

Spraakakoestiek

Gerrit Bloothoof

Te gebruiken bij het college Fonetiek van de Inleiding Taalwetenschap.

Tekst met een klein lettertype is optioneel. Dat geldt ook voor de paragrafen 8 en 9.

November 2008

1. Inleiding

Spraak bestaat uit luchttrillingen die bedoeld zijn om een betekenis van een spreker naar een hoorder over te brengen. De spraakakoestiek beschrijft hun akoestische eigenschappen.

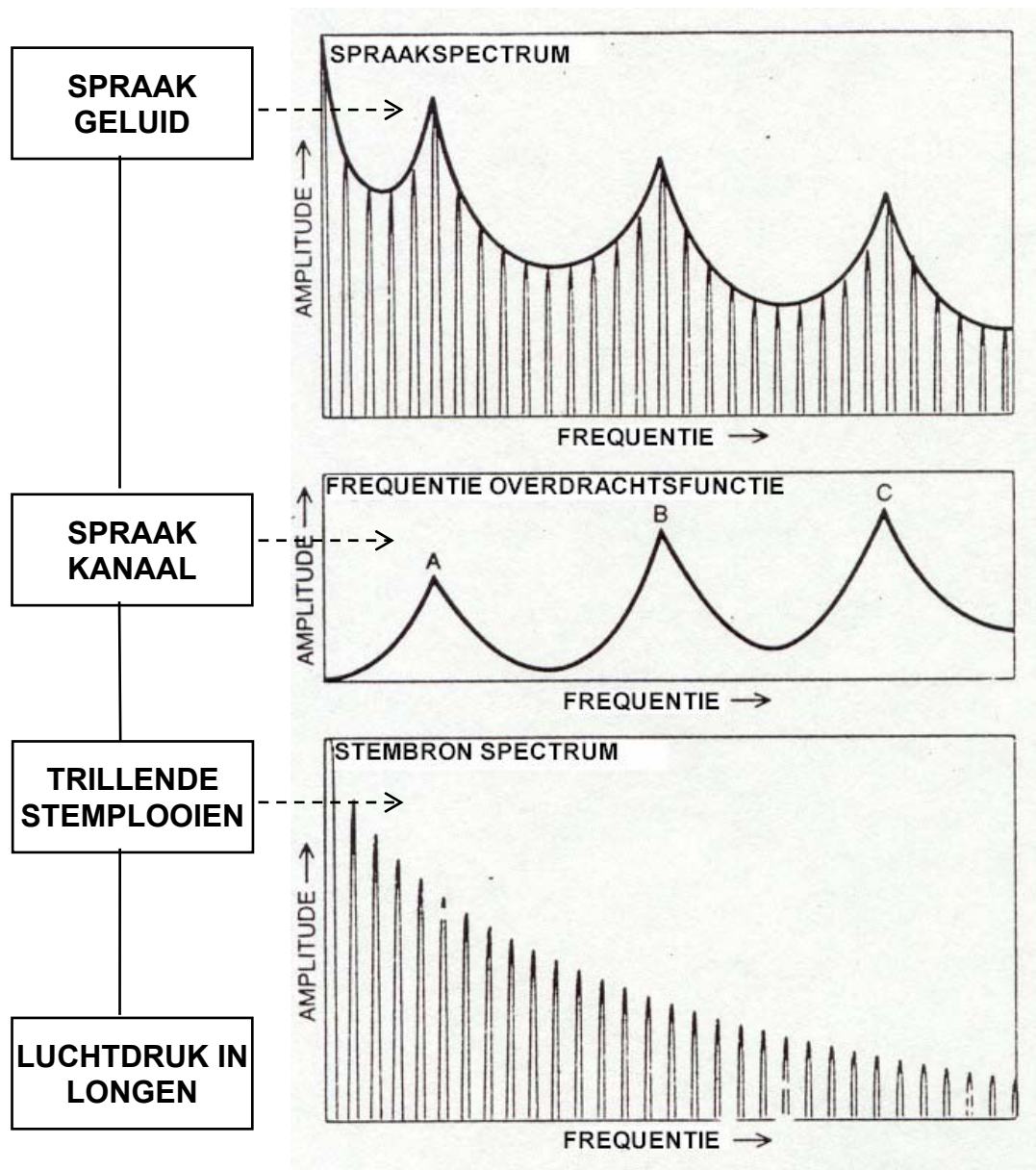
Figuur 1 geeft een schema van de belangrijkste fasen in de productie van spraak (opbouw van luchtdruk -> productie van brongeluid -> aanpassing van het brongeluid tot uiteindelijke spraakklank). We nemen aan dat de akoestische eigenschappen van de drie fasen elkaar niet beïnvloeden en afzonderlijk beschreven kunnen worden (Fant, 1961). Spraak is het eindresultaat van de drie fasen samen.

Om te beginnen vormen de longen de motor van de spraak door de opbouw van luchtdruk.

Vervolgens produceren in fase 2 de stemplooien het *stembrongeluid*, dat de basis vormt van al onze stemhebbende klanken. Wanneer de stemplooien regelmatig trillen bepaalt het aantal malen dat de stemplooien open en dicht gaan de *grondfrequentie*, die correspondeert met de toonhoogte van het geluid. Het geluidsspectrum van het brongeluid bestaat uit een reeks harmonischen die elk een veelvoud zijn van de grondtoon. In dit *bronspectrum* neemt de sterkte van elke harmonische met de frequentie af en wel zo dat het spectrum een aflopende helling heeft van -12 dB/octaaf.

In de fase 3 doorloopt het brongeluid vervolgens de keelholte en de mondholte (en soms de neusholte). Deze ruimte wordt ook wel *stemweg*, *aanzetstuk* of *spraakkanaal* genoemd. Het spraakkanaal vormt een akoestische resonator waarin bepaalde frequenties uit het stembrongeluid beter en andere frequenties juist slechter worden doorgegeven. We spreken ook wel van een akoestisch filter. Door articulatie kunnen we de filtereigenschappen van het spraakkanaal veranderen en daarmee de uiteindelijke spraakklank beïnvloeden. De akoestische filtereigenschappen van de verschillende vormen van het spraakkanaal staan in dit hoofdstuk centraal. De filtereigenschappen worden weergegeven in een *frequentieoverdrachts-functie* waaruit we kunnen aflezen welke frequenties door resonantie het beste worden doorgegeven (zie figuur 2).

De combinatie van stembrongeluid en de filterwerking van het spraakkanaal vormt het uiteindelijk spraakgeluid. Maar daarbij moet nog wel het gevolg van de akoestische afstraling van de mond verdisconteerd worden. Deze afstraling leidt tot een versterking van hogere frequenties volgens een helling van ongeveer + 6 dB/octaaf. De combinatie van een bronspectrum met een helling van -12 dB/octaaf en de afstraling van + 6 dB/octaaf leidt ertoe dat het spraakspectrum gemiddeld een helling heeft van - 6 dB/octaaf.



Figuur 1. De belangrijkste onderdelen van de spraakproductie met hun spectrale karakteristieken.

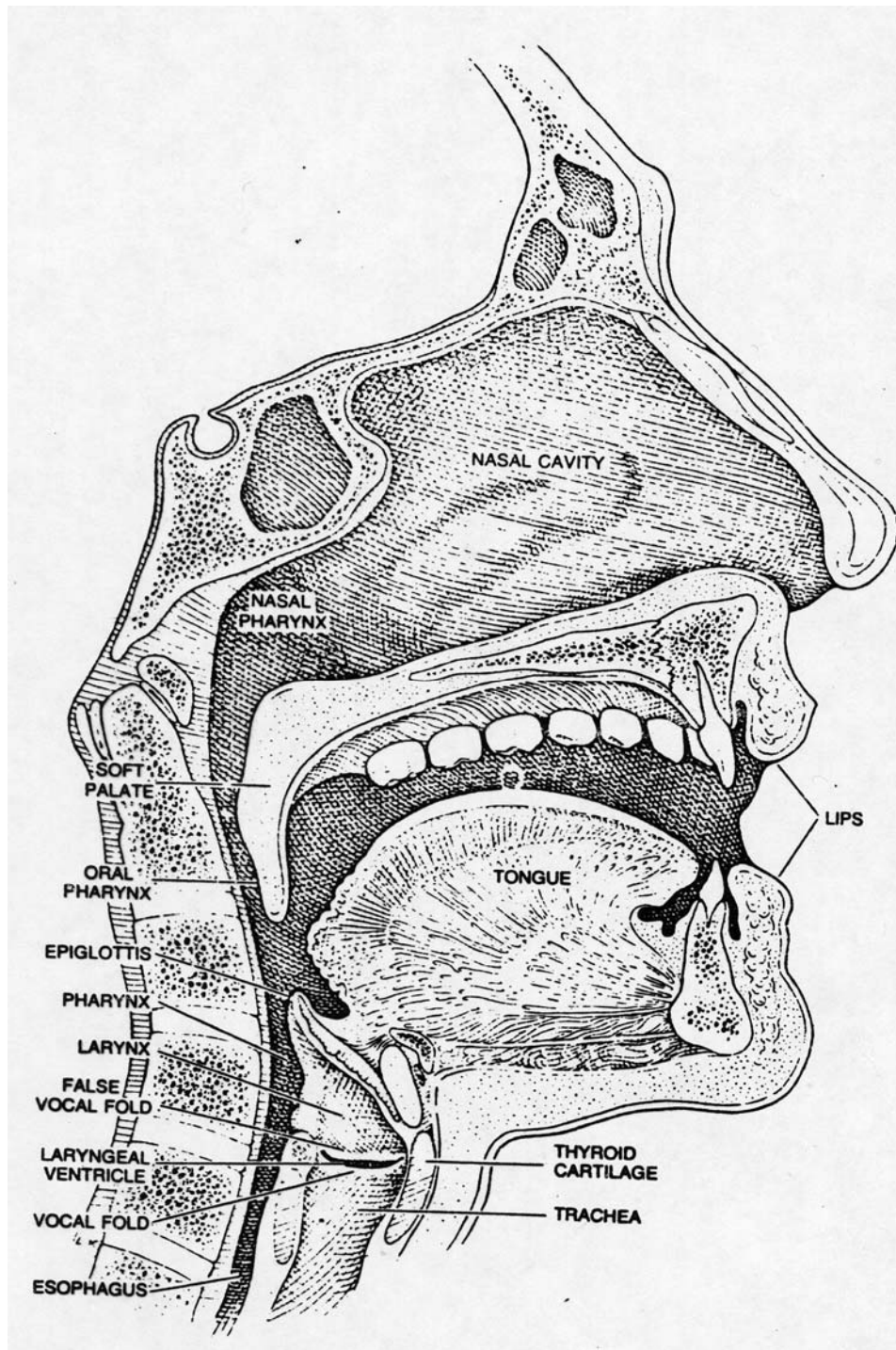
Naast de mogelijkheid om met de stemplooiën in het strottenhoofd lucht in trilling te brengen (voor alle stemhebbende spraakklanken) zijn er nog twee andere brongeluiden mogelijk:

- de luchtstroom wordt door een smalle doorgang tussen de stembanden geperst met als resultaat een turbulente luchtstroom (dit gebeurt bij gefluisterde spraakklanken), of
- de lucht wordt door een smal articulatiekanaal gestuwd, bijvoorbeeld tussen de omhoog gebrachte tong en het harde gehemelte, waardoor in dat articulatiekanaal turbulentie ontstaat (dat is het geval bij stemloze fricatieven).

2. Algemene vorm van keel-, mond- en neusholte; een beetje fysiologie

Laten we in vogelvlucht de holten van keel, mond en neus nader bekijken. Het gaat daarbij vooral om de afmetingen van de ruimtes, de mogelijke variatie in die afmetingen door middel van articulatie, waar we later de akoestische eigenschappen van zullen bespreken. De mond-keelholte is van stemplooien tot de lippen een flexibele buis met de bijzondere eigenschap dat de doorsnede van de buis op elk punt in de buis gevarieerd kan worden terwijl ook de lengte van de buis is te beïnvloeden. De buis heeft voor volwassenen gemiddeld een lengte van 17 cm. Voor vrouwen varieert deze lengte tussen 15 en 19 cm, voor mannen tussen 18 en 24 cm. Daarnaast kan de neusholte bijgeschakeld worden als een parallelle buis met vaste vorm.

Boven het strottenhoofd (*larynx*) bevindt zich de keelholte (*farynx*). Omdat het strottenhoofd met talrijke spieren los opgehangen is tussen het tongbeen (*hyoid*) en borstbeen (*sternum*) is de positie niet gefixeerd maar beweeglijk. Meestal is dat goed te zien als er van een lage naar hoge toonhoogte wordt gezongen. Het strottenhoofd heeft dan bij onge oefenden de neiging om een stukje te stijgen. Dat betekent dat de totale lengte van de stembuis enkele centimeters kan variëren al naar gelang de strottenhoofdpositie. Bij een hoge strottenhoofdpositie zal bovendien het volume van de keelholte in het algemeen kleiner zijn dan bij een lage strottenhoofdpositie. Vlak boven de stemplooien zitten de valse stemplooien als een vernauwing in de nog smalle laryngale buis (doorsnede $\sim 2 \text{ cm}^2$). Deze buis kan worden afgesloten met de strottenklep (*epiglottis*) om te voorkomen dat voedsel in de luchtweg komt. Achter de laryngale buis zit de ingang van de slokdarm (*oesophagus*). Deze zit wat lager dan de bovenzijde van de laryngale buis en de ruimte boven de slokdarmopening heet de *sinus piriformis*. Hoe klein deze ruimtes in de onderste keelholte ook zijn, ze hebben allen een specifieke invloed op de geluidsoverdracht van het spraakkanaal. Het bovenste gedeelte van de keelholte wordt aan de voorzijde begrensd en beïnvloed door de tongwortel, aan de bovenzijde door het zachte gehemelte. De tongwortel is het achterste gedeelte van de tong en kan ver naar achteren gedrukt worden en de keelholte bijna afsluiten, bijvoorbeeld bij de klinker /a/. Omgekeerd kan de tongwortel naar voren gaan en een ruime keelholte creëren, bijvoorbeeld bij de klinker /i/. De posities van het strottenhoofd en de tongwortel geven veel mogelijkheden tot volumevariatie in de keelholte. Meestal zijn we ons dat niet zo bewust. Toch kan de vorm van keelholte een grote invloed hebben op de klankkleur van spraak en zang.



Dwarsdoorsnede van de stemweg

De mondholte wordt van boven begrensd door het zachte en harde gehemelte. De zijkant en onderkant worden gevormd door de wangen en de tongrug terwijl van voren de tanden en lippen de mondholte afsluiten. De tong, onderkaak en lippen zorgen voor een grote vrijheid in de vormgeving van de mondholte. Door het omlaag en omhoog brengen van de onderkaak zorgen we voor een algemene vergroting of verkleining van de mondholte. Met de tongpunt en de tongrug kan de mondholte op verschillende punten (bijna) gesloten worden. Daarmee wordt de mondholte ruwweg in twee gedeelten verdeeld. De lippen bepalen tenslotte de grootte en vorm van de opening naar buiten. Dat kan door tuiten een ronde vorm zijn of door spreiding van de lippen een horizontale spleetvorm. Omdat getuite lippen (/y/) enigszins naar

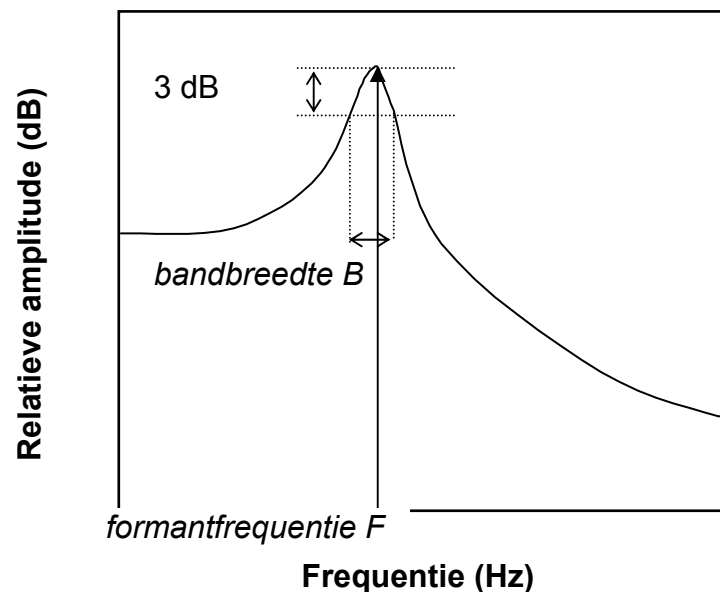
voren steken en in gespreide stand (/e/) iets meer teruggetrokken zijn, heeft de vorm van de lippen een kleine invloed op de totale lengte van het spraakkanaal.

De neusholte kan op de mond-keelholte worden aangesloten door het zachte verhemelte (*velum*) te laten zakken. De vorm van de neusholte is niet veranderbaar. We kunnen eenvoudig controleren of de neusholte bij de klankvorming een rol speelt door tijdens het spreken of zingen de neus open en dicht te drukken. Als we geen verandering van klankkleur waarnemen is het velum gesloten. Horen we wel een verschil dan is het velum geopend. De neusholte staat via kleine kanaaltjes in verbinding met de voorhoofdsholten. Omdat deze kanaaltjes heel klein zijn spelen de voorhoofdholtes akoestisch echter geen enkele rol.

De ruimtes beneden het strottenhoofd zijn de luchtpijp (*trachea*) en bronchiën met de steeds fijnere vertakkingen tot in de longblaasjes. Deze ruimtes hebben weinig invloed op de spraakklank. Alleen onder bijzondere omstandigheden is bij zingen op een bepaalde toonhoogte de (destabiliserende) invloed op de klank merkbaar.

3. Resonanties

De akoestische eigenschappen van het spraakkanaal worden louter en alleen door het volume, de vorm en de wandeigenschappen bepaald. Deze bepalen de *resonanties* van het spraakkanaal en ze worden *formanten* genoemd. De resonantiefrequentie heet voor spraak daarom *formantfrequentie*. Resonanties ontstaan wanneer er een goede afstemming is tussen een geluidsgolf en de dimensies van de ruimte. In dat geval blijft de energie in de geluidsgolf goed behouden. Het principe kan vergeleken worden met een schommel. Als op het juiste moment (die samenhangt met de lengte van de schommel) een duwtje wordt gegeven, dan zal de schommel uiteindelijk hoog opzwaaien. Je kan zeggen dat de opeenvolgende duwtjes en de schommel dan goed bij elkaar passen en elkaar versterken. Als daarentegen op verkeerde momenten tegen de schommel wordt geduwd dan komt de schommel niet hoog. Bij spraak ontstaan de duwtjes door de klapjes van de stemplooien. Als (één van de) geluidsgolven die door de stemplooibeweging ontstaan goed past bij de dimensies van (een deel van) het spraakkanaal, dan wordt deze geresoneerd door stapeling van energie. Als bij spraak een geluidsgolf niet goed overeenstemt met een resonantie dan verlaat die golf verzwakt de mond. In het algemeen is het zo dat hoe kleiner de betrokken ruimte is, hoe hoger de resonantiefrequentie is.



Figuur 2. Frequentieoverdrachtsfunctie van een formant, met formantfrequentie F en de bandbreedte B die wordt gemeten op 3 dB onder de formantpiek.

Hoe goed een geluidsgolf met een bepaalde frequentie zich kan handhaven in de spraakbuis wordt weergegeven in de *frequentieoverdrachtsfunctie*. Figuur 2 geeft een voorbeeld van de overdrachtsfunctie van een enkele resonantie. We zien dat bij de *formantfrequentie* F de energie het best behouden blijft en dat hogere en lagere frequenties veel meer gedempt worden. De mate van *damping* wordt uitgedrukt door de breedte van de piek in de frequentieoverdrachtsfunctie. De breedte wordt bepaald bij de overdracht die 3 dB zwakker is dan bij de piek. Dit is de *bandbreedte* B van de resonantie. Bij spraak neemt de bandbreedte van een formant toe van zo'n 50-100 voor F_1 tot 200-300 Hz voor F_4 en F_5 .

Formanten spelen een centrale rol in de beschrijving van spraakklanken, niet alleen omdat we er het spraakspectrum mee kunnen vastleggen, maar vooral omdat formantfrequenties bepalend zijn bij de *waarneming* van spraakklanken.

De eenvoudigste voorstelling van het spraakkanaal is die van een halfopen rechte buis. Zo'n buis heeft meerdere resonanties en die liggen in het voor spraak belangrijke gebied tot 5000 Hz bij 500, 1500, 2500, 3500 en 4500 Hz. Hun frequenties geven we aan met F_1 , F_2 , F_3 , F_4 en F_5 . Ook bij andere vormen van de keel-mondholte (en de neusholte) is het aantal formanten meestal vijf. Hun precieze frequentie hangt af van de vorm van het spraakkanaal.

Elke verandering in de vorm van het spraakkanaal leidt in principe tot een verandering in de formantfrequenties. Soms zijn er overigens kleine articulatorische veranderingen die grote akoestische effecten hebben, terwijl er ook articulatorische veranderingen kunnen zijn die maar kleine effecten hebben. Alhoewel de vorm van het spraakkanaal de formantfrequenties precies bepaalt, is het omgekeerd niet zo dat we uit spraakgeluid de vorm van het spraakkanaal precies kunnen vaststellen. Er kan worden aangetoond dat meerdere vormen van

het spraakkanaal tot een zelfde spraakspectrum met dezelfde vijf formanten aanleiding kunnen geven.

Alhoewel we voorzichtig moeten zijn om formanten in direct verband te brengen met de vorm van bepaalde delen van het spraakkanaal, dus de articulatie, kan het toch nuttig zijn om een aantal globale relaties te bespreken.

Lengte: Een belangrijke algemene variabele van het spraakkanaal is de lengte. Als we het spraakkanaal zien als een halfopen buis dan zal een verlenging van de buis een evenredige verlaging van alle resonantiefrequenties tot gevolg hebben en een verkorting een evenredige verhoging.

Mondopening: De eerste formant is erg gevoelig voor mate van mondopening. Toch kunnen we de eerste formant articulatorisch waarschijnlijk beter associëren met de kaakopening, ofwel de stand van de onderkaak. (Lindblom en Sundberg, 1971). Het is immers moeilijk om alleen de mondopening te veranderen. Het vergroten van de mondopening gaat gepaard met een verlaging van de stand van de onderkaak. En dat laatste heeft ook weer gevolgen voor bijvoorbeeld het volume van de mondholte, welke groter zal worden. Het gecombineerde effect blijkt vooral te leiden tot een verhoging van de frequentie van de eerste formant. F1 varieert tussen 200 Hz (gesloten kaak) en 800 Hz (open kaak).

Plaats van articulatie: De tweede formant wordt gewoonlijk geassocieerd met de plaats waar de tong een vernauwing in het spraakkanaal creëert (plaats van articulatie). Deze varieert van voor tegen het harde gehemelte bij /i/ tot achter tegen het zachte gehemelte bij /u/. De F2 verandert hierbij van zo'n 2000 Hz (voor) tot 800 Hz (achter). Daarnaast is de tweede formant veel gevoeliger dan andere formanten voor de doorsnede van de vernauwing. Een zeer nauwe opening kan zeker een hoge F2 aanzienlijk verlagen. De holte waarmee de tweede formant geassocieerd wordt hangt af van de plaats van articulatie. We moeten ons daarbij realiseren dat de tong niet alleen de plaats van de vernauwing bepaalt, maar dat de tong gelijktijdig de grootte van de holtes voor en na de afsluiting beïnvloedt. De achterste holte is het grootst bij een plaats van articulatie voor in de mond (/i/) en het kleinst bij een plaats van articulatie achter in de mond (/u/). Het omgekeerde geldt uiteraard voor de holte voor in de mond. De tweede formant kan een resonantie zijn van de voorste holte tussen de vernauwing en de lipopening, maar ook de resonantie van de holte achter de vernauwing. We moeten dus voorzichtig zijn bij het interpreteren van een formant in termen van een resonantieholte.

Naarmate de frequenties van de formanten toenemen zijn er steeds kleinere holtes mee gemoeid. Die kunnen gevonden worden in de keelholte, als de smalle doorgang bij de tongvernauwing of in het gebied rond tanden en lippen. Hun rol zal altijd bekeken moeten worden in relatie tot de effecten van grotere articulaties zoals kaakopening en plaats van articulatie.

Liprondding: Het ronden van de lippen creëert een verlenging van het spraakkanaal en verlaagt alle formantfrequenties. In vergelijking met gespreide lippen wordt bij geronde lippen in het bijzonder echter de holte tussen lippen en tanden verlengd, wat met name de derde formant verlaagt. De F3 varieert tussen 1500 en 3000 Hz.

Positie van het strottenhoofd: Ook het verlagen van het strottenhoofd verlengt de spraakbuis en verlaagt alle formantfrequenties. De grootte van de strottenhoofdbuis vlak boven de stemplooien en de diameter van de keelholte beïnvloeden de frequentie van de vierde en

vijfde formant (Sundberg, 1974). De vierde formant varieert tussen 2000 en 3500 Hz, de vijfde tussen 2500 en 4500 Hz.

5. De relatie tussen articulatie, spraakklank en formanten

We hebben gezien dat door articulatie van de mond-keelholte de akoestische eigenschappen ervan zeer kunnen variëren. Het aantal klanken wat we daarom kunnen maken is in principe enorm groot. Toch is het aantal *spraakklanken* (of fonemen) in een taal beperkt. In totaal onderscheiden we voor het Nederlands zo'n 44 spraakklanken, waaronder 12 *monoftonge* klinkers en 4 tweeklanken (*diftongen*). Het aantal spraakklanken kan per taal verschillen. Het zijn dan ook niet de eigenschappen van articulatie die de spraakklanken definiëren maar de keuzen die in een taal worden gemaakt. Er is geen reden om aan te nemen dat we niet in staat zouden zijn om alle spraakklanken die in alle talen van de wereld worden gebruikt te kunnen produceren.

De eigenschappen van een bepaalde spraakklank liggen niet precies vast. We weten allemaal hoe we ongeveer moeten articuleren om een /a/ klank te maken. Het is echter mogelijk om kleine veranderingen in de articulatie aan te brengen zonder dat het karakter van de /a/ verloren gaat. Dat we de klank als een /a/ blijven horen heeft alles te maken met onze *waarneming* van spraak. Bij elke verandering van articulatie zullen de akoestische eigenschappen van het geluid veranderen. Dat we dat niet altijd als een verandering van spraakklank opvatten zegt iets over de tolerantie in ons spraak*verstaan*. We maken in onze waarneming een onderscheid tussen een spraakklank en zoiets als een 'aanvullende' klankkleur (of timbre). We spreken van een lichte /a/, een scherpe /a/, een donkere /a/, enzovoorts. Akoestisch gezien is dit onderscheid tussen klinker en 'aanvullende' klankkleur echter kunstmatig want de klankeigenschappen vormen een continuüm, zonder specifieke markeringen.

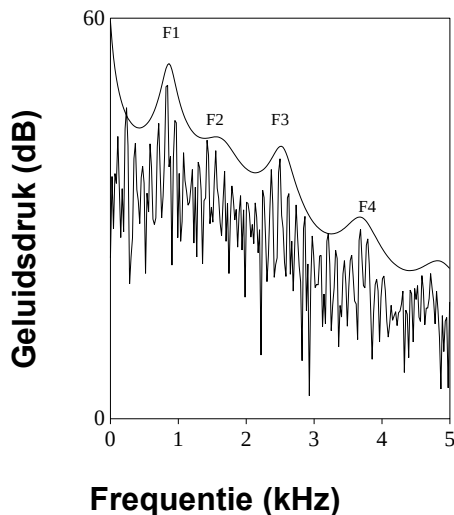
Bij het beschrijven van spraakklanken moeten we altijd uitgaan van de menselijke waarneming. We kunnen niet zeggen dat een bepaalde articulatie de klank 'aa' definieert. Wel kunnen we de articulatie en akoestische eigenschappen van een klank bestuderen die een grote groep luisteraars als een /a/ beschouwt. De akoestische eigenschappen blijken dan sterk afhankelijk te zijn van de waarden van de formantfrequenties. Onze geluidswaarneming is in het algemeen gevoelig voor geluiden die eruit springen en geluiden met sterke resonanties voldoen daaraan. Als we de frequenties van de vijf formanten meten voor heel veel spraakklanken, dan blijkt dat er vooral systematische relaties tussen spraakklank en formantfrequentie wordt gevonden voor de eerste twee formanten (en soms voor de derde formant).

6. De bepaling van formantfrequenties

We kunnen een spraakklank met een microfoon opnemen. Een microfoon zet de drukvariëaties in de geluidsgolf om in een gelijkvormig elektrisch signaal. Dit signaal kunnen we dan op een audioband vastleggen. We kunnen het oorspronkelijke geluid via een versterker en luidspreker weer hoorbaar maken, maar analyseren kunnen we het niet. Daarvoor moeten we het geluidssignaal in getallen vastleggen. Dat kan door de grootte van het signaal zeer vaak per seconde te meten (digitale opname). Dit noemen we *bemonsteren* (samen). De bemonsterfrequentie voor digitale opnamen op een DAT-recorder of op een compact disk is 44.100 keer per seconde. Hebben we spraak eenmaal op deze manier vastgelegd dan kunnen we er met de computer verder aan rekenen. We kunnen bijvoorbeeld het spectrum uitrekenen met behulp van *Fouriertransformatie* (*FFT*, *fast Fourier transform*). Dat doen we meestal over een kort stukje spraak van zo'n 25 ms waarvan we veronderstellen dat de

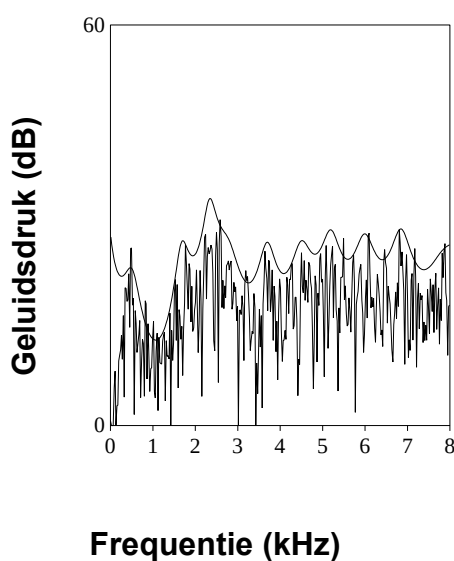
akoestische eigenschappen constant zijn. Het spectrum geeft de sterkte van elke frequentiecomponent in het signaal (tot een bovengrens die de helft is van de oorspronkelijke bemonsteringsfrequentie).

Grafisch staat in een spectrum horizontaal altijd de frequentie uit en verticaal de sterkte (als geluidsdruk). Als het spraaksignaal periodiek stemhebbend was dan zien we in het spectrum een duidelijke reeks pieken op gelijke frequentieafstand (de harmonischen of boventonen). Zie bijvoorbeeld het onderstaande spectrum van de klinker /a/.



Spectrum van een /a/

Als het spraaksignaal ruisig was (stemloos) dan zien we geen pieken maar een geheel gevuld patroon. Zie het onderstaande spectrum van een /f/.



Spectrum van een /f/

Het spectrum geeft precies aan wat de karakteristieken zijn van het spraakgeluid. Uit het spectrum zijn echter de formanten niet automatisch afleidbaar. De *effecten* van formanten drukken zich uit in het spectrum doordat harmonischen in de buurt van een formantfrequentie krachtiger zijn dan de harmonischen die juist tussen twee formantfrequenties in liggen. Dat betekent dat we uit het sterkteverloop van de opeenvolgende harmonischen, ofwel de *omhullende* van het spectrum, de formantfrequenties moeten *schatten*. Dat is bij een rijk gevuld spectrum, behorend bij een spraakklank met een lage grondfrequentie en dus met veel harmonischen, niet zo moeilijk. Bij hogere toonhoogten hebben we echter veel minder harmonischen die een indicatie geven voor de formantfrequentie en wordt de schatting ervan moeilijker zo niet onmogelijk. Formantfrequenties laten zich daarom gemakkelijker schatten bij mannenstemmen dan bij vrouwenstemmen omdat mannen op lagere toonhoogte spreken. Bij kinderstemmen en bij zang is het zelfs vaak onmogelijk om tot een goede schatting van formantfrequenties te komen.

Een veelgebruikte analyse om formanten te schatten is *linear predictive coding* of LPC. Deze techniek gaat uit van een model van de akoestische overdracht door het spraakkanaal (zie figuur 1) en probeert het resultaat van het model zo goed mogelijk in overeenstemming te brengen met het te analyseren geluid. Met de resultaten van de LPC analyse kunnen de formantfrequenties en bandbreedtes worden *geschat*. Die schatting kan er vooral bij hogere grondfrequenties soms flink naast zitten (zie figuur 12). De LPC resultaten kunnen worden weergegeven als de geschatte omhullende van het spectrum. Door het FFT spectrum met het LPC spectrum te vergelijken krijgen we een indruk van de betrouwbaarheid van de LPC resultaten.

Een enkel spectrum geeft de eigenschappen van een kort moment gedurende de spraakklank. Vaak is het belangrijk om zichtbaar te maken hoe de spectrale eigenschappen veranderen binnen een spraakklank of tussen spraakklanken. Het is grafisch lastig om naast frequenties en geluidsdruk ook nog de tijddimensie weer te geven. In een *spectrogram* is dat opgelost door horizontaal de tijd uit te zetten en verticaal de frequentie (onderaan de laag, bovenaan hoog, zie bijvoorbeeld figuur 7). In een stationaire periodiek klank vormen de harmonischen dan parallelle horizontale lijnen. De geluidsdruk wordt door middel van zwarting aangegeven: hoe zwarter, hoe hoger de geluidsdruk bij de betreffende frequentie. De zwartste gebieden liggen in de buurt van formantfrequenties. Formantverschuivingen binnen en tussen klanken zijn goed herkenbaar door de verschuiving van de zwarte gebieden.

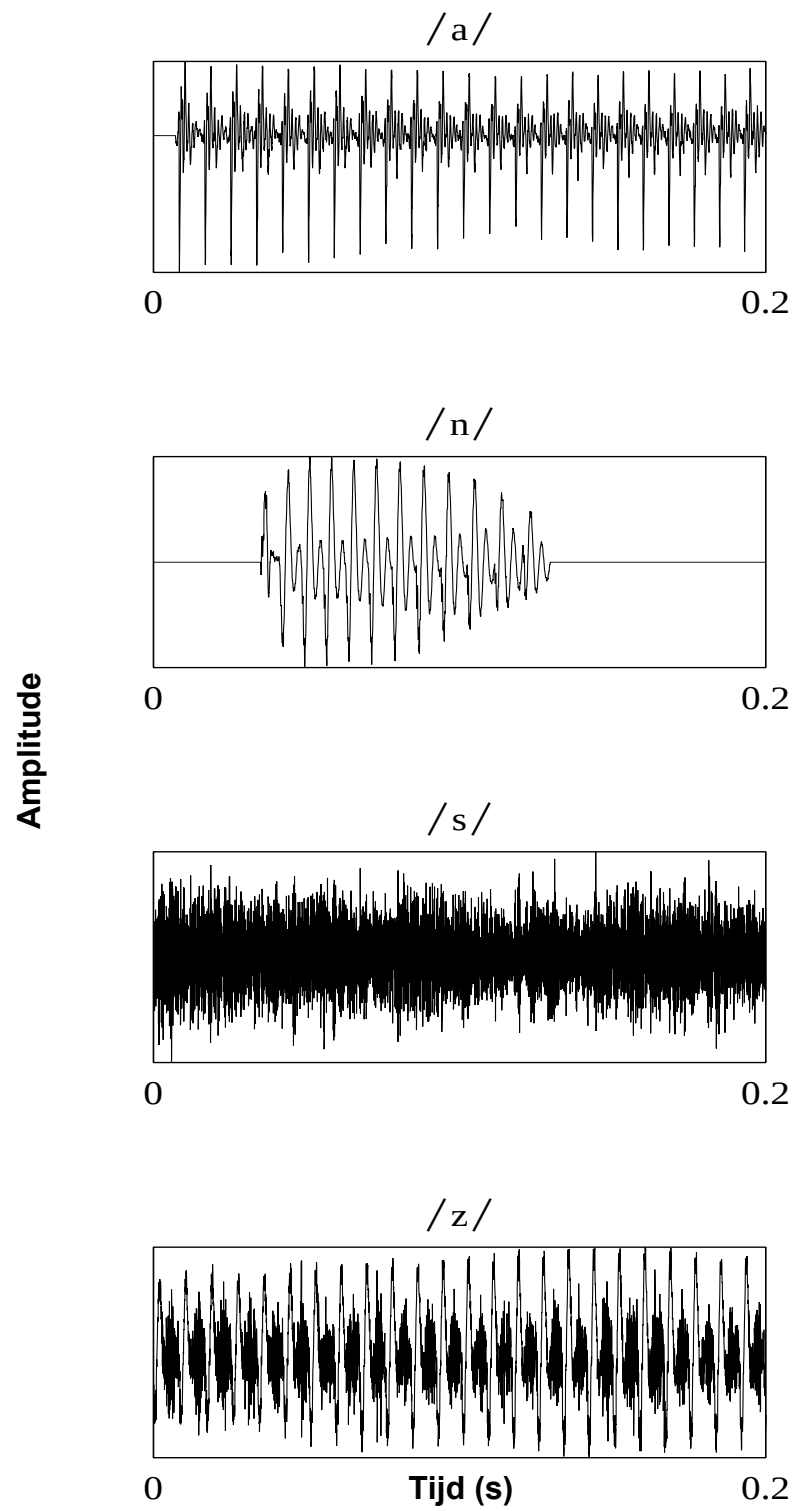
Een spectrogram wordt gemaakt door met een bandfilter de energie in een frequentie gebiedje te meten, en op die manier de frequenties van laag naar hoog te doorlopen. Als het bandfilter heel smal is zijn in het spectrogram de afzonderlijke harmonischen goed zichtbaar. We spreken van een smalle band spectrogram. Als het bandfilter breed is dan zijn de afzonderlijke harmonischen niet meer zichtbaar en zijn brede zwarte gebieden een indicatie voor de aanwezigheid van formanten. Spectrale veranderingen in de tijd zijn in breedband spectrogrammen beter te zien dan in smalle band spectrogrammen. In dit hoofdstuk geven we alleen smalle band spectrogrammen.

7. De akoestische eigenschappen van de Nederlandse spraakklanken

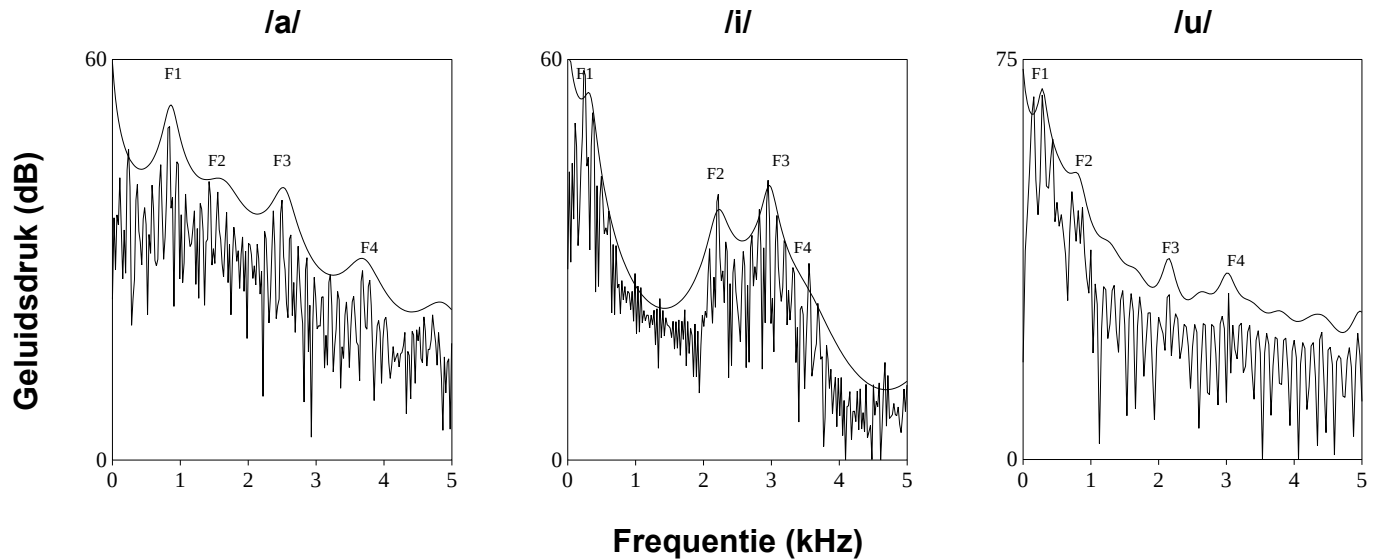
In de volgende paragrafen behandelen we de akoestische eigenschappen van de Nederlandse spraakklanken. We behandelen achtereenvolgens *klinkers* (waaronder *tweeklanken*, *nasale klinkers*) en *medeklinkers* (waaronder *plosieven*, *wrijfklanken*, *liquidae*, *nasale medeklinkers* en *halfklinkers*). Van veel spraakklanken laten we een voorbeeld zien van het geluidssignaal (*golfvorm* of *oscillogram*) zelf, het smalle band spectrogram, het gedetailleerde spectrum en het omhullende spectrum. Wanneer er in de tekst formantfrequenties worden gegeven, dan zijn die indicatief voor volwassen mannen.

7.1. Klinkers

Klinkers ontstaan bij stabiele, periodieke stemgeving en een stabiele articulatie. Het periodieke karakter is goed te zien in de golfvorm. Figuur 3 geeft o.a. de golfvorm van de klinker /a/. We zien een patroon dat zich heel regelmatig herhaalt. Elke periode correspondeert met het éénmaal sluiten en openen van de stemplooien. In het spectrum van een klinker zijn de waarde van de eerste en tweede formant meestal bepalend voor de waargenomen klinker. Figuur 4 geeft de spectra van de klinkers /a/, /i/ en /u/. De spectra tonen de sterk verschillende formantpatronen die met de LPC spectra redelijk omhuld worden. Voor /a/ is $F1 = 860$ Hz en $F2 = 1550$ Hz, voor /i/ is $F1 = 300$ Hz en $F2 = 2220$ Hz, en voor /u/ is $F1 = 300$ Hz en $F2 = 790$ Hz.

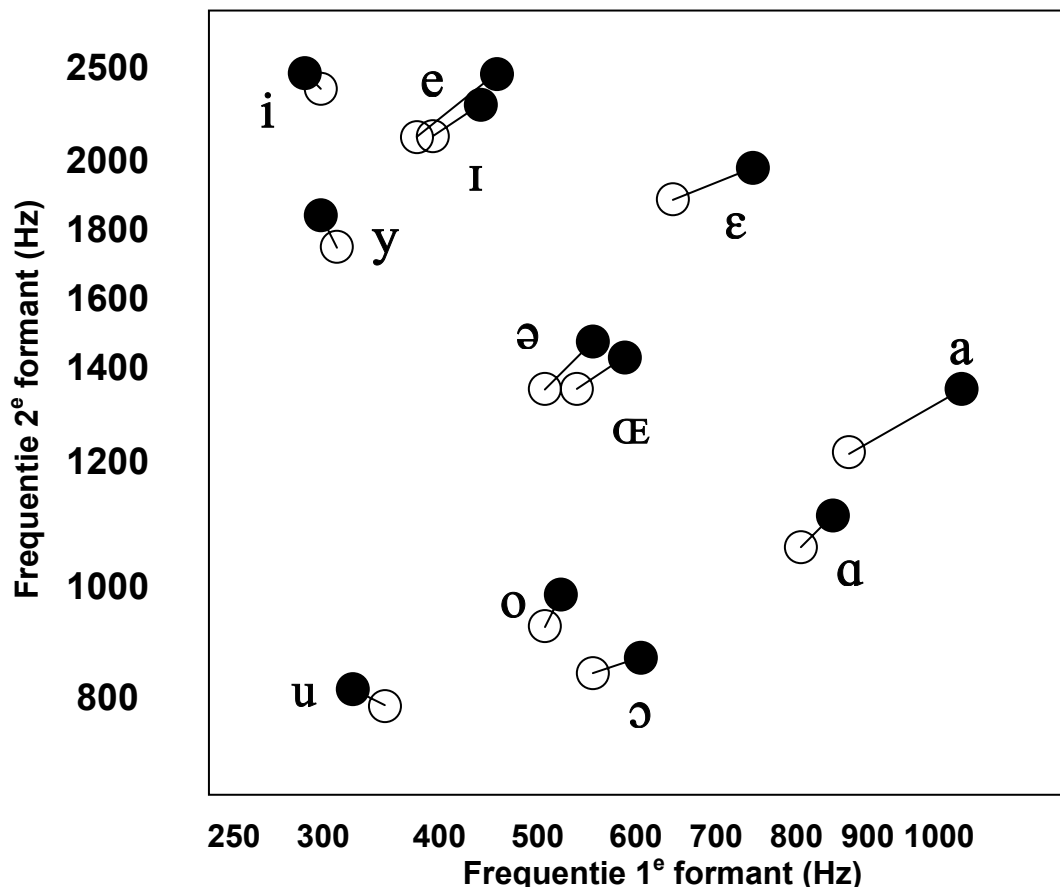


Figuur 3. Golfvormen van de klinker /a/, de nasaal /n/, de stemloze fricatief /s/ en de stemhebbende fricatief /z/.



Figuur 4. Spectra van de klinkers /a/, /i/ en /u/ met daarin aangegeven de posities van de eerste vier formanten.

In figuur 5 zijn de gemiddelde waarden van frequenties van de eerste en tweede formant van Nederlandse klinkers voor mannen en voor vrouwen tegen elkaar uitgezet. De waarden zijn afkomstig van 50 mannen en 25 vrouwen (Klein et al, 1970; van Nierop et al., 1973). We noemen een weergave zoals in figuur 5 het formantvlak. De klinkers vormen daarin ruwweg een driehoek, de *klinkerdriehoek*. De klinkers op de hoekpunten, /a/, /i/ en /u/, noemen we de kardinale klinkers. Dat de klinkers een begrensd gebied bestrijken komt door de fysiologische beperkingen die aan articulatie worden gesteld.



Figuur 5. *Klinkerdriehoek van Nederlandse klinkers. Gevulde cirkels zijn de gemiddelde waarde voor 25 vrouwen (van Nierop et al., 1973), de open cirkels zijn de gemiddelde waarde voor 50 mannen (Klein et al., 1970).*

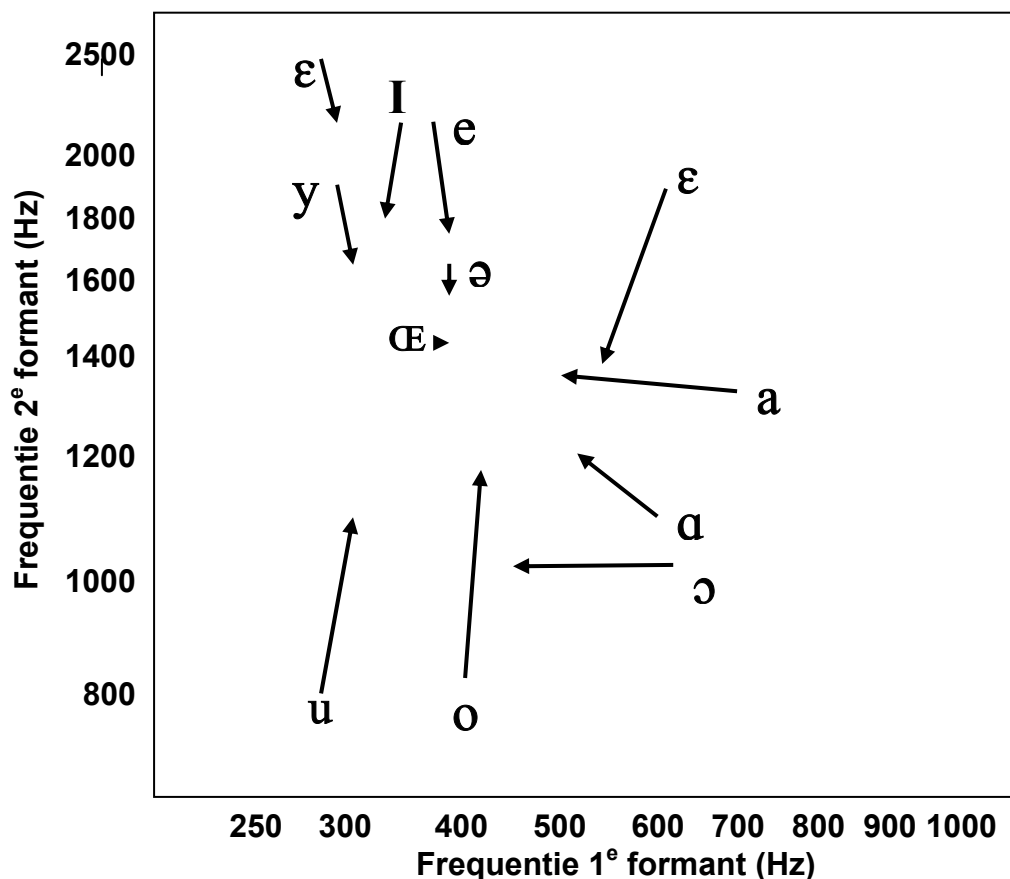
De as waarlangs de frequentie van de eerste formant is uitgezet kunnen we articulatorisch interpreteren als *gesloten* (/u/) tot *open* (/a/), in relatie tot de kaakopening. Langs de as met de waarde van de tweede formantfrequentie zien we verandering in de plaats van articulatie, van een afsluiting met de tongpunt *achter* in de mond (/u/), tot *voorin* (/i/). De neutrale articulatie van de sjwa (/ə/) bevindt zich midden in de driehoek, ongeveer bij de waarden 500 Hz en 1500 Hz. Dat zijn de waarden van de eerste twee resonanties in een halfopen buis.

De verschillen tussen de klinkers van mannen en vrouwen zijn het gevolg van de anatomische verschillen tussen mannen en vrouwen in het spraakkanaal. Het belangrijkste verschil zit in de totale lengte van de buis. Deze is voor vrouwen korter dan voor mannen. Voor vrouwen varieert deze lengte tussen 15 en 19 cm, voor mannen tussen 18 en 24 cm. Bij zangers wordt de kortste buis meestal gevonden bij sopranen en tenoren, de langste bij alten en bassen. Daarnaast zijn er natuurlijk nog belangrijke verschillen tussen mannen en vrouwen in de bouw van het strottenhoofd en stemplooien die tot uitdrukking komen in verschillen in spreekstemtoonhoogte en toonhoogtebereik.

Eerder hebben we uitgerekend dat bij een rechte buis van 17 cm lengte de resonanties gevonden worden bij 500, 1500, 2500, 3500, 4500,... Hz. Uit dat type berekening volgt dat bij een kortere buis de resonantiefrequenties hoger liggen, bij een langere buis lager. Voor een sopraan met een totale lengte van het spraakkanaal van 15 cm liggen de resonanties bij 567, 1701, 2835, 3969, 5103,... Hz. Voor een bas met een buislengte van 24 cm liggen ze bij 354, 1062, 1770, 2478, 3186,... Hz. Dit zijn echter de extremen. Gemiddeld zullen bij neutrale articulatie

de formantfrequenties van mannen ongeveer 10-20% lager liggen dan bij vrouwen. Uit figuur 5 zien we overigens dat de realiteit complexer is. Voor de meeste klinkers zijn zowel F1 als F2 van mannen inderdaad lager, maar dat geldt niet voor de gesloten klinkers /i/, /y/ en /u/ waarvoor F1 juist iets hoger is. Bij kinderen is het spraakkanaal nog veel korter dan bij vrouwen en liggen de formantfrequenties gemiddeld systematisch hoger.

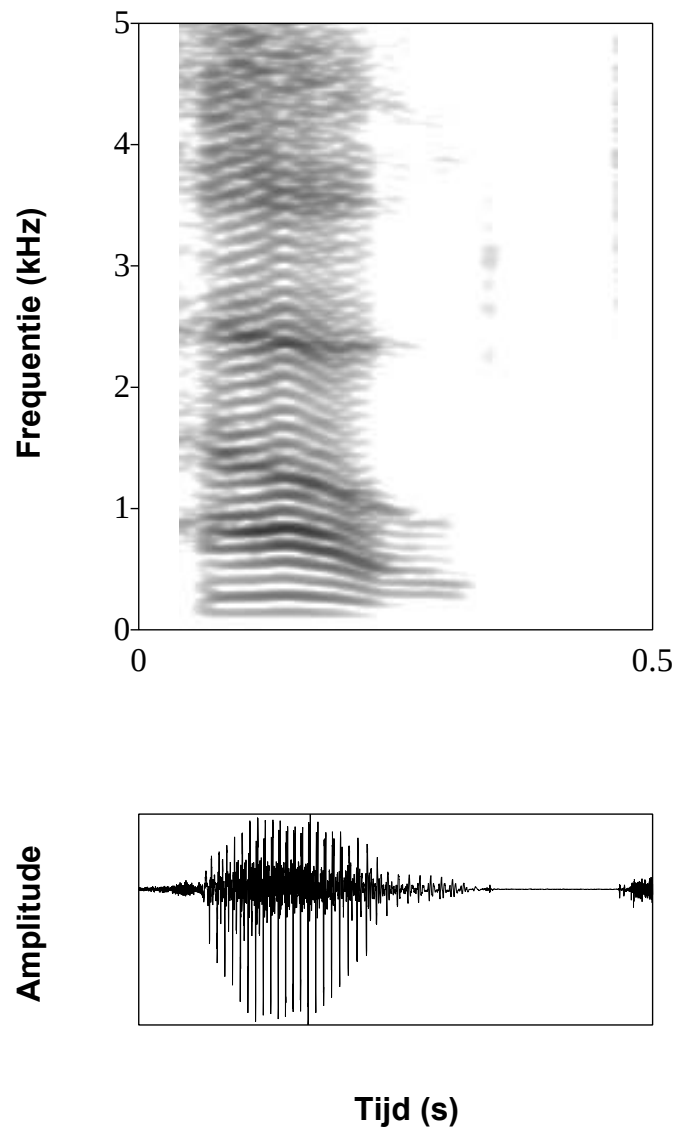
De tot nu toe besproken waarden van formantfrequenties van klinkers gelden wanneer de klinkers los worden uitgesproken. De articulatie is dan het meest precies. In lopende spraak is dat lang niet altijd het geval. Door de eisen die voorgaande of volgende spraakklanken aan de articulatie stellen en door de traagheid van onze articulatie is de realisatie van klinkers in lopende spraak minder nauwkeurig. Wanneer we daarbij ook nog snel moeten spreken dan neemt deze articulatorische precisie nog verder af. We noemen dit klinkerreductie. Ook onder die omstandigheden kunnen we de formantfrequenties bepalen en dan blijkt dat de formanten meestal veranderen in de richting van de sjwa, ofwel die van de neutrale buis. Dit wordt fraai gedemonstreerd in onderzoek van Koopmans-van Beinum (1980), waarin een proefpersoon klinkers sprak in allerlei contexten: los, in losse woorden, in voorgelezen tekst en in conversatie. Figuur 6 laat zien dat de formanten steeds meer verschuiven naar het centrum van de klinkerdriehoek. Als we gereduceerde klinkerrealisaties uit lopende spraak knippen en aan luisteraars ter beoordeling aanbieden dan blijkt dat de luisteraars de bedoelde klinker vaak niet meer herkennen. In lopende spraak wordt de klinkeridentiteit echter ook afgeleid uit de overgang van en naar naburige spraakklanken. Deze overgangen wijzen al naar de klinker die komt of naar de klinker die geweest is. En daar hebben we al voldoende aan. Daarnaast hebben we vaak verwachtingen ten aanzien van het gezegde. Op die manier kunnen we zelfs woorden herkennen waarin sommige klinkers nauwelijks gerealiseerd zijn.



Figuur 6. Klinkerreductie voor een mannelijke spreker naar Koopmans-van Beinum (1980). Het pijlbegin zijn de formantwaarden voor los uitgesproken klinkers, langs de pijl liggen de waarden voor achtereenvolgens klinkers in losse woorden, klinkers in teksten en aan het pijleinde de waarde voor klinkers in conversatie.

7.2. Tweeklanken

Het Nederlands kent drie echte tweeklanken, de /εI/ van ei, de /œy/ van ui, en de /au/ van au. Deze tweeklanken worden gevormd door een articulatorische verandering tussen twee klinkerposities. Dit leidt tot een verglijding van formanten van een start- naar een eindwaarde. Zoals de fonetische symbolen van de tweeklanken al uitdrukken verglijden we bij ei van /ε/ naar /I/, bij ui van /œ/ naar /y/ en bij au van /a/ naar /u/. Figuur 7 toont dit in het spectrogram van /εI/. Omdat tweeklanken zich taalkundig als klinkers gedragen worden ze hiertoe gerekend. De langzamere klinkercombinaties zoals in *aai*, *ooi* en *oei* worden niet tot de echte tweeklanken gerekend omdat de samenstellende klinkers nog als stabiele stukken in het spraaksignaal aanwezig zijn wat bij de echte tweeklanken niet het geval is.



Figuur 7. *Spectrogram en oscillogram van de tweeklank /ɛɪ/ gesproken in het woord heit. Merk de continue verandering in het formantverloop op.*

7.3. Nasale klinkers

Wanneer sprekers bij de realisatie van een klinker het velum laten zakken en daarmee een verbinding met de neusholte maken, dan spreken we van genasaleerde klinkers. Het is niet zo dat om een klank ‘nasaal’ te laten klinken altijd dezelfde mate van velumbeweging noodzakelijk is. Relevant bij nasaleren van spraakklanken is de ratio van de doorsneden van de toegang tot de neusholte en die tot de mondholte. Lage klinkers zoals de /a/ hebben een relatief grotere toegang tot de mondholte dan hoge klinkers zoals de /i/. Om die reden zal voor het produceren van een nasale /a/ een grotere neerwaartse velumbeweging noodzakelijk zijn dan om een nasale /i/ te maken (ongeveer 3 maal, cf. Clumeck, 1976).

Akoestisch gezien zijn er met een geopend velum twee kanalen die invloed op het geluid hebben: de mondholte en neusholte. De neusholte is een resonator met nasale formanten. Hoewel de neusholte een vaste vorm heeft is de frequentie van de nasale formanten toch enigszins variabel omdat die afhangt van de grootte van de opening van de doorgang naar de neusholte. De eerste nasale formant ligt ongeveer bij 250-500 Hz voor mannen, de overige formanten steeds ongeveer 1000 Hz hoger. De nasale formanten zijn echter sterk gedempt en de hogere nasale formanten zijn nauwelijks van belang.

Bij nasale klinkers wordt de nasale formant toegevoegd aan de formanten die ontstaan door de mond-keelholte (welke overigens ook een beetje wijzigt door het zakken van het velum). Daarboven gaat er echter ook energie verloren door de bijschakeling van de neusholte. Deze absorptie van energie noemen we antiresonantie of *antiformant*. De eerste nasale antiformant heeft een frequentie van ongeveer 400-1000 Hz (en volgende komen steeds weer 1000 Hz hoger maar zijn van geen belang). Wanneer een antiformant en een formant ongeveer een gelijke frequentie hebben dan kunnen ze elkaar geheel opheffen. De eerste nasale antiformant heeft vooral invloed op klinkers met een tweede formant tussen 500 en 1000 Hz. Deze kan sterk gereduceerd worden.

7.4. Medeklinkers

De medeklinkers hebben bij voorkeur een klinker nodig om goed tot hun recht te komen. Verschillende medeklinkers zijn daarnaast niet stationair te produceren. Ook doen ze niet alle een beroep op de stem als bron voor het geluid

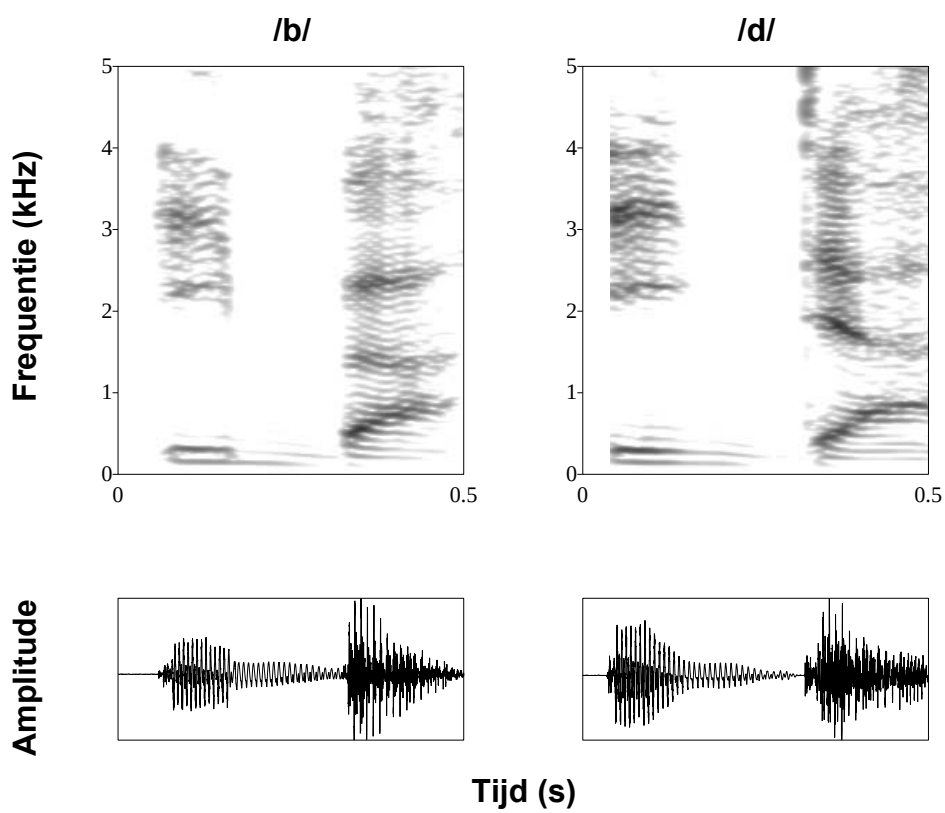
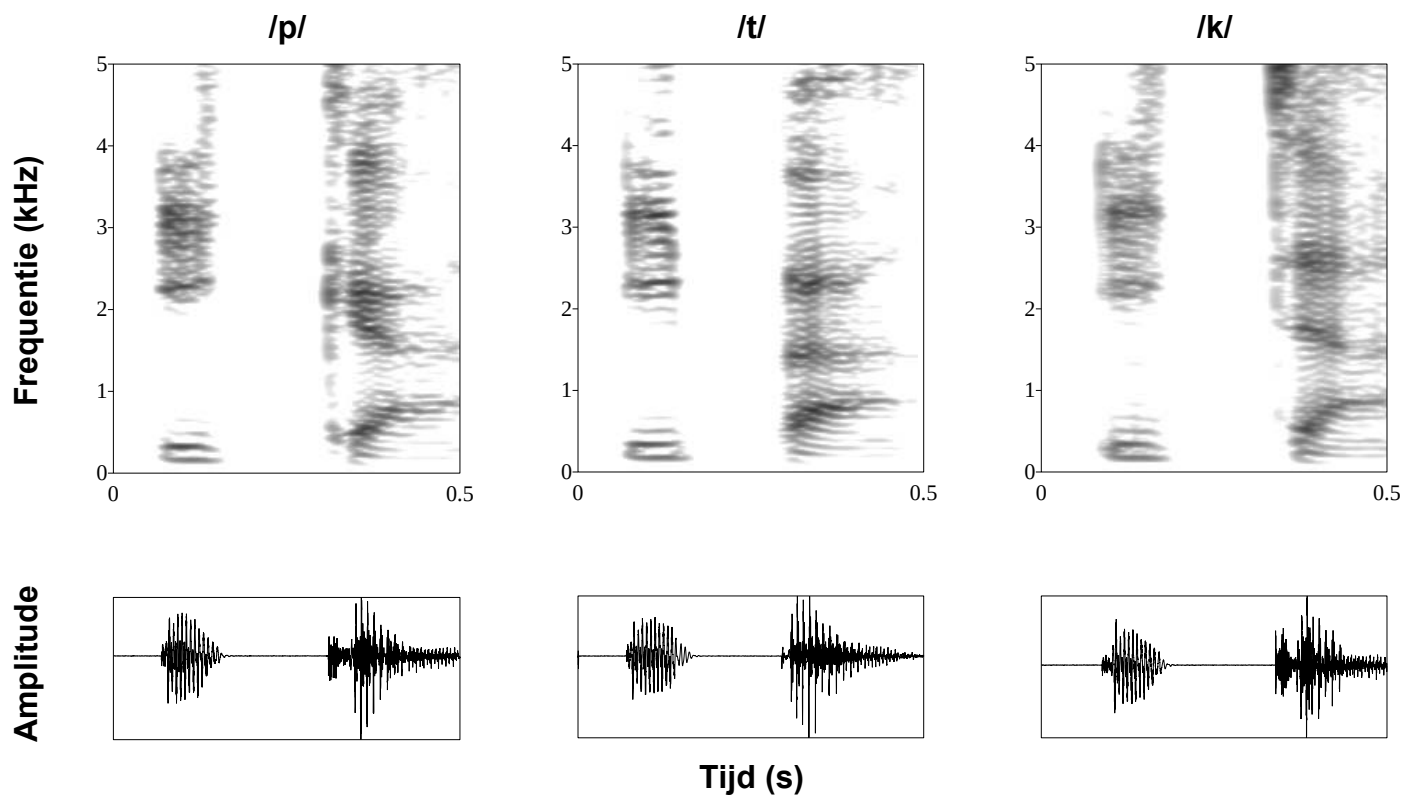
7.5. Plofklanken

De plofklanken /p t k b d/ kennen een abrupte articulatorische verandering. Het spraakkanaal wordt afgesloten met de tong of de lippen en er wordt druk opgebouwd die plotseling weer ongedaan wordt gemaakt door het opheffen van de afsluiting. Deze plotselinge verandering in druk leidt tot kortdurende sterke turbulenties van de ontsnappende lucht die we als een ruisploffje waarnemen. De plofklank kent daarom verschillende fasen. De eerste fase is die van de afsluiting. Omdat bij /p t k/ de stemplooiën niet trillen is het dan werkelijk even stil, het zijn de stemloze plofklanken. Bij de stemhebbende plofklanken /b d/ trillen de stembanden wel en het geluid lijkt dan op dat van de nasalen /m n ŋ/ (*vocal murmur*). Vervolgens komt er een plof die akoestisch gezien een uitbarsting van ruis is. Daarna volgt de overgang naar een volgende spraakklank waarin de akoestische kenmerken van de articulatie van de voorbije plofklank nog zichtbaar zijn. In het Engels kan die overgang in stemloze plofklanken geaspireerd uitgevoerd worden, dat wil zeggen met een korte /h/-achtige overgang met veel wilde lucht. In het Nederlands doen we dit echter niet. De stemplooiëtrilling start meestal alweer tijdens de ruisploff. In bijvoorbeeld het Engels en het Frans hoort in het rijtje plofklanken ook de /g/ thuis, als stemhebbende tegenhanger van /k/. Wij gebruiken deze klank alleen in leenwoorden (*goal, grand café*). In het Nederlands voeren we de 'g' echter uit als vellaire wrijfklank /χ/, zonder afsluiting.

Door het complexe karakter van de plofklank zijn de akoestische kenmerken divers. Bovendien zijn ze zeer afhankelijk van de omringende klanken. De duur van de stilte of vocal murmur kan variëren tussen 20 en 150 ms, de duur van de ruisploff tussen 10 en 50 ms. De stilte of de ruisploff kan ook geheel ontbreken. Het spectrum van de ruisploff hangt samen met de plaats waar het spraakkanaal wordt afgesloten (de plaats van articulatie). Bij de /p/ en de /b/ is dit tussen de lippen (*labiaal*). Dat betekent dat de ruis niet meer wordt beïnvloed door het spraakkanaal. Het spectrum kent geen echte resonantie pieken en is vlak tot dalend en lijkt op dat van de wrijfklank /f/. De /t/ en de /d/ zijn *alveolair*, de afsluiting vindt plaats achter de

tandkassen. Het spectrum is stijgend zoals bij de wrijfklank /s/. Bij de /k/ is de afsluiting achter in de mond (*velair*). Omdat de akoestische eigenschappen van de mondholte nog hun invloed hebben op de ruis zijn de middenfrequenties vertegenwoordigd met een formantachtig patroon.

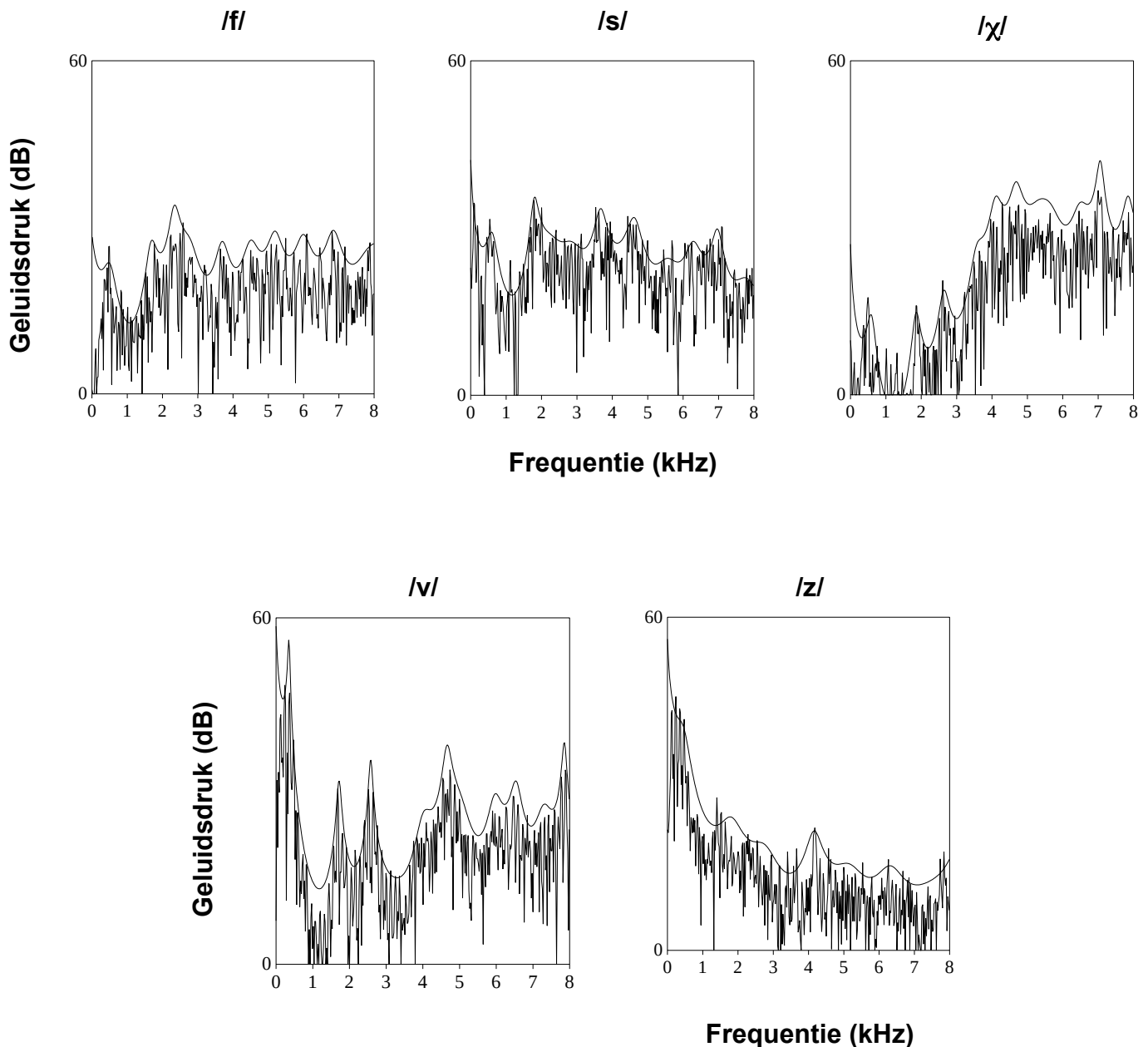
De overgang van een plofklank naar een volgende klinker laat een formantvergliding zien die indicatief is voor de plofklank. De eerste formant start meestal op een zeer lage frequentie als gevolg van de zeer nauwe doorgang in het spraakkanaal vlak na de opening van de afsluiting. Vervolgens vindt een stijging van de frequentie plaats naar die van de eerste formant van de volgende klinker. De frequentie van de tweede formant volgt een patroon als ware hij afkomstig van 600 Hz voor /p/ en /b/, van 1800 Hz voor /t/ en /d/ en van 3000 Hz voor /k/. Het hangt van de frequentie van de tweede formant van de volgende klinker af wat het uiteindelijk resulterend patroon is. Als we kijken naar de overgang van een klinker *naar* een plofklank dan zullen we in het formantpatroon van de klinker gelijksoortige maar gespiegelde patronen aantreffen. In Figuur 8 staan de oscillogrammen en spectrogrammen van de uitingen /ipa/, /ita/, /ika/, /iba/ en /ida/ waarin de kenmerken van de plofklanken in zowel de tijd als spectraal goed te zien zijn.



Figuur 8. *Oscillogram en spectrogram van de stemloze plofklanken /p/ in ipa, /t/ in ita en /k/ in ika, en de stemhebbende plofklanken /b/ in iba en /d/ in ida. Merk de verschillen in het formantverloop op van de plofklank naar de /a/; voor alle drie plofklanken start F1 in de ruisplof rond 300 Hz en stijgt tot 700 Hz, F2 start afhankelijk van de plofklank vanaf 1200 Hz voor /p/ en /b/, 1800 Hz voor /t/ en /d/ en van boven 2000 Hz voor /k/ om naar een waarde van 1200 Hz voor /a/ te gaan.*

7.6. Fricatieven

De fricatieven of *wrijfklanken* /f s v z χ/ worden gerealiseerd door een vernauwing van de stemweg. De doorgang is zo nauw dat de lucht er met grote snelheid doorgeperst wordt en er turbulenties ontstaan na de opening. Deze turbulenties vormen het brongeluid voor de fricatieven. Bij de /f/, /s/ en /χ/ trillen de stemplooien niet en ze worden daarom stemloze plofklanken genoemd, in tegenstelling tot /v/ en /z/ die stemhebbend zijn. In dat laatste geval is er sprake van twee bronnen: de stemplooien zelf met een periodiek brongeluid en de vernauwing met een ruisig brongeluid. De golfvorm van de stemloze /s/ en de stemhebbende /z/ in figuur 9 laat het verschil tussen het niet-periodieke en periodieke karakter van deze spraakklanken duidelijk zien. De plaats van vernauwing wordt *plaats van articulatie* genoemd en ligt voor /f/ en /v/ tussen de lippen en voor /s/ en /z/ achter de tanden, voor de 'g'-klank /χ/ achter in de mond. De /χ/ neemt in het Nederlands een bijzondere positie in omdat we de harde en zachte variant kennen. De zachte variant is een echte wrijfklank, maar in de harde variant wordt de vernauwing bij het zachte gehemelte zo verkleind dat het zachte gehemelte en tongrug door het Bernoulli effect in (onregelmatige) trilling kan raken. Spectraal gezien zijn er overeenkomsten tussen de wrijfklanken en de ruisplof van de plofklanken bij overeenkomstige plaatsen van articulatie. Bij /f/ is het spectrum vlak tot dalend. Bij /s/ is er een significante resonantieholte aanwezig tussen tong, tanden en lippen. De combinatie van de vernauwing en deze holte zorgt voor een eerste formant met een frequentie tussen 4000 en 6000 Hz en daarmee voor een stijgend spectrum. Het timbre van een /s/ is scherper dan dat van een /f/. Bij de /χ/ tenslotte heeft de mondholte invloed op het ruisspectrum dat daarom een formantpatroon heeft. Bij de stemhebbende wrijfklanken domineert het stembrongeluid over de ruis van de turbulenties wat resulteert in een spectrum voor /v/ en /z/ met relatief veel energie beneden 500 Hz. Daarnaast zorgen bij /v/ en /z/ resonanties in de mondholte achter de vernauwing nog voor hogere formanten die een sjwa-achtig patroon volgen. Rond 4-6 kHz domineert bij /z/ dezelfde resonantie als bij /s/. De spectra in figuur 9 tonen al deze karakteristieken.



Figuur 9. Spectra van de stemloze fricatieven /f/, /s/ en /x/ en de stemhebbende fricatieven /v/ en /z/. Bij de laatste klanken is de dominantie van de eerste harmonischen duidelijk zichtbaar.

De /h/ behoort ook tot de wrijfklanken. Het is een glottale wrijfklank waarbij de opening tussen de stemplooiën zo klein is dat er juist turbulenties ontstaan. De turbulenties zijn echter zwak en de /h/ heeft diensgevolge ook een geringe luidheid. De /h/ heeft geen eigen articulatie maar ontwikkelt zich in relatie tot de articulatie van voorafgaande en volgende klanken.

7.7. Liquidae

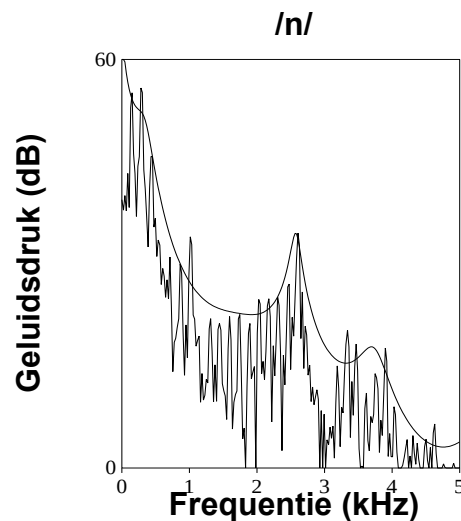
De /l/ en de /r/ worden liquidae (vloeiend) genoemd omdat ze worden gerealiseerd met een relatief geringe afsluiting van de stemweg. Ze zijn stemhebbend en hebben wat het spectrum betreft klinkerachtige eigenschappen.

De /l/ wordt lateraal genoemd omdat de tong het stemkanaal achter de tandkassen in het midden afsluit en de geluidstrillingen zich alleen aan weerszijden van de tong kunnen voortplanten. Bij de /l/ vormt de tong midden in de mond een aparte holte van zo'n 4.5 cm lang welke een antiresonantie vormt met een frequentie van ongeveer 2000 Hz. De /l/ is stemhebbend met de meeste energie beneden 1000 Hz. Onder invloed van de antiformant neemt het spectrum daarboven sterk af. Tussen 2000 en 5000 Hz kan door formantclustering overigens nog relatief sterke energie voorkomen welke de aangehouden /l/ een welluidend karakter kan geven.

Bij de /r/ vormt de tongpunt een afsluiting in het midden van harde gehemelte achter de tandkassen, waarna de tongpunt door ademdruk en het Bernouilli effect kortdurend in trilling wordt gebracht met een frequentie van zo'n 25 Hz. Het aantal malen dat de trilling wordt uitgevoerd kan variëren tussen precies één maal (*tap* of *flap*) en meerdere malen (*tril*). Alleen in het laatste geval is er sprake van een echte triklank. Omdat ook de stemplooien blijven trillen ontstaat een complexe klank. In de gesloten fase van de tongpunt is er alleen vocal murmur. In de open fase is er een klinkerachtig formant patroon met formantfrequenties bij 500 en 1000 Hz. Naast deze tongpunt /r/ kennen we ook de zogenaamde brauw /r/ die uvulair wordt gearticuleerd. De uvulaire afsluiting leidt net zoals bij de harde 'g' tot een trilling tussen de huig en de tongrug. In het Engels wordt een zogenaamde retroflexe /r/ geproduceerd waarbij de *onderzijde* van de tongpunt het gehemelte raakt. Deze versie heeft een zeer lage F3.

7.8. Nasale medeklinkers

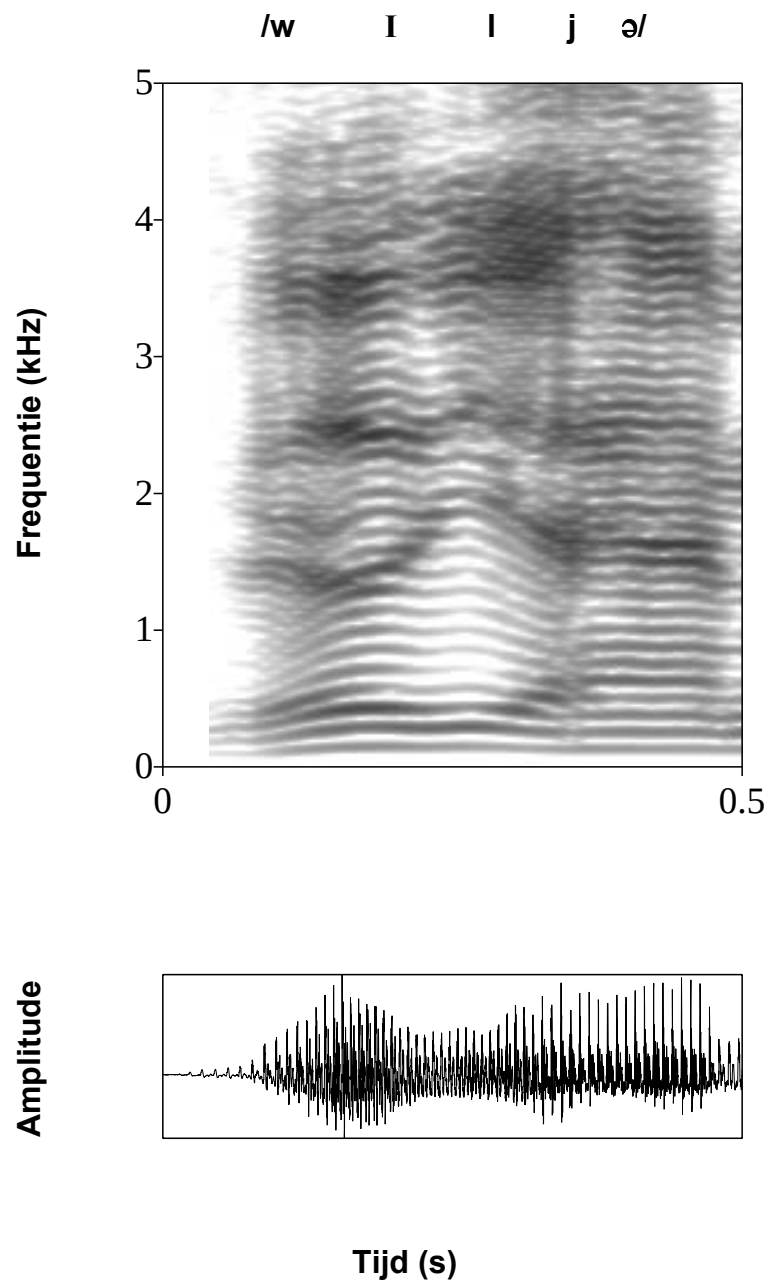
Bij de nasale medeklinkers /m/, /n/ en /ŋ/ wordt de neusholte bijgeschakeld en de mondholte geheel afgesloten. Door de lippen bij de /m/, door de tongpunt achter de tandkassen bij /n/ en achter in de mond bij /ŋ/. Het verschil tussen nasale medeklinkers wordt daarom veroorzaakt door het verschil in grootte van de bijgeschakelde mondholte. De daarbij behorende resonanties fungeren nu als antiformanten. De frequentie van de antiformant hangt samen met de grootte van de mondholte en ligt voor /m/ tussen 750-1250 Hz, voor /n/ tussen 1450-2250 Hz en voor /ŋ/ boven 3000 Hz. De /m/, /n/ en /ŋ/ hebben allen een geprononceerde eerste formant bij ongeveer 300 Hz die gerelateerd is aan de totale lengte tussen stemplooien en neusopening. Daarnaast zijn er nog een tweede formant bij 1 kHz en een derde formant boven 2 kHz te verwachten die gerelateerd zijn aan resonanties in keelholte en de neusholte zelf. Dit alles is zichtbaar in het spectrum van /n/ in Figuur 10.



Figuur 10. Sterk dalend spectrum van de nasale medeklinker /n/.

7.9. Halfklinkers

De medeklinkers /w/ en /j/ noemen we halfklinkers (*approximanten*). Ze zijn alleen goed te maken in combinatie met andere klanken. Dat zijn vooral opvolgende klinkers, vandaar de naam halfklinkers. In het Nederlands kennen we overigens ook combinaties met liquidae als in *wladimir* en *wreed*. De /j/ heeft een beginarticulatie die enigszins op die van /i/ lijkt. Bij /w/ is er een combinatie van een vernauwing bij de lippen en een vernauwing achter in de mond. In verband met dat laatste lijkt de /w/ ook enigszins op een /u/. Beide realisatie leiden tot een lage waarden van de eerste formant rond 300 Hz. Voor de /j/ heeft de tweede formant een frequentie bij 2000 Hz, bij /w/ ligt die lager rond de 1000 Hz. Het verdere verloop van de formantvergliding hangt af van de eigenschappen van de omringende spraakklanken. Figuur 11 geeft het spectrogram van de uiting '*wil je*' waarin die verglijdingen goed te zien zijn. De /w/, /l/ en /j/ manifesteren zich als een continue beweging van de zwarte formantgebieden. De formanten van de klinkers laten zich daarom moeilijk bepalen. Elk spraaksegment zal weer eigen waarden opleveren. Hooguit geeft een segment uit het midden karakteristieke waarden van formantfrequenties.



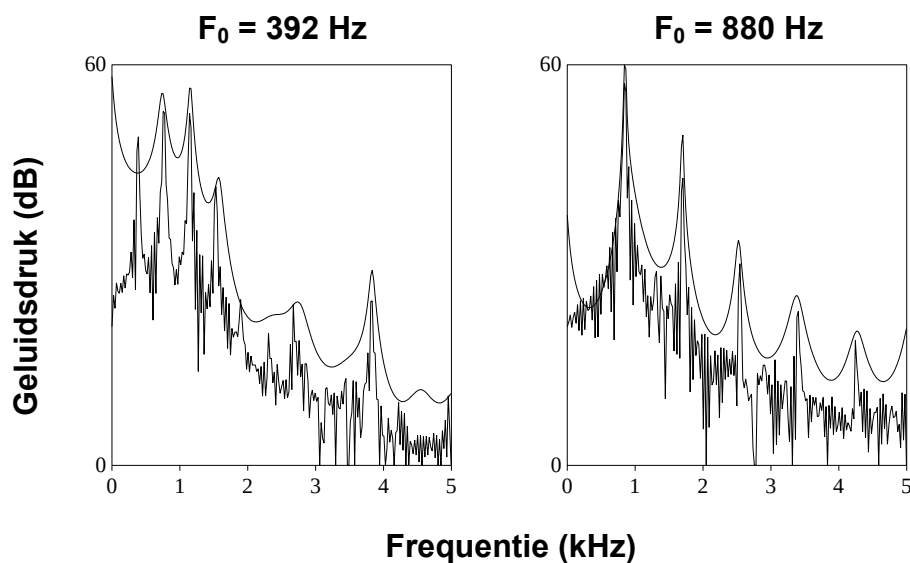
Figuur 11. *Spectrogram van de frase 'wil je'. De lage F1 bij 300 Hz stijgt pas bij de sjwa naar 500 Hz. De frequentie van F2 is voor /l/ het hoogst (2000 Hz) terwijl dan ook de dip (wit gebied) tussen F1 en F2 het grootst is door de aanwezigheid van een anti-formant.*

8. De stembron erbij (optioneel)

Tot nu toe hebben we ons wat betreft de stemhebbende klanken alleen bezig gehouden met de akoestische filtereigenschappen van het spraakkanaal. Voor het totaal resultaat van spraak, zoals we dat in de vele voorbeelden van spectra al zagen, moeten we de eigenschappen van het stembrongeluid, het akoestisch filter en de afstraling aan de mondopening samennemen. Het standaard stembrongeluid is een reeks harmonischen waarvan de amplitude afneemt met -12 dB/octaaf. De filterende werking van het spraakkanaal modificeert dit aflopend spectrum

doordat harmonischen met een frequentie rond een formantfrequentie beter doorgegeven worden dan andere (zie figuur 1). Deze harmonischen zullen een luisteraar meer opvallen en de luisteraar zal de klank op grond daarvan associëren met een bepaalde spraakklank. Het is belangrijk om op te merken dat de luisteraar formanten niet direct waarneemt maar alleen harmonischen die een indicatie geven bij welke frequenties formanten ongeveer aanwezig zijn in het spraakgeluid. Voor een goede verstaanbaarheid van spraak moet een luisteraar dus in staat zijn om de formanten af te leiden uit de harmonischen.

Bij hoge grondtonen is het schatten van formantfrequenties voor zowel mensen als machines een probleem. Als een sopraan zingt met een grondtoon van 880 Hz (a") dan zijn er in het geluid slechts harmonischen bij 880, 1760, 2540, 3520, 4400, 5280 ... Hz. Het is onmogelijk om op grond van de amplitudes van die harmonischen af te leiden welk formantpatroon bijgedragen heeft tot dat geluid. De sopraan is daarom bijna onverststaanbaar op die toonhoogte. Figuur 12 geeft het FFT spectrum en het LPC spectrum van een /a/, gezongen door een sopraan op 392 Hz en op 880 Hz. Bij een grondfrequentie (F_0) van 392 Hz corresponderen de eerste twee formanten waarschijnlijk nog wel met de 2e en 3e harmonische bij 784 Hz en 1176 Hz. Bij een grondfrequentie van 880 Hz wordt de 2e formant eigenlijk vrij zeker gemist door de harmonischen. Merk op dat het LPC spectrum in dit geval een formantpiek legt bij elke harmonische. Op die manier wordt in zekere zin de omhullende van het spectrum goed beschreven, maar als formantschatting deugt er niets van.



Figuur 12. Spectrum van de klinker /a/ gezongen door een sopraan op een grondfrequentie van 392 Hz (g') en 880 Hz (a''). De ruime spatiëring van de harmonischen maakt het onmogelijk de formanten goed te schatten. Dat blijkt uit het LPC spectrum welke niet de formanten maar de harmonischen volgt.

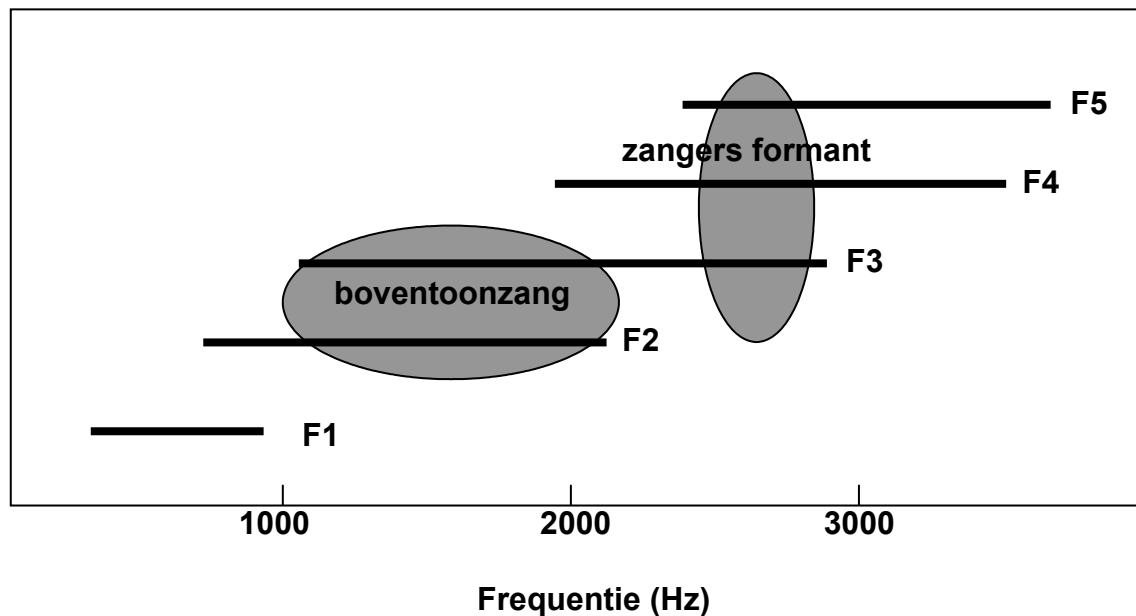
Omdat de eerste formantfrequentie een waarde kan hebben vanaf 250 Hz kan er bij de spreekstem van vrouwen, en in het bijzonder weer bij zingen een probleem optreden met de draagkracht van de stem. Omdat de stembron geluid produceert met een aflopend spectrum zijn het met name de lagere harmonischen die voor de intensiteit (waargenomen als luidheid) van het stemgeluid zorgdragen. De lagere harmonischen ondervinden de invloed van de eerste formant en meer precies zijn het dus de harmonischen in de buurt van de frequentie van de eerste formant die de luidheid van de spraakklank bepalen. Veronderstel dat een zanger een /i/

zingt met een F1 van 250 Hz bij een grondfrequentie van 440 Hz. Dan wordt de eerste formant helemaal niet aangesproken door een harmonische en zal de luidheid en draagkracht van de stem sterk verminderd zijn. Dat kan de zanger ondervangen door de articulatie van de /i/ aan te passen en de F1 te laten stijgen tot in de buurt van de grondtoon. Dat gebeurt als de zanger de mond wijder opent en de onderkaak laat zakken. De luidheid van de zang wordt daarmee herstelt, maar het gaat ten koste van de herkenbaarheid van de /i/ (Sundberg, 1978).

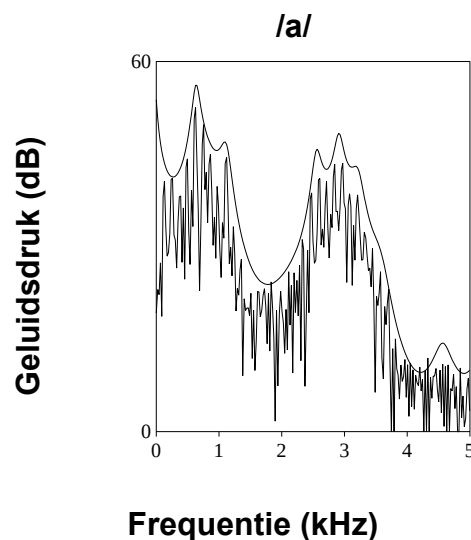
Onze stemplooien kunnen bij een zelfde grondtoon nog op veel verschillende manieren trillen. De sluiting kan geleidelijk gaan, zoals bij zachte stemgeving, of juist erg abrupt, zoals bij luide stemgeving. Het effect komt niet alleen tot uitdrukking in de luidheid van de stem maar ook in de klankkleur en het spectrum. Het bronspectrum is veel steiler bij zacht spreken dan bij luid spreken. Dat kan betekenen dat bij zacht spreken de harmonischen onder de hogere formanten erg zwak worden en dat het daarmee voor de luisteraar moeilijker wordt om de frequenties van die formanten te bepalen. De verstaanbaarheid wordt dan slechter, en niet alleen omdat de spraak zacht is.

9. Samenwerkende formanten (optioneel)

Als er zodanig wordt gearticuleerd dat formantfrequenties in elkaars buurt komen dan zullen harmonischen bij die frequenties extra krachtig in het resulterend spraakgeluid aanwezig zijn. In spraak maken we daar gewoonlijk geen gebruik van, maar in zang wel. De beste mogelijkheden voor samenwerking tussen formanten zijn er voor F2 en F3, en voor F3, F4 en F5, wat afgeleid kan worden uit figuur 13 waar de werkingsgebieden van de vijf formanten globaal zijn aangegeven. Het samenbrengen van F3, F4 en F5 wordt gepraktiseerd in westerse klassieke zang en staat bekend als de *zangersformant* (figuur 14). Omdat deze formanten minder van belang zijn voor klinkerdefinitie kan deze clustering van formanten steeds worden nagestreefd zonder de verstaanbaarheid geweld aan te doen. Het leidt tot een gewaardeerde constante factor in de klankkleur van klassieke zang en wordt wel egalisatie van het timbre genoemd. Het feit van een constante factor duidt erop dat de articulatorische vereisten in een relatief stabiel deel van het spraakkanaal moeten worden gezocht. Sundberg (1974) toonde aan dat die gevonden kunnen worden in de holtes vlak boven de stemplooien. Een verlaging van de positie van het strottenhoofd zou deze akoestische condities bevorderen. Mogelijk zijn de condities van nature aanwezig in welluidende spreekstemmen en in begiftigde zangers. Sundberg (1977) wijst erop dat naast egalisatie van het timbre een ander belangrijk effect van de zangersformant is dat het spectrum van de zanger afwijkt van het gemiddelde spectrum van een symfonie orkest. De zanger zal daarom door een luisteraar beter van het begeleidend orkest kunnen worden onderscheiden.



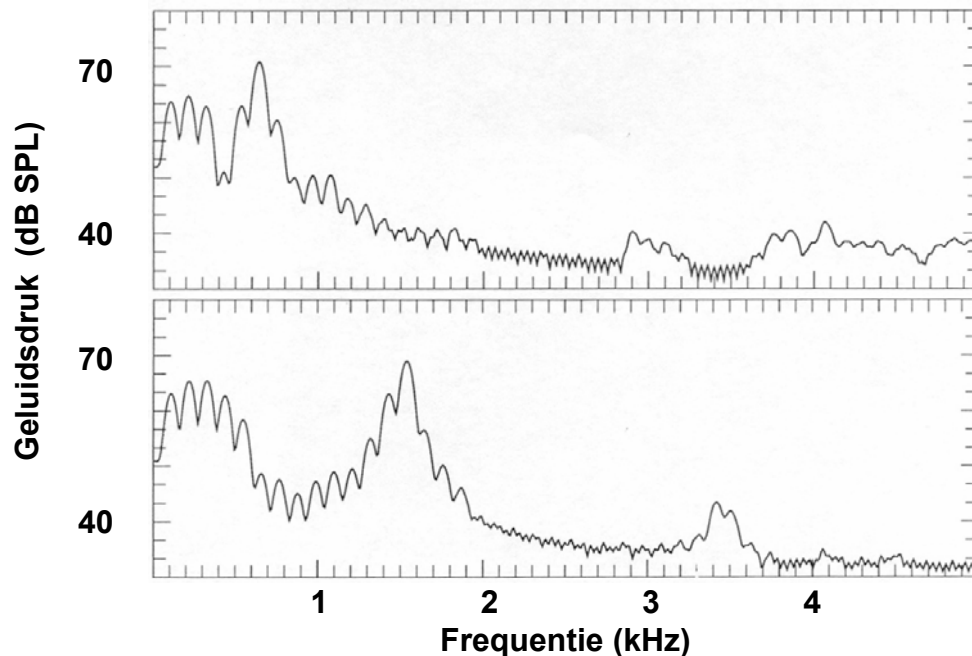
Figuur 13. Schets van de frequentiegebieden van de vijf formanten met daarin aangegeven de twee gebieden waar samenwerking tussen formanten kan voorkomen, toegepast in boventoonzang en in de zangersformant.



Figuur 14. Spectrum van de klinker /a/ gezongen door een bariton op 98 Hz (G). De clustering van formanten F3, F4 en F5 tot de zangersformant is duidelijk zichtbaar tussen 2.5 en 3.5 kHz.

De samenwerking van F3 en F2 wordt gebruikt in zogenaamde boventoonzang, een zangtechniek die afkomstig is uit Centraal Azië. In boventoonzang wordt gezongen met *constante* grondtoon. Met behulp van articulatie worden zowel F2 als F3 in de nabijheid van een gewenste harmonische (of boventoon) gebracht die daardoor sterk uitspringt boven alle andere harmonischen en hoorbaar wordt als een soort fluit. Door subtiele articulatorische veranderingen kan een hele reeks boventonen hoorbaar worden gemaakt. De articulatie loopt ongeveer als van /u/ via sjwa tot /i/ en volgt daarmee F2. Door een articulatie die lijkt op die van een Engelse retroflexe /ɹ/ wordt een zeer lage waarde van F3 bereikt die in de buurt komt

van F2. Voor de allerlaagste boventoonklanken wordt F1 gebruikt en wordt er nasaal gezongen om een scheiding met lagere boventonen te bewerkstelligen (Bloothoof et al., 1992). Figuur 16 geeft twee voorbeelden.



Figuur 16. Spectra van boventoonklanken waarin respectievelijk de 6^e en 14^e boventoon worden benadrukt. Bij de 6^e boventoon is de 4^e boventoon door nasaal zingen verzwakt waardoor een betere scheiding ontstaat met de laagste harmonischen. Voor beide klanken is de bedoelde boventoon 5 tot 10 dB luider dan alle andere boventonen.

Referenties

- Bloothoof, G., Bringmann, E., van Cappellen, M., van Luipen, J.M., and Thomassen, K.P. (1992). Acoustics and perception of overtone singing, *Journal of the Acoustical Society of America*, **92**, 1827-1836.
- Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, **4**, 337-351.
- Fant, G. (1960). *Acoustic Theory of Speech Production*, Den Haag: Mouton.
- Kent, R.D. en Read, C. (1992). *The Acoustic Analysis of Speech*, San Diego: Singular.
- Klein, W., Plomp, R. en Pols, L.C.W. (1970). Vowel Spectra, Vowel Spaces, and Vowel Identification. *Journal of the Acoustical Society of America*, **48**, 999.
- Koopmans-van Beinum, F.J. (1980). *Vowel contrast reduction*. Proefschrift, Academische Pers B.V. Amsterdam.
- Lindblom, B. en Sundberg, J. (1971). Acoustical Consequences of Lip, Tongue, and Larynx Movement, *Journal of the Acoustical Society of America*, **50**, 1166-1179.
- Nierop, D.J.P.J. van, Pols, L.C.W. en Plomp, R. (1973). Frequency Analysis of Dutch Vowels from 25 Female Speakers, *Acustica*, **29**, 110-118.
- Olive, J.P., Greenwood, A., Coleman, J. (1993). *Acoustics of American English Speech*, Heidelberg: Springer.
- Rietveld, A.C.M. en van Heuven, V.J. (1997). *Algemene Fonetiek*, Bussum: Coutinho

- Stevens, K.N. (1997). Articulatory-Acoustic-Auditory Relationships, in *The Handbook of Phonetic Sciences*, W.J. Hardcastle en J. Laver (editors), Oxford: Blackwell.
- Sundberg, J. (1974). Articulatory interpretation of the "singing formant", *Journal of the Acoustical Society of America*, **55**, 838-844.
- Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, **236**, 82-91.
- Sundberg, J. (1987). *The science of the singing voice*. San Diego: Singular.