by Cloarec-Heiss (1986), Cloarec-Heiss (1999) and (Arom and Cloarec-Heiss, 1976; Arom, 2007) examine another slit-log drum system of Central Africa: the Banda-Linda *lenga* of the Central African Republic. As in the Ewondo system, the *lenga* is used to encode the tonal phonology and certain segmental features of the spoken language. The data come from Ippy, a small community in the eastern Central African Republic where "any Banda-Linda speaker can understand drum messages . . . [but] the actual social function of drum language restricts its use to a few emergency situations requiring one or more people to go to the place from which the message originates" (146–148). The authors present an information-theoretical account of the way that Banda-Linda surrogate speech is encoded by the drummer and decoded by the listener. They identify patterns in the drummed messages that make them easier to decode, including fixed phrases and a formula for message organization. They also point out that decoding is a retroactive process relying on the short-term memory to continuously recast the interpretation of the signal based on new stimuli.

Seifart et al. (2018) analyze a corpus of slit-log drum messages from a Bora community in the northwest Amazon. They demonstrate that the communication system encodes both the tonal phonology and speech rhythm of the Bora language. Drummed texts are shown to contain several kinds of "enphrasing", or conventionalized elaborations that make words and phrases more identifiable. Small distinctions in vowel-to-vowel timing intervals on the drum are shown to systematically correlate with speech rhythms; these distinctions are observed to be informative when decoding drummed messages. A formulaic structure for organizing messages is also argued to reduce ambiguity in the system.

Sicoli (2016) studies the pragmatics of whistled conversation in a Chinantec-speaking community of Oaxaca, Mexico. The whistling system reproduces the surface realization of Chinantec tone as well as glottal stops and stress patterns. Sicoli argues that "this yields a very productive and flexible morphophonological system for both the spoken and the whistled registers" (413). The system is shown to be "generative . . . making it possible to chat, conduct business, and make plans" (413). Sicoli gathered a corpus of 40 whistled conversations in the community of San Pedro Sochiapam consisting of both naturalistic conversations and the results of an experimental communication task. Practitioners in both settings were highly successful at communicating pertinent information using short, simple utterances. However, localized communication failures were attested, which conversation partners repaired using a limited set of standard questions and interjections. Sicoli argues that the whistled modality limits conversation to single-proposition utterances, and that the attested conversational repairs help regulate this constraint.

It should be inferred that experimental work on speech surrogate communication is rare for instrumental systems and wholly unavailable for lexical ideograph systems. Whistling is logically the first avenue for experiment methods, since some systems are very well described, and the practice itself is not cumbersome to record and analyze. However, corpus-based work points to intriguing similarities and differences between whistled and instrumental communication. Future experimental work should include instrumental systems. Furthermore, what continues to be revealed through neuroimaging and linguistic processing studies should be balanced with interview and corpus-based approaches that can place these findings into context.

CHALLENGES FOR SPEECH SURROGATE COMMUNICATION

Modern implements used to directly extend the range of the human voice, like the telephone, sound recordings, or the public address system, are designed to preserve as much of its acoustic detail as possible. Given the importance of the visual signal in language (both spoken and, of course, signed), even more faithful is the video broadcast. A comparison is often drawn between these implements and speech surrogates, which scholars have called "telegraphic instruments" (Church 1898; Verbeken (1922)), "drum-telephones" (Verbeken, 1922), "loudspeaker[s]" (Neeley 1999), "ancient text messages" (Villepastour 2010), and "musical newspapers" (Bebey 1999). Practitioners of some systems have drawn the same connection: "Nekgini speakers playfully liken their slit-gongs to a telephone system, and it is common to hear one person say to another, "ring me on a slit-gong"" (Leach 2002). A 1970s Solomons Islands newspaper, *The Solomons News Drum,* was so named in reference to the island's indigenous drum-signaling practices (Linton 2013). But the analogy is not total. The acoustic fidelity of a cellular phone, allowing it to transmit a close acoustic rendering of the human voice, is a technological innovation that postdates the origins of the known speech surrogate traditions.

In its absence, practitioners face a serious challenge. Natural languages are complex systems, endowed with detail, nuance, and (in principle) limitless possibility. The tasks we entrust to our natural linguistic faculties are not easily undertaken by artificial means. Any literate person knows how hard it can be to express oneself clearly in the written word, and to interpret the writing of others, from the literary journal to the instruction manual or the text message. That is in spite of the sophistication of writing itself, which has only been invented three or so times in human history (Daniels and Bright 1996) and for which even the basic principles require years of schooling to acquire. The system of English writing employs, at minimum, 26 distinct two-dimensional forms; the conventions of capitalization add another 26, and punctuation at least nine more. Speech surrogates are practiced in channels of much narrower bandwidth, some of the most common being slit-log drums, trumpets of ivory or bamboo, and wooden flutes that produce only a handful of distinct tones. Even the anatomical whistle, which is produced by the vocal articulators just as in verbal speech, is far simpler than the human voice timbrally and occupies a more limited frequency range (Meyer 2015). There are of course some differences between the orthographic and auditory reductions that compensate in the opposite direction—for one, writing lacks the ability to distinguish meaning through small gradient variations in rhythm, as can be found in certain speech surrogate systems. Nevertheless, producing a wide range of sounds suitable for the nuances of human language can be a tall order.

This, then, is the essential problem facing speech surrogate practitioners: the possibility for meaningful contrast is less in speech surrogates than in natural language.

As a result, speech surrogates cannot and do not account for all of natural language's complexities. This is especially clear in the phonological domain: as Stern (1957) explains: "an abridging system, while preserving some phonic resemblance to the base utterance, represents only part of its phonemic qualities" (125). For systems based on tonal languages, that usually means stripping away segmental features to primary reproduce aspects of tonal phonology [as in Kele drumming (Carrington 1949)]. A handful of segmental features can be encoded in addition to tone, such as vowel length and syllable structure (McPherson 2018b). Non-tonal languages may be reduced to other prosodic features such as pitch-accent (Caughley, 1974), or to segmental features like vowel formants and consonants (Rialland, 2005). Seifart et al. (2018) find for Bora drumming that gradient rhythmic contrasts reintroduce some phonetic detail alongside categorical ones, but the overall effect of segment loss is still large. Lexical-ideograph systems do away with phonology altogether, leaving only lexical units that themselves tend to be arranged in simpler syntactic forms than in verbal speech. Translating from natural language to speech surrogate inevitably diminishes the linguistic content of the signal.

This process has several consequences for speech surrogate communicability. An important one is the need for acquisition: as a rule, surrogate speech does not superficially resemble verbal speech, and therefore needs to be acquired in addition to it. This is obviously true for lexical-ideograph systems, for which the acoustic signal diverges freely from verbal signals. But it is also true in varying degrees for abridgement systems. The sound of a slit-log drum, which is common in abridging speech surrogate traditions of Central Africa and South America, consists of transients with sharp attacks and quick decays, generally lacking all of timbral variations produced by the oral cavity. Other instruments make closer approximations of speech sounds: the Yorùbá "talking drum", which permits the drummer to regulate gradient pitch, follows the contours of Yorùbá post-lexical tone (Akinbo, 2020). These are nevertheless lacking in the timbral and articulatory contrasts making up segmental phonetics. Whistling systems have an easier route to phonetic detail, since they are produced from the vocal tract. Even so, as Meyer (2015) describes:

> [whistling] is very different from the human voice both in its mechanism of production and in its acoustic form. A whistle consists of a simple narrowband melodic line modulated in frequency and amplitude . . . the voice shows a complex distribution in frequency. (73)

Abridging systems can also diverge from the surface realization of a base utterance, further masking sonic resemblance. The Sambla balafon surrogate system encodes the underlying tonal phonology of the language (Seenku) as well as aspects of syllable structure (McPherson 2018b). Post-lexical contour tone simplification is present in the spoken language but not emulated on the balafon. Characteristic intonational patterns of downdrift and declination are also omitted from the balafon surrogate system, which is limited to discrete pitches. The surrogate forms are therefore perceptually distinct from the realization of tone in spoken Seenku.

Similarly, Ewondo drumming ignores downstep in the spoken language, reducing contrastive pitches on the surface from three (high/low/downstepped high) to two (high/low) (Neeley 1999). James et al. (2021) note the same pattern in at least two other African speech surrogates: the Efik double bell and Luba slit-log drumming. Both systems, which make use of only two pitches, were originally analyzed to be simplifications of three- or four-toned languages (Simmons, 1955/1976; Burssens, 1936). Later, the languages were later analyzed as two tones and a downstep process not encoded in the speech surrogate (NKongola and Maddieson, 1973; Glewwe, 2019).

As a result, even phonology-based surrogate speech is not easily intuited from knowledge of the spoken language. To those unfamiliar with the system, speech surrogates sound like musical instruments being played or pitches being whistled. This is borne out in the neurological evidence that listening to whistled speech does not activate speech processing areas in the brains of naïve listeners (Carreiras et al., 2005). As I discuss in *Compensation Strategies in Speech Surrogate Communication*, this means surrogate speech is a skill that must be learned independently from fluency in the base language.

Another challenge is ambiguity. The loss of contrast means (for abridging systems) that distinct units in speech translate to homophonous sequences in the surrogate channel. In a speech surrogate based solely on tone, two words with the same tone melody are identical, even if they are distinguished by segmental contrasts in speech. This effect has been noted extensively (e.g., Carrington, 1976, 1949/1976; Herzog, 1945/1976; Simmons, 1955/1976; Stern, 1957) to significantly increase the number of homophonous items in the speech surrogate lexicon. Homophony is rarely a problem for communication in everyday speech, but it is also rarely so widespread as it is in surrogate speech.

A third challenge, less commonly discussed in the literature, is in processing. Surrogate speech comprehension is taxing on the cognitive faculties. Both abridging and lexical ideograph systems require listeners to interpret complex auditory signals under much more difficult circumstances than everyday speech. Diminished redundancy means a listener must make use of every detail of the signal in order to interpret the message. A speech surrogate signal may be produced beyond a listener's sightline, eliminating the visual and contextual clues that aid in the interpretation of natural language. Local ambiguities mean that listeners have to hold larger chunks of information in their short-term memories at once rather than interpreting a signal segment by segment (Cloarec-Heiss 1999); unlike writing, a speech surrogate message cannot be revisited and reinterpreted indefinitely. These factors are all compounded in systems that have an overall lower frequency of use than natural language. There are no attested examples of communities for which surrogate speech is the primary means of verbal communication. Participants such as women or children in areas are unwilling or forbidden to practice the system themselves, and are given no formal instruction—yet acquire fluency in the system nevertheless.

Given these challenges, it is tempting to say that the comprehension of surrogate speech is overreported, and that speech surrogacy is typically most meaningful as musical performance or traditional pastime rather than as language. To be sure, early scholarly accounts erred on the side of exaggeration in suggesting that speech surrogates were as expressive as verbal speech and could permit (fantastically) an unbroken line of communication from one end of a continent to another. As Goodwin (1937) observed: "the drum language of West Africa has been built up by careless journalism into one of the wonders of the world" (234). But there is yet evidence that speech surrogates around the world are commonly used for communication, from relaying news and sending invitations (Cloarec-Heiss 1999) to cracking jokes and swapping insults (e.g., Taaken Sàmàari, Ames, Gregersen, and Neugebauer 1971). As *Previous Work on Speech Surrogate Communication* shows, this evidence is increasingly being supplemented by experimental and neurolinguistic methods showing how that speech surrogacy manifests linguistically in the brain. We are therefore left with the question not of *whether* speech surrogates can be effective linguistic communication systems, but *how*.

COMPENSATION STRATEGIES IN SPEECH SURROGATE COMMUNICATION

In this section, I present a preliminary typology of compensation strategies used to overcome the surrogate modality's shortcomings. As we shall see, the pathway to speech surrogate communication runs through cultural and structural factors alike. It blurs the formal lines between abridgment and lexical ideograph systems. Hearkening back to Viglioccoet al. (2014), speech surrogates rely on cultural context (*Constructing a Discourse: Cultural conventions*) as well as linguistic form (*Frames, Formulas, and Enphrasing: Structural constraints*) in the construction of meaning.

Constructing a Discourse: Cultural conventions

Speech surrogates, abridging and lexical ideograph types alike, are bounded by conventions that cannot be derived from their base languages. As discussed in *Challenges for Speech Surrogate Communication*, a speaker of a language is not automatically qualified to practice a speech surrogate based on their language. That is not to say that any person cannot *develop* one—from my own experience, if a reader does not already whistle a (spoken) language natively, they can choose to begin at any time, and may be fluent in minutes. For that matter, any musician ought to be equally capable of developing a speech surrogate on their instrument. But this does not mean that they will be understood or understand the surrogate speech of others, including those who speak the same language. Instead, inclusion in the discursive ecosystem of a particular speech surrogate requires insider knowledge of how it is used (*Internalizing the Rules*) and what it is used for (*Topic Limitations*).

Internalizing the Rules

Surrogate speech varies from community to community, even for abridging systems based on the same language. (Clarke, 1934).

remarks for the Congo that "[s]trangers going into a new locality, although their spoken language may have only slight differences, do not as a rule understand the language of the drum" (418).

Neeley makes a similar observation about two drummers in neighboring Ewondo-speaking communities, where the drumming systems are based on the spoken Ewondo language:

> Though residing only a few kilometers apart, the two drumming catechists have idiosyncratic ways of drumming and of verbally interpreting the drum patterns. Each one is oriented towards a local speech community and will probably be only partially understood anywhere else. (Neeley 1999)

The same is evidently even true of whistled speech, despite its roots in spoken articulatory phonetics (Meyer, 2015). As Wilken (1979) observes in his study of the whistled Spanish of Oaxaca, Mexico:

> [D]ifferent villages have distinctive methods of whistling, in essence whistle dialects, even though all are whistling Spanish. Thus residents of Tetlatlahuca say that they can more or less follow the whistle speech of other pueblos in the município but can pick up only a word or two in a sentence whistled in Tepeyanco (884).

Likewise, (Classe, 1957), notes two distinct varieties of Silbo Gomero which are "not always mutually intelligible . . . one expert silbador informed me that it took him three or 4 days to become sufficiently familiar with the style of whistlers from other parts of the island to understand everything they whistled" (974).

Clearly, then, one can only become a member of a speech surrogate community if one learns its conventions—what applies for one community does not apply to another. While we still know little about speech surrogate acquisition, it seems that it can be learned by passive exposure or explicit pedagogy. The former is observed in vibrant traditions, especially whistling systems, which are mastered in childhood (Cowan, 1948; 1976; Hurley 1968; Stern 1957; Wilken 1979). In communities where whistling is commonplace, non-practitioners may develop working knowledge of whistling out of necessity (Meyer, 2015). However, even in such environments, comprehension and production do not develop apace with natural language; children understand whistled speech only several years after they begin to talk (134).

For most systems, anything more than basic knowledge requires conscious effort to acquire, even when input is abundant. The residents of Mekomba, an Ewondo-speaking community studied by Neeley (1999), heard biweekly calls to church by the catechist, an expert speech surrogate drummer. As a result, "[m]any people have the receptive competence to understand roughly any short phrase that is commonly drummed" (154), but "the verbal formulas [of the drum language] are not well understood by the majority of Mekomba residents" (164), and "few people understand all the intended words" (161).

For the Sambla balafon surrogate system, Strand (2009) says that time spent away from the village and individual interest dictate a Sambla listener's understanding. "Most Sambla can understand their name and at least a handful of common phrases" (224). However, at Sambla musical celebrations, only a "core group" of attendees "approach the *baan* [balafon] between songs and engage in instrumental-verbal exchange with the soloist" (234).

In the Reite communities of Papua New Guinea, who use a slit-log drumming system of lexical ideographs, a similar distribution is reported:

> Only a few men and women are skilled in using the full range of beat combinations which enable one to say such complex things as, "the whiteman will come to eat banana in [a particular] hamlet tomorrow afternoon, as long as there is no rain". Everyone, however, is able to hear their own name, and simple instructions (as in the favourite, "Hurry up!") (Leach 2002).

(Carrington, 1976) reports that Lokele children began to understand drummed Kele at "five or 6 years old" (620), but it is not clear how often that was true in practice by the 1940s. His 1943 survey of Lokele schoolboys found only 36 percent could reproduce their own drum names on the drum, indicating a "marked decrease in drum-signaling . . . among the Lokele people" 552) in that period. He suggests this was because "Lokele youths and boys are becoming less and less anxious to learn the drum language," partly due to the rise of literacy and telecommunication.

According to Poss, limited receptive competence in the Hmong *raj* tradition is widespread among Hmong-Americans raised in Thailand or Laos. "Many native speakers of Hmong who claim not to understand *tshuab raj* can pick out certain common expressions" (Poss 2005). The most knowledgeable listeners are "highly successful at interpreting musical messages even when they are taken out of context" (Poss 2012). This level of competence involves some casual pedagogy but is mostly attained through extensive listening: "The process of learning to play and understand words on the raj was informal. Relatives or friends might demonstrate a few phrases, but much learning took place through the observation of performances" (146).

Evidently, a roughly binary hierarchy can be found in communities of practice consisting of 1) skilled members who learn the conventions of the system as a whole, and 2) unskilled members who learn the most common messages in the system by heart. In Meyer (2015) typology of whistlers, these categories correspond to ""fluent whistlers". . . [who] have mastered the production and perception of whistling . . . [and] "canonical whistlers". . . [who] know set phrases understood by nearly everyone" (57). All systems must have unskilled members, since that population includes those who have yet to learn it. But the vitality of a speech surrogate is a function of their proportion in the community: as Meyer points out, "when the population is mostly composed of canonical whistlers, the whistled language is nearly dead" (57).

Skilled members' competence comes from mastering the conventions of the tradition. McPherson, (2018a), McPherso, (2019), McPherson, (2020) work shows that a skilled practitioner of the Sambla balafon tradition can easily produce novel elicited forms, translating systematically between spoken language and surrogate speech. This is quite different from how the system is used in practice, where it appears in brief, mostly predictable verbal exchanges during traditional gatherings, as well as in instrumental adaptations of Sambla vocal music (McPherson and James, in press).

My preliminary study with Benjamin Nimbatara, a skilled Birifor *gyil* practitioner, offers additional evidence. The Birifor *gyil* of Ghana's Upper West region has some commonalities with the Sambla balafon tradition. Both are found within Africa's "western xylophone belt" (Mensah, 1982). Both are traditionally played in group ensembles at important cultural gatherings like weddings and funerals (Vercelli 2006; Strand 2009). Both reproduce the tonal phonology of their respective spoken languages on gourd-resonator xylophones. However, the "talking mode" found in the Sambla system is absent among the Birifor. Birifor surrogate speech is limited to traditional texts adapted into music via fixed rules of tonal text-setting, similar to the "singing mode" of the Sambla balafon (Vercelli, 2006). I nevertheless found that Benjamin translated most elicited words onto the *gyil* easily. His competence with the Birifor surrogate tradition was extensible to an impromptu "talking mode" he had never used before. While more research is needed to know how these elicited forms differ from traditional ones, this points to the way that skilled practitioners acquire fluency in the conventions of their speech surrogate.

We can conclude that, to comprehend surrogate speech, members of a speech surrogate community need first to acquire an understanding of its form. For unskilled members, attention is paid to the surface form of messages, with an emphasis on the most frequent; skilled members gain a mastery over the system's underlying organizing principles.

Topic Limitations

Another strategy delivered at the cultural level is the construction of a conventional discourse wherein the same topics come up many times. This is an important tool for combatting the problem of ambiguity, since it allows listeners to use context to their advantage when interpreting a message. This strategy is expressed differently depending on the practice: certain implements are better suited to particular discourses.

For instance, loud implements used in community settings are less often used for private or sensitive information. Instead, they are "generally intended for use in community life as a means of stimulating or guiding social action or social behavior" (Nketia, 1971). This is a healthy restraint on the Universe of possible messages, since listeners can expect only those that have some relevance to them: announcements relating to local public life, information on prominent members of the community, and other topics on which they already have background knowledge. This is a general feature of public discourse in oral cultures (Ong 1977).

The arrangement is aided by the specialization of surrogate speech in a given community. Part of the problem of ambiguity is that long-distance communication strips away information about the identity and context of the interlocutors. If there are only a few practitioners, or if the context of use is regulated, the signals are more easily attributed to a source. An example of both conditions is the Ewondo drumming of Mekomba, Cameroon. During the time of Neeley's study, "a handful of older men [were] recognized as having extensive ability on the *nkul* as a speech surrogate; several others [had] limited ability" (41). Moreover, there was only one catechist. Calls to church consisted of a single genre performed by a single individual at a predetermined time. This made messages heard at that time easy to interpret:

> When Mekomba residents awaken to the sound of the *nkul* on Tuesday and Friday mornings, they quickly construct in their minds a contextual configuration. They assign to it a field, recognizing that a scheduled, public communication is being drummed at dawn. They assign to it a tenor, recognizing the personal relationships involved. They assign to it a mode, recognizing specific meaning and broad intention communicated by the speech surrogate formulas. Through this mental construction of the contextual configuration, the audience knows what action is expected of them. (164)

Whistled speech and other quieter speech surrogates are not usually constrained by public discourse in the same way that louder implements are. But the topics are, probably to a greater degree, bounded by their immediate context of use. As Cowan (1948) describes for Mazatec whistling:

> In spite of the high probability of ambiguity, the actual instances where confusion occurs are amazingly few. This is due to the fact that whistling is most frequently (though not necessarily) concerned with topics immediately obvious to both parties to the conversation, and used in situations where cultural context plays a much greater part than in the spoken language. (1,390)

The context of use itself is mediated through social convention. Whistling is a "specialized version of a language used for specific purposes in particular social circumstances" (Sicoli 2016). More specifically, whistling systems are usually localized to outdoor environments and to the occupations of hunting and pastoralism (Sicoli, 2016), so topics will trend towards these areas. In Chepang whistling, for example:

> There is considerably more ambiguity in whistled communication than in the spoken equivalent. But the very strong limitations on cultural context means that most of these ambiguities can be resolved. In fact, not only is whistle speech limited to certain situations, those of animal and bird catching, but within these it appears to be used only for relatively few, more essential, communications, particularly those relating to movements of the prey. (Caughley, 1974)

Whistling and quieter instruments are more amenable to private or secretive conversation, since they have a limited range of transmission and may conceal meaning from casual or unskilled listeners. While the Universe of topics included in "private conversation" is limitless, in speech surrogates it often means something more specific: courtship. This theme is common to speech surrogates across Southeast Asia and the Americas, especially whistling (Hurley 1968), woodwind instruments (Hmong *raj,* Catlin 1982; Gavião *kotiráp*, Moore and Meyer 2014), and jaw harps (Proschan 1994; Pugh-Kitingan and Jacqueline, 1982).

The prominence of this theme is unsurprising. Social theory tells us that (along with public discourse) courtship is among the most conventionalized social practices; participants choose to enact or negotiate sexual scripts which "encourage[] the conservative, highly ritualized, or stereotyped character that sexual behavior often takes" (Simon and Gagnon 1986). Speech surrogates aid in the enactment of sexual scripts, since they distinguish courtship interactions from normative conversation (Catlin, 1982). But it can equally be said that they are aided by the script: messages limited to the careful constructions of a romantic exchange will tend to towards being frequent and predictable—and therefore more effective implements for communication.

Frames, Formulas, and Enphrasing: Structural constraints

The strategies described thus have been established on the cultural level: members of a speech surrogate community understand speech surrogate messages by gaining knowledge about the system's properties (*Internalizing the Rules*) and its uses (*Topic Limitations*). In this section, I examine the structure of individual messages in speech surrogates.

I find these strategies tend to operate in one or both of the following domains: roteness and elaboration. Some put constraints on the Universe of possible content, making speech surrogate messages more predictable and therefore easier to produce and interpret. I classify these as roteness-oriented strategies. Moreover, some systems regulate the complexity of messages, generating messages that are longer and more elaborate than the base language to eliminate ambiguities, or limiting their length to make them easier to parse. I classify these as elaborateness-oriented strategies. Several strategies operate in both domains.

Framing

What I call "framing" is a strategy for organizing messages in a surrogate system. It encourages messages to be structured in predictable, fixed forms. As such, it is roteness-oriented. However, in some cases, a framing strategy mandates additional elements—essentially discourse markers—that segment the flow of new content, making messages less ambiguous. This use is elaborateness-oriented.

Framing is widely attested in speech surrogates, displaying several cross-linguistic commonalities. Messages in many systems conventionally include an opening signal (Coulter, 2007), a closing signal (Sicoli, 2016), or both (Heepe, 1920; Carrington, 1944; Arom and Cloarec-Heiss, 1976; Carrington, 1944; Goethem 1976; Heepe, 1920; Neeley 1999). In discourse analysis, these signals correspond to the "aperture" and the "finis", which demarcate the termini of a discourse (Longacre 1996). Within these two brackets, the structure of a message may be bracketed further, with definite "slots" for the names of the addressed party (Rialland 2005; Burridge, 1959), the source of the signal (Burridge, 1959), or markers that identify a particular kind of message content (Seifart et al., 2018). Carrington (1976) provides a representative example of some of these types in Kele drumming:

> A drummer usually begins with the call: *ki*, *kɛ*, repeated two or three times. Then follows the name of the person or persons to whom he wishes to "speak". His business follows. He concludes by drumming out the name of the person for whom the call is made, and then a series of beats on the low note terminates the communication. (546)

Seifart et al. (2018) identify a similar pattern for the Bora *manguaré* system. *Manguaré* messages begin with a choice of two sequences corresponding to the message "type": *íkʲòòkáré tsà-ʔíhkʲà* "Come now!", or *íkʲòòkáré tsíβà-ʔíhkʲà* "Bring now!". The "type" is followed by the name of the "addressee", itself divided into several components associated with clan identifications. The "message content" follows, and the message then terminates with the "end" sequence (Seifart et al., 2018).

Lexical-ideograph systems use framing in much the same way. For example, in the non-phonoplogical Alamblak *nrwit* (slit-log drum) system of Papua New Guinea, messages begin with an aperture: an initial, indefinite striking of the drum that "alerts people within hearing range that someone is about to say something on the *nrwit*" (Coulter 2007). Then, the drummer plays the signal corresponding to the place of the message's intended recipient. That name itself constitutes a frame, comprising a signal "to inform the hearers that a place name will follow immediately" (Coulter 2007), a signal for the local clan's totem, and a specific regional identifier. The drummer then plays the addressee's name, also composed of distinct identifiers. Finally, the actual message content is given; once the drummed discourse is over, the *nrwit* player may add an optional "coda" to mark its completion (97).

In these cases, framing essentially imposes a "surrogate syntax": a set of rules more restrictive, and more predictable, than the base language. With regard to whistling systems, then, an important distinction must be made. Unlike many instrumental speech surrogates, whistled languages tend to adhere closely to the vocabulary and syntax of everyday speech. In fact, on these grounds Meyer (2008) asserts that "[c]ontrary to a "language surrogate", whistled speech does not create a substitute for language with its own rules of syntax" (70).

However, there are some instances where whistling systems do deviate from the syntax of the base language using framing. For example, Chintantec whistling employs an utterance-final particle *réi*[13] as a finis, working "similarly to "over" in radio operator talk" (Sicoli 2016). *Réi*[13] is not found in standard Chinantec speech contexts. Similarly, according to Rialland (2005), Silbo Gomero uses the vowel *a* as an aperture:

> The vowel *a* plays a special role in Silbo in that it provides a reference point for whistlers, who usually begin their messages by whistling a followed by the name of the addressee: *a Bernardo*, *a Maria*, *a Sebastian*, *a Domingo.* (5)

Moreover, though whistling systems adhere closer to a language's syntax, they appear to make use of *less* of it. Sicoli (2016) shows that Chinantec whistling displays only a few of the forms of conversational repair evident in the spoken language, selecting against those that produce syntactically complex forms. He argues that whistled speech in conversation is usually limited to short, semantically simplex utterances, disallowing "multiple actions and complex embeddings" in a single turn (427). This echoes Cowan (1948) observation for Mazatec whistling that "single utterances tend to be short" (5). This trend towards simplicity may even extend to morphology: Wilken (1979) notes that the whistled Spanish of Tlaxcala is mostly limited to the present tense (883).

This, too, is an elaborateness-oriented strategy: ruling out complex utterances and stripping away redundant linguistic features regulates the elaborateness of an utterance downward. While this likely makes whistled speech easier to produce and process, further research is required to assess the precise effects of discourse markers and pragmatic constraints in speech surrogate communication, particularly in whistled speech.

Formulas

In surrogate speech, formulas are groups of words combined into fixed, conventional phrases. To a large degree, formulas are characteristic of speech surrogate texts cross-linguistically. They have been attested for a great number of the documented abridging systems, particular the speech surrogates of Central Africa (Hulstaert 1935; Carrington, 1949; Dugast, 1955; Neeley 1999) and of West Africa (Herzog, 1945; Nketia, 1958; Ames et al., 1971), as well as in Hmong surrogate speech (Falk, 2004; Poss, 2005) and others. The arbitrary signals of lexical-ideograph systems are also essentially formulaic. With the evident exception of the Diola (Moreau, 1997) and Abuʔ-Wam (Nekitel, 1985) whistled languages, which apparently combine phonological encoding with more arbitrary conventional signaling, formulas are much less characteristic of whistling systems.

This strategy operates on the message level, crystallizing words, phrases and sentences so that they can be produced and processed in whole form. This makes it roteness-oriented, since it constrains the form a message can take. Since much of value has been written on this feature of surrogate speech, I refer the reader to Neeley (1999) and Ong (1977) for detailed analysis of their general properties; in *Formulas*, I discuss a variant of the formula in depth. This section is limited to addressing a point raised in those works that must be interpreted with caution.

Neeley (1999) and Ong (1977) each point to the oral-formulaic theory of Milman Parry and Albert Lord as a natural analogy for studying speech surrogate texts. Oral-formulaic theory, first popularized in Lord's work *The Singer of Tales* (1960), has had a profound impact on the study of oral poetry and related fields. Lord advanced a cross-cultural analysis that identified the formula as the defining feature of epic poetry. The comparison is compelling: formulas are the building blocks from which epic poetry is formed, making up a repertoire that defines a whole tradition. Likewise, for many speech surrogates, formulas are the atoms of meaning from which all discourse is assembled. However, this analogy only goes so far, at least without abandoning some of Parry and Lord's thesis.

In oral poetry, the formula is a "group of words which is regularly employed under the same metrical conditions to express a given essential idea" (Albert, 1960). They are constitutive of the texts of epic poetry in large part because they are the key to producing novel verses at great length: copious spontaneity is effortful, and improvisation is made more challenging by the constraints of poetic meter. Thus, rote phrases with the right metrical properties are relied upon to ease the poet's creative burden over hundreds or thousands of lines.

Lord's framing of this theory privileges the effort of the performer. For an epic poet in performance, the most important question is: what will I say next? The decision is made with immense pressure on the cognitive faculties, since it requires the working memory to fit every new development into the procession of a narrative which might last hours or days. Formulas are cognition-economizing devices, allowing a performer to readily develop verse under challenging conditions. Evidence from psycholinguistics confirms that formulas are effective tools for maintaining speech fluency when memory resources are tight (Kuiper, 2004).

We are currently without the benefit of psycholinguistic studies on surrogate speech production; we do not yet know the precise mental pathways practitioners take when they produce messages. Therefore, we cannot be sure if the performance context of speech surrogates is (always) comparably taxing to that of epic oral poetry. In my view, there is reason to think it is not.

One reason is that speech surrogate messages do not always share the epic poem's discursive context: long-form solo performance. Some speech surrogate traditions, such as the *qeej* funeral performances of the Hmong people (Falk, 2004), do approach the length and narrative involvement of the Serbo-Croatian epic poetry studied by Lord, in which individual songs numbered in the hundreds or even thousands of lines. But speech surrogate messages can also be short and simple, even while making use of formulaic language—the "speech mode" of Bora drumming in the Amazon, for instance, employs formulaic constructions in messages that amount to only a few phrases (Seifart et al., 2018). Moreover, as Neeley (1999) argues for Ewondo drumming, speech surrogate performance is not always subject to metrical constraints. To be sure, speech surrogates used in musical contexts often adhere to established metrical patterns (Nketia, 1963) or even reproduce interactions of meter and lyric from vocal music (as in the "sung mode" of the

Sambla balafon; McPherson and James forthcoming). But other abridging speech surrogates can be based entirely on speech rhythms, as determined by categorical factors like syllable structure (Neeley, 1999) in addition to gradient factors like vowel-to-vowel syllable timings (Seifart et al., 2018). In these cases, the rhythmic content of a message is determined purely by the choice of semantic content, not the reverse.

Surrogate speech may well require more cognitive effort to produce than everyday speech, given its unusual linguistic properties and its greater reliance on motor control to produce sounds. But it is not clear that this is what drives the prominence of the formula in surrogate speech, as Lord describes in the case of the epic poem. Lord writes: "the repeated phrases were useful not, as some have supposed, merely to the audience if at all, but also and even more to the singer in the rapid composition of his tale" (30). On the contrary, for the speech surrogate, the former seems more likely: formulas are of great benefit to the listener, as they do the essential work of disambiguating messages.

Enphrasing

Enphrasing is a special case of the formula that deserves special attention. The term is a coinage from Stern (1957), who defines it simply: "the lexical unit is replaced by a phrase" (127). Like other terms found in that paper, its definition has been (aptly) a bit elaborated by subsequent authors. A representative rewording can be found in Seifart et al. (2018): ""enphrasing", i.e., elaborating words and sentences to make them longer and less ambiguous" (2). Enphrasing is a strategy that operates on the roteness and the elaborateness of a message at once.

Variations on this description are attested throughout the literature of Central and West African drumming systems from early sources onwards. For example, Heepe (1920/1976) observed that, while some units in the phonological Ewondo drumming system were literal transpositions of a single word, others were figurative circumlocutions: for instance, Ewondo *awü* "death" corresponded to the drummed *Abüo (~abo) äsi äfgm* "He lies very quietly in the Earth" (333). Herzog ([1945] 1976) writes that the phonological Jabo drumming in Liberia consists largely of "periphrastic formula[s]" (555). Long lists of enphrasing texts are available for Kele drumming (Carrington, 1949, Carrington, 1976) and Beti drumming (Hulstaert, 1935). Outside of Central and West Africa, the phenomenon has been more recently described in Southeast Asia (Bradley, 1979; Poss, 2005) and the Amazon (Seifart et al., 2018).

Like other formulas, enphrasing is understood to be a strategy for targeting the ambiguity problem in abridging systems: "short words that would come out as homophones in drumming are replaced by longer, less ambiguous expressions" (Seifart et al., 2018). Carrington (1949) provides several examples to this effect in the phonological surrogate texts of Kele drumming. For instance, the word *songe* "moon" is conventionally replaced with a drummed sequence encoding *songe li tange la manga* "the Moon looks down on the earth" (Carrington, 1949). Carrington explains that *songe* (a word with two high-toned syllables) would be represented directly on the drum as a sequence of two high-pitched strikes; without enphrasing, it would be indistinguishable from *koko* "little bird", also represented by a sequence of two high notes. However, *songe li tange la manga* is tonally distinct from the enphrasing given to *koko*: *koko olongo la bokiokio* "the fowl, the little one which says *"kiokio"*", so their drummed representations are likewise distinct. This example shows that enphrasing is oriented to both roteness and elaborateness: the strategy transforms basic words into fixed, elaborate sequences to help the listener interpret otherwise ambiguous messages.

In its response to functional pressures, enphrasing makes the formal division between abridging and lexical ideograph systems less distinct. Just as framing creates its own independent "surrogate syntax", enphrasing systems constitute a "surrogate lexicon" which refers only conventionally to the base language. When Kele drummers want to refer to the Moon, and more specifically, to the word *songe*, they must check the word against the speech surrogate lexicon to find its equivalent. They then drum that phrase, following the rules of the abridging system. The resulting drum sequence matches only the phrase in the surrogate lexicon in terms of phonology—relative to the base utterance, it is arbitrary. While Carrington's *songe* and *koko* enphrasings both begin with the words themselves, that is not a requirement: a Kele enphrasing for *lotika* "orphan" is enphrased as "*wana ati la saŋgo la nyaŋgo* "the child has no father or mother"" (541). The surrogate phrase is therefore abstracted from the sound of the original word, even while retaining a connection to phonological structure. This echoes more typical lexical ideograph systems like the Tangu *garamut* signal-drumming of Papua New Guinea, where signals are acoustically abstracted from their referents, though they might still have roots in "linguistic [or] quasi-poetic rhythms" (Burridge, 1959).

Lord says that for the epic poet, "the formula means its essential idea" (65). In other words, an elaborated expression denotes only its most generic meaning; "The "drunken tavern" means "tavern"" (65). This applies equally to enphrasing. Regardless of what the enphrasing means in a literal sense, in the speech surrogate it denotes whatever word it is coindexed with in the base language.

Alternative interpretations for enphrasings are ruled out by fiat. The *songe*/*koko* pair provides a neat example: since *songe* and *koko* share the same tonal pattern, the first word of each enphrasing should sound the same. Therefore, the former could just as well be *koko li tange la manga* "the little bird looks down on the earth". Neither the phonological structure nor semantic content rules out the alternate interpretation on the drum, since a flighted bird may look down on the earth, as can a Moon. It is easy to see how other such semantic ambiguities could pose a problem when common referents have similar properties, such as animal species. If "rhinoceros" and "elephant" were homophonous in the surrogate language, it would be prudent to avoid an enphrasing meaning "the big gray creature", since the two animals could appear in interchangeable contexts. I can find no evidence that confusions of this sort ever happen in practice. With few attested exceptions, the surface form of an enphrasing is ascribed one meaning.

As a result, an enphrasing need not "make sense" in the base language. The meanings of the words in an enphrasing may change and fade. Carrington observes that "[a] second characteristic of the gong-phrases is that of the tendancy [sic] to use derogatory or diminutive words" (1949, 4), even for generic objects that are neither subpar nor small. For example, "the Kele gong-word for « fishing-net » is: *biléme yâwɛ́ŋgo*, which is translatable as: « a little bit of an old fishing net »" (4); similar constructions are attested in other drum languages of the Western/Central African group (Heepe, 1920; Hulstaert, 1935; Jacobs, 1954). Likewise, in Ewondo drumming "[a] few nouns are expanded within Antoine's performance paradigm with diminutive and pejorative descriptions: *man ntu eyie* "small old cloth"... *man etug nkul* "small, old, broken drum" ... *otutu dza* "poor little village"" (Neeley, 1999). Bora drumming has a similar pattern, using pejorative markers for nouns, as well as a repetitive morpheme for verbs. According to Seifart et al. (2018), "Bora speakers have no intuitions why elements with the literal meanings "deceased", "repeated" and "damaged" should be used in *manguaré* messages (12). They argue that these constructions "do not carry any semantic value, but function purely to identify the preceding sequences of beats as representing nouns or verbs" (12). Interestingly, this echoes the evidence that diminutive markers aid in word segmentation during spoken language acquisition (Kempe et al., 2007).

By the same token, an enphrasing need not be a phrase used in the spoken language. Carrington (1949) notes that enphrasings in various Bantu drum languages of the Congo often take the form of bare noun-noun compounds, often synonyms: a Kele drum phrase meaning "news" encodes "*mbóli sango* ... two words for « news »" (50), the Mbɔle drum phrase meaning "bird" encodes "*tofulú átɔnɔli* ... two words for « little birds »" (52), and the Olombo drum phrase meaning "oil" encodes "*sókó mainá* ... two words for « palm oil »" (53). Hulstaert (1935) also reports concatenations of two fish species names to refer to fish generally, and three antelope species names as the general term for animals (660) in Nkundo drumming.

To be sure, this kind of construction is not uncommon in spoken language. Co-hyponyms, constructions which name a category by compounding a few of its constituents, are attested cross-linguistically, as are synonym compounds (Wälchli, 2005). Though most common in the languages of East Asia (Wälchli 2005), Examples in English include "sol-fa" (Renner, 2008), and "subject matter". Spoken Hmong is rich with co-hypernyms and synonym compounds, particularly among the "elaborate expressions" of its formal repertoire: Mortensen (2004) reports *lab* "monkey" + *cuam* "gibbon" → *lab-cuam* "simian" and *ncauj* "mouth" + *lu* "mouth" → *ncauj-lu* "mouth", the latter "only used in flowery or ritual speech" (4). Unsurprisingly, these constructions appear in the Hmong surrogate languages. Poss (2005) lists elaborate compounds like *leej niam leej txi* "mother/father", used to "create characteristic melodic contours that skilled listeners recognize immediately [and] can then be employed formulaically" (129).

But no such elaborate spoken repertoire is attested in Carrington and Hulstaert's accounts. This is unsurprising, since such forms would be highly unusual for the language family. In Bantu languages NN compounding is usually unproductive and restricted to a few ancestral forms (Basciano et al., 2011). We can see, therefore, that enphrasings can include forms disallowed by a practitioner's own synchronic grammar.

In spoken language, lexical ambiguities are usually resolved through context and pragmatic cues. Still, in certain contexts, redundant elaborations have the same function in natural languages as in speech surrogates: making an utterance less ambiguous. In English, we sometimes use synonymous collocations when two senses of a word are easily confused. An everyday example is the pair "funny-haha/funny-peculiar" (Timothy Pulju p.c.), but this usage is most common in technical contexts that demand precision of language: "sanction approval" in regulatory documents and "handicap (dis)advantage" in golf are both examples where the second word is synonymous with one of the two ambiguous meanings of the first. Synonymous binomials like "cease and desist" or "last will and testament" are also prevalent in English legal language, where they were historically used to forestall alternative interpretations of medieval English and Latin or French synonyms (Crystal 2005) and now help to clarify a word's standard meaning in specific areas of law (Mellinkoff 2004).

Ong (1977) provides another example of the parallel processes between natural and surrogate language: Middle Chinese, an exception to the rarity of pernicious homophony in language. It is generally agreed that the ~1,500 years between Old Chinese and present-day Mandarin Chinese saw a phonological simplification that vastly reduced the number of contrasting syllables. Since the Old Chinese lexicon was largely made up of monosyllabic words, this meant that most syllables acquired new homophones. Moreover, as an isolating language, morphology could not usually differentiate homophones in context. As Sampson (2013) argues, this produced widespread ambiguity in the spoken language (though not the writing system), to the extent that "Classical Chinese read aloud cannot be understood, without sight of the text, even by scholars who are very familiar with that language" (8).

Consequently, something had to change to keep the language intelligible. What resulted was a large-scale shift from monosyllabic words to disyllabic words through compounding, especially synonym compounding. For instance: Mandarin *péngyǒu* "friend" derives from Old Chinese forms of *péng* "friend" and *yǒu* "friend"; Sampson notes that "*péng* is seven ways ambiguous as a morpheme of Mandarin and *yǒu* is three ways ambiguous, but the compound *péngyǒu* is unambiguous" (Sampson 2013, 9). This process overhauled the lexicon, pushing the rate of disyllabicity from roughly 20% in Old Chinese to over 80% in Modern Chinese (Shi 2002, 70–72). As a result, utterances in the language are longer and (diachronically) more redundant, but less ambiguous—just as in enphrased surrogate speech.

We can conclude that enphrasing lexicons are partially independent from spoken language, though they respond to functional pressures in much the same way. This independence makes it more difficult to draw the distinction between enphrasing systems and systems of lexical ideographs. Like Kele drumming, systems like the signal drumming systems of Oceania have a surrogate lexicon that runs parallel to the base language. Regarding the Tangu slit-log drumming of Papua New Guinea, Burridge (1959) writes:

> A child grows up knowing the call-signs of individuals, pigs, and dogs, and the phrases for localities and events, in the same way as he comes to know names and

> becomes aware of social situations. Standard signals and the events they refer to are associated in his mind as "whistle and train", "chime and clock", "hooter and factory", "church bell and prayers" are associated in the mind of a European. Special signals come into his understanding as do family jokes, sporting metaphor, and the rest. (1,238–9)

In the same way, practitioners of enphrasing systems learn to associate words of their base language with their lengthier, fixed equivalents. The orientation towards roteness and elaboration, used to render surrogate speech less ambiguous, supersedes the formal distinctions between the systems.

DISCUSSION

As Vigliocco et al. (2014) remind us, the study of linguistic structure cannot be disentangled from the structures that surround it—those of modality, culture, and context. In the case of speech surrogates, this holds a dual meaning. Speech surrogates are part of the linguistic ecosystem of the world; to fully understand language's "broader context of use", we must account for the surrogate modality. We cannot be surprised that the communicative pressures that shape speech apply their force in equal (or indeed greater) measure to the surrogate modality. At the same time, speech surrogates are a communication system of their own. To analyze speech surrogates in their full form, we need to account for their cultural context and structure in the same breath.

Shore (1991) writes: "[a] theory of meaning construction should account for the historical and local variability of conventional meaning systems" (11), and indeed it seems that speech surrogates are subject to the culture-specific negotiations that govern systems of meaning. What characterizes one speech surrogate does not apply to all, since such negotiations may well proceed differently from place to place. But more research is required to assess how exactly those negotiations take place. What accounts for the distribution and variety of speech surrogate structures around the world? It does not seem conducive to an explanation from innate factors, or strictly as a function of linguistic structure. The most common features—tone-based abridgement, elimination of consonant contrasts and so on—are nevertheless not universals. Their productive mechanisms, structures, and contexts of use all vary from culture to culture. If, say, tone is uniquely coded for speech surrogacy in the human language faculty, why are the surrogate systems of Papua New Guinea largely not based on tonal (or other) phonology, despite the presence of tonal languages in the region? And since whistling is a physically simple, effective means of communication, and is found in tonal and non-tonal languages alike, why do so few cultures have whistled speech? Meyer (2015) argument, sensibly, is that the practice is only developed by people for whom it is especially useful. Thus, as it stands now: the evidence suggests that speech surrogates have a conventional component that must not be ignored.

Evidence for the profound effects of functional pressures on speech surrogates ought to provoke reflection on all language use. If free, sophisticated communication were trivial in the surrogate modality, it would pose a challenge to functional explanations for language structure: why should language be so complex if its duties can equally be performed by a conch shell? Such a finding would lend credence to the theories of language origin by genetic accident or co-evolution. Instead, there is a wealth of evidence that speech surrogate communication is difficult, and that practitioners need to significantly adapt their communicative behavior to compensate for the challenge. Speech surrogates show how our linguistic faculties are molded by the circumstances of their use.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: ODSS: speechsurrogates.org/map.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

Dartmouth College—Presidential Scholars Program provided $2000 stipend during the course of this research.

REFERENCES

Akinbo, S. (2020). *Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis*. Vancouver, BC: PsyArXiv. doi:10.31234/osf.io/43gf6

Albert, Bates. Lord. (1960). *2000. The Singer of Tales*. Cambridge, Mass: Harvard University Press. Available at: http://archive.org/details/singeroftales00lord_0.

Ames, D. W., Gregersen, E. A., and Neugebauer., T. (1971). Taaken Sàmàarii: a Drum Language of Hausa Youth. *Africa* 41 (1), 12–31. doi:10.2307/1159675

Armstrong, Robert. G. (1953). "46. TALKING INSTRUMENTS IN WEST AFRICA," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. S. Thomas and J. Umiker-Sebeok (The Hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-017.xml.

Arom, Simha., and Cloarec-Heiss, France. (1976). Le Langage Tambouriné Des Banda-Linda (RCA) in Théories et Méthodes En Linguistique. Communications Au 11 Congrès de La SLAO. Yaoundé, Avril 1974. *Bibliothèque de La SELAF Paris* 54–55, 113–169.

Arom, Simha. (2007). Language and Music in Fusion: The Drum Language of the Banda Linda (Central African Republic). *Trans. Revista Transcultural de Música* 11, 0.

Bagemihl, Bruce. (1988). *Alternate Phonologies and Morphologies*. Vancouver, BC: University of British Columbia. doi:10.14288/1.0097954

Basciano, B., Kula, N., Melloni, C., and Meyer, J. (2011). Modes of Compounding in Bantu, Romance and Chinese. *Rivista di linguistica* 23(2), 203–249.

Bebey, Francis. (1999). *African Music: A People's Art*. Chicago, IL: Chicago Review Press.

Blench, R. (1987). Idoma Musical Instruments. *Amj* 6 (4), 42–52. doi:10.21504/amj.v6i4.1260

Bradley, D. (1979). Speech through Music: The Sino-Tibetan Gourd Reed-Organ. *Bull. Sch. Oriental Afr. Stud.* 42 (3), 535–540. doi:10.1017/S0041977X00135773

Burridge, Kenelm. O. L. (1959). *1976. 60. THE SLIT-GONG IN TANGU, NEW GUINEA*. The Hague: De Gruyter Mouton. Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-031.xml.

Burssens, A. (1936). "LE LUBA, LANGUE A INTONATION, ET LE TAMBOUR-SIGNAL," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The Hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804416/10.1515/9783110804416.474.xml/

Busnel, René-Guy. (1970). Recherches Experimentales Sur La Langue Sifflee de Kuskoy (Experimental Research on the Whistling Language of Kuskoy). *Revue de Phonetique Appliquee* 14-15, 70.

Busnel, R. G., Moles, A., and Gilbert, M. (1962). "UN CAS DE LANGUE SIFFLÉE UTILISÉE DANS LES PYRÉNÉES FRANÇAISES," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The Hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-018.xml.

Carreiras, M., Lopez, J., Rivero, F., and Corina, D. (2005). Neural Processing of a Whistled Language. *Nature* 433 (7021), 31–32. doi:10.1038/433031a

Carrington, John. (1949). *A Comparative Study of Some Central African Gong-Languages*. Bruxelles: Falk, G. van Campenhout, successeur./z-wcorg/.

Carrington, John. (1976). "Talking Drums of Africa," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The Hague: De Gruyter Mouton). doi:10.1515/9783110804423-001

Carrington, John. (1944). "The Drum Language of the Lokele Tribe," in *Speech Surrogates: Drum and Whistle System*. Editors A. Thomas and J. Umiker-Sebeok (The Hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-001.xml.

Catlin, Amy. (1982). "Speech Surrogate Systems of the Hmong: From Singing Voices to Talking Reeds," in *The Hmong in the West: Observations and Reports*, 170–197.

Caughley, Ross. Charles. (1974). "Chepang Whistle Talk.," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). doi:10.1515/9783110804423-026

Church, G. E. (1898). Notes on the Visit of Dr. Bach to the Catuquinaru Indians of Amazonas. *Geographical J.* 12 (1), 63. doi:10.2307/1774499

Clarke, Roger. T. (1934). "20. THE DRUM LANGUAGE OF THE TUMBA PEOPLE," in *Speech Surrogates: Drum and Whistled Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804416/10.1515/9783110804416.418.xml.

Classe, A. (1957). "Speech Surrogates: Drum and Whistle Systems," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-023.xml.

Cloarec-Heiss, France. (1986). *Dynamique et Équilibre d'une Syntaxe: Le Banda-Linda de Centrafrique. Descriptions de Langues et Monographies Ethnolinguistiques 2*. New York, NY, USAParis, France: Cambridge University Press; Éditions de la Maison des sciences de l'homme.

Cloarec-Heiss, France. (1999). From Natural Language to Drum Language: An Economical Encoding Procedure in Banda-Linda, Central African Republic. *Lang. Divers. Cogn. Representations*, 145–157. doi:10.1075/hcp.3.13clo

Coulter, Neil. R. (2007). "Music Shift: Evaluating the Vitality and Viability of Music Styles Among the Alamblak of Papua New Guinea,". PhD Thesis (Kent State University).

Coupé, Christophe., Yoon, Mi., Dediu, Dan., and Pellegrino, François. (2019). Different Languages, Similar Encoding Efficiency: Comparable Information Rates across the Human Communicative Niche. *Sci. Adv.* 5, eaaw2594. doi:10.1126/sciadv.aaw2594Oh

Cowan, George. M. (1948). "Mazateco Whistle Speech," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-038.xml.

Crystal, David. (2005). *The Stories of English*. Abrams.

Damm, Robert. J. (2003). Do You Udu? *Percussive Notes* 29–33. December 2003.

Daniels, Peter. T., and Bright., William. (1996). *The World's Writing Systems*. New York: Oxford University Press. Available at: http://archive.org/details/isbn_9780195079937.

Dugast, Idelette. (1955). "36. LE LANGAGE TAMBOURINÉ OU SIFFLÉ CHEZ LES BANEN," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-007.xml.

Dundes, Alan. (1980). *Interpreting Folklore*. Indiana University Press.

Falk, Catherine. (2004). Hmong Instructions to the Dead: What the Mouth Organ Qeej Says (Part One). *Asian Folklore Stud.* 63, 1–29.

Glewwe, Eleanor. (2019). "Efik Nominal Tonal Alternations as Phrasal Morphology," in *African Linguistics across the Disciplines* (Berlin: Language Science Press), 71–88. doi:10.5281/ZENODO.3520571

Goethem, L. Van. (1976). *18. LOKOLE of TAM-TAM BIJ DE NKUNDO-NEGERS. Part 1*. The Hague: De Gruyter Mouton. Available at: https://www.degruyter.com/view/book/9783110804416/10.1515/9783110804416.399.xml.

González, Villar., Pamela, Onur. Güntürkün., and Ocklenburg, Sebastian. (2020). Lateralization of Auditory Processing of Silbo Gomero. *Symmetry* 12 (7), 1183. doi:10.3390/sym12071183

Goodwin, Astley. John. Hilary. (1937). *Communication Has Been Established*. Methuen.

Gourlay, K. A. (1982). Long Trumpets of Northern Nigeria - in History and Today. *Amj* 6 (2), 48–72. doi:10.21504/amj.v6i2.1116

Güntürkün, O., Güntürkün, M., and Hahn, C. (2015). Whistled Turkish Alters Language Asymmetries. *Curr. Biol.* 25 (16), R706–R708. doi:10.1016/j.cub.2015.06.067

Heepe, M. (1920). "13. DIE TROMMELSPRACHE DER JAUNDE IN KAMERUN," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804416/10.1515/9783110804416.316.xml.

Herzog, George. (1945). "DRUM-SIGNALING IN A WEST AFRICAN TRIBE," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804416/10.1515/9783110804416.553.xml.

Hulstaert, P. Gustaaf. (1935). De Telefoon Der Nkundo (Belgisch Kongo). *Anthropos* 30 (5/6), 655–668.

Hurley, W. M. (1968). The Kickapoo Whistle System: A Speech Surrogate. *Plains Anthropologist* 13 (41), 242–247. doi:10.1080/2052546.1968.11908474

Jacobs, J. (1954). "34. SIGNAALTROMMELTAAL BIJ DE TETELA," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-005.xml.

James, Lucas., McPherson, Laura., and Rafkin, Emma. 2021. "Online Database of Speech Surrogates (ODSS)." Database. SpeechSurrogates.Org. 2021. Available at: speechsurrogates.org/map.doi:10.1017/9781108767484

Kaminski, Joseph. S. (2008). Surrogate Speech of the Asante Ivory Trumpeters of Ghana. *Yearb. Traditional Music* 40, 117–135.

Kempe, V., Brooks, P. J., Gillis, S., and Samson, G. (2007). Diminutives Facilitate Word Segmentation in Natural Speech: Cross-Linguistic Evidence. *Mem. Cogn.* 35 (4), 762–773. doi:10.3758/bf03193313

Kuiper, Koenraad. (2004). Formulaic Performance in Conventionalised Varieties of Speech. *Formulaic Sequences: Acquisition, Process. Use*, 37–54. doi:10.1075/lllt.9.04kui

Leach, J. (2002). Drum and Voice: Aesthetics and Social Process on the Rai Coast of Papua New Guinea. *J. R. Anthropological Inst.* 8 (4), 713–734. doi:10.1111/1467-9655.00130

Lewis, Tony. (2018). *Becoming a Garamut Player in Baluan, Papua New Guinea: Musical Analysis as a Pathway to Learning*. New York, NY: Routledge. doi:10.4324/9781315406503

Linton, Sean. (2013). "The Music of 'Are'are: Acoustemology, Environmental Influences and Sustainability,". Thesis (Dunedin, NZ: University of Otago). Available at: https://ourarchive.otago.ac.nz/handle/10523/4368.

Longacre, Robert. E. (1996). *The Grammar of Discourse*. Springer Science & Business Media.

McPherson, Laura. (2020). *A Grammar of Seenku*. Berlin: De Gruyter Mouton.

McPherson, Laura., and James, Lucas. "Artistic Adaptation of Seenku Tone: Musical Surrogates vs. Vocal Music," in Selected Proceedings of the 50th Annual Conference on African Linguistics, Vancouver, BC (Berlin: Language Science Press). In press.

McPherson, Laura. (2019). The Role of Music in Documenting Phonological Grammar: Two Case Studies from West Africa. *Proc. Annu. Meetings Phonology* 7 (0), 1–14. doi:10.3765/amp.v7i0.4493

McPherson, Laura. (2018a). The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. *Anthropological Linguistics* 60 (3), 255–294.

McPherson, Laura. (2018b). The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. *Anthropological Linguistics* 60 (3), 255–294.

Mellinkoff, David. (2004). *The Language of the Law*. Eugene, OR: Wipf and Stock Publishers.

Mensah, Atta. Annan. (1982). ' Gyil': The Dagara-Lobi Xylophone. *J. Afr. Stud.* 9 (3), 139.

Meyer, Julien., Dentel, Laure., Gerber, Silvain., and Ridouane, Rachid. (2019). "A Perceptual Study of CV Syllables in Both Spoken and Whistled Speech: A Tashlhiyt Berber Perspective," in 20th Annual Conference of the International Speech Communication Association (Interspeech 2019). doi:10.21437/interspeech.2019-2251

Meyer, Julien. (2008). Typology and Acoustic Strategies of Whistled Languages: Phonetic Comparison and Perceptual Cues of Whistled Vowels. *J. Int. Phonetic Assoc.* 38, 69–94. doi:10.1017/S0025100308003277

Meyer, Julien. (2015). *Whistled Languages: A Worldwide Inquiry on Human Whistled Speech*. Springer.

Moles, A. A. (1970). Etude Socio-Linguistique de La Langue Sifflee de Kuskoy (Sociolinguistic Study of the Whistling Language of Kuskoy). *Revue de Phonetique Appliquee* 14-15 (77-118), 70.

Moreau, Marie-Louise. (1997). La Communication Sifflée Chez Les Diola (Casamance, Sénégal). *DiversCité Langues* 2, Available at: https://www.planete-senegal.com/NOREF/communicationsiffleediola.html.

Mortensen, David. 2004. "Hmong Elaborate Expressions Are Coordinate Compounds,". doi:10.1057/9780230512207 Available at: http://www.davidmortensen.org/papers/elaborate_expressions.pdf/

Moore, D., and Meyer, J. (2014). The study of tone and related phenomena in an Amazonian tone language: Gavião of Rondônia. *Language Documentation* 8, 24.

Neeley, Paul. (1994). Discourse Peak and Poetic Closure in the Final Stanza of a Talking Drum Performance. *The J. West Afr. Languages* 24, 108.

Neeley, Paul. (1996). Drummed Transactions: Calling the Church in Cameroon. *Anthropological Linguistics* 38, 36.

Neeley, Paul. (1999). *People of the Drum of God–Come!*. Dallas, TX: Summer Inst of Linguistics.

Nekitel, O. I. M. S. 1985. *Sociolinguistic Aspects of Abu', a Papuan Lanquage of the Sepik Area*. doi:10.25911/5d723dfcf2f5a

Nketia, J. H. K. (1958). "41. DRUM PROVERBS," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-012.xml.

Nketia, J. H. K. (1963). "42. Drumming in Akan Communities," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-013.xml.

Nketia, J. H. K. (1971). "45. SURROGATE LANGUAGES OF AFRICA," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-016.xml.

NKongola, Muyumba., and Maddieson, Ian. (1973). *Luba Tonology and the Development of Downstep*. New York.

Ong, W. J. (1977). African Talking Drums and Oral Noetics. *New Literary Hist.* 8 (3), 411. doi:10.2307/468293

Poss, Nicholas. Frederick. (2012). "Hmong Music and Language Cognition: An Interdisciplinary Investigation,". PhD Thesis (Columbus, OH: Ohio State University).

Poss, Nicholas. Frederick. (2005). "The Communication of Verbal Content on the Hmong Raj: An Ethnographic Analysis of Performance Practice,". PhD Thesis (Columbus, OH: The Ohio State University).

Proschan, Frank. (1994). Khmu Play Languages. *Mon-Khmer Stud.* 23, 43–65.

Pugh-Kitingan, J. (1982). Language Communication and Instrumental Music in Papua New Guinea; Comments on Tse Huli and Samberigi Cases. *Musicology Aust.* 7 (1), 104–119. doi:10.1080/08145857.1982.10415938

Renner, V. (2008). On the Semantics of English Coordinate Compounds. *English Stud.* 89 (5), 606–613. doi:10.1080/00138380802011982

Rialland, A. (2005). Phonological and Phonetic Aspects of Whistled Languages. *Phonology* 22 (2), 237–271. doi:10.1017/S0952675705000552

Sampson, G. (2013). A Counterexample to Homophony Avoidance. *Dia* 30 (January), 579–591. doi:10.1075/dia.30.4.05sam

Sampson, G. (2015). "Writing Systems: Methods for Recording Language," in *The Routledge Handbook of Linguistics*. Editor Keith. Allan (Abingdon, United Kingdom: Routledge).

Sebeok, Thomas. A., and Umiker-Sebeok, Jean. (1976). *Speech Surrogates: Drum and Whistle Systems*. The hague: De Gruyter Mouton.

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing Language to Rhythm: Amazonian Bora Drummed Language Exploits Speech Rhythm for Long-Distance Communication. *R. Soc. Open Sci.* 5 (4), 170354. doi:10.1098/rsos.170354

Shi, Yuzhi. (2002). *Establishment of Modern Chinese Grammar: The Formation of the Resultative Construction and its Effects*. Amsterdam, NETHERLANDS: John Benjamins Publishing Company. Available at: http://ebookcentral.proquest.com/lib/dartmouth-ebooks/detail.action?docID=739946.

Shore, B. (1991). Twice-Born, once Conceived: Meaning Construction and Cultural Cognition. *Am. Anthropologist* 93 (1), 9–27. doi:10.1525/aa.1991.93.1.02a00010

Sicoli, M. A. (2016). Repair Organization in Chinantec Whistled Speech. *Language* 92, 411–432. doi:10.1353/LAN.2016.0028

Simmons, Donald. C. (1955/1976). "EFIK IRON GONGS AND GONG SIGNALS," in *Speech Surrogates: Drum and Whistle Systems*. Editors A. Thomas and J. Umiker-Sebeok (The hague: De Gruyter Mouton). Available at: https://www.degruyter.com/view/book/9783110804423/10.1515/9783110804423-008.xml.

Simon, W., and Gagnon, J. H. (1986). Sexual Scripts: Permanence and Change. *Arch. Sex. Behav.* 15 (2), 97–120. doi:10.1007/bf01542219

Stern, T. (1957). Drum and Whistle "Languages": An Analysis of Speech Surrogates. *Am. Anthropologist* 59 (3), 487–506. doi:10.1525/aa.1957.59.3.02a00070

Stone, Ruth. M. (1972). *Music of the Kpelle of Liberia*. Ethnic Folkways. Smithsonian Folkways. Available at: https://folkways.si.edu/music-of-the-kpelle-of-liberia/world/album/smithsonian.

Strand, Julie. Lynn. 2009. "The Sambla Xylophone: Tradition and Identity in Burkina Faso."

Verbeken, A. (1922) "LE TAMBOUR-TÉLÉPHONE CHEZ LES INDIGÉNES DE L'AFRIQUE CENTRALE." In, edited by A. Thomas and J. Umiker-Sebeok. Available at: https://www.degruyter.com/view/book/9783110804416/10.1515/9783110804416.336.xml.

Vercelli, Michael. (2006). "Ritual Communication through Percussion: Identity and Grief Governed by Birifor Gyil Music,". DMA Diss (Tucson, AZ: University of Arizona).

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a Multimodal Phenomenon: Implications for Language Learning, Processing and Evolution. *Phil. Trans. R. Soc. B* 369 (1651), 20130292. doi:10.1098/rstb.2013.0292

Villepastour, Amanda. (2010). *Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code*. Farnham, United Kingdom: Ashgate Publishing, Ltd.

Wälchli, Bernhard. (2005). *Co-Compounds and Natural Coordination*. Oxford University Press. doi:10.1093/acprof:oso/9780199276219.001.0001

Wilken, Gene. C. (1979). Whistle Speech in Tlaxcala (Mexico). *Anthropos, No. H.* 74, 881–888.

Winter, Y. (2014). On the Grammar of a Senegalese Drum Language. *Language* 90 (3), 644–668. doi:10.1353/lan.2014.0061

Wojtylak, Katarzyna. (2016). "Traversing Language Barriers. Murui Signal Drums from Northwest Amazonia," in Special Workshop 'Creativity in language: Secret codes and special styles, LCRC, JCU, Australia, August, 2016. doi:10.13140/RG.2.2.35413.22240

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

PERSPECTIVE
published: 11 August 2021
doi: 10.3389/fpsyg.2021.667599

Singing the Individual: Name Tunes in Oyda and Yopno

Azeb Amha[1]*, *James Slotta*[2] *and Hannah S. Sarvasy*[3]

[1] African Studies Centre Leiden, Leiden University, Leiden, Netherlands, [2] Department of Anthropology, University of Texas at Austin, Austin, TX, United States, [3] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, NSW, Australia

Music beats spoken language in identifying individuals uniquely in two disparate communities. In addition to their given names, which conform to the conventions of their languages, speakers of the Oyda (Omotic; SW Ethiopia) and Yopno (Finisterre-Huon; NE Papua New Guinea) languages have "name tunes," short 1–4 s melodies that can be sung or whistled to hail or to identify for other purposes. Linguistic given names, for both communities, are often non-unique: people may be named after ancestors or contemporaries, or bear given names common to multiple individuals. But for both communities, name tunes are generally non-compositional and unique to individuals. This means that each new generation is likely to bring thousands of new name tunes into existence. In both communities, name tunes are produced in a range of contexts, from quotidian summoning and mid-range communication, to ceremonial occasions. In their use of melodies to directly represent individual people, the Oyda and Yopno name tune systems differ from surrogate speech systems elsewhere that either: (a) mimic linguistic forms, or (b) use music to represent a relatively small set of messages. Also, unlike some other musical surrogate speech traditions, the Oyda and Yopno name tune systems continue to be used productively, despite societal changes that have led to declining use in some domains.

Keywords: Oyda, Yopno, surrogate speech, name tune, *konggap*, *moyzé*, music, whistled language

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Don Niles,
Institute of Papua New Guinea Studies, Papua New Guinea
Gerrit Bloothooft,
Utrecht University, Netherlands

****Correspondence:***
Azeb Amha
a.amha@asc.leidenuniv.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *13 February 2021*
Accepted: *15 July 2021*
Published: *11 August 2021*

Citation:
Amha A, Slotta J and Sarvasy HS (2021) Singing the Individual: Name Tunes in Oyda and Yopno. Front. Psychol. 12:667599. doi: 10.3389/fpsyg.2021.667599

INTRODUCTION

Philosophers have long distinguished proper names from common nouns on the basis of their semantics; in contrast to common nouns, which denote classes of things, proper names are said to denote individuals (Searle, 1958; Kripke, 1980). In this study, we describe musical practices from different sides of the world that share this property with names. Among the Oyda of Ethiopia and the Yopno of Papua New Guinea, a distinct musical tune is closely associated with each individual. In fact, these tunes are more uniquely associated with individuals than proper names are. In these communities, multiple individuals may share the same proper name, by virtue of being named after them, or by the name's popularity. What we term "name tunes," in contrast, are unique to individuals, at least normatively.

Speech surrogate systems are often divided into two main types: abridgment and ideographic (Stern, 1957). In abridgment systems, musical features correlate with features of the affiliated language: pitch with tone or vowel quality, or duration with vowel length. This is typical of "whistled languages," such as those corresponding to Spanish, Turkish, and Highland Mazateco (Cowan, 1948; Classe, 1957; Meyer, 2021), but abridgment systems in which musical instruments serve as proxies for speech are also attested (Niles, 2010). In ideographic systems, musical phrases are conventionally associated with words or expressions, bearing meaning without a phonic connection to the language. These are typified by some of the slit-gong drum messaging traditions of northern New Guinea (Burridge, 1959; Niles, 2010), and the "talking drums" of West Africa (Neeley, 1999; Winter, 2014).

The Oyda and Yopno name tune systems differ from both abridgment and ideographic systems in that the name tunes directly denote an individual, without recourse to language. That is, Oyda and Yopno name tunes reference individual human community members themselves; they are not "abridgments" or "ideographs" of people's given names. While there are minor stylistic differences between Oyda and Yopno name tunes, the two traditions are remarkably similar, and both are absent from most surrounding communities. This paper offers a first comparative analysis of the usage contexts and formal characteristics of Oyda and Yopno name tunes.

COMPARING OYDA AND YOPNO NAME TUNE TRADITIONS

Oyda Name Tunes: *moyzé*

The Oyda language of southern Ethiopia (about 45,000 speakers) belongs to the Ometo branch of Omotic. The Oyda region is dominated by mountains and valleys which are partly shaped by the Great Rift Valley system.

Most Oyda people have both a proper name and a *moyzé* "name tune."[1] Both men and women, as well as children from age five or six onward, use *moyzé*. While a proper name can simultaneously be used to designate different people, two or more living people cannot share the same *moyzé*. As such, *moyzé* are a more reliable personal identifier than a proper name.

Moyzé are typically whistled by blowing air between two fingers of a hand held against the mouth while the airstream is modified by the movement of fingers of the other hand (Amha in preparation). Some people blow air at a joint between their index-finger and middle finger, some between the middle finger and their ring-finger, still others between their ring-finger and little finger. This appears to have impact on the audibility of the whistle: air blown into the juncture between the index finger and the ring finger seems to have higher pitch, and is said by the Oyda to "reach farther," than that between the index finger and ring finger. Another difference among people involves the hand someone habitually uses to blow air into. Some prefer to use the left hand for this while using the right hand for modulating air; others prefer the reverse. People claim that they can tell who is whistling their *moyzé* name based solely on the sound. According to one consultant: "The whistling of different people differs as their fingerprint would differ."

An individual's *moyzé* is known and used by people close to him/her and is also introduced to new acquaintances. Such "teaching of *moyzé*" is done by singing the *moyzé* using consonant-vowel (CV) combinations, such as *léeteléetetóom*, for someone whose proper name is S'as'ima. Such sung "vocables" are not used to address people but are crucial for transmission of *moyzé* forms. Sung vocables can include any of the five vowels of the language: /i/, /u/, /e/, /o/, and /a/. Both short and long forms are attested. Consonants tend to be /m/, /n/, /l/, /w/, /h/, and /t/, but a few *moyzé* containing /b/ and /g/ are recorded, e.g., *tóógirgidáalíiim*.

Moyzé are most often used for communication at a distance, but also serve to invoke the memory of a deceased person. *Moyzé* are often used to summon or get the attention of family members at work in different places, to alert the addressee of some danger or emergency, and to greet someone when passing by his/her house. Until hunting was outlawed in the 1970s, *moyzé* were also widely used to assemble people for a hunting party. Among Oyda who still practice the traditional religion, *moyzé* play an important role in announcing death to family members in different villages and during funerals, where the *moyzé* name of the deceased is repeatedly called using a side-blown, trumpet-like instrument, also called *moyzé*, that is made from cattle horn. Indeed, it is possible that the term originally derived from *moys-* 'to see off, accompany someone for part of the journey,' and that the original function of Oyda name tunes was funerary. With the expansion of Protestant and Pentecostal faiths among the Oyda, the use of *moyzé* in funeral contexts is eroding, but its use in daily communication has been maintained. In some cases, young people may be innovating new arenas in which to use *moyzé*; one young man uses his deceased father's *moyzé* as a mobile ring tone.

Moyzé may be given (e.g., by a parent to a child, a husband to his wife), may be self-created, or inherited. Adults can choose to keep an inherited *moyzé* of a deceased family member alongside their own *moyzé,* thereby having more than one *moyzé* for some time. Either their own *moyzé* or the inherited one can then be passed on to one of their own descendants.

Yopno Name Tunes: *konggap*

The Yopno language (Reed, 2000; Slotta, 2014) is spoken by some 8,000 people living in the Yopno valley in the northeast of Papua New Guinea. Yopno is a Papuan language of the Finisterre branch of the Finisterre-Huon language family. Yopno is most closely related to the Nankina and Domung languages to its north and west. Speakers of these three languages (total population: 15,000) all use name tunes, known in Yopno as *konggap* (Nankina: *kunggwap*). *Konggap* have been discussed by Niles (1992), Slotta (2012), and Ammann et al. (2013).

Konggap are most often sung using the vowels /a/, /o/, and /e/. Sung in this way, *konggap* can be heard at great distances across the deep gorges and along the steep mountain slopes found throughout the region. This is also the form in which *konggap* are performed at funerals and in ceremonial dances. But on a daily basis in the confines of homes, villages, and gardens, *konggap* may also be whistled, hummed or sung in the middle of spoken conversation.

Virtually everyone in the region has one of these melodies uniquely associated with him or her. The *konggap* are sung, whistled, and hummed throughout the day as people summon others, alert them to their presence, or even think about them. The Yopno language is non-tonal and *konggap* bear no phonic relation to people's proper names. Rather, each is a melodic

[1] https://hdl.handle.net/1839/d63213ce-eab2-41cd-ae9a-ff3f4a8b1f40

TABLE 1 | Comparative summary, *moyzé*, and *konggap* traditions.

	moyzé	*konggap*
Location	Oyda district, Gofa zone of the Southern Nations, Nationalities, and Peoples Regional in Ethiopia	Nayudos LLG, Madang Province and the neighboring Yopno valley in Morobe province, Papua New Guinea
Environment	Parallel rolling hills, elevation between 1500 and 2600 m	Steep-sloped mountains and side valleys cut by streams flowing into the deep gorge shaped by the Yopno river, cloud rainforest and grassland, elevation between 1000 and 3000 m
Taboos on use of linguistic given names	Not strictly taboo. But often people are addressed using teknonyms, titles ("chief") and relational terms ("sister," "brother") even when no kinship relation exists	Avoid the names of in-laws. Older relatives and members of the community are typically not addressed by name
Acquisition	Given by parents, self-created or inherited: used by age five or six	Mothers innovate name tunes to their infants during infancy Later, individuals can change their own name tunes
Whistling production method	Air blown into hand and modulated by fingers of other hand	Whistled without hands
Sung method	CV sequences; C = tends to be /m/, /n/, /l/, /w/, /h/, /t/, but /b/, /g/ are also used. All five vowels of the language, /i/, /u/, /e/, /o/ and /a/ as well as their long counterparts are used	V sequences only: primarily /e/, /a/, /o/
Quotidian use	Hailing, summoning, alerting, recalling deceased	Hailing, summoning, alerting, recalling deceased
Special occasion use	In funerals, only through the use of a blown horn	Ceremonial dance, funerals
Taboos	Two living people cannot have similar-sounding *moyzé* names	Aside from ceremonial dances, people are reluctant to sing their own *konggap*
More than one per person?	Yes, possible: some may inherit a *moyzé* while they have their own	Yes, possible: an individual may be associated with different *konggap* in different places

and rhythmic sequence that identifies its bearer uniquely in a way that proper names do not, a point Yopno people routinely mentioned to the second author when discussing *konggap*. Traditional Yopno proper names are typically shared with living namesakes or ancestral forebears. Today, these names are often supplemented with names from English or Kâte, the former lingua franca of the Lutheran church in the region. But these too are drawn from a limited stock, so it is common for people to share a name with numerous others. In contrast, *konggap* are described as ideally unique to their bearer.

When still a baby, a child receives its first *konggap*, often composed by the mother while tending to the child. That *konggap* is later replaced by a new melody composed by a friend, relative, or often by the bearer him- or herself. The *konggap* of some is given by ancestors in dreams. Over their lives, people may cycle through multiple *konggap*, and even have more than one in use at the same time.

On a day-to-day basis, *konggap* are perhaps most often used to hail and summon, but, as with the funerary *moyzé*, they can also bespeak emotional associations with the people whose *konggap* are sung. Passing by a relative's house on the way to the forest, a man will sing his relative's *konggap* to alert him to his own passage. Seeing a friend at a distance along the steep mountain slopes in the region, a person will sing the friend's *konggap* as a greeting or summons. When one's thoughts turn to a person who has died or who one has not seen for a while, one expresses sorrow by singing or whistling the relative's *konggap*. Joy at the arrival of a friend is expressed with *konggap*. Shown a photo of the local preschool class by the second author, a group of children in Nian village burst into a mass of song—each singing out the *konggap* of a friend they recognized in the photo.

As with *moyzé, konggap* play a role in announcing death. Here, however, care must be taken, because singing a deceased person's *konggap* can also be taken as a sign that the singer is responsible for their death. At a funeral, where people gather around the body of the deceased for one or more days, women collectively sing the *konggap* of the deceased and of the relatives and friends of the deceased, sometimes cycling through dozens of *konggap*. The name tunes often elicit tears in the listening mourners.

Konggap are also sung in men's ceremonial dances (Niles, 1992; Ammann et al., 2013) to mark other moments of social importance—a new marriage, the conception of a first child, or an official event. A group of men sing their own individual *konggap* simultaneously, synchronizing the beginnings and endings of their very different *konggap* with a common beat. This is one of the few times a person sings his own *konggap*, which otherwise rings of great pretension. Indeed, an individual performer's *konggap* is said to potentially take on magical potency to seduce women in such performances.

Table 1 compares key features of the *konggap* with those of the *moyzé.*

Formal Qualities

Figure 1 shows pitch traces for a sample of 10 sung *konggap* (left) and 10 whistled *moyzé* productions (right), analyzed using Praat (Boersma and Weenink, 2019). Time (in seconds) is charted on

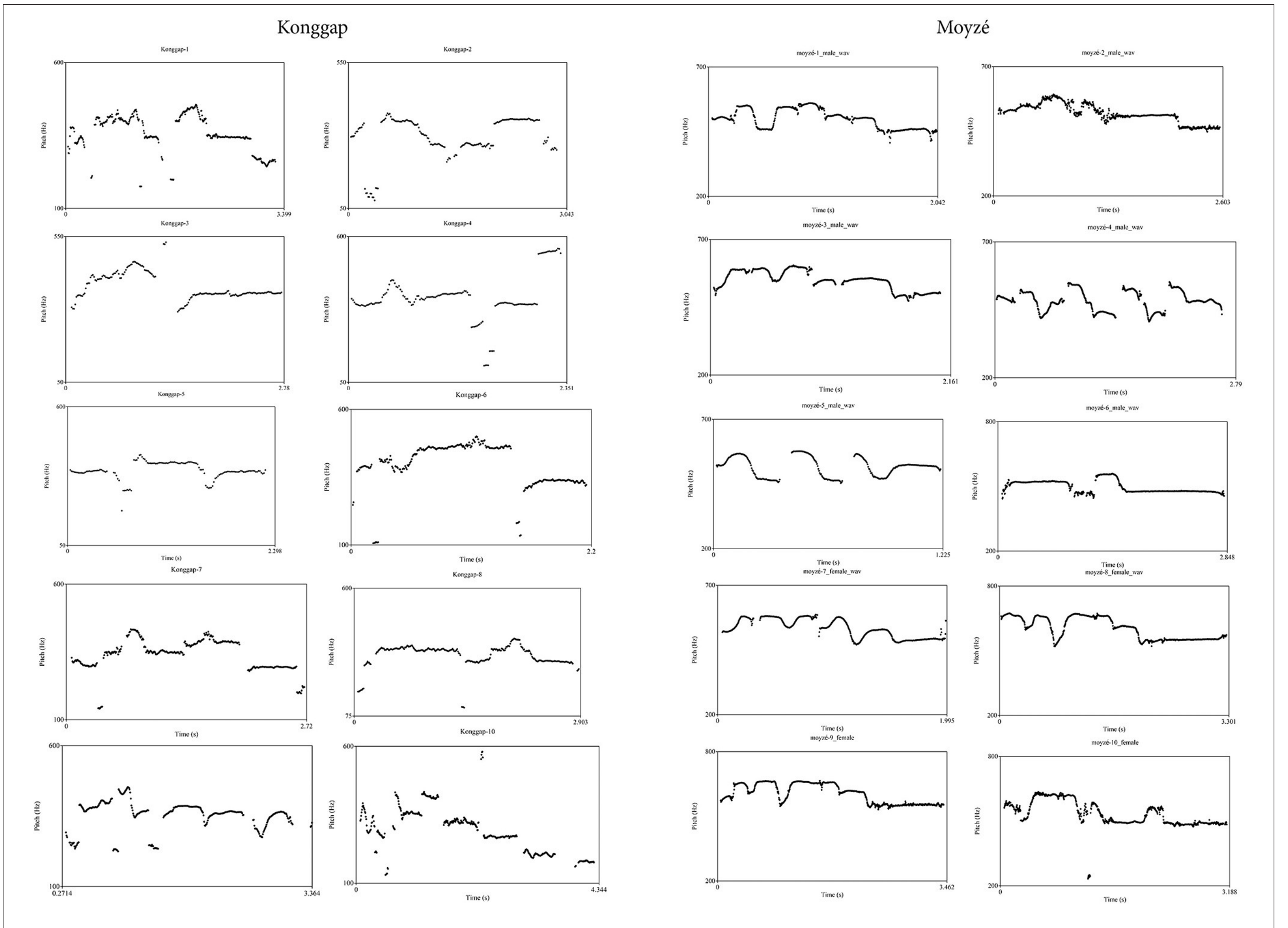


FIGURE 1 | Pitch traces for 10 sung *konggap* and 10 whistled *moyzé* name tunes.

the x-axes, and pitch range (Hz.) is charted on the y-axes. The *konggap* were recorded using a Sony MZ-R55 minidisc recorder at the funeral for a deceased woman. In the recordings, a group of women sing the *konggap* of relatives of the deceased, using the vowels /a/, /e/, and /o/. The *moyzé* were recorded using a Linear PCM LS-11 audio recorder at 44.1 kHz/16bit. Six were performed by men and four by women.

Figure 1 shows that the name tunes of both traditions often end with relatively lower pitch than the rest of the tune, and the final "note" is often sustained. These pitch contours are not directly reminiscent of the prosodic contours of any linguistic units in either language. Due to the unstaged group singing and noisy, uncontrolled recording context of the *konggap*, brief, spurious pitch trace fragments appear, primarily in the lower frequencies. These were not used for measuring pitch ranges, below.

Figure 2 visualizes tune duration, pitch range, and note durations of the productions in **Figure 1**, along with sung renditions of the same *moyzé*. **Figure 2A** compares durations of the *konggap* with those of both sung and whistled *moyzé*. The sung *konggap* and whistled *moyzé* have similar average durations; the sung *moyzé* are shorter than all *konggap*, and most whistled *moyzé*. **Figure 2B** gives the results of subtracting the least pitch value from the greatest pitch value in each tune, based on actual sung notes, not any spurious pitch readings by Praat that appear in **Figure 1**. The *konggap* and sung *moyzé* are very similar here, while whistled renditions of the *moyzé* feature smaller pitch ranges, overall. Note that the graph depicts the difference between minimum and maximum pitch in each name tune, not the actual pitch values.

Figure 2C aims to quantify the amount of rhythmic variation in each name tune. We segmented all "notes" in *konggap* and sung *moyzé* samples in Praat. A note was taken to be a sung vocoid or CV sequence with perceptually stable pitch; this was obtained by listening, along with visual inspection of the spectrogram. The number of notes per tune ranged from 4 to 14 (mean: 8.7; median: 8). Overall, there is more variation in note duration within individual *konggap* samples than within individual *moyzé* samples. Several *moyzé* are characterized by relatively even and short note durations (generally <0.4 s in duration), while all *konggap* include mixtures of short notes (lasting 0.2 s or less), and long notes (lasting 0.4 s or more).

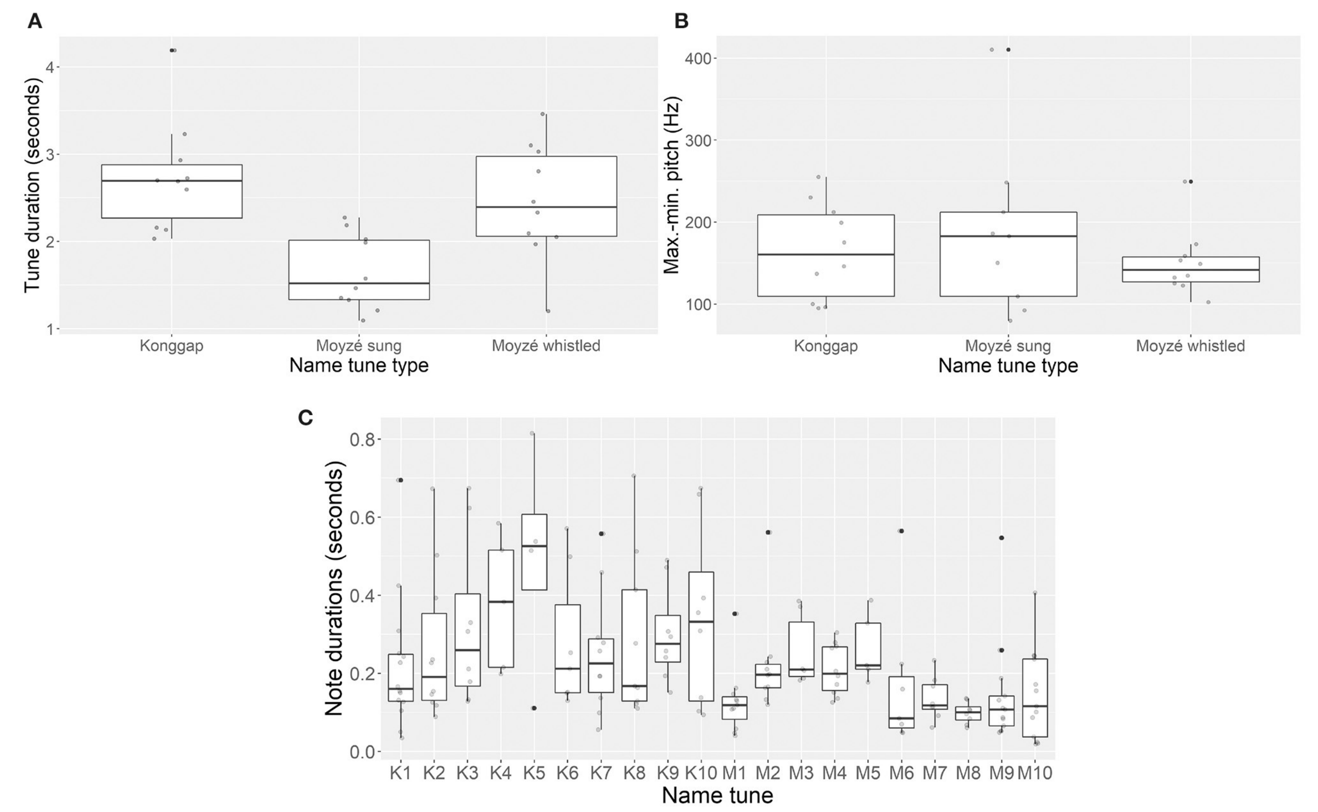


FIGURE 2 | Tune duration, pitch range, and average note duration for sampled *konggap* and *moyzé* (*K* = *konggap*, *M* = *moyzé*).

In sum, *konggap* and *moyzé* are formally very similar. They last for similar amounts of time (1.09–4.19 s) and feature similar pitch ranges. *Moyzé* vary more in length than *konggap*, and *konggap* tend to have more internal rhythmic variation than the *moyzé.* Durations and pitch contours of both *konggap* and *moyzé* suggest that, if they should be compared with any linguistic units, this should be the utterance or clause (not the word, or proper name). Future research should compare the forms of *konggap* and *moyzé* to Yopno and Oyda musical genres: we expect that musical styles may reveal more than language about the origins of *konggap* and *moyzé* stylistic conventions.

DISCUSSION

This has been the first comparative analysis of two name tune traditions. We have seen that in the Oyda and Yopno name tune traditions, a short musical sequence references an individual community member. These name tunes are maximally specific and individual; more so than proper names within each language, which can be identical in both communities to either ancestral figures' names or to living people's names. It appears that each new generation induces the generation of thousands of new *moyzé* and *konggap* name tunes. These name tune systems thus exemplify the human capacity for musical creativity, as well as for recognition and reproduction of musical sequences.

The Oyda *moyzé* and Yopno *konggap* are similar to each other in length and pitch range per tune, but diverge slightly in internal rhythmic variation (greater in the *konggap*). In both traditions, name tunes play an important role in funerary proceedings, evoking the essence of the deceased individual in a medium that could be said to speak directly to the emotions. The Yopno name tune tradition differs from the Oyda tradition in that *konggap* are also used in men's group song/dance performances.

Musical evocation of individuals is found beyond these two systems. The *yoik* of the Sami are short songs or chants that can be not only linked to individual people, but also to activities, animals, emotions, and nature (Anderson, 1984; Krumhansl et al., 2000: p. 19; Hanssen, 2011). At least two of the northern New Guinea communities with slit-gong drum messaging traditions (the Tangu and Reite) are also reported to have used particular drum sequences for each individual person—and sometimes individual domestic animals—in a community (Burridge, 1959; Leach, 2002).

In contrast, the Mehek, a Sepik area New Guinea community (Hatfield, 2016) have a small, apparently fixed inventory of *isi* "name whistles," which, crucially, are each associated with a given name, not with individual people. Individuals who share a given name also share an *isi*. Some *isi* also map onto multiple given names; only 94 *isi* are attested. The tunes themselves are generally shorter and less complex than either *konggap* or *moyzé.* Mehek further has a few

TABLE 2 | Comparison of Oyda and Yopno name tune systems with other surrogate speech systems.

		Formal relationship with words in language	Message compositionality	Creative freedom	Number of distinct forms
Abridgement systems	Whistled (Spanish, Turkish, Tepehuan, Hmong, Moba, and many others)	Phonic	Often extensive	Message creation	Somewhat less than the number of lexical items in the language
	Instrumental (Drum: Akan, Banda-Linda, Bora; Xylophone: Seenku; among others)	Phonic	Yes	Message creation	Less than the number of lexical items in the language
Ideograph systems	Instrumental (Senegalese drum language, New Guinea slit-gong signaling)	Lexical	Limited	Message creation	Significantly less than the number of lexical items in the language
	Mehek name whistles (*isi*), Papua New Guinea	Lexical	No	No	94 attested conventional whistled counterparts exist for certain given names (not individuals)
Name tunes	Oyda *moyzé*	None	No	Tune composition	Potentially tens of thousands: most individuals have a unique name tune
	Yopno *konggap*	None	No	Tune composition	Potentially tens of thousands: each individual ideally has a unique name tune

Whistled Abridgement Systems (Classe, 1957; Rialland, 2005; Meyer, 2021 inter alia), Instrumental Abridgement Systems (Nketia, 1971; McPherson, 2018; Seifart et al., 2018 inter alia), Ideographic Systems (Burridge, 1959; Zemp and Kaufmann, 1969; Winter, 2014; Hatfield, 2016).

longer name "songs" for some given names, which appear to be slightly longer and more complex than either *konggap* or *moyzé*.

Table 2 compares features of these other traditions with those of the *konggap* and *moyzé*.

Mountainous environments may have played roles in the development of name tune systems among both the Oyda and the Yopno, as may be the case with surrogate speech systems—especially whistled ones—elsewhere (Meyer, 2021). Indeed, in both Oyda and Yopno, name tunes coexst with small inventories of whistled phrases (in Yopno, these may also be sounded on a conch shell horn) with meanings like "yes," "come," and "go": these do correspond to the Oyda and Yopno languages through abridgment, just like other whistled languages around the world. Another factor in the development of these name tunes systems may be taboos on pronouncing certain proper names: in both communities, people avoid using some proper names, in either strict taboos on speaking in-laws' names (Yopno) or milder social preferences to address and reference using kin terms (Oyda). But such preferences are far from unique to these communities (Fleming, 2011). The fact remains that neighboring communities to the Oyda and Yopno, who live in similar terrains and have similar cultural practices, lack name tune traditions entirely.

The next populated area due east from the Yopno area is the Uruwa region. Although Uruwa people use no name tunes, just one of the 10 Uruwa villages uses birth-order terms (small sets of names that denote a person's sex and birth-order) to hail and refer to individuals, while the other villages do not (Sarvasy, 2013, 2017). Perhaps naming systems—musical or linguistic–lend themselves to clean breaks with neighboring groups, unlike other aspects of language and culture, which are more likely to evince clines.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://dobes.mpi.nl/projects/oyda/language/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Volkswagen Stiftung for the DoBeS project on Oyda; University of Chicago Institutional Review Board, University of Chicago. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

AA contributed to Oyda name tune data and analysis. JS contributed to Yopno name tune data and analysis. HSS contributed to formal analysis. All authors contributed to writing.

FUNDING

Grant from the DoBeS programme of Volkswagen Stiftung II/84 327 to AA. Fulbright-Hayes Doctoral Dissertation Research Abroad Grant, Wenner-Gren Dissertation Fieldwork Grant, Endangered Languages Documentation Programme Individual Postdoctoral Fellowship to JS. Grant from the Australian Research Council DE180101609 to HSS.

ACKNOWLEDGMENTS

AA wishes to thank the Oyda speakers who provided information on *moyzé* and her host family: Ato Tegegne Taddese and his wife w/o Birtukan. JS is grateful to his hosts in Nian, Weskokop, Gua, and Ganngalut villages, and to everyone in the Yopno valley who supported his research. Thanks also to Alexandra Y. Aikhenvald and Roger Dean for helpful comments.

REFERENCES

Ammann, R., Keck, V., and Wassmann, J. (2013). The sound of a person: a music-cognitive study in the Finisterre Range in Papua New Guinea. *Oceania* 83, 63–87. doi: 10.1002/ocea.5012

Anderson, M. (1984). Proper names, naming, and labeling in saami. *Anthropol. Linguist.* 26, 186–201.

Boersma, P., and Weenink, D. (2019). *Praat (Version 6.1.03) [Software]*. Available online at: www.praat.org (accessed July 25, 2021).

Burridge, K. O. L. (1959). The slit-gong in Tangu, New Guinea. *Ethnos* 24, 136–150. doi: 10.1080/00141844.1959.9980869

Classe, A. (1957). The whistled language of la gomera. *Sci. Am.* 196, 111–124. doi: 10.1038/scientificamerican0457-111

Cowan, G. M. (1948). Mazateco whistle speech. *Language* 24, 280–286. doi: 10.2307/410362

Fleming, L. (2011). Name taboos and rigid performativity. *Anthropol. Q.* 84, 141–164. doi: 10.1353/anq.2011.0010

Hanssen, I. (2011). A song of identity: Yoik as example of the importance of symbolic cultural expression in intercultural communication/health care. *J. Intercult. Commun.* 27. doi: 10.1111/j.1471-6712.2012.01021.x

Hatfield, A. (2016). *A Grammar of Mehek* [Ph.D. dissertation]. State University of New York, Buffalo, NY, Unites State. Available online at: http://www.acsu.buffalo.edu/~dryer/hatfield_mehek2016.pdf (accessed July 25, 2021).

Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., and Louhivuori, J. (2000). Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition* 76, 13–58. doi: 10.1016/S0010-0277(00)00068-8

Leach, J. (2002). Drum and voice: aesthetics and social process on the rai coast of papua new guinea. *J. R. Anthropol. Inst.* 8, 713–734. doi: 10.1111/1467-9655.00130

McPherson, L. (2018). Grammatical principles and documentary implications. *Anthropol. Linguist.* 60, 255–294. doi: 10.1353/anl.2019.0006

Meyer, J. (2021). Environmental and linguistic typology of whistled languages. *Annu. Rev. Linguist.* 7, 493–510. doi: 10.1146/annurev-linguistics-011619-030444

Neeley, P. L. (1999). *People of the Drum of God—come!* Dallas, TX: SIL International.

Niles, D. (1992). "Konggap, kap and tambaran: music of the Yopno/Nankina area in relation to neighbouring groups," in *Abschied von der Vergangenheit. Ethnologische Berichte aus dem Finisterre-Gebirge in Papua New Guinea*, ed J. Wassmann (Berlin: Reimer), 149–183.

Niles, D. (2010). *Editor's Introduction to Zemp and Kaufmann (1969). Kulele 4*. Boroko: Institute for Papua New Guinea Studies, 1–50.

Nketia, J. H. K. (1971). "Surrogate languages of Africa," in *Current Trends in Linguistics: Linguistics in Sub-Saharan Africa, Vol. VII*, ed T. A. Sebeok (Berlin: Mouton), 699–732.

Reed, W. (2000). *Yopno Grammar Essentials. Unpublished Manuscript*. Ukarumpa: SIL.

Rialland, A. (2005). Phonological and phonetic aspects of whistled languages. *Phonology* 22, 237–271. doi: 10.1017/S0952675705000552

Sarvasy, H. (2013). Across the great divide: how birth-order terms scaled the Saruwaged Mountains in Papua New Guinea. *Anthropol. Linguist.* 55, 234–255. doi: 10.1353/anl.2013.0012

Sarvasy, H. (2017). *A Grammar of Nungon: A Papuan Language of Northeast New Guinea*. Leiden: Brill.

Searle, J. R. (1958). Proper names. *Mind* 67, 166–173. doi: 10.1093/mind/LXVII.266.166

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *R. Soc. Open Sci.* 5:170354. doi: 10.1098/rsos.170354

Slotta, J. (2012). *On the Receiving End: Cultural Frames for Communicative Acts in Postcolonial Papua New Guinea* (Ph.D. thesis). The University of Chicago, Chicago, IL, United States.

Slotta, J. (2014). *Documenting Yopno Diversity: Dialect Variation in a Papuan Language*. London: SOAS, Endangered Languages Archive. Available online at: https://elar.soas.ac.uk/Collection/MPI43302 (accessed July 25, 2021).

Stern, T. (1957). Drum and whistle languages: an analysis of speech surrogates. *Amer Anthropol.* 59, 487–506. doi: 10.1525/aa.1957.59.3.02a00070

Winter, Y. (2014). On the Grammar of a Senegalese Drum Language. *Language* 90, 644–668. doi: 10.1353/lan.2014.0061

Zemp, H., and Kaufmann, C. (1969). Pour une transcription automatique des "langages tambourinés" mélanésiens: (Un exemple kwoma, Nouvelle-Guinée). *L'Homme* 9, 38–88. doi: 10.3406/hom.1969.367047

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fcomm.2021.650382

The Language of Gángan, A Yorùbá Talking Drum

Samuel Kayode Akinbo *

Department of Linguistics, University of British Columbia, Vancouver, BC, Canada

It is widely known that Yorùbá drummers communicate through their native drums. This paper investigates the grammar of gángan, which belongs to a family of Yoruba drums called dùndún. The results of this study show that Yorùbá drummers represent the phonetic realisation of lexical and grammatical tones of their language with the drum. Statistically, the speech tones and the acoustic correlate of the corresponding drum representations have a significant positive relationship. In both spoken and drum communication, vowel (V) and consonant-vowel (CV) prosodic units have different statuses. To conclude, Yorùbá drummers communicate via the gángan drum by transposing certain phonemic features and maybe phonological conditions of their language to musical forms.

Keywords: tone, syllable, music, language, talking drum, yoruba, speech surrogate

OPEN ACCESS

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Akinbiyi Akinlabi,
Rutgers University, United States
Amanda Villepastour,
Cardiff University, United Kingdom

****Correspondence:***
Samuel Kayode Akinbo
akinbosk@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *07 January 2021*
Accepted: *19 August 2021*
Published: *09 September 2021*

Citation:
Akinbo SK (2021) The Language of Gángan, A Yorùbá Talking Drum. Front. Commun. 6:650382. doi: 10.3389/fcomm.2021.650382

1 INTRODUCTION

Cultures around the world communicate through musical instruments by transposing linguistic features to music melodies (e.g., Bradley, 1979 on the Sino-Tibetan gourd organ, Poss, 2005; Poss, 2012 on Hmong raj in Southeast Asia, Kaminski, 2008 on Asante ivory trumpet in Ghana, Winter, 2014 on the Sabar of Senegalese, McPherson, 2018 on the balafon of the Sambla people in Burkina Faso, Seifart et al., 2018 on Amazon Bora). To distinguish speech surrogate systems that represent the phonemic aspects of a language from those that represent meaning without reference to phonemic inventories, Stern (1957) refers to the former as “abridged” and the latter as “lexical ideogram”. This work solely focuses on an abridged system.

Ranging from a bell to an electric guitar, many musical instruments possess the capacity to be a speech surrogate (Lo-Bamijoko and Joy, 1987; Agawu, 2016). For example, Yorùbá musicians communicate with both native and “imported musical instruments” (Waterman, 2000, p. 199), but the native drums are better known and more studied owing to their ability to encode the tones of Yorùbá (Beier, 1954; Euba, 1967, 1990; Villepastour, 2010). As a result of this capability, the dùndún drums of the Yorùbá people are popularly known as “talking drums”, the communicative capability of which is the focus of this work.

Although the drum communication is based on encoding the phonemic features of Yorùbá with the music instrument, the speech surrogate system is mostly studied from musical and anthropological perspectives. As McPherson (2019) notes, studying musical speech surrogates from a linguistic perspective “can offer valuable insights into what phonological aspects are being represented with the drum”, and the data from the system can serve as language-external “evidence for phonological representations and theory”. The present study investigates how Yorùbá musicians communicate with *gángan*, which belongs to a family of Yorùbá drums called *dùndún*. Unlike previous studies of *dùndún*, this study utilises linguistic instrumentation and methodology.

Given that the surrogate system is based on Yorùbá, it is essential to understand the basic sound inventory of the language. Before turning to the main focus of this work, the basic sound inventory is presented in **Section 2**. The description of the drum which is used in this study is presented in **Section 3**. The data in this work is from a linguistic experiment, so the data source and the method of data collection are discussed in **Section 4**. The articulatory results of the experiment are presented in **Section 5**. The acoustic results of drumming lexical tones are presented in **Section 6**. The section also focuses on the strength of the relationship between Yorùbá words and their corresponding drum renditions. In addition to the lexical tones, the participants also drummed words with grammatical tones. The results of drumming the grammatical tones are presented in **Section 7**. In the final section **Section 8**, the summary, discussion and conclusion are presented[1].

2 BACKGROUND ON YORÙBÁ SOUND INVENTORY

Yorùbá is a Volta-Niger language spoken in West Africa and most prominently South Western Nigeria (Blench, 2019). To understand what is encoded with a talking drum, it is imperative to understand certain aspects of the sound patterns of the source language used for the drum communication. For this reason, this section presents a description of the relevant sound patterns in Standard Yorùbá, which is the focus of this work.

2.1 Syllable in Speech

There is a consensus that a syllable in Standard Yorùbá is constructed of a vowel with an onset consonant (CV) (Awóbùlúyì, 1978; Orie, 2000; Orie and Pulleyblank, 2002; Pulleyblank, 2009). A vowel is the peak of a syllable, and the consonant is an onset 1a).

(1) Syllable patterns in Yorùbá

		Phonetics	Orthography	
a.	CV	[bà.bǎ]	bàbá	'father'
		[tɛ̀.lě]	tẹ̀lé	'follow'
b.	V	[à.g͡bà]	àgbà	'elder'
		[a.gɔ̀]	agọ̀	'stupidity'

Following the proposal in Orie, 2000, an onsetless vowel (V) such as the examples in 1b) does not constitute a syllable in Standard Yoruba. Orie, 2000 presents a whole range of evidence in support of this account. For example, word-initially, any tone can be associated with CV syllables, but onsetless vowels cannot bear a High tone in the same context 2). While there are V prefixes with a mid or low tone, the only high tone prefix has an onset 3). Orie, 2000 for more evidence.

(2) Word-initial onsetless vowel[2]

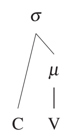
		CV		V	
1	H	rárà	'a genre of poem'	*íg̣bá	
2	L	ràrá	'dwarf'	ìgbá	'garden egg'
3	M	wolé	'enter'	igbá	'calabash'

(3) Prefixation: H-tone prefix must be a CV syllable

		M tone		L tone		H tone	
kú	'die'	i-kú	'death'	ì-kú	'dying'	kí-kú	'dying'
là	'split'	i-là	'tribal incision'	ì-là	'splitting'	lí-là	'splitting'

The syllable template is represented using a moraic structure in 4). Under standard moraic theory, a vowel projects a mora, which is a unit of length (Hyman, 1985; McCarthy, 1986; Hayes, 1989). Following the account in (Orie, 2000), only a vowel with an onset consonant (CV) can project a syllable in Standard Yorùbá. In this account, a V unit constitutes a mora, not a syllable.

(4) Yorùbá syllable



In addition to the forms in 1), there are VV units in the language. VV units like 5) constitute a sequence of two vowels not a long vowel or diphthong (Orie and Pulleyblank, 2002).

(5) VV sequences in Yorùbá

		Phonetics	Orthography	'elder'
a.	VV	[à.á.rɔ̀]	àárọ̀	'morning'
		[e.é.gṹ]	egúngún	'ancestral spirit'
b.	CVV	[jo.ò.bá]	Yoòbá	'Yorùbá'
		[wá.rì̃.í̃]	wá rì í	'come and sink it/him/her'

Under the moraic structure, VV units are bimoraic. There is very little (if any) evidence that VV sequences form single syllables (Orie, 2000; Orie and Pulleyblank, 2002). For example, VV sequences like 5a) are derived through intervocalic-consonant deletion and vocalic assimilation in the language (Akinlabi, 1993). The present work is based on Standard Yorùbá. In this work, I adopt the account in Orie (2000) that CV is the only syllable in this variety of Yorùbá. I will only refer to the moraic status of VV units in this work. Following from this, CV and V are monomoraic, but only the CV is syllabic.

2.2 Tone in Speech

Yorùbá is a tone language, which means pitch contrasts bring about lexical or grammatical distinctions in meaning (Yip, 2002; Hyman, 2018). As shown in 6), the language has three contrastive tones, namely H (igh), M (id) and L (ow) (Akinlabi, 1985;

[1]This is an extension of the preliminary version, which is published in LLA (Akinbo, 2019). While the version in LLA only focuses on lexical tones and syllables, the version here focuses on syllables, lexical tones and grammatical tones, and it compares spoken forms with their corresponding drum renditions. The present paper also improves on the description of the drum. This research is part of the project funded by SSHRC Insight grant (#435-2016-0369) awarded to Douglas Pulleyblank. For comments and suggestions on various aspects of this work, I thank Douglas Pulleyblank, Bukunmi Ogunsola, Kọ́lá Túbọ̀sún and Amanda Villepastour. Errors of fact or explanation are my own responsibility.

[2]Transcriptions are in standard Yorùbá orthography throughout, unless an example is enclosed in phonemic slants "//" or phonetic square brackets "[]". In Yorùbá orthography, [ɔ] = ọ [ɛ] = ẹ, nasalised vowel = Vn, [ʃ] = ṣ [kp] = p, High tone = bá, L tone = bà M tone = ba (unmarked for tone), syllabic nasal = a tone-marked nasal. Tone is indicated throughout, including tonal alternations

Pulleyblank, 1986).

(6) Yorùbá: Tonal minimal set

H	bá	'to meet'
M	ba	'to braid'
L	bà	'to land'

The tone-bearing unit in the language is a mora (Akinlabi and Liberman, 2000; Pulleyblank, 2004). Initial vowels in Yorùbá do not bear an H tone (Akinlabi, 1985; Pulleyblank, 2004; Pulleyblank, 2009). With sequences of H-L and L-H tones in the language, a contour tone is formed on the second tone[3]. However, this does not happen in H-M, M-H, L-M or M-L sequences (Ward, 1952; Akinlabi and Liberman, 1995). For example, words like /kpákò/ H-L "chewing stick" and /ìlú / "city" are realised as (kpákô) H-HL and (ìlǔ) L-LH respectively. The other relevant tonal process involves raising the pitch value of an H tone in a sequence of H-L tones and lowering the pitch of an L tone in a sequence of L-H tones (Akinlabi and Liberman, 1995; Laniran and Clements, 2003). For example, the pitch of H in /bájɔ̀/ "Báyọ̀" (a name) is higher than that of H in /bájɔ/ "exit through", and the pitch value of L in /ìlú/ "city" is lower than that of L in /ìlu/ "puncher".

Yorùbá also has a "subject marking H tone" Akinlabi and Liberman, 2000). To realise the H tone, an L-L noun phrase (henceforth NP) surfaces as L-LH 7a), but an H-H NP remains unchanged 7b). While an M-M NP surfaces as M-H 7c), an H-M NP surfaces with an extra H-tone mora (7d).

(7) Subject H tone

a.	ìlù	H	dára	→	ìlǔ dára	'a/the drum is good'
b.	síbí	H	dára	→	síbí dára	'a/the spoon is good'
c.	ọmọ	H	dára	→	ọmọ́ dára	'a/the child is good'
d.	ọmọ́le	H	lọ	→	Ọmọ́le é lọ	'Ọmọ́le went'

In genitive constructions, an M-tone extra vowel is realised on the possessum (Akinlabi and Liberman, 2001). Consider the examples below 8).

(8) M tone in genitives

ṣíbí	M	Délé	→	ṣíbí i Délé	'a/the spoon of Délé
bàtà	M	Ṣọlá	→	bàtà a Ṣọlá	'a/the shoe of Ṣọlá'

The basic description of the Yorùbá syllable and tone patterns has been detailed in this section. Throughout this paper, we will refer to the discussion in this section.

Most of the studies on Yorùbá drum communication are based on natural or semi-natural musical performance with limited or no acoustic evidence (e.g, Beier, 1954; Euba, 1967; Euba, 1990; Villepastour, 2010). Given that the communicative capability of the drums is based on encoding phonemic features of a drummer's language, it would be insightful to analyse the system like we analyse human language. By using linguistic instruments and methodology, the present work investigates how a Yorùbá drum communicates. The discussion in the present section will help us understand the linguistic features that are encoded with the talking drum. I now turn to the properties of the talking drum.

[3]A contour tone is a tone that slides from one pitch to another. In musical term, this is called a glissando.

3 THE COMPONENTS OF DÙNDÚN

Dùndún is the generic name for a family of Yorùbá drums, and the word dùndún literally means "sweet sound" (Ruskin, 2013, p. 1). The dùndún ensemble comprises ìyáàlù, *ìsáájú*, *kẹríkẹrì*, *gúdúgúdú* and *ìkẹ́yìn* (Euba, 1990). Ìyáàlù (lit. "mother drum") is the lead drum of the ensemble (Durojaye, 2020). The drums in the ensemble can be distinguished based on their structures, relative sizes and performance techniques and musical functions in the ensemble[4]. With the exception of gúdúgúdú, all the drums in the dùndún ensemble are hourglass-shaped pressure drums[5]. By considering the bells and cloth around the lead drum, Euba (1990) identifies nine components for the hourglass-shaped drums, but all the hourglass-shaped drums have seven components in common. The seven components are a wooden resonator, two surface membranes, tension cords that change the membrane pitches, a taut rope for tuning the drum, a strap to carry drum, leather rings ẹ̀gì for holding the membranes in place and a curved stick to play the drum. The image in **Figure 1** is adapted from Arewa and Adekola (1980), but the image and labels have been modified. This study solely focuses on gángan which belongs to the family of Yorùbá drums called dùndún[6].

The hourglass shape of the drums is from the carving of the wooden resonator into an hourglass shape with a tunnel-like hole linking the two ends. Traditionally, the ideal wood for making the drum resonator is "igi ọ̀mọ̀", which is a *Cordia alliodora (Linn)* tree (Lawal et al., 2010; Ọmọ́bọ́lá, 2019). Each opening of the drum resonator is covered with a drum membrane. The ideal material for the drum membranes is soft animal skins. So, the source of the drum membranes varies from one drum maker to another. Some drum makers prefer goatskin while others prefer the skin of a cow's foetus (BattaBox, 2017; dijiaderoGBA, 2018). My main consultant prefers the skin of a goat's foetus.

The drum membranes are stretched over the holed ends of the wooden resonator. The edges of the two drum membranes are connected and tightly held together with strings, which are made from a leather material. These strings form the tension cords or the wall of the drums, and the tensioning of these strings varies the tightness of the drum membrane. In order to communicate with the drum, the membranes of the drum need to be loose or slack on the resonator. However, when the drum has not been used for a while, the drum membrane shrinks and becomes tightly held to the resonator. To make the drum suitable for communication, the drum is tuned by using the taut rope to compress the tension cords for at least 4 hours. The taut rope can be made from leather or wool.

[4]See the performance in the YouTube video for the similarity in the features of the drums in the ensemble https://www.youtube.com/watch?v=kmxfxRmzVog.

[5]The term, ìyáàlù, are not exclusively used for the dùndún ensemble. In Yorùbá musical tradition, the term refers to musical function of the instrument playing the lead role. For example, ìyáàlù is also the lead drum in the bàtá ensemble and other Yoruba drum ensembles (Oludare et al., 2018).

[6]The components of the drums can be seen in this YouTube tutorial on how to make the drum: https://www.youtube.com/watch?v=1XP4ox_lORs.

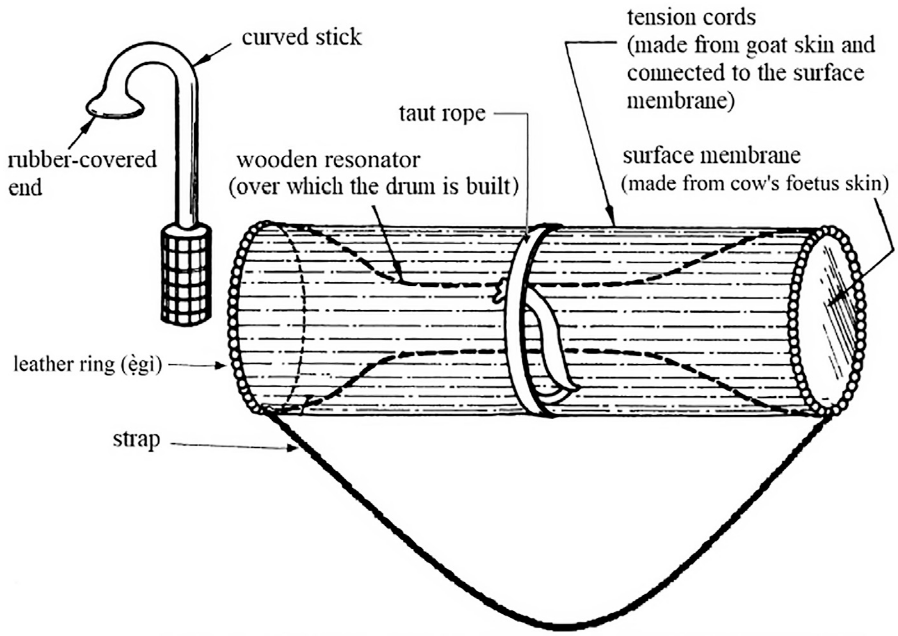


FIGURE 1 | Dùndún drum.

The drum is played with a curved stick. To avoid puncturing the drum, the head of the stick is covered with a flat rubber or leather material. To carry the drum, a strap is made from leather and padded cloth. Depending of the handedness of the drummer, the drum is suspended over the left or right shoulder with the strap and hangs under the opposite armpit.

Previous research suggests that Yorùbá drummers communicate by means of the drum by varying the compression of the tension cords (e.g. Beier, 1954; Euba, 1990; Villepastour, 2010). Unlike the previous studies, the present paper describes an articulatory and acoustic study of how Yorùbá drummers communicate with the talking drum by drawing insight from linguistic analysis of Yorùbá tones and syllables. The study is guided by the following specific questions: 1) How are tonal processes such low-tone lowering and high-tone raising represented by talking drum? 2) Are both lexical and grammatical tones represented with the talking drum? 3) How strong is the relationship between the spoken form of Yorùbá words and the corresponding drum rendition? To answer these questions, a linguistic experiment is conducted. The details of the experiment are presented in the following section.

4 METHODOLOGY

The data in this work were elicited at Diamond FM radio station, University of Ibadan, Nigeria from five male drummers who are native speakers of Yorùbá. Four of the drummers have at least 7–18 years of drumming experience, and one has 2 years of drumming experience. The data were recorded in a soundproofed room with a SHURE WH30XLR cardioid condenser (a headset microphone) and a Rode NGT2 supercardioid condenser (a shotgun microphone) at the sampling rate of 48.1 kHz in WAV format. The microphones were attached to a zoomQ8 camcorder. The headset microphone captured the speech of the drummer, and the shotgun microphone was pointed at the drum. The audio from the two microphones were saved as separate files at the same time as the video files. All the participants in the study used the same drum except for one who insisted on using his own drum.

The stimuli in this work consist of monosyllabic and bisyllabic words, and CVV units. The monosyllabic words with level tones are three words with each word bearing either H, L or M. The bisyllabic and VCV units with level tones cover nine tonal combinatorial possibilities 9). The CVV units cover three tone types: L-H rẹ̀ẹ́ "be tired" (lit. "tired it"), H-M ríi "see it" and M-H jẹẹ́ "eat it".

(9) Tone types in mono- and bisyllabic stimuli (Akinlabi and Liberman, 2000)

		H	M	L
H	rá 'disappear'	pákó 'plank'	kése 'infinity'	pákô 'chewing stick'
M	ra 'rub'	ọkọ́ 'hoe'	ọkọ 'husband'	ọkọ̀ 'vehicle'
L	rà 'buy'	ìlǔ 'city'	ìlu 'puncher'	ìlù 'drum'

In musical traditions with speech surrogates, [(Euba, 1990), p. 193] identifies three forms of drumming, namely "direct speech form", "musical speech form" and "song form". In this work, the drummers only drummed the stimuli in speech mode, which "involves the direct reproduction of the pitches and rhythms of spoken language" (Agawu, 2016, p. 128). Simply put, the drummers spoke the stimuli then drummed them. This was repeated three times for each stimulus. It bears mentioning that the experiment in this work is specifically designed to investigate the drumming of tones and syllable, so issues relating to segmental identity are only discussed in relation to tones.

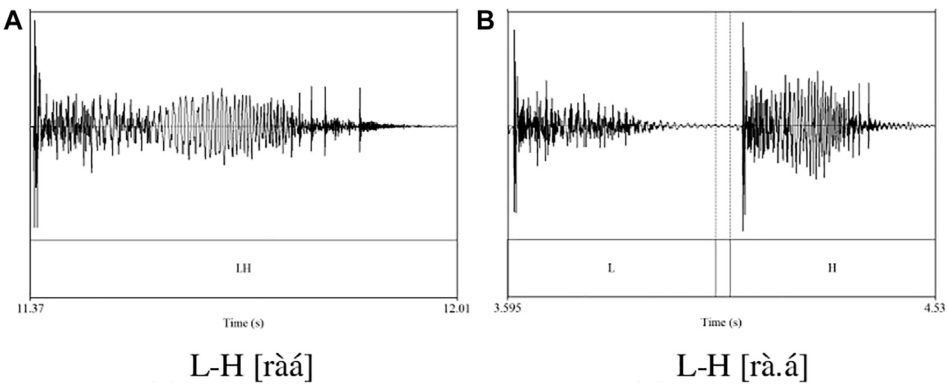


FIGURE 2 | Drumming ‘ràá’ as a monosyllabic form and a bisyllabic form. **(A)** L-H [r̀áa]. **(B)** L-H [r̀á.a].

5 RESULT 1: ARTICULATION OF YORÙBÁ WORDS WITH TALKING DRUM

The articulation of Yorùbá words with gángan is discussed in this section. Note, in drumming, only one of the surfaces of a dùndún drum is used. Regardless of the speech tone which was represented by the drum, the drummers struck the drum membrane once to produce a CV word. For a CVCV or VCV word, the drum membrane was struck twice. When the bimoraic CVV units were produced in natural speech, the drummers struck the drum membrane once. In careful or deliberate speech, the drum membrane were struck twice. These drumming options for the CVV units are presented in **Figure 2**.

As discussed in **Section 3**, varying the compression of the tension cords affects the tightness of the drum membrane. To understand the representation of tones with the talking drum, the drumming of the minimal set [rá] “disappear”, [ra] “rub” and [rà] “buy” are observed in isolation. To drum the H-tone word [rá] ‘disappear’, the tension cords were tightly compressed, then the drum membrane is struck. For the drumming of the M-tone word [ra] “rub”, the tension cords of the drum were loosely compressed before the drummer struck the drum membrane. By striking the drum membrane without compressing the tension cords, the L-tone word [rà] “buy” was drummed. Considering that the pitch value of the initial L drum tone in L-L and L-M sequences is higher than the pitch value of the initial L in the L-H sequence (see **Section 6.2**), the drummers may have compressed the tension cords for the initial L tone in L-M and L-L sequences but not for the initial L in the L-H sequence. In this case, the drummers rendered the L-tone lowering in the L-H sequence by producing the lowest tone possible. However, the compressions in L-M and L-L sequences were not visually perceptible in the video of the drumming experiment.

To drum the H-M sequence of a CVV unit in a normal speech, the tension cords of the drum were tightly compressed then the drum membrane was struck for the initial H. As the membrane was vibrating from the strike, the tension cords were then slightly loosened for the M tone. For a word with an M-H sequence of a CVV unit in a normal speech, the tension cords of the drum were slightly compressed then the drum membrane was struck. As the drum resonates from the initial strike, the drummer tightly compressed the drum. The drummers articulated words with an L-H of a CVV unit by striking the drum membrane without compressing the tension cords. As the drum vibrates from the strike, they tightly compressed the drum. In **Table 1**, the summary of the drumming is presented.

As shown in the table, the drummers produced H with a tight compression of the tension cords, M tone with a light compression and a L tone with no or minimal compression. In this case, the tones in the Yorùbá words are distinctly encoded by varying the compression of the tension cords. The summary also shows that the drummers obligatorily struck the drum membrane for CV and word-initial V. However, the drum membrane is optionally struck for word-medial V. The asymmetry between the drumming of CV and V can be accounted if we consider the status of CV and V in Standard Yorùbá speech. Based on the proposal in Orie (2000), V and the syllabic CV are monomoraic, but only CV is syllabic in Standard Yorùbá. The proposal is based on the asymmetry in the distribution of V and CV (**Section 2.1**). In line with the account of Standard Yorùbá syllable in Orie (2000), this shows that V and CV have different status in both drumming and speech.

6 RESULT 2: ACOUSTIC RESULTS OF DRUMMING WORDS WITH LEXICAL TONES

This section presents the acoustic results of drumming the words with only lexical tones. In order to determine

TABLE 1 | Summary: Articulation of words.

Tone	Tension-Cord Compression	Segments	Strike
L	No/minimal compression	CV	1
M	Light compression	VCV	2
H	Tight compression	CVCV	2
—	—	CVV	1 or 2

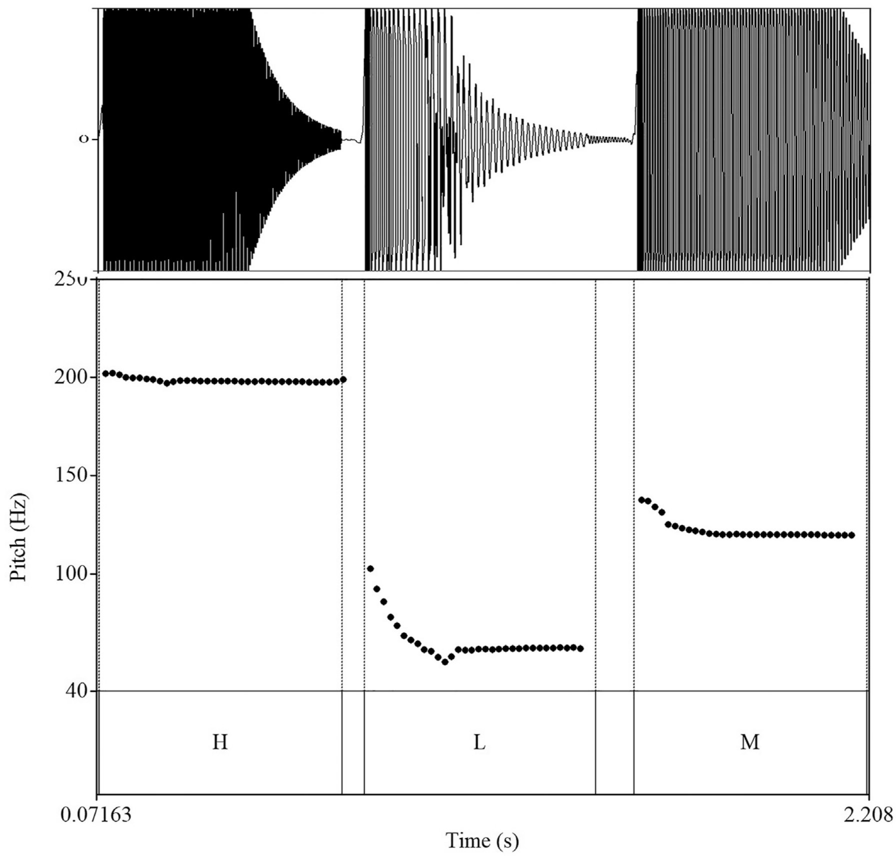


FIGURE 3 | Drumming H (rá) 'disappear', L (rà) 'buy' and M (ra) 'rub' in isolation.

the strength of the relationship between the speech tones and the corresponding drum representation, this section also compares the acoustic results of speech tones to those of their musical correspondents. Throughout this work, the drum representation of speech tones are called drum tones.

6.1 Acoustic Correlates of Tones in Talking Drum

The acoustic correlates of the drum strokes and tensioning are presented in this section. This is based on the words that were produced in **Section 5**. The acoustic results in this section are based on the data from three drummers who used the same drum. The results of this study show that the acoustic cues for drum tones are the pitch contours. As a result of this, the discussion mainly focuses on pitch tracks.

The musical representations of the contrastive speech tones in Yorùbá, namely H, L and M, have distinctive pitch tracks. As shown in **Figure 3**, the pitch of H tone is higher than those of M and L tones, and the pitch of M tone is higher than that of L tone.

The drum tones were manually annotated in Praat (Boersma, 2001). Using a script written by Riebold (2013), F_0 values of the pitch tracks were extracted at 50% point for the three drum tones. The mean F_0 values, which are presented in **Table 2**, show that the drum tones have distinctive F_0 values.

The correlation between speech tones and drummed tone are also shown for the phrases in (10). The speech and musical forms of the utterances were produced at least four times by the participants.

(10) Yorùbá utterances (source Laniran and Clements, 2003)

a. ògúnbọ̀dé gbà 'gbá
Ogunbode collect garden egg
'Ogunbode collected a/the garden egg'

b. ìgbéga dára
promotion Dára
'Dára's promotion'

The mean F_0 values of the drummed tones are plotted against those of the speech tone. In **Figure 4**, the y-axis indicates the acoustic measurement of pitch contour in

TABLE 2 | Mean F_0 values of H, M and L drum tones.

Drum tones	Tokens	Mean F_0
H	12	172
M	12	125
L	12	61

F_0 (Hz), and the x-axis indicates the sequences of tone in the Yorùbá phrase in 10). As shown in **Figure 4**, the pitch contours of the speech are similar to those of their drum renditions. Relative to the pitch values of the tones in speech, the pitch values of tones in drumming are amplified. By this, I mean that the pitches of H and M tones are higher and that the pitch of an L tone is lower in drumming. The difference between the pitch curves of the speech and drum in **Figure 4B** might be an effect of the intrinsic F_0 of vowels (Hombert, 1977).

Using *ggpubr* (Kassambara, 2018), the correlation coefficient of the speech and drum tones are calculated for the utterances in 10). The results of the calculation are presented in **Figure 5**.

In **Figure 5**, the results of the correlation test show that the relationship between the speech tones and the drum tones are positively strong ($R \geq 0.98$) and statistically significant ($p \leq 0.0043$).

In sum, the pitch contours are the acoustic correlates of drum tones, and the pitch values clearly distinguishes the drum tones. There is a strong positive relationship between the speech tones and the drum tones, and that this relationship is statistically significant.

6.2 Acoustic Results of Tonal Co-occurrence

The present section focuses on the tonal co-occurrence and tonal processes in speech and drumming. The discussion in **Section 2** shows that a sequence of H-L tone in Yorùbá is realised as H-HL. Similarly, a sequence of L-H is realised as L-LH. This pattern of tone transfer is referred to as pitch delay in the linguistic literature (Akinlabi and Liberman, 1995; Yip, 2002).

In **Figure 6**, the tone transfer is shown with the pitch tracks of the words pákò "chewing stick" and ìlú "city". When we compare the pitch trajectories of these words to those of their drum rendition in **Figure 7**, we observe that the pitch trajectories of the drum rendition are similar to those of the corresponding speech.

As shown in **Figures 6**, **7**, the sequence of H-L tones, a LH contour is formed on the second tone in both speech and drum. Similarly, in the sequence of L-H tones, a HL contour is formed on the second tone in both speech and drum. This contour formation does not occur on the second tone in H-M, M-H, M-L or L-M. Quantitative data on the pitch contours of the second tone in two sequences of tones are presented in **Figure 8**. In the graph, the y-axis indicates the acoustic measurement of pitch contour in F_0 (Hz), and the x-axis indicates the proportional duration of the tones. There are three panels in the graph, where the three tones form one panel each.

If we consider that the second tone in the H-L sequence has to be articulated on the drum with no compression of the tension cords on the L syllable (**Section 5**), one way of explaining the contour formation on the L drum tone in the H-L sequence is as follows. The drummers strike the drum membrane to encode the first syllable, they then compress the tension cords for the articulation of the initial tone as the drum resonates. However, the initial compression is sustained until after striking the drum membrane for the second syllable. As the drum resonates from the second strike, the tension cords are not compressed for the articulation of the second drum tone. By sustaining the initial compression (or lack of compression for the L-H sequence) until after striking the drum membrane for the second syllable, the drummers should be able to articulate the contour formation. While this is assumed to be the explanation for the contour, this hypothesis should be tested experimentally in future research.

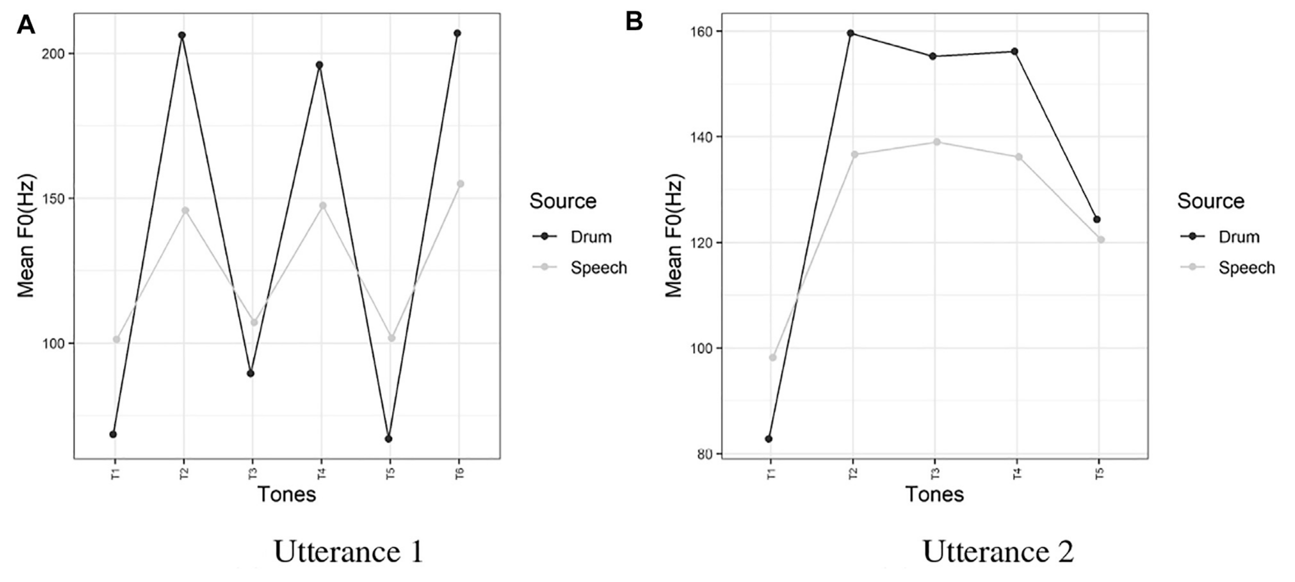


FIGURE 4 | Mean F_0 contours for the drum and speech tones. **(A)** Utterance 1. **(B)** Utterance 2.

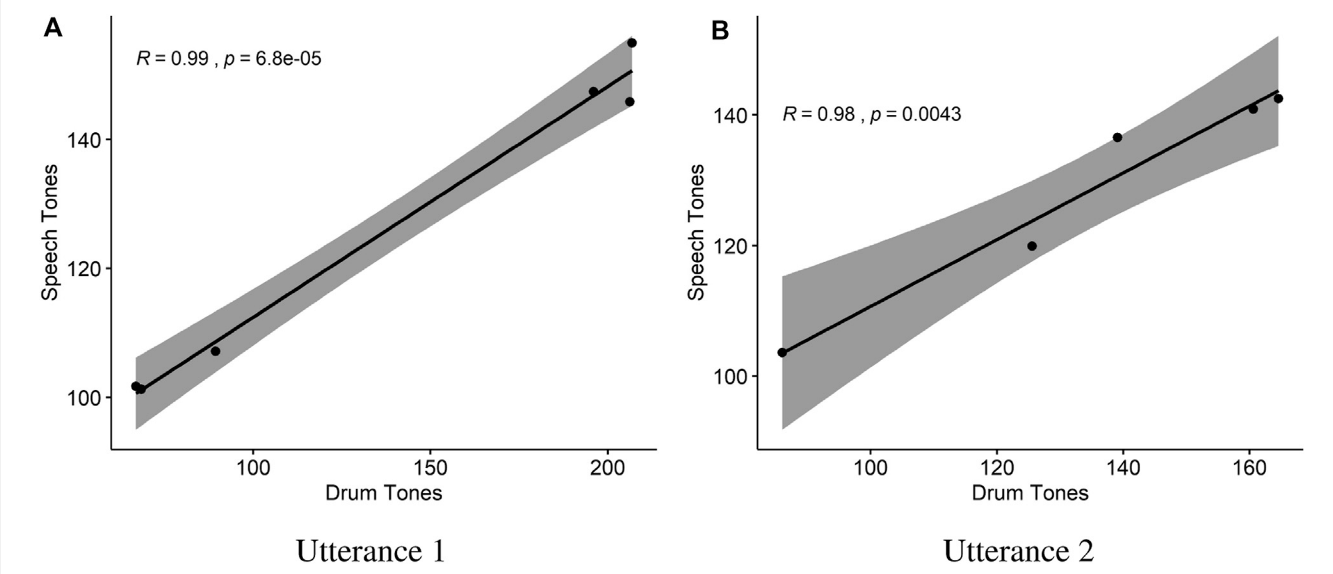


FIGURE 5 | Correlation between Mean F_0 contours of the drum and speech tones. **(A)** Utterance 1. **(B)** Utterance 2.

To check if there is any effect of the following drum tone on a word-initial drum tone, F_0 values of word-initial drum tones in a sequence of two tones were boxplotted[7] in **Figure 9** by using ggplot2 (Wickham, 2016; Kassambara, 2018).

As shown in **Figure 9**, H, L and M drum tones are clearly distinctive. The F_0 value of the H drum tone is higher when it precedes an L drum tone, and that of the L tone is lower when it precedes an H drum tone. For the M drum tone, it has a higher F_0 value when it precedes an M drum tone. The H-raising and L-lowering are consistent with the patterns in speech (see the discussion in **Section 2**). The difference between the F_0 values of the H tone before H and L tones is statistically significant, but the difference of the H-tone values before M and L tones is not statistically significant. Similarly, the difference between the F_0 values of the H tone before H and M tones is not statistically significant. When we compare the following H to the following L, there is no significant effect on the F_0 value of the preceding L. Similarly, comparing the following M to the following L shows no significant effect on the F_0 value of the L tone. However, comparing the following H to the following M shows a significant effect on the F_0 of the L tone. The comparisons of H, L or M have no significant effects on the F_0 value of the preceding M tone.

[7]The y-axis contains the F_0 of the drum tone while the distributional characteristics of the F_0 range of each tone in an environment is represented in the box. The box contains the middle 50% of the F_0 range. The mid-line in the box marks the median. The top cell of the box contains the maximum 25% of the box and the bottom cell of the box contains the minimum 25% of the box. Each of the lines below and above the box contains 25% of the F_0 values which are outside the middle 50%. The dots represent the outliers in the data. The numbers on the bars connecting the compared groups is the p(robability) values of the observed differences between the groups. *p*-values that are ≦0.05 indicates that the differences are statistically significant.

In sum, the results in this section suggest that the contour formation on the second speech tone in the sequence of H-L and L-H is represented with the talking drum, and that the H-raising and L-lowering in speech are also represented with the drum.

7 RESULT 3: GRAMMATICAL TONES IN TALKING DRUM

The discussion so far has focused on the drumming of lexical tones. In this section, we focus on the drumming of the subject H tone and the extra vowel with a M tone in the genitive constructions, which are discussed in **Section 2**. Examples of the subject H tone and the M-tone extra vowel are presented in 11) and 12) respectively.

(11) Subject H tone
ọmọ H dára → [**ɔ.mɔ́**] dára 'a/the child is good'
ìlù H dára → [**ì.lǔ**] dára 'a/the drum is good'

(12) M-tone extra vowel in genitives
pákó V Ṣọlá → [**k͡pá.kó.o**] Ṣọlá 'Ṣọlá's plank'
ọkọ V Ṣọlá → [**ɔ.kɔ.ɔ**] Ṣọlá 'Ṣọlá's husband'

To investigate whether the subject H tone and the extra vowel with a M tone are encoded in drumming, sentences with NPs in subject and possessum positions were drummed four times by each participant. The NPs with the sequence of H-L and L-H tones are excluded in order to control for the contour formation on the second tone (**Section 6.2**). The mean F_0 trajectories of the subjects and the possessa are respectively plotted against a control group, which is the NPs in isolation.

Figure 10 contains NPs with the isolation sequences of L-L and M-M tones. When we compare the occurrence of these NPs in isolation to their occurrence in a subject position, we see that the second half of the pitch trajectory for the NPs is

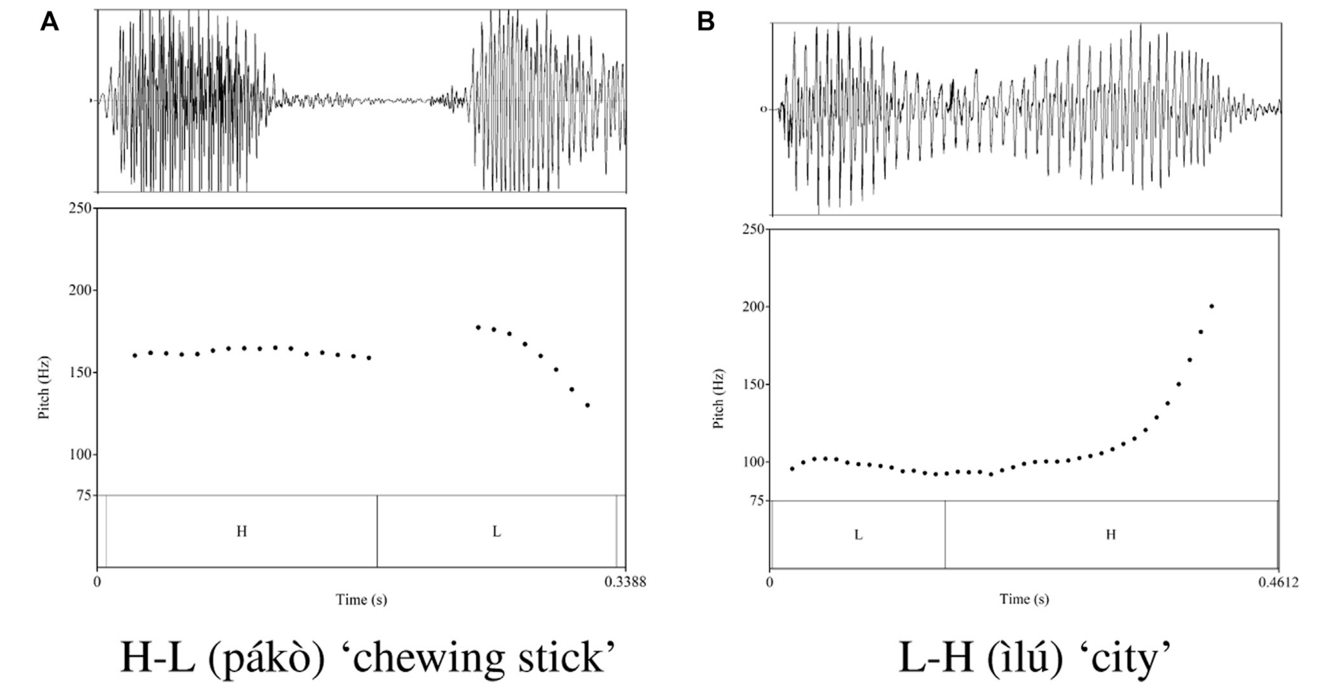


FIGURE 6 | Pitch tracks of H-L and L-H speech-tone sequences. **(A)** H-L (p'ak'o) 'chewing stick'. **(B)** L-H ('l'u) 'city'.

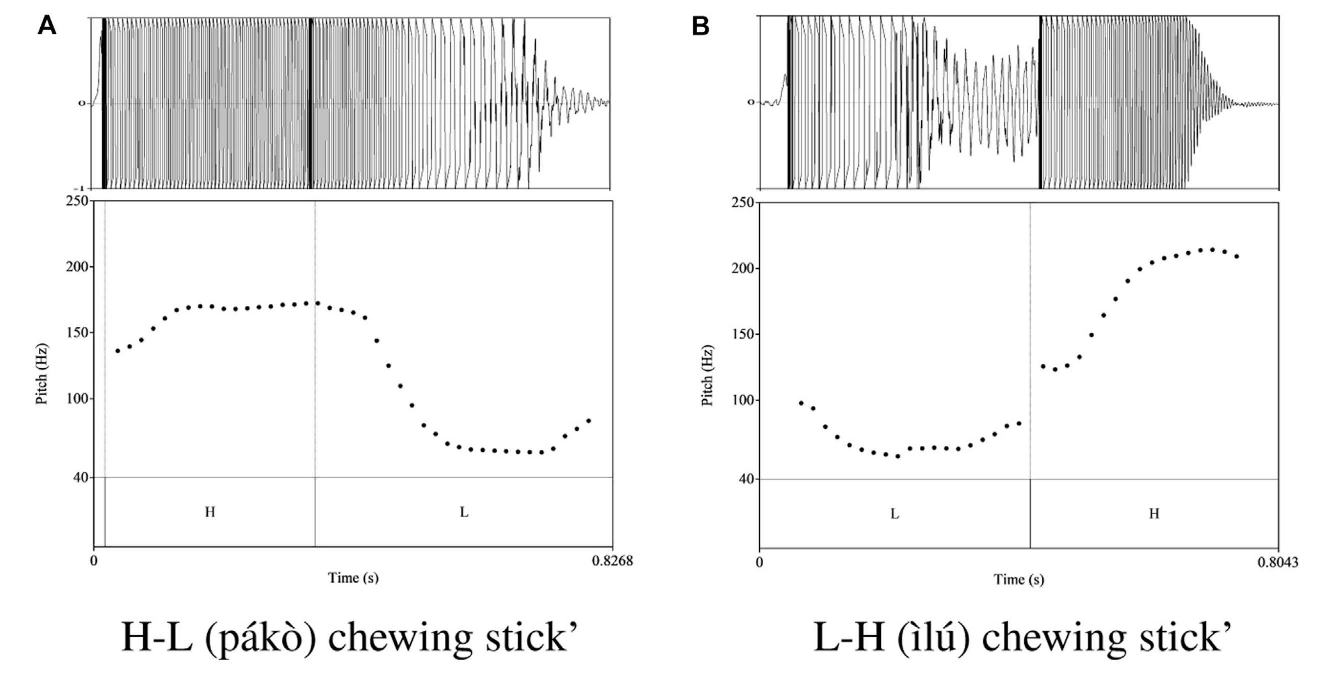


FIGURE 7 | Pitch tracks of H-L and L-H drum-tone sequences. **(A)** H-L (p'ak'o) chewing stick'. **(B)** L-H ('l'u) chewing stick'.

raised when they occur as the subject. The trajectory is consistent with the subject H tone found in speech. This suggests that the drummers did not only represent but exaggerate the subject H tone with the drum.

Before discussing the results of the M-tone extra vowel in **Figure 11**, we need to note that the bimoraic possessum and the M-tone extra vowel were produced as trimoraic (C) VCVV forms in natural speech, so the drummers only struck the drum

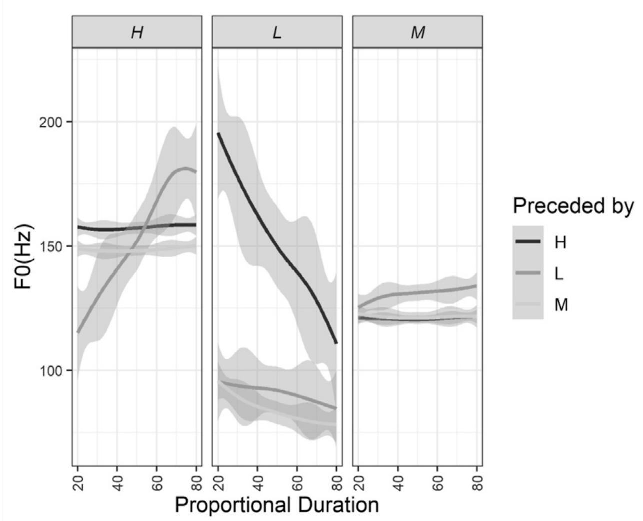


FIGURE 8 | Pitch contour of the second tone in a sequence of two drum tones.

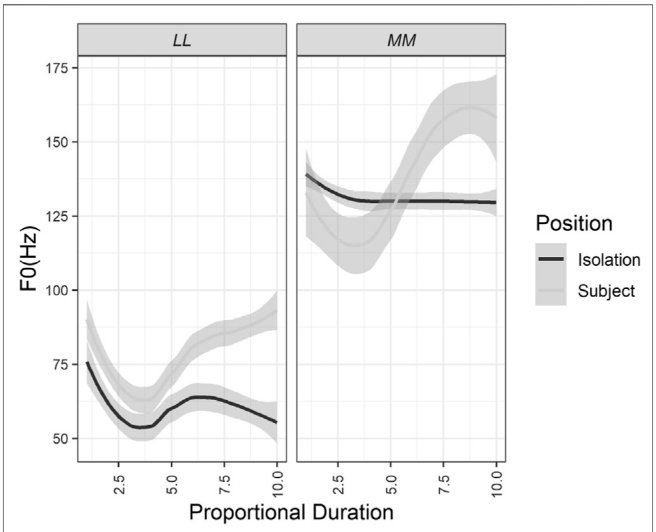


FIGURE 10 | Mean F_0 contours for the drummed NPs in isolation and subject position.

membrane twice for this constituent. This is consistent with the account in **Section 5**.

I now turn to the description of the results in **Figure 11**. The Yorùbá forms that were used for the five patterns in **Figure 11** are presented in A. With the exception of the NP with the sequence of M-M tones, the F_0 trajectories and the duration of the possessa are consistent with the presence of the M-tone extra vowel in speech. However, the curves in ML and MH sequences in isolation are not found speech contours (Akinlabi and Liberman, 1995). For the NP with the sequence of M-M tones, the F_0 trajectory is consistent with the spoken form, but the duration of the drummed form is not consistent with the trimoraic status of the possessum with the M-tone extra vowel. This is because the tension cords of the drum is only compressed once for CVV moras with the same tone. So, why are the tension cords not compressed for a longer duration like the other forms? To answer this question, we have to look at the tone of the object pronouns in Yorùbá 13).

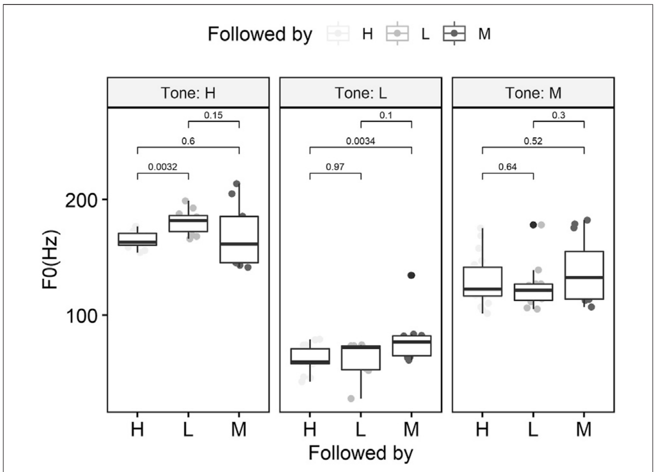


FIGURE 9 | Word-initial H L and M followed by H L or M.

(13) Tone of the object pronoun "me" in Yorùbá (Akinlabi and Liberman, 2000)

	Input	Output	
a.	kà + mí	kà **mí**	'count me'
b.	ká + mí	ká **mi** (*ká mí)	'caught me'
c.	jẹ + mí	jẹ + **mí**	'eat me'

The object pronouns in Yorùbá have an H tone underlyingly. When the pronouns follow a tone-bearing unit with an H tone, they surface with an M tone (Akinlabi, 1985; Akinlabi and Liberman, 2000). 13) is an example of this pattern. As argued in (Pulleyblank, 2004), the tonal alternation in 13) is the effect of the obligatory contour principle (OCP), which requires adjacent tones to be distinct at the melodic level of the grammar (e.g, Leben, 1973; Goldsmith, 1976; McCarthy, 1986; Archangeli and Pulleyblank, 1994). Lowering the H tone of the pronominal clitic to an M tone is the solution that is adopted for the form in (13-b).

The co-occurrence of the NP-final vowel and the M-tone extra vowel would have resulted in an OCP violation if the NP-final vowel and the extra vowel associate with different M tones (14a-b). For this form, the best way to satisfy OCP involves the adjacent morae associating with 1 M tone (14-c). As proposed in Pulleyblank (2004), this is the solution that is adopted in Yorùbá. Based on this, it is possibly the case that the drummers only encode 1 M tone, which associates with the three morae in the bisyllabic form.

(14) OCP satisfaction for **ọkọ + V**... → [ɔ.kɔ.ɔ]

a.		b.		c.
M	M	M	M	M
ɔ kɔ	ɔ	ɔ	kɔ ɔ	ɔ kɔ ɔ

To summarise, the grammatical tones are encoded with the talking drum. This is expected given that the phonetic realisation

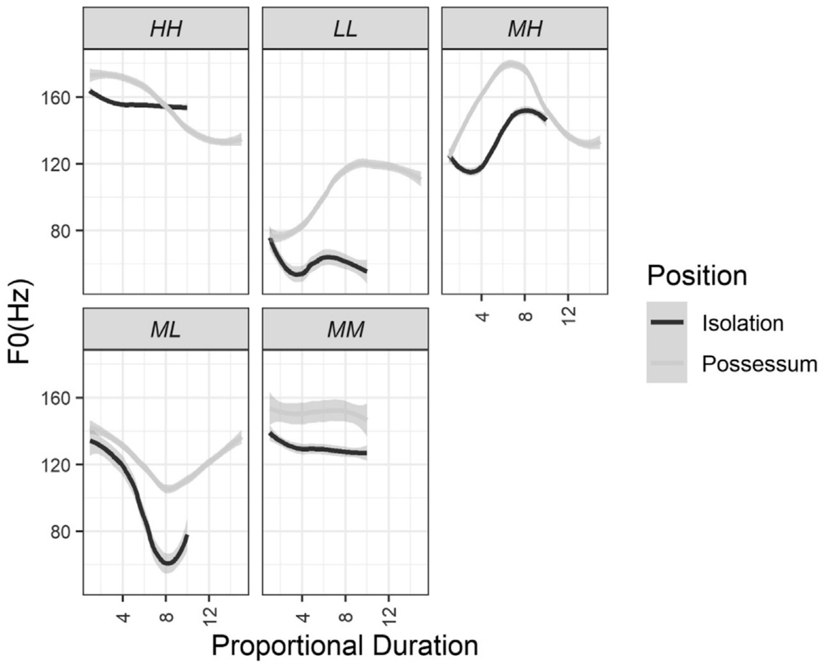


FIGURE 11 | Mean F_0 contours for the drummed NPs in isolation and possessum position.

of the tones in Yorùbá is encoded with the drum. However, the duration of tension-cord compression in the M-M possessum suggests that the drummers might be sensitive to the OCP condition.

8 SUMMARY, DISCUSSION AND CONCLUSION

The present study shows that the drummers obligatorily struck the drum membrane for a CV syllable or a word-initial V form. When the V form occurs word-medially, the drum was optionally struck. By varying the compression of the tension cords, the drummers represented the tones of the words. The acoustic results show that H, M and L drum tones have distinctive pitches. In sequences of H-L and L-H drum tones, a contour is formed on the second drum tone. The pitch of the H drum tone is significantly raised in a sequence of H-L drum tones. However, the pitch of the L drum tone is significantly lowered in a sequence of L-H drum tones presumably because it is produced with no tension. The pitch tracks of the grammatical tones in the drumming are consistent with those of the grammatical tones in speech. Comparing the pitch contours of the speech tones to those of the corresponding drum tones shows that there is a significantly strong positive relationship between the speech tones and the corresponding drum tones.

The results support previous findings that the lexical and grammatical tones are distinctively represented with the talking drum (Euba, 1990; Villepastour, 2010). Note that Villepastour (2010) considers the tones of segmental morphemes [e.g. the negative marker (kò)] as grammatical tones, but in the phonology literature, the term grammatical tone is confined to grammatical morphemes with tones as their phonetic exponents (Akinlabi and Liberman, 2000; Akinlabi and Liberman, 2001; Rolle, 2018). That V and CV forms have different status in drumming is in line the account in Orie (2000) that there is syllable asymmetry in Standard Yorùbá.

The present study has multiple implications. First, the musical representation of the lexical tones, the grammatical tones and the phonetic details of the tones could serve as language-external evidence for linguistic theory, such as the Theory of Adaptive Dispersion (TAD) (Liljencrants and Lindblom, 1972; Diehl and Lindblom, 2004). In the TAD, "preferred phoneme and feature inventories reflect the listener-oriented selection criterion of auditory distinctiveness", which is "achieved through maximal dispersion of phonemes in the available phonetic space" (Liljencrants and Lindblom, 1972; Diehl and Lindblom, 2004). Although Liljencrants and Lindblom, 1972) proposes the theory for segmental features, Yoshida (2011) extends it to tone inventories and argues that the motivation for H-raising and L-lowering in languages like Yorùbá and Japanese is the maximisation of contrast. That the drummers encoded H-raising and L-raising with the drum can also be the effect of contrast maximisation. Encoding these tone processes with the drum could serve as language external evidence for TAD.

Studies show that vowel and consonant types affect the pitch value of a tone (Hombert, 1977; Whalen et al., 1995; Whalen et al., 1999). If we take into consideration the claim that the tone-based speech surrogates like Yorùbá do not encode segmental identity (McPherson, 2019), the second implication of the present study is that tone-based speech surrogates create a segment-neutral ground for testing linguistic theory on tones. Third, while the link between music and language in African music is emphasised by musicologists and music teachers (e.g. Westphal, 1948; Nketia, 1963; Nketia, 1970; Ekwueme, 1974; Agawu, 1988; Agawu, 2001; Agawu, 2016), African-music education rarely incorporates linguistic courses, especially phonetics and phonology (Oehrle, 1991; Horton, 1997; Nzewi, 1999). That the drummers encode the phonetic details implies that phonetics and phonology courses could be essential in African-music education. In other words, to play an instrument like gángan successfully, a level of understanding of the source language for the instrument (and by that learning to speak the language) is important.

Researchers have studied the effect of musical knowledge or performance on second language acquisition, and the results of the studies suggest that music experience, rhythmic language and leisurely activities contribute to the success of perceiving and producing tones in second language acquisition (e.g. Orie, 2006; Gottfried, 2007; Wayland et al., 2010; Cooper and Wang, 2012). Considering that abridged surrogate systems such as the one in Yorùbá is based on the tone, it would be interesting to investigate whether training second language learners to perceive or encode speech tones with talking drums could aid the acquisition of tones.

Just as the pitch values of H and L tones in Yorùbá speech, the pitch values of the H and L drum notes vary depending on the following drum note in speech mode. This suggests that the pitch values of the gángan notes are also flexible. For example, the talking drum can produce infinite pitch possibilities between its highest and lowest pitches, but the pitch ranges can be categorised into three drum tones, namely H, M and L. In fact, Yorùbá musicians teach the talking drum with reference to the three drum tones. If we consider that drummers might not compress the tension cords at the same rate twice for a specific drum tone, the pitch value of drum notes might vary on different occasions.

It bears mentioning that "the dùndún drum does not have a language of its own; it is the drummer who speaks through the drum" (Eluyefa, 2011, p. 76). In this sense, if a dùndún drummer speaks English, the drummer can definitely present English phrases through the drum. In fact, Eluyefa (2011) reports cases of dùndún drummers representing English phrases with the drum. A similar example comes from the performance of the Yorùbá musician, King Sunny Ade, in Essen Germany. The musician instructed his drummer to drum the English phrase "hello ladies and gentlemen" with a dùndún drum[8]. Considering that English, unlike Yorùbá, is not a tone language, it would be interesting to investigate the use of a tone-based speech surrogate on non-tonal languages.

The limitation of the present study is that it is based on laboratory-like conditions in the studio. As a result of this, the result might not represent speech-surrogate data in natural musical performance. The study is also limited to words and short phrases. Future research on the language of talking drum should compare the representation of Yorùbá phrases in direct speech mode and musical speech mode.

To conclude, this study shows that Yorùbá drummers represent lexical tones, grammatical tones and the phonetic realisation of these tones with a talking drum. The relationship between the drum tones and the corresponding speech tones is positively strong. Just as in speech, V and CV units have different status in drumming.

[8]The footage of the performance can be found in this YouTube video: https://youtu.be/1nNE5CL1VEw?t=703

A STIMULI FOR GENITIVE CONSTRUCTIONS

The Yorùbá forms, which were drummed for the five patterns in **Figure 11**, are presented in this appendix.

1.	HH	pákó o Ṣọlá	'Sọla's plank'
		pákó o Yọ̀mí	'Yomi's plank'
		pákó o Délé	'Dele's plank'
2.	LL	ìlù u Ṣadé	'Ṣade's drum'
		ìlù u Gbàmílà	'Gbamila's drum'
3.	MH	ọkọ́ ọ Bọ̀dé	'Bọde's hoe'
		ọkọ́ ọ Ṣọlá	'Sọla's hoe'
4.	ML	ọkọ̀ ọ Bọ́lá	'Bọlá's vehicle'
		ọkọ̀ ọ Ṣọlá	'Sọla's vehicle'
5.	MM	ọkọ ọ Bídèmí	'Bidemi's husband'
		ọkọ ọ Bọ̀dé	'Bọde's husband'

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Behavioural REB, The University of British Columbia. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SA collected designed the stimuli, collected the data, analysed the data and wrote the results.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.650382/full#supplementary-material

REFERENCES

Agawu, K. (2001). African Music as Text. *Res. Afr. Literatures* 32 (2), 8–16. doi:10.2979/ral.2001.32.2.8

Agawu, V. K. (2016). *The African Imagination in Music*. Oxford: Oxford University Press.

Agawu, V. K. (1988)., 58. *Tone and Tune: The Evidence for Northern Ewe Music*. Africa: Journal of the International African Institute, 127–146. doi:10.2307/1160658

Akinbo, S. (2019). Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis. *LLA* 11–23.

Akinlabi, A., and Liberman, M. (1995). "On the Phonetic Interpretation of the Yoruba Tonal System," in *Proceedings of the International congress of Phonetic Sciences*. Editors K. Elenius and P. Branderud (Stockholm: Stockholm University), 42–45.

Akinlabi, A., and Liberman, M. (2000). "The Tonal Phonology of Yoruba Clitics," in *Clitics in Phonology, Morphology and syntaxBirgit Gerlach and Janet Grijzenhout* (Amsterdam: John Benjamins Publishing), 31–62.

Akinlabi, A., and Liberman, M. (2001). Tonal Complexes and Tonal Alignment. *Proceedings-nel* 31, 1–20.

Akinlabi, A. (1985). *Tonal Underspecification and Yoruba Tone. PhD Thesis*. Ibadan: University of Ibadan.

Akinlabi, A. (1993). Underspecification and the Phonology of Yoruba/r. *Linguistic Inq.*, 139–160.

Archangeli, D., and Douglas, P. (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.

Arewa, O., and Adekola, N. (1980). Redundancy Principles of Statistical Communications as Applied to Yoruba Talking-Drum. *Anthropos* 1./2, 185–202.

Awobuluyi, O. (1978). *Essentials of Yorùbá Grammar*. Ibadan: Oxford University Press Nigeria.

BattaBox (2017). Amazing: How to Make a Yoruba Talking Drum. URL: https://www.youtube.com/watch?v=Z4D_mP6wsZg.

Beier, U. (1954). The Talking Drums of the Yoruba. *Amj* 1, 29–31. doi:10.21504/amj.v1i1.223

Blench, R. (2019). *Niger-congo: An Alternative View*. Manuscript.

Boersma, P. (2001). Praat, a System for Doing Phonetics by Computer. *Glot Int.* 5 (9–10), 341–345.

Bradley, D. (1979). "Speech through Music: the Sino-Tibetan Gourd Reed-Organ," in *Bull. Sch. Oriental Afr. Stud.* (London: University of London), Vol. 42, 535–540. doi:10.1017/s0041977x00135773

Cooper, A., and Wang, Y. (2012). The Influence of Linguistic and Musical Experience on Cantonese Word Learning. *The J. Acoust. Soc. America* 131 (6), 4756–4769. doi:10.1121/1.4714355

Diehl, R. L., and Lindblom, B. (2004). "Explaining the Structure of Feature and Phoneme Inventories: the Role of Auditory Distinctiveness," in *Speech Processing in the Auditory System*. Editors S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (New York: Springer-Verlag), 101–162.

dijiaderoGBA (2018). Ìlù Gángan (Talking Drum) – a Tale of a Drummer. URL: https://www.youtube.com/watch?v=ClUx_wN1IZI.

Durojaye, C. (2020). The Dùndún Talking Drum of the Yorùbá Community in South-West Nigeria. *J. Arts Humanities* 9 (7), 11–19.

Ekwueme, L. N. (1974). Linguistic Determinants of Some Igbo Musical Properties. *J. Afr. Stud.* 1 (3), 335–353.

Eluyefa, D. O. (2011). *Recontextualisation of the Dùndún Drumming Tradition in HampshirePhD Thesis*. Winchester: University of Winchester.

Euba, A. (1990). *Bayreuth: Bayreuth African Studies*.Yorùbá Drumming: The Dùndún Tradition

Euba, A. (1967). Multiple Pitch Lines in Yoruba Choral Music. *J. Int. Folk Music Counc.* 19, 66–71. doi:10.2307/942189

Goldsmith, J. A. (1976). *Autosegmental Phonology. PhD Thesis*. Cambridge, MA: Massachusetts Institute of Technology.

Gottfried, T. L. (2007). "Music and Language learningLanguage Experience in Second Language Speech Learning," in *Honor of James Flege*. Editors O-S. Bohn and M. J. Munor (Amsterdam/Philadelphia: John Benjamins Publishing Company), 221–237. doi:10.1075/lllt.17.21got

Hayes, B. (1989). Compensatory Lengthening in Moraic Phonology. *Linguistic Inq.* 20 (2), 253–306.

Hombert, J-M. (1977). Consonant Types, Vowel Height and Tone in Yoruba. *Stud. Afr. Linguistics* 33, 173–190.

Horton, C. D. J. (1997). Educational Resources on Traditional African Music: An Annotated Bibliography of Contemporary Offerings and Suggestions for Their Use in the Classroom. *J. Negro Edu.* 66, 16–28. doi:10.2307/2967248

Hyman, L. M. (1985). *A Theory of Phonological Weight*. Dordrecht: Foris.

Hyman, L. M. (2018). What Tone Teaches Us about Language. *Language* 94 (3), 698–709. doi:10.1353/lan.2018.0040

Kaminski, J. S. (2008). Surrogate Speech of the Asante Ivory Trumpeters of Ghana. *Yearb. Traditional Music* 40, 117–135.

Kassambara, A. (2018). *ggpubr:"ggplot2" Based Publication Ready Plots*. R package version 0.1.7. URL: https://CRAN.R-project.org/package=ggpubr.

Laniran, Y. O., and Clements, G. N. (2003). Downstep and High Raising: Interacting Factors in Yoruba Tone Production. *J. Phonetics* 31 (2), 203–250. doi:10.1016/s0095-4470(02)00098-0

Lawal, I. O., Uzokwe, N. E., Abi IgboanugoAdio, A. F., Awosan, E. A., Jo, N., et al. (2010). Ethno Medicinal Information on Collation and Identification of Some Medicinal Plants in Research Institutes of South-West nigeria. *Afr. J. Pharm. Pharmacol.* 4 (1), 1–7.

Leben, W. R. (1973). Suprasegmental Phonology. *PhD Thesis*. Cambridge, MA: Massachusetts Institute of Technology.

Liljencrants, J., Lindblom, B., and Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48 (4), 839–862. doi:10.2307/411991

Lo-Bamijoko, J. N., and Joy, N. (1987). Classification of Igbo Musical Instruments, Nigeria. *Amj* 6 (4), 19–41. doi:10.21504/amj.v6i4.1259

McCarthy, C. F. (1986). Enzymes in the Small Intestine. *Curr. Opin. Gastroenterol.* 2 (2), 207–211. doi:10.1097/00001574-198603000-00006

McPherson, L. (2019). Musical Adaptation as Phonological Evidence: Case Studies from Textsetting, Rhyme, and Musical Surrogates. *Lang. Linguistics Compass* 13 (12), 1–15. doi:10.1111/lnc3.12359

McPherson, L. (2018). The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. *Anthropological Linguistics* 60 (3), 255–294.

Nketia, J. H. K. (1963). *Drumming in Akan Communities of Ghana*. Edinburgh: Thomas Nelson and sons.

Nketia, J. H. K. (1970). Music Education in Africa and the West: We Can Learn from Each Other. *Music Educators J.* 57 (3), 48–55. doi:10.2307/3392906

Nzewi, M. (1999). Strategies for Music Education in Africa: Towards a Meaningful Progression from Tradition to Modern. *Int. J. Music Edu.* 33 (1), 72–87. doi:10.1177/025576149903300109

Oehrle, E. (1991). Emerging Music Education Trends in Africa. *Int. J. Music Edu.* 18 (1), 23–29. doi:10.1177/025576149101800103

Oludare, O. E., Layiwola, D., and Olusegun, M. (2018). "Yoruba Traditional Musical Instruments and Gender Stereotyping: Forms, Functions and Factors," in *FELA after 1997: Challenges, Justification and Lessons for Africa's Development (Proceeding of the 2017 Fela Kuti International Conference on African Homelands and Social theory) Biodun Olamusu. Ibadan* (Ibadan: Global Image Books and Publisher), 181–202.

Orie, O. O. (2006). "L2 Acquisition and Yoruba Tones: Issues and Challenges," in *36th Annual Conference on African Linguistics*. Editors O. F. Arasanyin and M. A. Pemberton (Somerville, MA: Cascadilla Proceedings Project), 121–128.

Orie, O. O. (2000). Syllable Asymmetries in Comparative Yoruba Phonology. *J. Linguistics* 36 (01), 39–84. doi:10.1017/s0022226799008130

Orie, O., and Pulleyblank, D. (2002). Yoruba Vowel Elision: Minimality Effects. *Nat. Lang. Linguistic Theor.* 20 (1), 101–156. doi:10.1023/a:1014266228375

Ọmóbọ́lá, A. A. (2019). Colonization and Cultural Values of Yorùbá People: A Case of Traditional Drums in Yorùbá Land. *J. Lang. Lit.* 19 (2), 148–156.

Poss, N. F. (2012). *Hmong Music and Language Cognition: An Interdisciplinary Investigation. PhD Thesis*. Columbus, OH: The Ohio State University.

Poss, N. F. (2005). *The Communication of Verbal Content on the Hmong Raj: An Ethnographic Analysis of Performance Practice*. Ma thesis: The Ohio State University.

Pulleyblank, D. (2009). "Patterns of Reduplication in Yoruba," in *The Nature of the Word: Essays in Honor of Paul KiparskyKristin Hanson and Sharon Inkelas* (Cambridge, MA: MIT Press), 311–357.

Pulleyblank, D. (1986). *Tone in Lexical Phonology*. Dordrecht, Holland: D. Reidel Publishing Company.

Pulleyblank, D. (2004). A Note on Tonal Markedness in Yoruba. *Phonology* 21 (3), 409–425. doi:10.1017/s0952675704000326

Riebold, J. (2013). Vowel Analyzer. Praat Script. URL: https://github.com/jmriebold.

Rolle, N. R. (2018). *Grammatical Tone: Typology and Theory*. Berkeley: PhD Thesis, University of California.

Ruskin, J. D. (2013). *The Yorùbá Dùndún in Local and Transnational Perspective: A Cosmopolitan Tradition in the makingPhD Thesis*. University of California Los Angeles.

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing Language to Rhythm: Amazonian Bora Drummed Language Exploits Speech Rhythm for Long-Distance Communication. *R. Soc. Open Sci.* 5 (4), 170354. doi:10.1098/rsos.170354

Stern, T. (1957). Drum and Whistle "Languages": An Analysis of Speech Surrogates. *Am. Anthropologist* 59, 487–506. doi:10.1525/aa.1957.59.3.02a00070

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code*. Surrey, England: Ashgate Publishing Limited.

Ward, I. C. (1952). *An Introduction to the Yoruba Language*. Cambridge: V. Heller and Sons Ltd.

Waterman, C. A. (2000). “Yoruba Popular Music,” in *The Garland Handbook of African Music*. Editor R. M. Stone (Oxford: Routledge), 169–185.

Wayland, R., Herrera, E., and Kaan, E. (2010). Effects of Musical Experience and Training on Pitch Contour Perception. *J. Phonetics* 38 (4), 654–662. doi:10.1016/j.wocn.2010.10.001

Westphal, E. (1948). Linguistics and the African Music Research. *Afr. Music Soc. Newsl.* 1 (1), 15–21.

Whalen, D. H., Gick, B., Kumada, M., and Honda, K. (1999). Cricothyroid Activity in High and Low Vowels: Exploring the Automaticity of Intrinsic F0. *J. Phonetics* 27 (2), 125–142. doi:10.1006/jpho.1999.0091

Whalen, D. H., Levitt, A. G., and Levitt, A. G. (1995). The Universality of Intrinsic F0 of Vowels. *J. Phonetics* 23 (3), 349–366. doi:10.1016/s0095-4470(95)80165-0

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Winter, Y. (2014). On the Grammar of a Senegalese Drum Language. *Language* 90 (3), 644–668. doi:10.1353/lan.2014.0061

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yoshida, K. (2011). Contrast Maximization in F0 Declination: Japanese Shiki-Accent Dialects. *ICPhS*, 2228–2231.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 December 2021
doi: 10.3389/fpsyg.2021.674289

A Flute, Musical Bows and Bamboo Clarinets that "Speak" in the Amazon Rainforest; Speech and Music in the Gavião Language of Rondônia

Julien Meyer[1] and Denny Moore[2]*

[1] *Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France,* [2] *Museu Paraense Emílio Goeldi, Ministério de Ciência, Tecnologia e Inovação, Belém, Brazil*

OPEN ACCESS

Edited by:
Laura McPherson,
Dartmouth College, United States

Reviewed by:
Morgan Sleeper,
Macalester College, United States
Annie Rialland,
UMR 7018 Laboratoire de Phonétique et Phonologie (LPP), France

***Correspondence:**
Julien Meyer
julmeyer.lab@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *01 March 2021*
Accepted: *01 November 2021*
Published: *21 December 2021*

Citation:
Meyer J and Moore D (2021) A Flute, Musical Bows and Bamboo Clarinets that "Speak" in the Amazon Rainforest; Speech and Music in the Gavião Language of Rondônia. Front. Psychol. 12:674289. doi: 10.3389/fpsyg.2021.674289

The Gavião, a native Amazonian group in Rondônia, Brazil, use three different traditional musical instruments that they identify as "speaking" ones and that are characterized by a very tight music-lyric relation through similar pitch patterns: a flute (called *kotiráp*), a pair of mouth bows (*iridináp*), and three large bamboo clarinets (*totoráp*), played by three different players, each one playing a single-note clarinet. They show in different ways the relation of acoustic iconicity which exists between the words of the songs' lyrics and the music played on such instruments to "sing" the songs. Linguistic analysis makes it possible to understand the phonetic and phonological nature of the iconicity. The sung speech form, being intermediate between the spoken and the instrumental forms, is useful for both learning and explaining the musical notes. In a language with distinctive tone and length, such as Gavião of Rondônia, the first question about speech that is played by musical instruments is the relation between the melodies and the supersegmental phonology of the corresponding words in sung speech and in modal spoken speech. It is influenced by the phonological possibilities of the spoken form and by the musical possibilities of the instrumental form. The description and analysis of Gavião instrumental speech and song practices are found to be a noteworthy contribution to the typology of instrumental language surrogates associated with a tone language, one that calls for a reexamination of hypotheses about which aspects of the phonological/phonetic structure can be transposed in instrumental speech and how this can be done. The role of this kind of instrumental sung speech is artistic and also practical as it contributes to maintain the oral heritage. Such practice represents a little-studied and threatened cultural heritage of the traditional substratum of the cultures of Amazonia.

Keywords: Gavião of Rondônia, speech-music relation, Instrumental speech, talking musical instruments, tone language, archaic speech, musical acoustics, speech surrogates

INTRODUCTION

Play, Sing, Speak, Sound Structures in Common

Playing musical instruments and singing are considered as independent activities in some cultures/traditions. In fact, songs can exist which are not accompanied by played music and played music can exist without lyrics. However, there are various possible relations between the two. For example, in many musical styles of the world, there is music which has an associated song, even if the song and the music are not always produced together. In various cultures, there is such a strong link between played music and its associated lyrics that they share several sound structures. Among the Gavião of Rondônia ("gavião" means "hawk" in Portuguese, which is a direct translation of the self-denomination "*ikṍló*"—the plural "*ikṍlééỳ*" is often preferred by the community) living in the western Amazon in Brazil, a subset of the musical traditions is based on such a strong link. There exist three types of instrumental music—corresponding to three different types of musical instruments: a pair of musical mouth bows, a flute and an ensemble of three one-note bamboo clarinets—in which the instrumental melodies are based on the structure of the language via the emulation of some aspects of the lyrics. The perception of the Gavião is that the instruments are speaking, or, more exactly, that they are expressing the sung form of speech. From this comes the idea of "singing" instruments. The songs are taught to facilitate learning the instrumental music, but the players rarely use them on other occasions. Hearing the melodies played, non-Gavião-speaking observers don't suspect the relation that they have with the associated lyrics.

Singing Instruments in the Amazon and in the World

Present in various parts of the world, the phenomenon of instruments which "sing" (imitate aspects of the phonetics and phonology of the words associated with the music) represents a verbal art of traditional musical instruments. This practice has been described for a very large diversity of instruments, ranging from percussion instruments (such as bells, drums, gongs and traditional xylophones; e.g., Sebeok and Umiker-Sebeok, 1976; Zemp and Soro, 2004; McPherson, 2018) to wind instruments [such as flutes (e.g., Ritzenthaler and Peterson, 1954; Hill, 1993; Mindlin et al., 2001), mouth organs (e.g., Catlin, 1982), bamboo clarinets (e.g., Moore and Meyer, 2014), etc.] and to several types of string instruments [such as mouth bows (e.g., Arom, 1970; Yegnan-Touré, 2008, Moore and Meyer, 2014), or even traditional guitars (e.g., Wedekind, 1983)]. Such instrumental emulation of sung speech has been little studied in Latin America, though the Amazonian basin is one of the rare areas with a great diversity of indigenous languages still expressed in this manner. Some publications indicate that this particular practice was observed in the past in the region (Whiffen, 1915; Izikovitz, 1935), while other more recent references show that it still exists in different forms at least in the cultures of the Bora (Thiesen, 1969; Meyer et al., 2012), Cinta Larga (Ermel, 2004), Gavião (Mindlin et al., 2001; Meyer and Moore, 2013; Moore and Meyer, 2014), Kuikuro (Franchetto and Montagnani, 2011), Pirahã (Everett, 1983), Wakuenai (Hill, 1993), and Aikanã (van der Voort, 2016).

Considering other parts of the world, the singing mode of musical instruments was largely studied on the west coast of Africa and is recognized as one of the important characteristics of the musical heritage of the cultures of this region (Carrington, 1949; Nketia, 1976; Sebeok and Umiker-Sebeok, 1976; Locke and Agbeli, 1981; Zemp and Soro, 2004; Kawada, 2014; Winter, 2014; McPherson, 2018; Dentel and Meyer, 2020). This linguistic-musical phenomenon has also been indicated as very common in the southeast of Asia (Stern, 1957; Catlin, 1982; Poss, 2005; Meyer, 2007), and in Papua New Guinea (Niles, 2010), with various types of instruments in each place. The notion of a *singing mode* of the musical instruments was clearly defined by Nketia (1976) to distinguish, in the Akan culture of Africa, the beats of drums used for telecommunication (*speaking mode* imitating spoken speech), from the beats tied to the songs (*singing mode* imitating sung speech). This distinction between speaking mode and singing mode of the drums also exists in Amazonia, and it has been mentioned for example in the culture of the Bora (Thiesen, 1969; Meyer et al., 2012; Seifart et al., 2018). For the Gavião of Rondônia, an equivalent distinction exists between whistled speech, which serves to imitate spoken speech in the case of long-distance communication, and instrumental speech, which imitates songs presented in this study (and see also Meyer and Moore, 2013; Moore and Meyer, 2014; Meyer, 2015).

Context and Objectives of the Present Study

Although the works cited above represent an important advance, they are still few, and many questions remain without answers about the linguistic functioning of these speech systems. In the case of the Amazonian languages, there are hardly any systematic studies in scientific linguistics about this subject. This fact alone is sufficient to justify the development of an initiative to document sung instrumental speech. Also, an important fact is that, in general, instrumental forms of speech are lost before the forms of speech which only involve the voice. In the case of the Gavião of Rondônia, this situation is worrisome, since it is symptomatic of the threat which weighs on all the oral heritage.

Given the reduced degree of transmission between generations, the Gavião who care about the maintenance of traditional culture think that it is important to have scientific accompaniment for preservation. The capacity to play music which imitates the sounds of its associated song requires a complex understanding, involving the control of various different techniques, for example, the manufacture of the instruments, the learning of the vast repertoire of songs or the playing of musical notes corresponding to words. Sometimes it also requires coordination between different people involved in the music or the simultaneous performance of dances by the players in the context of a festival. Aside from this, the texts regularly make reference to the traditional cosmology of the Gavião and to the relation between the people and the Amazonian environment; participating in its living day-to-day activities. Because of this, these cultural practices suffer invasive and destructive pressure from religions that have been recently imported, who denigrate them and, at times, prohibit them. By substituting texts of traditional beliefs by other texts, the non-native religions eliminate traditional forms of indigenous arts, aside from the traditional mythical vision of the natural and supernatural world of the people.

As part of the response to this urgency, both scientific and cultural, the present article provides a general view of the process of documenting three principal musical instruments identified as "singing" by the Gavião collaborators and describes the relation of acoustic similarity which exists between the musical melodies and the corresponding words in their associated songs, or even in modal spoken speech. Linguistic analysis is important for understanding the phonetic and phonological nature of this acoustic iconic relation, which is one of the bases of the creativity of the musical memory of the Gavião people. Their instrumental speech and singing show previously undescribed typological properties that have scientific importance.

METHODOLOGY

Steps of the Research

The multidisciplinary methodology that was adopted registered various complementary aspects of the traditional use of the three instrumental speech forms under study. It was carried out in various villages on the Indigenous Territory Igarapé Lourdes in the state of Rondônia in western Brazil. Each step was observed and recorded in video and/or audio. The steps can be replicated and adapted to other cultures with similar practices.

The first step was to identify the instruments involved by making an ethnographic enquiry with various traditional indigenous collaborators in various villages. After the identification of the instruments and of the few players regarded as skillful representative performers, the second step consisted of documenting the players' preparation for playing in several villages: gathering the materials in the forest, manufacturing the instruments, and tuning the notes that they produce.

In the third step, documentation was carried out of each musical piece and its associated song in various ways, each focusing on a different aspect. First, the instrumental music was played, recorded, and described in its natural context, or, at least in a simulation of the natural context. Second, the associated song, without the instrumental music, was sung, recorded, and transcribed. The content of the lyrics was described ethnographically. Fourth, the lyrics of the songs were recorded in their modal spoken form (not sung). They were then transcribed and glossed as texts using methods which presuppose a reasonably complete description of the language as well as the help of an indigenous consultant to pronounce and translate the words, to respond to analytic questions, and to help identify some aspects of the language used for singing, which is different from modal spoken use. One problem in this step is the difficulty in suppressing the impulse of the informants to sing the lyrics instead of speaking them. Another interesting difficulty is identifying and understanding the archaic or rare forms in the lyrics that are quite common in old songs. Lastly, the three melodies were compared, that of the music played by instruments, that of the associated song, and that formed by the tone and length of the associated song lyrics as they are normally spoken.

Three other steps were: (1) storing the data in two professional linguistic archives [one international (Meyer, 2014) and one national (Moore and Meyer, 2015)], (2) editing video documents for the Gavião villages to return the data to the community in a ready-to-use form (edited documents on CDs, DVDs, pen drives, printed forms that can be used in families, schools, associations), and (3) making the results of the study primarily accessible both locally to the Gavião community—in accordance with the fieldwork authorizations delivered (see next section "Ethics and Permits")—and nationally to the Brazilian community by publishing them in Portuguese (except young Gavião children and very old people, all speak some Portuguese) (Meyer and Moore, 2013). These steps were important because these singing Gavião musical instruments represent a sociocultural patrimony threatened by extinction. It has equivalents in other Amazonian cultures that are rarely documented due to the fact that they are only maintained in rather isolated cultural communities. Another objective was therefore to stimulate the investigation of instrumental speech by Brazilian linguists and anthropologists. Finally, concerning the Gavião community, the collaborative work was well accepted, especially among the players, who saw their traditional knowledge gaining prestige; among the young people, who recorded sounds and videos and followed the process of edition of DVDs; and among the children, who learned the ancient songs and followed the manufacturing of the traditional instruments. It was also an opportunity to interact with indigenous school teachers about the tone in their language.

Ethics and Permits

Research was conducted in accordance with the Declaration of Helsinki. In Brazil, there was at that time no standing ethics committee at the Museu Goeldi, our host institution in Belém. Ethical questions, should they arise, were addressed by an internal investigative committee called a "sindicância." In Brazilian law and practice, the participating indigenous community indicates, either orally or in writing, their informed consent to the proposed research to the local office of the National Indian Foundation (FUNAI), which in turn transmits that consent, in the form of a document, to the central FUNAI office in the national capital. This office issues written research permits. Our research followed these established procedures. Native local authorities authorized our work in all of the visited communities. Permits were obtained from the National Indian Foundation (FUNAI) and the National Research Council (CNPq). Copies of the recordings and the research results were given to the community.

Some Phonetic and Phonological Characteristics of the Language of the Gavião of Rondônia

It is essential to know some technical aspects of the language of the Gavião to understand the relation between modal spoken speech, the song and the instrumental music associated with each song. Basically, all the Gavião (population of approximately 800 people) speak the language, which is part of the Mondé branch of the Tupi family. In the phonology there are 18 consonant phonemes, not much different from the consonants of English. There are only five contrastive vowels. Nasalization is contrastive on vowels and spreads to the right under certain conditions. There are no stress contrasts (Moore, 1984). The IPA symbols for the consonant phonemes are [*p, t, c, k, ʔ, b, d, ɟ, g, m, n, ɲ, ts, dz, β, ɾ, l, j*]; for the vowels [*i, e, a, o, ɨ*]. In the practical transcription used here, the symbols *c* and *j* denote palatal stops, variably affricated, *y* the palatal glide, and *s* and *z* dental affricates. The voiced bilabial fricative is indicated by *v* and the glottal stop by an apostrophe. The high central vowel is indicated by *u*.

The supersegmental phonology is complicated and quite complex to represent in practical transcription. There are contrasts between high, low and rising tones and also between short and long vowels. Long vowels are phonologically different from two short vowels in the same syllable, though they are phonetically the same. Also, some long syllables have a floating low tone finally. A more detailed analysis of the tonal system is given in Moore (1999) and Moore and Meyer (2014). The transcription used here, originally adapted to the Brazilian keyboard and carried over, for the sake of consistency, from earlier stages of phonological analysis, is adequate for our purposes.

In our transcription low tone is unmarked; high tone is marked by an acute accent and rising tone by a circumflex. Long vowels and sequences of two vowels in the same syllable are both transcribed as sequences of two vowels. In the transcribed examples below all syllables with two equal vowels with low tone are long vowels, not short vowel sequences, for example, *pagátaa* "cut us." A grave accent is used on the second vowel to indicate long tones which have final floating low tones. The common short and long tones and their phonetic manifestation are given below:

Low tone	sep	[tsēp]	"leaflike object"
High tone	sép	[tsep]	"hairlike object"
Short syllable	aka	[ākā]	"kill"
Long syllable	aa-kaa	[ā:kā:]	3-go

Some long tones can be realized phonetically as contours, for example, the long rising tone or the two long tones which drop as the floating low tone attaches to the end of the syllable when nothing follows:

Raising tone	ãá	[ã:]	"this"
Falling tone	kîit	[ki:t]	"white"
Rising-Falling tone	pêèp	[pe:p]	"black"

Another complication is that certain syllables (those with an underlying long syllable with a final floating low tone) trigger a fall in the phonetic level of any immediately following high tone to a mid-level. This lowers the register of all following high tones. If nothing follows the syllable, the floating low tone attaches to the end of the syllable and the result is a falling tone, for example: H:(L) > HL. If the syllable is followed by a high tone, the floating low tone attaches to the following high tone and the combination is realized phonetically as a mid-level, for example: H:(L) H > H:[!] H. So this *downstep*, which is well known in Africa, occurs only in these cases in Gavião and the language does not display downdrift, i.e., H pronounced lower after an overt L tone. Downstep tones, which are systematically rendered by instrumental speech in Gavião (see details in section "Correspondence Between Musical Notes and Phonetic Tone—Playing Flattened Surface Tone") are quite frequent in Gavião: the final tones of adjectives are generally downstep tones; long final tones in nouns are usually downstep tones, except in monosyllable nature items, where they are level. Long final verb tones are also level, with no floating low tone.

Without going into the details of Gavião tonal phonology, it is worth noting that another phenomenon that is relevant for instrumental singing is low-tone dissimilation, which follows the pattern of the Obligatory Contour Principle (Odden, 1986). Long low tones before immediately following low tones, across word boundaries, optionally (but frequently) become long rising tones: L: L > LH L. This is another postlexical tone alternation which is played by instruments (see details in the Results section "Correspondence Between Musical Notes and Phonetic Tone—Playing Flattened Surface Tone") as well as spoken and sung.

Speakers of languages without tone and length contrasts generally have considerable difficulty hearing the tone and length. The Gavião, however, easily hear the tone and the length, which have contrastive weight in their language. These phonological traits are the principal base of the whistled speech of the Gavião and also the iconic resemblance between speech, songs and instrumental music (Moore and Meyer, 2014; Meyer, 2015) as we will describe more thoroughly in the present paper.

RESULTS

During the first step of our inquiry we identified the three different types of instruments that have associated songs. We also verified that there is a large diversity of styles in the musical repertoire of the Gavião: for example, there are also songs that are not played on musical instruments, with a freer relation to linguistic pitch pattern. There is also music that is only played and not sung and that has therefore little to do with linguistic tone. Such music is played on other types of musical instruments (flutes or clarinets with different names in Gavião and different manufacturing processes, for instance). The recording and transcription sessions of music, associated songs and spoken lyrics confirmed the indigenous claim that the *iridináp* musical mouth bows (**Figure 1**), the *kotiráp* flute (**Figure 2**), and the *totoráp* clarinets (**Figures 3**, **4**) are expressing the language, and more precisely, the singing mode of speech. All the steps described in the methodology helped identify what

FIGURE 1 | The *iridináp* mouth bows are played exclusively by women. This musical instrument is based on the principle of rubbed strings: the player applies with the fingers three levels of tension to a single cord rubbed by the other bow. The mouth of the player is also used as a resonance chamber (Photos Laure Dentel/Julien Meyer).

FIGURE 2 | The *kotiráp* flute is an open bamboo stick with five holes: one at one extremity of the instrument in which the players blows and four that are used to produce four different notes (Photos Julien Meyer).

the relation is between the music, the song and the spoken pronunciation of the lyrics. The key to this relation is tone and length in Gavião phonology.

We first present here the contexts of use of each instrument and subsequently provide some representative examples of the repertoire of songs that was collected for each type of musical instrument. Next, we extend our pluridisciplinary approach with insights from musical acoustics to characterize the sounds played with each instrument, and with insights from linguistics to understand the music-language relation underlying the musical phrases.

Ethnographic Enquiry

Context of Use and Transmission

Each instrument has its specific contexts of use. For example, the *iridináp* bows are used by the girls and serve exclusively to express messages dealing with love (seduction, refusal and marriage). They are used in private contexts if the boyfriend is near, or during some festivals with all the girls playing and dancing in a line, each one with her instrument. The *kotiráp* flute serves for all types of poetry; it is played by boys and men. It is in general very common for the expression of sentiments of friendship or love or for the description of a distinctive event. It is used in private contexts or in festivals, sometimes to respond to the bows of the girls (Mindlin et al., 2001). Finally, the *totoráp* clarinets are used by boys and men for entertainment in festivals with drink, recounting important events in the life of the community, such as ceremonies, war, poetic impressions or hunting.

Although the children are still learning their maternal language, the instrumental forms of the language are used less and less. The words of these songs frequently refer to traditional knowledge and practices, which now are not well known by a large part of the Gavião population, owing to the reduced frequency of traditional cultural events. Also the Gavião collaborators say that few instrumental songs have been invented since the time of contact, in the decades of 1940–1950. Of the three instruments documented, only the *totoráp* continues with a certain vitality, since it is systematically linked to dances and festivals.

Up till the present, in spite of excellent cooperation with the community owing to confidence in the study of the language, initiated in the decade of the 1970s, only four songs with *iridináp* were found and recorded (with five informants) during the study reported here, eight songs with *kotiráp* (with five informants) and nine songs with *totoráp* (with three groups of three players).

All the informants of the census agree that few songs remain for *iridináp* and *kotiráp*. The eldest at the time of the enquiry

FIGURE 3 | The three bamboo clarinets that constitute the *totoráp* musical instrument are always played together while dancing or sitting. Each player plays a single note clarinet (Photo: Laure Dentel/Julien Meyer).

FIGURE 4 | Details of the reed (right), its tuning with a vegetal string and its insertion in the bamboo stick (left) of the *totoráp* clarinets (Photos: Julien Meyer).

(above 70 years old in 2009–2013), who perhaps still knew other songs, had difficulty in passing on the knowledge because they no longer had the force to blow an instrument or the motor coordination necessary to play one. Since much time has passed without playing the instruments of seduction, it is also difficult for the ladies to remember the less popular songs of the past.

Repertoire and Some Examples of Songs

The most surprising aspect of the music played by these instruments is that the Gavião spontaneously indicate that it has a meaning. By asking specific questions it is possible to get more details since there is nothing secret about the phenomenon. Generally the players state that the instruments speak words. If one insists, they frequently recount the story told by the lyrics of the music. Other times they sing them. After analysis we realized that the sung version is the key to the instrumental technique because it is exactly that which serves as a model for the music played by the instruments. As mentioned, a linguistic analysis of the words of the sung version shows how they correspond to the notes and rhythms of the played music (see next subsections).

Some examples of the songs are given below. They are short and repeated several times in a musical session. Different players may recombine the components/verses in differing ways, or sometimes delete some parts. Each instrument has its own repertoire of songs. We identified some ancient words in the lyrics and they are in italics in the texts (a), (b), and (c) below. Additionally the songs of *totoráp* and *iridináp* intersperse syllables without meaning (for example, *iriri*, *toy péréré* for the iridináp mouth bows, *sẽ́t sẽ́t, o sẽ́t o sẽ́t* for the totoráp clarinets) for rhythmic instrumental melodies between sung phrases. This characteristic is sometimes found in the songs of the *kotiráp* flute but only with some players (with *o sẽ́t o sẽ́t*). There are other rhythmic and esthetic processes linked to the tonal melody, which will be explained later in the article.

(a) The *iridináp* musical bows

The *iridináp* song which everyone knows is the following:

ãá bó *õbút* mága (2x)	This is mine.
iriri iriri	(instrumental rhythm)
ãá bó alimé *akút* mága (2x)	This (one) is a good hunter of spider monkeys.
iriri iriri	(instrumental rhythm)
ãá bó ikṍló *akút* mága (2x)	This (one) is a good hawk hunter.
iriri iriri	(instrumental rhythm)

Sometimes the melody is associated with another phrase which is interspersed between two repetitions of the preceding stanza:

amapit atíní atí mága ṍgay	This woman wants her daughter for herself, not for me.
iriri iriri	(instrumental rhythm)

Other songs make subtle reference to the prestige of the hunter, like this one:

toy péréré toy péréré	(instrumental rhythm)
ẽbéré boràát kõ̂òp ma'ã́ papá	Get medicine powder for the hunt, Uncle.
toy péréré toy péréré	(instrumental rhythm)

(b) The *kotiráp* flute

The *kotiráp* flute frequently speaks of love and marriage:

ṍgalí atí mápit *bósara* kî káre béá (2x)	I will lay down with her daughter again.
õot bó mága citá ẽmápit táá atíá	I will stay with your daughter, Madam.

aaná õóbiri biriá	Now below me. *(women sleep in a hammock under their husband)*
õot bó mága citá ẽmápit táá atíá	It is I who will stay with your daughter, Madam.
aaná õóbiri biriá	Now, below me.

It is also common for the words to refer to a myth (recounting here a summary of the myth of the end of the domestication of the boa constrictor).

aajaay va teé bay póòy máà áleá (2x)	Boa constrictor ate his owner
abóta pérúup mábuúre káá	because he failed to shoot the deer for him.
aajaay va eénaá	Ate his owner like that
abóta pérúup mábuúre káá	because he failed to shoot the deer for him.
aajaay va eénaá	Ate his owner like that

(c) The *totoráp* ensemble of three Clarinets

At this time the *totoráp* song refers to the historical event of the invasion of the Gavião land.

ãá bó zaat máà pagátaa	The white man invaded our land.
o sẽ́t o sẽ́t	(instrumental rhythm)
ãá bó zaat máà pagátaa	The white man invaded our land.
o sẽ́t o sẽ́t	(instrumental rhythm)
zérék kîit zérék kîit máà pagátaa	The white skin the white skin invaded our land.
o sẽ́t o sẽ́t	(instrumental rhythm)
zérék kîit zérék kîit máà pagátaa	The white skin the white skin invaded our land.
o sẽ́t o sẽ́t	(instrumental rhythm)
ẽzáká ẽebíí teé kay áleá zérék kîit kay áleá	Don't be afraid of him of the white skin.
o sẽ́t o sẽ́t	(instrumental rhythm)
ẽzáká ẽebíí teé kay áleá zérék kîit kay áleá	Don't be afraid of him of the white skin
o sẽ́t o sẽ́t	(instrumental rhythm)

The *totoráp* can also speak poetically to restore a natural environment:

ikábiit ábi ká bi ká coliléèy mága	The swallows fly above the creek.
sẽ́t sẽ́t	(instrumental rhythm)
ikábiit ábi ká bi ká coliléèy mága	The swallows fly above the creek.
sẽ́t sẽ́t	(instrumental rhythm)
zólṍp tîì zólṍp tîì kat mága	Pineapple, there is a patch of pineapple.
sẽ́t sẽ́t	(instrumental rhythm)
zólṍp tîì zólṍp tîì kat mága	Pineapple, there is a patch of pineapple.
sẽ́t sẽ́t	(instrumental rhythm)
ikãáybîìt ci tára	On the bank of the creek.
sẽ́t sẽ́t	(instrumental rhythm)
ikãáybîìt ci tára	On the bank of the creek.
sẽ́t sẽ́t	(instrumental rhythm)

Musical Acoustics

Acoustic Characteristics of the Three Instruments

The music played with the *iridináp*, *kotiráp*, or *totoráp* is characterized by the particular timbre of each instrument (**Figures 5–7**), by the melody of the notes played and by the rhythms of complex musical phrases interspersed with simple rhythms. We won't provide here an ethno-musicological description, which would be beyond the scope of the present article, but rather an acoustic description of the main points of interest for the music-language comparison. For a direct audio impression of the music of each instrument under study, some audio recordings corresponding to some of the figures illustrating this paper are provided in Meyer (2021), which serves as an audio annex.

First, each instrument has its own acoustic signature, characterized by its timbre. The *iridináp* is an instrument which—like a violin—is based on the principle of rubbed strings: the player applies three levels of tension to a single cord rubbed by the other bow and in this way produces three different musical notes. The mouth of the player is also used as a resonance chamber (**Figure 1**). The acoustic result is a fundamental frequency of 200–350 Hz and various harmonics equally reinforced (**Figure 5**). The *kotiráp* is a wind instrument, a flute with 4 holes corresponding to four different notes (**Figure 2**) with simple timbre, reduced to a sinusoidal signal in the band of 500–900 Hz with its harmonics (**Figure 6**). The *totoráp* clarinets constitute a wind reed instrument composed of three monotone clarinets (**Figures 3**, **4**), characterized by dense and complex harmonic sounds, with a fundamental frequency in the band of 100–200 Hz with its odd harmonics reinforced (**Figure 7**). Since the three instruments are harmonic, the height of each note played on any instrument is defined by its fundamental frequency (F0). Note that the perceptual attribute of F0, pitch, constitutes an independent entity parallel to timbre in human perception (Schouten et al., 1962; Risset, 1968; Schwartz and Purves, 2004).

It is through the perception of pitch that the players tune their musical instruments. The manufacturing of these three Gavião instruments is often completely public and a collective work, particularly for the *totoráp*, for which each participant –including the dancers who won't play the instruments—is temporarily specialized in preparing one of the elements of the three clarinets. All instruments are made with great attention but the manufacturers do not measure them with much precision. In the flute, the four holes are evenly spaced and realized by burning the bamboo with a wooden stick. The two bows of the mouth bow are approximately of the same size and are interchangeable. Concerning each clarinet of the *totoráp* ensemble, they consist of a tube and a reed. The three tubes have generally slightly different sizes. Since we worked in several villages and since our enquiry lasted several years, we had occasion to verify that for all types of

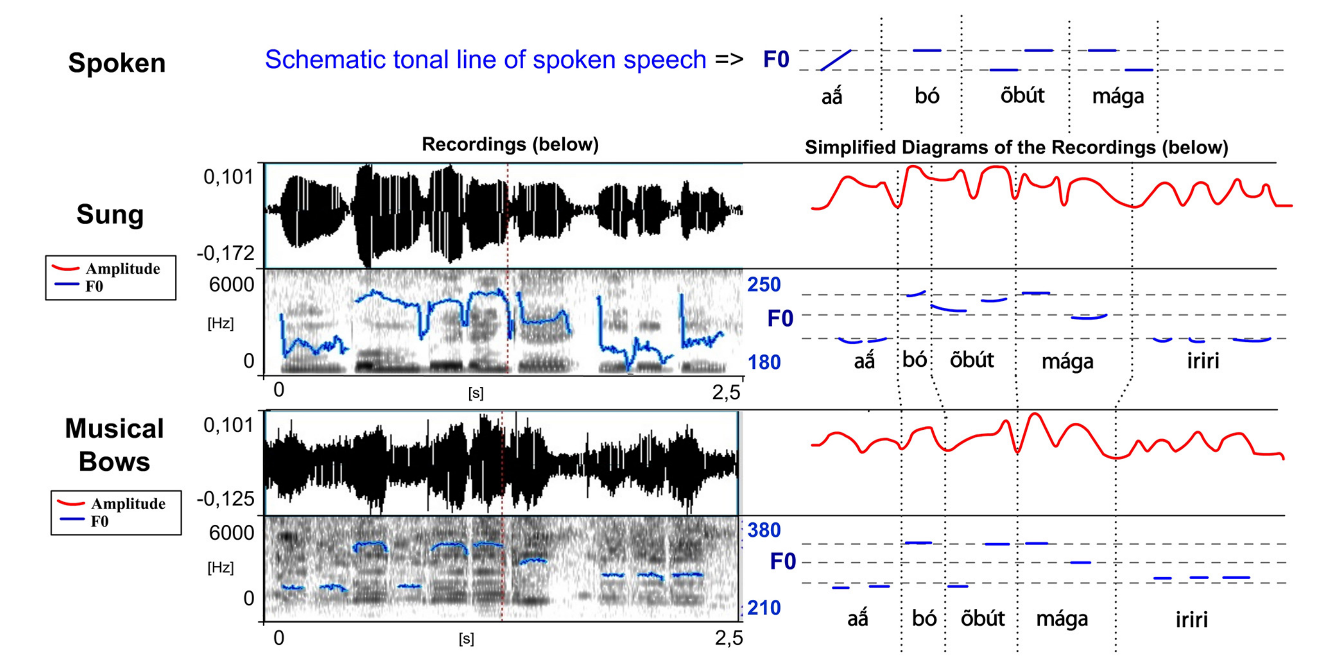


FIGURE 5 | Waveforms and spectrograms of the sung and musical forms of a love song extract played on the Gavião mouth bows (meaning "This [hunter] is mine," and where *iriri* is just a musical rhythm). The blue lines on the spectrograms represent the extracted F0 (the same parameters have been used for the sung and the instrumental form which explains some jumps in detection in the sung form). For this reason and to see more clearly the similarities between different speech types, they are reported in simplified charts (right), with the amplitude in a red line. The tonal line of the spoken form is also schematically represented (top) (adapted from Moore and Meyer, 2014; listen to sound extracts in Meyer, 2021).

instruments (flute, mouth bows and clarinets) a large variation of notes is tolerated. For the mouth bows and the flute, it is because the instruments can change in size and diameter depending of the pieces of wood that are chosen. For the flute, the longer the bamboo, the wider is the space between holes, with the fixed condition that the holes have equal spacing in the lower part of the bamboo. For the clarinets, what is the most important for determining the tonal height is the tuning of the reed, which imposes its frequency to the tube (this is well described for similar instruments among the Wayãpi by Beaudet, 1997). Moreover, differences between different ensembles of clarinets show that three clearly distinct notes are necessary for the instrument to be acceptable but it is not necessary that they are equally distant in frequency or that they keep the same relation from one ensemble to another. **Figure 7** shows three stable notes for one *totoráp* ensemble with F0s at respectively 130, 145, and 155 Hz. Other Gavião *totoráp* clarinet ensembles have been documented and show a different absolute distribution of notes each time. For example the clarinet ensemble recorded in the "totorap-AudioSupplement" file available in the audio annex of the present paper (see Meyer, 2021) is characterized by three stable notes at respectively 115, 126, and 134 Hz. This other ensemble is also the one recorded in Moore and Meyer (2014).

In musical acoustics, the form of the variation of amplitude is called "envelope of the wave." This refers to the graph of the amplitude of the signal, which shows the variation of amplitude as a function of time. The envelope graph of the melodies played by each type of instrument described here is quite similar to the envelope graph of the words of each associated song (and therefore also of their rhythm, which is strongly characterized by the peaks of sound energy), with some differences between the type of musical instruments. The *kotiráp* flute and the *iridináp* double-bow tend to reproduce the outlines of the amplitude graph of the associated song, with the *kotiráp* flute being more precise in this aspect (**Figures 5**, **6**). In the case of the *totoráp*, each of the three clarinets has a level of amplitude that is relatively stable and distinct from the others, which corresponds to the force of the wind applied by the player to vibrate the reed and produce the sound. As explained above, the reeds of the three clarinets are tuned differently: the less the vibrating part of the reed (sometimes adjusted with a vegetal string as shown in **Figure 4**), the greater the force necessary to make it vibrate and the higher the frequency of the sound which results from the vibration. This constraint in the *totoráp* thus influences and limits the possibilities of emulation of the amplitude envelope of the lyrics.

Adaptation of the Sung Mode to the Instruments

The sung forms corresponding to the melodies played with *iridináp, kotiráp* and *totoráp* have particular acoustic characteristics which indicate that they were standardized by the Gavião in order to be played on the instruments. The principle lines of evidence which converge in support of this observation are the following:

In principle, spoken and sung Gavião forms of speech have no limits on the modulation of the fundamental frequency. For

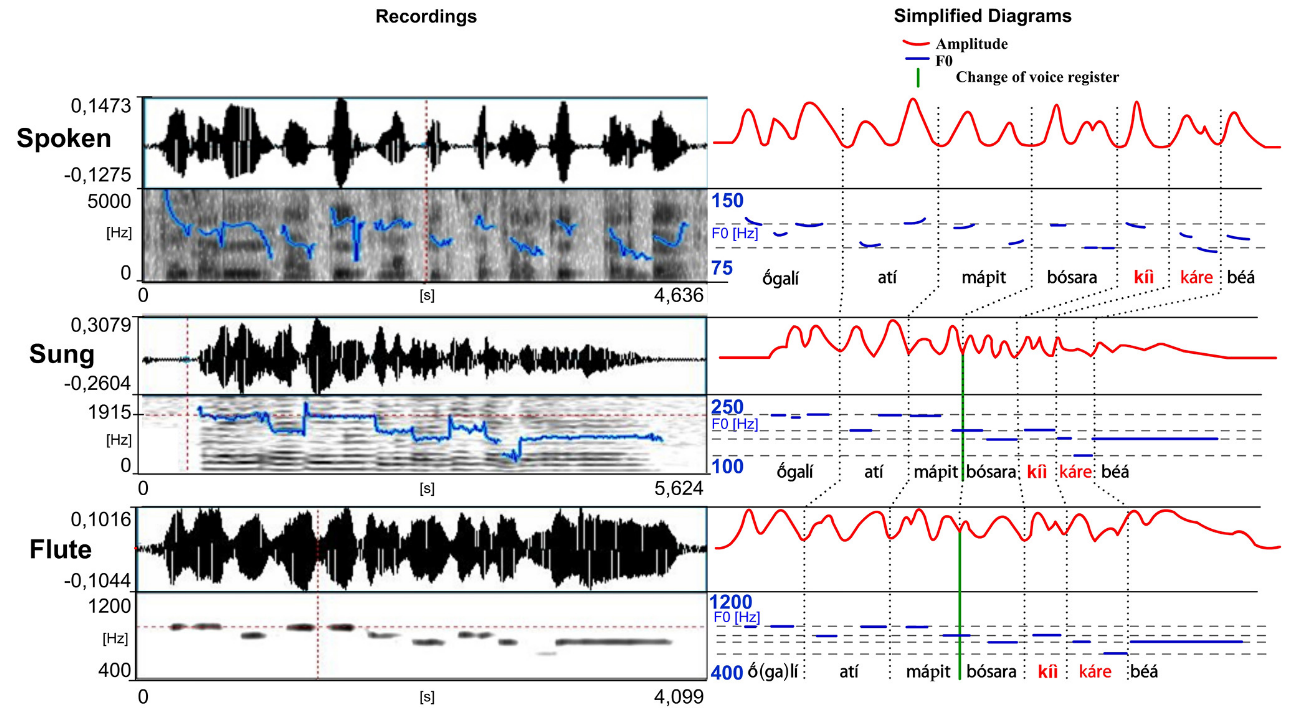


FIGURE 6 | Waveforms and spectrograms of the spoken, sung and musical forms of a song extract played on the *kotiráp* flute (meaning "I will lay down with her daughter again"). The blue lines on the spectrograms represent the extracted F0 (the same parameters have been used for the spoken and the sung speech types which explains more jumps in detection in the spoken form). To see more clearly the similarities between different speech types, they are reported with a focus on vowel nuclei in simplified charts (right), along with the amplitude in a red line. Note the tone downstep after *kîi* in *kîi káre béá*. Note also the change of voice register in the middle of the verse (delimited by a green line) (adapted from Meyer and Moore, 2013; listen to sound extracts in Meyer, 2021).

example, there are in Gavião culture genera of songs, such as those of the shaman or of festivals, which are not associated to the playing of a musical instrument and which use modulation of F0 both for musical ornamentation and for encoding the tones (Moore and Meyer, 2014). By contrast, the tones of the sung forms of the songs of *iridináp*, *kotiráp* and *totoráp* are realized without internal contours (modulations). For the three bamboo one-tone (one-note) clarinets, the impossibility of producing frequency modulations with such an instrument would explain why the associated songs produce sequences of flat notes (the rising spoken tone of the initial *ãá* of the verse *ãá bó zaat máà pagátaa* is played as flat low note (**Figure 7**) and the falling spoken tones of the repeated expression *zérék kíìt* are played as flat lower notes than the ones of the preceding word, see **Figure 8**). The music with flute and mouth bows shows the same absence of modulations, even in the associated sung forms, despite the fact that modulations are technically possible to produce with such instruments, at least micromodulations. Interestingly, collected data shows also that even with the flute, such modulations are not easy to produce. For example, when we asked the players to try to "speak with the flute" the common sentence *jaá pavíjíá* ("shall we go to bathe," as shown in **Figure 9**) that does not appear in any *kotiráp* flute traditional song, we observed that instead of producing a long note emulating the long rising tone spoken for *jaá*, they produced two level notes with a micromodulation in between (**Figure 9**, middle row). By contrast, when asked to produce the same sentence as it would be traditionally sung, they rather produced a high flat note (**Figure 9**, lower row). It appears that in the latter case relative height is being maintained in relation to the surrounding tones (the *jaá* is higher than the *pa*).

For the flute and the mouth bows, the Gavião players thus deliberately choose to focus instead on the production of level notes for singing the verses. This has important consequences for playing the tone in Gavião language, including adaptations of several tonal rules (see section "Correspondence Between Musical Notes and Phonetic Tone—Playing Flattened Surface Tone").

A second important aspect in relation to tone levels is that the sung form is adapted to the given number of different notes played on each instrument, as illustrated in **Figures 5–7** (three notes for the *totoráp* and *iridináp* instruments and four notes for the *kotiráp*). In the case of the *totoráp*, since each player represents a part of the instrument and executes only one note, this also influences the melody, with the goal of having participation by everyone in the phrase that is played (**Figures 7–9**). Notably, when a low tone is played on vocables, these low tones are played with the lower note of the instrument, as if it were a kind of style marker of nonsense vocables.

Some Specific Esthetic Effects

First, like all the types of Amazonian songs, the music played by the three analyzed instruments uses various types of repetition:

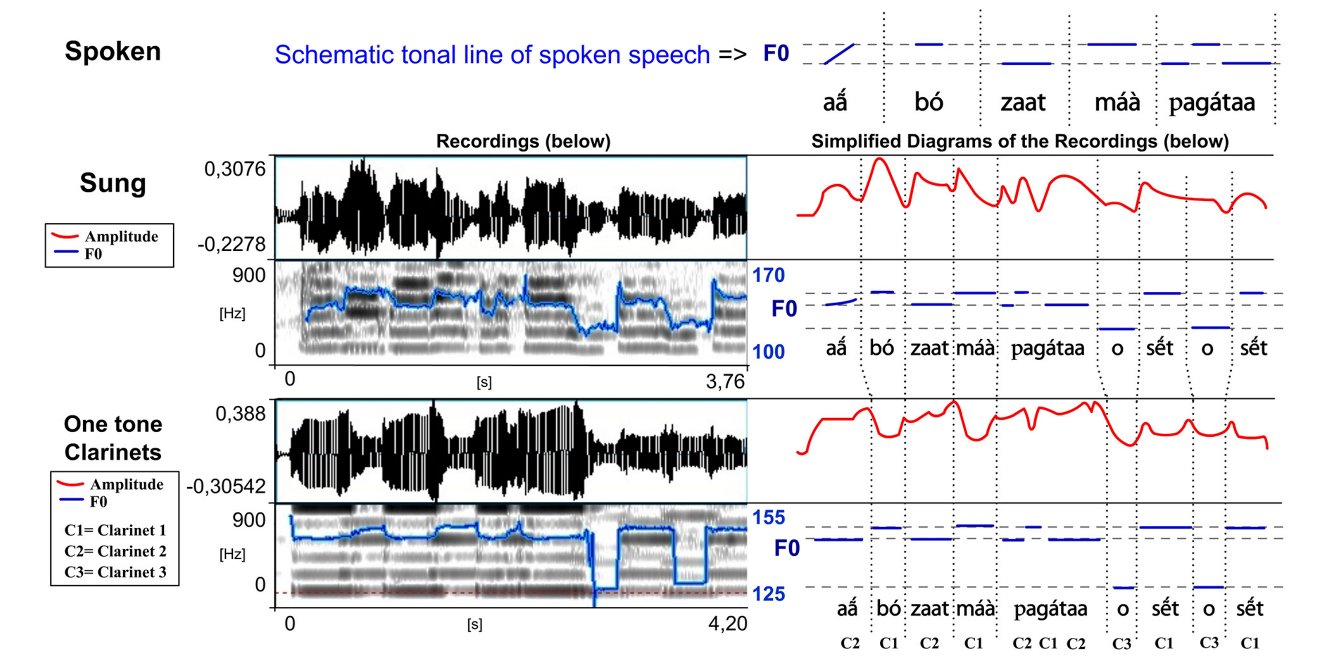


FIGURE 7 | Waveforms and spectrograms of the sung and musical forms of a song extract played on the *totoráp* clarinets (meaning "The white man invaded our land," and where *o sɛ̃́t o sɛ̃́t* is just a musical rhythm). The blue lines on the spectrograms represent the extracted F0 (the same parameters have been used for the sung and musical speech types which explains more jumps in detection in the sung forms). To see more clearly the similarities between different speech types, they are reported in simplified charts with a focus on vowel nuclei (right), along with the amplitude in a red line (adapted from Meyer and Moore, 2013; listen to sound extracts in Meyer, 2021).

- repetition of a whole phrase (several examples can be found in the music lyrics above in section "Repertoire and Some Examples of Songs");
- repetition of one part of a phrase, as in *ikábiit ábi ká bi ká*, but also in *zólɔ̃́p tîì zólɔ̃́p tîì* (in the second *totoráp* song of section "Repertoire and Some Examples of Songs") or in *zérék kîìt zérék kîìt* (see the first song of section "Repertoire and Some Examples of Songs" and the illustration in **Figure 8**).
- repetition of the whole song several times
- the music of these singing instruments also widely uses motif repetitions—a recurring set of notes with the same temporal intervals—which may be the result of different lyrics in the same song. That is to say, words are chosen which will repeat a musical pattern already established by other words (this is well illustrated by the **Figure 10**, where the three notes of *póòy máà á-leá* follow the same melody and rhythmic patterns as the three notes of *jaay va eé-naá*; see also details on how these examples are played in section "Correspondence Between Musical Notes and Phonetic Tone—Playing Flattened Surface Tone").

Another common esthetic technique found in these instruments is a sudden lowering or rising of the whole scale of notes used to transpose the tonal levels. Such a technique is akin to the effect of change of vocal register usually found in sung speech, which is characterized by frequency breaks between lower and upper registers [in classical singing, a person's vocal range is usually thought of in terms of different sections or registers, which arise from different types of vibratory patterns of the vocal chords and thus are categorized in different domains of frequency pitch which may correspond to different pharyngeal shapes and/or different areas of resonance of the voice in the body (e.g., head vs. chest)].

- in the *kotiráp* flute this is manifested as a frequency lowering of one note in the scale of its four musical notes. This lowering happens in the middle of each long verse of music. This can be well observed in the sentence *ɔ̃́gali atí mápit bósara kîì káre béá* presented as an example in **Figure 6**. Interestingly here the lowering happens between the low tone of the syllable *pit* and the high tone of the syllable *bó* which are thus at the same level and played in one sole long continuous note;
- in the case of the *totoráp* and *iridináp* music, this esthetic effect may happen between repetitions of the same phrase or of part of a phrase. This is well illustrated for the *totoráp* clarinets in **Figure 8** showing a focus on the difference between two repetitions of the expression *zérék kîìt* (here, the second repetition of this expression is higher in pitch). On both repetitions, the long vowel of the syllable *kîìt* is played with the clarinet as a long flat tone that is lower than the ones of the preceding high tone syllables *zérék*—adapting the music to the phonological rules explained in section "Some Phonetic and Phonological Characteristics of the Language of the Gavião of Rondônia" for a floating

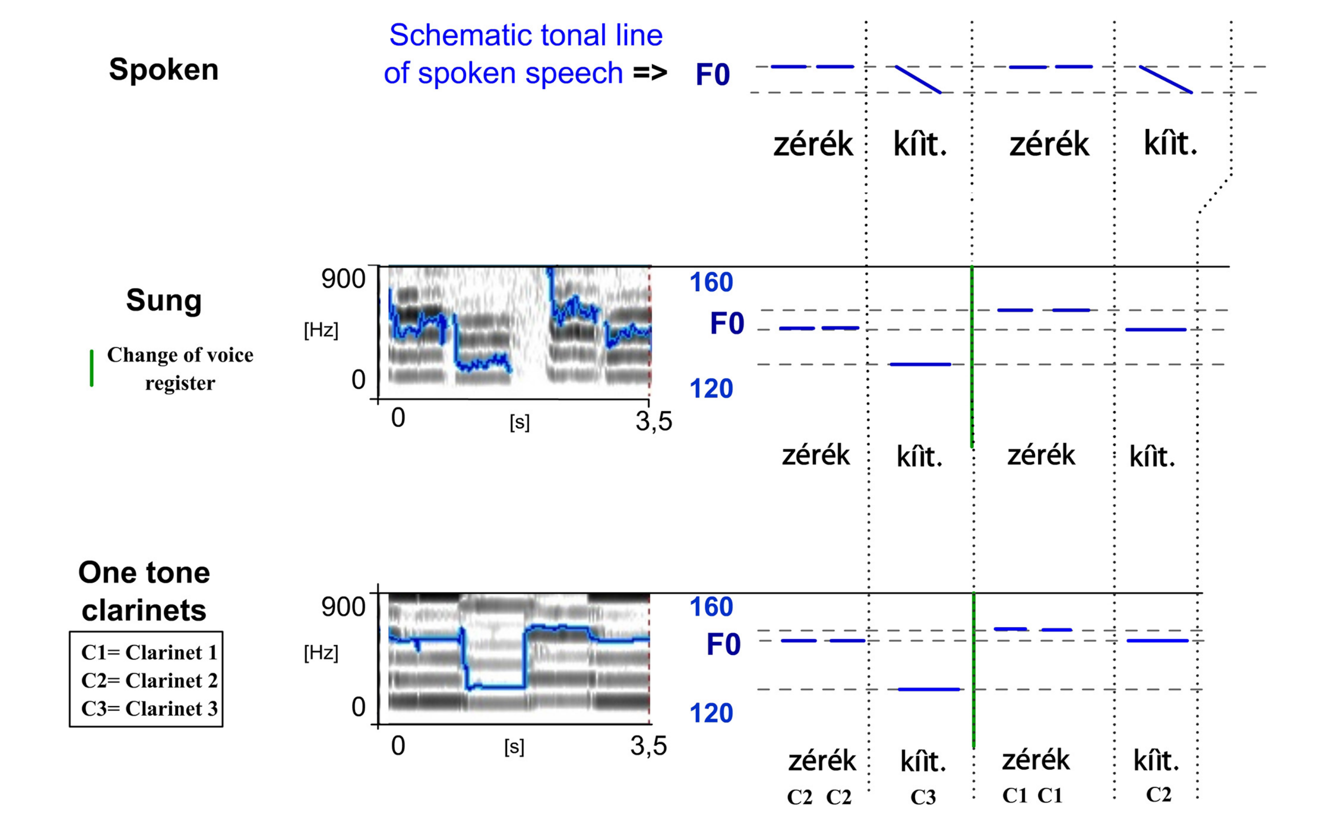


FIGURE 8 | Spectrograms of the sung and musical forms of a song played on the *totoráp* clarinets (meaning "The white skin the white skin." The blue lines on the spectrograms represent the extracted F0 (the same parameters have been used for the sung and musical speech types which explains more jumps in detection in the sung forms). To see more clearly the similarities between different speech types, they are reported in simplified charts with a focus on vowel nuclei (right). Note here the change of voice register in the middle of the verse (delimited by a green line) and the repetition (adapted from Meyer and Moore, 2013).

low tone at the end of an utterance, but also adapting to the constraints of the instrument in terms of modulation as we just explained earlier (see more details in "The Linguistics of the Music-Language Relation" for linguistic details/implications of this adaptation). Note also that the *totoráp* and the *iridináp* music are based on three discrete notes and the musical transposition of a change of voice register is thus more limited than for the *kotiráp* flute.

The Linguistics of the Music-Language Relation

General Considerations, Similarities and Differences Between Spoken and Musical Lines

In speech there is a distinction between fundamental frequency (F0)—which corresponds to the vibration of the vocal cords and carries the supersegmental prosody (including the lexical tone in the case of tone languages)—and its harmonic resonances in the vocal tract, which define the quality of the vowels and of the consonants. We already saw that a distinction between timbre and F0 exists also in the acoustics of the instruments. Based on this parallel, one of the most notable similarities observed between the music played by the *kotiráp*, *totoráp,* and *iridináp* instruments and spoken speech is the syllable-by-syllable correspondence between the respective tones and notes of the two forms.

However, the correspondence is from a spoken tonal line that allows frequency curves (modulations), and an instrumental music line that is based on flat notes (musical constraint). Indeed, the modulations of the melodic line of modal spoken Gavião speech produce curved tones that are hardly ever imitated in the instrumental form, not even in the associated song form (see the topic "Adaptation of the sung mode to the instruments"). This affects the rendering of tonal patterns of the Gavião language by the instruments and is a central part of our analysis.

As we just saw, some other differences observed between the instrumental form and the spoken modal speech form also occur because of esthetic effects that are also present in the sung modality of speech (see the topic of section "Some Specific Esthetic Effects"). Finally, differences come from the use of archaic forms of words in songs which no longer exist in the common spoken form (see the following topic in section "Archaic Forms of Words in the Old Songs"). All of these aspects are emulated by the players of the instruments and coded in the tonal line (F0) of the played melody.

In terms of resemblance, the similarity between normal speech and music is also exhibited in the duration of the spoken syllables

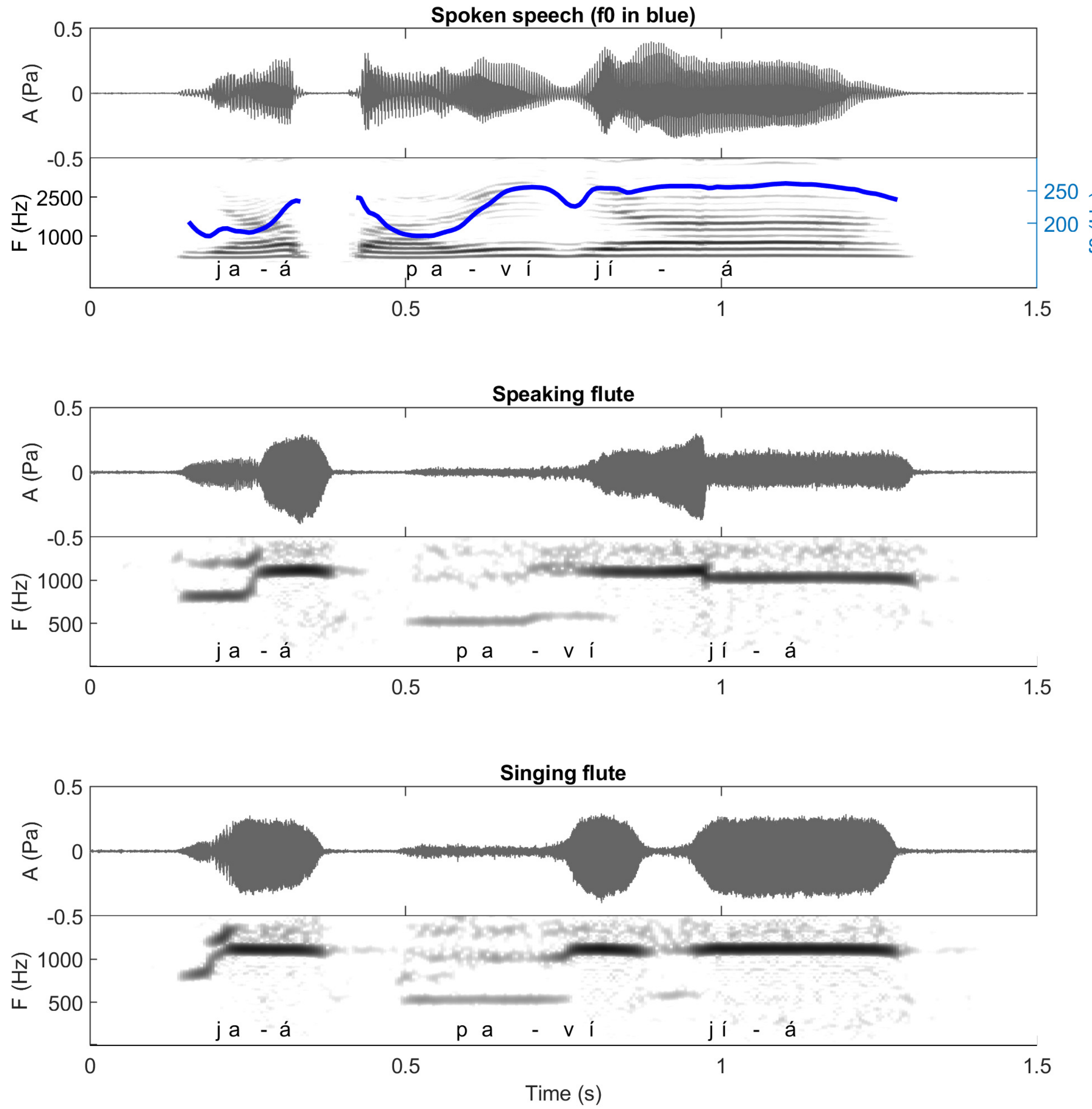


FIGURE 9 | Waveforms and spectrograms of the sentence *jaá pavíjíá* (meaning "Let's go to bath") played in spoken speech (upper row), with the flute as if speaking (middle row), and with the flute as in traditional *kotiráp* songs (lower row). The blue line on the spectrograms of spoken speech represents the extracted F0, highlighting the surface tonal line of speech.

and their corresponding notes: the long syllables are made with long sounds on the musical instruments, respecting the quantity distinctions of the Gavião phonology. This is very clear for the *totoráp* clarinets which emphasize precisely this coding level (as shown in **Figure 7** for example). For the *kotiráp* flute, relative durations are also respected but with less emphasis than with the *totoráp* clarinets (as shown, for example, in **Figure 6** for *kíì* which is one of longest syllables played in this verse). Finally, with the *iridináp* bows a long vowel is sometimes played with two successive notes (see, for example, the two first notes of **Figure 5**). Moreover, as explained earlier, at the rhythmic level, the dynamic of the amplitude envelope of the instrumental modality of speech parallels the sung one and even the spoken one, with more precision for the flute and less for the clarinets (**Figures 5–7**).

Finally, the phonetic and phonological nature of the acoustic iconicity which exists between the words of the songs and the music played with the *totoráp*, *kotiráp* and *iridináp* is simply based on the cognitive association between, on one side, the tone and syllable length of the spoken language and, on the other side, the notes played with musical instruments. What could not be foreseen is just how the flat notes reflect the tone and length. As it turns out, there are quite consistent rules for this, and they are common to the three types of singing Gavião instruments.

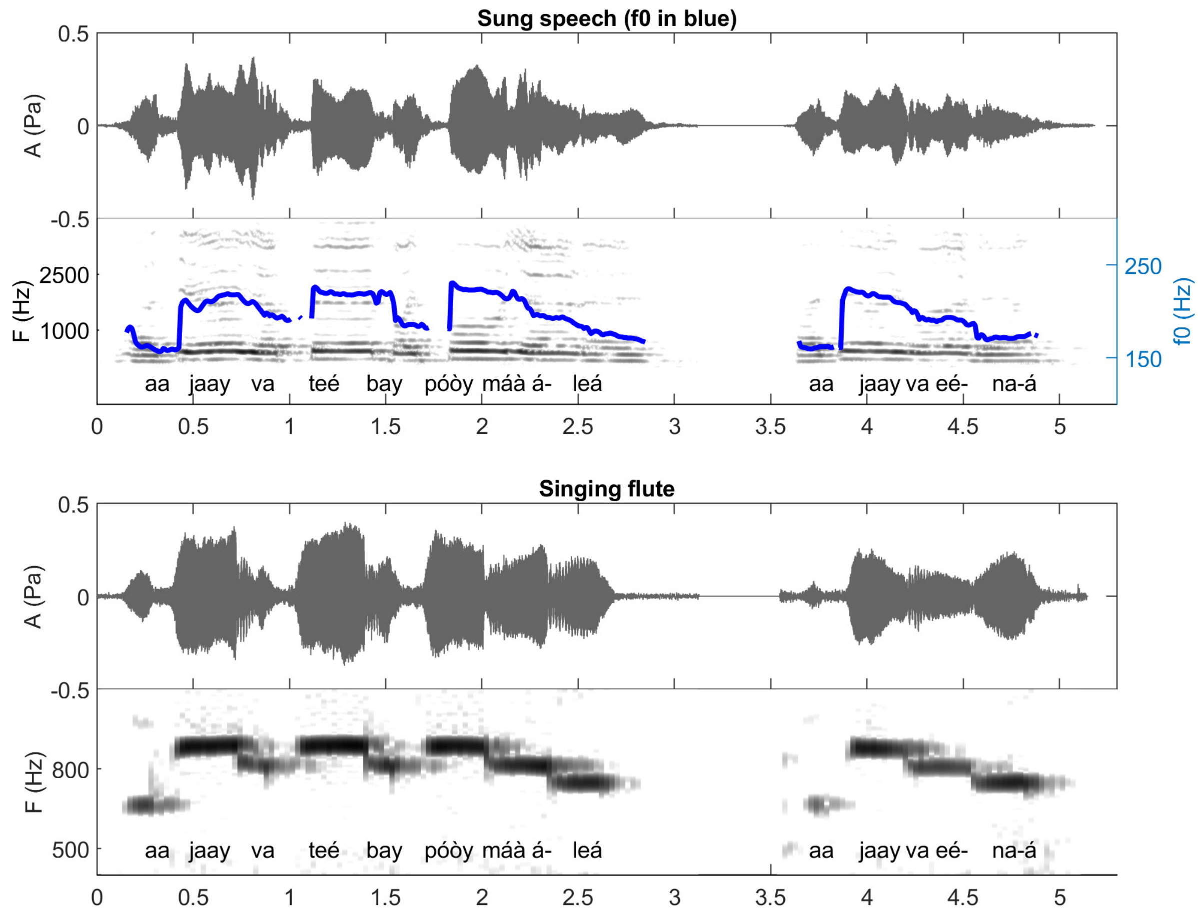


FIGURE 10 | Waveforms and spectrograms of two different verses of the same "Boa constrictor" song traditionally played with the *kotiráp* flute among the Gavião people. The first verse is *aajaay va teé bay póòy máà áleá* (meaning "Boa constrictor ate his owner"), the second verse is *aajaay va eénaá* (meaning "ate his owner like that"). The blue line on the spectrogram of sung speech (upper row) represents the extracted F0 that is emulated by the flute on the lower row. Note that *máà-á* are played together as realized in normal speech, and that *va eé* are also played together, as realized in fast speech (see explanation in section "Correspondence Between Musical Notes and Phonetic Tone—Playing Flattened Surface Tone" listen to sound extracts in Meyer (2021).

Correspondence Between Musical Notes and Phonetic Tone—Playing Flattened Surface Tone

The variations of F0 of the musical notes correspond directly to the variations of F0 of the lyrics of the associated song (see **Figure 5–7**), which also follow the variations of F0 of the lexical tone of normal speech. Two major patterns of correspondence are discussed below.

Downstep

The complex phenomenon of the downstep of the Gavião phonology is reproduced by notes which reproduce the effects of this phonological rule. This is exemplified for the *kotiráp* flute in **Figure 6** and schematized below in (1) (the downstep occurs after *kíì* in *kíì káre béá*). It is illustrated for the *totoráp* clarinets in **Figure 11** and schematized below in (2) (the downstep occurs after *éèy* in *coliléèy mága*). We will see in the discussion (section "Discussion and Conclusion") that the finding that downstep is preserved in the instrumental speech forms of the Gavião has an important implication: it directly questions previous theoretical hypotheses maintaining that only underlying forms of tonal systems are emulated in speech surrogates (as stated by Bagemihl (1988), who had built such a conclusion on the fact that at the time of his review on the subject he had not found in the literature any mention of surface form imitation in such instrumental systems). Gavião instrumental speech forms are among the rare speech surrogate practices where downstep has been found so far (see other references in the Discussion section 4).

kíì káre béá (1) (see also **Figure 6**)

coliléèy mága (2) (see also **Figure 11**)

In the case of (1), the floating low tone of the syllable *kíì* is followed by the high tone of the syllable *ká*, and thus attaches to the following high tone: the combination is realized phonetically as a mid-tone in spoken speech and sung speech, and as a middle note in instrumental speech played with the flute (this is also highlighted in red on **Figure 6**). This example also shows that the

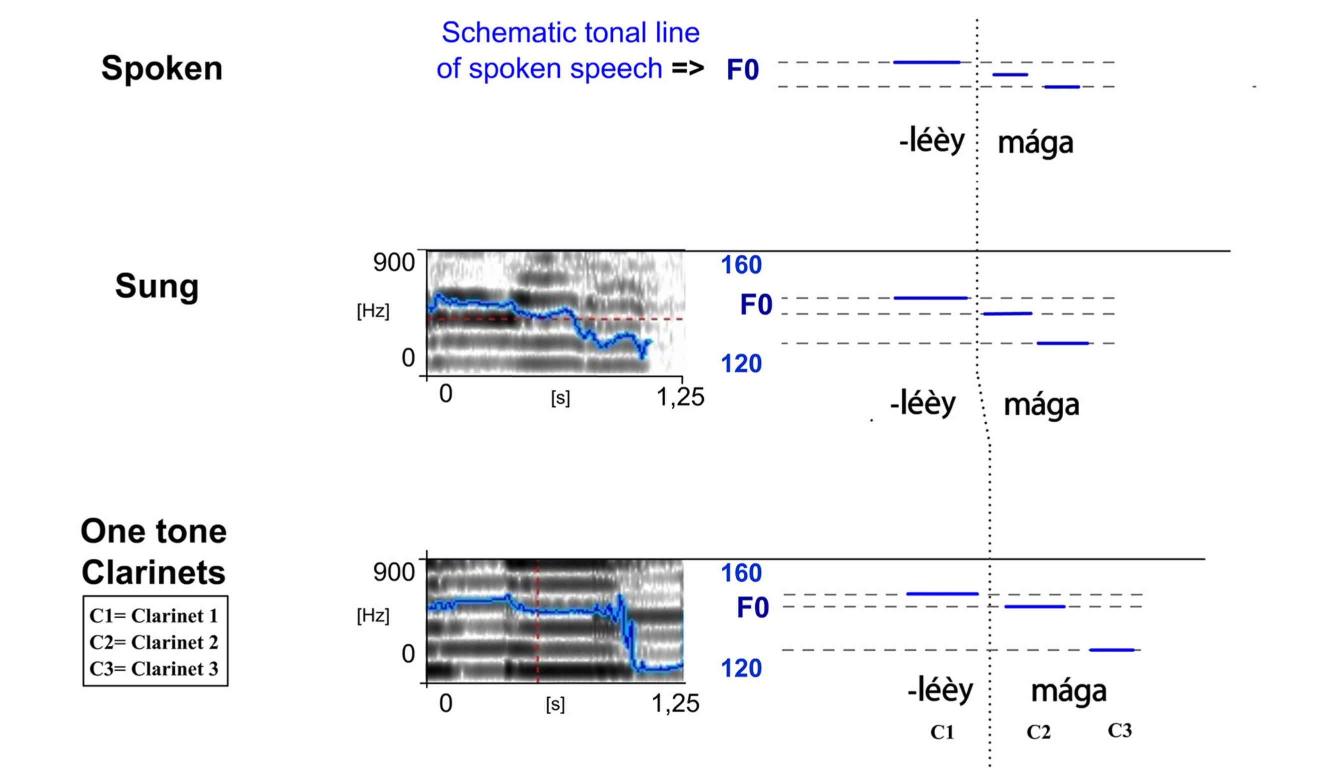


FIGURE 11 | Spectrograms of the sung and musical forms of the song extract *léèy mága* played on the *totoráp* clarinets (from *colilléèy mága* meaning "there are swallows"). The blue lines on the spectrograms represent the extracted F0 (the same parameters have been used for the sung and musical speech types which explains more jumps in detection in the sung forms). To see more clearly the similarities between different speech types, they are reported in simplified charts with a focus on vowel nuclei (right). Note here the tone downstep after *éèy* (adapted from Meyer and Moore, 2013).

consecutive high tones of the same sentence are also lowered to a phonetic mid-level (here and in **Figure 6** see the middle note played to encode *béá*). The case of (2) follows the same pattern and together with the **Figure 11** provides a zoom on how these syllables are sung and played with the three clarinets.

Rising Tones

Another interesting aspect of the correspondence between musical notes and phonetic tone is that long contour tones are reproduced as flat notes, according to certain rules which take into account the immediate tonal environment, maintaining approximate relative heights. Phonetic long rising tones, whether from fused LH sequences, constant long rising tones, dissimilating long low tones or rising downstep tones, are played as long middle notes when following high or rising phonetic tones and preceding low phonetic tones, as middle or high notes between low phonetic tones and as long low notes finally. These patterns also apply to phonetic curved tones formed by fusing syllables in rapid speech that are separate in slow speech.

aajaay va teé bay póòy máà áleá (3) (see also **Figure 10**)

aajaay va eénaá (4) (see also **Figure 10**)

In example (3) and **Figure 10** the second syllable of *aajaay* first becomes a phonetic long rising tone because of the low dissimilation rule and then, being a long rising tone between two low tones, is played as a long high note. The long rising tone of *teé* also is played as a long high note between two low tones (note that with the flute the low tones of *va* and *bay* are played as middle notes, on an absolute scale, keeping the relative relation between tones in their environment, while the initial long low and final long rising tones of the sentence are played as long low notes on an absolute scale). The "future" particle *ále* cliticizes to the (downsepped) auxiliary *máà*, producing a descending tone sequence which is played as a long middle note. The final rising tone of *áleá* is played as a long low note.

In example (4) the form *eénaá* cliticizes to the verb *va* and in the resulting form, *veénaá,* the phonetic long rising tone is played as a long middle note followed by a final long low note.

ẽzáká ẽebíí teé kay áleá (5)

In (5) the two short high tones of *ẽzáká* are played together as a long high note. The long rising tone of *teé* is played as a long middle note after a high and before a low. That level carries over to the first note of *áleá*. (listen to audio "totorap-Equation5" in audio annex Meyer (2021).

Archaic Forms of Words in the Old Songs

As noted earlier, the lyrics of the associated songs contained archaic forms, which require additional explanations. It is interesting to observe what the differences are between the archaic forms (marked in italics in the texts in "Repertoire and some examples of songs") and the modern forms. In the phrase presented in **Figure 4**, for example, there appears the verb *bósara*, translated as "going to bed (retiring to a hammock)," which is a lexical item that does not exist in the modern Gavião language. Unexpectedly, the phrase in **Figure 6** contains the form *õbút* "1s-thing + diminutive," which retains the prefix *õ-*. In the evolution of the languages of the Mondé branch, this prefix disappeared some time ago in the speech of the Gavião and Zoró while persisting in the language of the Paiter (Suruí) and, optionally, in the speech of the Aruá. So, one of the differences in the common archaic forms is in the lexicon and another is in the morphological processes. Since contemporary music composed by the Gavião does not contain these archaic forms they may indicate that the composition of the associated songs occurred centuries ago. However the Gavião can pronounce these old lexical items in normal speech and they conform to modern Gavião phonological patterns.

DISCUSSION AND CONCLUSION

The Gavião singing instruments, aside from constituting an important cultural heritage threatened by extinction, represent a phenomenon in the interface between language and music. In Amazonia this type of practice is still little known by linguists and musicologists because it is on the margins of each of the two disciplines. In the present article we show the principal steps of our methodology for documenting and studying it. The resulting archives, the associated publications and the collaborative work with the local community represent a contribution to the cultural and linguistic memory of this group, important for revitalization. The Gavião know that the songs played with musical instruments have the rare quality of imitating the sung voice and have a certain pride in their capacity to imitate linguistic properties with musical sounds.

The analytic perspectives are multiple and pluri-disciplinary. We show that techniques of linguistic analysis make it possible to explain, in detail, the nature of the music-language relation in the *totoráp, iridináp* and *kotiráp*, in which lexical tone and syllable length constitute fundamental traits. We highlighted that this relation is due to the fact that linguistic tone is important in a tone language for the understanding of the meaning of the lyrics of the songs (much more than in a language without tone). Very strikingly, because spoken surface tonal downstep (though not the underlying tones with their floating lows) is systematically transposed in this kind of Gavião music, our description also places Gavião instrumental speech practices in a noteworthy situation in the typology of instrumental language surrogates associated with a tone language. Indeed, according to the few existing surveys of the literature about instrumental speech and language surrogates with musical instruments (Sebeok and Umiker-Sebeok, 1976; Bagemihl, 1988), the phenomenon of downstep was not found until recently in instrumental speech forms emulating a tone language. Up to now, as far as we know, it was described first in Serepewa Luth playing Akan language (Nketia, 1994), in Gavião instrumental speech forms (Meyer and Moore, 2013) and, more recently, in Yoruba talking drums (Akinbo, 2020), as well as in Northern Toussain Balafon (Struthers-Young, 2021) and sometimes in singing Balafon forms of Seenku language (McPherson and James, in press). Some recent results analyzing the statistical correspondence between tones of Mooré language and the way they are beaten with the traditional royal Bendré drum also suggests congruence with the surface terracing properties of the tonal Mooré system (Dentel and Meyer, 2020) but confirmation with detailed linguistic analysis remains to be done.

It does appear that the great majority of speech surrogates diverge from the surface realization of a base utterance and instead encode the underlying tonal phonology of the language that is emulated with the instruments [as reviewed by Bagemihl (1988) and described for example in Ewondo Drumming (Neeley, 1999), or more recently in Sambla Balafon emulating the spoken form of Seenku language (McPherson, 2018)]. So it is something of a rarity that downstep is played in instrumental speech music. Yoruba talking drums were, for example, described as permitting the drummer to regulate gradient pitch, following the contours of Yoruba post-lexical tone. Such versatility in production may explain the behavior observed on downstep with this speech surrogate. Notably, the talking drums have the particularity of enabling diverse frequency modulation patterns because of the ability of the drummer to change the tension of the skin by pressing chords with his arm.

The explanation for Gavião instrumental speech forms is necessarily different. The Gavião downstep is also post-lexical, but it is even emulated on instruments such as the ensemble of three one-tone clarinets that do not have the capacity for any modulation in frequency. Its realization is consistent with a strong generalization about Gavião song emulation: the tone of the surface phonetic form is replicated, but only with flat notes, which reflect the relative levels of the syllables. This constraint affects all contour tones, rising and falling. If the underlying H:(L) sequences of high downstep tones were reproduced, for example, a HL contour within one syllable would be produced, violating the constraint. However, a surface drop to flat mid-level on the following syllable is acceptable and can be played. Interestingly, even a late rule like low tone dissimilation can occur in lyrics to produce a surface rising tone that is then flattened in its instrumental replication, like any other phonetic rising tone. More surprising is that the results of fast speech fusions, which are far from the underlying forms, can be played as flat notes according to the same patterns.

The example of Gavião suggests that the instrumental realization of downstep may depend, across languages, on general constraints on instrumental song representation in the particular language and also on what kinds of underlying and surface forms are involved. In the Gavião words the downstep-triggering syllables have underlying floating lows, which only surface as lows in isolated words or elsewhere when they are utterance final (as in the final L tone in the spoken form of the second word in *zérék kîìt*), not medially in connected text. Interestingly, all these cases confirm the need for a reexamination of the theoretical basis for the hypothesis that tonal downstep (as well as downdrift) is absent in instrumental speech surrogates (see Bagemihl, 1988).

In the context of a current effort to compare such "talking/singing" musical practices worldwide, the Gavião language clearly provides new key elements about which levels of the phonological/phonetic structure are possibly accessed and highlighted by talking musical instruments. In that sense, the analysis of the Gavião "singing" instruments contributes to general theory about the principles of instrumental imitation of speech (e.g., McPherson, 2019) and more generally to the typology of tone-tune association in songs based on tonal languages (e.g., Wong and Diehl, 2002; Meyer, 2007; Schellenberg, 2012; McPherson and Ryan, 2018; Ladd and Kirby, 2019). More precisely, by describing new patterns, it adds empirical data to understand better which levels of the phonological/phonetic structure may be transposed in instrumental speech, and how parts of these levels may be represented.

Another point of interest raised by the sounds associated to the music analyzed in this study deals with the syllables without meaning interspersed in the songs. These are akin to vocables because the singers affirm that they represent the sounds and the rhythm played on the instrument itself. They were found for the *iridináp* and the *totoráp* instruments and in each case under two different forms. The vocables corresponding to *iridináp* mouth bows were *iriri* (with first syllables in common with the name of the instrument) and *toy péréré*, while for the *totoráp* ensemble of clarinets they were *sẽ́t sẽ́t* and *o sẽ́t o sẽ́t*. This is a very limited sample that does not enable us to draw definitive conclusions about how and why the vowels and consonants are chosen for each kind of vocable.

When comparing these data with results presented in other studies dealing with larger collections of vocables in traditional song around the world (e.g., Hinton, 1980, 1984; Hughes, 2000; Tuttle, 2012), we did not find a tendency indicating a possible convergence in the acoustic iconicity of these Gavião vocables with common hypotheses of either intrinsic pitch or intrinsic F0 (stating for example that the Formant 2 (F2) or the fundamental frequency (F0) of vowels may influence the choice of the vowel type for such vocables). There are other Gavião musical instruments which also employ vocables and the study should be extended to these in order to have a clearer view on this topic.

Another important factor in the music-language iconic relation is the fact that the songs associated with the instrumental melodies of *totoráp, iridináp,* and *kotiráp* adapt themselves to the particular acoustic properties and modes of playing the instruments (the number of possible discrete notes and/or the lesser possibility of modulating the frequency of the notes), and to certain requirements in the manner of playing them (as in the case of the *totoráp*, which needs the coordination and participation of three people). This absence of frequency modulations in Gavião instrumental singing appears to be a deliberate choice for the flute and the musical bows because it is technically possible to produce at least some elements of modulations with these two instruments, though with some difficulty. We could think of many different reasons to explain that musical behavior: the influence of the *totoráp,* a technical choice to simplify the corpus for non-expert players, a compositional rule for all "singing" instruments or a result of an ancient evolution of singing styles. The closest dialect to the speech of the Gavião, that of the Zoró, as well as the sister language of both, Paiter (Suruí), have no rising tones.

Aside from the average age of the few skilled representative players, linguistic facts also converge to demonstrate the antiquity of the phenomenon. There are archaic forms in the Gavião words in the lyrics of the songs associated with the played music, principally in the vocabulary and in the morphological processes. Interestingly, we did not observe any indication of diachronic change in the phonology of the lyrics of the analyzed songs. On the basis of these observations we consider two hypotheses about the songs associated with instrumental speech. First, it is possible that the songs with archaic forms were composed two or more centuries in the past and have maintained their original form. Second, it is also possible that there exists a poetic system of musical composition characteristic of these instruments, capable of producing archaic forms in contemporary times, that is known by the trained players. The next step will be to investigate the possible existence of archaic forms in music of other genera, of recent composition. To clarify these points it is necessary to continue the documentation and the collaboration with the Gavião and other peoples of the Mondé branch of the Tupi family (Cinta Larga, Zoró, Aruá, Paiter) to extend the repertoire of played songs and compare the phenomena typologically.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://soundcloud.com/user-28976943/sets/sounds-ofa-flute-musical-bows-and-bamboo-clarinets-that-speak-in-the-amazon-rainforest.

ETHICS STATEMENT

There was at the time of our research no standing Ethics Committee at the Museu Goeldi, our host institution in Belém. Ethical questions, should they arise, were addressed by a "sindicância" (an internal investigative committee). In Brazilian law and practice, the participating indigenous community indicates, either orally or in writing, their informed consent to the proposed research to the local office of the National Indian Foundation (FUNAI), which in turn transmits that consent, in the form of a document, to the central FUNAI office in the national capital. This office issues written research permits. Our research followed these established procedures. Native local authorities authorized our work in all of the visited communities. Permits were obtained from the National Indian Foundation (FUNAI) and the National Research Council (CNPq). Thus, research was conducted in accordance with the Declaration of Helsinki. Copies of the recordings and the research results were given to the community. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JM proposed the study, did the fieldwork inquiry, collected and recorded the musical pieces and their associated sung and spoken forms, and prepared the figures. DM provided the contacts for authorization of the study and corrected the transcriptions. JM and DM did the manuscriptwork for the working permit in the indigenous territory, transcribed the texts with local Gavião collaborators, analyzed the data, and wrote the manuscript.

FUNDING

Fieldwork documentation and laboratory analyzes were conducted thanks to funding of two institutions: The Endangered Language Documentation Program (Post-doc project IPF0136, The School of Oriental and African Studies—University of London) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (Pós-Doutorado Júnior Grant n. 159747/2010-4, CNPq Brazil). The writing of the present version of the manuscript and the reorganization of the data were funded by the International Emerging Action CNRS 2020-2021 "AmazonLing." Research activity was also partly financed in 2020-2021 in the framework of the project GeoTeleLing with the support of the CNRS 80Prime interdisciplinary programme.

ACKNOWLEDGMENTS

We are grateful to the Gavião people of Rondônia, and particularly to the players and dancers who accepted to be recorded; to all the Gavião language speaking teachers and to the Gavião language assistants of the project; to Laure Dentel and to Aïto Meyer-Dentel for their contribution during fieldwork documentation; to the colleagues of the Museu Paraense Emilio Goeldi, to the Endangered Language Documentation Programme—ELDP; to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the authorization of scientific expedition n. EX032/08-CR; and to the Fundação Nacional do índio (FUNAI), in Brasília and in Ji-Paraná. We would like also to thank Jimena Felipe Beltrão, Scientific Editor of the Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas for her prompt reply authorizing us to translate some elements previously published in Portuguese. Finally we thank the two reviewers and the editors of this volume for their work and helpful suggestions. This is an English language expanded version of "Arte verbal e música na lingua Gavião de Rondônia: metodologia para estudar e documentar a fala tocada com instrumentos musicais" originally published in Portuguese in the Boletim do Museu Paraense Emílio Goeldi; Ciências Humanas, volume 8, número 2, p. 307–324, in 2013. JM and DM prepared this version with support from the CNRS as part of its International Emerging Action AmazonLing and of its 80PRIME interdisciplinary programme. Permission for the English version was granted by Jimena Felipe Beltrão, Scientific Editor of the Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas.

REFERENCES

Akinbo, S. (2020). Representation of Yorúbá Tones by a talking drum: an acoustic analysis. *PsyArXiv [Preprint]*. doi: 10.31234/osf.io/43gf6

Arom, S. (1970). *L'arc Musical Ngbaka, République de Centrafrique*. Paris: CNRS-Comité du Film Ethnographique.

Bagemihl, B. (1988). *Alternate Phonologies and Morphologies*. Vancouver, BC: University of British Columbia.

Beaudet, J. M. (1997). *Souffles d'Amazonie, Les Orchestres tule Des Wayãpi*. Nanterre: Société d'Ethnologie.

Carrington, J. F. (1949). *Talking Drums of Africa*. London: Carey-Kingstate Press.

Catlin, A. (1982). "Speech Surrogate Systems of the Hmong: from singing voices to talking reeds" in *The Hmong in the West: Observations and Reports*, eds B. T. Downing and D. P. Olney (Minneapolis, MI: University of Minnesota Press). 170–197.

Dentel, L., and Meyer, J. (2020). "How Rhythm and Timbre Encode Mooré Language in Bendré Drummed Speech" in *Proceedings of Interspeech 2020- 21st Annual Conference of the International Speech Communication Association* (Germany: DBLP), 1898–1902. doi: 10.21437/Interspeech.2020-2532

Ermel, P. B. (2004). *L'Arc et la Lire: proposition d'un Narratif Visuel et Sonore*. Paris: Collège Iconique de l'Institut National de l'Audiovisuel (INA).

Everett, D. (1983). *A língua Pirahã e a Teoria da Sintaxe: descrição, Perspectivas e Teoria*. Campinas: Universidade Estadual de Campinas.

Franchetto, B., and Montagnani, T. (2011). Flutes des hommes, chant des femmes. Images et relations sonores chez les Kuikuro du haut Xingu. *Gradhiva* 13, 94–111. doi: 10.4000/gradhiva.2052

Hill, J. (1993). *Keepers of the Sacred Chants: the Poetics of Ritual Power in an Amazonian Society*. Tucson: The University of Arizona Press.

Hinton, L. (1980). "Vocables in Havasupai music" in *Southwestern Indian Ritual Drama*, ed. F. Charlotte (Albuquerque, New Mexico: University of New Mexico Press). 275–305.

Hinton, L. (1984). *Havasupai Songs: a Linguistic Perspective*. Amsterdam: John Benjamins Publishing Company.

Hughes, D. W. (2000). No nonsense: the logic and power of acoustic-iconic mnemonic systems. *Br. J. Ethnomusicol.* 9, 93–120. doi: 10.1080/09681220008567302

Izikovitz, K. G. (1935). "Idiophones" in *Musical and Other Sound Instruments of the South American Indians: a Comparative Ethnographical Study*, ed. K. G. Izikovitz (Göteborg: Elanders Boktryckeri Aktiebolag), 7–33.

Kawada, J. (2014). Notes on the Drum Language of the Ancient Mossi Kingdoms (Burkina Faso). *Japanese Rev. Cult. Anthropol.* 15, 199–215.

Ladd, D. R., and Kirby, J. (2019). "Tone-melody matching in tone language singing" in *The Oxford Handbook of Prosody*, eds C. Gussenhoven and A. Chen (Oxford: Oxford University Press), 676–687. doi: 10.1093/oxfordhb/9780198832232.013.47

Locke, D., and Agbeli, G. (1981). Drum language in Adzogbo. *Black Perspect. Music* 9, 25–50. doi: 10.2307/1214103

McPherson, L. (2018). The Talking Balafon of the Sambla: grammatical Principles and Documentary Implications. *Anthropol. Linguist.* 60, 255–294. doi: 10.1353/anl.2019.0006

McPherson, L. (2019). "The role of music in documenting phonological grammar: two case studies from West Africa" in *Proceedings of the Annual Meeting on Phonology 2018* (New York, NY: Linguistic Society of America). doi: 10.3765/amp.v7i0.4493

McPherson, L., and James, L. (in press). "Artistic adaptation of Seenku tone: musical surrogates vs. vocal music" in *Celebrating 50 Years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, eds A. Akinlabi., L. Bickmore., M. Cahill., M. Diercks., L. J. Downing., J. Essegbey., K. Franich., et al. (Berlin: Language Science Press).

McPherson, L., and Ryan, K. (2018). Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94, 119–156. doi: 10.1353/lan.2018.0003

Meyer, J. (2007). "Acoustic features and perceptive cues of songs and dialogues in whistled speech: convergences with sung speech" in *Proceedings of the International Symposium on Musical Acoustics 2007* (Spain: Universitat Politechnica de Catalunya). 1-S4-4, 1–8.

Meyer, J. (2014). *Documentation of Gavião and Surui Languages in Whistled and Instrumental Speech, Endangered Languages Archive*. London: University of London.

Meyer, J. (2015). *Whistled Languages: A Worldwide Inquiry About Whistled Speech.* Berlin: Springer.

Meyer, J. (2021). "Sound extracts of 'A Flute, Musical Bows and Bamboo Clarinets that "Speak" in the Amazon Rainforest," in *The World Whistles*. Available online at: https://soundcloud.com/user-28976943/sets/sounds-ofa-flute-musical-bows-and-bamboo-clarinets-that-speak-in-the-amazon-rainforest (accessed September 13, 2021).

Meyer, J., Dentel, L., and Seifart, F. (2012). "A methodology for the study of rhythm in drummed forms of languages: bora of Amazon" in *Proceedings of Interspeech 2012 Annual Conference of the International Speech Communication Association* (Brazil: Universidade Federal do Pará), 686–690. doi: 10.21437/Interspeech.2012-210

Meyer, J., and Moore, D. (2013). Arte verbal e música na língua Gavião de Rondônia: metodologia para estudar e documentar a fala tocada com instrumentos musicais. *Bol. Mus. Para. Emílio Goeldi Ciênc. Hum.* 8, 307–324. doi: 10.1590/S1981-81222013000200006

Mindlin, B., Tsoraba, D., and Sebirop, C. (2001). *Couro Dos Espíritos: namoro, Pajés e Cura Entre os Índios Gavião-Ikolen de Rondônia*. São Paulo: Terceiro Nome.

Moore, D. (1984). *Syntax of the Language of the Gavião Indians of Rondônia (Brazil)*. New York, NY: City University of New York.

Moore, D. (1999). "Tonal System of the Gavião Language of Rondônia, Brazil" in *Tupian Perspective Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics*. ed. K. Shigeki (Tokyo: ILCAA/Tokyo University of Foreign Studies), 297–310.

Moore, D., and Meyer, J. (2014). The study of tone and related phenomena in an Amazonian tone language, Gavião of Rondônia. *Lang. Document. Conserv.* 8, 613–636.

Moore, D., and Meyer, J. (2015). *Gavião de Rondônia. Archive of Indigenous Languages of the Museu Goeldi (ALIM)*. Belém: Museu Goeldi.

Neeley, P. (1999). *People of the Drum of God—Come!*. Dallas, TX: Summer Inst of Linguistics.

Niles, D. (2010). Editor's Introduction to Zemp and Kaufmann. *Kulele Occasion. Pap. Pacif. Music Dance* 4, 1–50.

Nketia, J. H. K. (1976). "Surrogate languages of Africa" in *Speech Surrogates: drum and Whistle Systems*, eds T. A. Sebeok and D. J. Umiker-Sebeok (Mouton: La Hague and Paris), 825–864.

Nketia, J. H. K. (1994). "Generative processes in Seperawa music" in *To the Four Corners: a Festschrift in Honor of Rose Brandel*, ed. E. C. Leichtman (Michigan: Harmonie Park Press), 127–151.

Odden, D. (1986). On the role of the obligatory contour principle in phonological theory. *Language* 62, 353–383. doi: 10.2307/414677

Poss, N. (2005). *The Communication of Verbal Content on the Hmong Raj: an Ethnographic Analysis of Performance Practice*. Ohio: Ohio State University.

Risset, J. C. (1968). Sur certains aspects fonctionnels de l'audition. *Ann. Télécommun.* 23, 91–120. doi: 10.1007/BF02996236

Ritzenthaler, R. E., and Peterson, F. A. (1954). Courtship Whistling of the Mexican Kickapoo Indians. *Am. Anthropol.* 56, 1088–1089. doi: 10.1525/aa.1954.56.6.02a00110

Schellenberg, M. (2012). Does language determine music in tone languages? *Ethnomusicology* 56, 266–278. doi: 10.5406/ethnomusicology.56.2.0266

Schouten, J. F., Ritsma, R. J., and Cardozo, B. I. (1962). Pitch of the residue. *J. Acoustic Soc. Am.* 34, 1418–1424. doi: 10.1121/1.1918360

Schwartz, D. A., and Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hear. Res.* 194, 31–46. doi: 10.1016/j.heares.2004.01.019

Sebeok, T. A., and Umiker-Sebeok, D. J. (1976). *Speech Surrogates: drum and Whistle Systems*. Mouton: La Hague and Paris. doi: 10.1515/9783110804416

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: amazonian Bora drummed language exploits speech rhythmfor long-distance communication. *R. Soc. Open Sci.* 5:170354. doi: 10.1098/rsos.170354

Stern, T. (1957). Drum and whistle languages: an analysis of speech surrogates. *Am. Anthropol.* 59, 487–506. doi: 10.1525/aa.1957.59.3.02a00070

Struthers-Young, A. (2021). A preliminary account of the Northern Toussian balafon surrogate language. *Front. Commun.* doi: 10.3389/fcomm.2021.652740

Thiesen, W. (1969). The Bora signal drums. *Lore* 19, 101–103.

Tuttle, S. (2012). Language and music in the songs of Minto, Alaska. *Lang. Document. Descrip.* 10, 88–112.

van der Voort, H. (2016). "Aikanã and Kwaza: their ethno-historical and sociolinguistic context in Rondônia, Brazil" in *Endangered Languages and Languages in Danger: Issues of Documentation, Policy, and Language Rights*, eds L. Filipoviæ and M. Pütz (Amsterdam: John Benjamins), 203–230. doi: 10.1075/impact.42.09van

Wedekind, K. (1983). A Six-Tone language in Ethiopia: tonal Analysis of Benchnon. *J. Ethiop. Stud.* 16, 129–156.

Whiffen, T. (1915). *The North-West Amazons: Notes of Some Months Spent Among Cannibal Tribes*. New York, NY: Duffield and Co.

Winter, Y. (2014). On the Grammar of a Senegalese Drum Language. *Language* 90, 644–668. doi: 10.1353/lan.2014.0061

Wong, P. M., and Diehl, R. L. (2002). How can the lyrics of a song in a tone language be understood? *Psychol. Music* 30, 202–209. doi: 10.1177/0305735602302006

Yegnan-Touré, A. (2008). La technique et le jeu de l'arc musical. *Cahiers Ethnomusicol.* 21, 205–223.

Zemp, H., and Soro, S. (2004). Paroles de balafons. *L'Homme* 17, 313–332. doi: 10.4000/lhomme.24930

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fcomm.2021.652542

The Speech Surrogacy Systems of the Yoruba Dùndún and Bàtá Drums. On the Interface Between Organology and Phonology

Mariano González[1]* *and Olupemi Oludare*[1,2]

[1]Utrecht Institute of Linguistics UiL OTS, Utrecht University, Utrecht, Netherlands, [2]UNILAG, University of Lagos, Lagos, Nigeria

This paper explores the interdependence between organology and phonology in the Yoruba *dùndún* and *bàtá* drums. We analyze how the specific features of these drums, such as corpus shape, size, kind and number of membranes, and playing techniques affect their systems of speech surrogacy. The study relies on field recordings collected by the authors in Lagos, Nigeria, in February 2020, featuring drummed performances of Yoruba sentences previously unknown to the informants. The recorded sentences were transcribed and analyzed comparatively, which allows us to characterize systematic regularities in the speech-to-drum mapping. Observing how the intrinsic characteristics of language sounds (pitch, duration, intensity and spectrum) are addressed by means of the organologic and acoustic properties of the *dùndún* and the *bàtá*, we conclude that these drums' different properties foster distinct speech surrogacy systems. Alongside a consideration of native perspectives on speech surrogacy, we propose an understanding of drum languages as platforms capable of supporting the development of native theories on sound and language.

Keywords: drum languages, speech surrogacy, bàtá, dùndún, Yoruba, drum music, organology

Edited by:
Laura McPherson,
Dartmouth College, United States

Reviewed by:
Samuel Kayode Akinbo,
University of British Columbia, Canada
Olabode Omojola,
Mount Holyoke College, United States

Correspondence:
Mariano González
m.gonzalez@uu.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *12 January 2021*
Accepted: *05 November 2021*
Published: *21 February 2022*

Citation:
González M and Oludare O (2022) The Speech Surrogacy Systems of the Yoruba Dùndún and Bàtá Drums. On the Interface Between Organology and Phonology. Front. Commun. 6:652542. doi: 10.3389/fcomm.2021.652542

INTRODUCTION

This paper presents the current state of an ongoing investigation into the speech surrogacy systems of the *dùndún* and *bàtá* drums of the Yoruba people in Southwestern Nigeria. Used for communication in a series of daily situations, as well as in civic and religious rituals, the *dùndún* and *bàtá* drums are capable of mimicking the spoken Yoruba language, thus being popularly referred to, both in Nigeria and abroad, as "talking drums."

The main focus of interest of the investigations resides in the fact that the *dùndún* and *bàtá* drums, besides playing distinct cultural roles in the Yoruba community, are very different instruments, and are related to the spoken language in different ways, yet still fulfill a similar function as "talking" drums, or speech surrogates (Beier, 1954; Arewa and Niyi, 1980; Euba, 1990; Klein, 2000; Ruskin, 2013; Oludare, 2020, 2021, and others). On a cultural level, the *bàtá* and *dùndún* differ largely on their contextual applications, as well as on their traditional and religious customs. While the *bàtá* is generally regarded as a religious instrument protected by specific ritual codes of the *orisá* devotion, the *dùndún* is present both in diverse religious settings, as well as in urban popular music. On a musical level, these drums have different tuning systems, number and kind of membranes, as well as varied playing techniques. Each of them displays its own phonological system, which is related to the

Yoruba language based on distinct parameters. The present study aims at comparing the linguistic usages of the *bàtá* and *dùndún*. This is done by studying the regularities found in the mapping between Yoruba phonology and the speech surrogacy practices on the two drums. To this end, we analyze how drummers make use of the specific features of their instruments when performing texts, and which general rules can be identified in the translation from a spoken utterance to a drummed utterance.

Like most African languages, Yoruba is a tonal language, where pitch variations distinguish lexical entries and grammatical processes (Pulleyblank, 2004; Ajiboye, 2011). Yoruba has three relative pitches – Low, Mid and High – known as *ohùn ìsàlè*, *àárín*, and *òkè* in Yoruba terminology. In present-day Nigeria, the three Yoruba speech tones are more commonly referred to as Do, Re and Mi, respectively. These terms are inspired by Western music theory, yet with no particular relation to their conventional frequencies or interval ratios. Additionally, the Yoruba speech tone system is characterized by a series of glides connecting these three basic pitches.

The studied drums display distinct organological features that determine how they are used in rendering Yoruba speech. The *dùndún* drums are flexible tension drums, and are used mostly to imitate the tonal inflections of Yoruba, with three tones, or pitches, and different glides between them. At present, two types of *dùndún* drums are used regularly: *ìyáàlù dùndún* and the *gángan*. These two types of the *dùndún* share the same basic construction, and differ mostly in size, with the *gángan* being about half the size of the *ìyáàlù dùndún*. Unlike the *dùndún*, *bàtá* drums have fixed membranes: two on the *ìyáàlu bàtá* and three (playable) on the *omele bàtá*. The different membranes may be played simultaneously, or separately from one another.

Most actual performances with *dùndún* and *bàtá* drums involve a traditional repertoire of texts, including praise poetry (*oríkì*) and proverbs (*òwe*). However, the musicians are capable of drumming any given text in Yoruba. This paper concentrates on this trait. Our aim is to contribute to the on-going formal description of these speech surrogates (Oyelami, 1989, Euba, 1990 and 2011, Villepastour, 2010, and Akinbo, 2019), by comparing the way in which previously unknown sentences are played on the four types of drums we study.

The research is based on a collection of 12 sentences in Yoruba as performed on the *ìyáàlù dùndún* (Mr. Ayanlere Alajede), *gángan* (Mr. Kangan Bamidele), *ìyáàlù bàtá* and *omele bàtá*[1]. Each drummer was asked to perform the sentences on his instrument. The sentences, previously unknown to the drummers, were composed by the researchers and display varied phrasal structures and vocabulary, and were read out loud to the drummers by a native speaker in the moment of recording. The drummers were allowed to take some time, if necessary, to internalize the sentences before playing them on their drums, but, with the exception of some of the longer sentences, most of the drummers responded immediately after the linguistic stimuli. Each sentence was recorded in three rhythmically distinct versions. The first of them corresponds to what was characterized by Akin Euba (1990) as "speech mode," i.e., without a specific rhythmic background the drummer has to relate to. In the second and third versions, the drummers were requested to perform the same sentences in the "dance mode" (ibid.), which matches two common rhythmic accompaniments, with both 16- and 12-pulse time-line patterns (Kubik, 2008).

The recordings were transcribed and annotated. This was used to identify how different drummers, using distinct instruments in various (musical and linguistic) situations, transfer the linguistic data to a drummed equivalent, and what techniques they use in doing so. The results were used for comparing the speech surrogacy systems supported by the four instruments.

By means of a side-by-side comparison of the *dùndún* and *bàtá* speech surrogacy systems, we focus on similarities between distinct drum sounds, and the sounds of language. For this reason, beyond exploring the mapping between units of speech and drum sounds, we stress the acoustic ground on which this system relies, and how humans perceive these sounds. For that, we use both spectrograms – as an objective tool for the visualization and analysis of sound phenomena – and fragments of interviews with Yoruba musicians, who describe, by means of their own knowledge and sensibility, how the sounds of the drums they play relate to the sounds of the language they speak.

THE DÙNDÚN SPEECH SURROGACY

The Yoruba *dùndún* drums are hourglass-shaped bi-membranophones manufactured in distinct sizes. Here, we concentrate on the *gángan* and on the *ìyáàlù dùndún*, whose basic construction is the same. In a traditional *dùndún* ensemble, the *ìyáàlù* takes the leading role and is responsible for most text-based passages.

The *dùndún* ensemble comprises at least five instruments, belonging to three sub-families: 1) the *ìyáàlù* sub-family, consisting of the *ìyáàlù*—the "mother drum," leader of the ensemble, as well as the *keríkerì*, *ìsáájú*, and *ìkehìn*; 2) the gángan sub-family, consisting of the *gángan*, the *àdàmò* (often called "*gángan*" too; see below), and the *kànàngó*; and 3) the *gúdúgúdú* sub-family, which only has one type of drum. The instruments of the *ìyáàlù* and *gángan* families are all tension drums in distinct sizes, whereas the *gúdúgúdú* is a kettle (non-tension) drum.

Sometimes, two *ìyáàlù* drums may be engaged in the ensemble, with the second referred to as the *èjìn*, providing an effect of a two-tone bass melo-rhythmic pattern. While the *ìyáàlù*, *gudugudu*, *keríkerì*, *ìsáájú*, and *ìkehìn* are the standard drum members of *dùndún* ensembles, any one or all of the *gángan*, *àdàmò*, and *kànàngó* may also be included in the ensemble, often duplicating one of the *keríkerì*, *ìsáájú*, and *ìkehìn* drums, as required by the musical performance (King, 1961; Adegbite, 1988; Euba, 1990; Olaniyan, 2008; Oluga and Babalola, 2012; Osundina, 2015; Olatunji, 2017; Oludare, 2018). It is important to note that the traditional name "*gángan*" is now used interchangeably for two different drums. According to Kangan

[1]Our main *bàtá* informant, a master drummer from *Àyán* tradition, requested to remain anonymous.

Bamidele and Ayanlere Alajede (personal communication on February 3, 2020), the *àdàmò* is now referred to as the *gángan*, which is ubiquitous in the urban settings, favored among young drummers, and adopted in popular and gospel music. In the present paper, we opted to use this current terminology.

The attachment and tuning of the membranes are achieved in the *dùndún* drums using multiple cords made from animal skin, which are fastened on the flexible rims surrounding the membranes, and connect the two membranes to each other without contact to the wooden corpus of the drum. The drums are carried horizontally on the musician's shoulder with a strap and played with the curved stick *opá*, a wooden beater in "J" form. The bigger *ìyáàlù* drum is carried on the height of the musician's waist, while the (smaller) *gángan* is carried in a similar way to a purse, under the axil. The player of the *ìyáàlù dùndún* embraces some of the tuning cords of the drum with the left hand[2], thus they can increase or release the tension of the membranes by squeezing the drum between their fingers and against their waist. On the *gángan*, a similar technique applies, but with the drum being compressed between the drummer's upper arm and ribs. The left hand of the *gángan* player is kept on the upper side of the drum's rim. On this instrument, also the fingertips and nails of the left hand are used to provide percussive sounds and to mute the drum membrane (when required) in-between drum strokes.

This flexible tuning mechanism, combined with the unique fashion in which each of the drums is carried, allows drummers to vary the tension of the membranes continuously by applying more or less pressure to the tuning cords, which affects the drum's pitch. Players use this technique to imitate the three basic Yoruba tones, as well as the tone inflections that occur on rising or falling gliding tones or between syllables of distinct tones.

Playing the *dùndún* drums involves an interplay between the *opá* strokes and changes in the membrane tension. Before striking the membrane, the drummer can adjust the membrane to three basic levels of tension: a loosened stage, without applying any tension to the membrane (producing a low pitch that mimics the low speech tone); a half-way tensioned stage, applying some pressure on the tuning cords (mid speech tone); or a tight stage, when full pressure is applied (high speech tone). After striking the membrane, the drummer can make further modifications to the tuning of the instrument, causing the pitch to rise or fall while the membrane continues to vibrate. This generates gliding pitches that are used to imitate the tonal inflections used in Yoruba speech. Describing this interplay requires isolating single units of action within a sequence of strokes and movements, which we have achieved using the annotation method described below.

The Dùndún Phonology

In his comprehensive work on the *dùndún* drums, Akin Euba (1990) did not have access to the use of spectrograms or any other technology that would facilitate a detailed description or graphic visualization of the nuances of intonation of the *dùndún* drums. Due to these technical limitations, this aspect of the *dùndún* phonology is treated only in a rather rudimentary way in his transcriptions and analysis. Akinbo (2019) takes an initial crucial step into filling this gap, describing a range of distinct possibilities of tonal inflections that can be performed on a *dùndún* drum to represent short words of one or two syllables. The present work contributes to the study of the *dùndún* phonology by describing the regularities found in the ways *dùndún* playing techniques are used to render diverse phonological aspects of Yoruba in longer, varied sentences, largely relying on detailed annotation of spectrograms.

Representing the recordings in spectrograms allows us to isolate the *opá* stroke (or onset) and the sustained sound produced by the vibrating membrane, thus facilitating the description of the sounds used in the course of a *dùndún* performance. On a spectrogram, each stroke appears as a vertical line crossing the whole image. This is due to the fact that the impact of the drum-stick on the skin produces a fair amount of "noise" (diffused sound frequencies), which decays within a very short period of time. Immediately after the onset, the vibrating membrane emits a clearly defined pitch, which can be read on the spectrogram as a horizontal line or as a curve, varying according to the tension applied to the membrane by the player. Accordingly, the spectrogram of a *dùndún* performance contains data on 1) the moment of impact of the *opá*, 2) the initial pitch of the sound, 3) the tuning, i.e., tension variations of a sound, including the moment and rate these variations take place, as well as 4) the intensity (amplitude) and 5) the overall duration of a sound.

Figure 1 shows the spectrogram for an *ìyáàlù dùndún* rendition of the sentence "*Bàbá àgbàlagbà kaán jókòó lórí àga pupa*" ("There is an old man sitting on a red chair"). With a few exceptions, which will be commented on later, each syllable corresponds to one stroke (vertical line) and the subsequent horizontal line or curve to its right.

The shape of each of these lines and curves can be described by observing how the pitch is caused to rise, fall or stagnate as a result of changes in the membrane tension. The sounds produced by a *dùndún* drum can thus be annotated as sound shapes resulting from a sequential combination of specific movements of both arms, and in direct correlation with the phonemes of the Yoruba language. Each of these shapes, as described in the example below, contains precise information on the kind, intensity and speed of the left-arm movements related to the tuning of the membrane, as well as the moment in time in which the *opá* strokes and arm movements take place.

Table 1 describes examples for the annotation of spectrogram shapes encountered on *dùndún* performances and their mapping to the corresponding playing techniques and the spoken syllables they represent. Each shape is formed by one or more vertical lines, indicating the moment of onset (impact of the *opá* on the membrane), and one or more lines or curves indicating the course of the sound with its pitch variations.

The column "playing technique" indicates the initial stage of the membrane (loosened, half-way tight or tight), the position of each stroke (Pos. 1 = first action), and the presence (+) or absence (–) of movement (pressure increase/release/oscillation) and its

[2]All references to the handedness of playing techniques in the present text describe the most common practice for right-handed drummers. There can be differences between individual players.

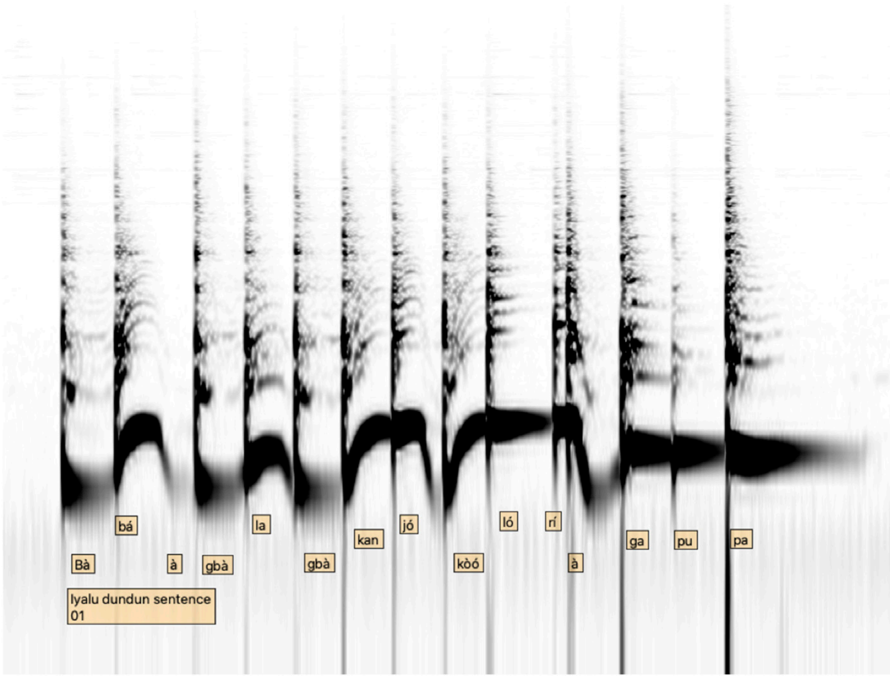
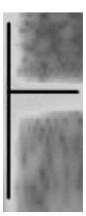


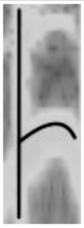
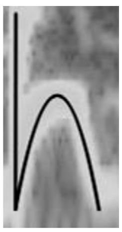
FIGURE 1 | *Ìyáàlù dùndún* spectrogram for the sentence *"Bàbá àgbàlagbà kaán jókòó lórí àga pupa"* ("there is an old man sitting on a red chair"). All spectrograms in the present study were generated with the freeware Sonic Visualizer, available under https://www.sonicvisualiser.org/ (downloaded on January 08, 2021).

TABLE 1 | The phonology of the *dùndún*. The black shapes on the spectrogram images were manually added by the researchers to facilitate reading.

Spectrogram shape	Playing technique	Acoustic resultant	Phonological meaning	Notation	Example
	Membrane: loosen Pos. 1: strike (–) movement	Sound on low pitch	Syllable on low tone (do)	[do]	**Bà**bá
	Membrane: half-way tight Pos. 1: strike (–) movement	Sound on mid pitch	Syllable on mid tone (re)	[re]	À**ga**
	Membrane: tight Pos. 1: strike (–) movement	Sound on high pitch	Syllable on high tone (mi)	[mi]	**Ló**rí

(Continued on following page)

TABLE 1 | (*Continued*) The phonology of the *dùndún*. The black shapes on the spectrogram images were manually added by the researchers to facilitate reading.

Spectrogram shape	Playing technique	Acoustic resultant	Phonological meaning	Notation	Example
	Membrane: half-way tight Pos. 1: strike (+) movement: oscillation up-down	Sound on mid pitch with oscillation	Syllable on mid tone (re)	[re~]	àgbà**lag**bà
	Membrane: tight Pos. 1: strike (–) movement (+) movement: pressure release: quick	Sound on high pitch dropped after sounding	Syllable on high tone before syllable on low tone	[mi ∨]	**Jó**kòó
	Membrane: loosen Pos. 1: strike (+) movement: pressure increase: quick (+) movement: pressure release: quick	Upwards glide followed by downwards glide	Syllable on high tone followed by syllable on low tone on the same vowel (diphthong)	[∧ mi] (vdo)	Bàb**á àg**bà
	Membrane: loosen Pos. 1: strike (+) movement: pressure increase: slow	Sound on low pitch followed by upwards glide	Syllable on low tone followed by syllable on high tone on the same vowel	[do] (∧ mi)	Jó**kòó**
	Membrane: tight Pos. 1: strike (–) movement Pos. 2: strike (+) movement: pressure release: slow	Double stroke on high pitch followed by downwards glide	"quick" syllable on high pitch followed by syllable on low tone	[(mi) (vdo)]	Ló**rí àg**a

rate (slow or quick) with the left arm of the player. These shapes and their classification form a basic repertoire of sound units that can be used to describe and compare *dùndún* performances of speech texts, with every shape corresponding to one or sometimes two syllables. The columns "acoustic resultant" and "phonological meaning" describe how each drum sound is perceived acoustically and how it is related to the sounds of spoken Yoruba.

The "notation" column represents drum sounds with minimal number of symbols, as follows:

- The notes do, re and mi refer respectively to the low, mid and high tones of Yoruba speech as used conventionally today by Yoruba speakers in allusive manner, without any correspondence to the actual pitches or interval ratios they represent in Western music theory.
- The symbols ∧ and ∨ placed before a tone name signalize that the tone in question was reached through an up- or downwards glide, respectively.
- The symbols ∧ and ∨ placed after a tone name signalize an up- or downwards glide after the tone in question.
- The symbol ~ represents a pitch oscillation.
- The square brackets [] indicate that the sound between them was produced by a strike with the playing stick opá.
- The round brackets () indicate that the sound between them was concatenated to the previous sound through a glide without a new *opá* strike immediately before it.

- Double strokes, produced by the rapid re-bounce of the opá stick, are represented as e.g., [re re] or [re mi].

Departing from the description of the shapes seen above, it is possible to establish a generative model, or finite state machine, describing how each membrane tension stage is assigned to one Yoruba speech syllable in the *dùndún* speech surrogacy, as well as how single tension stages can be concatenated between strokes, rendering the intonational glides of Yoruba speech.

In the model above (**Figure 2**), the L, M and H states represent the different levels of membrane tension, which are related to the three tones of Yoruba speech, commonly referred to as do, re and mi, respectively. The arrows represent the possible transitions from one tension state to another or the permanence on one state, as a result of the increase, decrease or maintenance of the pressure applied by the player's left arm and hand to the tuning cords. The transitions are annotated according to their interpretation in Yoruba speech and identified with the help of a number and letter combination (1a, 2a etc.). Each sound is understood as the transition or the concatenation of transitions occurring after a stroke and before the next stroke. Normally, not more than two transitions are concatenated, due to the limited duration of the audible vibration of the membrane.

Through a nuanced stick playing technique, the player of the *ìyáàlù dùndún* is also able to take vowel intrinsic intensity into account when modulating the force with which the membrane is hit (Villepastour, 2010, p. 83). On the amplitude wave in **Figure 3**, it is noticeable that the syllables containing the vowels high vowels /i/ and /u/ are played in a significantly softer manner (lower amplitude), when compared to the surrounding strokes.

This tendency can be observed throughout the analyzed data, even there, where achieving this contrast in amplitude demands a complex playing technique. In the beginning of the double strokes "*sí ojà*" and "*mi ra*", shown on **Figure 4**, the achievement of the contrast between soft and strong strokes requires the first attack of each double or re-bouncing stroke to be played softer than the second attack, which means that the player must skillfully increase the beater's energy between both strokes within a very short time frame.

Although the representation of vowel height is present throughout the *ìyáàlù dùndún* data that we collected, it should be pointed out that it is not necessarily recognized by the artists. Our main informant for the *ìyáàlù dùndún*, Mr. Ayanlere Alajede, a prestigious senior *dùndún* master drummer, suggested that vowel height is not part of the *ìyáàlù dùndún* speech surrogacy. For him, this practice relies solely on the reproduction of the pitch component of Yoruba speech. Evidently, the relation between vowel height and amplitude cannot be expected to be consciously perceived by any speaker

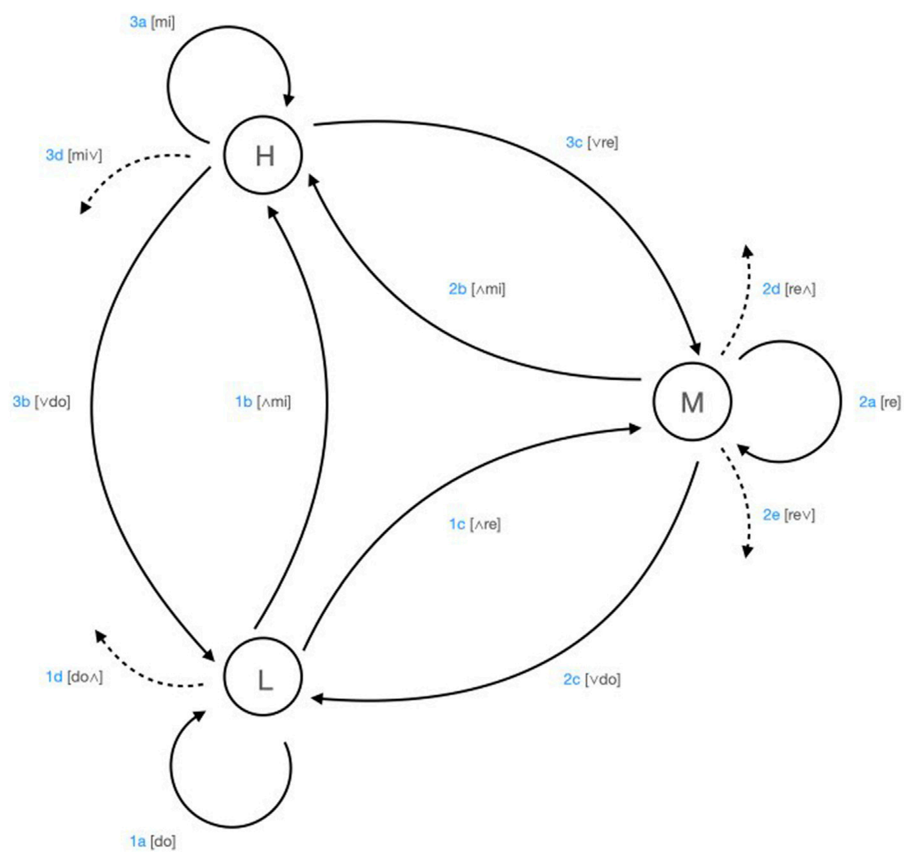


FIGURE 2 | Graphic representation of the dùndún speech surrogacy system as a finite state machine.

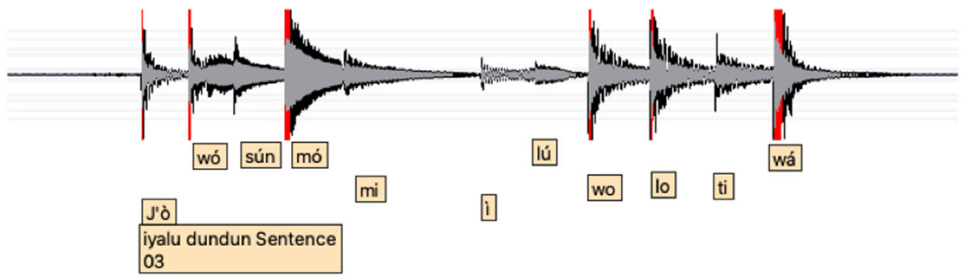


FIGURE 3 | *Ìyáàlù dùndún* amplitude wave for the sentence "*J'òwó sún mó mi. Ìlú we lo ti wá*" ("Please come here. Where are you from?").

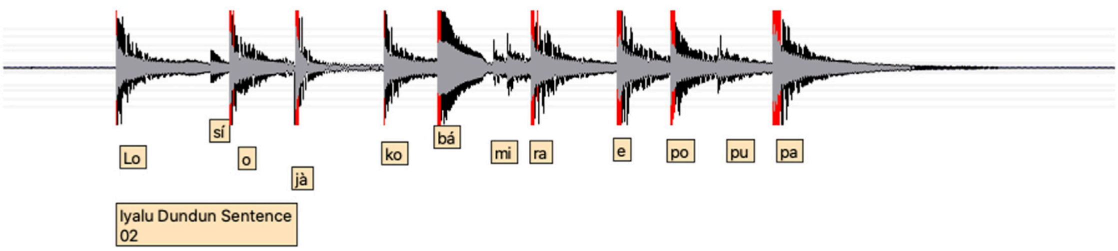


FIGURE 4 | *Ìyáàlù dùndún* amplitude wave for the sentence "*Lo sí ojà k'o bámi ra epo pupa*" ("Go to the market and buy me palm oil").

of any natural language, as this relation is hardly conceived without means for acoustic measurement. The relation between vowel height and amplitude can be considered as an acoustic by-product of vowel intrinsic intensity[3], rather than a phonological distinction. The fact that the imitation of speech by highly skilled drummers is sensitive to such relations gives us a hint as to the extent to which "talking" and "drumming" are kept close by masters of this craft.

The *gángan*, apart from being smaller in size than the *ìyáàlù dùndún*, features the same construction elements and a similar playing technique. The most relevant differences between the two drums, as seen in our dataset, concern the glides and the rendering of the /r/ consonant, present to a lesser extent on the *gángan*, as well as the differentiation between high and non-high vowels by means of amplitude, absent on our *gángan* recordings. The absence of these traits on the *gángan* speech surrogacy can be tracked back mostly to the cultural context of this instrument in present-day urban Nigeria. According to Kangan Bamidele (personal communication on February 4, 2020), most *gángan* players today only learn the instrument in an urban environment, thus learning only the necessary playing techniques required to apply the drum in the *orin àsìkò* (popular music) or *ìgbàlódé* (modern music) settings. Ayanlere Alajede also argues that, while rendering intonational glides and the /r/ consonant should be regarded as the most correct practice for all *dùndún* drums, including the *gángan*, the modern day *gángan* players often omit them due to their limited knowledge of the tradition. He added that only those who learn through the Ayan (Yoruba indigenous knowledge system of drumming) apprenticeship will know the *àṣà* (culture) and *ìṣe* (practice) of the drum, with regards to communicating in Yoruba Language and mastery of its full performance techniques (personal communication on February 4, 2020). Besides cultural aspects of performance context and apprenticeship, the small membrane size of the *gángan* can be regarded as an additional organological reason for the different speech surrogacy practices on the two studied *dùndún* drums. With a much shorter sustain time—the time during which the membrane keeps vibrating and sounding after a stroke—the gángan membrane is limited in its possibilities of performing glides, for which a long sustain time is advantageous. Equally, its small size imposes limitations on tuning and amplitude variations.

THE ÌYÁÀLÙ BÀTÁ SPEECH SURROGACY

Organology

The *ìyáàlù bàtá* is a large sized (ca. 80 cm length) bi-membranophone drum with a conic wooden corpus. The two membranes of the *ìyáàlù bàtá* are of different sizes and are called *ojú òjò*, and *ṣáṣá*, with their diameters measuring around 35 and 20 cm, respectively. The membrane attachment system of the *bàtá* drums consists of a net of animal skin ropes connecting both membranes to each other. Unlike the *dùndún* drum, the membrane attachment system of the instruments of the *bàtá* family is tightly tied to the

[3]See the spectral analysis of Yoruba vowels in the discussion on *bàtá* vowel representations below.

drum corpus. As a result, the drummer is not able to adjust the drum's tuning on the go.

The traditional *bàtá* ensemble consists of the *ìyáàlù bàtá*, the *omele bàtá*, and either the *omele méjì* (*omele ako* and *kúdi*, tied together), or the *omele méta* (*omele ako*, *kúdi*, and *àdàmò*, also known as *aféré*, tied together). Like in the *dùndún* ensemble, a second *ìyáàlù* might also be used to provide the *èjìn*. Both the *omele méjì* or *omele méta* are played by one person, and only one of the pairing is used in a *bàtá* ensemble (Oyelami, 1991; Villepastour, 2010; Omojola, 2012). According to Euba (2011), and corroborated by our informants, the *omele méta* is a recent development in the *bàtá* drumming tradition, in response to the wish to give the *omele bàtá* some liberty to perform independent of the *ìyáàlù*. A third and different pitched drum was thus added, so the *omele méta* can imitate the three tones of the Yoruba language in achieving speech surrogacy, hence our choice of the *omele méta* over the *omele méjì* for this study.

To achieve the desired deep sound of the *ojú òjò*, the large skin of the *ìyáàlù bàtá*, a portion of a tuning paste called *ìda* is stuck close to the center of the drum, forming a circle or a ring around 3 mm thick and 5 cm wide. This paste is usually kept on the membrane during storage, and drummers frequently make small adjustments on the shape, placement or even amount of paste applied to their drums, striving to achieve the desired sound.

The instrument is carried on the drummer's right shoulder in horizontal position in front of the player, approximately on the waist's height, with the large skin *ojú òjò* on the left, and the small skin *ṣáṣá* on the right side of the player. The playing technique of the *ìyáàlù bàtá* is different for each membrane. The *ojú òjò* is played with the bare hand using a palette of three basic strokes described below. For the *ṣáṣá*, a flexible strap of dry animal skin called *bílálà* is used as beater.

The *ojú òjò* has the following playing techniques:

- Open tone: The four fingers of the flat (left) hand are held together, hitting the membrane evenly and close to the drum rim, bouncing back after the stroke. This allows the membrane to vibrate freely, revealing the fundamental frequency of its sound spectrum. This technique produces a low and sustained sound.
- Muted tone: This stroke is similar to the open tone described above, however, here the hand is maintained on the membrane after the stroke, muting the skin. With this technique, the membrane is not allowed to vibrate freely, cancelling the fundamental frequency of its sound spectrum and producing a muted sound with short sustain time.
- Slap: In this technique, the four fingers of the left hand are held slightly spread and hard, striking the skin towards its center, with the fingertips reaching the region covered by the *ìda* paste. In this stroke, the membrane is muted both by the fingers and by the region around the palm of the hand. The combination of the whip-like finger technique and the large contact surface between the player's hand and the membrane creates a bright and short sound of high frequency spectrum.

The *ṣáṣá* is played with the *bílálà* using one single technique, in which the skin strip is embraced by the closed four fingers of the right hand and supported by the stretched thumb. The membrane is hit hard by the *bílálà* with a quick wrist or lower arm movement, producing a loud and high-pitched sound.

Basic Rules of the *Ìyáàlù Bàtá* Speech Surrogacy

On the *ìyáàlù bàtá*, the Yoruba speech tones are generally rendered using the three playing techniques of the ojù òjò described above, with the open, mute and slap strokes rendering the low, mid and high tones, respectively.

The following general regularities of the *ìyáàlù bàtá* speech surrogacy were first described by Muraina Oyelami (1992), and have been observed throughout our collected data with only rare exceptions:

1. The three *ojú òjò* sounds (open tone O, mute tone M and slap S) are assigned to the three tones of Yoruba language (*ohùn ìsàlè*, *àárín*, and *òkè*, or do, re and mi) when performing syllables containing the high vowels /i/ and /u/ (1a, 2a and 3a on **Figure 5**).
2. The simultaneous playing (represented by the double line) of the *ṣáṣá* stroke Sa and one of the *ojú òjò* sounds is assigned to the respective Yoruba tone when performing syllables containing the non-high vowels /a/, /e/, /ẹ/, /o/ and /ọ/, (1, 2b, and 3b).
3. The *ṣáṣá* stroke, when performed alone, is assigned to the Yoruba tone mi, represented here as mi*, and can appear in some situations described below (4a).

The *ṣáṣá* stroke has two functions: 1) as a vowel intensity marker, when performed additionally to an *ojú òjò* stroke; and 2) as a high pitch (mi), when performed alone. In specific situations shown below, this second function may overrule the first, when a high-pitched syllable is represented by the *ṣáṣá* stroke only, even on syllables that would by default be represented by a joint *ojú òjò* and *ṣáṣá* strokes.

On the *ìyáàlù bàtá*, both membranes are attached to the two ends of the same conic corpus, thus connected to each other by the same air column. A clear pattern of resonation can be seen between both membranes, causing their frequency spectra to overlap and their sounds to blend in an optimal way. An aural interpretation of this phenomenon would be, that on intense vowels, the *ṣáṣá* is added as to enhance the high frequency spectrum of the *ojú òjò* sound.

Additionally to the rules above, the following situations have proven to appear on a consistent basis:

1. Vowel sequences

Most double vowel sequences in spoken Yoruba require an intonational continuity, or, in the case of H-L and L-H sequences, an intonational glide. When rendering such double vowel sequences on the *ìyáàlù bàtá*, both the *ojú òjò* stroke of the high-pitched syllable and the (eventual) *ṣáṣá* stroke of the low or

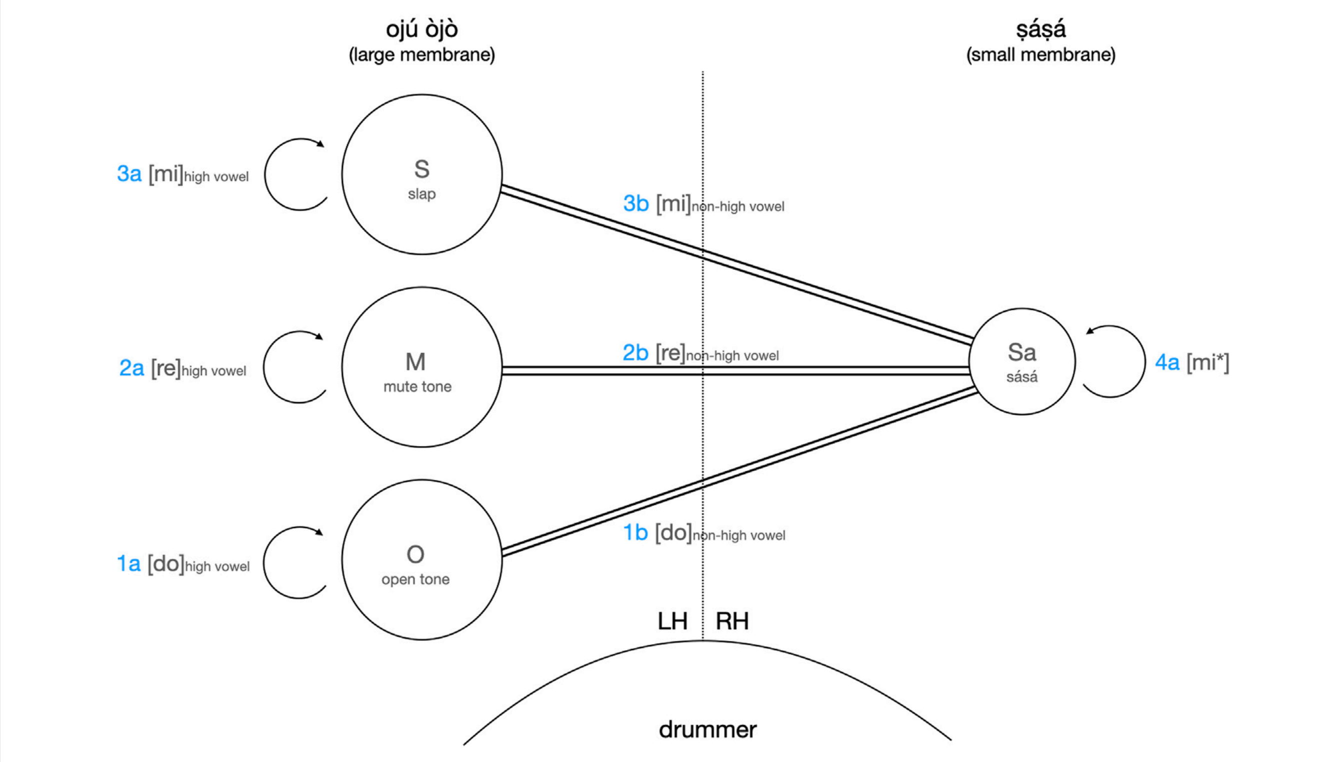


FIGURE 5 | Graphic representation of the *ìyáàlù bàtá* speech surrogacy system.

mid pitched syllable are left out. This can happen both in upwards (mid-high or low-high) sequences, as well in downwards (high-mid and high-low) glides (Villepastour, 2010, p. 54).

2. Contractions

In spoken Yoruba, contractions between syllables and words are frequent, such as in *kí o*—becoming *ko*. In these cases, drummers represented either the full expression or the contracted version. In one example, when asked to perform a sentence containing the contraction *ko* in a sentence, as pronounced by the researcher, the drummer opted to perform the full expression *kí o*, which requires two, instead of one single stroke. Following the handling of vowel sequences described above, the syllable *kí* was performed on the *ṣáṣá* only – contradicting the "rule" that syllables containing the high vowel /i/ are performed on the *ojú òjò* membrane, and the syllable o was performed as a mute stroke on the *ojú òjò*, thus dropping the *ṣáṣá* stroke that would be otherwise required when rendering a syllable on an non-high vowel /o/.

3. Contrast and compensation

In some cases, drummers may choose not to articulate (by not playing a stroke on) one or a few syllables. This happens mostly due to sound similarity (either by vowel, in CVV sequences, or by pitch) to the preceding syllable(s), a procedure common also in spoken Yoruba (Orie and Pulleyblank, 2002). After an omitted syllable, a high-pitched syllable is likely to be represented by a *ṣáṣá* stroke only, regardless of its vowel. The function of this transformation could be understood as a compensation of the preceding omission, as rendering a high-pitched syllable on *ṣáṣá* only would enhance its contrast to the surrounding syllables. Since a single *ṣáṣá* stroke is always understood as a mi, regardless of vowel, using this stroke after an omitted syllable could be regarded as a way of reinforcing the pitch of the subsequently played syllable.

A glance on a *ìyáàlù bàtá* spectrogram (**Figure 6**) reveals that its speech surrogacy is based solely on combinations of the sounds shown above:

The sustained low fundamental frequency of the open tone can be recognized on the spectrogram above as a horizontal line at the bottom of the sound spectrum (e.g., syllable *Bàbá*, at the beginning of the sentence). The mute tone appears in the shape of a white vertical line and slightly higher on the spectrum, indicating an intense sound of short duration and covering a rather wide range of medium-low frequencies (e.g., the three syllables *àga* pupa in the end of the phrase). The slap can be identified on the spectrogram as a thin vertical line of low intensity and short duration, covering a range of medium-high frequencies (e.g., syllable *Bàbá*, at the second position). The *ṣáṣá* strokes appear on the top of the spectrum and can be visually distinguished from the *ojú òjò* strokes as a range of intense high frequencies. From all syllables on this sentence, only *ló**rí*** and ***pu**pa* are rendered without the addition of the *ṣáṣá* stroke, following the rule

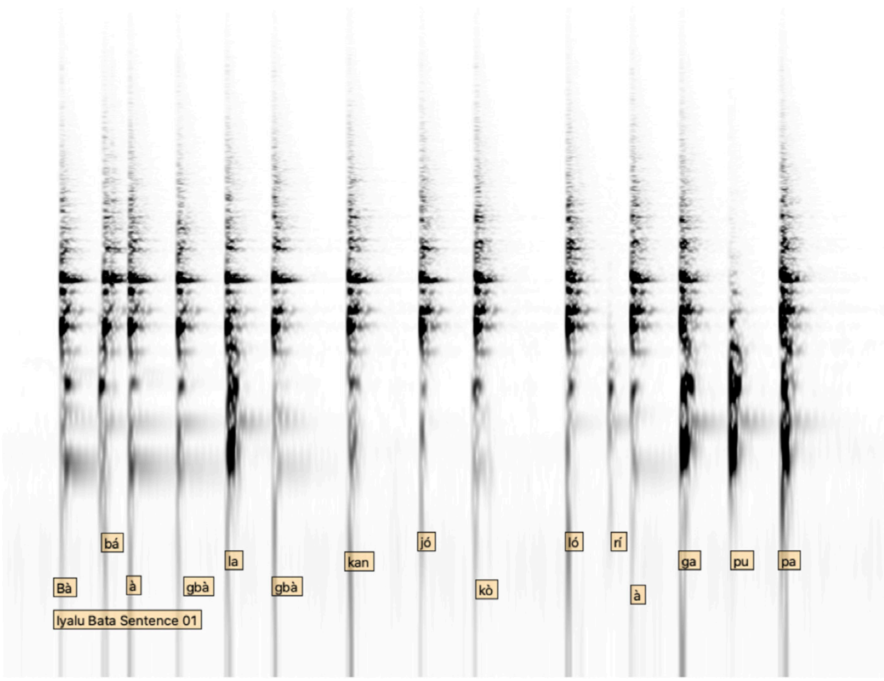


FIGURE 6 | *Ìyáàlù bàtá* spectrogram for the sentence "*Bàbá àgbàlagbà kaán jókòó lórí àga pupa*" ("there is an old man sitting on a red chair").

described on the table above regarding the representation of syllables containing the vowels /i/ and /u/.

The Phonology of the *Ìyáàlù Bàtá*

Of the set of 12 sentences (version 1, in speech mode) performed on the *ìyáàlù bàtá* for the present study, 51 of the 55 occurrences of the vowels /i/ and /u/ (including the tonal variations /í/, /ì/, /ú/ and /ù/)) were performed according to the rules stated above, proving this to be a consistent trait of the *ìyáàlù bàtá* speech surrogacy. However, the fact that the *ìyáàlù bàtá* does not render the language tones through pitch distinctions (as it is the case on the dùndún speech surrogacy), but rather also through contrasting sound spectra emerging from the different hand playing techniques, leads to further questions regarding the nature of the mapping between drum and language sounds seen here.

Different from the low, mid and high pitches and their contours as performed on the *dùndún* drum, which consist of different frequency modulations of the same basic timbre, much like the different keys of a piano, or different notes on a violin, the basic sounds of the *ìyáàlù bàtá* are mapped to the Yoruba tones based on their timbre. Timbre is acoustically described as a combination of factors including attack and decay rates, overtone structures (the sound spectrum), and wave shapes. As such, it is not per se a matter of pitch, but of sound quality, and thus phonetically closer to the characterization of vowel sounds than to aspects of linguistic tone. Thus, the different degrees to which the spectra of the open tone, the mute tone and the slap of the *ojú òjò* contain the fundamental frequency of the vibrating membrane (present virtually only in the open tone), and its overtones (present in increasing degree on the mute tone and slap), correspond acoustically to the overtone spectra responsible for the distinction between vowels in ordinary speech under the influence of formants. The question arising from this observation is why and how pitch and timbre seem to appear here as exchangeable representations of speech tone on the drums.

One important hint can be extracted from the emic perspective represented by our main *bàtá* informant, a master drummer from the Ayan tradition, who opted to remain anonymous. He states that the general practice is to play the *ṣáṣá* simultaneously with the *ojú òjò* anytime a word with *ohùn òkè* (high pitch "mi") is to be rendered, and that it is only on rare occasions that a high-pitched syllable is performed only the *ojú òjò* (personal communication on February 5, 2020). According to him, this tends to happen during fast performance, as a facilitating technique in quick rhythms. Thus, the connection between drum strokes and vowel qualities that our dataset shows is explained as a practical matter of performance speed.

This same apparent contradiction is visible in Villepastour's argumentation (2010, p. 51), whose main informant also conceptualizes the playing of the *ojú òjò* slap sound without the *ṣáṣá* on high-pitched syllables as a matter of performance tempo. Villepastour, in trying to understand the vowel height representation on the *ìyáàlù bàtá*, refers to Hughes (2000) observations on the usage of the so-called "nonsense-syllables" in mnemonics of diverse musical styles across the globe, defending some universals in the way (instrumental) musical sounds are vocalized, whereby low sounds, e.g., large drums or deep string registers, tend to be imitated by syllables containing the vowels /i/ and /u/. Her justification for this mapping, is that "while these (the /i/ and /u/) vowels take less time for a human to utter, they also take less time for an *alubàtá* to play, as he only uses the strong hand to do so" (Villepastour, 2010, p. 51). This justification, while still lacking on quantifiable evidence, appears to be in line with a specific native perspective on the *bàtá* speech surrogacy, which links the dropping of the *ṣáṣá* strokes in some cases to concerns of performance tempo.

Yet, whereas time, according to the arguments above, seems to be the central factor from the perspective of the physical processes of sound production, an aural-acoustic approach offers an alternative justification to the mapping between high-vowels and the absent *ṣáṣá* stroke evident in the data. Oyelami states that "what they [the strokes on the *ojú òjò* only (without *ṣáṣá*) and the high vowels] have in common is the fact that the /i/ and /u/ are similar when you think of softness . . . when you don't have the *bílálà* together with it" (personal communication to Villepastour, 2010, p. 49), arguing in favor of a mapping based on sound perception and aural sensibility, rather than one based on physical constrains of sound production.

It is relevant to ponder on what Oyelami could be in fact alluding to, when referring to the "softness" of the vowels /i/ and /u/. Besides being considered vowels of low intrinsic intensity – a relevant factor for the *ìyáàlù dùndún* speech surrogacy, as described above – the closed vowels /i/ and /u/ also share a similar organization of their harmonics, or overtones, which differs from the spectra of open or non-high vowels. The image below (**Figure 7**) shows the contrast between the high vowels /i/, /u/ and the non-high /a/ vowel as pronounced by a Yoruba native speaker. It is noticeable that the overtone structure of the /a/ vowel occupies a much higher frequency range, while the spectra of the /i/ and /u/ vowels are more compact. Besides, both the /i/ and /u/ spectra concentrate more energy on the first and third overtones, which vibrate in a 2/1 and 4/1 harmonic ratio to the fundamental, respectively, and are perceived as a consonance or even as the same sound (*unisono*). The spectrum of the /a/ vowel, instead, activates a higher number of frequencies of diverse, more dissonant proportions. Both these characteristics of high vowels–the more compact spectrum, and the concentration of energy in more consonant overtones – contribute to a perceived "softness," when compared to the diffuse spectrum of non-high vowels, which can be perceived as harsh or shrill.

Similarly, the spectral analysis of *ìyáàlù bàtá* strokes (**Figure 8**) shows that, while *ojú òjò* strokes – as used to render the /i/ and /u/ vowels – display a rather low and compact overtone spectrum, the *ṣáṣá*, commonly added to the *ojú òjò* when performing all further (non-high) vowels, displays a high and more diffuse spectrum.

While addressing this question, additionally, it has to be kept in mind that the *ìyáàlù bàtá* is a two-headed conical drum (in the Sachs-Hornbostel systematics of musical instruments: 211.252.818, Sachs et al., 1986, pp. 174–181), whereby each membrane is at one end of the same (closed) air column. This means that every vibration caused on one membrane will naturally resonate on the opposite membrane, interfering with its own vibration and causing the sound spectra of both

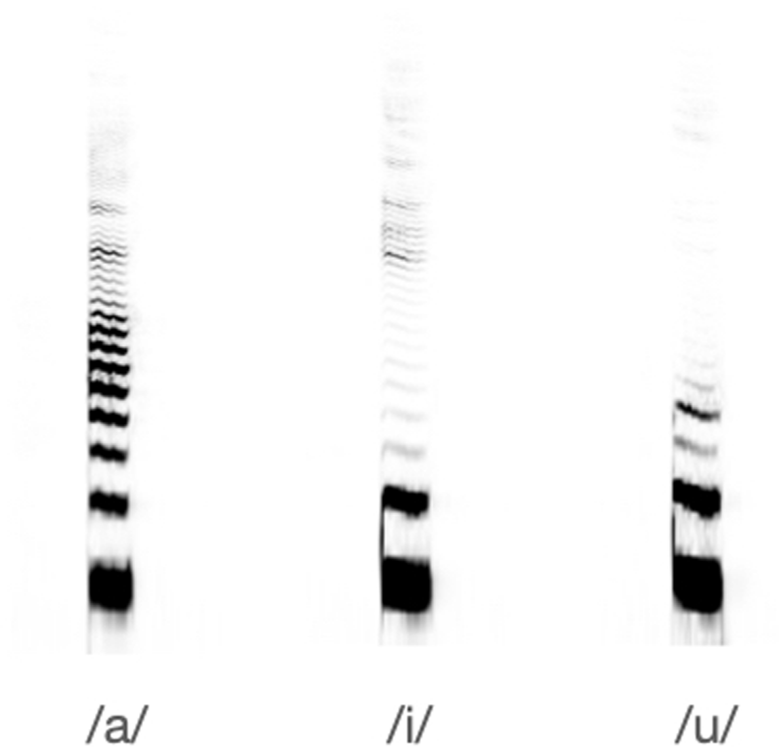


FIGURE 7 | Spectra of the vowels /i/, /u/ and /a/ as pronounced by a Yoruba native speaker.

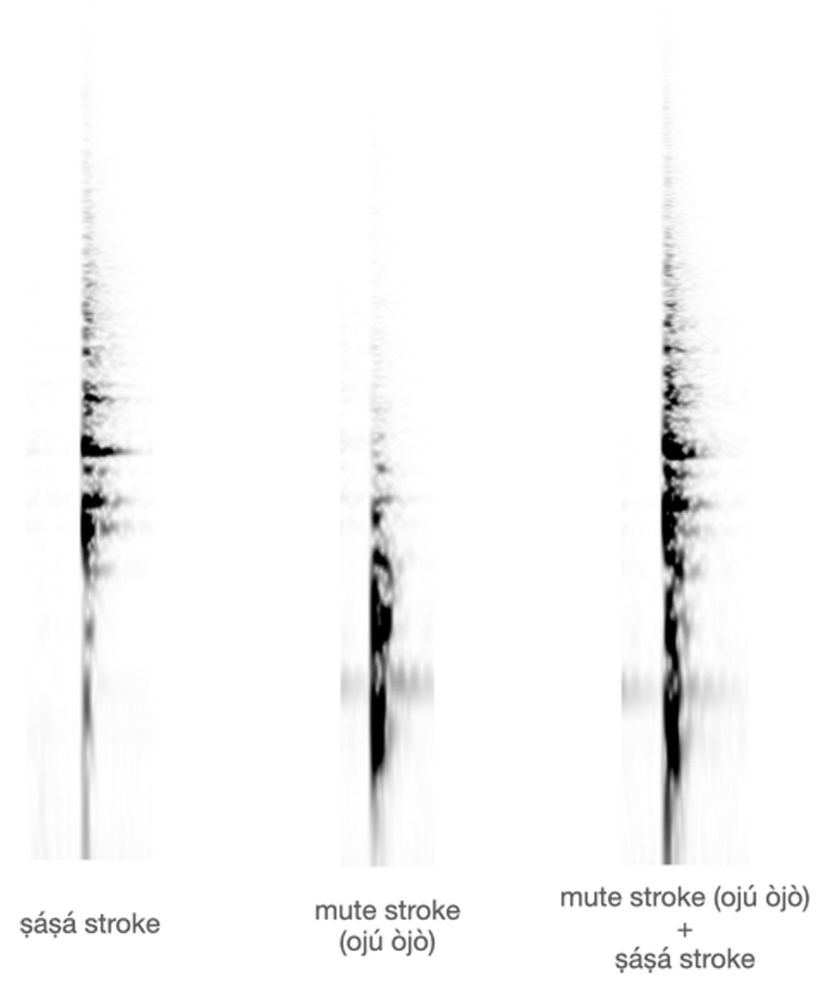


FIGURE 8 | Spectra of *ìyáàlù bàtá* strokes.

drumheads to merge. This is noticeable on the spectrogram images of the different *ìyáàlù bàtá* stroke combinations, showing that the frequency spectra of the *ṣáṣá* and the *ojú òjò* membrane do overlap each other partially on the mid region. Thus, more than simply combined, the two independent membrane sounds can be said to be acoustically intertwined.

The consideration of the spectral characteristics of the /i/ and /u/ vowels in relation to this particular organological feature of the *ìyáàlù bàtá* reveals that, while the imitation of the low, mid and high speech tones on this drum is achieved by means of low, mid and high spectra, respectively, its vowel representation maps vowel spectra to actual drum sound spectra, imitating them through their own acoustic means. This observation is crucial in questioning Villepastour's classification of the *dùndún* and the *bàtá* drums as standing on opposite sides of a continuum between "mimicking" and "encoding," whereby the bàtá would rather "encode" than "mimic" language (2010, p. 90). Understanding the *ìyáàlù bàtá* as an "encoder" obscures the factual sound similarities between language and drum sounds which underlie its speech surrogacy system.

Yet, in spite of the sound similarities mentioned above, the *bàtá* drum language is considered, even among Yoruba cultural insiders, as difficult to understand, especially when compared to the *dùndún*. The *bàtá*, then, is often referred to as an *akólòlò*, a polysemous Yoruba term (Bello 2020), whose interpretation and most adequate English translation itself have become subject of discussion among scholars. Villepastour devotes a great deal of attention to this term, refuting the translation of *akólòlò* by Láoyè—himself a drummer, music researcher and Yoruba king—as "stammerer" (Láoyè 1959 cited by Villepastour, 2010, p. 122) and suggesting an interpretation as "the one who speaks in a staccato manner" (Villepastour, 2010, p. 122), shifting the focus to the acoustic impression of the *bàtá* drum language, which consists indeed of sequences of short sounds, in contrast to the sound continuum promoted by the long sustain time of the *dùndún*.

However, the term *kólòlò* means more than "stammer" or "stutter." It is also used, for instance, for someone who, in spite of having no speech impediment as such, is not articulate or confident in speech, and especially someone not fluent in communicating in Yoruba Language. In other words, just as a person who stammers due to a speech impediment is referred to as "*akólòlò*," so also is a person (native or non-native speaker) called "*akólòlò*," if they cannot express themselves clearly enough in Yoruba Language for another native speaker of the language to comprehend the message or communication easily and fully. Yoruba scholars such as Euba (2011), Omojola (2012), and Olatunji (2012), discussing the reason why the *bàtá* is called *akólòlò* in Yoruba culture, support that the word *akólòlò* does not imply that the *bàtá* has a speech impediment, but that a single *bàtá* drum cannot independently communicate in Yoruba. This is because, in the traditional performance setting of a *bàtá* ensemble, the *ìyáàlu bàtá* is supported by the *omele abo bàtá*, whose player "doubles" the *ìyáàlù*'s mute strokes with its own representation of the mid speech tone, enhancing the intelligibility of the message. Evidently, this combination of two drums played by two drummers to render one same sentence can only be achieved within performances of previously known texts, which is different from the experimental approach pursued by the present study but represents the majority of the conventional practice of a bàtá ensemble. The semantic pluralism in the use of *akólòlò* in Yoruba culture can be better understood by one of many rich Yoruba proverbs—"*b'ópé b'óyá, akólòlò á pe bàbá*" ("eventually, a stammerer will say the word father"). One of the proverb's interpretation is that a stammer can and will eventually communicate his/her message, if given time, space, and opportunity. In this context, the *bàtá* drum can and will talk if given the time (it is longer and more difficult to comprehend the *bàtá* drum text than the dùndún), space (the *ìyáàlù bàtá* engaging different timbre on the *ojú òjò* and *ṣáṣá* strokes for its speech surrogacy), and opportunity (the *ìyáàlù* and *omele abo bàtá* combining for complete communication of sentences). According to this interpretation, *bàtá* drums are referred to as "*kólòlò*" (stammers) not because they cannot "mimic" the Yoruba Language, but because it is more difficult for a single bàtá to communicate efficiently than it is for a single dùndún drum. This is also one of the reasons why in Yoruba culture, the *dùndún* is more popular for musical arts and speech communication than the *bàtá*.

The *Bàtá* Players' Perspective on Yoruba Speech Tones

Yoruba speakers refer to the three speech tones as *ohùn ìsàlè* ("low/down voice"), *ohùn àárín* "mid voice"), and ohùn òkè ("high/upper voice"). However, in addition to these native terms, Yoruba drummers also view and relate to these words in a native acoustic concept, that is, the way it appears and sounds to them from their drumming practice. For example, due to the deep reverberations of the *ohùn ìsàlè* on the drums, they often refer to such tones as "*ohùn kèkè*." The *kèkè* is an onomatopoeic word representing the bass acoustic reverberation, gotten from the repetition of the Yoruba word "*kè*" (deep, bass sound) with a low pitch. The *ohùn àárín* is referred to as kele, two words with the same (mid pitch) tone sound, with the "*le*" being an onomatopoeic word representing a similar or repeated soft acoustic sound. The drummers refer to *ohùn òkè* as either *kúrún* or *aago* (or *agogo*). In the Yoruba language, just as a high pitch sound is perceived as *kúrún* (with two high pitch syllables), any acoustic sound with high pitch is also said to "*ró bíi aago*" (ring like a gong/bell). Similarly, a person (or instrument) with a high-pitched voice is said to have a voice like a gong/bell (*al'óhùn bíi aago*). Hence, both linguistically and acoustically, the Yoruba people (speakers and drummers) conceive speech intonation in their language and (drum) music both in space (o*hùn ìsàlè, àárín*, and òkè) and sound (o*hùn kèkè, kele,* and *kúrún or aago*).

The Secret Language *Ẹná Bàtá*

Ẹnà bàtá is a secret language used by *bàtá* players and plays a major role in Villepastour's monograph on *bàtá* speech surrogacy. Villepastour presents *ẹnà bàtá* as a "formal language" that acts as an "interface" (p. 107) between spoken Yoruba and the "machine language" (idem) represented by the bàtá drum speech rendering. This view, however, is contested by Euba (2011, p. 518), who considers the ẹná bàtá to be of lesser importance for the speech surrogacy on bàtá drums.

Villepastour's account of the *ẹná bàtá* does not constitute a proof of the necessity of this artificial language for the praxis of *bàtá* speech surrogacy. On the contrary, the linguistic information contained in sentences played on a *ìyáàlù bàtá* can be regarded as even closer to its spoken equivalent in Yoruba language, than its *ẹná bàtá* version. Although Villepastour's main informant, Ayandokun, seems to employ *ẹnà bàtá* systematically as an important step in the translation from a spoken into a drummed phrase, Villepastour does not extensively argue for the need she sees "for an interface (that is, an intermediate language) between the natural language of Yorùbá and the machine language of the bàtá" (Villepastour, 2010, p. 108). As a form of "oral notation" (Kubik, 2010), it is clear that *ẹná bàtá*, when used for this purpose, can facilitate the memorizing and teaching of a specific drum pattern. However, it cannot be regarded as a linguistically necessary step for the production the *bàtá* drum language as such, but as an – undoubtedly interesting and culturally important—related subject.

Instead of a crucial interface between spoken Yoruba and the *bàtá* drum language, as postulated by Villepastour, we propose that *ẹná bàtá* could be better understood as a helpful method for teaching and memorizing and as a specific form of *code talking* (Isola, 1982), which is a common practice among the Yoruba. In this sense, in *ẹná bàtá*, *bàtá* drummers may be taking advantage of one aspect of their art when transforming the Yoruba language to create their own form of code talking.

THE OMELE BÀTÁ SPEECH SURROGACY

The *omele bàtá* consists of three small sized drums (ca. 20 cm height) that are bound together and hung with a strap on the drummer's shoulder. The construction of the three drums is identical and resembles a miniature version of the larger bi-membranophones of the *bàtá* family, such as the *ìyáàlù bàtá* discussed in this study, featuring the same drum shape, type of drumhead and rim, as well as a similar attachment system with animal skin. In spite of having two membranes, the drums are carried vertically in front of the player, whereby only the up-facing larger membranes are played. The membranes are hit with two flexible strips of dried animal skin called *bílálà*, identical to the one used to hit the *ṣáṣá* (small membrane) of the *ìyáàlù bàtá*.

Unlike the *dùndún* drums discussed above, the tuning of the drums of the *bàtá* family is fixed, and no changes of tuning can be undertaken by the player on the go. However, a tuning paste called *ìda*, made from tree resin and charcoal, is applied to the center of two of the larger membranes of two of the drums, lowering their pitches. Normally, drummers will check the tuning before playing, and adjust the position, shape and, if necessary, also the quantity of *ìda* applied to each drum, in order to get the desired sound.

Basic Rules

The speech surrogacy of the *omele bàtá* can be said to unite some of the advantages of both the *dùndún* and the *ìyáàlù bàtá* drums. With three different membranes tuned to three contrasting pitches, the *omele bàtá* is able to represent the three tones of the Yoruba language by means of pitch differences. In this aspect, *omele bàtá* sounds can be considered acoustically closer to spoken Yoruba than the speech surrogacy of the *ìyáàlù bàtá*, which represents tone by means of contrasting sound spectra. Moreover, the possibility of striking two membranes simultaneously with the 2 *bílálà* allows the drummer to partially apply the i-and-u-rule described above for the *ìyáàlù bàtá*. However, these rules are applied less consistently on the *omele bàtá* than on the *ìyáàlù bàtá*.

Different from the *ìyáàlù bàtá*, the *omele bàtá* is also used to render the /r/ consonant by means of rapid flam strokes (rapid sequence of two strokes on the same membrane, one hand after the other). Although this trait is observed in our dataset only on the high-pitched membrane (mi), our informants support that it can occur in the mid and low membranes as well.

General regularities of the *omele bàtá* speech surrogacy:

1. The three drums (*omele ako, kúdi,* and *àdàmò*) are assigned to the three Yoruba speech tones (*ohùn ìsàlè, àárín*, and *òkè*), representing the L, M, and H speech tones (do, re and mi) (1a, 2a and 3a on **Figure 9**).
2. When two of the drums are played simultaneously (double lines), the combinations H-L and H-M render the speech tones do and re, respectively, in non-high vowels.
3. For the speech tones do and re, when rendering syllables containing the soft vowels /i/ and /u/, the representations in rule 1 are preferred.
4. Syllables on high speech tone containing intense vowels are mostly rendered as a double stroke on the high-pitched membrane (3b)
5. The double mi stroke (3b) can also be played sequentially in a quick rhythm, rendering the /r/ consonant in a high-pitched syllable.

When considering the rules above, it is noticeable that the *omele bàtá*, as a more recently created instrument, displays a speech surrogacy system easily relatable to the playing patterns of the *ìyáàlù bàtá*, which relies largely on the parallel playing of two membranes to render vowel height. On the omele, the high-pitched membrane, just like the *ṣáṣá* on the *ìyáàlù*, serves a double function, acting both as a representation of the high speech tone, and as a vowel quality marker on non-high vowels. However, three fundamental differences between the two drums should be highlighted here. Firstly, the *omele bàtá*, with three independent membranes, is able to render speech tone through actual pitch distinctions, i.e., through distinct fundamental frequencies. On the *ìyáàlù bàtá*, as mentioned before, this differentiation is largely based on timbre—i.e., sound spectra. Secondly, the omele consists of three independently constructed drums, which are not attached to each other but through the carrying strap. Thus, they are also acoustically largely independent from each other, causing their sounds, when played simultaneously, not to merge with each other in the same way the *ojú òjò* and *ṣáṣá* sounds of the *ìyáàlù bàtá* do. Thirdly, on the *omele bàtá*, the three membranes have similar sounds, varying virtually only in the pitch they are tuned to. Therefore, when played simultaneously, they tend to sound like an interval (i.e., the simultaneous playing of two pitches or "notes" in the conventional sense of Western music theory), instead as two complementary timbres, as it is the case on the *ìyáàlù bàtá*.

Villepastour (2010, p. 89) argues that as a speech surrogate, the *omele bàtá* is advantageous over the *ìyáàlù bàtá*. Her reasoning is based mostly on the fact that the *omele bàtá* renders speech tones by means of actual pitch distinctions, which leads, arguably, to more clearly distinguishable representations of speech tone. On this instrument, additionally, it is possible to render the /r/ consonant. However, the differences listed above indicate that the *omele bàtá* is not fully capable of imitating vowel spectra, as seen on the *ìyáàlù bàtá*. The high pitched *ṣáṣá* of the *ìyáàlù bàtá* appears to complement the high frequency spectrum on intense vowels, which is not fully applicable for the *omele bàtá*. On the *omele bàtá*, unlike the *ìyáàlù bàtá*, single sounds of individual membranes do not merge, but are rather perceived as intervals, which is distant from the original aurally rooted system of vowel imitation of the *ìyáàlù bàtá*. This means that the vowel rendering technique of the omele bàtá could be understood as a formal device borrowed from the *ìyáàlù bàtá* technique, employing similar sound connections, yet without its underlying aural groundwork. It may be concluded that the *omele bàtá* is used to (partially) implement phonological distinctions originating from an aural effect that is not present on this instrument.

EXAMPLE OF FULL SENTENCE ANALYSIS

Our main aim in this study was to provide adequate data for comparing the speech surrogacy of the most popular Yoruba "talking" drums. Our dataset is made up of renditions of 12 different sentences on the *gángan, ìyáàlù dùndún, ìyáàlù bàtá* and *omele bàtá*. In this section we illustrate these renditions for the sentence "*Bàbá àgbàlagbà kaán jókòó lórí àga pupa*" ("Father – old – a certain – sits on – chair – red" or "there is an old man sitting on a red chair").

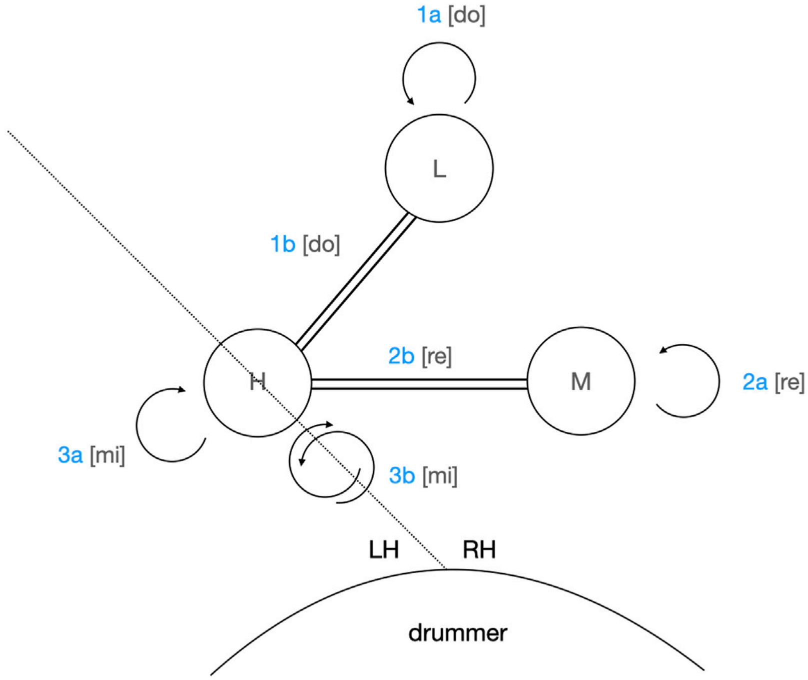


FIGURE 9 | Graphic representation of the *omele bàtá* speech surrogacy system.

In this sentence, both recordings of *dùndún* drums display a similar overall melodic pattern with a clear correspondence between speech syllables and drum strokes. However, some differences between the patterns played demonstrate how *dùndún* players can choose between the representation of distinct phonological features in rendering one drummed phrase and how the distinct sizes of the *ìyáàlù dùndún* and the *gángan* affect their acoustic possibilities when mimicking Yoruba speech (**Table 2**).

The words "*Bàbá àgbàlagbà kaán*" are rendered in the same way in both *dùndún* samples:

- "*Bàbá*" is represented with an upwards gliding pitch in both versions [do][∧ mi].
- Both drummers represented the vowel sequence "*Bàbá àgbà*" by slowly dropping the tuning tension without a new onset before the next syllable, thus creating a ligature between the high pitch [∧ mi] and the following pitch (∨ do).
- Both drummers represented the word "*kaán*" ("a certain") with an upwards gliding pitch. The *gángan* version includes a prolonged pause after this syllable.
- Both versions differ on the representation of "*jókòó*". The *ìyáàlù dùndún* version features a slow upwards glide [do](∧ mi), whereas the gángan version seems not to represent the last rising contour of the word but with an extended pause after the low [do].
- The *ìyáàlù dùndún* version represents the /r/ in "*lórí àga*" by performing the pitches [mi][do] on a fast bouncing (double) stroke. This representation is left out on the gángan version.
- On the *ìyáàlù dùndún* version, the /u/ vowel in "pupa" is rendered with a softer sound than the surrounding words. On the *gángan* version, no such intensity contrast is found.

Both *bàtá* drums, as played by the same drummer, largely rely on the basic rules of syllable-tone-pitch correspondence and the i-u-rule as described above. The few exceptions are interesting and reveal some phonologically relevant features of the *bàtá* speech surrogacy. Notably, two very distinct applications of rhythmically shorter strokes, or "flams", can be observed in the *ìyáàlù bàtá* versions (**Table 3**).

- On the *ìyáàlù bàtá*, all syllables were performed according to the i-u-rule described, with the exception of the strokes involved in glides (see below). On the *omele bàtá*, the rule is applied to only one third of the strokes. The basic correspondence between syllable speech tone and played stroke is observed throughout both recordings.
- The *ìyáàlù bàtá* version renders the vowel sequence "*Bàbá àgbà*" with a flam from the ṣáṣá [mi] and *ojú òjò* slap [mi] to the *ojú òjò* [do] (open tone). Seemingly, the short rhythmic figure is used here to represent the vowel sequence.
- Both versions feature a longer pause after "*kaán*". This could be a way of representing the rising intonation of this word, as seen on both the dùndún and spoken versions of the sentence, as well as the prolonged duration of this word due to the nasal /n/.
- Both versions use quick rhythms in the representation of the /r/ in "*lórí àga*". In the *ìyáàlù* version, due to the presence of the /i/ vowel, this is achieved through a quick movement from the slap to the open tone (*ojú òjò*), a technically demanding stroke sequence at this speed.
- Both versions render the syllable "*jókòó*" on a single stroke, in a similar fashion to the *gángan* version, whereby the last high pitch contour is not displayed with a sound (e.g., a flam), but rather with the insertion of a pause after the low pitch.

TABLE 2 | Transcriptions of the sentence "*Bàbá àgbàlagbà kaán jókòó lórí àga pupa*" ("there is an old man sitting on a red chair") on the *dùndún* drums. (A) *Ìyáàlú dùndún*. (B) *Gángan*.

(A) *Ìyáàlú dùndún*

do	Λ mi	(V do)	do	re	do	re Λ	mi	do	(Λ mi)	mi		mi	do	re	re	re
Bà	bá	à	gbà	la	gbà	kaán	jó	kò	ó	ló		rí	à	ga	pu	pa
+	+	+	+	+	+	+	+	+	+	+		+	+	+	-	+

(B) *Gángan*

do	Λ mi	(V do)	do	re	do	re Λ		mi	do		Λ mi	mi	do	re	re	re
Bà	bá	à	gbà	la	gbà	kaán		jó	kò		ló	rí	à	ga	pu	pa

TABLE 3 | Transcriptions of the sentence "*Bàbá àgbàlagbà kaán jókòó lórí àga pupa*" ("there is an old man sitting on a red chair") on the *bàtá* drums. (A) *Omele bàtá*. (B) *Ìyáàlù bàtá*.

(A) *Omele bàtá*

Left Hand	mi	mi	mi	mi	mi	mi	mi		mi			mi		mi				mi
Right Hand	do	mi	do	do	re	do	re		mi	do		mi		mi	do	re	re	re
	bà	bá	à	gbà	lag	bà	kaán		jó	kòó		ló		rí	à	ga	pu	pa

(B) *Ìyáàlù bàtá*

Sásá	mi	mi	mi	mi	mi	mi		mi		mi			mi			mi	mi		mi
Ojù òjò	do	mi	do	do	re	do		re			do		mi		mi	do	re	re	re
	bà	bá	àg	bà	lag	bà		kaán		jó	kòó		ló		rí	à	ga	pu	pa

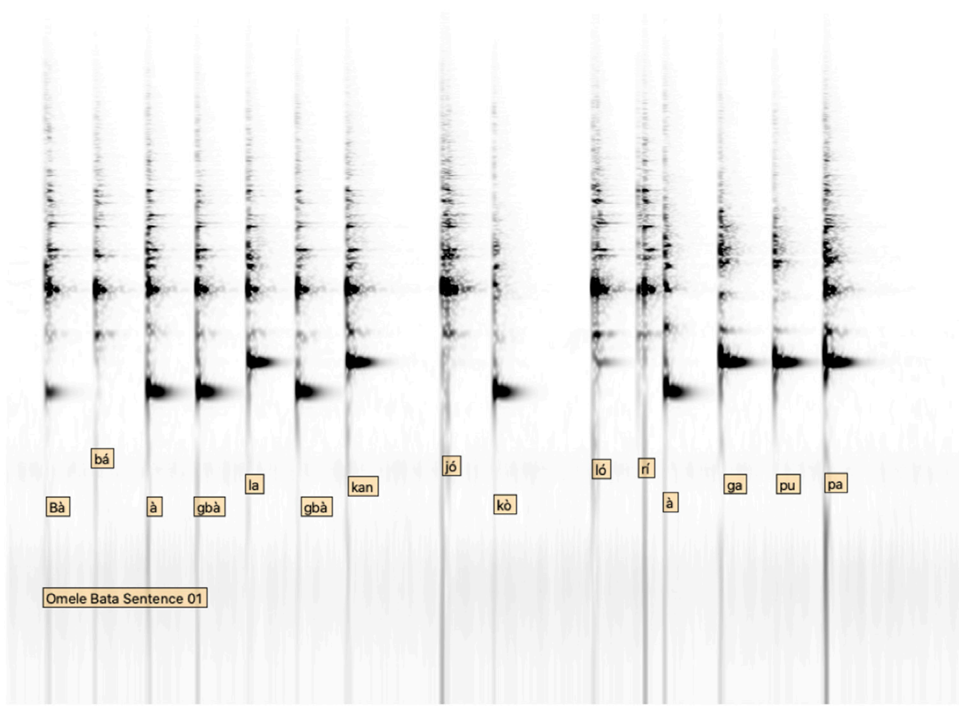


FIGURE 10 | *Omele bàtá* spectrogram for the sentence "*Bàbá àgbàlagbà kaán jókòó lórí àga pupa*" ("there is an old man sitting on a red chair").

- The spectrogram of the *omele bàtá* version (**Figure 10**) shows a great rhythmic regularity (average IOI—Inter Onset Intervals), with clearly longer IOI after "*kaán*" (re with rising pitch) and "*kòó*" (vowel sequence with upwards glide on the same vowel), and a shorter time interval rendering the /r/ consonant. Apart from these exceptions, all IOI are virtually equal.
- Similarly, the rhythm used in the *ìyáàlù bàtá* (see **Figure 6**) version is constant, being interrupted by the flams and prolonged pauses mentioned above. One additional pause can be seen after "*àgbàlagbà*" for a reason that is yet unclear.

The analysis of this sentence, as recorded by three drummers ad hoc, without knowing the sentences in advance or having substantial time for preparation, shows that the regularities described above largely apply on their drum performances. On the *dùndún* drums, both drummers respected the basic tone contours of Yoruba. As expected, both versions differ in their use of glides and of the double stroke on the imitation of the /r/ consonant. Furthermore, the *gángan* version does not differentiate between vowel intensities. Similarly, the *bàtá* versions largely obey to the regularities demonstrated above, with the *omele bàtá* versions proving to apply the vowel related rules less regularly.

FINAL CONSIDERATIONS

The *dùndún* and *bàtá* drums that were discussed in this paper differ in their organological features and sound production. Accordingly, in spite of being both used as speech surrogates, they differ in their means of doing so. The *dùndún* drums have a flexible tuning system and can be used to quite accurately depict the tonal inflections of spoken Yoruba. The *ìyáàlù bàtá* has fixed membranes and therefore cannot perform gliding tones in the same fashion. Nonetheless, the richness of the *ìyáàlù bàtá* timbre enables the differentiation between vowel qualities through the distinct sounds emerging from diverse playing techniques and their combinations. By contrast, the *dùndún* employs a single playing technique, with the *opá* producing a uniform timbre over the whole tonal range of the instrument. Accordingly, these drums cannot be used to distinguish between vowels through timbre, and they mark the difference in vowel intensity by means of amplitude. The *omele bàtá* is only able to produce three different sounds and their combinations, but its speech surrogacy practices represent both tonal aspects of spoken Yoruba and some vowel distinctions. Additionally, each of the three types of drums is able to imitate the sounds of language relating to specific constraints of time, for instance using flams and re-bouncing strokes to resemble the sound of the rolling tongue when of pronouncing an /r/, or prolonging sounds or pauses to mark, for instance, the length of a nasal coda consonant. In our analysis, we have focused on the organological features and playing techniques of the studied drums, on the way they favor certain phonological elements of the Yoruba language over others. Thus, we propose that drums are not placed along a continuum between "encoding" and "mimicking", as supported by Villepastour (2010, p. 90) but rather between their tendencies to address and modulate each of the intrinsic characteristics of

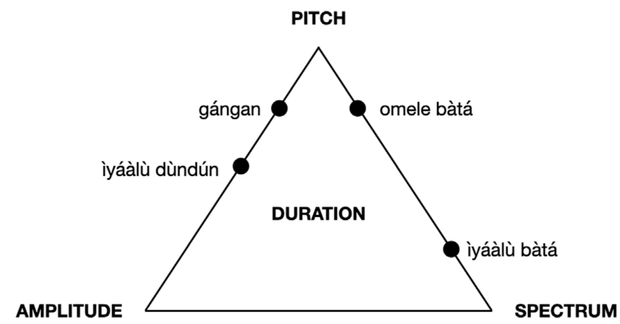


FIGURE 11 | Placement of the studied drums according to their addressing of pitch, amplitude, duration and spectrum.

language sounds: intensity, pitch, duration and spectral composition[4] when imitating the sounds of spoken Yoruba language (**Figure 11**).

The *ìyáàlù dùndún* addresses both the tonal inflections of Yoruba and the intensity contrasts between vowels by modulating both pitch and intensity (amplitude) when mimicking them. The *gángan*, due to its smaller membrane, is less capable of modulating amplitude, but still derives its speech surrogacy from the direct imitation of speech tones and their inflections. The *ìyáàlù bàtá*, although it can hardly mimic the three speech tones, is capable of imitating vowel spectra thanks to a nuanced playing technique and the advantages on a conic drum corpus acting as a closed air column between the two membranes. The *omele bàtá*, finally, with three membranes tuned to different pitches, is used to imitate the three speech tones of Yoruba language directly. The *omele bàtá* is also able to mimic to some extent the vowel characterizing system of the *ìyáàlù bàtá*, however without relying on a nuanced modulation of sound spectra. Furthermore, it is needless to say that duration—be it in form of musical rhythm, or when imitating the /r/ consonant – is one element of speech or musical sound necessarily present and modulated in any form of speech or music.

It can be observed that the notions of pitch height, tone spectrum and vowel quality are deeply intertwined in the speech surrogacy systems of the Yoruba "talking" drums. They may be regarded as reflecting an oral culture with high aural sensibility but without the need for a theoretical dissociation of these single acoustic concepts. Furthermore, the great acoustic sensibility in oral cultures (Ong, 1982) offers the possibility to understand drum languages not as surrogates that imitate established (i.e., consciously formalized) language structures, but rather as technologies capable themselves of participating in making these structures tangible. Inasmuch as phonetics and phonology are Western disciplines devoted to the systematization of knowledge around language sounds, drum language systems such as the Yoruba *dùndún* and *bàtá* systems are themselves native, concrete strategies of systematization of language sounds

[4]*Lexicon of Linguistics*, art. "Intrinsic Characteristics" <https://lexicon.hum.uu.nl/?lemma=Intrinsic+characteristics&lemmacode=1223&lemma=Intrinsic+characteristics&lemmacode=1223> (downloaded on 08.01.2021).

through their transformation and reinterpretation within the acoustic systems made possible by the organological features of musical instruments and their playing techniques. We may therefore view the transformation of sound from speech to drum as a platform of developing a native theory of language, in which the sounds of language are reconsidered inside the framework of an artificial instrument, which helps to represent native introspective aspects of spoken language.

In his early account of the Yoruba drum languages, Ulli Beier seems to find necessary to state that "the talking drum does *not* use a kind of Morse system, as imagined by most non-Africans" (Beier, 1954, p. 30), showing that the perhaps most natural assumption made by a (Western) outsider, when first confronted with a seemingly unintelligible practice of speech surrogacy, is that it must contain a specific secret "code" to be revealed. The Yoruba drum languages presented here, however, as we hope to have shown, do not rely on any hermetic coding process, but on the outstandingly cultivated aural sensibility of both drummers and listeners, who are able to navigate and creatively employ diverse relations of sound similarity between the language they speak and the drums they play.

DATA AVAILABILITY STATEMENT

The transcripts of the audio recordings supporting the conclusion of this article can be made available by the authors upon request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Faculty Ethics Assessment Committee, Faculty of Humanities, Utrecht University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

MG: First author, responsible for research design, field data collection, data analysis and redaction. OO: Second author: Responsible for field research organization, collaboration in data collection, help with data analysis and with article redaction.

ACKNOWLEDGMENTS

We would like to thank first and foremost our informants, Ayanlere Alajede, Kangan Bamidele and our *bàtá* master drummer, who opted to remain anonymous, for their kindness in sharing their art and knowledge with us, as well as Prof. Yoad Winter, for his insatiable curiosity and competent support. Finally, we thank the Dutch Research Council (NWO), for the financial support (project 360-89-060, "When language has a beat"), and Utrecht University, for being our home during this research.

REFERENCES

Adegbite, A. (1988). The Drum and its Role in Yoruba Religion. *J. Religion Afr.* 18 (1), 15–26. doi:10.1163/157006688x00207

Ajiboye, O. (2011). Disambiguating Yoruba Tones: At the Inferface between Syntax, Morphology, Phonology and Phonetics. *Lingua* 121, 1231–1648. doi:10.1016/j.lingua.2011.05.008

Akinbo, S. (2019). Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis. *Linguistique et Langues Africaines* 5, 11–23.

Arewa, O., and Niyi, A. (1980). Redundancy Principles of Statistical Communications as Applied to Yoruba Talking-Drum. *Anthropos* 75, 185–202.

Beier, U. (1954). The Talking Drums of the Yoruba. *Afr. Music Soc. J.* 1, 29–31. doi:10.21504/amj.v1i1.223

Bello, A. M. (2020). "Stylistics of Identity and Ideology in Alter-Native Yoruba Music Textuality. PhD Thesis. Nigeria: Department of Linguistics, African and Asian Studies, University of Lagos.

Euba, Akin. (2011). Review: Villepastour, Amanda. *Ethnomusicology* 55, 516–520. doi:10.5406/ethnomusicology.55.3.0516

Euba, Akin. (1990). *Yoruba Drumming: The Dùndún Tradition*. Bayreuth: Bayreuth University.

Hughes, D. W. (2000). No Nonsense: The Logic and Power of Acoustic-iconic Mnemonic Systems. *Br. J. Ethnomusicology* 9, 93–120. doi:10.1080/09681220008567302

Isola, A. (1982). Ena: Code Talking in Yoruba. *J. West Afr. Languages* 12 (1), 43–51.

King, A. (1961). *Yoruba Sacred Music from Ekiti*. Ibadan: Ibadan University Press.

Klein, D. (2000). *Yoruba Bàtá: Politics of Tradition in Erin-Osun and Overseas*. Santa Cruz: Diss. University of California.

Kubik, G. (2010). *Theory of African Music*, 2. Chicago and London: University of Chicago Press.

Kubik, G. (2008). "Zur Mathematik und Geschichte der afrikanischen Time-Line Formeln," in *Systematic and Comparative Musicology: Concepts, Methods, Findings*. Editor A. Schneider (Frankfurt am Main: Lang), 359–398.

Lexicon of Linguistics (2021). Art. Intrinsic Characteristics. Available at: https://lexicon.hum.uu.nl/?lemma=Intrinsic+characteristics&lemmacode=1223&lemma=IntrIntri+characteristics&lemmacode=1223 (Accessed 01 08, 2021).

Olaniyan, O. (2008). The Manifestation of the Poetic Aptitude of the Yoruba Dundun Master Drummer within the Framework of Dundun Music. *Awka J. Res. Music Arts* 5, 95–110.

Olatunji, M. (2017). "A Historiographical Study of Gudugudu Drum – A Member of Dundun Ensemble in Yorubaland," in *Musical Horizons in Africa (Essays and Perspectives): A Festchrift in Honour of Daniel Chikpezie Christian Agu* (Ile-ifé: Association of Nigerian Musicologists), 247–253.

Olatunji, M. (2012). Yorùbá Proverbs and Musicality. *Legon J. Humanities* 23, 125–140.

Oludare, O. E. (2020). "The Musical and Cultural Leadership Roles of Yoruba Master Instruments," in *Governance and Leadership Institutions in Nigeria*. Editors E. Aniche and T. Falola (London: Routledge), 65–74. doi:10.4324/9781003111405-7

Oludare, O. E. (2021). "Yoruba Traditional Instrumental Ensemble and Indigenous Knowledge Systems," in *Pathways to Alternative Epistemologies in Africa*. Editors A. Afolayan, O. Yacob-Haliso, and S. O. Oloruntoba (Cham: Palgrave Macmillan), 205–219. doi:10.1007/978-3-030-60652-7_12

Oludare, O. E. (2018). "Yoruba Traditional Musical Instruments and Gender Stereotyping: Forms, Functions and Factors," in *Fela After 1997: Challenges, Justification and Lessons for Africa's Development* (Ibadan: Global Image Publishers), 181–202.

Oluga, S., and Babalola, H. (2012). Drummunication: The Trado-Indigenous Art of Communicating with Talking Drums in Yorubaland. *Glob. J. Hum. Soc. Sci. Arts, Humanities* 12 (11), 38–48.

Omojola, B. (2012). *Yoruba Music in the Twentieth Century: Identity, Agency, and Performance*. Rochester, NY: University of Rochester Press.

Ong, W. (1982). *Orality and Literacy: The Technologizing of the Word*. London and New York: Routledge.

Orie, O. O., and Pulleyblank, D. (2002). Yoruba Vowel Elision: Minimality Effects. *Nat. Lang. Linguistic Theor.* 20 (1), 101–156. doi:10.1023/a:1014266228375

Osundina, O. (2015). *Dundun: The Talking Drum of the Yoruba People of Southwest Nigeria*. London: Austin Macauley Publishers.

Oyelami, M. (1991). *Yorùbá Bàtá Music: A New Notation with Basic Exercises and Ensemble Pieces*. Bayreuth: Iwalewa Haus.

Oyelami, M. (1989). *Yoruba Dundun Music: A New Notation with Basic Exercises and Five Yoruba Drum Repertoires*. Bayreuth: Iwalewa Haus.

Pulleyblank, D. (2004). A Note on Tonal Markedness in Yoruba. *Phonology* 21, 409–425. doi:10.1017/s0952675704000326

Ruskin, J. (2013). *Diss the Yorùbá Dùndún in Local and Transnational Perspective: A Cosmopolitan Tradition in the Making*. Los Angeles: Diss. University of California.

Sachs, C., Hornbostel, V., and Moritz, E. (1986). "Systematik der Musikinstrumente. Ein Versuch," in *Tonart und Ethos: Aufsätze zur Musikethnologie und Musikpsychologie* (Leipzig: Reclam).

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code*. Farnham: Ashgate.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Communication

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fcomm.2021.652740

Check for updates

A Preliminary Account of the Northern Toussian Balafon Surrogate Language

*Anthony Struthers-Young**

Department of Linguistics, University of California San Diego, San Diego, CA, United States

Northern Toussian is a language spoken in southwest Burkina Faso. Like many of the ethnicities in the region, members of the musical caste, the griots, have developed a musical surrogate language which is played on the balafon. This article is a preliminary documentation of the Northern Toussian balafon surrogate language, describing its cultural usage, analyzing how the tones of the spoken language are encoded in the surrogate language and comparing the Northern Toussian surrogate language with the neighboring Sambla surrogate language. It was found that the Northern Toussian surrogate language encodes significantly more post-lexical features of the spoken language than the Sambla surrogate language, such as downdrift.

Keywords: Northern Toussian, surrogate language, balafon, tone, Burkina Faso

OPEN ACCESS

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Christopher Green,
Syracuse University, United States
Lee Bickmore,
University at Albany, United States
Arto Anttila,
Stanford University, United States

****Correspondence:***
Anthony Struthers-Young
astruthe@ucsd.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *13 January 2021*
Accepted: *01 October 2021*
Published: *09 March 2022*

Citation:
Struthers-Young A (2022) A Preliminary Account of the Northern Toussian Balafon Surrogate Language. Front. Commun. 6:652740. doi: 10.3389/fcomm.2021.652740

INTRODUCTION

Musical surrogate languages are encodings of speech using musical instruments. They have been found on diverse instruments such as drums, jaw harps, flutes, mouth organs, and a type of xylophone played in West Africa called the balafon (Proschan, 1993; Villepastour, 2010; McPherson, 2019a). However, few musical surrogate languages have been studied comprehensively by linguists and have instead been analyzed only through an ethnomusicological lens. This is unfortunate, since a holistic understanding of a surrogate language is predicated on understanding the structure of the language being represented in addition to the musical culture. Moreover, the study of surrogate languages has much to contribute to the field of linguistics. Speech surrogates, another term for surrogate languages, can represent an utterance word-for-word, but will only encode certain structures of speech, such as tone and prosody. Because of this, they can be useful tools for phonological analysis of tonal languages—often, they pare away much of the phonetic variation found in the tones of speech, allowing researchers to more quickly and efficiently analyze the tonology of the language (McPherson, 2019b; McPherson, 2019c). Additionally, surrogate languages give insight into the speaker's perception of the language, showing what linguistic structures they encode. Learning how speech surrogates function and when they are used can be a useful tool when developing engaging resources for preservation and revitalization programs.

This paper documents the balafon surrogate language of the Northern Toussian language spoken in southwest Burkina Faso. It is based on data elicited in Bobo-Dioulasso with the musician and griot Emile Diabaté in the summer of 2019. The research was exploratory, aimed at determining how the surrogate language functions; this paper is a preliminary description of the system. It demonstrates what linguistic structures are represented in the surrogate language and the strategies for encoding said structures. The primary element of speech which is encoded with the surrogate language is tone, where tones are mapped to specific bars on the balafon. Other elements such as syllable structure and phrase boundaries are represented: syllable structure by two quick strikes called flams, and phrase boundaries by striking octaves or extra low notes. Similar work has been done on the Sambla balafon surrogate language (McPherson, 2019a). The Sambla and Toussian are neighboring ethnicities who speak different Niger-Congo languages—according to oral history, the Sambla learned how to play the balafon from the Toussian, likely in the late 19th century (Strand, 2009). A comparison of the structure of the Seenku and Toussian balafon surrogate languages is made to examine the relationship of these speech surrogates.

The following topics are discussed: the Toussian people and languages (*The Toussian people and languages* Section), the construction and cultural significance of the Toussian balafon (*Griots and the balafon* Section), relevant elements of Northern Toussian grammar (*Relevant Northern Toussian grammar* Section), the Northern Toussian balafon surrogate language (*Surrogate language* Section) and finally the Sambla balafon tradition and its historical and cultural ties to the Toussian (*Comparison to the Sambla balafon surrogate language* Section).

THE TOUSSIAN PEOPLE AND LANGUAGES

The Toussian[1] people reside in the southwest of Burkina Faso, near Bobo-Dioulasso. Bordered by the Banfora escarpment, a line of cliffs which spans from Banfora to Bobo-Dioulasso, the Toussian live on a plateau which becomes increasingly hilly as one travels west. They are neighbored in the north by the Dzuungo, in the northeast by the Sambla and the Tiefo, in the west by the Siamou, in the south by the Turka and Karaboro, and in the east by the Dioula (Hammarström et al., 2020).[2] There are no urban centers in Toussian country; the people are spread among villages of varying size and most Toussian are subsistence farmers. They grow diverse crops, including their staples corn, millet, and sorghum, fruits such as mangoes and bananas and cash crops like cotton. They are a small minority in Burkina Faso; the most recent survey in 1995 listed the population of speakers of Toussian at around 40,000, though the number of speakers has likely grown substantially since then due to population growth (Eberhard et al., 2020).

The Toussian people speak languages which belong to the Toussian family. Two Toussian languages have been reported in published literature: Northern Toussian and Southern Toussian, which are separated by a sparsely populated area of hilly forests. Within these languages, there is considerable dialectal variation. In addition to Northern Toussian and Southern Toussian, I have suspicions that a variety of Toussian spoken only in the villages Moami and Tien is divergent enough to be classified as a separate language, as the variety they speak is not mutually intelligible with speakers of Northern or Southern Toussian. Based on preliminary research I conducted with three speakers, this variety shares more similarities in vocabulary and phonotactics with Southern Toussian than Northern Toussian—more research is necessary to determine if it is a separate language or a dialect of Southern Toussian. **Figure 1** is a map of the Toussian languages, including the variety spoken at Moami and Tien, as well as neighboring languages.

The vitality of Northern Toussian and Southern Toussian is still strong in most villages—children are exposed to the language from infancy. Most young Toussian are bi- or trilingual, as they speak Dioula from early childhood and learn French in school. However, the elderly (~85 or older) might be monolingual in Toussian and middle-aged speakers (~50 or older) might have limited or nonexistent knowledge of French. Linguistic vitality is not so robust in every village—In Moami and Tien, which are isolated in Dioula- and Seenku-speaking regions, there are indications of loss of vitality. According to individuals from Moami, their language is often not passed on to children. When I elicited preliminary data there, I noted that some middle-aged speakers were found to exhibit linguistic attrition. There have been no surveys to determine how many speakers of the language remain. Moami and Tien appear to be outliers; in most Toussian villages, Toussian is widely and often used, though more sociolinguistic work is necessary to truly evaluate the vitality of Toussian throughout the region.

The genetic affiliation of the Toussian languages has long been debated and there is no consensus. In the 1960s, they were grouped as members the Gur language family and were noted for their similarities with Senoufo languages, which were considered Gur at the time (Prost, 1964). Later, Naden (1989) reanalyzed the Senoufo languages and removed them from Gur, asserting that they belong to their own family. However, he still analyzed the Toussian languages as Gur languages. Recently (Güldemann, 2018) argued that the evidence grouping Toussian into the Gur family is scant and the Toussian languages should be considered an unaffiliated Niger Congo language family until more research has been done.

There is currently no orthography for Northern Toussian, though one exists for Southern Toussian, developed by the SIL missionary Hannes Wiesmann while he studied Southern Toussian (Wiesmann, 2000; Wiesmann et al., 2004). Though Wiesmann primarily studied the dialect spoken at Nianha, the orthography was designed to accommodate all Southern Toussian dialects. Nevertheless, it appears that Southern Toussian literacy is still quite low.

GRIOTS AND THE BALAFON

Construction and Use

Griots are members of a musical caste who fill important cultural roles in many West African societies. In southwest Burkina Faso, most men who are griots will become musicians, mastering the

[1]The word "Toussian" appears to be autonymous, as speakers of the Northern Toussian dialect of Kourinion use the term Toussian (tusiyã). However, there are other words for the Toussian ethnicity, including the Southern Toussian Wì n (wǹ̩) and the Northern Toussian dialect of Djigouera Trū (trū). More details on the languages and dialects of the Toussian people are found below.

[2]Map made with ArcGIS Online by ESRI.

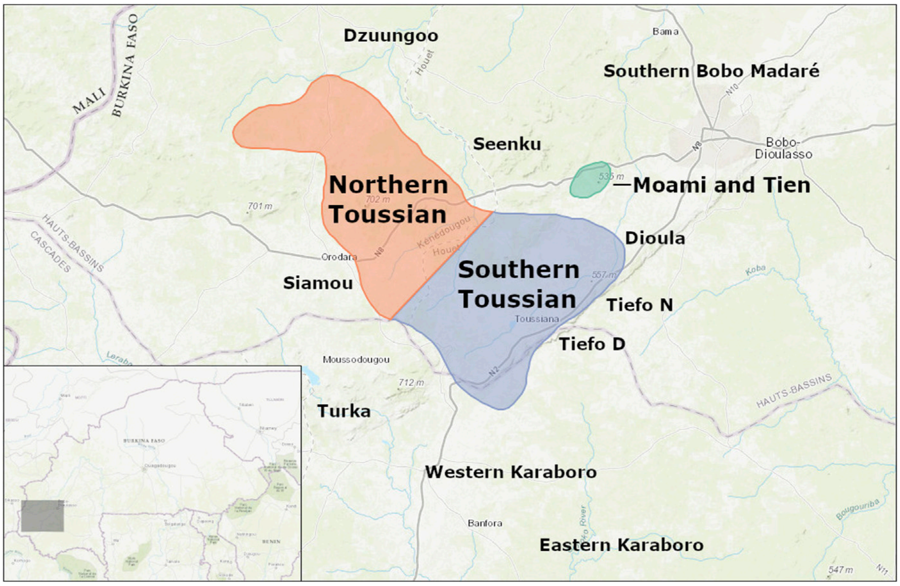


FIGURE 1 | Map of languages of southwest Burkina Faso.

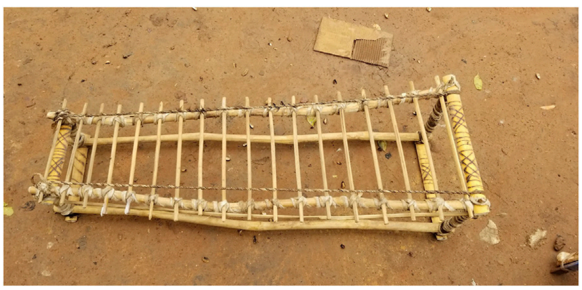


FIGURE 2 | Balafon frame.

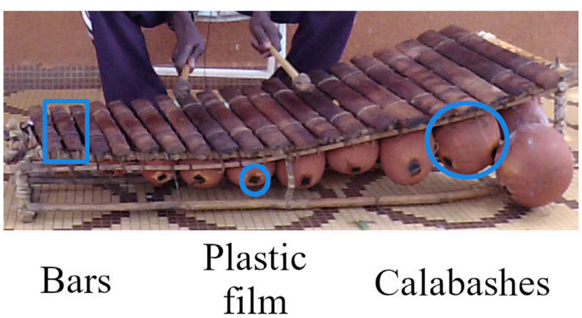


FIGURE 3 | Parts of the balafon.

instrument their fathers played and learning traditional songs. Women who belong to griot families, often called griotes, might sing or play the rattle, but will not traditionally play instruments like the balafon or drums. There are two distinct groups of griots among the Toussian: balafon griots called *ɲɔ́pī* and drum griots called *kə̄tɔ̰̄*[3]. Membership in each of these groups is hereditary; whereas a person from a non-griot family can become a balafonist or drummer and learn how to play the balafon or drums, they cannot become a *ɲɔ́pī* or *kə̄tɔ̰̄* and fulfill the cultural roles that *ɲɔ́pī* and *kə̄tɔ̰̄* do. Likewise, the two types of griots are independent—the child of a *ɲɔ́pī* can only be a *ɲɔ́pī*, not a *kə̄tɔ̰̄*. Only balafon griots of certain families are permitted to construct balafons. At Djigouera and the neighboring villages, a single family from Djigouera has this right, meaning there are only a handful of people in and around Djigouera who can build and supply balafons to all the other griots.[4] In addition to Toussian balafons, this family also makes Sambla, Dioula, Siamou, Senoufo and Manding[5] balafons, each of which are constructed similarly to the Toussian balafon, but use a different scale.

The Toussian balafon consists of a row of 18–22 bars bound to a flat wooden frame which houses a series of resonating chambers. The bars are cut from a hardwood made from the wood of a *ɲīrɛ̰̄* [ɲḭ̄nɛ̰̄] (*Pterocarpus erinaceus*) tree and then are dried over a fire and under the sun for a period of time. Once they have dried, they are tuned by shaving off portions of the underside with an adze. The frame of a balafon is constructed with struts of wood lashed together, and can be seen in **Figure 2**. The lashings are traditionally made of strips of goat hide, but now the griots will often use leather, rope, or twine, depending on what is available. When the bars have been dried and tuned, they will be lashed individually to the frame.

[3]Nasalization of vowels will be indicated by a tilde under the vowel to avoid typographic issues representing both tone and nasalization above the letter.

[4]Not all villages have griots—of villages close to Djigouera, there are only griots in Serekeni, Kleni and Kouini.

[5]The term used by the griot recounting this. He specified this as a type of balafon which came from Guinea but did not state the language or ethnicity of people who use this balafon.

There is one calabash per bar and each calabash is tuned to the resonant frequency of the bar, greatly amplifying its sound. The tuning is accomplished by removing material from the top of the calabash, lowering its pitch. Each calabash has a hole drilled in its side which is covered by a thin film—traditionally from the egg sack of a particular species of spider, but now usually from a thin piece of plastic. This distorts the note slightly, giving the balafon its characteristic buzzing sound. When the calabashes have been finished, the griot will string them up inside the frame so that the mouth of the calabash sits directly below the bar it is is tuned to. **Figure 3** shows a completed balafon, highlighting different components of it.

The balafon is pentatonic and typically has a span of around four octaves. For the Toussian balafon, this range is approximately 90–1120 Hz. There does not appear to be a designated root or note which acts as the musical center; songs can be played in different modalities by choosing any of the five bars as a root. The bars bear the following names:

(1) Bar names and intervals

bar	cents[6]	name
1 (lowest)[7]		tā-ɲɛ́pwɛ̄ 'Ta (a name) bar'
	322	
2		dú-fɔ̄-ɲɛ́pwɛ̄ 'Major initiation bar'
	214	
3		yā̰-tɛ̰̀-ɲɛ́pwɛ̄ 'rattle player bar'
	242	
4		ɲɛ́pwɛ̄-kà 'unripe bar'
	309	
5		kɔ̄tɔ̰̄-ɲɛ́pwɛ̄ 'drum griot bar'
	81	
1 (octave)		tā-ɲɛ́pwɛ̄ 'Ta (a name) bar'

The word *ɲɛ́pwɛ̄* (pl. *ɲɛ́plɔ̄*), present in all these names, means 'bar.' The word *tā* refers to a wealthy, important man from long ago. The compound *dú-fɔ̄* refers to a group of people who play an important role in the *dú* initiation ritual, which serves to teach Toussian children important lessons about morals, nature and Toussian culture. It is often held to be the most significant cultural tradition for the Toussian and is discussed below (*Culture and use* Section). The word *yā̰-tɛ̰̀* is a gourd rattle that accompanies the griots as they play. The name of the fourth bar, *ɲɛ́pwɛ̄-kà* means 'the unripe bar.' It plays a note that is unique to the Toussian balafon, not shared by any other ethnicities in the region, and as such acts as an identifier of Toussian balafons. Finally, *kɔ̄tɔ̰̄-ɲɛ́pwɛ̄* refers to the drum griots. They are the only people permitted to repair a damaged *kɔ̄tɔ̰̄-ɲɛ̰́pwɛ̄*; reserving a bar of the balafon for the *kɔ̄tɔ̰̄* is a sign of respect.

Culture and Use

Balafons have historically been important fixtures in Toussian culture and religion; this remains the case for many, if not most, Toussian individuals. However, as the population of Muslims and Christians has increased in recent years, this has affected the cultural role of the balafon, discussed below. At traditional marriages, funerals and festivals, several balafonists will likely be accompanied by a full band comprised of drummers, rattlers, and singers. In these concert settings, the balafonists will often play in tandem on the same balafon, with up to three people playing the instrument from both sides. Additionally, for some events like planting season and the harvest, the griots will join the farmers out in the field with their balafons strapped to them, playing while walking alongside the cultivators. Historically, the griots did no farming; there was a social contract where the rest of the populace would supply them with food in exchange for their music, but this practice is in decline and now griots must often supplement their finances by farming (Trost, 1999). There are specific songs which accompany each of these events. Some of these songs are structured in a call-and-response way, where a portion of the song is sung and then repeated using the surrogate language.

The griot plays an active role in many Toussian customs. An example of this would be during the major initiation *dú~dó*, which the griots help organize and lead certain sacrifices in. The major initiation is a coming-of-age tradition which boys take part in once in their life, starting at the age of around 15 or older. These boys are brought out into the bush for several months and taught skills and lessons important for Toussian society and culture. These lessons are varied, from teaching practical skills such as identifying plants and hunting, to building cultural knowledge with history, stories, songs and dances. During this initiation, the children will be exposed to the surrogate language and learn how to interpret it. Likewise, there is a corresponding initiation for girls, where they spend time learning important lessons, and they will also learn how to interpret the balafon (Trost, 1999). The frequency of these initiations varies by village. According to Trost, the major initiation occurs approximately every 10 years in Toussiana; my consultant from Djigouera said that the major initiation historically happened every 40 years, but there has not been an initiation since 1953 due to lack of interest. In many villages, these initiations and other events are becoming less frequent, in part because the organizers have often converted to Islam or Christianity and prioritize traditional religion less or view their new beliefs and the traditional beliefs as incompatible. Additionally, these initiations are difficult to plan around children's school schedules. Typically, the only extended break children get from school is from August to the end of September. During the school year, they are not able to devote such a long period of time away from their classes and during August and September, their parents often need their help planting or harvesting the fields.

As Islam and Christianity have gained a wider following in the villages, the cultural role of the balafon has been changing for certain groups of Christians and Muslims. Often, individuals will choose to have Muslim or Christian weddings and funerals, where the traditional Toussian balafon plays no major role. There is a rise in non-griot Toussian musicians who play popular music and they are often sought instead of griots for these events. They will usually play Dioula songs, as such songs are not associated with traditional religion, and frequently these musicians are unable to use the Toussian balafon surrogate language.

[6]Cents are normalized musical intervals. 100 cents correspond to one semitone; an octave is 1,200 cents.

[7]The numbering of the bars was chosen not for any inherent properties of the scale, but rather was numbered starting at the lowest bar on the balafon used by the griot with whom I worked.

The Surrogate Language

The Northern Toussian balafon surrogate language is a tone-based speech surrogate, where the tones of speech are represented by striking bars of certain pitches on the balafon. Additionally, aspects of the syllable structure, such as the presence of a consonant cluster or coda, as well as phrase boundaries can be encoded. Many other phonological features, such as vowel quality and type of consonant are not represented.

The speech surrogate can be quite ambiguous in many contexts because certain features are not represented; therefore, it is necessary to learn how to interpret it. As discussed above, this was traditionally taught during the great initiations, which have become increasingly infrequent. Very few people other than griots can still understand the speech surrogate, according to my surrogate language consultant Emile Diabaté and others whom I have interviewed on the subject. He has said that there are currently only a handful of older women who are able to understand the surrogate language. However, 'understanding the surrogate language' can have different connotations, as the surrogate language takes different forms depending on how it is used.

Many speech surrogates are found to employ several 'modes,'[8] which are different encodings of speech (Moore and Meyer, 2014; McPherson, 2019a). In the case of Seenku, there is a 'singing mode' and a 'speech mode,' where the surrogate language encodes aspects of speech differently in each mode. The speech mode is used when a griot communicates with another individual, and it follows the rhythms and structure of speech closely. The singing mode is used in songs, and its rhythm matches whichever song it is played in; moreover, the system is subject to stylistic modification and extra notes and passages might be improvised, further distancing the surrogate language from speech (McPherson, 2019a).

It is not currently understood how many modes the Northern Toussian balafon surrogate language has or when they are used. There appear to be at least two, a speech mode and a singing mode. The singing mode is used during songs—like the Seenku singing mode, it varies substantially from the speech mode. This is the mode which is most used in the public sphere, as the traditional Toussian songs employ it. However, these songs are often both sung and played with the surrogate language. Therefore, it is possible that many people have learned the lyrics of the songs and have learned the portions played in the speech surrogate by means of the lyrics, rather than learning how the system itself functions and being able to understand novel utterances of the surrogate language.

It is unclear how much, if at all, the speech mode is used in the public sphere–it is my impression that the speech mode is largely restricted to griot households, currently. Griots teach their children how to interpret the speech mode young, starting by asking the children to bring them food or drinks, building up proficiency until the children can chat and communicate with other griots using the surrogate language. However, outside of these households, few people can understand it. Moreover, it is equally unclear whether the people Emile claims can understand the surrogate language can interpret the speech mode, or instead are familiar with the various songs played in the singing mode. Historically, it is possible that the speech mode was used at many of the cultural events mentioned above, where conversations would be had with the people present at the events, as is still seen in Sambla villages (McPherson, 2019a). At present though, it appears that speech mode is no longer used in most contexts outside of Emile's home, at least in Djigouera, his home village.

[8]Not to be confused with the term "mode" used in musicology to refer to types of musical scales.

RELEVANT NORTHERN TOUSSIAN GRAMMAR

Basic Grammatical Features

The order of constituents in Northern Toussian is subject-object-verb (SOV) and word order is quite strict; there are only a handful of constructions which can alter this order. Various grammatical elements such as tense, aspect and mood (TAM) morphemes, as well as other elements like auxiliary verbs are found between the subject and the object. These words will henceforth be called 'auxiliaries.' Adjuncts, such as adverbials and postpositional phrases, are found following the verb. An example Toussian sentence is given below.

(2)	pə̀	wū	ɲīn	pὲ	sā̰	tə̂r=ɲīŋ
	3PL	EVID	name	good	place (v)	3SG.EMPH=on

'(It is known) that they gave him a great name.'[9] (0194)

Syllable Structure

Most words—excluding compounds—are monosyllabic in Northern Toussian. Syllables can be of the shape (C)CVC. The consonant inventory is found in (3); consonants in parentheses are exceedingly rare—some, like g͡b, are only found in borrowings. The symbols ⟨j⟩ and ⟨y⟩ correspond to IPA [ɟ] and [j], respectively.

(3) Consonant inventory

	Bilabial	Alveolar	Palatal	Velar	Labiovelar
Oral Stops	p b	t d	j	k	k͡p (g͡b)
Nasal Stops	m	n	ɲ	ŋ	
Fricatives	f (v)	s (z)	ʃ	ɣ	
Trills		r			
Approximants		l	y		w

Northern Toussian has an eight-vowel system, /i e ɛ ə a u o ɔ/, with a nasal series which lacks close mid vowels and the schwa, /ĩ ɛ̃ ã ũ ɔ̃/. Vowel length is not contrastive. Further research is required to determine whether vowel-glide and glide-vowel sequences are phonologically diphthongs; for the purpose of this paper, the glides will be transcribed as consonants.

Northern Toussian appears to have true consonant clusters; they do not seem to be reduced disyllabic words as is found in some languages in the region like Dioula, i.e., there is not a phonological rule which reduces CVCV to CCV. The first element of a consonant cluster is an obstruent or, in limited contexts, a sonorant. The second element can only be a sonorant—specifically, only *l, r, y,* or *w*. Examples of words with obstruent-initial consonant clusters are found in 4).

(4)	Cluster	example	gloss
	bl-	blê	‘to hang’
	bw-	bwɛ̄y	‘side’
	fl-	flê	‘woman’
	fw-	fwɛ̄	‘cloth’
	dr-	drɛ̄	‘tail’
	kr-	krɛ̂	‘that’

When both elements of a consonant cluster are sonorants, only *mw-* and *ny-*are permitted.

(5)	Cluster	example	gloss
	ny-	nyą̄	‘beehive’
	mw-	mwǫ	‘rat’

There are minimal pairs of *ɲ* and *ny*, such as *ɲār* ‘porridge,’ and *nyār* ‘oil.’

The following codas are permissible: *n, ŋ, m, r, y* and ɣ, though ɣ might not be a true coda. The sound ɣ is only found word-finally, as in the words below.

(6)	Word	gloss
	dāɣ	‘tomb’
	búmblâɣ	‘hyena’
	fə̀tàɣ	‘mud’
	nâɣ	‘ear’

However, ɣ has two allophones, [x] found phrase-finally and [ɣ] found phrase-medially. Whenever ɣ occurs phrase-medially, it is followed by a copy of the last vowel, which is always *a*. It is cognate to Southern Toussian /k/, which surfaces as [g] intervocalically. Historically, it appears that **k* lenited first to **g*, then to ɣ which was devoiced to *x* following the loss of the final vowel.

(7) Distribution of ɣ

a. Phrase-final

[mə́	émíl	pyáx]
1sg	Emile	greet

‘I greeted Emile’

b.

[mə́	émíl	pyáɣá	kàrì]
1sg	Emile	greet	today

‘I greeted Emile today’

Vowel-initial words are only found in auxiliaries and borrowings. When ɣ occurs before such a word phrase-medially, the *a* following the ɣ does not surface.

(8)

[blêmpàɣ	émíl	wé]
orphan	Emile	see

‘The orphan saw Emile’

Tonology

The balafon surrogate language encodes aspects of both lexical and postlexical tone in Northern Toussian. Lexical tones are the tones which are associated with specific words. Postlexical tone, on the other hand, is the set of tonal processes which are not tied to individual words, but to larger prosodic structures. This includes sandhi, when tones of certain words change when the words are adjacent to one another; phrasal effects, when the tones of words change due to their placement in a phrase; and downdrift, which will be elaborated below.

Lexical Tonology

Like most languages in the region, Northern Toussian has a complex tonal system—a considerable amount of lexical and grammatical information is conveyed by tone. In Northern Toussian, there are three distinct tone heights, high (H), mid (M) and low (L), written here with the diacritics *á*, *ā* and *à*, respectively.

A minimal set of these tonal contrasts are found in 9).

(9)	Word	tone	gloss
	yí	H	‘year’
	yī	M	‘laughter’
	yì	L	‘hair’

In addition to the level tones, there are three lexical contour tones which can be found on monosyllables, HM, HL and LH, written *a᷇*, *â* and *ǎ*.

(10)	Word	tone	gloss
	bá	H	‘porridge’
	bà	L	‘limit’
	bâ	HL	‘leg’
	bǎ	LH	‘poison’
	dɛ̂	HL	‘rock’
	dɛ᷇	HM	‘dream’

Contours of three tones can be found on a single vowel, but only as the result of the combination of a grammatical L tone and the HM contour.[10] For example, this is attested when intransitive verbs are used in a declarative context, as shown in (11a), where the underlyingly HM verb *ko᷇* surfaces with a LHM contour.

(11) Intransitive grammatical tone

a. ko᷇
walk.imp
‘walk!’

b.

kàrímù	ko᷈[11]
Karim	walk

‘Karim walked’

The tone bearing unit (TBU) appears to be the syllable in Northern Toussian; short open syllables can bear contour tones, and there appear not to be structural differences in tone assignment due to coda consonants. Nasal codas do not bear separate tones, though syllabic nasals can bear tone, as in the word *ǹʃɛ᷇* ‘ashes.’

[10]When this process occurs with a word which has a HL contour, the word surfaces as L instead of the expected LHL.

[11]There is no IPA diacritic for a LHM contour, so the LHL diacritic will be used instead.

Most combinations of tones can be found in disyllabic words, except for L.M, which is currently unattested in my lexicon. Here are several combinations of different tones in disyllabic words.

(12)

Word	tone	gloss
frībú	M.H	'scarab beetle'
înɔ̄	HL.M	'thing'
dúmpwɛ̂	H.HM	'testicle'
jibɛ̂y	L.HM	'esophagus'

It is exceedingly rare for monomorphemic disyllables to bear two different contour tones, though it is attested in the word *tǎrtâ* "Wednesday."

(13)

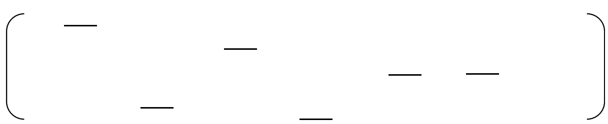
Word	tone	gloss
tǎrtâ	LH.HL	'Wednesday'

Postlexical Tone

Any combination of tones is permissible across word boundaries; there are no restrictions on which tones can occur on adjacent words. Moreover, few consistent tone sandhi rules have been identified, though some contour tones might be simplified at fast rates of speech. This is common for LH syllables, which will often surface as M or L, e.g., *bǒ* 'father,' which can be realized as *bō* when spoken quickly.

Downdrift and downstep play an important role both in speech and in the surrogate language. They are often characterized as tonal processes that lower the pitch of high tones following low tones; this lowering effect applies to all subsequent high tones within a particular domain (Connell, 2011). This is schematized in (14), where each horizontal line represents the pitch of the utterance. When a H follows a L, its pitch will be lower than the last H in the phrase. If there are two H tones following a single L, as we see with the final three tones of the diagram, both H tones will be at the same pitch level.

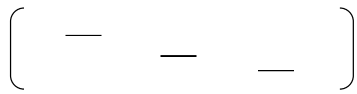
(14) H L H L H H



This lowering effect can be triggered both by surface L tones, where it is called "downdrift" or "automatic downstep," or by floating tones, where it is called "downstep" or "non-automatic downstep" (Connell, 2011). The terms "downdrift" and "downstep" will be used in this paper. The following discussion will focus on downdrift and not downstep, as many properties of the surrogate language's representation of downstep are still uncertain.

When languages have more than two tones, there is variation in which tones can trigger downdrift. In Yala, a three-tone language, a L and M can trigger downdrift of a H, and a L can trigger downdrift of a M (Armstrong, 1968). The language Seenku has four distinct tones and the lowest two tones can trigger downdrift on the highest tone (McPherson, 2020). In Northern Toussian, like Yala, L and M trigger downdrift on H tones, but unlike Yala, both L and M trigger downdrift on M tones, i.e., a series of mid tones will form a pitch track that resembles a staircase, as each M is lowered following the previous M, shown below.

(15) M M M



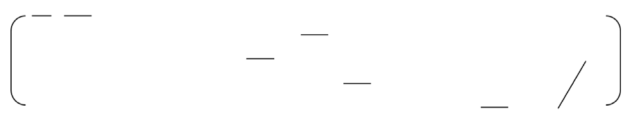
The amount that a tone drops in pitch from downdrift varies by language. In Northern Toussian, as well as Yala, a H lowered by downdrift has a higher pitch than a M. This is different from other languages such as Bimoba, where a lowered H is indistinguishable from a M, or Supyire, where the downdrift of a H can fall to the level of a M or land somewhere between a H and M depending on the speaker (Carlson, 1994; Snider, 1998).

In (16), we see approximations of the surface pitch of the phrase *tə́ = sú kɛ̄y ń = pī yɔ̀ wě* "His father's wife saw my child's broom." Following the word *kɛ̄y*, the surface pitches of the following words has been lowered due to downdrift. The occurrence of downdrift will be indicated by the symbol ⟨ꜜ⟩.

(16)

tə́=sú	kɛ̄y	ꜜń=pī	yɔ̀	wě
3SG.HUM.POSS=father	wife	1SG.POSS=child	broom	see

"His father's wife saw my child's broom"



SURROGATE LANGUAGE

Methodology

In this section, I will discuss the methodology used for data collection. I worked exclusively with the *ɲɔ́pī* Emile Diabaté from Djigouera in August 2019 and this description is based on how he plays the surrogate language. Most data were collected by elicitation, though there are several phrases which Emile produced unprompted—usually idiomatic expressions or jokes. Several recordings of songs were made, but it is currently unclear how the melody in a song might differ from the speech mode, so no songs were used as data for determining how the speech surrogate functions. I worked with a translator, Karim Traoré, who is my primary consultant for the spoken language. I would either ask for phrases in French which Karim would then translate for Emile, or I would say the phrases myself in Toussian. Emile would then produce the sentence on the balafon. All elicitation sessions were recorded as videos; there was approximately 22 h of elicitation, producing a corpus of around 700 phrases spoken on the balafon. Many of these were simple one- or two-word phrases, eliciting the tones of individual words or tonal paradigms. This methodology allowed me to investigate specific tonal phenomena; most of the structures I elicited were to better understand the structure of the balafon surrogate language, but many were to verify certain aspects of the tonology of the spoken language, for example eliciting a particular word whose tone had been difficult for me to hear.

There are notable limitations when using only elicited sentences for evaluating how the speech surrogate functions. For example, the duration of syllables often varies quite substantially, from as short as 80 ms to as long as 350 ms. Some of this variation is certainly due to inherent differences in syllable length, though much of it might be attributable to other factors, such as Emile slowing down so I could accurately detect which bars were being struck. Therefore, the duration of the syllables might not be representative of actual speech rate in certain phrases. Other features, such as the tones which are encoded on the balafon, do not vary at all when phrases are repeated and are likely unaffected by elicitation.

Ideally, to determine how much elicitation affects the production of the surrogate language, naturalistic data would be analyzed in conjunction with the elicited data. Naturalistic data would reveal how the surrogate language is used under normal circumstances, such as at cultural events or at home—if Emile does, in fact, vary his production of the speech surrogate to accommodate me, this would reveal which, if any, features' productions are affected. Conversely, perhaps the elicited data do reflect how the speech surrogate is typically used, and syllable length is always variable. Until naturalistic data have been analyzed, it is not possible to understand more deeply the nuances of the surrogate language.

However, this work was done in the regional capital Bobo-Dioulasso, not in Djigouera, so I have little data of the surrogate language being played naturally in either cultural events or in daily life beyond the accounts Emile has described to me. Another trip to gather naturalistic data in Djigouera was planned for the summer of 2020 but has been postponed due to the COVID-19 pandemic.

The following sections demonstrate how the surrogate language functions, starting with its basic structure in the *Basic surrogate language structure* Section. Following this, representation of syllable structure by the use of flams—two notes struck in very quick succession—for encoding consonant clusters and codas is shown in the *Codas and consonant clusters* Section. The surrogate language encodes phrase boundaries in certain circumstances, detailed in the *Phrase-final syllables* Section. Certain postlexical tonal processes are encoded with the balafon, namely downdrift, which are elaborated in the *Downdrift* Section. The inherent ambiguity of the system and disambiguation strategies are described in the *Managing ambiguity* Section. Finally, I discuss the linguistic structures encoded by the speech surrogate in the *Linguistic structure of the surrogate language* Section.

The Encoding of the Surrogate Language

Basic Surrogate Language Structure

The Northern Toussian balafon surrogate language functions by selecting a frame of three adjacent bars which appear to correspond to the three tones of speech. Emile generally used the bars *dú-fɔ̄-ɲɛ́pwɛ̄*, *yā̰-tɛ̰̀-ɲɛ́pwɛ̄*, and *ɲɛ̄pwɛ̄-kà*, which are 440, 473 and 533 Hz respectively, though sometimes he starts a phrase one bar higher or lower. This is approximately two octaves higher than the average pitch of his voice. This frame can be shifted up or down due to downdrift, but no notes will be played outside of this frame, with the exception of notes which represent phrase boundaries.

To represent a level tone, a single bar will be struck; a H by striking the bar with the highest pitch within the frame, a M by striking the middle bar and a L by striking the bar with the lowest pitch. Examples of the speech surrogate such as (17) will be schematized following the methodology of McPherson (2019a): the names of the bars and their respective scale heights are shown vertically on the left. The fifth bar is the highest scale degree, and the first is the lowest. There are some examples where the notes played span more than one octave; the numbering will restart for the next octave, e.g., (25b). The rows at the bottom show the sentence in Northern Toussian, a gloss, and the duration in milliseconds from the start of the note to the beginning of the following note. Since the griot will not stop the bar from resonating at the end of a phrase, the final note cannot provide any useful duration data, so it is represented with ellipses. The shaded cells show which bar is played for what word.

Example (17) shows a sentence of three monosyllables, each bearing a different tone. The word *à* is played on *dú-fɔ̄-ɲɛ́pwɛ̄*, *lɛ̰̄* on *yā̰-tɛ̰̄ɲɛ́pwɛ̄*, and *bɛ́y* on *ɲɛ́pwɛ̄-kà*. The widths of the cells do not represent duration.

(17) 'She/he sang the words'

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄			
2	dú-fɔ̄-ɲɛ́pwɛ̄			
1	tā-ɲɛ́pwɛ̄			
	words	à	lɛ̰̄	bɛ́y
	gloss	3SG	speech	sing
	durations (ms)	297	271	...

Example (18) shows two phrases which have disyllabic words.

(18) Phrases with disyllabic words

a. 'Sesame'

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄		
4	ɲɛ́pwɛ̄-kà		
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄		
2	dú-fɔ̄-ɲɛ́pwɛ̄		
1	tā-ɲɛ́pwɛ̄		
	words	dèmíŋ	
	gloss	sesame	
	durations (ms)	227	...

b. 'Your father saw a squirrel'

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄					
4	ɲɛ́pwɛ̄-kà					
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄					
2	dú-fɔ̄-ɲɛ́pwɛ̄					
1	tā-ɲɛ́pwɛ̄					
	words	á	sú	kɪ̰dí		wé
	gloss	2sg	father	squirrel		see
	durations (ms)	225	194	164	299	...

If the tone is a contour tone, it will be represented by a flam, a word adopted from percussion terminology which refers to a note preceded closely by a grace note. In the context of the balafon surrogate language, it refers to two strikes on the balafon played closely together. Flams are used to represent contour tones, as well consonant clusters and codas, detailed in *Codas and consonant clusters* Section. In (19), the LH contour in the first syllable of the word *bǎ* 'poison' is represented by the lowest and highest of the three bars of the frame, and the duration of the first note is significantly shorter than the other notes in the phrase since the contour is played with a flam.

(19) 'Your father ate poison'

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄					
4	ɲɛ́pwɛ̄-kà					
3	yā-tɛ̰̀-ɲɛ́pwɛ̄					
2	dú-fɔ̄-ɲɛ́pwɛ̄					
1	tā-ɲɛ́pwɛ̄					
	words	á	sú	bǎ		tá
	gloss	2sg.POSS	father	poison		eat
	durations (ms)	217	227	77	249	..

The first component of a contour tone is usually around 25–50% of the duration of a syllable with a level tone, as seen in (19), as well as (20a) and (20b). The second component of the contour tone is typically longer, sometimes as long as level-toned syllables. In (20b), three notes are struck, played on the scale degrees 4, 2 and 3. The notes played on 4 and 2 correspond to the syllable *î*, which has a HL contour tone. The first note of the contour tone is 77 ms, almost half the duration of the second note, which is 143 ms.

(20) Disyllabic contours

a. 'I saw a needle'

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄					
4	ɲɛ́pwɛ̄-kà					
3	yā-tɛ̰̀-ɲɛ́pwɛ̄					
2	dú-fɔ̄-ɲɛ́pwɛ̄					
1	tā-ɲɛ́pwɛ̄					
	words	mə́	mísɛ̌n			wé
	gloss	1SG	needle			see
	durations (ms)	481	197	96	218	...

b. 'Thing'

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā-tɛ̰̀-ɲɛ́pwɛ̄			
2	dú-fɔ̄-ɲɛ́pwɛ̄			
1	tā-ɲɛ́pwɛ̄			
	words	îɲɔ̄		
	gloss	thing		
	durations (ms)	77	143	...

The frame of three notes has been shifted down in (20a); the frame of (20a) is between the scale degrees 1 and 3, whereas the frame of (20b) is one note higher, between 2 and 4. The shift in frame above is not a systematic frame shift; there are examples of systematic frame shifting due to downdrift, which is discussed below (*Downdrift* Section). It is difficult to determine why (20a) has been shifted down a note—it could be changed to better represent the pitches of speech, or it might be for some non-linguistic reason such as stylistic variation.

Codas and Consonant Clusters

There are several additional properties of the surrogate language. Level-toned syllables with codas (*n, r, ŋ*, etc.) are often, but not always, struck with a flam. In (21), Emile played the sentence *mə́ sú wɔ̄n* "my father left" two different ways, one where *wɔ̄n* "leave" was struck once, the other where the same note is struck twice with a flam.

(21) 'My father left.'

a. With one strike for *wɔn* 'leave' in 'my father left'

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā-tɛ̰̀-ɲɛ́pwɛ̄			
2	dú-fɔ̄-ɲɛ́pwɛ̄			
1	tā-ɲɛ́pwɛ̄			
	words	mə́	sú	wɔ̄n
	gloss	1SG	father	leave
	durations (ms)	221	329	...

b. With two strikes for *wɔn* 'leave' in 'my father left'

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄				
4	ɲɛ́pwɛ̄-kà				
3	yā-tɛ̰̀-ɲɛ́pwɛ̄				
2	dú-fɔ̄-ɲɛ́pwɛ̄				
1	tā-ɲɛ́pwɛ̄				
	words	ń=[12]	sú	wɔ̄n	
	gloss	1SG	father	leave	
	durations (ms)	202	229	70	...

For the sentence *à pɔ̰̂* 'he/she is selling a pot' in example (22), Emile gave three versions, differing by the representation of two coda consonants. In the first, the *tár* in *tárkó* and the word *ɲɔŋ* were both struck twice (22a); in the second, only *ɲɔŋ* was struck twice (22b); in the final one, only *tár* was struck twice (22c). This variation is due to the presence of the coda consonant; since the subject does not have a coda (or consonant cluster), it will never be struck with two notes—the TAM marker is represented with a flam because of its contour tone.

(22) 'She/he is selling a pot'

a. Flams for both *tár*- and *ɲɔŋ*

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄								
4	ɲɛ́pwɛ̄-kà								
3	yā-tɛ̰̀-ɲɛ́pwɛ̄								
2	dú-fɔ̄-ɲɛ́pwɛ̄								
1	tā-ɲɛ́pwɛ̄								
	words	à	pɔ̰̂		tárkó			ɲɔ̄ŋ	
	gloss	3SG	IPFV .NPST		pot			sell	
	durations (ms)	287	199	260	84	168	284	94	...

b. Flam for only *ɲɔŋ*

5	kɔ̄tɔ̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā-tɛ̰̀-ɲɛ́pwɛ̄							
2	dú-fɔ̄-ɲɛ́pwɛ̄							
1	tā-ɲɛ́pwɛ̄							
	words	à	pɔ̰̂		tárkó		ɲɔ̄ŋ	
	gloss	3SG	IPFV .NPST		pot		sell	
	durations (ms)	199	109	192	198	284	108	...

c. Flam for only taŕ-

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄							
2	dú-fɔ̄-ɲɛ́pwɛ̄							
1	tā-ɲɛ́pwɛ̄							
	words	à	pə̂		tárkó			ɲɔ̄ŋ
	gloss	3SG	IPFV .NPST		pot			sell
	durations (ms)	184	126	191	85	137	252	...

Similarly, syllables beginning with consonant clusters such as *kr-*, *bl-*, *bw-*, etc. will sometimes be struck with a flam. As with codas, this is optional, as demonstrated in (23b) and (23c), where the former uses a flam and the latter does not.

(23) Consonant Clusters

a. 'My uncle got stuck'

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄				
4	ɲɛ́pwɛ̄-kà				
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄				
2	dú-fɔ̄-ɲɛ́pwɛ̄				
1	tā-ɲɛ́pwɛ̄				
	words	ń	lè	blè	
	gloss	1SG	uncle	stick	
	durations (ms)	184	126	191	...

b. 'Cough, father!' with a flam

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄			
2	dú-fɔ̄-ɲɛ́pwɛ̄			
1	tā-ɲɛ́pwɛ̄			
	words	sú	bwɛ́	
	gloss	father	cough	
	durations (ms)	270	118	...

c. 'Cough, father!' without a flam

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄		
4	ɲɛ́pwɛ̄-kà		
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄		
2	dú-fɔ̄-ɲɛ́pwɛ̄		
1	tā-ɲɛ́pwɛ̄		
	words	sú	bwɛ́
	gloss	father	cough
	durations (ms)	326	...

These double strikes might serve the useful function of disambiguating level-toned syllables with codas or consonant clusters from those without. Since flams tend to have about half the duration of full syllables, the two strikes for a consonant cluster or coda can usually be disambiguated from a series of strikes representing two separate syllables. However, (23a) shows a word with a consonant cluster where the duration between the first note of the flam is longer than the duration of the words before it. Therefore, while flams might often be useful cues, they do not unambiguously differentiate consonant clusters/codas and word/syllable boundaries.

All the data here are from elicitation and are not naturalistic; therefore, it is difficult to know how systematic the use of flams is. When eliciting a word or phrase, both the griot and the person eliciting know which words are played, so there is no ambiguity as to which words are being encoded on the balafon. However, if the griot is producing spontaneous speech with the balafon, the interlocutor will have to infer from context which specific words are being said. I hypothesize that the griots employ this doubled strike to disambiguate sentences such as *mə́ yɔ̌ mə́ nə̀ ɲār nə́ŋ* 'bring me food' from *mə́ yɔ̌ mə́ nə̀ tɔ̰̄ wó* 'bring me beer,' both of which are tonally identical and differ only in that the second-to-last word in the former sentence has a coda and the latter does not. The usage of flams in Northern Toussian is quite similar to their use in the Sambla balafon surrogate language in representing contour tones and coda consonants (McPherson, 2019a). Further comparisons between the two surrogate language systems will be made in the *Comparison to the Sambla Balafon Surrogate Language* Section.

Phrase-Final Syllables

Emile often ends phrases whose final tones are H or M by striking the same scale note across two octaves. This is presumably to demonstrate phrase boundaries, allowing interlocutors to better determine the ends of sentences. In elicitation sessions, Emile often varied when he would end phrases with octaves. Usually, he reserved this for more natural sentences which he would say when demonstrating idioms or making jokes. During elicitations where I would ask him to say certain words or sentences, he would not usually end them with octaves. Example (24) was an inside joke between Emile and my consultant Karim which Emile told when I was eliciting examples of the word *fwó* "to farm."

(24) 'Karim works the fields'

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄							
2	dú-fɔ̄-ɲɛ́pwɛ̄							
1	tā-ɲɛ́pwɛ̄							
5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄							
	words	Kàrímù			pə̂		swā	fwó
	gloss	Karim			IPFV .NPST		field	farm (v)[13]
	durations (ms)	184	126	191	85	137	252	...

In addition to striking octaves at the end of phrases, Emile sometimes struck phrase final L with a note much lower than what came before it. This occurs with phrase-final syllables bearing L or HL tones. Example (25a) demonstrates an extra-low note with a HL contour and (25b) with a word bearing a low tone. The latter uses a flam because of the consonant cluster.

[12]Northern Toussian employs both emphatic and clitic pronouns. Here, the clitic form *ń* = is used instead of *mə́*, seen earlier.

[13]This was not a perfect octave and the lower note was likely struck one note too low by accident.

[14]A single strike was used for a disyllabic word, perhaps reflecting a reduction to *snɔ̄* in speech.

(25) Extra-low notes

a. ‘I looked for father’

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄				
4	ɲɛ́pwɛ̄-kà				
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄				
2	dú-fɔ̄-ɲɛ́pwɛ̄				
1	tā-ɲɛ́pwɛ̄				
5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄				
4	ɲɛ́pwɛ̄-kà				
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄				
	words	mə́	sú	jâ	
	gloss	1SG	father	look	
	durations (ms)	281	244	68	...

b. ‘I dug’

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄			
2	dú-fɔ̄-ɲɛ́pwɛ̄			
1	tā-ɲɛ́pwɛ̄			
5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄			
	words	mə́	krɛ̀	
	gloss	1SG	dig	
	durations (ms)	333	79	...

In speech, phrase boundaries are not indicated by such a precipitous drop in pitch, therefore this appears to be a balafon-specific strategy of showing the end of phrases. These two patterns, striking octaves for H and M tones and striking extra-low notes for L and HL tones might be a consistent encoding—in my corpus, there are no examples of octaves struck for L and HL or extra low notes for H or M tones. Toussian additionally has the contours LH and HM, however there are no examples in my dataset of these tones using either method. Since there are no such examples it is unknown whether griots employ a specific method for representing phrase-final LH or HM tones.

The use of octaves or extra low notes is the only example of notes being played outside of the frame of notes. Generally, Emile picks a frame which spans the scale degrees 2–4, as seen in many of the figures above. There have been no examples of phrases which have the sequence of tones H L H which are played with the notes 4 1 4. Likewise, the sequence L H has not been represented by the notes 2 5. Therefore, it appears that each note within the frame maps to a particular lexical tone, rather than representing the surface pitch of speech. This hypothesis will be elaborated in the *Linguistic Structure of the Surrogate Language* Section below.

Downdrift

Downstep and downdrift can lead to many more surface pitches than there are lexical tones: when H is lowered due to downdrift, it surfaces higher than a M; likewise, when a M is lowered by a M or L, it surfaces higher than a L. Each instance of downdrift alters the pitch range, which means that there can be many more surface pitches than lexical tones.

Since the balafon can only play discrete notes, it is unable to precisely represent the subtle difference in pitch found as a result of downstep or downdrift while keeping within the three-bar frame. The Sambla balafon surrogate language resolves this issue simply by not representing downdrift or downstep. Northern Toussian does encode these phenomena in certain circumstances, using several different strategies. For the following examples, the gloss will show each instance where downdrift is found in speech. In the transcription of speech, downdrift will be indicated by the symbol ⟨ꜜ⟩. The realization of downdrift on the balafon can be seen in the diagrams and will be indicated in the prose below.

Most instances of downdrift which are encoded on the balafon surrogate language are found in sequences of M tones. In (26a), downdrift is encoded after every instance of a M, as *kɛ̰̄y* ‘wife’ is lower than *ɲɔ̄ŋ* “person,” *nīŋ* ‘water’ is lower than *kɛ̰̄y,* and *wé* ‘see’ is one note higher than *nīŋ,* which is lower than the phrase initial M on *ɲɔ̄ŋ* However, in (26b) downdrift is not encoded in the surrogate language following *kɛ̰̄y*, though it is found in speech.

(26) Downdrift

a. “The person’s wife saw the water”

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄				
4	ɲɛ́pwɛ̄-kà				
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄				
2	dú-fɔ̄-ɲɛ́pwɛ̄				
1	tā-ɲɛ́pwɛ̄				
	words	ɲɔ̄ŋ	ꜜkɛ̰̄y	ꜜnīŋ	ꜜwé
	gloss	person	wife	water	see
	durations (ms)	359	302	280	...

b. “The person’s wife saw the stream’s water”

5	kɔ̄tɔ̰̄-ɲɛ́pwɛ̄					
4	ɲɛ́pwɛ̄-kà					
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄					
2	dú-fɔ̄-ɲɛ́pwɛ̄					
1	tā-ɲɛ́pwɛ̄					
	words	ɲɔ̄ŋ	ꜜkɛ̰̄y	ꜜyɔ̄	ꜜnīŋ	ꜜwé
	gloss	person	wife	stream	water	see
	durations (ms)	265	305	291	296	...

In (27), the *-wú* of ‘hunter,’ *wé* ‘see,’ the past tense morpheme *á,* and *k͡pɔ̄* ‘kill’ are all subject to downdrift in speech, because they are preceded by a L or M, though the pitch lowering is not represented in the surrogate language. If downdrift were consistently encoded, one would expect the frame of three notes to shift down at each of these words.

(27) ‘If the hunter had seen the squirrel, he would have killed it.’

5	K															
4	N															
3	YT															
2	DF															
1	T															
	words	bānkô-ꜜwú				á	sə́nɔ́[14]	tɛ̰̆ʃwār			ꜜwé	à	ꜜá	nə̀	kə̀	ꜜk͡pɔ̄
	gloss	hunter				PST	COND	squirrel			see	COND	PAST	SS	3SG. NON HUM	kill
	(ms)	330	48	101	179	192	197	77	156	196	197	156	175	254	187	...

Generally, it appears that downdrift is encoded often for series of M tones, but rarely elsewhere. Downdrift between two different tones such as L H or L M will not substantially change the second

[15]An immediate sequencing morpheme, indicating that the following clause occurs immediately after the events of the first.

tone; there will be a slight difference in relative height between a H and a lowered H, but the H will still be high relative to a L or M. However, the difference between two M tones which are lowered and two which are not is much more significant, as the former will have two different pitches and the latter will have two identical pitches. It appears that downdrift between two M tones is more salient than downdrift between two different tones and is therefore more likely to be encoded.

When there are long sequences of M tones, in addition to not representing downdrift, as seen in (26b), the frame of tones might be reset and shifted up one bar. There does not appear to be a corresponding pitch rise in speech. In example (28a) between *kɛ̄y* 'wife' and *lɔ̄* 'cailcedrat,' the bars two and three are struck in the surrogate language, showing a rise in pitch on the balafon where the frame of notes is shifted up, but in the spoken language there is a drop in pitch between the two words due to downdrift. Similarly, in (28b), we find that a higher note is played for the word tɔ̰̄ even though it has a lower pitch than the preceding word in speech from downdrift.

(28) Frame shift

a. "The person's wife saw the cailcedrat's flower."

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄					
4	ɲɛ́pwɛ̄-kà					
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄					
2	dú-fɔ̄-ɲɛ́pwɛ̄					
1	tā-ɲɛ́pwɛ̄					
	words	nə̄ŋ	ꜜkɛ̄y	ꜜlɔ̄	ꜜfə̄ŋ	ꜜwé
	gloss	person	wife	cailcedrat	flower	see
	durations (ms)	321	330	282	297	...

b. "The person's wife insulted the odor of the millet beer (made) from the stream's water."

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄							
2	dú-fɔ̄-ɲɛ́pwɛ̄							
1	tā-ɲɛ́pwɛ̄							
	words	nə̄ŋ	ꜜkɛ̄y	ꜜyɔ̄	ꜜnīŋ	ꜜtɔ̰̄	ꜜpān	ꜜfī
	gloss	person	wife	stream	water	millet beer	odor	insult
	durations (ms)	359	300	338	348	300	362	

Instead of not representing downdrift, as we saw in (26b), the frame of notes is shifted up in (28). If each instance of downdrift is encoded on the balafon by shifting the frame down, then after only a few instances of downdrift, the pitch of the balafon will have diverged widely from speech. Resetting the frame up one note is a way to realign the pitch of the balafon so that it is more in accordance with speech.

There are three ways that the balafon surrogate language interacts with downdrift: 1) it ignores it and does not encode it; 2) it faithfully encodes downdrift by shifting the frame down and playing a lower-pitched note; 3) it resets the frame of notes up one, realigning the pitch of the balafon with the pitch of speech so that downdrift can be encoded on the next note. The variance in encoding seems to revolve around one central fact about the balafon surrogate language: it cannot accurately represent the many phonetic pitches of speech. If downdrift were encoded on the balafon at each instance of it in speech, then the pitches played by the balafon would be quite divergent from the pitches in speech. After each subsequent instance of downdrift in speech, the relative difference in pitch decreases, i.e., the pitch drops less and less after each time downstep is triggered. However, the intervals between the notes of the balafon stay the same across octaves and the balafon cannot encode the subtle changes in speech faithfully. Therefore, the intermittent encoding of downdrift and frame shifting seems to be ways of preventing the divergence of the pitches of speech from the pitches played on the balafon.

Managing Ambiguity

Since the surrogate language does not encode every aspect of speech, many utterances are ambiguous. For example, the speech surrogate utterance in (17), reproduced below, would resemble any series of monosyllabic words which bear the tonal sequence L M H. Several such examples are given in (30).

(29) 'She/he sang the words' (ZOOM0003 2019-8-15 13:02)

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄			
4	ɲɛ́pwɛ̄-kà			
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄			
2	dú-fɔ̄-ɲɛ́pwɛ̄			
1	tā-ɲɛ́pwɛ̄			
	words	à	lɛ̰̄	bɛ́y
	gloss	3SG	speech	sing
	durations (ms)	297	271	...

(30) Phrases with the tonal sequence L M H

a. pə̀ bō tá
3pl.hum African eggplant eat
'They ate the eggplant'

b. dɔ̀ dā wé
buffalo shea nut see
'The buffalo saw the shea nut.'

c. lò pɔ̄ dó
third son IS[15] sleep
'When the third son slept...'

There are a number of strategies which griots employ to reduce the ambiguity of a phrase. One such strategy is to use idiomatic constructions which are partially or completely lexicalized. These idiomatic phrases will be easily recognizable by many people, even those who cannot understand the surrogate language, and they can give context that will help the interlocutor understand the topic of discussion. Many of these idioms are found only in the speech surrogate and are not used in speech. For example, to call people for food, the idiomatic phrase 'an insect is flying because it is hungry,' perhaps better rendered in colloquial English as 'bugs are buzzing around because they are hungry.'

[17]Where a flam would be expected for encoding the contour tone, instead only a single L is struck. This is another example of how the idiomatic expressions do not always abide by the same rules detailed above.

[18]An example of downstep.

(31) 'Come eat food,' lit. 'an insect is flying because it is hungry.'

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄								
4	ɲɛ́pwɛ̄-kà								
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄								
2	dú-fɔ̄-ɲɛ́pwɛ̄								
1	tā-ɲɛ́pwɛ̄								
5	kə̄tɔ̰̄-ɲɛ́pwɛ̄								
4	ɲɛ́pwɛ̄-kà								
	words	sə̄mâ			ń=	dyɛ́	kə́	dyē	ꜜsəŋ
	gloss	insect			NPST	fly	3SG.NONHUM .POSS	intestines	because
	durations (ms)	350	231	285	300	255	320	434	...

It is still encoding Northern Toussian, rather than being a code completely divorced from language, but the phrase it encodes is used exclusively in the surrogate language, not in speech. Tonally, it differs from what we would expect in speech, as downdrift/downstep does not occur between *dyɛ́* and *kə́* in speech, but it is found within the surrogate language. This tonal difference appears not to be systematic in the speech surrogate and might reflect that, when the idiom was originally devised, the phrase used was slightly different, or that the spoken language has changed tonally while the idiom has resisted change.[16]

Some of these idiomatic expressions are partially productive, meaning that certain elements are replaceable. For example, when a griot requests something to eat or drink, they will use a phrase *mə́ yə̌ mə́ nə̀ N V*, literally 'I said that I will V N.' The phrase 'bring me a beer,' literally 'I said I will drink beer,' is given below.

(32) 'Bring me a beer,' lit. 'I said I will drink beer.'

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄							
2	dú-fɔ̄-ɲɛ́pwɛ̄							
1	tā-ɲɛ́pwɛ̄							
	words	mə́	yə̌[17]	mə́	nə̀	bíɛ́r		wó
	gloss	1SG	say	1SG	NPST	beer		drink
	durations (ms)	99	170	98	218	67	238	...

This is not the typical way a beverage is requested in speech, instead, a phrase such as 'I want N,' or 'give me N' will be employed, as seen in (33).

(33) Ways of requesting something in speech

a. mə́ ꜜpé[18] bíɛ́r=rɛ́
1sg COP beer=LOC
'I want beer,' lit. 'I am at beer'

b. mə́ kɔ̰́ bíɛ́r=sɛ̄
1SG give.IMP beer=with
'Give me beer,' more literally, 'supply me with beer'

The object and verb of this idiomatic expression can be replaced with other elements, including different drinks or types of food. Example (34) shows this construction with *tɔ̰̄* 'millet beer' and (35) with *dì vɛ̰́* 'some wine (*du vin*).'

[16]McPherson (2019a) posited that the singing mode of the Seenku balafon surrogate language might represent an older stage of the language, as the language of the singing mode is symbolic and based on proverbs.

(34) 'Bring me millet beer,' lit. 'I said I will drink millet beer.'

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄						
4	ɲɛ́pwɛ̄-kà						
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄						
2	dú-fɔ̄-ɲɛ́pwɛ̄						
1	tā-ɲɛ́pwɛ̄						
	words	mə́	yə̌	mə́	nə̀	tɔ̰̄	wó
	gloss	1SG	say	1SG	NPST	millet beer	drink
	durations (ms)	71	165	74	209	232	...

Other idioms include phrases for asking people to come, as well as ways of addressing different types of people. There are specific phrases used when addressing balafon griots, drum griots and blacksmiths.

(35) 'Bring me some wine,' lit. 'I said I will drink some wine.'

5	kə̄tɔ̰̄-ɲɛ́pwɛ̄							
4	ɲɛ́pwɛ̄-kà							
3	yā̰-tɛ̰̀-ɲɛ́pwɛ̄							
2	dú-fɔ̄-ɲɛ́pwɛ̄							
1	tā-ɲɛ́pwɛ̄							
	words	mə́	yə̌	mə́	nə̀	dì	vɛ̰́	wó
	gloss	1SG	say	1SG	NPST	some	wine	drink
	durations (ms)	116	185	106	222	152	318	...

A common feature of these idioms is redundancy; they express ideas in sometimes circuitous ways, so that there is more input which a listener can analyze and interpret. As stated above, there is not yet naturalistic data to analyze, but redundancy is potentially employed even in fully productive and fluent use of the surrogate language to aid in comprehension.

In addition to the use of redundancy and idioms, representation of phrase boundaries, consonant clusters and codas, as discussed in the *Codas and consonant clusters* Section and *Phrase-final syllables* Section, are potentially other ways of reducing the ambiguity of the phrase. Moreover, if they are viewed as disambiguation strategies, rather than systematic, structural features of the surrogate language, this could explain why their use is optional. If a particular phrase is being encoded which can only be read in one way, there is no reason to encode consonant clusters or codas. However, if two or more phrases might be valid for a particular utterance, representing consonant clusters or codas could differentiate the phrases. For example, the phrases in (36) are both phrases with three monosyllabic words which have a sequence of tones L M H. None of the words (36a) can be represented by flams, since there are no contour tones, consonant clusters, or codas. In (36b), both *ɲār* 'porridge' and *ɲə́ŋ* 'eat,' have codas and can be represented with flams. Using a flam for either *ɲār* or *ɲə́ŋ* will necessarily indicate to the listener that (36a) cannot be what the speaker is saying. If no flams are used, it will still be ambiguous between the two readings, but the fact that no flam was present in an ambiguous phrase might intimate to the listener that the coda-less phrase was the intended one.

(36) Tonally identical phrases

a. mə́ tɔ̰̄ wó
1SG millet beer drink
'I drank millet beer.'

b. mə́ ɲār ɲə́ŋ
1SG porridge eat
'I ate porridge.'

Though the surrogate language might at first appear very ambiguous, it employs several strategies which can help the listener interpret what is being represented. These include a number of useful idioms, ample use of redundancy and representation of linguistic features other than tone, namely syllable structure and phrase boundaries.

Linguistic Structure of the Surrogate Language

As has been demonstrated, the primary linguistic feature encoded by the Northern Toussian surrogate language is tone. Both lexical and postlexical tone is encoded, depending on the utterance. The frame of three notes seems to correspond to the lexical tones of speech and the griot will not strike notes outside of this frame except for when phrase boundaries are encoded. This might indicate that, for most utterances, the lexical tones of speech are encoded, rather than just the surface pitches. However, downdrift, a postlexical feature of tone, is at times encoded, though largely restricted to series of M tones. It is either faithfully represented by shifting the frame of notes down by one scale degree, or unfaithfully represented by shifting the frame of notes up, allowing faithful representation of downdrift later in the phrase. The shifting of frames has many parallels with the register shifting effects of downdrift in speech. In effect, these different approaches to representing downdrift appear to allow the realization of surface-level tonal phenomena while maintaining the lexical identity of the tones. Moreover, they accommodate the limitations of the medium used to play the speech surrogate, namely the discrete nature of the notes of a balafon.

All the other features which are encoded by the balafon surrogate language appear to be part of the prosody of speech, including aspects of syllable structure, namely complex onsets and codas, as well as phrase boundaries. The encoding of all these features is optional, though there are trends for when they are used.

As codas are represented by flams, one might be tempted to posit that the surrogate language is encoding moraic structure. Complex onsets are also represented by flams, which presents some issues for this hypothesis, as onsets are typically analyzed as not being moraic (Hayes, 1989). Recent work by (Topintzi and Nevins (2006; 2017)) posits that onsets can be moraic; however, there are no clear cases of CCV being heavier than CV, which would be expected if the balafon represents CCV differently from CV. Instead of representing moras, the speech surrogate appears to use flams whenever the syllable diverges from a level-toned, CV structure.

In sum, the balafon surrogate language utilizes many aspects of the phonological structure of speech, including both lexical and postlexical tone, the syllable structure of words and phrase boundaries.

COMPARISON TO THE SAMBLA BALAFON SURROGATE LANGUAGE

The Sambla people have a balafon surrogate language which is reported to have been brought to them by the Toussian (Strand, 2009). This raises several questions about how similar they are structurally—whether the Sambla learned how the Toussian surrogate language functions and fit this system to their language, or if they saw the balafon being used for a speech surrogate and devised a new system independent of the Toussian system.

Genetically, Northern Toussian and Seenku are quite different. Seenku is a Mande language, whereas Northern Toussian has generally been classified as a Gur language. Seenku has four contrastive tone levels, phonemic vowel length, it only allows a single noncontrastive nasal phoneme as coda, and many words have sesquisyllabic structure, meaning the words have a CəCV structure where the first vowel is a greatly reduced schwa which does not bear its own tone (McPherson, 2020). Northern Toussian exhibits certain differences, as it has three contrastive tone levels, non-contrastive vowel length, and many different sonorants as codas. It does have examples of words of the shape Cə(C) CV, though the schwa can bear its own tone, as in the word *kə̀tyɛ̄* 'courtyard.' However, regardless of the structural differences, the core fundamentals of the Toussian surrogate language are quite similar to the Seenku surrogate language. Like the Northern Toussian surrogate language, syllables with level tones are encoded by single strikes and contours as well as codas are encoded by flams; Seenku extends the use of flams to long vowels, and sesquisyllabic words are encoded in the same way as onset consonant clusters in Northern Toussian. It appears that in both languages, flams can be used for any complex syllable which is not of the shape CV.

There is one important structural difference between the two surrogate languages: the Northern Toussian balafon surrogate language often represents postlexical processes, whereas the Sambla balafon surrogate language generally encodes only lexical processes—the Sambla balafon surrogate language does not encode downdrift. This can give credence to the hypothesis that, if the Sambla did indeed learn how to play the balafon from the Toussian and based their balafon surrogate language on the Toussian one, they did not learn it structurally, but rather created their own system. That is, they were not trying to replicate how precisely the speech surrogate functions at a basic level—instead, they created a speech surrogate which aesthetically resembled the Toussian balafon surrogate language but did not wed itself closely to the inner-workings of Toussian surrogate language. With that said, the Toussian balafon surrogate language does not strictly encode all postlexical processes. In Northern Toussian, especially in rapid speech, monosyllabic LH syllables tend to flatten and sound similar to M. This is especially true for the postposition *=rɔ̌* 'in,' which I only learned bears a LH tone from hearing it played on the balafon. However, I have not found any words bearing a LH contour which were represented by a single M strike on the balafon. Therefore, it cannot be said that the Northern Toussian balafon surrogate language only represents surface tone; rather it exists on a gradient where the Northern Toussian surrogate language tends to encode more postlexical processes but does not strictly represent the surface form of the spoken language.

Beyond structural differences, there are several notable differences between the balafon culture of the Toussian and the Sambla. The Sambla view the balafon as speaking for itself—it is not the griot who chooses the words, but rather the instrument itself. A griot will never speak Seenku while playing a balafon, only the surrogate language. If a griot who is using the speech surrogate and an interlocutor are speaking and the griot insults the other person, the person cannot become angry at the griot, as it is not viewed as the griot speaking, but the balafon itself (Strand, 2009). The Toussian do not hold this perspective; they view the griot as speaking with the instrument, and Toussian griots have no qualms accompanying the balafon with speech or singing. Though they will not speak, the Sambla griots will hum when playing the surrogate language—this is also found with Toussian griots (McPherson, personal communication).

Certain differences between the Northern Toussian balafon surrogate language and the Sambla balafon surrogate language have been underscored—it must be noted that, while the balafon is rather new to the Sambla, present only for ~ 130 years, the Sambla have played other instruments long before the balafon was introduced. These include a flute and a horn, both of which have their own speech surrogates. The horn, unlike the balafon, can play any pitch the musician can vocalize and therefore is not restricted to the discrete notes that the balafons are. However, the flute has the same tuning as the balafon (McPherson, 2019a). Perhaps, then, when the Toussian brought the balafon to the Sambla, they adopted certain features of the Toussian balafon surrogate language such as the use of flams, but also transferred aspects of their flute or horn speech surrogate to the balafon, such as the tuning and the degree of postlexical representation. The use of flutes and horns could also explain the prohibition on speaking while playing the balafon surrogate language—if they had a well-established surrogate language tradition based on instruments where it is physically impossible to speak and play at the same time, that could have developed into a cultural prohibition.

CONCLUSION AND FUTURE DIRECTIONS

This paper has described the musical culture of the Toussian and demonstrated how the surrogate language functions at a basic level, showing that lexical tone is the primary feature represented by the balafon. Likewise, it has described the other aspects of prosody which are represented by the speech surrogate, namely syllable structure, phrase boundaries, and downdrift. Complex syllables are represented optionally by flams and might serve to disambiguate phrases. Downdrift is sometimes represented as found in speech, by lowering the pitch of a word, but other times, it is represented by raising the frame of notes even though the pitch of the spoken word was lowered from downdrift allowing downdrift to be encoded faithfully later in the phrase. In addition to the structure of the surrogate language, this paper has described the cultural and musical exchange between the Toussian and Sambla.

The most pressing work that remains to be done on the Northern Toussian balafon surrogate language is to gather a corpus of naturalistic data. All data used for this study were elicited and it is possible that there are substantial differences between naturalistic and elicited data. Emile explicitly stated that he frequently talks with his family using the speech surrogate, though it is currently unclear what its function currently is within the wider culture. According to Emile and Karim, it seems that most people who are not griots can no longer understand the speech surrogate, but the rate of comprehension remains to be determined. Naturalistic data would be of interest for linguists when studying Northern Toussian phonology or the encoding of speech by surrogate languages, as well as for ethnomusicologists, but most importantly, they would preserve an endangered tradition that seamlessly unites music and language.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Michigan. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This material is based upon work supported by the National Science Foundation-National Endowment for the Humanities Documenting Endangered Languages Program and the National Science Foundation Program in Linguistics, under National Science Foundation Award No. BCS-1943896 awarded to PI Laura McPherson.

REFERENCES

Armstrong, R. (1968). Yala (Ikom): A Terraced-Level Language with Three Tones. *J. West Afr. Languages* 5, 49–58.

Carlson, R. (1994). *A Grammar of Supyire*. Berlin: DE GRUYTER MOUTON. doi:10.1515/9783110883053

Connell, B. (2011). "Downstep," in *The Blackwell Companion To Phonology*. Editors M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (Oxford, UK: John Wiley & Sons), 1–24. doi:10.1002/9781444335262.wbctp0035

Eberhard, David. M., Simons, Gary. F., and Fennig, Charles. D. (2020). *Ethnologue: Languages of the World*. 23rd edn. Dallas: SIL International. Avaiable at: http://www.ethnologue.com.

Hammarström, Harald., Forkel, Robert., and Haspelmath, Martin. Sebastian Bank (2020). *Glottolog 4.3*. Jena. doi:10.5281/zenodo.4061162

Hayes, B. (1989). Compensatory Lengthening in Moraic Phonology. *Linguistic Inq.* 20 (2), 253–306.

McPherson, Laura. (2020). A Grammar of Seenku *(Mouton Grammar Library 83)*. 1st edn. Boston: De Gruyter.

McPherson, L. (2019c). Musical Adaptation as Phonological Evidence: Case Studies from Textsetting, Rhyme, and Musical Surrogates. *Lang. Linguist. Compass* 13 (12). doi:10.1111/lnc3.12359

McPherson, L. (2019b). The Role of Music in Documenting Phonological Grammar: Two Case Studies from West Africa. *Amp.* 7. doi:10.3765/amp.v7i0.4493

McPherson, L. (2019a). The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. *Anthropological Linguistics* 60 (3), 255–294. doi:10.1353/anl.2019.0006

Moore, Denny., and Meyer, Julien. (2014). The Study of Tone and Related Phenomena in an Amazonian Tone Language: Gavião of Rondônia. *Lang. Documentation* 8, 24.

Naden, J. A. (1989). "Gur," in The Niger-Congo Languages: A Classification And Description Of Africa's Largest Language Family, *140–168*. Editor John. Bendor-Samuel (Lanham: University Press of America).

Proschan, F. (1993). Khmu Play Languages. *Mon-Khmer Stud.* 23, 43–65.

Prost, A. (1964). "Le Toussian,". Contribution à l'étude des langues voltaïques *(Mémoires de l'Institut Français d'Afrique Noire)*. Editor A. Prost (Dakar: Institut Fondamental d'Afrique Noire, Université de Dakar), 70, 239–342.

Snider, K. L. (1998). Phonetic Realisation of Downstep in Bimoba. *Phonology* 15 (1), 77–101. doi:10.1017/S0952675798003534

Strand, J. L. (2009). *The Sambla Xylophone: Tradition and Identity in Burkina Faso*. Middletown, CT: Wesleyan University Doctor of Philosophy. doi:10.14418/wes01.3.3

Tom. Güldemann (Editor) (2018). The Languages and Linguistics of Africa *(The World of Linguistics* (Berlin; Boston: De Gruyter Mouton), 11.

Topintzi, I. (2006). *Moraic Onsets*. London: University of London Dissertation.

Topintzi, N., and Nevins, A. (2017). Moraic Onsets in Arrernte. *Phonology* 34 (3), 615–650. doi:10.1017/S0952675717000306

Trost, F. (1999). Musik und Xylophontexte der südlichen Tussian (Burkina Faso). *Anthropos. Anthropos. Inst.* 94 (1/3), 163–180.

Villepastour, A. (2010). Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code *(SOAS Musicology Series). Farnham, England*. Burlington, VT: Ashgate.

Wiesmann, H. (2000). in Eléments du discours narratif dans les textes win (toussian du sud) *(Cahiers Voltaïques (Gur Papers), Hors Série)*. Editor H. Wiesmann (Bayreuth: Universität Bayreuth), 2.

Wiesmann, H., Barro, K., and Coulibaly, S. (2004). *Wín wɛlɔ dúkúrí : Dictionnaire toussian-français, français-toussian. Banfora, Burkina Faso: Equipe toussian de la Société Internationale de Linguistique*.

Conflict of Interest: The author declares declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership