

# Speech and Language Processing

## Chapter 8 of SLP Speech Synthesis / Prosody

### Prosody:

from words+phones to boundaries, accent, F0, duration

- Prosodic phrasing
  - Need to break utterances into phrases
  - Punctuation is useful, not sufficient
- Accents:
  - Predictions of accents: which syllables should be accented
  - Realization of F0 contour: given accents/tones, generate F0 contour
- Duration:
  - Predicting duration of each phone

5/19/2011

Speech and Language Processing, Jurafsky and Martin 2

### Defining Intonation

- Ladd (1996) "Intonational phonology"
- "The use of **suprasegmental phonetic** features
  - Suprasegmental = above and beyond the segment/phone
    - F0
    - Intensity (energy)
    - Duration
- to convey **sentence-level pragmatic meanings**"
  - i.e. meanings that apply to phrases or utterances as a whole, not lexical stress, not lexical tone.

5/19/2011

Speech and Language Processing, Jurafsky and Martin 3

### Three aspects of prosody

- **Prominence**: some syllables/words are more prominent than others
- **Structure/boundaries**: sentences have prosodic structure
  - Some words group naturally together
  - Others have a noticeable break or disjuncture between them
- **Tune**: the intonational melody of an utterance.

5/19/2011

From Ladd (1996)

Speech and Language Processing, Jurafsky and Martin 4

### Prosodic Prominence: Pitch Accents

A: What types of foods are a good source of vitamins?

B1: Legumes are a good source of VITAMINS.

B2: LEGUMES are a good source of vitamins.

- Prominent syllables are:
  - Louder
  - Longer
  - Have higher F0 and/or sharper changes in F0 (higher F0 velocity)

5/19/2011

Slide from Jennifer Venditti

Speech and Language Processing, Jurafsky and Martin 5

### Stress vs. accent (2)

- The speaker decides to make the word **vitamin** more prominent by accenting it.
- Lexical stress tell us that this prominence will appear on the first syllable, hence **Vitamin**.

5/19/2011

Speech and Language Processing, Jurafsky and Martin 6

## Which word receives an accent?

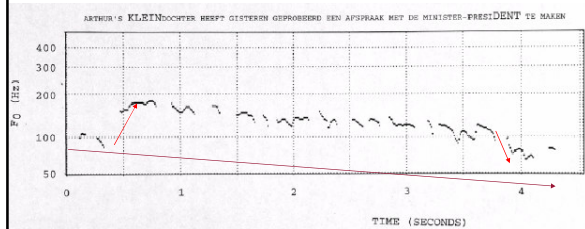
- It depends on the context. For example, the 'new' information in the answer to a question is often accented, while the 'old' information usually is not.
  - Q1: What types of foods are a good source of vitamins?
    - A1: LEGUMES are a good source of vitamins.
  - Q2: Are legumes a source of vitamins?
    - A2: Legumes are a GOOD source of vitamins.
  - Q3: I've heard that legumes are healthy, but what are they a good source of?
    - A3: Legumes are a good source of VITAMINS.

5/19/2011

Slide from Jennifer Venditti

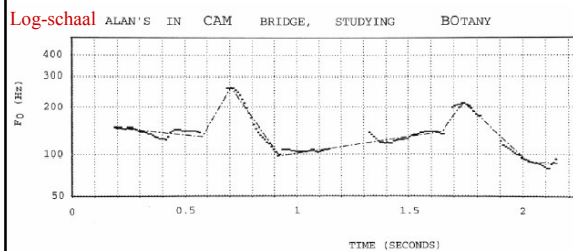
Speech and Language Processing: Jurafsky and Martin 7

## Welke variatie doet ertoe?



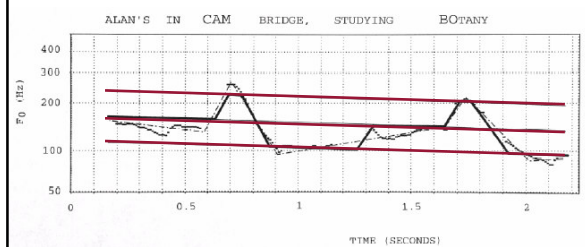
- Rechte lijn benadering, met minimaal aantal lijnstukken
- Verschil niet hoorbaar

## Close-copy stylization



- Rechte lijn benadering, met minimaal aantal lijnstukken
- Verschil niet hoorbaar

## De rol van declinatielijnen

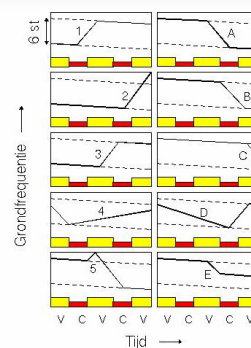


- Geleidelijk drukverlies in longen: algemene toonhoogtedaling
- Vereenvoudigd beschrijving

## declinatie & patronen

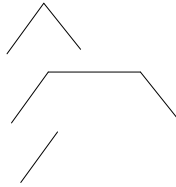
- Aantal niveaus (# declinatielijnen)
- Aantal relevante toonhoogtebewegingen
  - Nederlands 10 2 niveaus
  - Brits-Engels 27 3 niveaus
  - Duits 11 2 niveaus

## NL: 10 bewegingen, stijgend / dalend (IPO)



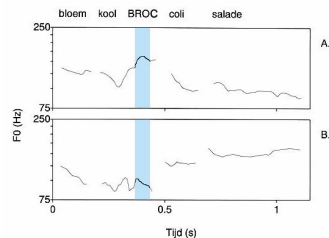
## Belangrijke patronen

- Punthoed
- Platte hoed
- Vraagstijging



Maar **waar** komt een stijging, waar een daling

## accentverlenend en grensmarkerend



Welk verloop signaleert welke functie?

## Regels (Odijsk)

- Benadruk alleen inhoudswoorden, niet de functiewoorden
  - *Wel*: zelfstandige naamwoorden, werkwoorden, bijwoorden, bijvoeglijke naamwoorden
  - *Niet*: lidwoorden, voorzetsels, voegwoorden, hulwerkwoorden
- mbv Part-Of-Speech tagging
- 80% succesvol

## maar

- Hij ZAG iemand
- Hij ZAG iedereen (Hij zag IEDEREEN)
- Modificeer "functiewoorden"
- Zelfstandig werkwoord krijgt alleen accent als geen ander woord accent krijgt

## Nog moeilijker: partikels krijgen accent

- Hij belde haar OP
- Hij zat *op* een STOEL
- Hij heeft er *tegen* OP gezien

Dit vereist een complexe linguïstische analyse

## Complex Noun Phrase Structure

- Sproat, R. 1994. English noun-phrase accent prediction for text-to-speech. *Computer Speech and Language* 8:79-94.
- Proper Names, stress on right-most word
  - New York CITY; Paris, FRANCE
- Adjective-Noun combinations, stress on noun
  - Large HOUSE, red PEN, new NOTEBOOK
- Noun-Noun compounds: stress left noun
  - HOTdog (food) versus HOT DOG (overheated animal)
  - WHITE house (place) versus WHITE HOUSE (made of stucco)
- examples:
  - MEDICAL Building, APPLE cake, cherry PIE.
  - What about: Madison avenue, Park street ???
- Some Rules:
  - Furniture+Room -> RIGHT (e.g., kitchen TABLE)
  - Proper-name + Street -> LEFT (e.g. PARK street)

## State of the art

- **Hand-label** large training sets
- Use **CART, SVM, CRF**, etc to predict accent  
(Classification And Regression Trees, Support Vector Machines, Conditional Random Field)
- Lots of rich features from context (parts of speech, syntactic structure, information structure, contrast, etc.)
- Classic lit:
  - Hirschberg, Julia. 1993. Pitch Accent in context: predicting intonational prominence from text. Artificial Intelligence 63, 305-340

5/19/2011

Speech and Language Processing, Jurafsky and Martin 19

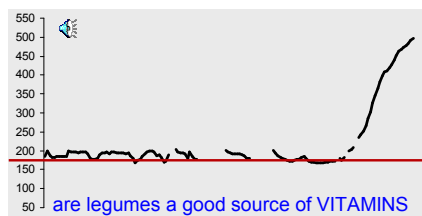
## Levels of prominence

- Most phrases have more than one accent
- The last accent in a phrase is perceived as more prominent
  - Called the **Nuclear Accent**
- **Emphatic** accents like nuclear accent often used for semantic purposes, such as indicating that a word is contrastive, or the semantic focus.
  - The kind of thing you represent via \*\*\*s in IM, or capitalized letters
  - 'I know **SOMETHING** interesting is sure to happen,' she said to herself.
- Can also have words that are **less** prominent than usual
  - Reduced words, especially function words.
- Often use 4 classes of prominence:
  1. **emphatic accent**,
  2. **pitch accent**,
  3. **unaccented**,
  4. **reduced**

5/19/2011

Speech and Language Processing, Jurafsky and Martin 20

## Yes-No question



Rise from the main accent to the end of the sentence.

5/19/2011

Slide from Jennifer Venditti

Speech and Language Processing, Jurafsky and Martin 21

## 'Surprise-redundancy' tune



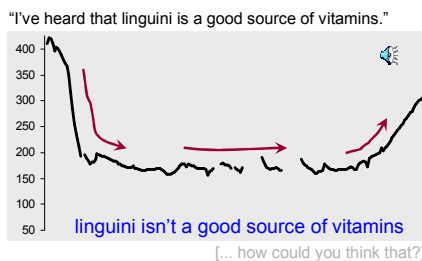
Low beginning followed by a gradual rise to a high at the end.

5/19/2011

Slide from Jennifer Venditti

Speech and Language Processing, Jurafsky and Martin 22

## 'Contradiction' tune



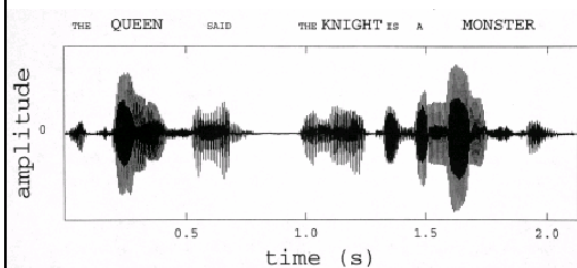
Sharp fall at the beginning, flat and low, then rising at the end.

5/19/2011

Slide from Jennifer Venditti

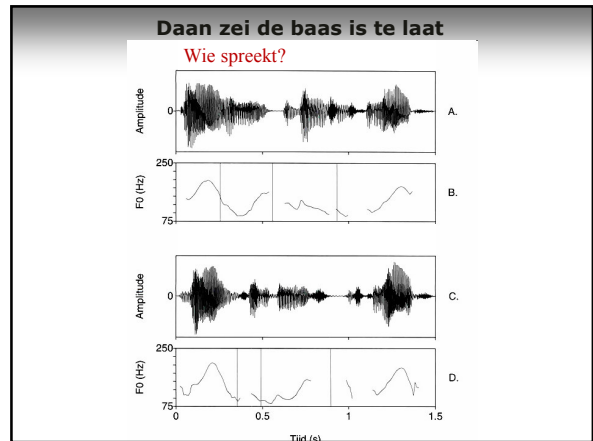
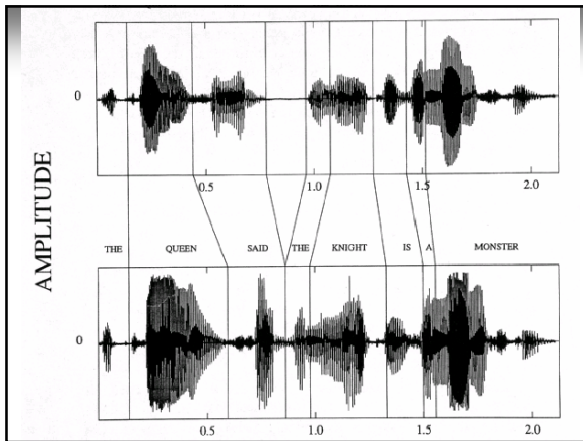
Speech and Language Processing, Jurafsky and Martin 23

## Duren & ritme



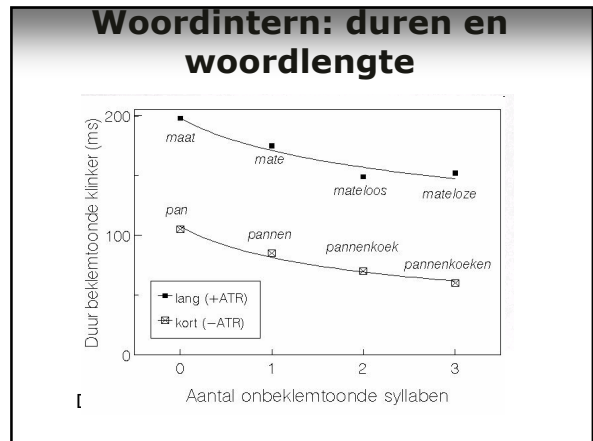
Vgl The QUEEN, said the Knight, is a MONSTER

5/19/2011



### Oude discussie

- Stress-timed - ~ gelijke duren tussen beklemtoonde syllaben (germaanse talen)
- Syllable-timed - ~ gelijke duur van syllaben (romaanse talen)



### Temporele effecten

Domein	Effect
<b>klank</b>	Intrinsieke duren van V (lange en korte klinkers)
<b>syllabe</b>	Intrinsieke duren van C Co-intrinsiek: verlenging van V voor stemhebbende C Co-intrinsiek: verlenging van C na korte V Verkorting van C indien optredend in cluster
<b>voet</b>	Verlenging van syllabe voor einde van frase/zin
<b>woord</b>	Sterke syllabe binnen voet langer dan zwakke syllabe Sterke voet langer dan zwakke voet Syllaben korter naarmate meer syllaben in een woord Proportionele verlenging van alle klanken indien woord [+accent]
<b>zin</b>	Hoger tempo naarmate meer woorden in een zin Functiewoorden korter dan inhoudswoorden
<b>alinea</b>	Hoger tempo naarmate meer zinnen in de alinea

### Duration

- Simplest:**
  - fixed size for all phones (100 ms)
- Next simplest:**
  - average duration for that phone (from training data). Samples from SWBD in ms:

aa	118	b	68
ax	59	d	68
ay	138	dh	44
eh	87	f	90
ih	77	g	66
- Next Next Simplest:**
  - add in phrase-final and initial lengthening plus stress:

Dutch durations (next simplest):  
[www.let.uu.nl/~Gerrit.Bloothoof/personal/onderwijs/TST/TST\\_MBROLA.htm](http://www.let.uu.nl/~Gerrit.Bloothoof/personal/onderwijs/TST/TST_MBROLA.htm)

## Higher level prosody

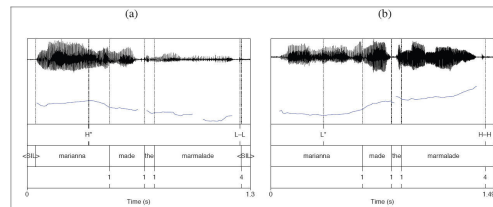
### ▪ Tobi (Tone and Break Indices)

Pitch Accents		Boundary Tones	
H*	peak accent	L-L%	"final fall": "declarative contour" of American English
L*	low accent	L-H%	continuation rise
L*+H	scooped accent	H-H%	"question rise": cantonal yes-no question contour
L+H*	rising peak accent	H-L%	final level plateau (plateau because H- causes "upstep" of following)
H*!H*	step down		

Todi (for Dutch)

<http://todi.let.kun.nl/>

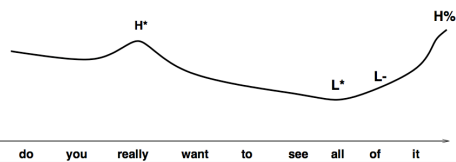
## Two patterns



## Intermediate representation: using Festival

### ▪ Do you really want to see all of it?

do	you	H* really	want	to	see	L* all	of	L- H% it
d	uw	r	w	t	s	ao	ah	v
110	110	50	75	64	57	82	57	50
72	41	43	47	54	130	76	90	44
62	46	220						



5/19/2011

Speech and Language Processing: Jurafsky and Martin 33

## Pauzes

- Benadrukken frasescheiding en prosodische grenzen
- Finale verlenging voor pauze
- Zijn vaak optioneel

## Prosodische frasering

Syntactische structuur >

Metrische structuur >

Fraserings regels

Intonatie frasen met zware en lichte grenzen

## Tekst structurering (alinea)

**F0**

**duur**

### Markering van de eerste zin van een alinea

relatief hoge eerste F0-piek  
relatief hoge inzet van de lage declinatie lijn

### Markering van de laatste zin van een alinea

relatief lage eerste F0-piek  
relatief lage inzet van de lage declinatie lijn

hoge spreesnelheid  
(behalve op de finale lettergrepen)  
extra lange finale rekking  
langere pauzeduur tussen alinea's  
dan tussen zinnen