

WERKBOEK VOOR HET
PRACTICUM

*INLEIDING IN DE
FONETIEK*

cursuscode 200300287

Opleiding Taalwetenschap, Studierichting Fonetiek
Onderwijsinstituut Vreemde Talen, Faculteit Letteren
Universiteit Utrecht

najaar 2004

INHOUDSOPGAVE

HOOFDSTUK 1 INLEIDING	4
1.1 Doelstelling van het practicum	4
1.2 Wat gaan we doen	4
1.3 Het verslag	4
1.4 Waar is welke informatie te vinden	4
HOOFDSTUK 2 BASISTHEORIE SIGNALEN	6
2.1 Wat is geluid	6
2.2 Registratie en weergave	6
2.3 Fysische oorzaken voor verschil in waarneming	6
2.4 Enkelvoudige signalen	11
Nieuwe begrippen	12
VERVOLG HOOFDSTUK 2 BASISTHEORIE SIGNALEN	12
2.5 Samengestelde signalen	12
2.6 Samengestelde signalen als functie van tijd en frequentie	14
2.7 Ruis	17
2.8 Spraaksignalen	18
2.9 Formanten	19
Nieuwe begrippen	23
Opdrachten hoofdstuk 2	24
HOOFDSTUK 3 DIGITAAL GELUID	26
3.1 Inleiding	26
3.2 Opnemen en vastleggen	26
3.3 Opslaan in de computer	26
3.4 Het binaire stelsel	26
3.5 A/D-en D/A-conversie	27
3.6 Bemonsteringsfrequentie	28
3.7 Amplitude-resolutie	29
3.8 De DAT-recorder	30
Nieuwe begrippen	31
Opdrachten hoofdstuk 3	32
HOOFDSTUK 4 ANALYSE VAN SPRAAK	34
4.1 Filters	34
4.2 Spectrografische analyse	36
4.3 Analyse van grondfrequentie en intensiteit	38
Nieuwe begrippen	40
Opdrachten hoofdstuk 4	41
HOOFDSTUK 5 ANALYSE EN SYNTHESE VAN SPRAAK VIA LPC	42
5.1 Wat biedt LPC?	42
5.2 LPC-analyse en synthese	42
5.3 PSOLA	45
Nieuwe begrippen	46
Opdrachten hoofdstuk 5	47
Opdrachten 6	48
Verantwoording	49

APPENDIX I WERKEN OP HET UNIX-SYSTEEM	50
Inloggen.....	50
Uitloggen	50
Wachtwoord wijzigen.....	50
Directorystructuur.....	50
Commando's.....	50
Afdrukken.....	51
APPENDIX II GEBRUIK SPRAAKPROGRAMMATUUR	52

HOOFDSTUK 1

INLEIDING

1.1 Doelstelling van het practicum

In de experimentele fonetiek wordt gebruik gemaakt van verschillende methoden om spraak te analyseren, te manipuleren en te synthetiseren. In het practicum behorend bij de cursus *Inleiding in de Fonetiek* leer je een aantal van deze methoden kennen en gebruiken. Het practicum heeft de volgende doelen:

1. Het vertrouwd raken met de benodigde apparatuur en programmatuur.
2. Het krijgen van inzicht in de meest gangbare methoden van spraakanalyse, manipulatie en -synthese.
3. Het leren zorgvuldig en gestructureerd te werken en schriftelijk verslag te doen van de uitgevoerde opdrachten.

1.2 Wat gaan we doen

Dit *Werkboek* bevat naast deze inleiding 4 hoofdstukken, bestaande uit theorie en opdrachten. Het is de bedoeling dat je elke week, voorafgaand aan de practicumbijeenkomst, de theorie gelezen hebt. Hoofdstuk 2 is erg groot en daarom verdeeld over twee weken. Het is echter wel één hoofdstuk, omdat de stof een duidelijk geheel is.

Tijdens de practicumbijeenkomsten wordt in groepjes van twee gewerkt aan het uitvoeren van de opdrachten. In de eerste week is er nog geen practicum. Voor de bijeenkomsten hierna, die een dagdeel per week beslaan, geldt een aanwezigheidsplicht. Naast de practicumbijeenkomsten, die begeleid worden door een practicumassistent, is er gelegenheid tot zelfstandig werken.

1.3 Het verslag

De uitwerkingen van de opdrachten worden ingeleverd en beoordeeld. Zorg dat in de uitwerkingen duidelijk is waar wat te vinden is, en dat het geheel er verzorgd uitziet. Het laatste betekent dat het in ieder geval getypt of op een tekstverwerker gemaakt moet worden. Afbeeldingen kan je het beste verkleind kopiëren en invoegen in de tekst. Beschrijf eventuele verschillen tussen afbeeldingen in de tekst en geef in de figuren zelf altijd aan wat de assen representeren, en welke waarden relevant zijn voor de betreffende opdracht. Programma- en filenamen noemen is overbodig. Geef echter wel steeds een korte(!) beschrijving van wat het programma doet. Tenslotte is het altijd raadzaam om een kopie in te leveren en de originelen zelf te houden! Zet je naam en studentnummer op alles wat je inlevert, en vertel de practicumassistent waar je postvakje te vinden is, zodat de laatste opdracht op die manier teruggegeven kan worden.

1.4 Waar is welke informatie te vinden

Het grootste deel van de tijd zullen we achter de computer zitten. Dat wil zeggen, achter een beeldscherm van een zgn. 'werkstation'. In de practicumruimte zijn een aantal werkstations. Alle werkstations bewaren hun bestanden (zoals spraakfiles) echter op één harde schijf, die zich elders bevindt. Om elkaar niet in de weg te zitten en om je spullen te beveiligen, krijgt elk practicumkoppel een eigen 'account'. Dit is je eigen stukje geheugenruimte in de computer, waar je toegang toe krijgt door een naam (username) en een toegangscode (password) in te typen. De practicumassistent zal je vertellen wat er ingetypt moet worden.

Als je eenmaal in de computer zit (ingelogd bent), heb je verschillende programma's tot je beschikking. Je start een programma op door de naam ervan in te typen in het venster getiteld 'Console', of door het programma te selecteren uit het menu linksboven op het scherm. In de practicumruimte staan klappers met informatie over alle beschikbare programmatuur. Van elk programma staat beschreven wat het precies doet en hoe het werkt. Bij de opdrachten in dit

werkboek wordt niet vermeld welke programma's je kan gebruiken, omdat de programmatuur vaak verandert. De practicumassistenten zullen je ter plekke vertellen welke programma's je voor de verschillende opdrachten moet gebruiken.

Bij het werken met spraaksignalen wordt veelvuldig gebruik gemaakt van computers. Geluidssignalen kunnen opgeslagen en op elk moment weer opgeroepen worden, we kunnen een signaal zichtbaar maken op het scherm en er is allerlei programmatuur voorhanden om signalen te analyseren en te manipuleren. De computer is met andere woorden een goed hulpmiddel.

Voor dit practicum is voor elk koppel een klein stukje geheugenruimte op de centrale computer gereserveerd; een account. Hier kan je gebruik van maken door 'er op in te loggen'. Voor meer informatie over het computersysteem en het gebruik ervan verwijzen wij naar de informatie op de schouw in de practicumruimte. Bovendien bevat dit *Werkboek* een nuttige Appendix, met daarin informatie om de computer-werkstations te gebruiken (de werkstations gebruiken het Linux operating systeem).

Meer informatie over de spraak-programmatuur kan je vinden via Internet op de volgende locatie: http://www.let.uu.nl/~Hugo.Quene/personal/faq/spraakprogr.html
--

Veel plezier en succes bij het practicum!

Opleiding Taalwetenschap, Studierichting Fonetiek
Universiteit Utrecht
najaar 2004

HOOFDSTUK 2 BASISTHEORIE SIGNALLEN

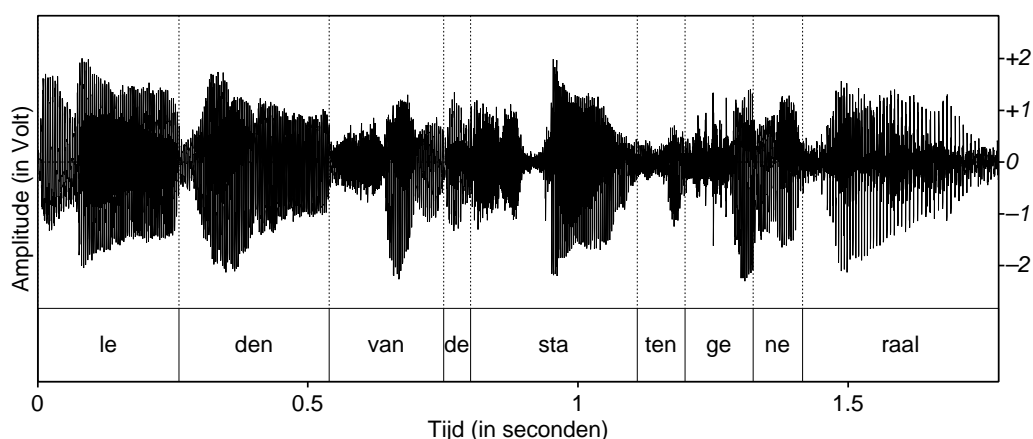
2.1 Wat is geluid

In dit hoofdstuk zal een korte introductie gegeven worden over geluid in het algemeen en spraakgeluid in het bijzonder. Strikt genomen bestaat geluid alleen bij de gratie van een waarnemer. Geluid wordt veroorzaakt door een trillende bron. Deze bron veroorzaakt luchtdrukwisselingen. Lucht bestaat uit atmosferische deeltjes; door het geluid worden die deeltjes afwisselend dichter op elkaar gedrukt en uit elkaar getrokken. In een geluidstrilling bewegen de luchtdeeltjes zich over een minieme afstand rond hun gemiddelde positie; die afstand is ca. 10^{-11} tot 10^{-5} m, afhankelijk van de intensiteit en frequentie van de trilling. De gemiddelde positie van ieder luchtdeeltje blijft onveranderd. De luchtdeeltjes zelf verplaatsen zich niet naar je oor toe, zoals bij wind, maar zij geven de drukwisselingen door aan de omliggende deeltjes. Daardoor verplaatsen de luchtdrukwisselingen zich in de lucht: de geluidsgolf plant zich voort.

2.2 Registratie en weergave

Om metingen te kunnen doen aan geluid, moet dit op de een of andere manier opgeslagen worden. Dit gebeurt bijvoorbeeld op een cassettebandje, een geluidsband, een DAT-bandje of in de computer. Hiertoe worden met behulp van een microfoon luchttrillingen omgezet in een elektrisch signaal; een membraan in de microfoon wordt in trilling gebracht en deze trillingen worden omgezet in een variërende elektrische spanning (wisselspanning).

De voordelen van elektrische spanning ten opzichte van luchtdrukverschillen zijn dat je het signaal kan opslaan, kan versterken, verzwakken en op andere manieren kan manipuleren. Het verloop van de elektrische spanning (de golfvorm) kan ook zichtbaar gemaakt worden. Een voorbeeld wordt gegeven in figuur 1, waar op de horizontale as het verloop in de tijd (in seconden) te zien is, en op de verticale as het verloop van de elektrische spanning (in millivolt). Zo'n weergave wordt een oscillogram genoemd.



Figuur 1: Verloop van de elektrische spanning van een spraaksignaal.

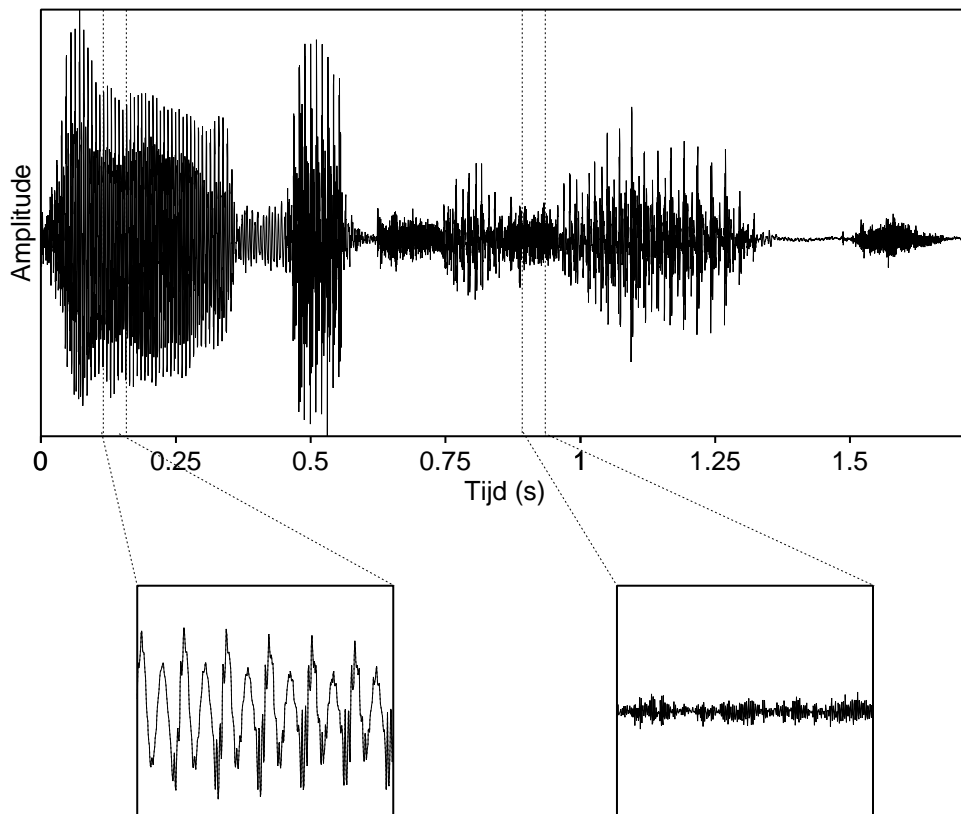
2.3 Fysische oorzaken voor verschil in waarneming

De waarneming van geluid resulteert in verschillende gewaarwordingen. Denk bijvoorbeeld aan de luidheid, of het geluid bijvoorbeeld dof of schel is, de toonhoogte, en of er überhaupt een

toonhoogte aan het geluid toegekend kan worden. Op dit punt moet een onderscheid gemaakt worden tussen de gewaarwording door het gehoor en de gemeten waarde van een bepaalde fysische eigenschap van een geluidssignaal. De fysische oorzaken van verschillen in waarneming kunnen het beste duidelijk gemaakt worden aan de hand van plaatjes.

Toon of ruis

Geluiden kunnen grofweg onderscheiden worden in periodiek en a-periodiek. De indruk die beide categorieën op ons gehoor maken, zijn respectievelijk aan te duiden met tonaal en ruis. Een aangehouden klinker /e/ is een voorbeeld van een tonaal geluid, terwijl een stemloze medeklinker als /s/ een ruisachtige indruk geeft.



Figuur 2: Twee uitvergrotingen van de golfvorm van een spraaksignaal, van respectievelijk een periodiek en een a-periodiek fragment.

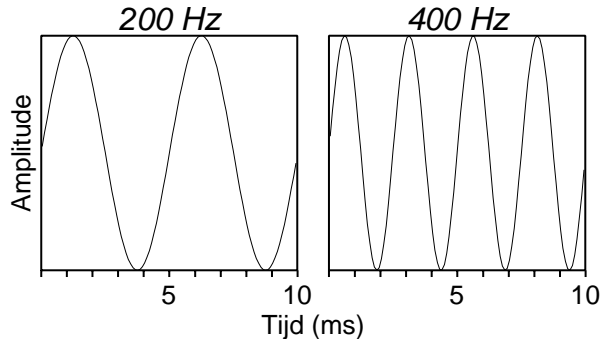
In Figuur 2 is respectievelijk een periodiek en een a-periodiek signaal weergegeven. Een signaal is periodiek als de golfvorm binnen het signaal zich steeds herhaalt. Het kleinste zich herhalende deel van de golfvorm heet de periode (periodetijd T).

Toonhoogte

Er bestaat een direct verband tussen de periodeduur van een signaal en de waargenomen toonhoogte. De fysische correlaat van toonhoogte wordt meestal aangeduid met de term frequentie (f). De frequentie van een signaal wordt bepaald door het aantal perioden per seconde, en uitgedrukt in eenheden die Hertz (Hz) heten.

$$f = 1 / T$$

Als een periode 4 milliseconden duurt ($T=1/250$ seconde), dan is de frequentie van dat signaal $f = 1 / (1/250) = 250$ Hz. Twee signalen met verschillende frequentie zijn weergegeven in Figuur 3 hieronder.

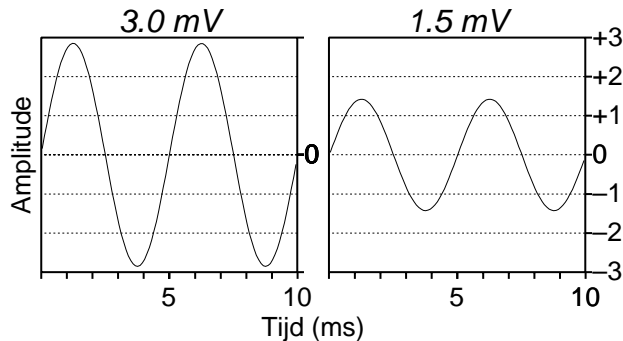


Figuur 3: Twee periodieke signalen, met frequenties van respectievelijk 200 Hz en 400 Hz.

Bij een periodiek signaal van 200 Hz duurt één periode: $1 / 200 = 0.005$ s = 5 ms.

Luidheid

De maximale uitwijking van de golfvorm ten opzichte van de ruststand (ofwel de maximale afwijking van de geluidsdruk) wordt aangeduid met de term amplitude (A). De amplitude van een signaal wordt waargenomen als de luidheid van dat signaal. In Figuur 4 hieronder zijn twee signalen weergegeven met verschillende amplitudes.



Figuur 4: Twee periodieke signalen, met amplitudes van respectievelijk 3.0 mV en 1.5 mV.

Er is een relatie tussen de amplitude van een elektrisch signaal en de luidheid die we waarnemen als dat signaal via een luidspreker ten gehore wordt gebracht. De uitslag van de luidsprekerconus volgt de momentane waarde van het elektrisch signaal. Dus een grote signaalamplitude geeft een grote conusuitslag, en een kleine amplitude een even zo kleine uitslag. Doordat de conus rechtstreeks luchtmolekelen in trilling brengt is de geluidsdruk (p) (die immers de mate van luchtmolekuultrilling voorstelt) recht evenredig met de amplitude van het aangeboden elektrisch signaal. Als we echter iets over luidheid willen zeggen moeten we eerst naar geluidsintensiteit (I) kijken. Geluidsintensiteit komt overeen met het kwadraat van de geluidsdruk.

Voor de liefhebbers: de exacte relatie tussen intensiteit en druk is
$$I = \frac{1}{2} \cdot \frac{p^2}{\rho \cdot v},$$

waarbij p = maximale geluidsdruk, ρ = soortelijk gewicht van medium, voor lucht 1,29 kg/m³, en v = geluidssnelheid, voor lucht 331 m/s (soortelijk gewicht en

geluidssnelheid zijn afhankelijk van temperatuur en luchtdruk; deze waarden gelden bij 0°C).

Geluidsdruk wordt uitgedrukt in eenheden kracht per oppervlakte, ofwel Pascal: 1 Pa = 1 Newton/m². De trillende luchtdeeltjes transporteren akoestische energie¹, als ze botsen met hun naburige deeltjes. De intensiteit is de hoeveelheid energie van een geluidstrilling, per seconde en per oppervlakte dwars op de voortplantingsrichting. Intensiteit wordt uitgedrukt in eenheden energie per seconde en per m², ofwel Nm/(s·m²) = Watt/m². Een sinusvormige geluidstrilling met frequentie $f=1000$ Hz en met geluidsdruk $p=2 \cdot 10^{-5}$ N/m² heeft een intensiteit van $4,7 \cdot 10^{-13}$ Watt/m².

Uit proefnemingen is gebleken dat de gevoeligheid van het menselijk gehoor ongeveer correspondeert met de logaritme van de intensiteit van het geluid (zie hieronder). Dit gegeven, plus het feit dat de maximaal toelaatbare geluidsdruk wel een factor 1000000 (=10⁶) groter kan zijn dan de minimaal waarneembare geluidsdruk, leidt tot het gebruik van een logaritmische schaal voor geluidsintensiteiten of geluidsdrukken. Op zo'n schaal kunnen zowel heel kleine als heel grote waarden (in grafieken bijv.) met dezelfde relatieve nauwkeurigheid worden weergegeven.

Logaritme

De logaritme van een getal x is de macht waartoe een grondgetal verheven moet worden, om x te verkrijgen. Het grondgetal is meestal 10. Voor zgn. 'natuurlijke' logaritmes met grondgetal $e \approx 2.7$ wordt doorgaans de aanduiding $\ln(x)$ gebruikt.

voorbeeld: $^{10}\log(1000)=3$
 want $\text{grondtal}^{\text{macht}}=10^3=10 \cdot 10 \cdot 10=1000$

Een belangrijk voordeel van logaritmes is dat vermenigvuldiging van twee getallen versimpeld wordt tot optelling van hun logaritmes:

$100 \cdot 1000 = 100000$
 $10^2 \cdot 10^3 = 10^5$
 $^{10}\log(100) + ^{10}\log(1000) = 2 + 3 = ^{10}\log(100000) = 5$

Als we praten over verhoudingen, bv. van geluidsdrukken of intensiteiten, dan zijn grote verhoudingen makkelijker uit te drukken in logaritmes:

$10 : 1000 = 1 : 10^3$
 $^{10}\log(10) : ^{10}\log(1000) = 1 : 3$

Decibel

In de praktijk meten we zelden absolute geluidsintensiteiten, maar met verhoudingen van geluidsintensiteiten. Gelijke verhoudingen van geluidsintensiteiten worden door verschillende mensen namelijk waargenomen als gelijke luidheidsverschillen. We zijn daarom vaak meer geïnteresseerd in verhoudingen dan in de werkelijke geluidsintensiteiten, geluidsdrukken, elektrische spanningen etc. Een dergelijke verhouding wordt doorgaans uitgedrukt in decibels [dB] en wordt niveauverschil (level) genoemd. Een decibel geeft de verhouding van twee getallen weer, en wel op logaritmische schaal. Meestal wordt de verhouding bepaald van een meetwaarde ten opzichte van een afgesproken referentiewaarde.

Voor de relatie tussen intensiteit I en niveau L geldt:

$$L_I = 10 \cdot \log\left(\frac{I}{I_0}\right) \quad [\text{dB}]$$

waarbij L_I verschil in intensiteitsniveau, in dB
 I intensiteit van geluid, in Watt/m²

¹ De natuurkundige grootheid *kracht* (F) wordt uitgedrukt in eenheden Newton, afgekort N.

Energie (E), ook wel *arbeid* genoemd, is gedefinieerd als kracht maal afstand, met als eenheid

Nm. *Vermogen* (P) is gedefinieerd als energie per tijdseenheid, met als eenheid Nm/s = Watt.

I_0 intensiteit van referentie-geluid, in Watt/m²,
 deze referentie $I_0 = 10^{-12}$ Watt/m²

Voor de relatie tussen geluidsdruk p en niveau L geldt:

I is evenredig met p^2 dus:

$$L_p = 10 \cdot \log\left(\frac{I}{I_0}\right) = 10 \cdot \log\left(\frac{p}{p_0}\right)^2 = 20 \cdot \log\left(\frac{p}{p_0}\right) \quad [\text{dB}]$$

waarbij L_p : verschil in geluidsdrukniveau, in dB
 p : druk van geluid, in Newton/m²
 p_0 : druk van referentie-geluid, in Newton/m²,
 deze referentie $p_0 = 2 \cdot 10^{-5}$ Newton/m².

Uit het bovenstaande volgt dat een intensiteitsniveauverschil van 10 dB gelijk is aan een geluidsdrukniveauverschil van 20 dB. Een verdubbeling van de geluidsdruk (of signaal-amplitude) betekent een toename van het geluidsdrukniveau met 6 dB ($20 \cdot \log(2) \approx 6$). Een verdubbeling van de intensiteit betekent een toename van het intensiteitsniveau met 3 dB ($10 \cdot \log(2) \approx 3$). Op de dB-schaal komt 0 dB per definitie overeen met de gekozen referentiewaarde. Zie voor een uitleg over referentiewaarde het einde van deze paragraaf.

Ook verhoudingen van amplitudes van elektrische spanningen kunnen we analoog aan geluidsdrukken in decibels uitdrukken:

$$H = 20 \cdot \log\left(\frac{U}{U_0}\right) \quad [\text{dB}]$$

waarbij H : verschil in amplitudeniveau, in dB
 U : amplitude van signaal, in Volt
 U_0 : referentie-amplitude (zelf gekozen), in Volt

In dit geval wordt als referentie vaak het maximale uitsturingniveau van een recorder gebruikt waarbij nog net geen vervorming optreedt. U is dan dus kleiner dan U_0 , waardoor H negatieve dB-waarden geeft.

In de akoestiek is voor de referentie-geluidsdruk $p_{\text{ref}} = 2 \cdot 10^{-5}$ Newton/m² gekozen. Deze waarde is het laagste geluidsdrukniveau dat mensen met een normaal gehoor nog kunnen waarnemen². We spreken dan over een geluidsdruk van X dB SPL (*sound pressure level*; zie Tabel 1 op blz. 11). Het 0 dB SPL niveau is een fysische referentie. Er bestaat ook een psycho-akoestische referentie genaamd SL (*sensation level*). Deze is afhankelijk van de luisteraar en wordt bepaald door, voor verschillende frequenties, de kleinste geluidsdruk te meten die voor de luisteraar nog waarneembaar is. Deze meting heet een gehoordrempelmeting. Is de gehoordrempel bekend dan kun je een geluidsdrukniveau aanbieden die bijvoorbeeld 60 dB boven SL ligt. Dit kan bv. nuttig zijn als je waarnemingen van slechthorenden wilt vergelijken.

Om je een indruk te geven van verschillende geluidsdrukniveaus volgen een aantal voorbeelden, in Tabel I (p.11).

² Bij dit referentie-geluid verplaatsen de luchtmoleculen zich over een zeer kleine afstand rond hun evenwichtspositie, nl. ongeveer de diameter van één luchtmolecuul! Als onze oren nog iets gevoeliger zouden zijn, dan konden we de random Brownse beweging van luchtmoleculen horen.

2.4 Enkelvoudige signalen

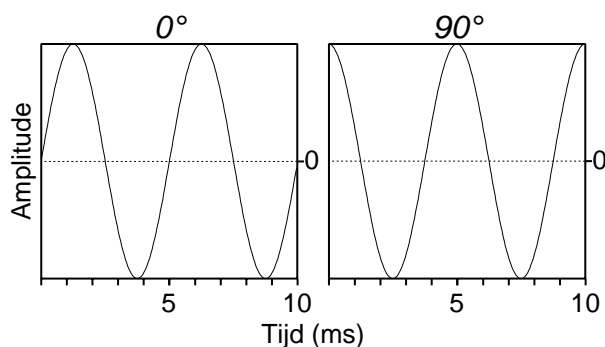
Enkelvoudige signalen als functie van tijd

Het signaal dat het eenvoudigst is om te beschrijven, is de sinus of het enkelvoudige signaal. In je normale omgeving komen sinusvormige geluidssignalen bijna niet voor. Afbeeldingen van sinussen zag je al in Figuur 3 en 4. Om een beschrijving te kunnen geven van een enkelvoudig geluidssignaal, hebben we drie grootheden nodig: naast de in de vorige paragraaf beschreven grootheden frequentie (f) en amplitude (A), de nog niet beschreven grootheid fase (φ).

Tabel 1: Voorbeelden van geluidsdrukniveaus in verschillende omstandigheden.

dB (SPL)	omstandigheden
0	Absolute drempel ³ van een sinustrilling met $f=1000$ Hz
20	Bladgeritsel in het bos
30	Fluisteren
35	Buiten in woonwijk bij nacht
40	Rustige kamer
45	Mechanische typemachine
60	Gemiddelde conversatie (op 1,5 m afstand)
65	Kraaiende haan (op ca. 1,5 m afstand)
80	Schreeuwen, zingen (op 1,5 m afstand); straatlawaaï
100	Perron van de ondergrondse als de trein aankomt; symfonie-orkest
120	Pijngrens; versterkte popmuziek in discotheek
130	Motor van een straalvliegtuig (op 30 m afstand)

Twee sinusvormige signalen die alleen in fase van elkaar verschillen hebben dezelfde frequentie en maximale amplitude, maar zijn ten opzichte van elkaar in de tijd verschoven (zie Figuur 5). Het begrip fase is nog niet in de vorige paragraaf beschreven, omdat fase geen duidelijk perceptief correlaat heeft. Dat betekent dat het menselijk gehoor faseverschillen slecht waarneemt.



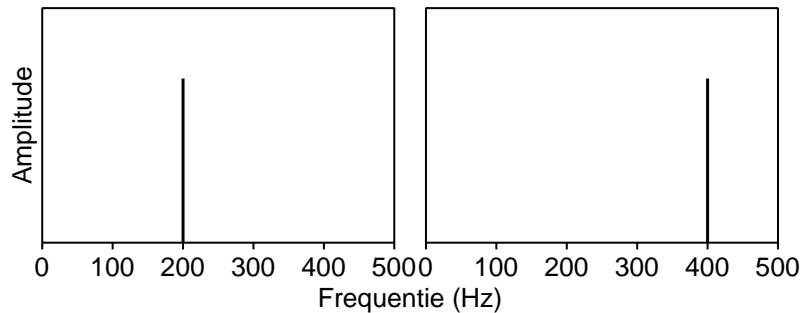
Figuur 5: Twee periodieke signalen, met een fasehoek van respectievelijk 0° en 90° .

Bij een weergave van een (enkelvoudig) signaal als functie van de tijd wordt op de horizontale as de tijd (t) uitgezet en op de verticale as de uitwijking of amplitude (A) van de geluidsdruk. Zo'n weergave van de golfvorm noemt men een oscillogram. Oscillogrammen zijn heel geschikt om duurmetingen aan een spraaksignaal te doen. Alle plaatjes die je tot nu toe hebt gezien, zijn oscillogrammen.

³ Dit is de *minimale* gehoordrempel, die slechts bereikt wordt door 1% van de luisteraars. De *gemiddelde* gehoordrempel, die bereikt wordt door 50% van de luisteraars, ligt op ca. +16 dB. [H. Fletcher (1995) *Speech and Hearing in Communication*. Woodbury, NY: Acoustical Society of America. Originally published: New York: Van Nostrand, 1953. Fig.96, p.135].

Enkelvoudige signalen als functie van frequentie

Van elk repeterend signaal kun je de grondfrequentie berekenen. Ook dit kun je weergeven in een plaatje. Hierin wordt langs de x-as de frequentie (f) en langs de y-as de amplitude (A) uitgezet (de fase blijft buiten beschouwing). Zo'n plaatje heet een spectrum. In een spectrum wordt dus weergegeven welke frequenties in het signaal aanwezig zijn, en wat de amplitude van elk van die frequenties is. In het spectrum van een zuivere sinus komt maar één frequentie voor, daarom zie je maar één "paaltje". In de figuren 6a en 6b zijn spectra afgebeeld van twee sinussen met een verschillende frequentie en een gelijke amplitude.



Figuur 6: Spectra van twee periodieke signalen, met frequenties van respectievelijk 200 Hz en 400 Hz.

Enkelvoudige signalen kunnen dus verschillen in luidheid, toonhoogte en fase. Een zuivere sinus zul je in werkelijkheid nooit tegenkomen. Alle geluiden die je om je heen hoort zijn zogenaamde samengestelde signalen.

Nieuwe begrippen

- periodiek, a-periodiek
- frequentie (f)
- amplitude (A)
- decibel (dB)
- geluidsintensiteit (I)
- geluidsdruk (p)
- logaritme (\log)
- gehoordrempel
- pijngrens
- sinus
- enkelvoudig signaal
- fase (j)
- oscillogram
- spectrum
- samengesteld signaal

VERVOLG HOOFDSTUK 2 BASISTHEORIE SIGNALLEN

2.5 Samengestelde signalen

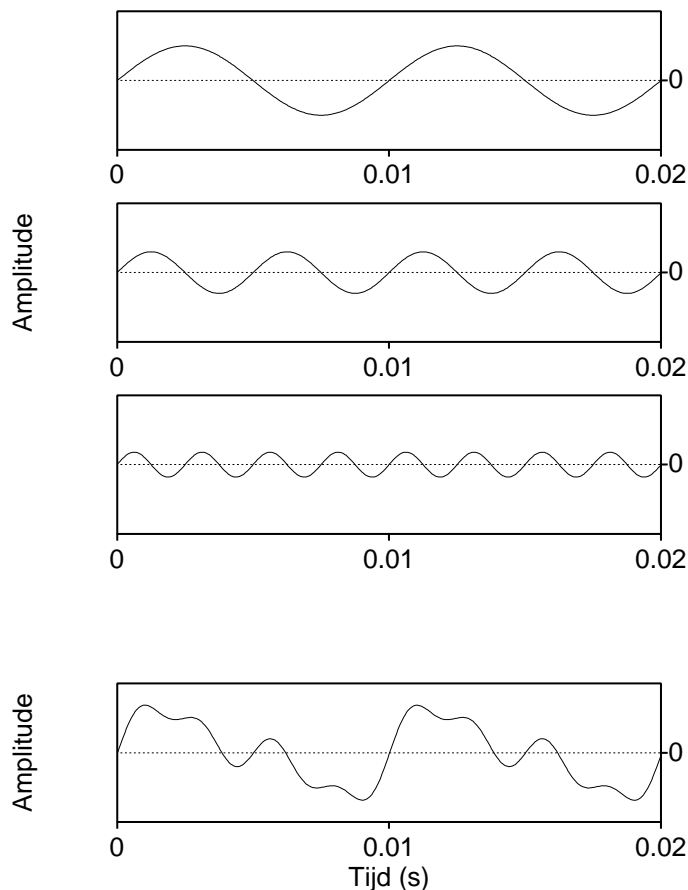
Alle geluiden die geen sinussen zijn noemen we samengestelde geluiden. Alle samengestelde geluiden kunnen beschreven worden als de som van een aantal sinussen. Deze relatie is ontdekt

door de Franse wiskundige, baron J.B.J. Fourier, 1768-1830. De samenstellende sinussen noemen we dan de frequentiecomponenten van die (samengestelde) geluiden.

Als zo'n samengesteld geluid zelf ook weer een golfvorm heeft die zich herhaalt, dan spreken we van een periodiek samengesteld geluid. Daarvan is een voorbeeld gegeven in Figuur 7. We zien daar een periodiek geluid, verkregen door optelling van drie componenten. De frequenties van deze componenten zijn 100 Hz, 200 Hz en 400 Hz. Merk op dat de frequentie waarmee de golfvorm van het samengestelde geluid zich herhaalt (de herhalingsfrequentie), identiek is aan de frequentie van de laagste component, die van 100 Hz.

Die laagste component noemen we nu de grondtoon (f_0) van het samengestelde geluid, de overige componenten noemen we boventonen. Grondtoon en boventonen worden samen ook wel aangeduid met de term harmonischen. De grondtoon is de eerste harmonische, de eerste boventoon is de tweede harmonische etc. Voor ieder periodiek samengesteld geluid geldt dat alle boventonen frequenties hebben die hele veelvouden zijn van de frequentie van de grondtoon.

De eigenschappen van een samengesteld signaal worden bepaald door de amplitudes, frequenties en fasen van de samenstellende sinussen.



Figuur 7: Sinussen met frequenties van respectievelijk 100 Hz (relatieve amplitude 0.5), 200 Hz (amplitude 0.3) en 400 Hz (amplitude 0.2), en het somsignaal van deze drie enkelvoudige signalen.

Klankkleur

Twee geluiden die dezelfde sterkte en toonhoogte hebben, kunnen toch voor het gehoor nog sterk van elkaar verschillen in wat we zouden kunnen noemen het 'karakter' van het geluid. De algemene naam hiervoor is klankkleur of timbre. Aangezien voor een periodiek geluid elke verandering in de amplitudes van de harmonischen leidt tot een nieuwe klankkleur, is de mogelijke variatie in klankkleur zeer groot. Een geluid kan bijvoorbeeld dof klinken of schel, of

warm of metalig. Ook het verschil bijvoorbeeld tussen een /a/ en een /i/ uitgesproken door dezelfde spreker, op dezelfde toonhoogte en luidheid, is een verschil in timbre.

Timbre is niet zoals luidheid en toonhoogte een eendimensionaal verschijnsel. Geluiden kunnen wat betreft luidheid gerangschikt worden op de eendimensionale schaal van zacht naar luid. Wat toonhoogte betreft kunnen ze geplaatst worden op een schaal van laag naar hoog. Klankkleur daarentegen wordt een multi-dimensionele perceptieve grootheid genoemd.

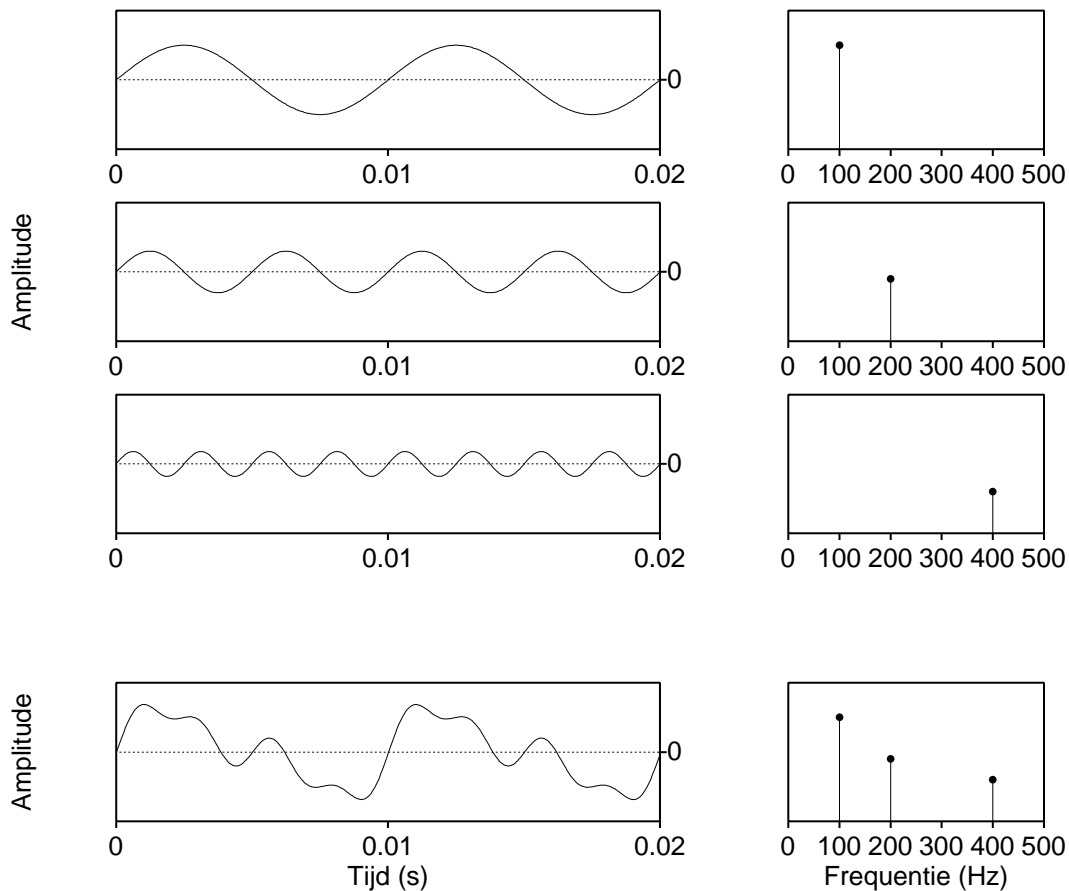
Octaaf

Een verdubbeling van de grondtoon (F_0) betekent een toename van de grondtoon met een octaaf. De verhouding tussen twee opeenvolgende octaven is dus 1:2, bijvoorbeeld 200 en 400 Hz.

2.6 Samengestelde signalen als functie van tijd en frequentie

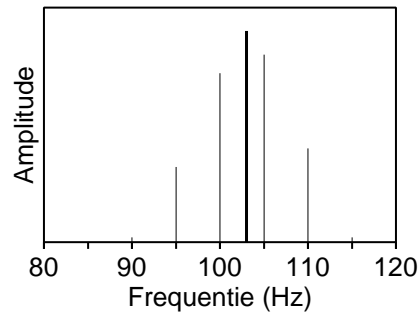
Een samengesteld signaal kan, evenals een enkelvoudig signaal, op twee manieren gerepresenteerd worden: als functie van de tijd of als functie van de frequentie. In Figuur 8 zijn de oscillogrammen van Figuur 7 herhaald. Dit is weergegeven in het tijddomein. In het rechtergedeelte van Figuur 8 is van elke golfvorm het spectrum afgebeeld. Dit is weergegeven in het frequentiedomein.

In het spectrum van het samengestelde signaal zie je voor iedere frequentie-component de frequentie (op de horizontale as) en de amplitude (hoogte van een 'paaltje', d.w.z. plaats op de verticale as). In een spectrum kun je dus zien wat de frequentie en de amplitude zijn van de verschillende harmonischen. Het is dan niet nodig om het signaal op te splitsen in alle sinussen en van elke sinus een oscillogram te tekenen.

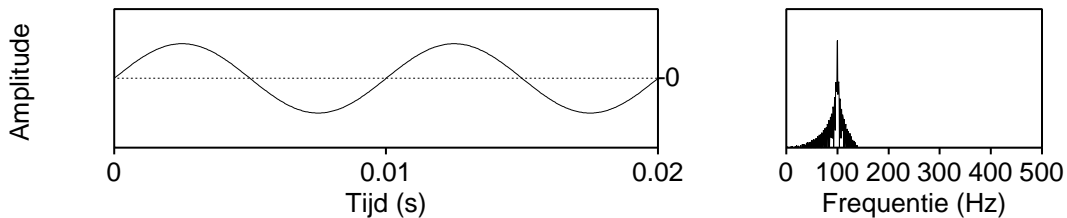


Figuur 8: Oscillogrammen uit Figuur 7, met rechts daarvan de bijbehorende spectra.

Een spectrum wordt in theorie afgebeeld als een rij 'paaltjes' die de aanwezigheid van bepaalde frequentiecomponenten (harmonischen) voorstellen. In de praktijk zal je een dergelijke weergave niet vaak tegenkomen. Harmonischen tonen zich vaak in de gedaante van bergachtige pieken. De oorzaak hiervan ligt in het feit dat computerprogramma's die spectra berekenen moeten uitgaan van een discreet spraak-sigitaal, dat wil zeggen een sigitaal dat bestaat uit een eindige reeks getallen (bemonsteringen van de amplitude-waarden, of 'samples'). Nu wil de theorie dat je in het spectrum precies evenveel punten krijgt in het frequentiedomein (op onderling gelijke afstand) als het aantal meetwaarden dat je bekijkt in het tijdsdomein. Dit houdt automatisch in dat frequenties niet precies kunnen worden weergegeven. Wanneer een frequentiecomponent eigenlijk tussen twee frequentiepunten in ligt, dan wordt de energie van die component weergegeven bij de omliggende frequentiepunten. Deze 'uitsmering' wordt geïllustreerd in Figuur 9. Het dikgedrukte paaltje geeft de frequentie en amplitude van de echte harmonische weer; deze wordt echter alleen gemeten op de omringende punten van het frequentiedomein. In Figuur 10 zie je het daadwerkelijk gemeten spectrum van de bovenste sinus uit Figuur 8.

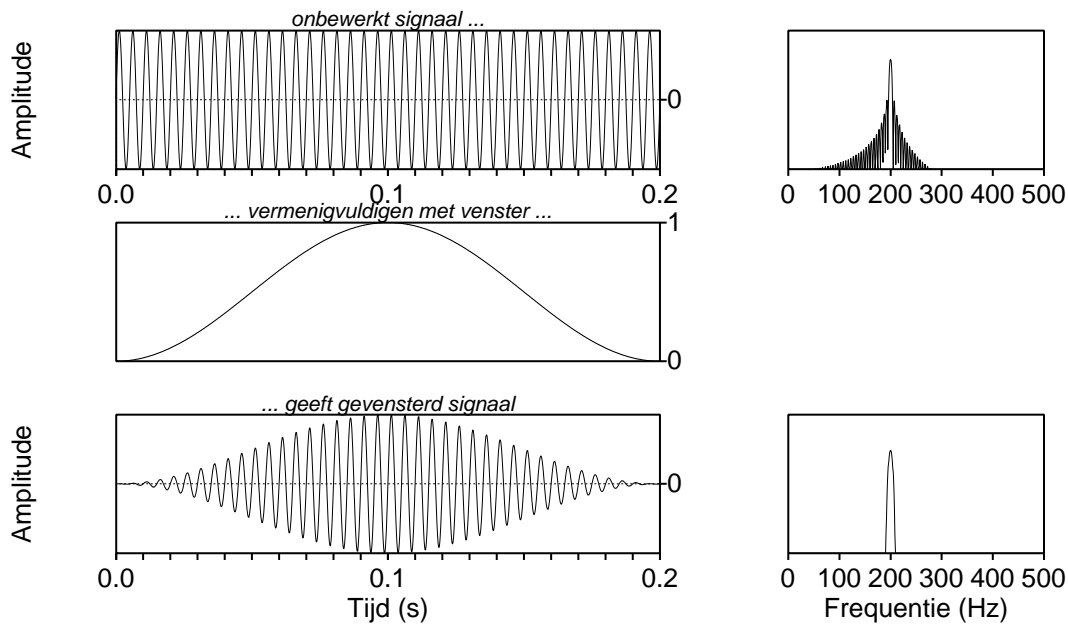


Figuur 9: Uitsmering van een harmonische van 103 Hz over de omringende punten in het frequentiedomein (met een resolutie van 5 Hz).



Figuur 10: Oscillogram van een sinus van 100 Hz, met rechts daarvan het bijbehorende gemeten spectrum. Merk op dat het gemeten spectrum verschilt van het theoretische spectrum in Figuur 8.

In de meeste spraakprogrammatuur wordt voor het berekenen van een spectrum een stukje signaal gebruikt met een standaard tijdsduur (bijvoorbeeld 25 ms). Hoogstwaarschijnlijk zullen de begin- en de eindgrens van het uitgeknipte stukje spraaksignaal niet precies samenvallen met begin en einde van een periode. Het algoritme dat de frequentie-analyse doet verwacht dit echter wel. Het gevolg is dat er frequentiecomponenten in het spectrum worden opgenomen die eigenlijk niet in het signaal voorkomen. Om dit probleem op te lossen wordt gebruik gemaakt van 'venstering' ('windowing'): het signaalsegment wordt vermenigvuldigd met een venster. Dat is een wiskundige functie met de vorm van een heuvel, zoals te zien in Figuur 11. Venstering zorgt ervoor dat het begin en het einde van het stukje signaal op de nul-as komen te liggen. Het segment is daardoor kunstmatig periodiek gemaakt, waardoor de frequentie-analyse beter gaat. Meestal wordt gebruik gemaakt van een zgn. 'Hanning'-window (bedacht door Hanning).



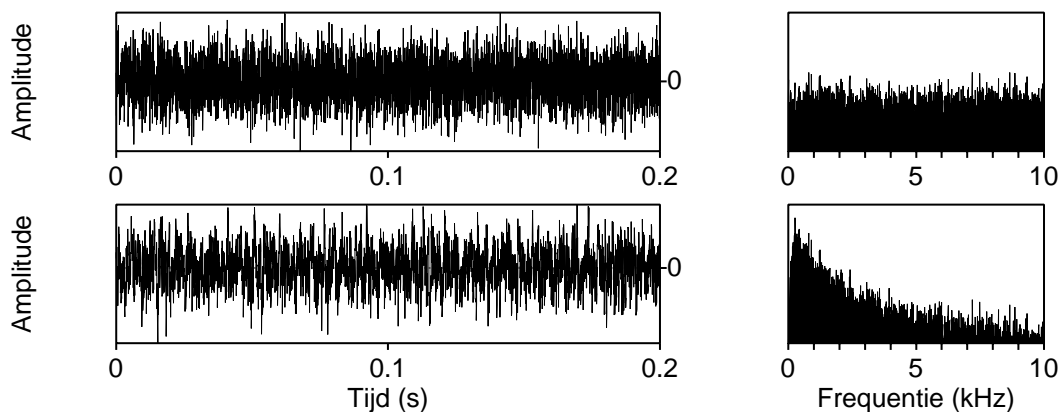
Figuur 11: Een fragment van een onbewerkt signaal (0,2 s van een sinus van 200 Hz) met zijn spectrum, een Hanning-venster (met een duur van 0,2 s), en het resulterende gevensterde signaal-fragment (0,2 s) met zijn spectrum.

Sommige programma's waarmee spectra berekend worden geven alleen de zgn. 'omhullende' van het spectrum weer. Aan de omhullende kun je niet meer zien uit welke afzonderlijke harmonischen het signaal bestaat. Je kunt dan wel in een oogopslag zien welke frequentiegebieden het sterkst vertegenwoordigd zijn. Bij klinkergeluiden in spraak worden deze gebieden 'formanten' genoemd. Hierover meer in §2.9.

2.7 Ruis

We spreken van ruis wanneer een signaal niet periodiek is. Ook ruis is samengesteld uit een aantal spectrale componenten, maar er is geen eenvoudige relatie tussen die componenten. Bij periodieke geluiden zijn de frequenties van de boventonen altijd gehele veelvouden van de frequentie van de grondtoon (zie p.13). Bij ruis-geluiden wordt de signaalfunctie geheel of grotendeels door het toeval bepaald. Ruissignalen hebben theoretisch een vlak spectrum. Dit komt omdat een ruissignaal een oneindig aantal frequentie-componenten bevat.

Twee belangrijke soorten ruis die we kunnen onderscheiden zijn 'witte' en 'roze' ruis. Witte ruis wordt zo genoemd naar analogie met licht, omdat in witte ruis alle frequentiecomponenten evenredig voorkomen. In roze ruis komen meer lage dan hoge frequenties voor (evenals in rood licht). Het spectrum van roze ruis loopt af met 3 dB per octaaf, zoals je kunt zien in Figuur 12. In Figuur 12 is de frequentie-as lineair getekend. De spectrale omhullende van roze ruis zou dan theoretisch een continue en (logaritmisch) dalende lijn moeten zijn. Net als in Figuur 10 wijken de gemeten spectra af van de theoretische spectra. Die afwijking komt vooral doordat het spectrum is bepaald over een eindig aantal samples, resulterend in een eindig aantal gemeten frequentiecomponenten (die in dit tijdsfragment toevallig niet allemaal even sterk zijn).



Figuur 12: Oscillogram en spectrum van witte ruis (boven) en roze ruis (onder).

Natuurlijke ruissignalen hebben vaak een 'wit' spectrum. Bij onze waarneming van witte ruis overheerst de hoogfrequente energie. Deze ruis heeft een scherpe sisklank. Roze ruis is voor onze oren een veel evenwichtiger (prettiger) ruissignaal.

2.8 Spraaksignalen

Spraakgeluiden zijn altijd samengestelde geluiden. Spraakgeluid ontstaat doordat een brongeluid wordt opgewekt dat vervolgens gefilterd wordt. Men spreekt ook wel van bron-filter-model (naar G. Fant). Er zijn in normale spraak twee typen brongeluiden.

Ten eerste het geluid dat ontstaat door het trillen van de stembanden. Dit brongeluid resulteert na filtering in stemhebbende klanken zoals de klinkers en stemhebbende consonanten zoals de **/m,n,b,d,z,l/**.

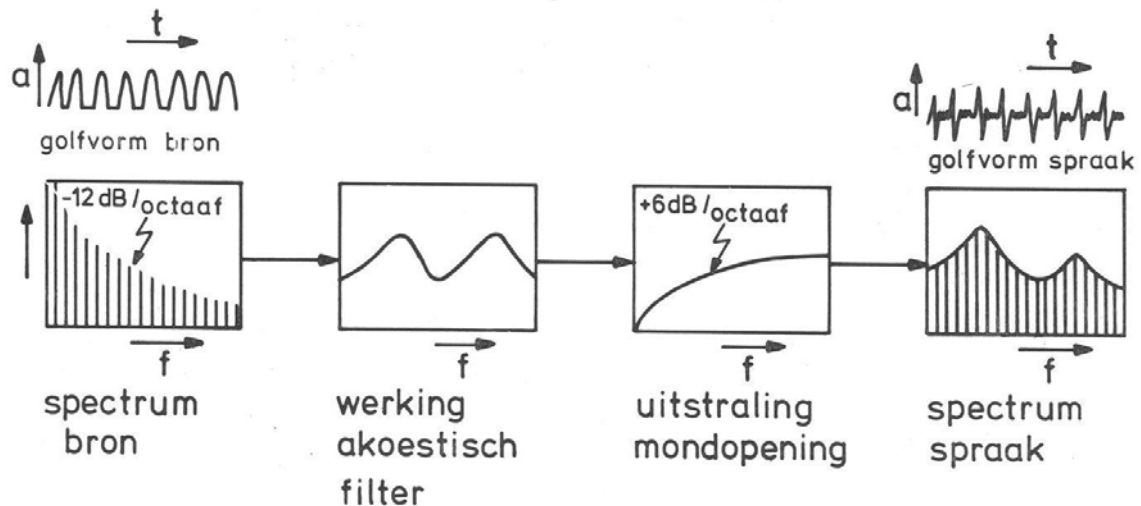
Het tweede soort brongeluid ontstaat niet bij de stembanden, maar in de mondholte. Ruisgeluid ontstaat als de luchtstroom uit de longen door een sterke vernauwing in de mondholte wordt geperst, of als de mondholte tijdelijk volledig wordt afgesloten en daarna plotseling geopend. Klanken met dit niet door de stembanden voortgebrachte brongeluid worden de stemloze klanken genoemd. Dit zijn **/p,t,k,f,s,x/**. De eerste drie zijn plofklanken (plosieven), de andere drie wrijfklanken (fricatieven).

Het periodieke brongeluid van stemhebbende spraakklanken, d.w.z. het stembandgeluid, heeft een spectrale helling van -12 dB/octaaf. Dit betekent dat de hogere frequentiecomponenten zwakker zijn dan de lage. Het (stemband)brongeluid zul je overigens nooit direct horen, omdat dit signaal vervormd wordt door de mond. Dit vervormen heet filteren. Het filter dat gevormd wordt door het spraakkanaal werkt als een 'resonator': sommige frequentiecomponenten worden versterkt, andere verzwakt. De uitstraling van de mond veroorzaakt een algehele versterking van het spraaksignaal (bronsignaal plus filter-effect) met $+6$ dB/octaaf. De uiteindelijke afval van het spraaksignaal (bronsignaal plus filter-effect plus uitstraling) bedraagt dan -6 dB/octaaf.

In het geval van stemhebbende spraakgeluiden wordt het filter gevormd door de gehele mond-keel-neusholte. Het effect van dit filter wordt bepaald door de vorm van de resonantieholtte, d.w.z. door de stand van de articulatie-organen. Denk maar aan de verschillende mondstanden bij de uitspraak van een **/a/** en een **/i/**. Bij een **/a/** is de kaak laag (mond open) en de grootste vernauwing van de tong achter; bij een **/i/** is de kaak hoog (mond gesloten), de lippen gespreid, en de grootste vernauwing van de tong voor in de mond.

Het ontstaan van de stemloze medeklinkergeluiden kunnen we op dezelfde manier beschrijven als het ontstaan van de stemhebbende klanken. Er is weer een verkleurend filter, nu gevormd door de resonantieholtten die zich tussen de ruisvormende vernauwing of afsluiting en de buitenlucht bevindt. Deze resonantieholtte is bijvoorbeeld vrij groot voor de **/x/** vrij klein voor de **/s/**.

Er zijn ook spraakklanken waarbij twee brongeluiden betrokken zijn, ten eerste het geluid dat opgewekt wordt door de stembandtrilling, ten tweede een ruisgeluid dat ontstaat in een vernauwing in de mondholte. Zo'n klank noemen we een stemhebbende wrijfklank. Voorbeelden zijn /v,z/. Het ruisgeluid kan ook ontstaan door afsluiting en plotselinge opening van de mondholte; zo'n klank noemen we een stemhebbende plofklank. Voorbeelden zijn /b,d/. Omdat de luchtstroom uit de longen sterk geremd wordt, door de trillende stembanden, is de ruis die wordt opgewekt bij het maken van stemhebbende wrijf- en plofklanken veel zwakker dan de ruis van stemloze wrijf- en plofklanken.

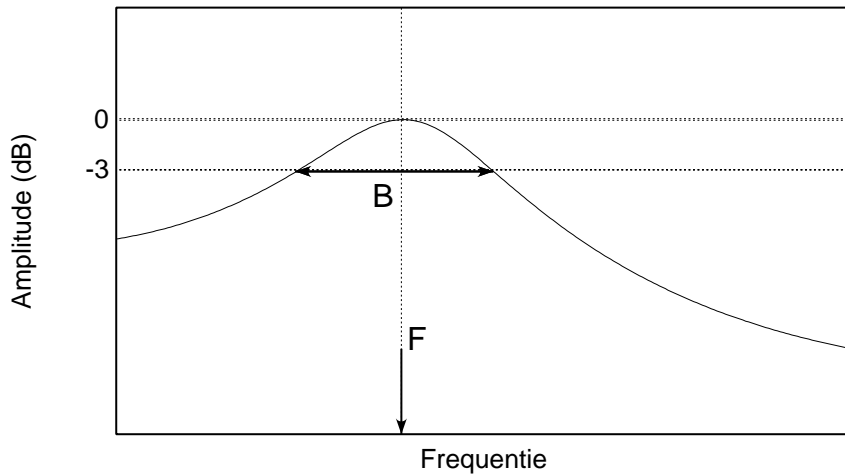


Figuur 13: Het bron-filter-model: het bronsignaal (oscillogram en spectrum), het effect van het resonantiefilter en van de uitstraling, en het uiteindelijke spraaksignaal (oscillogram en spectrum). Ontleend aan: S.G. Nootboom & A. Cohen, *Spreken en Verstaan*, Fig.3.7. (p.59).

2.9 Formanten

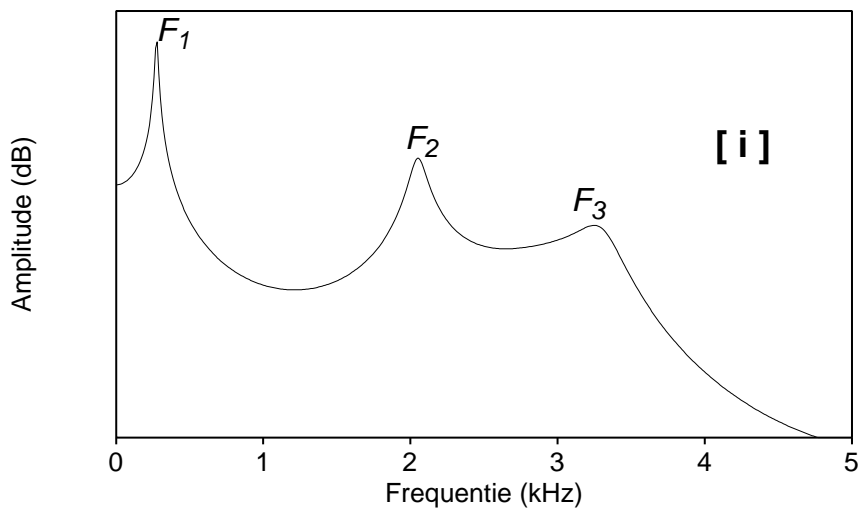
Zoals al in §2.8 is beschreven, worden door de vorm van de mondholte bepaalde frequentiecomponenten verzwakt en andere juist versterkt. Dit heeft gevolgen voor het spectrum. De versterkingen zijn in het spectrum terug te vinden als pieken, de verzwakkingen als dalen. Vooral bij klinkers is de resonantie groot genoeg om echt pieken en dalen te kunnen onderscheiden in het spectrum. De plaatsen van de pieken en de dalen hangen samen met de vorm van het filter en zijn dus afhankelijk van de stand van de spraakorganen. De stand van de spraakorganen wordt voornamelijk bepaald door de klank die de spreker voort wil brengen. Natuurlijk zijn ook verschillen in de bouw van de spraakorganen van belang. Om deze reden zijn de plaatsen van de pieken en de dalen bij vrouwen, mannen en kinderen enigszins verschillend.

De pieken worden formanten genoemd. De formanten worden gekarakteriseerd door hun frequentie, amplitude en bandbreedte. De plaats van de piek op de horizontale as van het spectrum is de formantfrequentie. De maximale hoogte is de formantamplitude. De bandbreedte is de breedte van de formant, gemeten 3 dB onder de maximale amplitude van de piek, zoals geïllustreerd in Figuur 14.

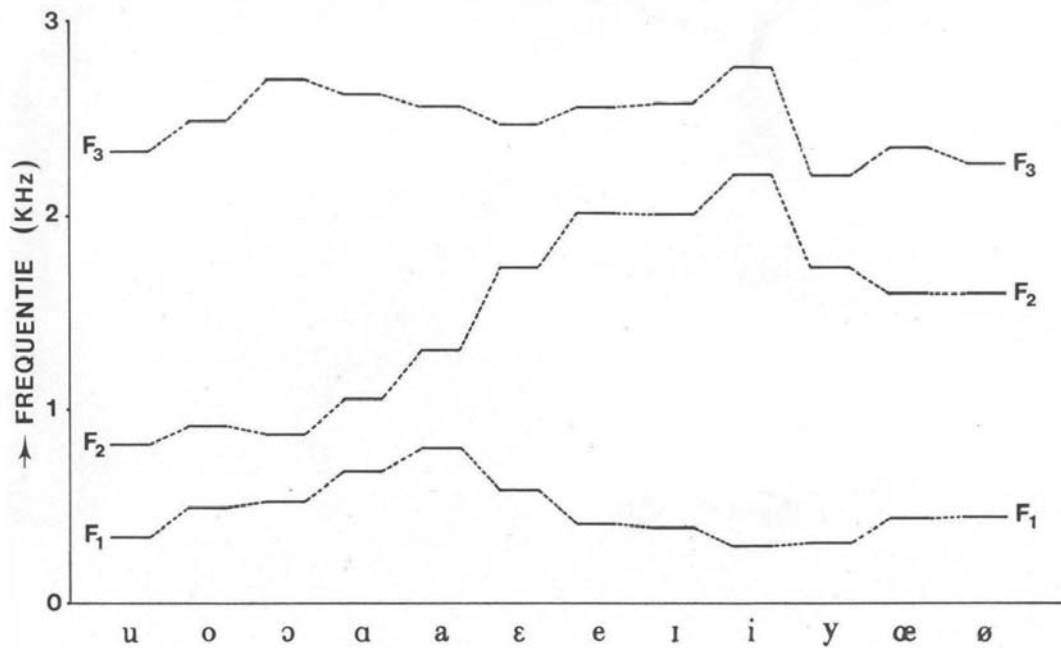


Figuur 14: Spectrale omhullende van een formant, met daarbij aangegeven de piekfrequentie (F) en bandbreedte (B) van de formant.

Formanten worden vooral gebruikt bij de beschrijving van klinkers. Voor de beschrijving wordt het meest gebruik gemaakt van de formantfrequenties. De piek met de laagste frequentie heet de F_1 , de volgende F_2 enzovoort. Vooral de laagste drie formanten zijn verantwoordelijk voor de waargenomen klinkerkleur. In Figuur 15 zie je de spectrale omhullende van de klinker /a/, uitgesproken door een man, met de formanten aangegeven. Merk op dat de formanten verschillende bandbreedtes hebben. Zie Figuur 16 voor de gemiddelde formantfrequenties van de Nederlandse klinkers.



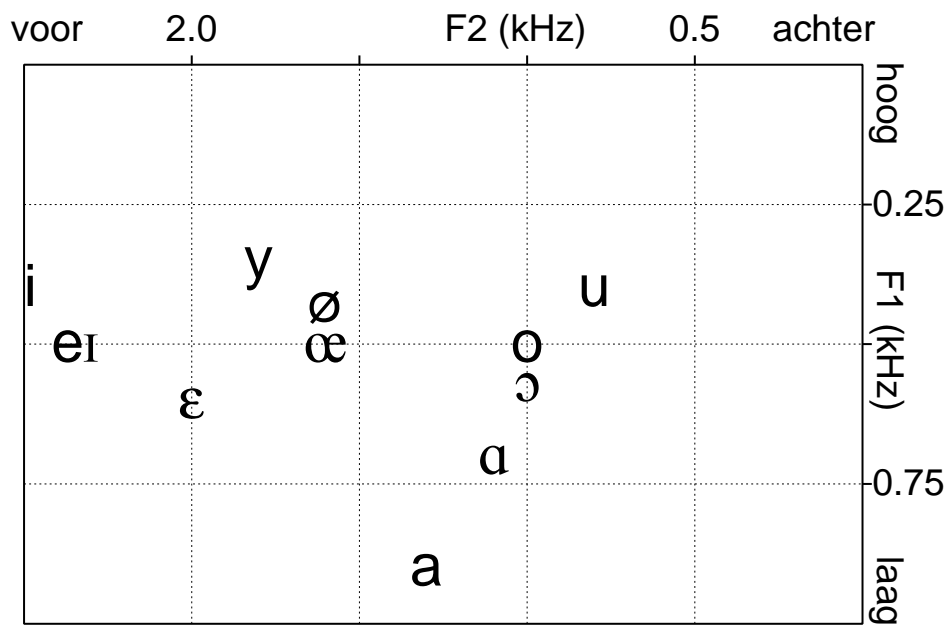
Figuur 15: Spectrale omhullende van de klinker /a/ uitgesproken door een man, met daarbij aangegeven de eerste vier formanten.



Figuur 16: Piekfrequenties van de laagste drie formanten, F1, F2 en F3, van 12 Nederlandse klinkers, gemiddeld over 50 mannelijke sprekers. Ontleend aan: S.G. Nootboom & A. Cohen, *Spreken en Verstaan*, Fig.3.17 (p.75).

De formanten F4 en F5 zijn vooral van belang voor de natuurlijkheid van het spraakgeluid. Nog hogere formanten, F6 enz, zijn in het algemeen niet meer terug te vinden in het spraakgeluid. De formantfrequenties van de verschillende klinkers verschuiven niet als de frequentie van de grondtoon verandert.

Zoals je in Figuur 16 kunt zien, geven de frequenties van de eerste en tweede formant genoeg informatie om de verschillende klinkers van elkaar te kunnen onderscheiden. Het is daarom gebruikelijk om de klinkergeluiden weer te geven in een plat vlak, met de frequentie van F1 langs de verticale as en de frequentie van F2 langs de horizontale as, zie Figuur 17. De klinkers blijken dan alle een plaats te vinden in een driehoek met /a/, /i/ en /u/ als hoekpunten. Deze driehoek wordt de akoestische klinkerdriehoek genoemd.



Figuur 17: Akoestische klinkerdriehoek, met daarin aangegeven de 12 Nederlandse klinkers uit Figuur 16.

Nieuwe begrippen

- grondtoon, boventonen, harmonischen
- klankkleur, timbre
- spectrale omhullende
- witte ruis, roze ruis
- bron-filter-model
- klinker
- stemhebbend, stemloos
- formant, formantfrequentie, formantamplitude
- bandbreedte
- klinkerdriehoek

Opdrachten hoofdstuk 2

Bij de opdrachten van deze week wordt gebruik gemaakt van een aantal spraakbestanden ("files"). Deze bestanden kun je vinden op www.let.uu.nl/~audiufon/data/pracfon1.html. De bestanden moet je kopiëren naar je eigen account. Hoe dat moet vind je op dezelfde internetpagina. Bewaar alle bestanden in je data-directory.

Voor een gebruiksaanwijzing van een spraakprogramma kun je kijken op www.let.uu.nl/~Hugo.Quene/personal/faq/spraakprogr.html.

<u>File</u>	<u>Inhoud</u>
norma	klinker /a/, normaal uitgesproken door een man
normi	klinker /i/, normaal uitgesproken door een man
normu	klinker /u/, normaal uitgesproken door een man
norme	klinker /e/, normaal uitgesproken door een man
normy	klinker /y/, normaal uitgesproken door een man
zinga	gezongen /a/ (man)
zingi	gezongen /i/ (man)
fluisa	gefluisterde /a/ (man)
fluisi	gefluisterde /i/ (man)
hooga	klinker /a/, hoog uitgesproken door een man
hoogi	klinker /i/, hoog uitgesproken door een man
witte ruis	een stukje ruis
zin1	gesproken zin "Het leven is mooi als de zon schijnt"
wiekust.aifc	gesproken ambigue zin "Kees zei Koos kust met het buurmeisje van Toos"
woord1.aifc	deel van ambigu woordpaar
woord2.aifc	het andere deel van hetzelfde ambigue paar

Opdracht 1

a. Maak sinussignalen aan met de volgende frequenties en amplitudes:

- (I) frequentie 440 Hz en relatieve amplitude 0.25 (een kwart van de maximale amplitude),
- (II) frequentie 1200 Hz en amplitude 1 (maximale amplitude). Sla dit bestand op als `sin1200.aifc`,
- (III) frequentie 1200 Hz en amplitude 6 dB lager dan (II),
- (IV) als (I), maar met frequentie twee octaven hoger, en amplitude 6 dB hoger.

b. Luister ze af, en vergelijk ze. Druk alle sinussen af voor in het verslag.

c. Hoeveel is de amplitude van de sinus uit opdracht (III)?

d. Wat zijn de frequentie en de amplitude van de sinus uit opdracht (IV)?

Opdracht 2

a. Maak spectra van de vier sinussen die je in opdracht 1 hebt gemaakt. Druk deze af.

b. Beschrijf de verschillen tussen de spectra.

Opdracht 3

(I) Vraag bestand `sin1200.aifc` op. **Maak een kopie van dit bestand.** Open het bestand in een (golfvorm)editor. Selecteer een halve periode, in het midden van het signaal (bij 0.5 s), lopend vanaf een maximum tot een minimum. Verwijder dit fragment. Noteer de cursor-positie van het snijpunt. Sla het resultaat op onder een nieuwe naam.

(II) Vraag `sin1200.aifc` op. **Maak een kopie van dit bestand.** Open het bestand in een (golfvorm)editor. Verwijder weer een halve periode in het midden van het signaal, nu lopend vanaf een positieve tot een negatieve nuldoorgang. Noteer weer de cursor-positie van het snijpunt. Sla het resultaat op onder een nieuwe naam.

(III) Vraag `sin1200.aifc` op. **Maak een kopie van dit bestand.** Open het bestand in een (golfvorm)editor. Verwijder nu een hele periode in het midden van het signaal, lopend vanaf een positieve nuldoorgang tot de volgende positieve nuldoorgang. Noteer weer de cursor-positie van het snijpunt. Sla het resultaat op onder een nieuwe naam.

a. Maak afdrukken van de golfvormen. Kies hierbij een inzichtelijk deel van de tijdsas, en een tijdsduur van ongeveer 2 ms..

- b. Beluister de drie bewerkte sinus-signalen, en maak spectrogrammen. Druk deze spectrogrammen af
- c. Wat valt je op?

Opdracht 4

Vraag `zin1.aiff` op in een golfvorm-editor. Probeer de verschillende woorden en klanken terug te vinden in het signaal. Geef de verschillende spraakklanken aan met labels (zie handleiding spraakprogrammatuur op internet). Doe dit met behulp van oren en ogen. Geef aan welke klanken stemhebbend en welke stemloos zijn. Druk je resultaat af.

HOOFDSTUK 3 DIGITAAL GELUID

3.1 Inleiding

In het vorige hoofdstuk hebben we de belangrijkste eigenschappen van (spraak)geluid en de weergave van die eigenschappen beschreven. Om die eigenschappen te bepalen, is het nodig dat de spraak op de een of andere manier is opgeslagen. In dit hoofdstuk wordt besproken hoe (spraak)geluid vastgelegd kan worden op band en opgeslagen kan worden in de computer.

3.2 Opnemen en vastleggen

Geluiden kunnen alleen langs elektronische weg opgeslagen en later weer hoorbaar gemaakt worden. Voor het opnemen van (spraak)geluid wordt gebruik gemaakt van microfoons. De microfoon zet luchtdrukverschillen om in elektrische spanning, een zogenaamde wisselspanning. Wanneer de uitgang van de microfoon is aangesloten op de ingang van een recorder, wordt dit signaal vastgelegd op band, cassetteband of DAT-band.

Er zijn verschillende soorten microfoons. In de practicumruimte ligt een *Instrumentariumboek*, waarin allerlei apparaten uitgebreid beschreven staan. We gaan er hier dan ook niet verder op in.

Geluidsdruk en elektrische spanning zijn grootheden die wel snel, maar niet sprongsgewijs kunnen variëren. Een spanning die oploopt van 0V naar 5V moet alle tussenliggende waarden doorlopen, hoe kort elk van die waarden ook aangenomen wordt. Zulke continue signalen worden analoog genoemd, en de bewerkingen die er op toegepast worden staan bekend onder de term analoge signaalverwerking. Bandrecorders, cassetterecorders en versterkers zijn voorbeelden van apparaten die signalen analoog opslaan of bewerken.

3.3 Opslaan in de computer

Bij het werken met spraak is een computer een onmisbaar hulpmiddel. Echter, een computer kan geen analoge signalen verwerken, omdat een computer alleen met getallen werkt; een analoog signaal moet beschouwd worden als een oneindige reeks getallen. Om toch met geluid te kunnen werken, wordt een signaal in een computer weergegeven als een reeks getallen. Dit heet dan een discreet of digitaal signaal. De getallen waarmee een spraaksignaal beschreven wordt, zijn voor de computer binaire getallen. Om deze reden is enige kennis omtrent het binaire getallenstelsel vereist.

3.4 Het binaire stelsel

Het decimale stelsel heeft 10 als grondtal. Dit getal is gelijk aan het aantal beschikbare cijfers (0 t/m 9). In een decimaal stelsel kan elk getal weergegeven worden door sommatie van machten van 10, bijvoorbeeld:

$$135 = 100 + 30 + 5 = 1(10^2) + 3(10^1) + 5(10^0)$$

waarbij $10^2=10 \cdot 10=100$, $10^1=10$, en per definitie $10^0=1$.

Een meer algemene vorm van deze notatie is als volgt:

$$X = \dots + N(10^N) + \dots + D(10^3) + C(10^2) + B(10^1) + A(10^0)$$

waarbij N, D, C, B, A, enz. de waarden 0 t/m 9 kunnen aannemen.

We kunnen echter een getal ook in het binaire stelsel weergeven. Elk getal bestaat uit een combinatie van de cijfers 0 en 1. Het grondtal, dat gelijk is aan het aantal beschikbare cijfers, is dus 2. Een getal wordt dan weergegeven als een sommatie van machten van 2. Het cijfer 1 geeft aan dat de betreffende macht wel (1 keer) voorkomt en het cijfer 0 geeft aan dat de betreffende macht niet (0 keer) voorkomt.

In formule:

$$X = \dots + N(2^N) + \dots + C(2^3) + B(2^2) + A(2^1)$$

waarbij N, C, B, A, enz. ieder alleen de waarden 0 of 1 kunnen aannemen.

In het tweetalig stelsel wordt 135 dus:

$$\begin{aligned} 135 &= 128 + 0 + 0 + 0 + 0 + 4 + 2 + 1 \\ &= 1(2^7) + 0(2^6) + 0(2^5) + 0(2^4) + 0(2^3) + 1(2^2) + 1(2^1) + 1(2^0) \\ 135 \text{ (decimaal)} &= 10000111 \text{ (binair)} \end{aligned}$$

In het getal 10000111 heet elke 0 of 1 een bit (van 'binary digit'). In het tweetalig stelsel is 135 dus een 8-bits getal. In het tientallige stelsel zijn er drie cijfers nodig om 135 op te schrijven. In het tweetalig stelsel kost het 8 bits. Het maximale getal dat je kunt weergeven in een 3-cijferig decimaal getal = 999 (decimaal) = $10^3 - 1$. Het maximale getal dat je kunt weergeven in een 8-bits binair getal = 11111111 (binair) = $2^8 - 1 = 255$ (decimaal). Het maximale getal dat je kunt weergeven in een 16-bits getal = 1111111111111111 (binair) = $2^{16} - 1 = 65535$ (decimaal) — zie voetnoot ⁴.

Om een getal om te rekenen van het decimale naar het binaire stelsel, zoek je dus eerst de grootste macht van 2 die kleiner is dan het om te rekenen getal. Je schrijft een 1 op. Als de volgende macht van 2 in het restgetal past, schrijf je weer een 1 op, zo niet, een 0. Zo ga je door tot je het hele getal op hebt gedeeld in machten van 2.

Ter vergemakkelijking een lijstje met machten van 2:

$n=$	0	1	2	3	4	5	6	7	8	9	16
$2^n=$	1	2	4	8	16	32	64	128	256	512	65536

En tot slot volgen hier nog een paar voorbeelden:

$$\begin{aligned} 279 \text{ (dec.)} &= 100010101 \text{ (bin.)} \\ 279 &= 1(256) + 0(128) + 0(64) + 0(32) + 1(16) + 0(8) + 1(4) + 0(2) + 1(1) \end{aligned}$$

$$\begin{aligned} 11101000 \text{ (bin.)} &= 1(128) + 1(64) + 1(32) + 0(16) + 1(8) + 0(4) + 0(2) + 0(1) \\ &= 128 + 64 + 32 + 8 = 232 \text{ (dec.)} \end{aligned}$$

3.5 A/D-en D/A-conversie

Om een analogoog signaal om te zetten in voor de computer hanteerbare getallen wordt gebruik gemaakt van een Analoog/Digitaal-omzetter (A/D-converter). Deze converter bepaalt in een vast ritme steeds de spanning van het analoge signaal en zet dit om in een getal dat evenredig is met de gemeten spanning. Dit proces heet digitaliseren of bemonsteren. Een Digitaal/Analoog-omzetter (D/A-omzetter) doet precies het omgekeerde.

Het door A/D-conversie verkregen digitale signaal is een benadering van de originele analoge golfvorm. Er zijn twee aspecten die de nauwkeurigheid van de bemonstering bepalen. Dit zijn het aantal metingen dat gedaan wordt per tijdseenheid: de bemonsteringsfrequentie, en het

⁴ In alle getallenstelsels moet ook het getal 0 opgeslagen kunnen worden. Inclusief de nul kan je dus resp. $10^3=1000$, $2^8=256$ en $2^{16}=65536$ *verschillende* getallen opslaan.

aantal bits dat beschikbaar is voor de beschrijving van de amplitude van het digitale signaal: de amplitude-resolutie.

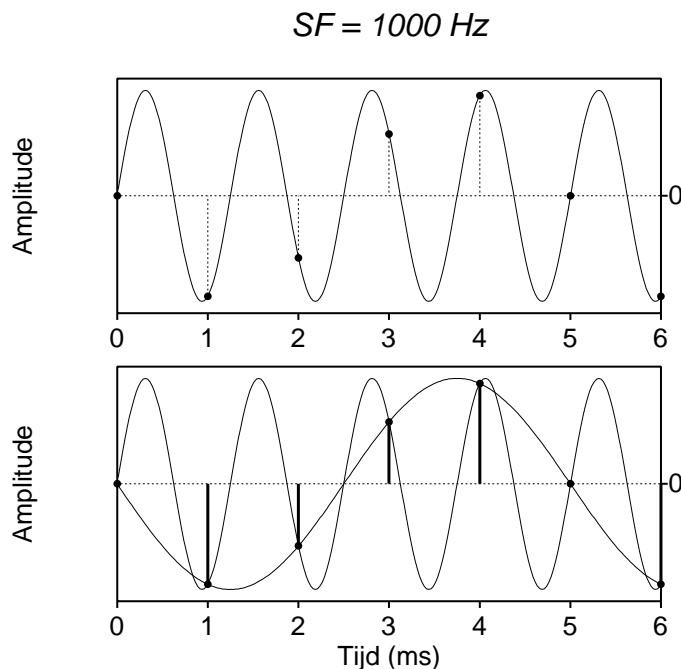
3.6 Bemonsteringsfrequentie

Hoe meer metingen per tijdseenheid gedaan worden, hoe beter de benadering van het originele signaal is. Wanneer het tijdsinterval tussen de momenten waarop de A/D-converter de waarde van het analoge signaal bepaalt 1 seconde is, zeggen we dat de bemonsteringsperiode 1 seconde is ('sampling time', T_s). Meestal echter, spreken we over de bemonsteringsfrequentie ('sampling frequency', f_s). We bedoelen dan het aantal malen per seconde dat de A/D-converter de waarde van het analoge signaal bepaalt. De formule voor de bemonsteringsfrequentie is

$$f_s = 1 / T_s$$

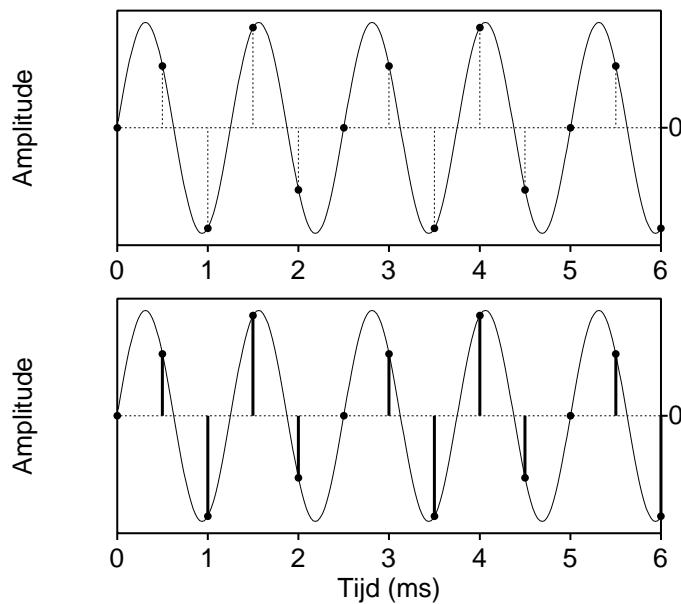
De f_s wordt opgegeven in Hz. Een f_s van 2 kHz betekent dus dat het signaal 2000 keer per seconde wordt bemonsterd.

Het gebruik van een te lage bemonsteringsfrequentie heeft niet alleen negatieve gevolgen voor de nauwkeurigheid van de omzetting, maar kan zelfs leiden tot een verkeerde representatie van het originele signaal. Een gulden regel hierbij is dat de bemonsteringsfrequentie altijd minimaal twee keer zo hoog moet zijn als de hoogste in het signaal voorkomende frequentiecomponent. Als de bemonsteringsfrequentie te laag is, gaan hogere frequenties zich voordoen als lagere frequenties. In Figuur 18-a zie je wat het effect is van een te lage bemonsteringsfrequentie. Hetingangssignaal is een enkelvoudige toon (sinus) met frequentie f (met schending van de gulden regel: niet $f_s > 2f$). Het signaal in Figuur 18-a blijkt na bemonstering niet meer de juiste gedaante te hebben. Het resulterende digitale signaal lijkt wel op het origineel, maar er is een andere frequentie-component bijgekomen, met een frequentie van $f_s - f$. Deze component wordt een spiegelfrequentie genoemd (Engels: 'alias frequency'). In Figuur 18-b, waar met een hogere frequentie wordt bemonsterd (waardoor nu wel $f_s > 2f$), correspondeert de resultante keuring met het aangeboden signaal.



Figuur 18-a: Boven: Oscillogram van eeningangssignaal, een sinus met $f=800$ Hz, bemonsterd met $f_s=1000$ Hz. Onder: De meetwaarden komen tevens overeen met een spiegelfrequentie met $f_a=200$ Hz. Deze spiegelfrequentie is een hoorbare component in het resulterende signaal.

$SF = 2000 \text{ Hz}$



Figuur 18-b: Boven: Oscillogram van een ingangssignaal, een sinus met $f=800$ Hz, bemonsterd met $f_s=2000$ Hz. Onder: De alias met $f=200$ Hz “past niet meer” bij de meetwaarden. De resultante bevat alleen de frequentie-component van het ingangssignaal.

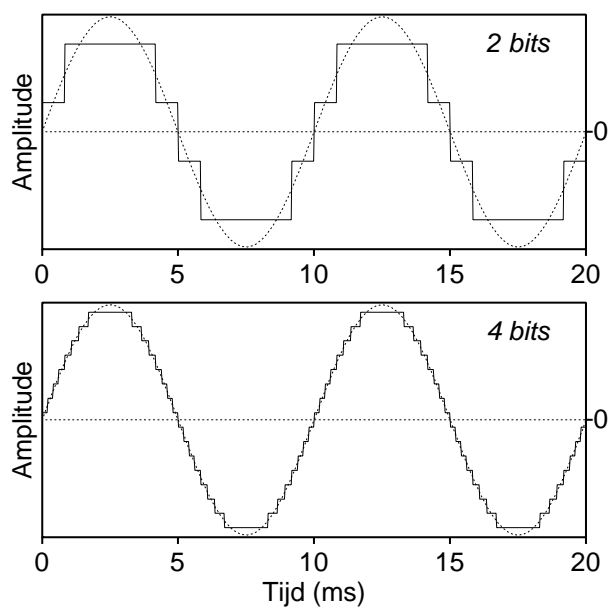
In de praktijk wordt de gulden regel ($f_s > 2 \cdot f_{\max}$, bemonsteringsfrequentie is groter dan tweemaal de hoogste frequentie in het signaal) afgedwongen door het te bemonsteren signaal eerst te filteren met een laagdoorlaat-filter, alvorens het wordt bemonsterd (zie §4.1). De afsnijfrequentie van het filter wordt dan ingesteld iets onder de helft van de bemonsteringsfrequentie (de zgn. nyquist-frequentie, f_n).

Een gebruikelijke bemonsteringsfrequentie voor spraak is 10 kHz. De reden hiervoor is dat men er in de praktijk vanuit gaat dat de belangrijkste informatie (F1 t/m F5, zie §2.9) van het spraaksignaal zich bevindt in het gebied tot 5 kHz. Tegenwoordig wordt voor spraak ook wel gebruik gemaakt van een bemonsteringsfrequentie van 16 kHz of nog hoger.

Aan muziek worden echter veel hogere kwaliteitseisen gesteld. Ter vergelijking: voor CD's wordt meestal een bemonsteringsfrequentie van 44,1 kHz gebruikt, omdat de hogere frequenties in muziek veel belangrijker zijn dan bij spraak.

3.7 Amplitude-resolutie

Het aantal bits dat een A/D-converter beschikbaar heeft voor een zo nauwkeurig mogelijke benadering van het analoge signaal wordt aangeduid met de term ‘amplitude-resolutie’. Met een 2-bits converter kan het analoge signaal in slechts $2^2 = 4$ niveaus worden verdeeld. Dit betekent dat het digitale signaal maximaal 4 amplitude-waarden kan aannemen. Met een 4-bits A/D-converter zijn er 16 niveaus beschikbaar en bij een 12-bits converter 4096. Elk toegevoegd bit verbetert de geluidskwaliteit (signaal/ruis-verhouding) met 6 dB.



Figuur 19: Oscillogrammen van een ingangssignaal (gestippelde curve) en het bemonsterde signaal (doorgetrokken curve). *Boven:* Na bemonstering met amplitude-resolutie van 2 bits (4 mogelijke amplitude-waarden). *Onder:* Na bemonstering met amplitude-resolutie van 4 bits (16 mogelijke amplitude-waarden).

Digitalisatie levert altijd afronding van de amplitude op. Elke amplitude-waarde die tussen twee met behulp van het aantal beschikbare bits weer te geven amplitude-waarden ligt wordt naar boven of naar beneden afgerond. Deze afronding zorgt voor (hoorbare!) ruis. In de praktijk is het digitale signaal van een 8-bits converter visueel (in een oscillogram) al niet meer te onderscheiden van het origineel. Bij Fonetiek wordt standaard gewerkt met 16-bits converters. Je zal dus niet te maken krijgen met problemen door een te lage resolutie. Om deze reden gaan we in dit *Werkboek* er niet verder op in.

3.8 De DAT-recorder

Naast analoge opname-apparatuur, is er ook digitale opname-apparatuur; de DAT-recorder. DAT staat voor *Digital Audio Tape*. Dit is kwalitatief het beste bandopname-apparaat dat momenteel beschikbaar is.

De DAT-recorder slaat analoge signalen digitaal gecodeerd op. Het ingangssignaal wordt in de recorder meteen bemonsterd en als getallen (digitaal) op de band geschreven. Ook hier is de kwaliteit weer afhankelijk van de resolutie en de sampling-frequentie van de ingebouwde converter. De amplitude-resolutie van de A/D- en de D/A-converter in de recorder is 16 bits. De bemonsteringsfrequentie is 48 kHz. Het digitale signaal van de tape kan direct in de computer worden ingelezen, d.w.z. in digitale vorm. Dat resulteert dan in een stereo geluidsbestand met bemonsteringsfrequentie van 48 kHz. Voor meer informatie verwijzen we je naar het *Instrumentariumboek*.

Nieuwe begrippen

- analoog
- digitaal
- bit
- Analoog/Digitaal-omzetter ('A/D-converter')
- Digitaal/Analoog-omzetter ('D/A-converter')
- bemonsteringsfrequentie ('sampling frequency', f_s)
- nyquist-frequentie, f_n
- afsnij-frequentie
- amplitude-resolutie
- spieglfrequentie, 'alias frequency'

Opdrachten hoofdstuk 3

Opdracht 5

- (I) Luister de vijf normaal uitgesproken klinkers die in je account staan af. Knip van elke klinker het begin en het eind af, zodat je een stuk over houdt met een constante amplitude, en constant spectrum. Dit stuk heet de 'steady-state' van de klinker. Let bij het knippen op de nuldoorgangen! Sla de verschillende steady-states op in aparte bestanden (met de extensie `.st.aifc`).
- Maak spectra van de vijf klinkers. Druk de spectra af.
 - Bepaal van elke klinker de F1 en de F2 en zet de waarden tegen elkaar uit (F₂ op X-as).
 - Wat is het resultaat? Laat hierbij zien dat je de overeenkomsten tussen de plaats van de klinkers in de klinkerdriehoek en de vernauwing in het spraakkanaal snapt.

Opdracht 6

- Luister het bestand af waarin een stukje ruis staat. Bekijk het in een (golfvorm)editor. Tel het aantal nuldoorgangen.
- Tel ook het aantal nuldoorgangen van een even lang fragment van de normaal uitgesproken /a/.
- Wat constateer je?
- Hoe zijn de verschillen te verklaren?

Opdracht 7

- Maak opnames van de woorden, zinnen en lettergrepen die op de volgende pagina staan. Zorg hierbij voor voldoende uitsturing op de recorder! Laat zin 2 zowel door een man als door een vrouw uitspreken.
- Zet de opnames in de computer en sla ze op in bestanden. Maak een apart bestand (file) voor elke groep opnames. Geef de bestanden zinnige namen.

Opdracht 8

- Segmenteer de los opgenomen syllaben uit het syllabenbestand (zonder stiltes) en plak het woord *delicatesse*. Label de verschillende syllaben. Sla dit op en noem het `delicatesse1.aifc`. Druk het oscillogram met de labels af.
 - Segmenteer vervolgens de hieronder onderstreepte syllaben uit zin 5:
"De degelijke linkse politicus had genoeg van het kabaal van zijn tekkel."
Knip op basis van wat je ziet en ook op basis van wat je hoort. Let op nuldoorgangen! Ook hiermee plak je weer het woord *delicatesse* en sla dit op onder de naam `delicatesse2.aiff`. Vergeet de labels niet. Druk oscillogram en labels af.
 - Segmenteer ook het normaal uitgesproken woord *delicatesse* uit je woordenbestand en noem deze `delicatesse3.aiff`. Zet ook hier labels bij en druk het af.
- Beluister nu de drie versies. Waarom klinken 1 en 2 zo heel anders dan 3? Bespreek de verschillen in termen van formanten, temporele organisatie etc.

Opnames

Zinnen:

- De delicatesse zal door de minister(2) worden opgegeten.
- Deze fotograaf maakt altijd dia's(5).
- Wie heeft er nog een tafel nodig(9)?
- Voor mij heeft(3) hij koffie meegebracht.
- De degelijke linkse politicus had genoeg van(6) het kabaal van zijn(7) tekkel.
- De(1) minister heeft de dia's van zijn vader(8) nodig.
- De(4) boer had een baaldag.

Woorden: (spreek ze los van elkaar uit!)

- | | |
|-----------|---------|
| 1. senaat | 8. de |
| 2. gebied | 9. zijn |

- | | | | |
|----|----------|-----|-------------|
| 3. | gezoet | 10. | delicatesse |
| 4. | piloot | 11. | heeft |
| 5. | beleid | 12. | vader |
| 6. | schavuit | 13. | van |
| 7. | heraut | 14. | minister |
| | | 15. | nodig |
| | | 16. | dia's |

Syllaben:

- | | |
|------|----------------|
| /ka/ | als in "kaars" |
| /s^/ | als in "suf" |
| /tɛ/ | als in "tent" |
| /li/ | als in "lied" |
| /de/ | als in "deeg" |

Klinker: (aangehouden /a/)

1. normaal gesproken
2. gezongen (hoog)
3. gezongen (laag)
4. gefluisterd
5. luid gesproken
6. zacht gesproken

HOOFDSTUK 4

ANALYSE VAN SPRAAK

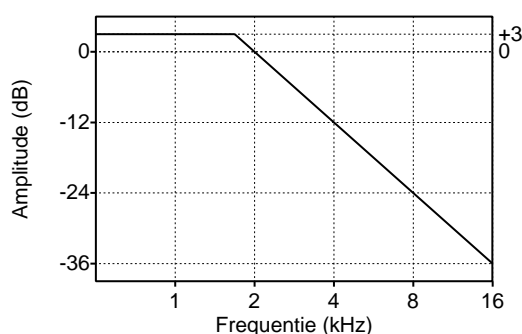
4.1 Filters

Het is mogelijk een signaal door een apparaat te sturen dat slechts een deel van het spectrum van dat signaal doorlaat. Zo'n apparaat heet een filter. Bij het werken met en het analyseren van spraak wordt veelvuldig gebruik gemaakt van filters. We zagen hiervan al een voorbeeld op p.29: bij AD-conversie worden frequenties boven de nyquist-frequentie weggefilterd (bv. boven 5 kHz), om aliasing te voorkomen.

Filters worden meestal gebruikt om frequentiegebieden weg te filteren of te verzwakken: frequentiefilters. Dit betekent dat de amplitude in die frequentiegebieden verzwakt wordt. De mondkeelholte is ook een filter. Denk maar aan de spectra van de verschillende klinkers, waarbij steeds verschillende frequentiegebieden versterkt of verzwakt worden (bron-filter-model, zie §2.8).

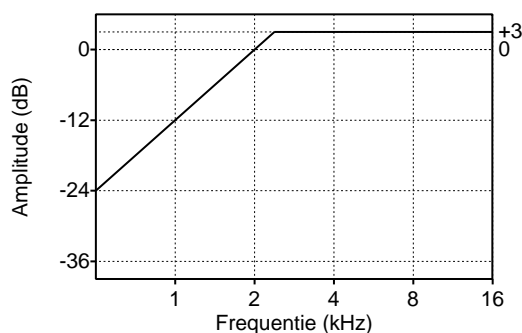
We onderscheiden vier basale typen frequentiefilters. In de volgende figuren zijn de zgn. frequentiekaracteristieken of doorlaatkaracteristieken afgebeeld van de verschillende filters. Een frequentiekaracteristiek van een filter geeft weer welke frequenties door een filter worden doorgelaten en welke frequenties worden verzwakt.

Ten eerste zijn er filters die lage frequenties doorlaten, en hoge frequenties verzwakken. Dit zijn laagdoorlaatfilters (low-pass), zie Figuur 20.



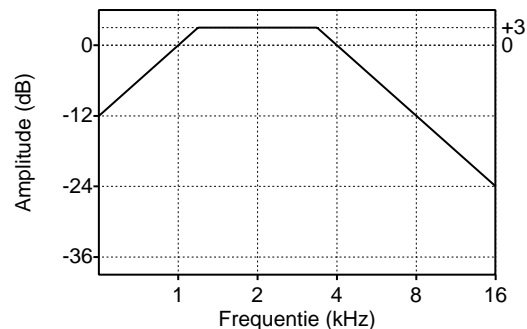
Figuur 20: Geïdealiseerde frequentie-karakteristiek van een laagdoorlaat-filter ('low-pass') met een afsnij-frequentie van 2 kHz.

Ten tweede zijn er filters die hoge frequenties doorlaten, en lage frequenties verzwakken. Dit zijn hoogdoorlaatfilters (high-pass), zie Figuur 21.



Figuur 21: Geïdealiseerde frequentie-karakteristiek van een hoogdoorlaat-filter ('high-pass') met een afsnij-frequentie van 2 kHz.

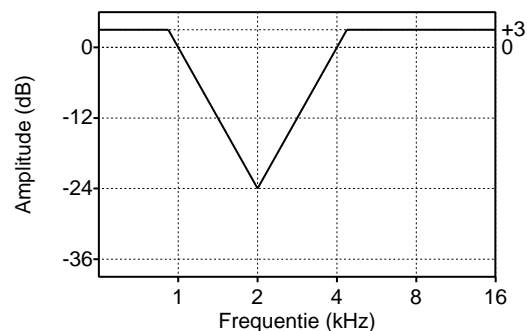
Ten derde zijn er filters die een bepaald frequentie-gebied doorlaten, en frequenties buiten deze doorlaat-band verzwakken. Dit zijn banddoorlaatfilters (band-pass), zie Figuur 22.



Figuur 22: Geïdealiseerde frequentie-karakteristiek van een banddoorlaat-filter ('band-pass') met een bandbreedte van 1 tot 4 kHz (2 octaven).

Ten vierde zijn er filters die een bepaald frequentie-gebied verzwakken, en frequenties buiten deze sperband doorlaten. Dit zijn bandsperfilters (band-reject, notch), zie Figuur 23.

Een filter wordt gekarakteriseerd door twee eigenschappen, namelijk door de afsnijsteilheid en de afsnijfrequentie. Banddoorlaat- en sperfilters worden mede gekarakteriseerd door hun bandbreedte. Hieronder worden deze begrippen toegelicht.



Figuur 23: Geïdealiseerde frequentie-karakteristiek van een bandsper-filter ('notch') met een bandbreedte van 1 tot 4 kHz (2 octaven) en afsnij-steilheid van -24 dB per octaaf.

Afsnijsteilheid

De afsnijsteilheid is een maat voor de scherpte van de overgang tussen het doorgelaten en het afgezwakte deel van het spectrum. Dit wordt uitgedrukt in decibel per octaaf. Een octaaf is de afstand tussen een bepaalde frequentie en zijn tweevoud. De afstand tussen 100 Hz en 200 Hz is een octaaf, zo ook de afstand tussen 200 Hz en 400 Hz.

Een afsnijsteilheid van -12 dB/octaaf wil zeggen dat bij een laagdoorlaatfilter de uitgangsamplitude 12 dB zwakker wordt, bij elke verdubbeling van de frequentie van hetingangssignaal (zie Figuur 20). Bij een hoogdoorlaatfilter met dezelfde afsnijsteilheid wordt de uitgangsamplitude 12 dB sterker, bij elke verdubbeling van de frequentie van hetingangssignaal (d.w.z. als hetingangssignaal 1 octaaf hoger wordt).

Afsnijfrequentie

De afsnijfrequentie is een maat voor de plaats van de overgang tussen het doorgelaten en het afgezwakte deel van het spectrum. De afsnijfrequentie wordt gemeten op het punt waar de

amplitude van het signaal 3 dB zwakker is dan die van het onverzwakte signaal. Dit wordt ook wel het -3 dB punt genoemd.

Bandbreedte

Banddoorlaat- en bandsperfilters worden gekarakteriseerd door hun bandbreedte. Dat wil zeggen de grootte van het frequentiegebied dat door een filter wordt doorgelaten, respectievelijk tegengehouden. Dit gebied wordt begrensd door de twee -3 dB punten.

Het is gebruikelijk om bandbreedte te specificeren door middel van de twee afsnijfrequenties (of -3 dB punten). Zo heeft geluid uit een telefoon een bandbreedte van 300 - 3400 Hz. Het is ook mogelijk om het verschil tussen de grensfrequenties aan te geven. Om aan te geven waar dat frequentiegebied ligt, wordt ook de centrumfrequentie (CF) gegeven.

Tenslotte wordt bandbreedte ook wel uitgedrukt door middel van het muzikale interval. Dat is de verhouding tussen de beide grensfrequenties. Dan wordt de term relatieve bandbreedte gebruikt. De relatieve bandbreedte van een bandfilter met grensfrequenties van 600 en 1200 Hz is een octaaf. Hetzelfde geldt voor een bandfilter met grensfrequenties van 100 en 200 Hz. Hieruit blijkt dat de absolute bandbreedtes verschillen, terwijl er in beide gevallen sprake is van een relatieve bandbreedte van een octaaf. Dit zijn octaafbandfilters.

Als de grensfrequenties van de band zich verhouden als 4:5, spreekt men van een tertsbandfilter. Een tert is 1/3 octaaf.

4.2 Spectrografische analyse

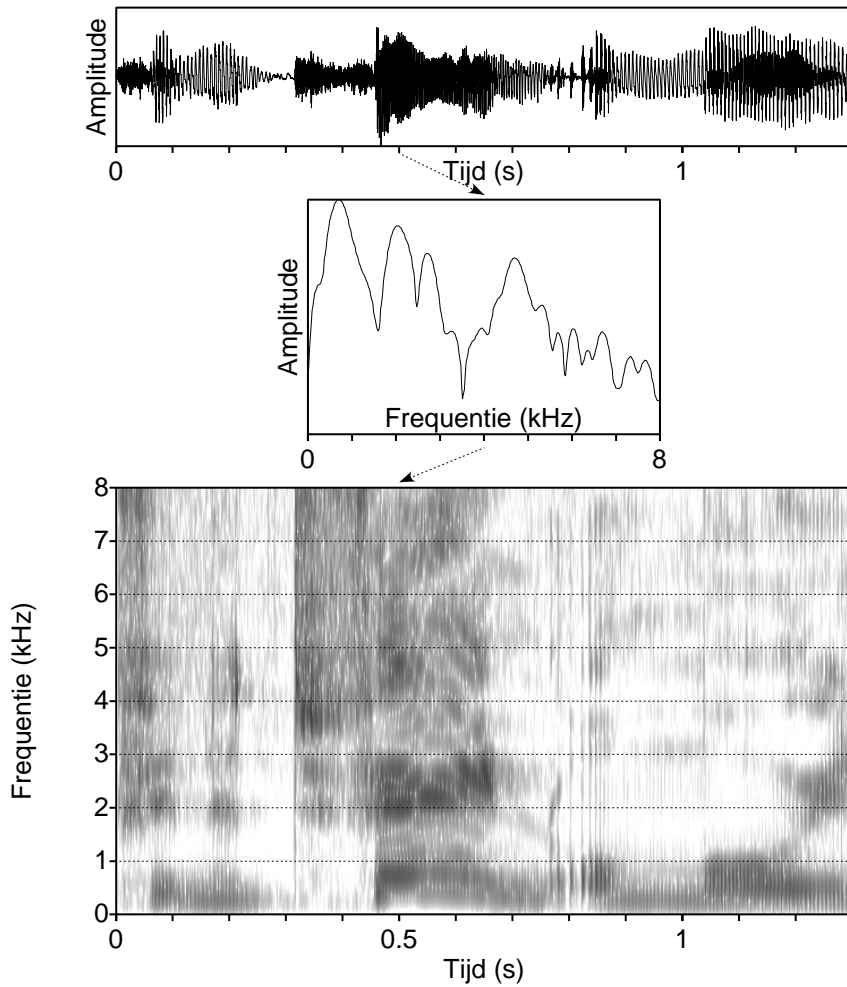
Zoals bekend kunnen met behulp van een spectrum de verschillende frequentiecomponenten van een samengesteld signaal worden weergegeven (zie §2.6, p.14). Je kunt bijvoorbeeld van een spraakklank de formanten zien (zie §2.9, p.19). Echter, een spectrum is maar één analyse op een bepaald moment in de tijd. Het wordt berekend over een kort stukje signaal (van meestal 25 ms).

Om een indruk te krijgen van het verloop van het spectrum in de tijd, kun je het spectrum laten berekenen over een stukje spraak van 25 ms. Daarna verschuift het analyse-venster met 10 ms, en je berekent opnieuw het spectrum over het nieuwe venster. Het is mogelijk om een tweedimensionale weergave te maken van de drie dimensies tijd, frequentie en amplitude. Zo'n plaatje heet een spectrogram, wat ruwweg hetzelfde is als een heleboel spectra achter elkaar. De hoogte van een spectrale piek correspondeert dan met de mate van zwarting in het spectrogram. Een instrument dat deze analyse uitvoert is de spectrograaf of sonagraaf.

Het hart van de spectrograaf is een bandfilter met een betrekkelijk kleine bandbreedte. Het filter van de spectrograaf heeft een bijzondere eigenschap: de frequentie waarop het is afgestemd kan gewijzigd worden. Hierdoor is het mogelijk om met één filter bij een groot aantal verschillende frequentiewaarden te meten wat de amplitude van het geluid daar is. Het analysefilter begint bij een heel lage frequentie en schrijft na elke meting iets op. Dit gebeurt net zo lang tot het hele frequentiebereik (traditioneel meestal van 0 tot 8 kHz) doorlopen is.

Het resultaat is een spectrogram. Hierin staat de tijd langs de horizontale as, net als in een oscillogram, en de frequentie langs de verticale as. De zwarting of kleuring is een maat voor de sterkte van het geluid. Er worden dus drie dimensies van het geluid weergegeven: tijd, frequentie en amplitude. Het spectrogram geeft daarom een goede indruk van de sterkte van de frequentiecomponenten over het hele frequentiebereik van 0 tot 8 kHz, en van de veranderingen van de sterkteverhoudingen in de tijd.

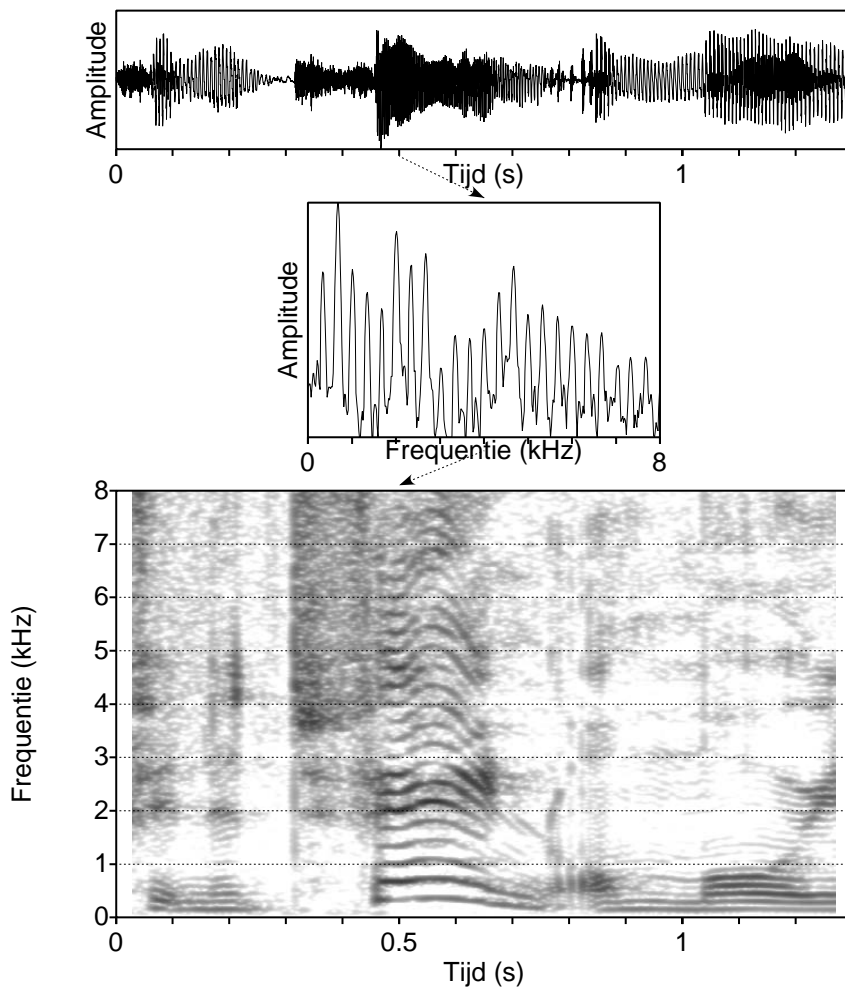
Er zijn twee soorten spectrogrammen: bredeband- en smalleband-spectrogrammen. Figuur 24 is een voorbeeld van een bredeband-spectrogram. Hierin kun je wel formanten herkennen, maar geen afzonderlijke harmonischen. Hier is de bandbreedte van het analysefilter (d.w.z. het frequentiegebied waarin het filter meet hoe sterk het geluid is) betrekkelijk ruim, ongeveer 300 Hz. Hierdoor is de nauwkeurigheid waarmee frequenties van elkaar gescheiden worden betrekkelijk slecht, maar de nauwkeurigheid waarmee verschijnselen in de tijd van elkaar gescheiden worden betrekkelijk goed. Dit heeft te maken met een algemene eigenschap van filters: naarmate het oplossend vermogen voor frequenties slechter wordt, wordt het oplossend vermogen in de tijd beter. Ter illustratie: fotografen kennen een dergelijke relatie tussen het oplossend vermogen in tijd (sluiter tijd), en in scherptediepte (diafragma), bij een vaste filmgevoeligheid.



Figuur 24: Boven: oscillogram. Midden: spectrum bij $t=0.5$ s. Onder: breedband-spectrogram.

Wanneer je vooral geïnteresseerd bent in het oplossend vermogen voor frequenties, dan maak je gebruik van een smalleband-spectrogram. De bandbreedte van het analysefilter is dan betrekkelijk smal, ongeveer 45 Hz. Het oplossend vermogen in de tijd is nu veel slechter. Zie Figuur 25, waarin de afzonderlijke boventonen nu goed zichtbaar zijn.

In beide soorten spectrogrammen kun je goed het onderscheid zien tussen stemhebbende en stemloze delen van het spraakgeluid. In een spectrogram van een hele spraakuiting valt goed op dat het spraakgeluid voortdurend in verandering is.



Figuur 25: Boven: oscillogram. Midden: spectrum bij $t=0.5$ s. Onder: smalleband-spectrogram, van hetzelfde spraakfragment als in Figuur 25.

Er zijn nog verschillende andere analysemethoden om iets over de formanten in een stuk spraak te weten te komen. De LPC-analyse is een veel gebruikte analysemethode. In het volgende hoofdstuk gaan we daar verder op in.

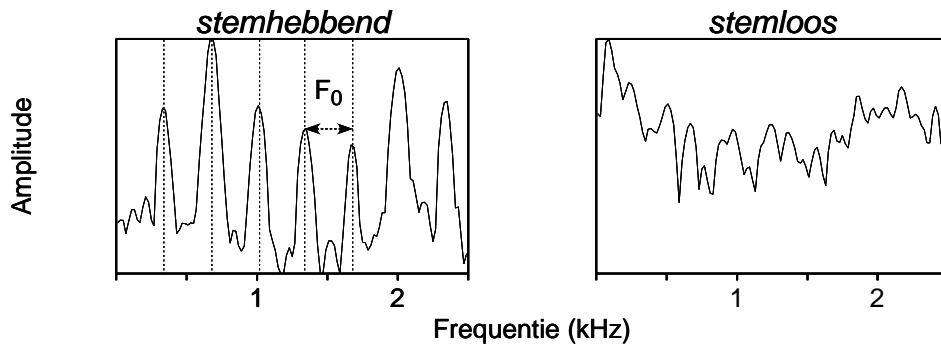
4.3 Analyse van grondfrequentie en intensiteit

De spectrografische analyse van een spraaksignaal toont de verglijdingen van de formanten tijdens de uiting. Met andere woorden: je ziet hoe de resonantiefrequenties van het spraakkanaal veranderen in de tijd. Als we even het bron-filtermodel in gedachten nemen, dan zie je dat een spectrogram vooral informatie over het filter geeft. We kunnen natuurlijk ook metingen verrichten die informatie geven over het bronsignaal.

De belangrijkste kenmerken van het brongeluid zijn stemhebbend-/stemloosheid, grondfrequentie en intensiteit. Dit heten de bronparameters. Je kunt de bronparameters net als de filterparameters uitzetten tegen de tijd. De bron- en filterparameters zijn afhankelijk van de stand van de 'spraakorganen' en veranderen in hetzelfde ritme als waarin de klanken worden uitgestoten. Voor het gemak stelt men meestal dat er elke 10 à 25 ms een nieuwe 'stand' is. We noemen zo'n kort stukje spraak een analysevenster.

Hoe wordt de grondfrequentie binnen een analysevenster bepaald? Tot op heden is er geen ultieme methode gevonden om de grondfrequentie (F_0) foutloos te bepalen. Vrijwel elke aanpak geeft wel eens foute resultaten. De fouten die vaak optreden zijn octaaffouten (er wordt dan een F_0 berekend die een octaaf te hoog of te laag ligt) of fouten in de stemhebbend-

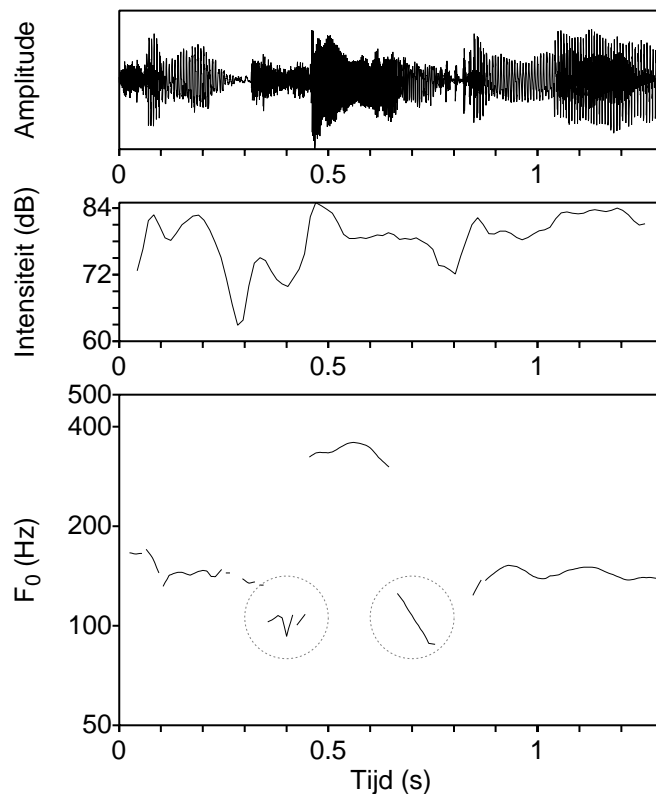
/stemloosdetectie (een stukje spraak wordt als stemloos benoemd terwijl het juist stemhebbend had moeten zijn, of andersom). Een van de methoden om de grondfrequentie te bepalen is door een frequentiespectrum van het venster te berekenen, en daarna de afstand in Hz te meten tussen twee opeenvolgende harmonischen. Dit kan uiteraard alleen als er duidelijk harmonischen herkenbaar zijn.



Figuur 26: Links: De grondtoon (F_0) wordt afgeleid uit de afstand tussen harmonischen. Rechts: Indien er geen duidelijke harmonischen zijn, wordt het spraakfragment als stemloos beschouwd.

Intensiteitbepaling is een stuk eenvoudiger. Dat komt neer op het berekenen van de zgn. effectieve waarden (root mean square) van de analysevensters. Per venster worden alle meetwaarden (amplitudes) gekwadrateerd en vervolgens gesommeerd, waarna de wortel van de som wordt genomen.

Figuur 27 toont de bronparameters zoals ze zijn bepaald van het spraakfragment uit Figuren 24 en 25. De bovenste curve is het oscillogram; direct daaronder de intensiteits-curve (met markeringen per 3 dB). Het onderste paneel bevat het F_0 -contour. Stemloze spraakfragmenten zijn hierin gekenmerkt door het ontbreken van een F_0 -waarde. We hebben door middel van omcirkelingen aangegeven waar het analyse-algoritme zich heeft vergist. De eerste fout (bij $t=0.4s$) betreft de vermeende stemhebbendheid: het analyse-venster bevindt zich hier midden in een stemloze [f]. De tweede fout (bij $t=0.7s$) is een octaaf-fout: de aangegeven F_0 is precies een octaaf te laag. In werkelijkheid daalt de F_0 vloeiend vanaf de piek bij $t=0.5s$, zoals ook te zien is in Figuur 25 (aan het vloeiende verloop van de harmonischen).



Figuur 27: Bron-parameters van hetzelfde spraakfragment als in Figuur 24 en 25. Boven: oscillogram. Midden: verloop van intensiteit. Onder: verloop van grondtoon (F_0). Analyse-fouten zijn omcirkeld; zie tekst.

Nieuwe begrippen

- frequentiefilter
- frequentiekaracteristiek
- laagdoorlaatfilter, hoogdoorlaatfilter
- banddoorlaatfilter, bandsperfilter
- afsnijsteilheid, afsnijfrequentie
- octaaf
- -3-dB-punt
- bandbreedte
- centrumfrequentie (CF)
- octaafbandfilter, tertsbandfilter
- spectrogram
- spectrograaf, sonograaf
- bredeband-spectrogram, smalleband-spectrogram
- filterparameters, bronparameters
- analysevenster

Opdrachten hoofdstuk 4

Opdracht 9

(I) Maak een ruissignaal aan van witte ruis, maak een spectrum en druk het af.

- a. Beschrijf het verschil tussen witte en roze ruis.
- b. Hoe zou je witte ruis moeten filteren om het op roze ruis te laten lijken?

Opdracht 10

Maak een spectrogram van

(I) de normaal uitgesproken versie van *delicatesse* (delicatesse3)

(II) een geplakte versie van *delicatesse* (delicatesse2)

a. Label de verschillende lettergrepen.

b. Maak een afdruk van het spectrogram met daaronder het oscillogram. Doe dit voor beide *delicatessen* en beschrijf de verschillen.

Opdracht 11

(I) Laat het frequentiespectrum van een stukje steady-state van de klinker /i/ berekenen, eenmaal met en eenmaal zonder Hanning-venster. Druk de spectra af.

(II) Druk de oscillogrammen af van beide versies.

Beschrijf het effect dat je waarneemt bij venstering.

Opdracht 12

(I) Maak een spectrogram van de file telefoon.aifc. Deze file kan je kopiëren volgens de aanwijzingen van de practicumassistent.

(II) Kopieer het bestand en filter uit het nieuwe bestand de frequenties die niet door de telefoon worden doorgelaten. Sla het bestand op onder een nieuwe naam.

Hoe klinkt de nieuwe file?

HOOFDSTUK 5

ANALYSE EN SYNTHESE VAN SPRAAK VIA LPC

5.1 Wat biedt LPC?

Bij de analyse van spraaksignalen zijn we vaak geïnteresseerd in de spectrale eigenschappen van de signalen. In het bijzonder kijken we naar formantsporen in een spectrogram, en hoe ze verglijden tijdens de overgang van de ene naar de andere klank. Een spectrogram is echter niet erg geschikt voor de bepaling van concrete filterparameters. Je kunt in een spectrogram hier en daar wel ongeveer zien wat de frequenties zijn van (enkele) formanten, maar de bandbreedtes zijn nauwelijks te schatten. Een betere analysemethode om kwantitatieve informatie over formanten te verkrijgen is de LPC-methode. LPC staat voor *linear predictive coding*; wat deze term inhoudt wordt straks beschreven.

De LPC-analyse (en synthese) gaat uit van het bron-filter-model (zie §2.8, p.18). In dit model wordt de spraak beschouwd als een gefilterd bronsignaal; om precies te zijn een door mond-, keel- en neusholte gefilterde stembandtrilling of geruis. De analyse neemt aan dat het spectrum van het brongeluid afloopt met een helling van -12 dB/octaaf, en dat het effect van de uitstraling +6 dB/octaaf bedraagt. Op grond van deze aannames kan de LPC-analyse het aangeboden spraaksignaal splitsen in een bronsignaal en een filter. Simpel gezegd (voor een stemhebbend fragment): het bronsignaal moet zorgen voor een spectrale helling van -6 dB; alle spectrale karakteristieken die daarvan afwijken worden toegeschreven aan de werking van het filter. Het bron-filter-model is echter niet perfect, en mede daarom moeten de gevonden filterparameters (formantfrequenties en bandbreedtes) als *schattingen* worden geïnterpreteerd.

De LPC-analyse geeft tesamen met de F_0 -analyse, stemhebbend/stemloos-detectie, en de intensiteit-meting een redelijk goede karakterisering van de spraak. Dat blijkt uit de LPC-resynthese, waarbij de berekende parameters gebruikt worden om het originele spraakgeluid te reconstrueren. Deze resynthese levert doorgaans spraak van merkbaar slechtere kwaliteit, maar de geresyntheseerde spraak is nog wel aanvaardbaar en verstaanbaar.

5.2 LPC-analyse en synthese

Via de LPC-analyse kan de filtering van een brongeluid door het spraakkanaal worden bepaald (geschat). De wiskundige methode die hierbij toegepast wordt heet 'lineaire predictie' en werkt als volgt. We beschouwen een stukje spraak (ca. 25 ms van een bemonsterde golfvorm), en we willen proberen een wiskundige beschrijving te maken van de meetwaarden (samples) in dat analysevenster. We werken dus in het tijds-domein. Iets concreter gezegd gaan we de gemeten amplitudewaarden in het venster *voorspellen* aan de hand van vorige meetwaarden. Een simpel voorbeeld: De getallenreeks y heeft de waarden $\{0, 2, 8, 26, 80, \dots\}$. In formuleform kan dit worden opgeschreven als

$$y_{\text{huidig}} = 2 + 3 \cdot y_{\text{vorig}}$$

Een willekeurige spraakgolfvorm is natuurlijk niet zo eenvoudig en vooral ook niet zo precies te beschrijven, maar het voorbeeldje toont wel aan dat het (soms) mogelijk is meetwaarden te voorspellen uit vorige bemonsteringen. Wat moeten we nu met deze voorspellingen of predicties, zal je zeggen. Welnu: het gaat niet om de voorspelde meetwaarden, maar om de coëfficiënten in de predictieformule (in het voorbeeld: 2 en 3). Uit de resulterende LPC-coëfficiënten kunnen namelijk de frequenties en bandbreedtes van formanten worden afgeleid. Hoe deze afleiding in z'n werk gaat, en hoe de coëfficiënten precies worden berekend, valt buiten het bestek van dit *Werkboek*. De LPC-analyse richt zich erop de coëfficiënten zodanig te kiezen dat het verschil tussen de voorspelde en de werkelijke amplitudewaarde zo klein mogelijk is. In de praktijk houden we echter altijd een foutsignaal over (het residu).

Voor de liefhebbers:

$$y_n = \sum_{i=1}^{i=m} c_i \cdot y_{n-1} + e$$

waarbij y_n het n-de sample in het analyse-venster

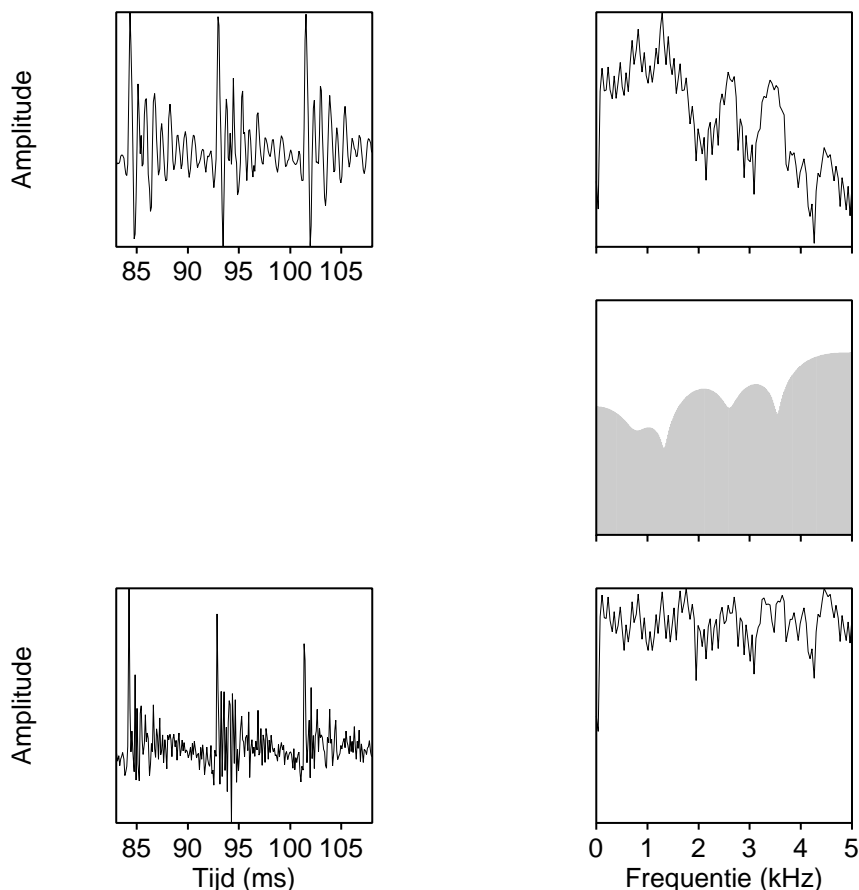
c_i de i-de LPC-coëfficiënt

m totaal aantal LPC-coëfficiënten

e verschil tussen voorspelling en werkelijke waarde;
dit moet zo klein mogelijk worden

Meestal gebruikt de LPC-analyse 10 meetwaarden om het 11^e te voorspellen. Dit geeft dan $m=10$ coëfficiënten, die kunnen worden omgerekend naar 5 formanten (5 piekfrequenties + 5 bandbreedtes). Omdat het spraaksignaal nu eenmaal verandert in de tijd, wordt er elke 10 ms een nieuwe analyse gedaan over een venster van 25 ms. Zo'n stukje van 10 ms noemen we een 'frame'. De 10 LPC-coëfficiënten in een frame geven een goede benadering van het spraaksignaal op dat moment, onder de aanname dat het bronsgaunaal een vaste spectrale helling heeft van -6 dB/octaaf.

In Figuur 28 zien we wat de LPC-analyse in het frequentiespectrum doet. Het ingangssignaal (linksboven) heeft een spectrum (rechtsboven) met formanten. De LPC-analyse kiest de filter-karakteristiek nu zodanig (rechtsmidden), dat het residu een vlakke spectrale helling heeft van -6 dB/octaaf, kenmerkend voor brongeluiden in spraak. (Maar let op: in dit geval is die helling reeds verwerkt in de filter-karakteristiek, en het residu heeft een vlakke helling van 0 dB/octaaf, rechtsonder). Dit spectrum correspondeert met een puls-achtig bronsgaunaal (linksonder), dat te beschouwen is als een grove benadering van de stembandtrilling.



Figuur 28: LPC-analyse in het frequentiedomein.

Op deze manier is het ingangssignaal dus gesplitst in een bronsgaunaal (het foutsignaal of residu) en een filter. De resonanties in het spectrum van het ingangssignaal zijn overgedragen op het analyse-filter. Uit de coëfficiënten die dit filter beschrijven kunnen formanten worden afgeleid. Het

gereconstrueerde bronsignaal (residu) bevat dus geen informatie meer over deze formanten. We nemen aan dat het residu geen nuttige informatie meer bevat, en dus mag worden weggegooid⁵. Voor ieder frame van 10 ms bewaren we dan 10 filter-coëfficiënten (A1..A10), de grondfrequentie (F0), de intensiteit (Gain), plus de stem/stemloosheid (VUV). Er zijn dus slechts dertien parameters nodig per frame om een spraaksignaal vast te leggen. De spraak is dan geparametriseerd ofwel gecodeerd.

frm	tijd	parameters												
		bron			filter									
		G	VU	F0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
1	0								
2	10									
3	20	...												
...

Figuur 29: Opslag van parametrische spraak.

Het aantal coëfficiënten m in de LPC-analyse is meestal zelf in te stellen; de optimale waarde voor m is afhankelijk van de bemonsteringsfrequentie van het signaal. Zoals hierboven gezegd, bepaalt het aantal coëfficiënten uiteindelijk het aantal spectrale pieken ('formanten') dat gevonden wordt in hetingangssignaal. Als je een zuivere sinus analyseert met $m=10$ coëfficiënten, dan levert de standaard LPC-analyse toch $m/2=5$ spectrale pieken op — hetgeen duidelijk onjuist is.

Als je kijkt naar een spectrogram van een mannenstem, zie je meestal zo'n 5 formanten in het gebied van 0 tot 5 kHz. Dat is niet toevallig. Op grond van de bron-filter-theorie kan je uitrekenen waar de formanten liggen, bij een neutrale stand van de mondholte. We nemen aan dat het aanzetstuk van een man 17 cm lang is, van glottis tot lippen. De formanten liggen dan ruwweg bij 500, 1500, 2500, 3500 en 4500 Hz, dus ca. 1 kHz uit elkaar. Bij niet-neutrale standen van de mondholte veranderen wel de frequenties en bandbreedtes van de formanten, maar het aantal formanten blijft gelijk.

Bij een bemonsteringsfrequentie van 10 kHz is de maximale frequentie van het signaal 5 kHz. In dit frequentiegebied verwachten we voor een mannenstem 5 formanten, dus gebruik je 10 LPC-coëfficiënten. Het juiste aantal coëfficiënten is dus doorgaans gelijk aan de bemonsteringsfrequentie in kHz. Maar: formanten boven F5 zijn niet meer betrouwbaar in het signaal op te sporen, zoals je zelf kan zien in een spectrogram. Het is dus niet verstandig om LPC-analyse uit te voeren met veel meer dan 10 coëfficiënten: de gevonden spectrale pieken corresponderen dan niet meer met echte formanten (die zijn immers te zwak om opgespoord te worden) maar met andere spectrale karakteristieken van het signaal. Zeer grote bandbreedtes van de gevonden 'formanten' vormen een indicatie voor dergelijke 'fouten' in de LPC-analyse.

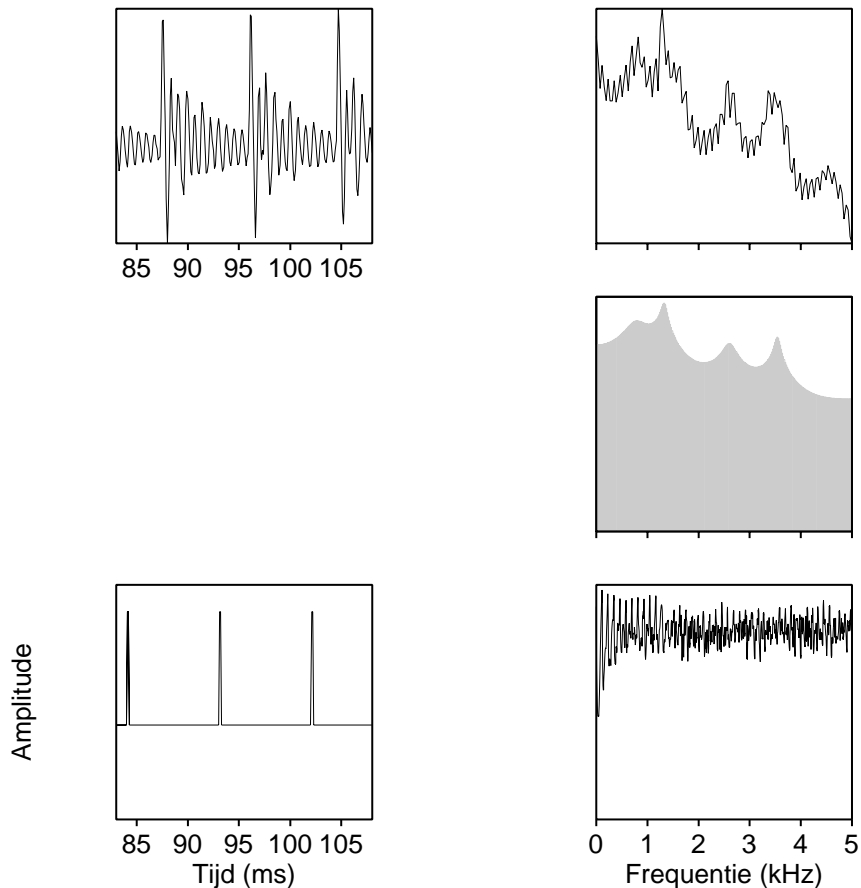
Voor sommige toepassingen is het geen probleem dat bron-kenmerken geïnterpreteerd worden als filter-kenmerken. Het gaat pas fout indien de LPC-uitvoer geïnterpreteerd wordt in termen van piekfrequenties en bandbreedtes van formanten (want de coëfficiënten stellen immers geen formanten meer voor, maar andere spectrale karakteristieken). Zo'n toepassing is bijvoorbeeld het gebruik van LPC-resynthese om de prosodie van een uiting te manipuleren.

In de praktijk willen we dus niet meer dan 10 coëfficiënten opgeven, althans voor mannenstemmen. We moeten dan ervoor zorgen dat het invoer-signaal alleen het frequentiegebied tot 5 kHz bevat, door herbemonstering ('downsampling') tot 10 kHz.

Met LPC-synthese kan een spraaksignaal worden gereconstrueerd vanuit de opgeslagen parameters. De synthese bewandelt de omgekeerde weg t.o.v. de analyse, zoals geïllustreerd in Figuur 30. We beginnen dus links-onder in de figuur: hetingangssignaal is nu een (pulsvormig) bronsignaal met de frequentie van parameter F0 (linksonder); dit signaal heeft een vlakke spectrale helling (rechtsonder). Dit kunstmatige signaal is te beschouwen als een grove benadering van de stembandtrilling. We voeren dit signaal toe aan een zgn. synthese-filter. De karakteristiek van dit filter (rechtsmidden) is de inverse van het analyse-filter. Met het synthese-filter worden formanten aangebracht in het vlakke bronspectrum (rechtsboven); het resulterende

⁵ Toch vormt het foutsignaal soms nog verstaanbare spraak; het foutsignaal of residu bevat dan blijkbaar nog voldoende spectrale informatie voor de menselijke spraakperceptie.

geluid lijkt op het originele spraaksignaal dat de invoer van de LPC-analyse vormde (linksboven). De spraak klinkt wel een beetje zoemerig (buzzy); dat komt omdat er een kunstmatige puls als bronsignaal is gebruikt in plaats van het bij de LPC-analyse gereconstrueerde bronsignaal (residu).



Figuur 30: LPC-synthese in het frequentiedomein.

5.3 PSOLA

Wanneer je in PRAAT een toonhoogtemanipulatie uitvoert wordt standaard echter niet van LPC gebruik gemaakt, maar van PSOLA. PSOLA staat voor Pitch-Synchronous-OverLap-Add. Over het signaal wordt een analysevenster gelegd. De lengte van dit venster is gerelateerd aan de periodeduur (= toonhoogte) van het signaal en omvat meestal exact 2 of exact 4 perioden. Hieraan dankt PSOLA een deel van zijn naam, namelijk Pitch-Synchronous. In het spectrum van een analysevenster mag immers geen informatie te vinden zijn over de afzonderlijke harmonischen, aangezien dat een correcte toonhoogtemanipulatie in de weg staat. De lengte van het analysevenster mag hierdoor niet te lang zijn. Het venster mag echter ook niet te kort zijn, want dan kun je het origineel niet goed reconstrueren.

Voorwaarde is natuurlijk dat er een goede toonhoogte-detectie op het signaal uitgevoerd kan worden. Dat is niet altijd het geval. Met behulp van de analysevensters kun je de tijd (duur) en de toonhoogte van een signaal manipuleren. De tijd wordt gemanipuleerd door of vensters weg te gooien of vensters te herhalen. In het eerste geval maak je de duur van het signaal korter; in het tweede geval verleng je de duur. Toonhoogtemanipulatie gebeurt door vensters in elkaar te schuiven of juist uit elkaar te trekken. Bij de resynthese worden de (vershoven of aangepaste) vensters weer bij elkaar opgeteld. Hieraan is het andere deel van de naam PSOLA ontleend, namelijk OverLap-Add. Manipulatie werkt het beste als het bereik waarin het gebeurt tussen halvering en verdubbeling van de oorspronkelijke vensterduur ligt. Dat geldt zowel voor de duur als voor de toonhoogte.

Nieuwe begrippen

- LPC (linear predictive coding)
- LPC-coëfficiënten, filter-coëfficiënten
- foutsignaal, residu
- parameter, parametrische spraak
- frame
- PSOLA

Opdrachten hoofdstuk 5

Opdracht 13

- (I) Bekijk en beluister het F_0 -verloop van zin 3 (opnames). Label de zin per woord. Druk de gelabelde golfvorm af.
- a. Geef eventuele punthoeden en platte hoeden in het intonatieverloop aan.
b. Omcirkel eventuele octaaffouten in het F_0 -verloop en foute VUV-beslissingen (VUV = voiced-unvoiced).

Opdracht 14

- (I) Bekijk de F_0 -patronen van de zinnen die door een man en een vrouw zijn uitgesproken (zin 2) en druk ze af.
- a. Zoek octaaffouten en VUV-fouten en haal ze eruit.
b. Vergelijk de intonatiepatronen en bespreek eventuele verschillen.

Opdracht 15

- (I) Zoek de bestanden `woord1.aiff` en `woord2.aiff`. Deze woorden vormen een ambigu woordpaar. Manipuleer het F_0 -patroon en de duur van de syllaben om van *kànon* *canòn* te maken, en andersom.
- (II) Maak afdrukken van beide originele versies en van de gemanipuleerde versies (golfvorm en toonhoogteverloop).

Beschrijf wat voor een manipulaties je hebt uitgevoerd en wat de problemen waren die je daarbij tegenkwam.

Opdracht 16

- (I) Bepaal een frequentiespectrum van een stukje steady-state van een klinker. Druk dit af.
(II) Laat op hetzelfde klinkerfragment ook een LPC-analyse los. Let daarbij op de instelling van de coëfficiënten (prediction order).

Vergelijk het frequentiespectrum en het LPC-spectrum. Kun je de formanten die in het LPC-spectrum te zien zijn ook in het frequentiespectrum herkennen?

Opdrachten 6

Deze bijeenkomst worden practicumopdrachten uitgevoerd die betrekking hebben op alle theorie uit dit werkboek.

Opdracht 18

- (I) Segmenteer uit de zinnen die je hebt opgenomen in week 3 de onderstreepte woorden en voeg ze aaneen tot de zin.: "*De minister heeft de dia's van zijn vader nodig.*" Label deze zin per woord.
- (II) Maak dezelfde zin met de losse woorden, die je labelt. Vergelijk beide zinnen met de normaal uitgesproken versie. Beluister ze en beschrijf de verschillen.

Opdracht 19

- (I) Zoek het bestand `wiekust.aiff` met de zin "*Kees zei Koos kust met het buurmeisje van Toos*" en maak twee kopieën. Noem de eerste `kooskust.aiff` en de tweede `keeskust.aiff`.
- (II) De zin is ambigu. Zowel Kees als Koos kunnen zoenen met het buurmeisje van Toos. Manipuleer het F_0 -patroon en de duur van pauzes en van syllabes van beide bestanden zodanig dat beide betekenissen duidelijk worden.
- (III) Label de twee zinnen, druk de twee patronen af (gelabelde golfvorm en toonhoogte) en vermeld de bijbehorende betekenissen.

Opdracht 20

- (I) Segmenteer de onderstreepte klanken uit de opnames van de losse woorden:
- | | |
|----------------------|------------------------|
| <i>sen<u>aa</u>t</i> | <i>be<u>ei</u>d</i> |
| <i>ge<u>bi</u>ed</i> | <i>schav<u>ui</u>t</i> |
| <i>ge<u>zo</u>et</i> | <i>her<u>au</u>t</i> |
| <i>pi<u>lo</u>ot</i> | |
- (II) Maak een LPC-analyse van de zeven klanken en druk de LPC-spectrogramen af. Beschrijf de verschillen tussen de klanken onderling en tussen de klanken uit het eerste rijtje en uit het tweede rijtje.

Verantwoording

Dit *Werkboek* is een herziene versie van het *Werkboek bij het Practicum Fonetiek I*, uit 1993. Aan deze herziening versie hebben meerdere personen een bijdrage geleverd: Mirjam Wester en Denise Bruin hebben de vragen en opdrachten aangepast en gemoderniseerd. Bovendien hebben zij model-antwoorden en -uitwerkingen gemaakt voor de practicum-assistenten. Olga van Herwijnen heeft de model-uitwerkingen verder aangepast. Eva Sittig heeft het werkboek, de modeluitwerkingen en de handleiding van de spraakprogammatuur opnieuw aangepast. Daarnaast hebben alle practicum-assistenten in de afgelopen jaren veel nuttig commentaar geleverd, dat op enige wijze in deze versie is verwerkt. De eindredactie en supervisie van dit *Werkboek* was in handen van Hugo Quené.

APPENDIX I

WERKEN OP HET LINUX-SYSTEEM

Inloggen

De werkstations in de practicumruimte staan al aan als je binnenkomt. Zet deze nooit uit! Inloggen doe je door je gebruikersnaam (bijv. `prak01`) in te typen achter 'login name:', gevolgd door `<ENTER>`, en vervolgens je wachtwoord. Gebruikersnaam en wachtwoord krijg je bij aanvang van het practicum van de practicumassistent. Verander je wachtwoord direct. Het is onmogelijk het wachtwoord dat je krijgt te onthouden, en je mag het *niet* opschrijven in verband met de veiligheid van het systeem.

Uitloggen

Je logt uit door met de rechtermuisknop op de background te klikken en vervolgens de optie 'logout' te kiezen. Ook kun je op de balk (beneden op het scherm) links van de klok een logout-button vinden. Daarna bevestigen dat je wilt uitloggen.

Wachtwoord wijzigen

Je wachtwoord wijzig je door het commando `passwd` in te typen in de LINUX-shell. Daarna wordt je gevraagd je oude wachtwoord in te typen, en vervolgens twee keer je nieuwe wachtwoord.

Directorystructuur

De directorystructuur van LINUX is hiërarchisch. De hoofddirectory bevat de subdirectory `/prak`, welke zelf de subdirectories `prak01`, `prak02` etc. bevat. Ieder groepje heeft zo'n directory met dezelfde naam als de gebruikersnaam. Deze directory wordt de 'homedirectory' genoemd. Wanneer je inlogt met je gebruikersnaam sta je onmiddellijk in deze homedirectory (dat wil zeggen: je homedirectory is dan je huidige directory). Deze directory hoort specifiek bij jouw account. Deze homedirectory bevat de subdirectory `/data`. In deze directory `/prak/prak0?/data` zet je je geluidsbestanden. De volledige naam van zo'n bestand is bijvoorbeeld

```
/prak/prak12/data/stotter.aiff
```

Commando's

Hieronder volgen enkele essentiële LINUX-commando's. Je kunt deze opdrachten echter ook uitvoeren via de File Manager. Deze wordt geactiveerd door met de linker muisknop enkel te klikken op het directory-icoon (een mapje). De practicumassistent kan je uitleggen hoe dit werkt.

ls	toon inhoud van de huidige directory (<u>l</u> ist)
pwd	toon huidige directory (<u>p</u> ath to <u>w</u> orking <u>d</u> irectory)
cd	wijzig huidige directory (<u>c</u> hange <u>d</u> irectory), bv naar:
cd ..	moeder van huidige directory
cd	homedirectory
cd ~	homedirectory
cd prak01	subdirectory <code>prak01</code> onder huidig
mkdir <naam>	<u>m</u> aak een nieuw <u>d</u> irectory
rmdir <naam>	<u>r</u> emove een <u>d</u> irectory
cp <bron> <doel>	Kopiëer bronbestand naar doelbestand (<u>c</u> opy)
mv <bron> <doel>	Hernoem bronbestand tot doelbestand (<u>m</u> ove)

In de practicumruimte liggen uitgebreide handleidingen.

Afdrukken

Tekstbestanden

Je kunt tekstbestanden op twee manieren uitprinten (in de LINUX-shell):

a2ps <bestandsnaam> Zo krijg je twee blzz. op 1 pagina.
lpr <bestandsnaam> Eén bladzijde per pagina.

Printqueue

Er zijn enkele handige commando's waarmee je kunt kijken hoeveel opdrachten ('jobs') er staan in de wachtrij ('queue'), en waarmee je printjobs kunt weghalen. Dit voorkomt lange wachttijden voor jezelf èn voor andere gebruikers!

lpq Laat zien welke jobs er in de wachtrij staan.
lprm <jobnr> Haal de job met nummer <jobnr> weg.
Het jobnummer vind je met het commando lpq.

APPENDIX II

GEBRUIK SPRAAKPROGRAMMATUUR

<http://www.let.uu.nl/~Hugo.Quene/faq/spraakprogr.html>