

## MASTER

### Transformaties van de klinkerdriehoeken van verschillende sprekers

Simons, W.F.

*Award date:*  
1980

[Link to publication](#)

#### **Disclaimer**

This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

4217

Instituut voor Perceptie Onderzoek

Den Dolech 2 - Eindhoven

15.8.1980

Rapport nr. 388

Transformaties van de klinkerdriehoeken  
van verschillende sprekers

W.F. Simons

Verslag van een afstudeeronderzoek

verricht onder leiding van

prof.dr. H. Bouma

prof.ir. F.J. Kylstra

ir. L.F. Willems

### Dankwoord

Een woord van dank zij hier gericht aan prof. Bouma en prof. Kylstra, die mijn onderzoek met veel belangstelling gevolgd en geleid hebben. Ook en vooral ben ik dank verschuldigd aan mijn dagelijkse begeleider ir. Willems, die mij in vele gesprekken en discussies steeds een bereidwillig oor geleend heeft, en mij door zijn enthousiasme gesteund en gestimuleerd heeft.

W.F. Simons

Eindhoven, 15 augustus 1980

## Samenvatting

Zeven lange nederlandse klinkers, afkomstig van vier sprekers, en geplaatst in de context h-klinker-s, zijn geanalyseerd met behulp van LPC-analyse, en van iedere klinker zijn de eerste twee formant-frequenties  $F_1, F_2$  bepaald. Voor iederspreker is een klinkerdriehoek verkregen. Deze verschillen in grootte, plaats en vorm. Getracht is, om met een spreker-afhankelijke transformatie de klinkerdriehoeken van de sprekers op elkaar te leggen.

Onderzocht zijn enkele tamelijk eenvoudige lineaire transformaties, en ter beoordeling ervan is een maat gedefinieerd. Een bepaald type transformatie, waarmee  $F_1$  en  $F_2$  afzonderlijk bewerkt worden, levert duidelijk de beste resultaten.

De gevonden transformatie zou toegepast kunnen worden in een automatisch spraakherkenningsysteem, om aanpassing aan verschillende sprekers te bereiken.

Verder is aandacht besteed aan het Newton (binomiaal) venster, en aan de toepassing van een autocorrelatie-venster bij de spectrale analyse van spraak.

Ook zijn enkele merkwaardige en tegenstrijdige punten in de theorie van LPC-analyse naar voren gekomen.

Inhoudsopgave

Inleiding	3
Analyse van het spraaksignaal	6
Toegepaste analyse	6
Spectra bij de analyse	7
Het autocorrelatie-venster	12
Problemen bij de analyse van spraak	23
Klinkerdriehoeken van vier sprekers	29
Transformaties van de klinkerdriehoeken	36
Een maat voor de afwijkingen tussen sprekers	49
Discussie	51
Literatuurlijst	52
Appendix	54
Het Newton-venster	54
Beschouwingen over LPC-analyse	64
Autocorrelaties en spectrum van het restsignaal	64
Een paradox bij LPC-analyse	65
Spiegeling van polen en nulpunten aan de eenheidscirkel	66
Een andere normalisering van het filter	67
Tijdsomkering	69

## Inleiding

Bij automatische spraakherkenning bestaat de eerste stap uit het bewerken en analyseren van het spraaksignaal, en worden belangrijke grootheden en parameters bepaald. Bijvoorbeeld kan dit zijn: het vermogen, het aantal nuldoorgangen per tijdseenheid, de uitgangsniveaus van een filterbank, autocorrelaties, filtercoëfficiënten volgens LPC-analyse. De drie laatstgenoemde stelsels parameters geven informatie over de spectrale samenstelling van het signaal. Bij klinkers zijn de maxima in het spectrum (formanten) belangrijk, en met name zijn de frequenties van de eerste twee formanten (F1 en F2) relevant voor het onderscheiden van klinkers (Peterson & Barney, 1952; Pols, Tromp & Plomp, 1973).

In figuur 1 zijn de klinkers van 50 mannelijke sprekers afgebeeld, met de bijbehorende F1 en F2. Er wordt een zogenaamde klinkerdriehoek gevormd. De gebieden van de klinkers zijn niet scherp gescheiden en overlappen elkaar gedeeltelijk. Wanneer F1 en F2 gebruikt worden bij automatische herkenning, zal in de grensgebieden verwarring optreden, en zal een keuze gemaakt moeten worden met behulp van andere informatie.

Voor één spreker is de situatie veel gunstiger: er is geen overlapping, en de gebieden van de klinkers zijn goed gescheiden (Peterson & Barney, 1952). De overlappingsen worden dus veroorzaakt door de variaties tussen de sprekers onderling.

De ~~ham~~vraag van dit onderzoek is: zijn er transformaties in het F1-F2-vlak te vinden, zodanig dat de klinkerdriehoeken van de individuele sprekers beter op elkaar gelegd worden, en de overlappingsen verdwijnen.

In figuur 2 zijn gemiddelde klinkers van mannelijke en vrouwelijke sprekers apart weergegeven. Men zou ook kunnen vragen naar een transformatie, die het mannelijke gemiddelde omzet in het vrouwelijke gemiddelde, of omgekeerd.

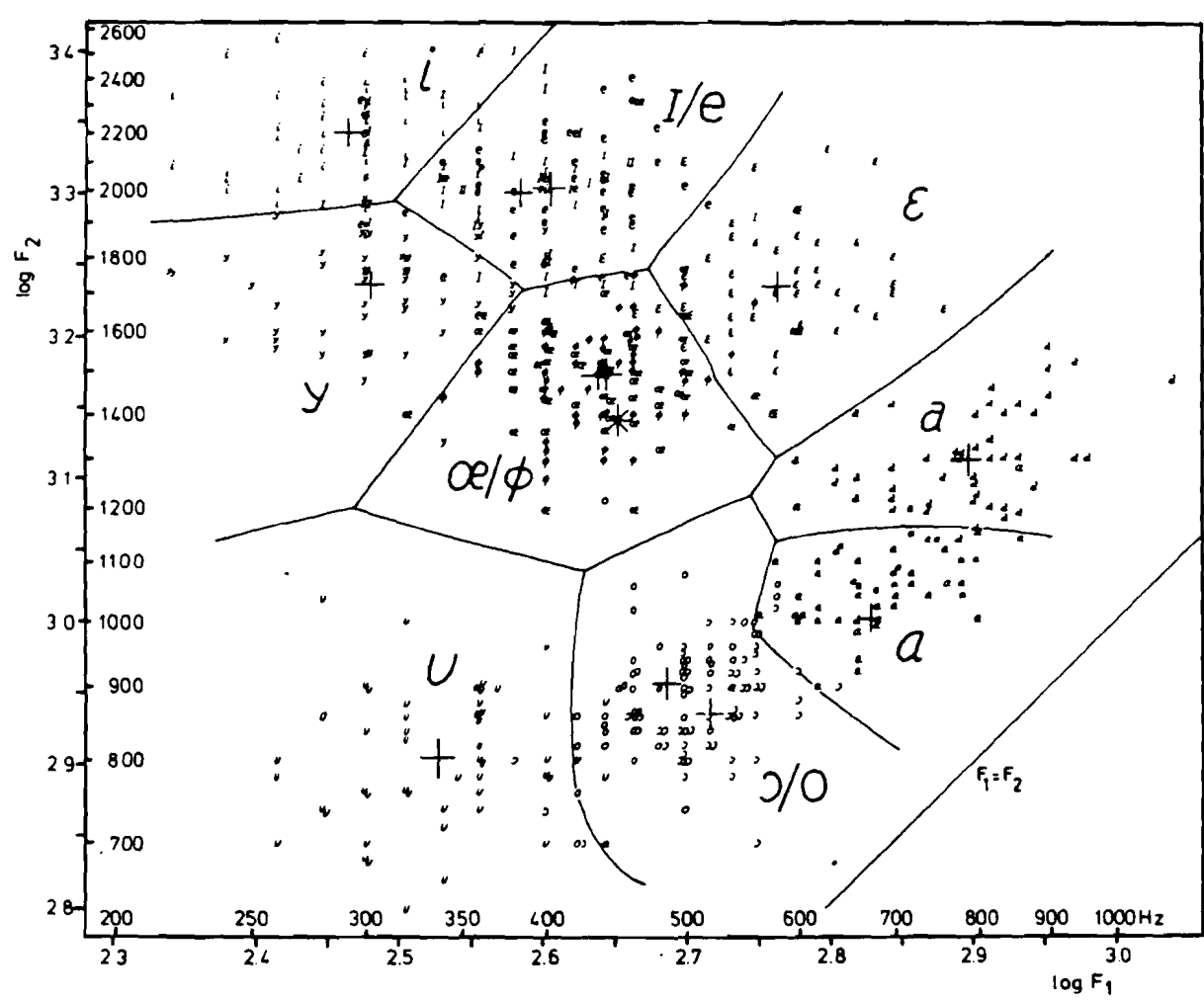


FIG. 1.  $\log F_2$  vs  $\log F_1$  of the  $50 \times 12$  individual vowel sounds. The + symbols indicate the average values of the 12 vowels. The curves represent the boundaries of the maximum-likelihood regions of the grouped data (three pairs of vowels grouped together). In this graph, no speaker-dependent correction is applied. (uit: Pols e.a. 1973)

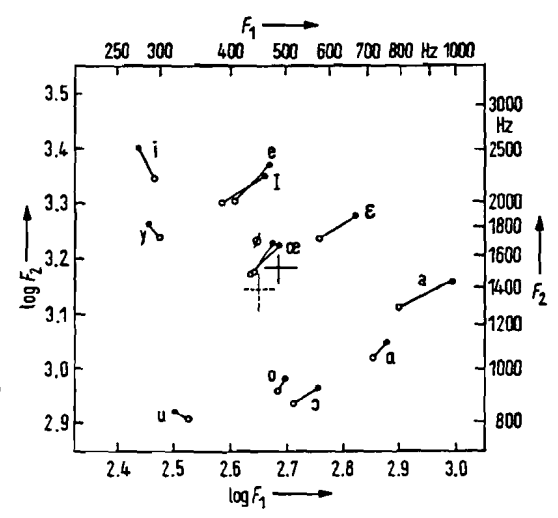


Fig.2. Average position of the vowel sounds in the  $\log F_1$ - $\log F_2$  plane both for the male and female speakers.  
 ● Female,  
 ○ male.  
 (uit: van Nierop e.a. 1973)

Dit soort vragen is reeds door verscheidene onderzoekers behandeld. Pols, Tromp & Plomp (1973) hebben een verschuiving toegepast, zodanig dat de zwaartepunten van de individuele klinkerdriehoeken in een voorgeschreven vast punt komen te liggen.

Wakita (1977) heeft een lineaire rek van de frequentie-as gebruikt, van het type  $f' = a f$ , en wel voor F1 en F2 in dezelfde mate.

De factor  $a$  wordt in verband gebracht met de "vocal tract length".

Barth & Chulliat (1978) hebben een lineaire transformatie van de vorm  $f' = a f + b$  onderzocht, voor F1 en F2 gecombineerd. De coëfficiënten  $a, b$  worden verkregen met orthogonale regressie.

Matsumoto & Wakita (1979) hebben niet-lineaire, monotone vervorming van de frequentie-as toegepast ("frequency warping").

In dit onderzoek zijn, zoals reeds aangeduid in de samenvatting, een aantal lineaire transformaties onderzocht, waaronder ook twee van de hierboven genoemde.



### Analyse van het spraaksignaal

In het volgende wordt beschreven, hoe het spraaksignaal geanalyseerd is, en welke problemen daarbij voorgekomen zijn.

Ook wordt aandacht besteed aan het autocorrelatie-venster, dat interessante mogelijkheden biedt bij spectrale analyse.

### Toegepaste analyse

Op het van microfoon of bandrecorder afkomstige spraaksignaal worden achtereenvolgens de hierna genoemde bewerkingen en berekeningen toegepast.

1 - samples, 8 kHz : het signaal wordt bemonsterd met een frequentie van 8 kHz, en met 12 bit per sample; de samples worden bewaard in het computer-geheugen, zodat alle overige bewerkingen zonder tijdnood uitgevoerd kunnen worden.

2 - preemphasis,  $a = 0.9$  : het signaal wordt digitaal gefilterd volgens

$$s'_n = s_n - a s_{n-1}$$

en hierdoor worden de hoge frequenties versterkt tov. de lage.

3 - tijdvenster,  $N = 198$ , Hamming : het signaal wordt vermenigvuldigd met een tijdvenster  $w_n$  volgens  $s'_n = s_n w_n$  ; er blijven  $N^{+1}$  achtereenvolgende samples over, die een zogenaamd frame vormen; gekozen is een Hamming-window

$$w_n = .54 - .46 \cos(2\pi n/N) \quad , \quad n \in [0, N] \quad .$$

4 - autocorrelaties,  $p = 12$  : er worden  $(p+1)$  autocorrelaties  $R_k$  berekend volgens

$$R_k = \sum_n s_n s_{n-k} \quad , \quad k \in [0, p] ;$$

de sommatie-index  $n$  doorloopt in principe alle gehele waarden, maar hier zijn slechts eindig veel termen van belang.

- 5 - LPC analyse,  $p = 12$  : er worden  $p$  filtercoëfficiënten  $a_k$  en het bijbehorende spectrum berekend volgens de LPC analyse; de theorie hiervan is tamelijk ingewikkeld, en verwezen zij naar Makhoul (1975) en naar Markel & Gray (1976) ; hierop zal nog teruggekomen worden.
- 6 - toppen van het spectrum : de toppen van het spectrum worden gezocht, en de frequenties in de maxima worden beschouwd als formantfrequenties; hiervan worden alleen de eerste twee, F1 en F2, verder gebruikt.

De analyse levert dus uiteindelijk "slechts" twee getallen op, de formantfrequenties F1 en F2, voor ieder frame van 25 ms . De volgende procedure is vrijwel standaard in het huidige spraakonderzoek.

#### Spectra bij de analyse

Er zijn bij de analyse verscheidene spectra aan de orde.

- 1 - Het spectrum van het na preemphasis en tijdvenster verkregen signaal  $s_n$

$$S(\omega) = \sum_n s_n e^{-j\omega n}$$

Hierbij hoort een energiespectrum  $P(\omega)$  volgens  $P(\omega) = |S(\omega)|^2$  .

- 2 - De benadering van het energiespectrum  $P(\omega)$  volgens de LPC-analyse

$$\hat{P}(\omega) = \frac{G^2}{|A(\omega)|^2}$$

Hierin is  $A(\omega)$  het spectrum van het LPC-filter  $a_k$

$$A(\omega) = \sum_{k=0}^p a_k e^{-j\omega k}$$

en  $G$  een evenredigheidsfactor.

De filtercoëfficiënten  $a_k$  worden in de LPC-analyse gedefinieerd als oplossing van het stelsel vergelijkingen

$$\sum_{k=0}^p a_k R_{k-n} = 0 \quad \text{voor } n \in [1, p] \quad ; \quad a_0 = 1$$

waarin  $R_k = \sum_n s_n s_{n-k}$  de autocorrelaties van het signaal  $s_n$  zijn.

3 - Het spectrum  $P(\omega)$  kan ook berekend worden volgens

$$P(\omega) = \sum_k R_k e^{-j\omega k}$$

Omdat er na het tijdvenster slechts eindig veel, namelijk  $N+1$  opeenvolgende samples overblijven, zijn er ook slechts eindig veel autocorrelaties, dwz.  $R_k = 0$  voor  $k \notin [-N, N]$ . Bovendien geldt steeds  $R_{-k} = R_k$ .  $P(\omega)$  wordt dus gegeven door een eindige Fourier-reeks

$$P(\omega) = \sum_{k=-N}^N R_k e^{-j\omega k} = R_0 + 2 \sum_{k=1}^N R_k \cos(\omega k)$$

Bij de analyse van het signaal  $s_n$  worden echter slechts de eerste  $(p+1)$  autocorrelaties berekend en gebruikt ( $p < N$ ). Dit betekent, dat de Fourier-reeks afgebroken wordt

$$P'(\omega) = \sum_{k=-p}^p R_k e^{-j\omega k} = R_0 + 2 \sum_{k=1}^p R_k \cos(\omega k)$$

Volgens de theorie van de Fourier-reeks is  $P'(\omega)$  de beste benadering voor  $P(\omega)$ , in de zin van de kleinste kwadratische afwijking, wanneer alleen de eerste termen tot en met  $p$  gebruikt mogen worden. Figuur 3 a - e toont een voorbeeld van de spectra  $P$ ,  $\hat{P}$  en  $P'$ . De toppen van  $\hat{P}$  zijn betrekkelijk smal, die van  $P'$  veel breder.  $P$  en  $\hat{P}$  zijn altijd positief, maar  $P'$  kan ook negatief worden.

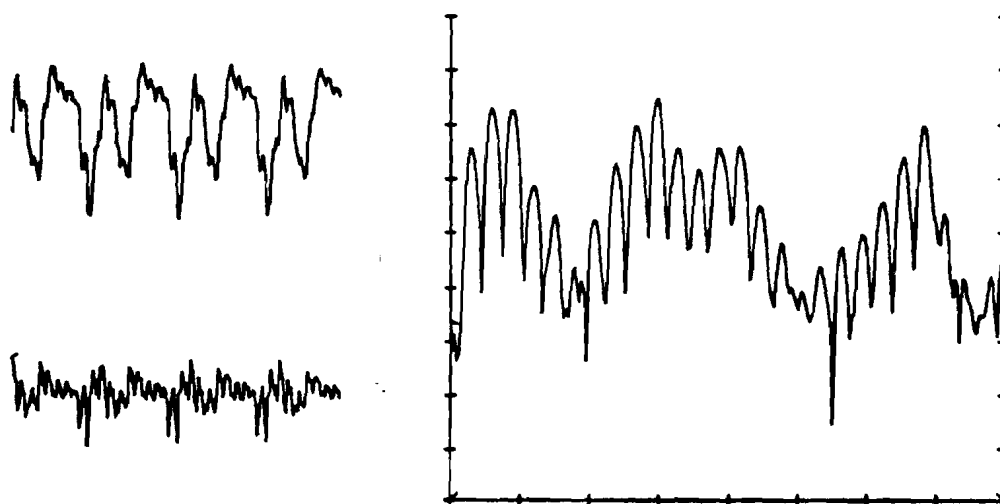


Fig. 3 a. Analyse van een klinker /p/ van een mannelijke spreker.  
Linksboven: tijdpatroon van een geselecteerd frame (25 ms) .  
Linksonder: tijdpatroon ná preemphasis. Hierna wordt een tijdvenster  
(Hamming) toegepast van dezelfde lengte.  
Rechts: energiespectrum van het signaal ná preemphasis en tijdvenster.  
Horizontale as: 0 - 4 kHz . Verticale as: 10 dB per schaaldeel.

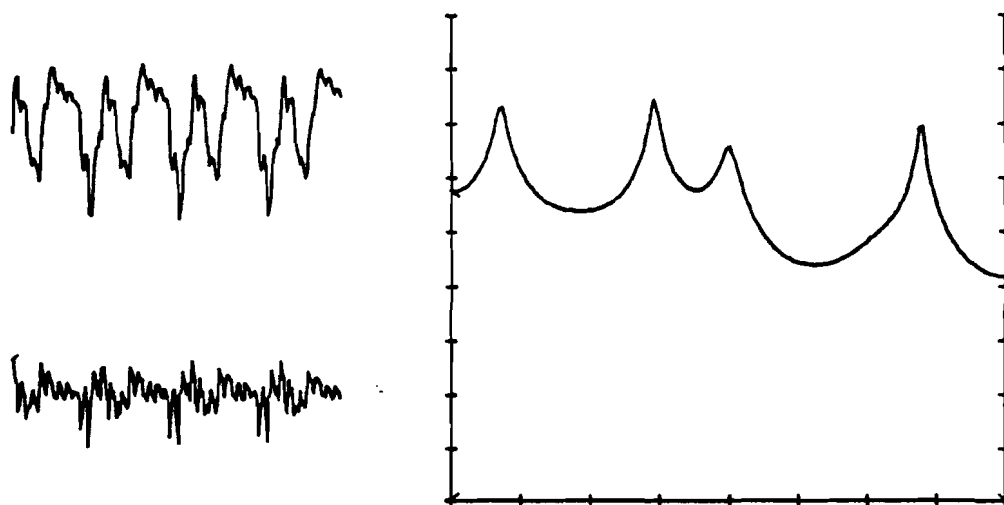
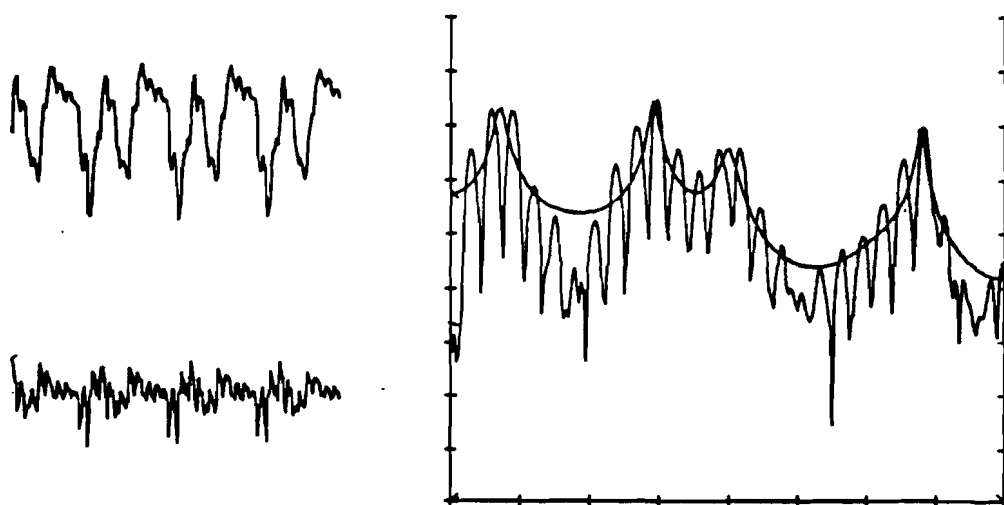


Fig. 3 b.

Benadering van het energiespectrum

Fig. 3 c. volgens LPC-analyse ( $p = 12$ )

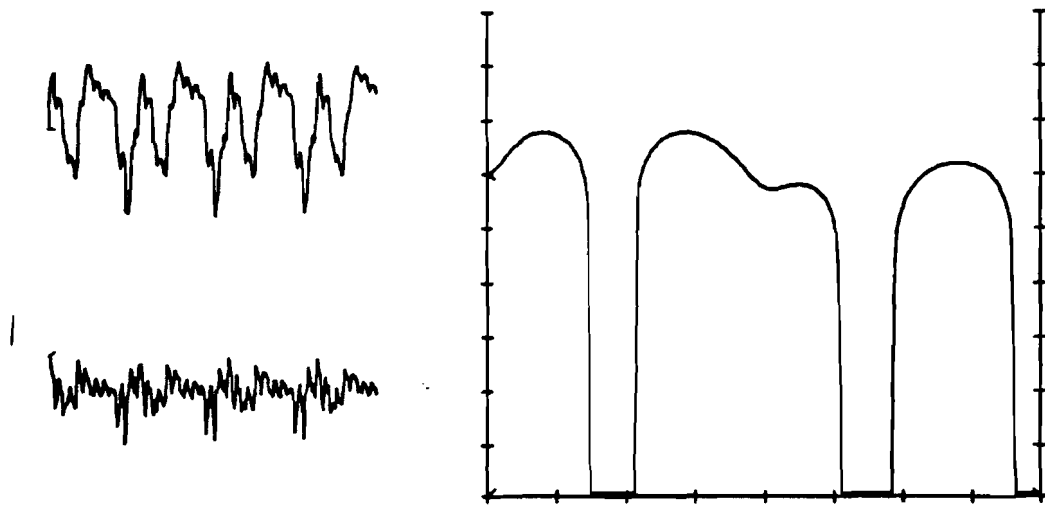
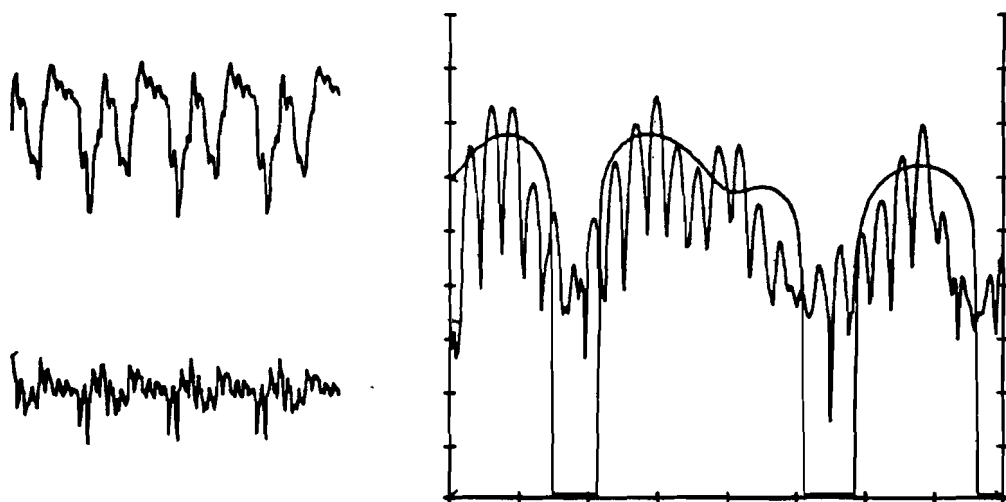


Fig. 3 d.

Benadering van het energiespectrum via transformatie

Fig. 3 e. van de  $(p + 1)$  autocorrelaties.



### Het autocorrelatie-venster

Het afbreken van de reeks kan ook beschouwd worden als het leggen van een rechthoekig venster over de autocorrelaties heen

$$R'_k = w_k R_k \quad , \quad \text{met} \quad w_k = \begin{cases} 1 & \text{als } k \in [-p, p] \\ 0 & \text{als } k \notin [-p, p] \end{cases}$$

Hiermee wordt

$$P'(\omega) = \sum_k R'_k e^{-j\omega k}$$

en deze formule is geheel analoog aan de formule voor  $P(\omega)$ . De sommatie-index  $k$  doorloopt hier in principe alle gehele waarden. Blijkbaar is dit rechthoekige autocorrelatie-venster ook reeds impliciet aanwezig in stap 4 van de analyse (p.6). Men kan zich, net als bij het tijdvenster, afvragen of een ander, minder abrupt venster niet gunstiger is. Wat gunstig is, hangt natuurlijk af van het doel van de analyse, en een algemeen antwoord is moeilijk te geven. In ieder geval zijn er op dit terrein interessante mogelijkheden tot experimenteren. Hier lijkt een fraaie toepassingsmogelijkheid te liggen voor het Newton-venster (binomiaal venster, zie ook de appendix)

$$w_k = \binom{2M}{k+M} \quad \text{voor } k \in [-M, M]$$

Het spectrum van dit venster wordt gegeven door

$$W(\omega) = \sum_k w_k e^{-j\omega k} = \left( 2 \cos(\omega/2) \right)^{2M}$$

en dit is steeds positief. Het spectrum  $P'$  kan geschreven worden als convolutie van  $W$  en  $P$ , en is dus ook steeds positief. Daarentegen kan het spectrum van het rechthoekige venster wél negatieve waarden aannemen, en dit verklaart waarom  $P'$  eveneens negatieve waarden kan krijgen, zoals in figuur 3 d. In het algemeen zal men dus een autocorrelatie-venster met een positief spectrum wensen.

Vensters, die geschreven kunnen worden als autocorrelatie-functie van een ander venster, hebben altijd een positief spectrum. Tot deze soort behoort het driehoekige Fejér-venster, als autocorrelatie-functie van een rechthoekig venster (Harris, 1978), en bovengenoemd Newton-venster, als autocorrelatie-functie van een ander Newton-venster (appendix).

Het autocorrelatie-venster heeft uiteraard invloed op het LPC-spectrum  $\hat{P}$ , wanneer de filtercoëfficiënten  $a_k$  berekend worden met behulp van  $R'_k$  in plaats van  $R_k$ .

Dit is reeds onderzocht door Tohkura, Itakura & Hashimoto (1978), en zij hebben gevonden, dat het Newton-venster zeer geschikt is, om de toppen van  $\hat{P}$  af te vlakken. Dit is gewenst bij de resynthese van spraak.

In figuur 4 a - b is het gewone LPC-spectrum met  $p = 24$  te zien.

In figuur 4 c - e is een Newton-venster met  $M = 104$  toegepast, gevolgd door een rechthoekig venster met ongewijzigde  $p$  (dus in feite een afgekapt Newton-venster).

Merkwaardig en verrassend is, dat het spectrum van de autocorrelaties  $P'$  (c) en het LPC-spectrum  $\hat{P}$  (d) nu nauwelijks verschillen (e).

Figuur 5 a - e toont hetzelfde, met  $p = 50$  en  $M = 416$ .

Het aantal toppen in c en d is veel kleiner dan het theoretisch maximale aantal ( $\approx \frac{1}{2}p$ ).

De getallen  $M$  zijn zó gekozen, dat de effectieve breedte van de vensters gelijk is aan, resp. tweemaal zo groot is als de effectieve breedte van een rechthoekig venster met  $p = 12$ . Onder effectieve breedte wordt hier verstaan: de spreiding van de verdeling  $w_k$ , zoals die berekend wordt in de statistiek.



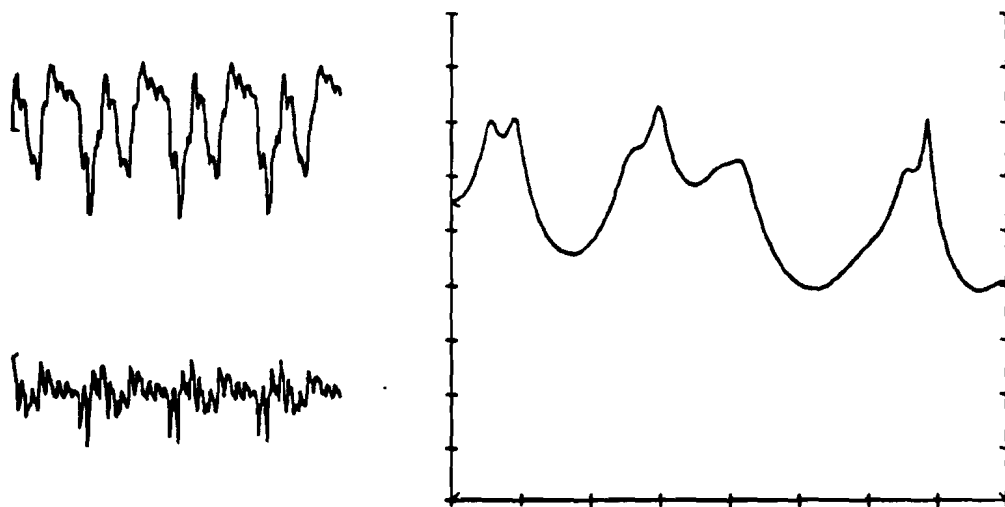
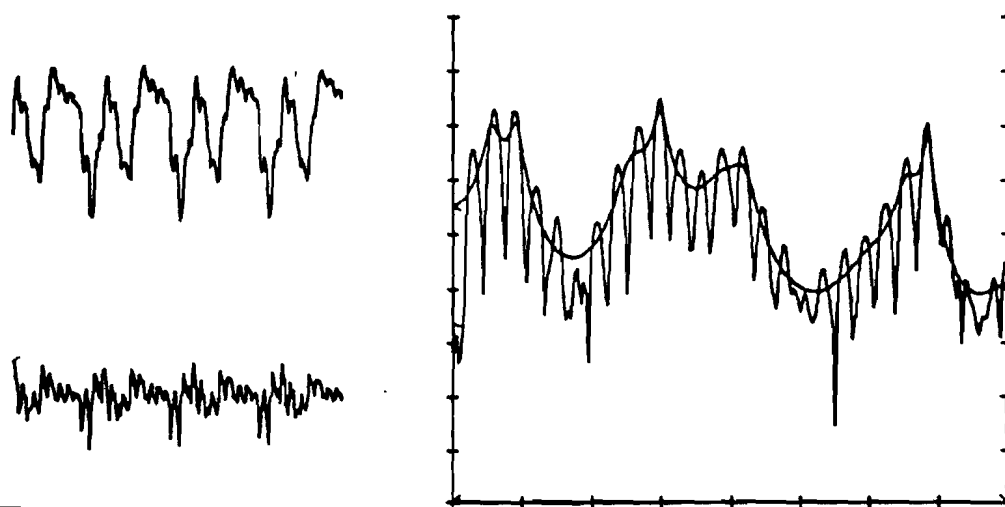


Fig. 4 a. Normaal LPC-spectrum met  $p = 24$  .

Fig. 4 b.



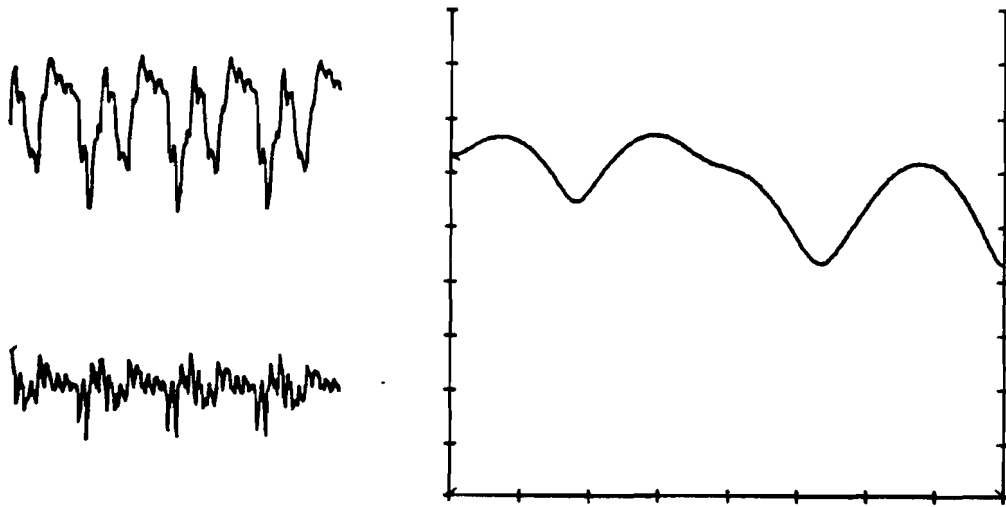
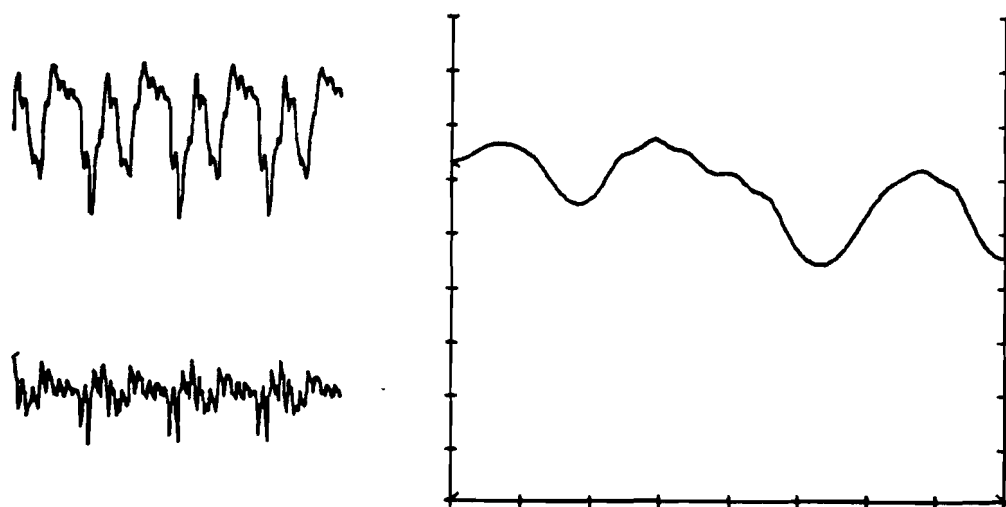


Fig. 4 c.

Toepassing van een Newton-venster als autocorrelatie-venster (zie tekst), met  $M = 104$ ,  $p = 24$ .

Fig. 4 d.

Boven het spectrum van de autocorrelaties, onder het LPC-spectrum, beide ná toepassing van het venster.



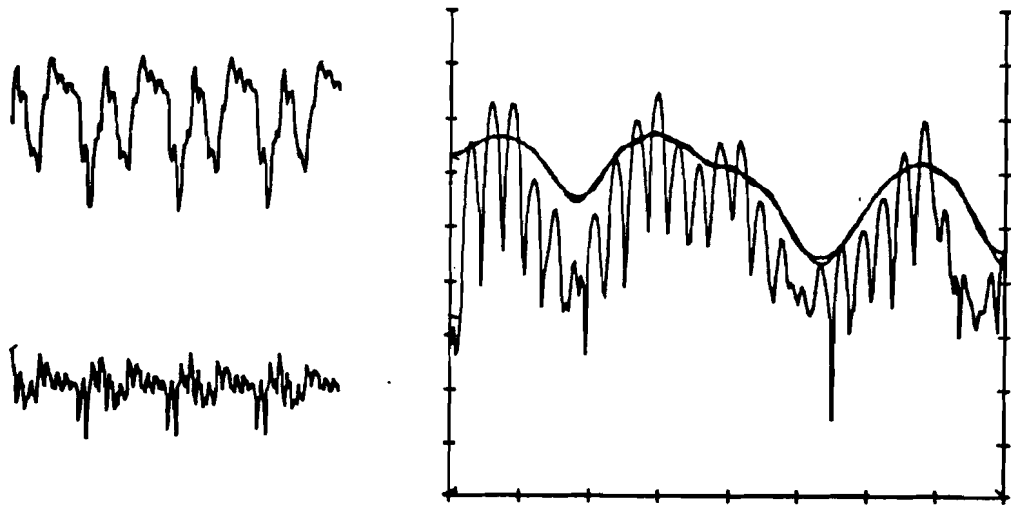


Fig. 4 e. De spectra van figuur c en d , samen met het spectrum van het signaal ná het tijdvenster.

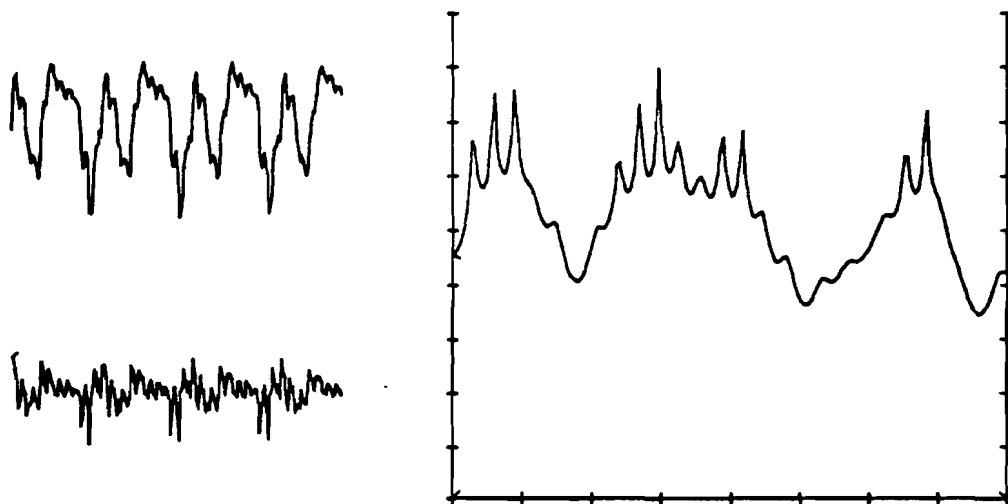
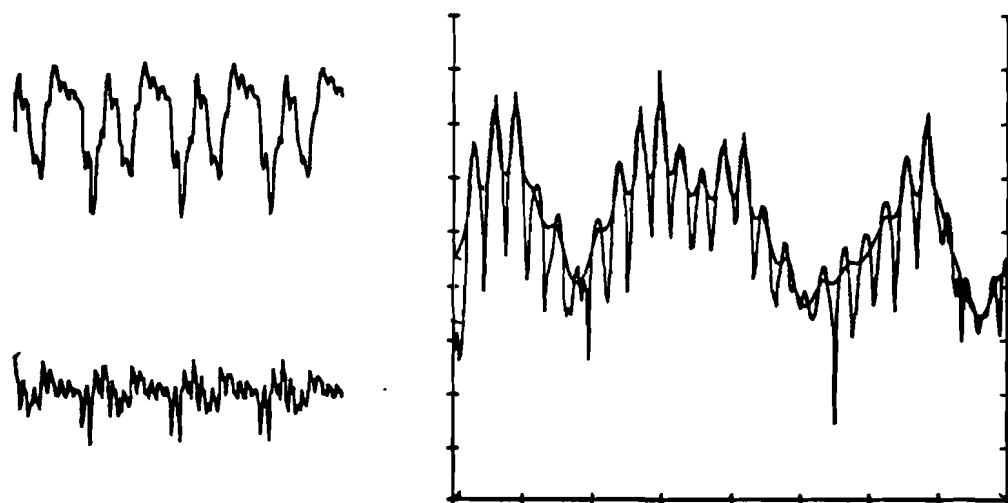


Fig. 5 a.

Normaal LPC-spectrum met  $p = 50$  .

Fig. 5 b.



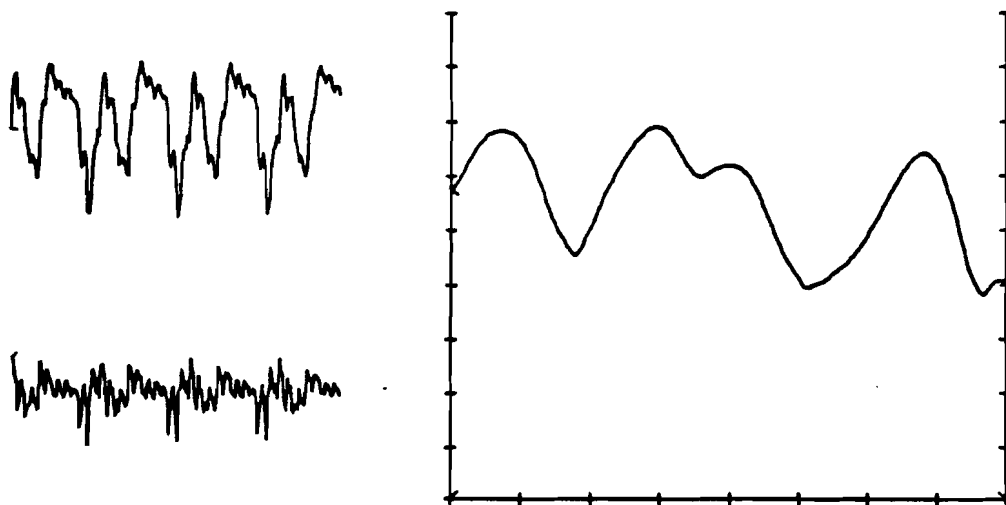
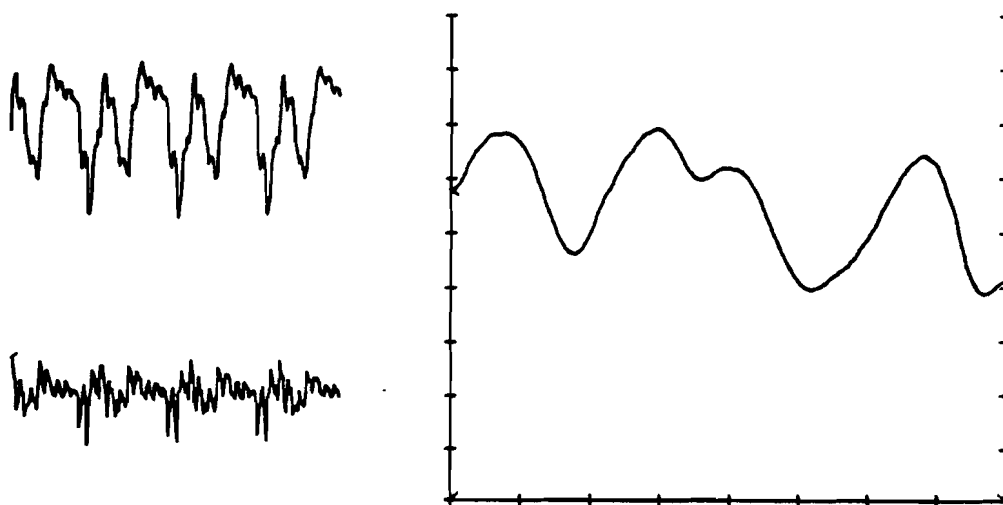


Fig. 5 c.

Toepassing van een Newton-venster als autocorrelatie-venster (zie tekst), met  $M = 416$ ,  $p = 50$ .

Fig. 5 d.

Boven het spectrum van de autocorrelaties, onder het LPC-spectrum, beide ná toepassing van het venster.



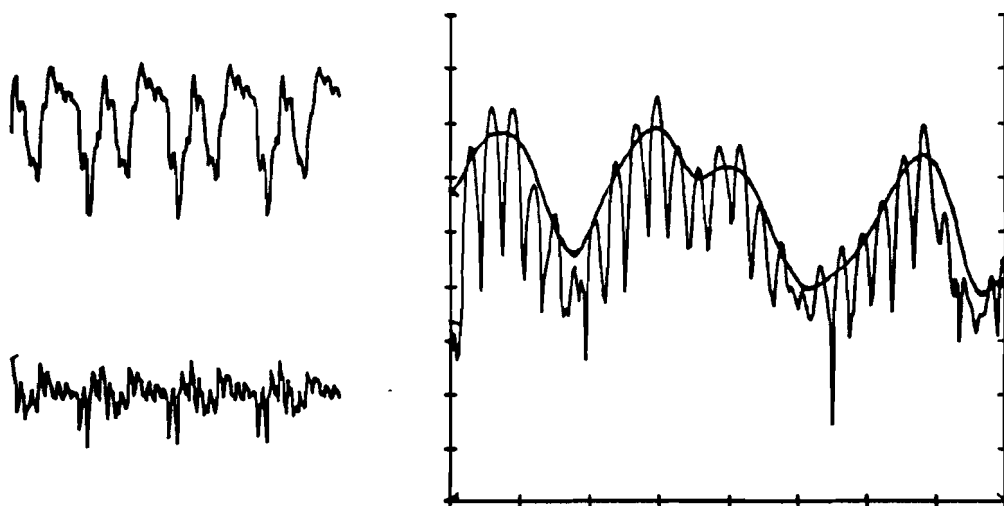


Fig. 5 e. De spectra van figuur c en d , samen met het spectrum van het signaal ná het tijdvenster.

Interessant is ook de toepassing van een parabolisch autocorrelatievenster (figuur 6 a - d )

$$w_k'' = \left( 1 + \frac{a}{p^2} k^2 \right) w_k \quad , \quad w_k \text{ rechthoekig op } [-p, p]$$

Hiermee gaat  $P'(\omega)$  over in

$$\begin{aligned} P''(\omega) &= \sum_k \left( 1 + \frac{a}{p^2} k^2 \right) w_k R_k e^{-j\omega k} \\ &= \left( 1 - \frac{a}{p^2} \frac{d^2}{d\omega^2} \right) \sum_k w_k R_k e^{-j\omega k} \\ &= P'(\omega) - \frac{a}{p^2} \frac{d^2}{d\omega^2} P'(\omega) \end{aligned}$$

Door de invloed van de tweede afgeleide zullen bij positieve  $a$  de toppen hoger en scherper worden, bij negatieve  $a$  vlakker, en de buigpunten aan weerskanten van een top blijven op hun plaats (in een lineair assenstelsel) .

Dit opscherpende of afvlakkende effect werkt blijkbaar door in het LPC-spectrum  $\hat{P}$  (figuur 6 a - b , zie ook figuur 3 b ) . De toppen verschuiven maar weinig in de richting van de frequentie-as (figuur 6 c - d ) .

Bij een zogenaamde verborgen formant, waar bijna een top aanwezig is, kan met een positieve  $a$  dikwijls een echte top geproduceerd worden.

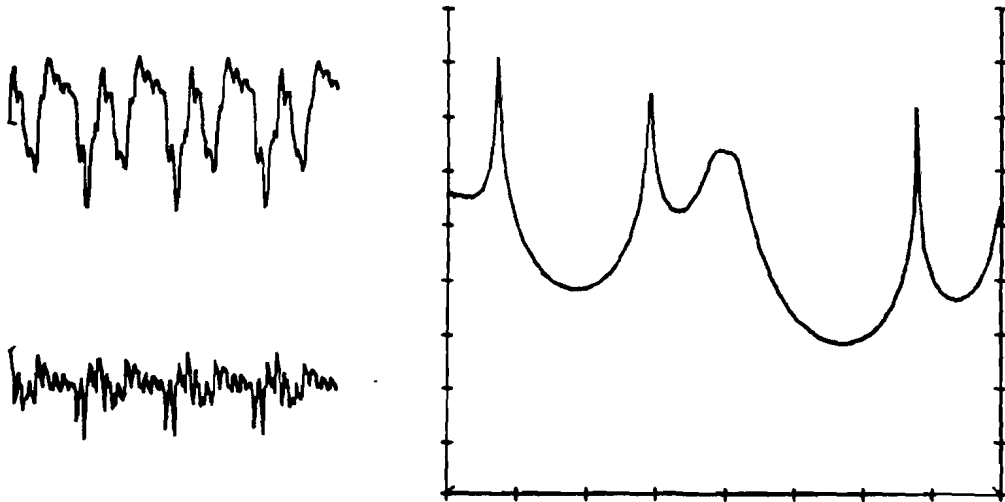
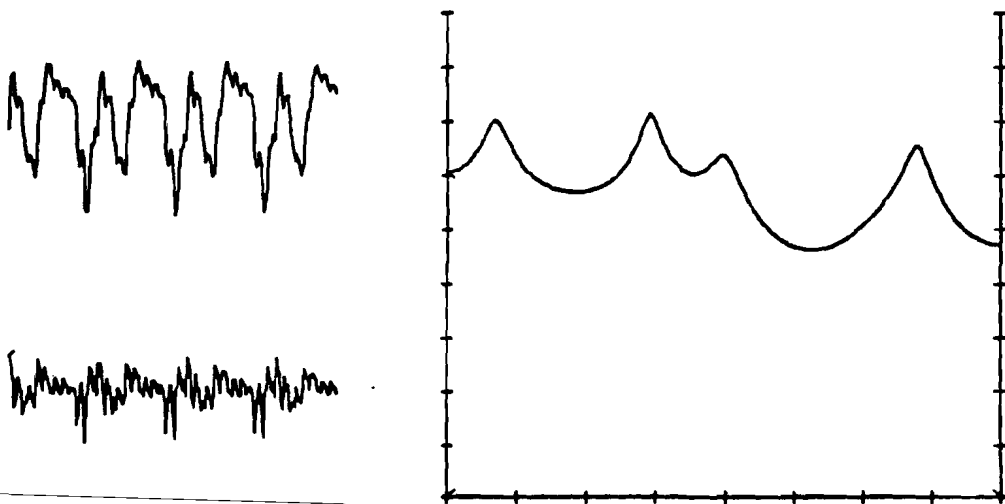


Fig. 6 a.

LPC-spectrum na toepassing van een parabolisch  
 Fig. 6 b. autocorrelatievenster (zie tekst) met  $a = +.5$  (boven)  
 resp.  $a = -.5$  (onder) en  $p = 12$ .





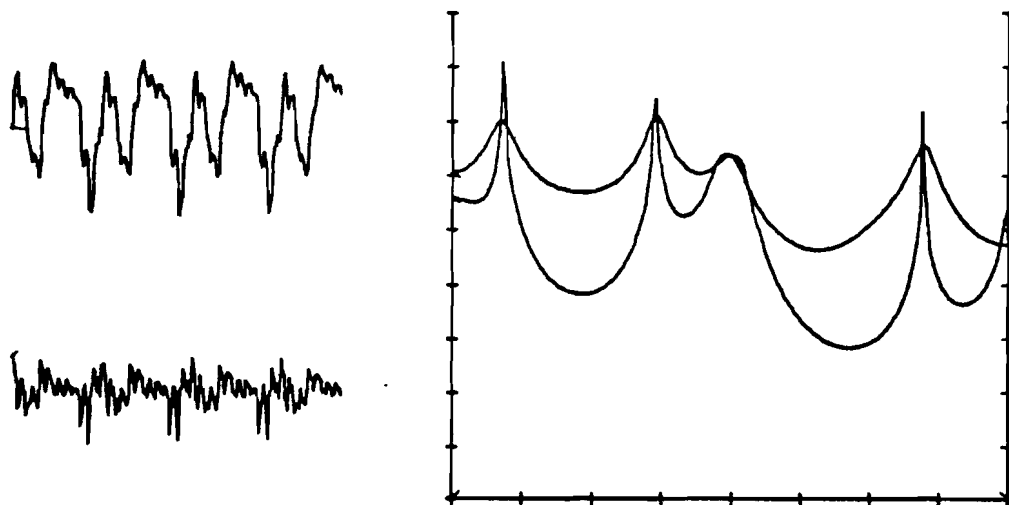
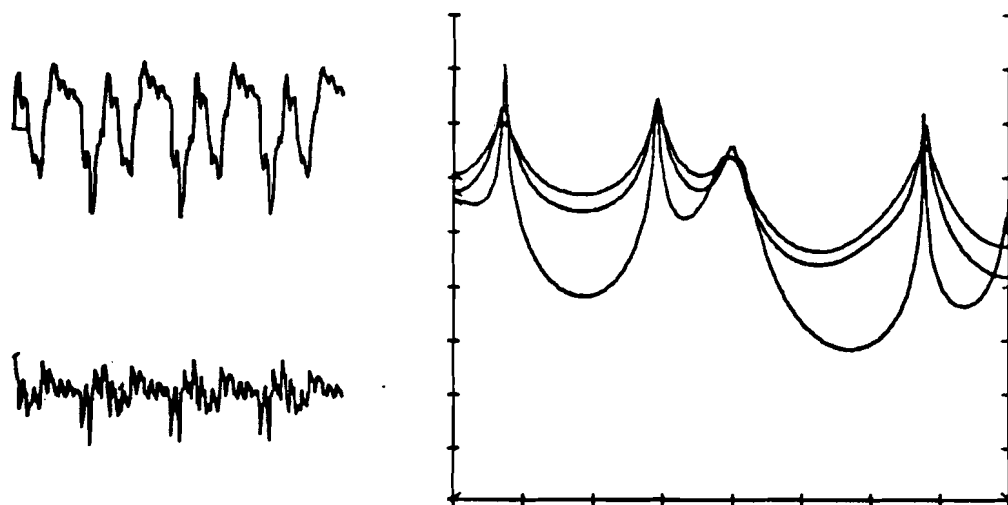


Fig. 6 c.

Idem, en onder tevens het gewone LPC-spectrum

Fig. 6 d. van figuur 3 b (middelste lijn) .



### Problemen bij de analyse van spraak

Het resultaat van de analyse voldeed niet altijd aan de verwachtingen. Bij sommige frames was één van de formantfrequenties F1,F2 duidelijk fout, in vergelijking met andere frames van eenzelfde klinker.

Figuur 7 a toont een klinker /i/ van een mannelijke spreker. De formanten F1,F2 liggen hier ver uit elkaar. In het dal ertussenin is een tamelijk vlakke top zichtbaar, die normaal niet als formant beschouwd wordt. Dit komt ook wel voor bij de klinker /y/ .

Het was niet moeilijk, om de ongewenste top te elimineren, via een eenvoudig criterium: wanneer het niveau van de tweede top lager was dan 0.01 maal het niveau van de eerste top, werd hij overgeslagen en de volgende top gekozen als formant. Dit criterium werd bij twee sprekers gebruikt (sprekers 3 en 4) , en wel bij alle klinkers, dus niet alleen bij /i/ en /y/ . Bij de andere klinkers leverde dit geen problemen op.

Figuur 7 c toont een klinker /u/ van een mannelijke spreker. De formanten F1,F2 liggen hier dicht bij elkaar. De tweede formant is niet echt als top aanwezig, alleen als een soort uitstulping, verbreding. Dit verschijnsel komt ook voor bij de klinkers /o/ en /u/ . Het bleek moeilijk, om hier iets aan te verbeteren. Weliswaar kon met een andere keuze van de parameter p of met gebruik van een parabolisch autocorrelatie-venster een top geproduceerd worden, maar dan ging er bij andere klinkers wel weer iets mis. De mislukte gevallen zijn derhalve gewoon weggelaten, het waren er maar enkele. Er was ook een geval van een klinker /u/ dat op geen enkele manier "goed" te krijgen was (figuur 7 e) . Dit frame volgt in de tijd onmiddellijk op het frame van figuur 7 c . Bij beschouwing van het oorspronkelijke spectrum na het tijdvenster (figuur 7 f) ziet men, dat er nauwelijks een tweede formant aanwezig is, althans daar waar men er een zou wensen.

Bij vrouwenstemmen is de grondtoon, dwz. de frequentie van de stembandtrilling, relatief hoog. Het spectrum van klinkers bestaat uit hogere harmonischen van de grondtoon, en de afstand tussen de componenten is dus vrij groot (figuur 7 g) .

De toppen van de LPC-benadering zullen dikwijls samenvallen met een top van het spectrum, of in de buurt ervan liggen.

De formantfrequenties, zoals de analyse die oplevert, zijn dus gekoppeld aan de grondtoon, in die zin, dat er een voorkeur is voor gehele veelvouden van de grondtoon. De grondtoon kan bijvoorbeeld 300 Hz zijn, en in de klinkerdriehoek is dit een zeer grote stap (zie bijv. figuur 1) . Een dergelijke stap is voldoende, om van een /u/ een /o/ te maken.

Samengevat zijn er dus de volgende problemen en tekortkomingen bij de gekozen analyse: bij /i,y/ zijn er ongewenste toppen; bij /a,o,u/ worden toppen gemist; bij vrouwenstemmen geeft de hoge grondtoon moeilijkheden.

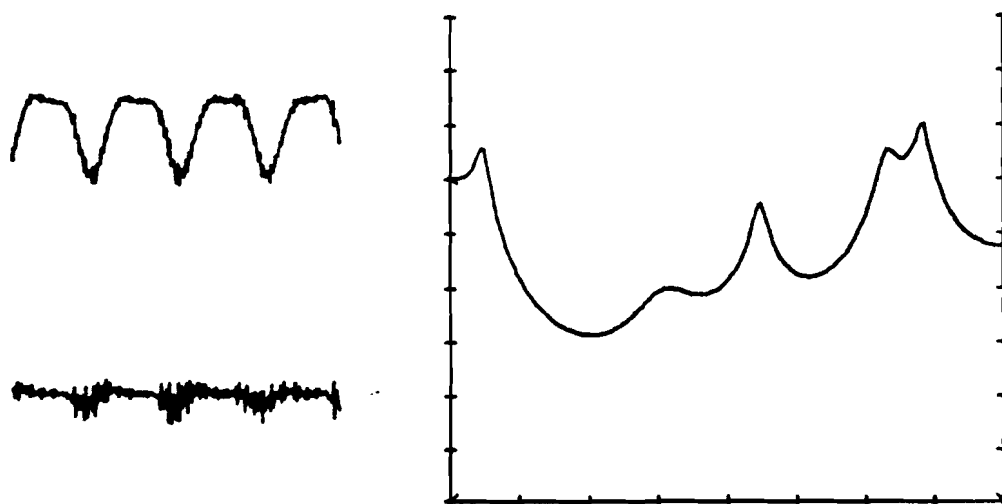
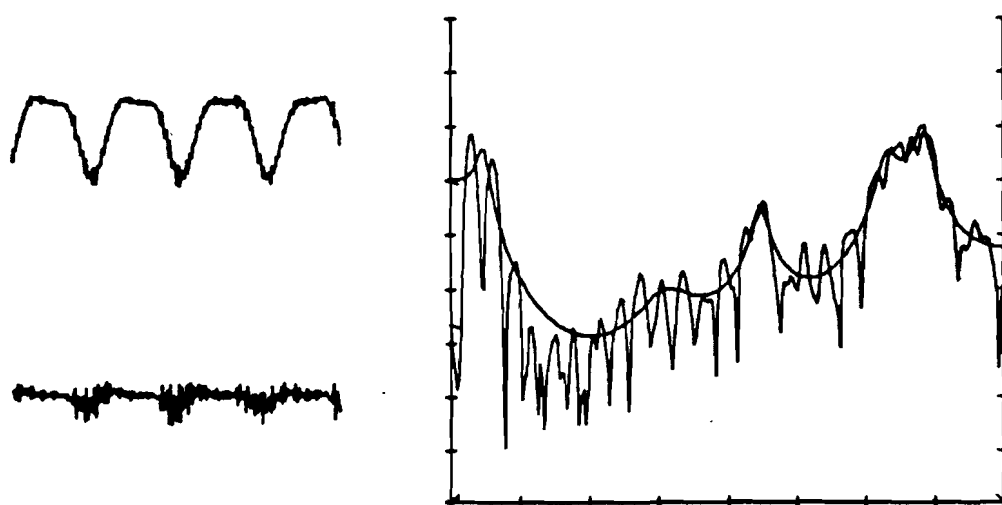


Fig. 7 a.

Klinker /i/ van een mannelijke spreker. Tussen F1 en F2

Fig. 7 b. is een ongewenste top aanwezig.



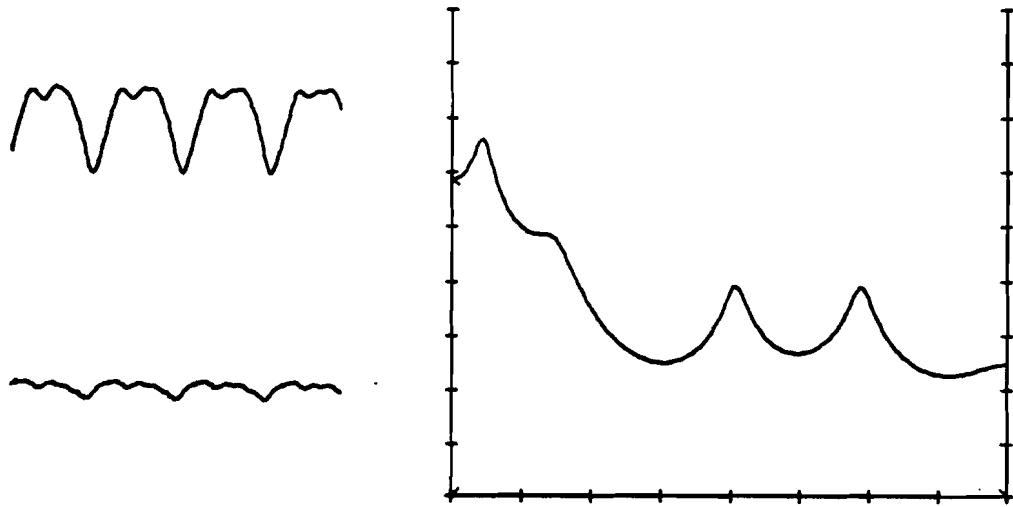
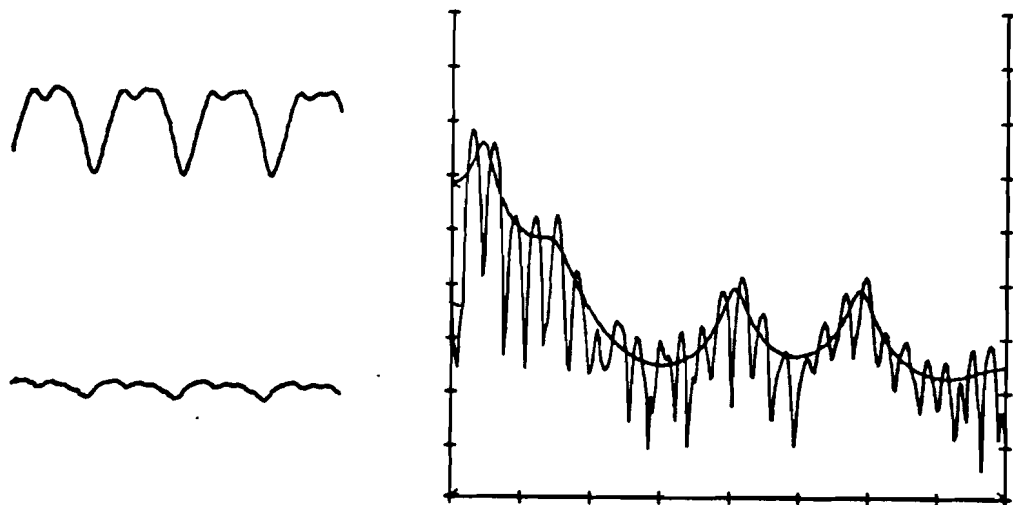


Fig. 7 c.

Klinker /u/ van een mannelijke spreker. F2 is niet  
 Fig. 7 d. als echte top aanwezig.



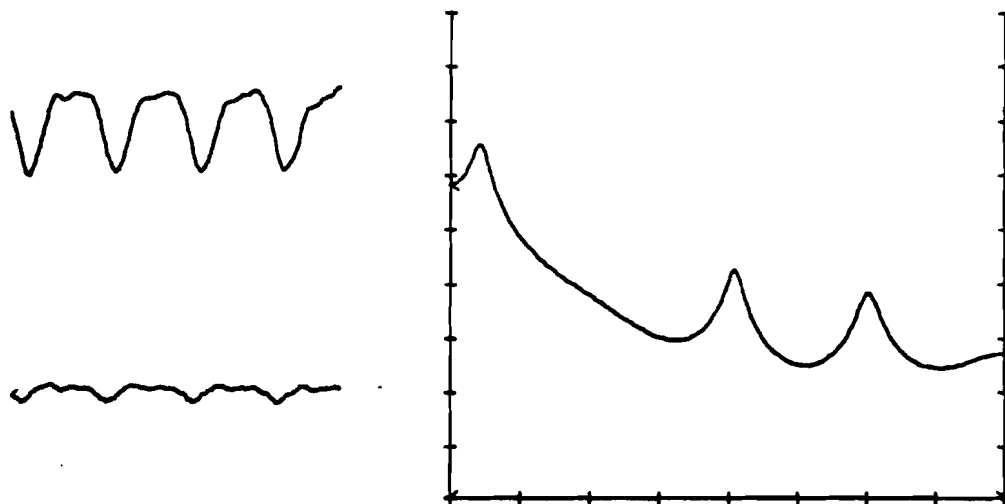
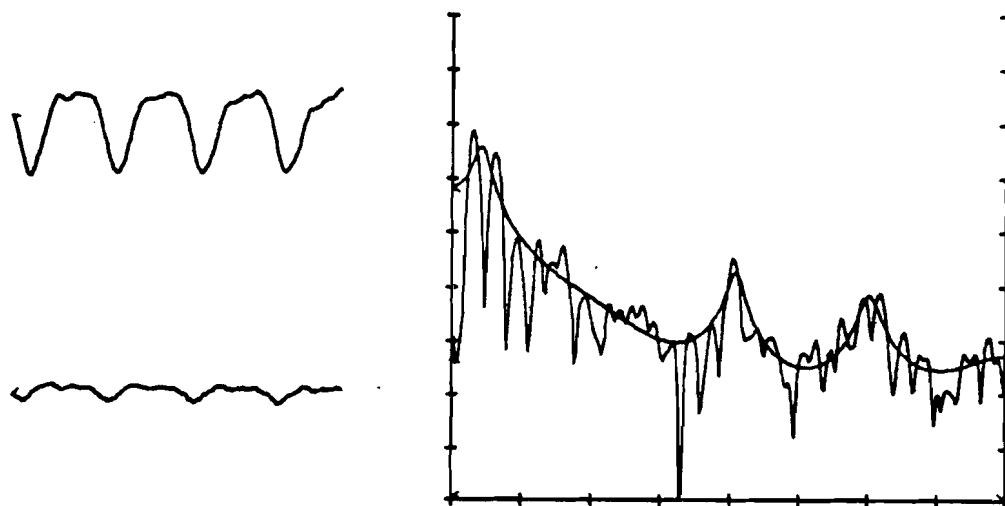


Fig. 7 e.

Klinker /u/ van een mannelijke spreker. Dit frame volgt  
 Fig. 7 f. onmiddelijk op het frame van figuur c en d. F2 lijkt  
 geheel verdwenen te zijn.



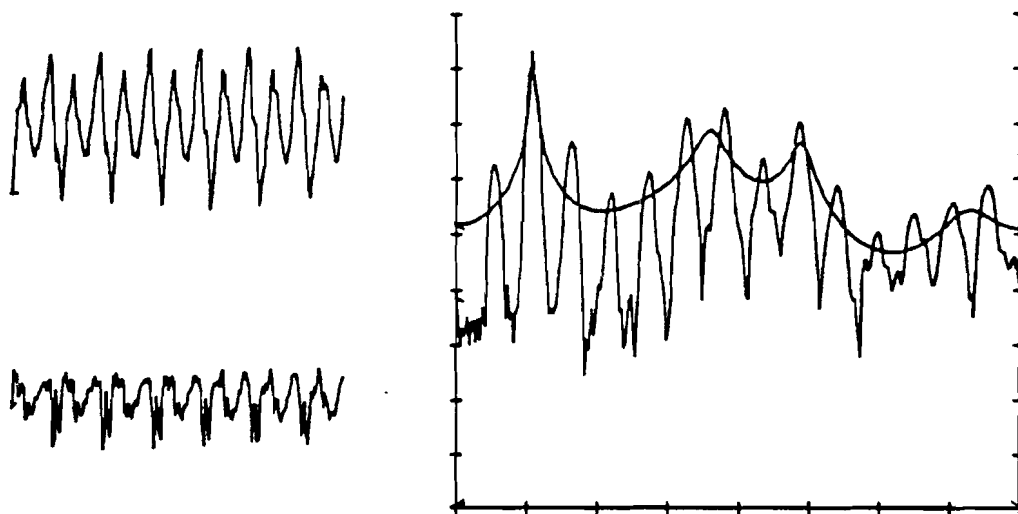


Fig. 7 g. Klinker /ø/ van een vrouwelijke spreker. De hoge grondtoon geeft problemen (zie tekst) .

### Klinkerdriehoeken van vier sprekers

In het volgende wordt beschreven, welk materiaal gebruikt is en hoe dit verder verwerkt is, resulterend in de klinkerdriehoeken van vier sprekers.

Onderzocht werden de zeven lange klinkers /u, o, a, e, i, y, ø/. Deze klinkers waren steeds in eenzelfde context geplaatst: voorafgegaan door /h/ en gevolgd door /s/. Er werden dus de volgende woorden gebruikt: hoës, hoos, haas, hees, hies, huus, heus.

Met deze woorden werden de volgende vier proefzinnen gevormd, waarin de woorden in een wisselende volgorde voorkomen.

```
regel 1 :   hoës  hoos  haas  hees  hies  huus  heus  hoës
regel 2 :   haas  hies  hoës  heus  huus  hoos  hees  haas
regel 3 :   heus  hees  hoos  huus  hoës  hies  haas  heus
regel 4 :   hees  huus  hies  hoos  heus  haas  hoës  hees
```

Aan de proefpersonen werd gevraagd, om de woorden langzaam en afzonderlijk uit te spreken. Het laatste woord van iedere regel werd naderhand niet meer gebruikt, omdat aan het einde van een zin de toonhoogte nogal kan dalen.

De tekst werd voorgelezen door vijf proefpersonen, 3 mannelijke en 2 vrouwelijke, en op de band opgenomen. Vanwege problemen bij de analyse, zoals reeds eerder beschreven, moest één van de vrouwelijke sprekers verder buiten beschouwing blijven.

De bandopname werd bemonsterd en verder met behulp van de computer verwerkt en geanalyseerd.

Allereerst werd de plaats van de klinkers in het tijdsignaal bepaald, en werden uit elke klinker drie opeenvolgende frames (tijdvakken) van 25 ms gekozen. Dit gebeurde langs visuele weg, met behulp van het beeldscherm van de computer en de cursor, als volgt.

De computer tekent een tijdpatroon van ongeveer 10 sec (400 frames), en de gebruiker wijst met de cursor de woorden aan (figuur 8 a).

Daarna worden de aangewezen woorden vergroot afgebeeld (figuur 8 b), telkens 325 ms (13 frames).



Het periodieke karakter van de klinkers is goed zichtbaar. Men kan nu het middelste van de drie gewenste frames aanwijzen.

De frames werden gekozen in het midden van de klinkers, waar de amplitude maximaal is en tamelijk stationair.

Om geheugenruimte te sparen, werden nadien alleen de geselecteerde frames bewaard. Er bleven dus voor iedere spreker 84 frames over, 12 voor iedere klinker.

De frames werden geanalyseerd met behulp van LPC-analyse, zoals reeds eerder beschreven (p. 6), en van ieder frame werden uiteindelijk de eerste twee formantfrequenties  $F_1, F_2$  bepaald.

De twee formantfrequenties van een frame kunnen ook weergegeven worden door één punt in het  $F_1$ - $F_2$ -vlak. Zo worden de klinkerdriehoeken verkregen (figuur 9 a - d).

In plaats van punten zijn gecentreerde symbolen gebruikt, die aangeven om welke klinker het gaat. De kruisen geven de gemiddelden (zwaartepunten) per klinker aan.

In de figuren 10 a - d met logaritmische schaal is eerst de logaritme genomen, en daarna het gemiddelde berekend.

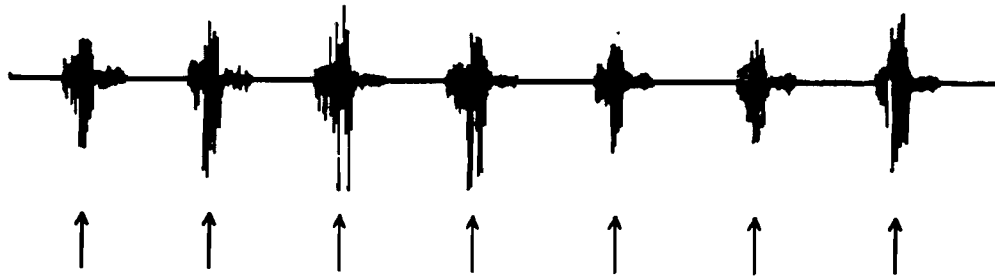


Fig. 8 a. Tijdpatroon van de woorden: hoes hoos haas hees hies huus heus. Dit zijn 450 frames à 25 ms (11.25 sec). Met behulp van de cursor kunnen de woorden aangewezen worden.



Fig. 8 b. Tijdpatroon van een gedeelte van het woord heus. Dit zijn 13 frames à 25 ms. Met behulp van de cursor worden drie opeenvolgende frames geselecteerd.

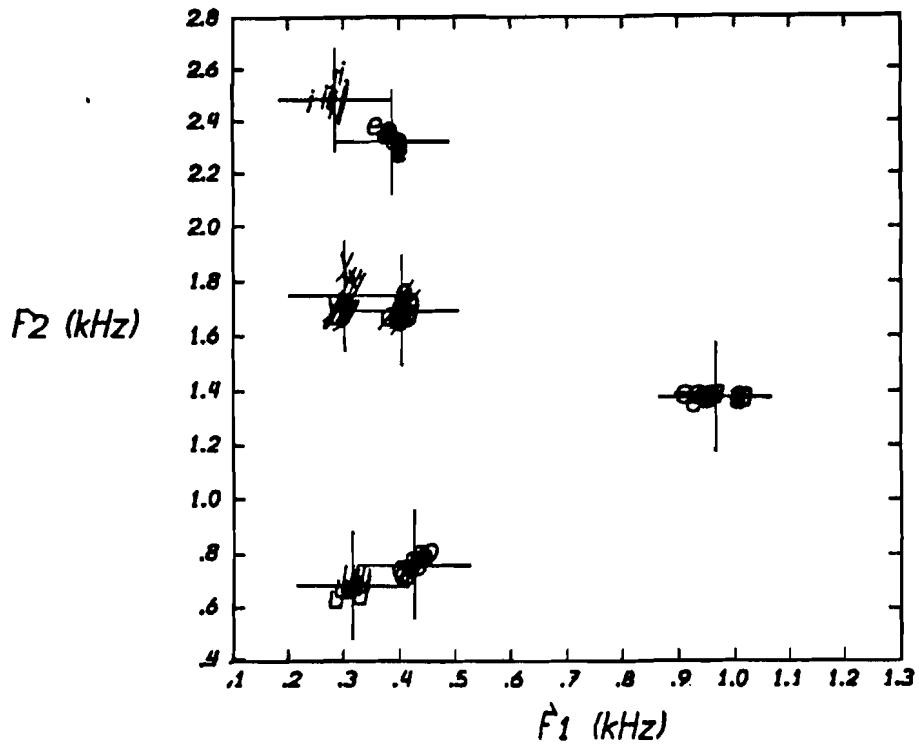
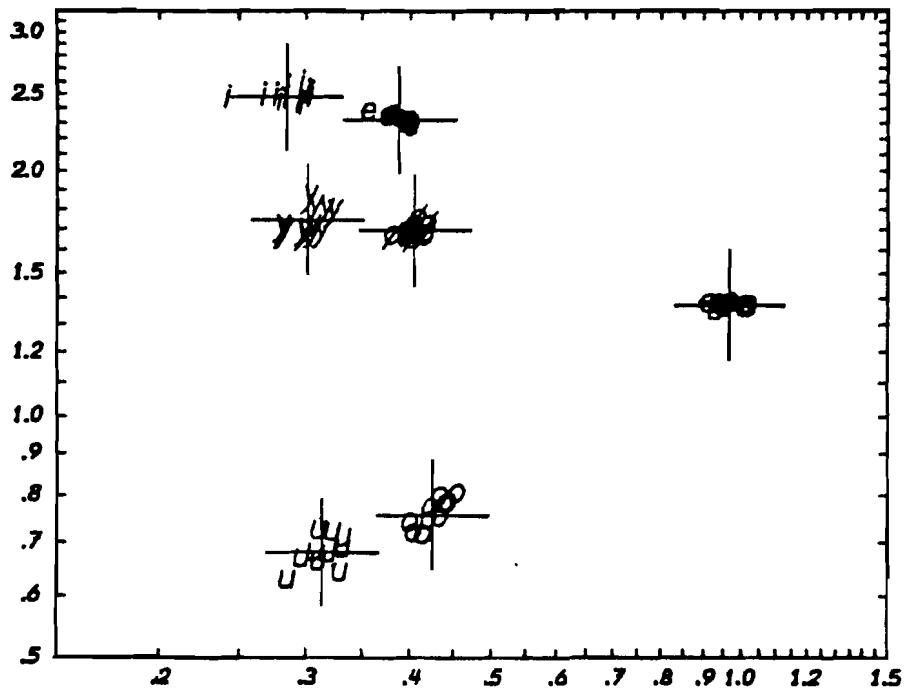


Fig. 9 a.

Klinkerdriehoek van spreker 1 (m.)

Fig. 10 a.



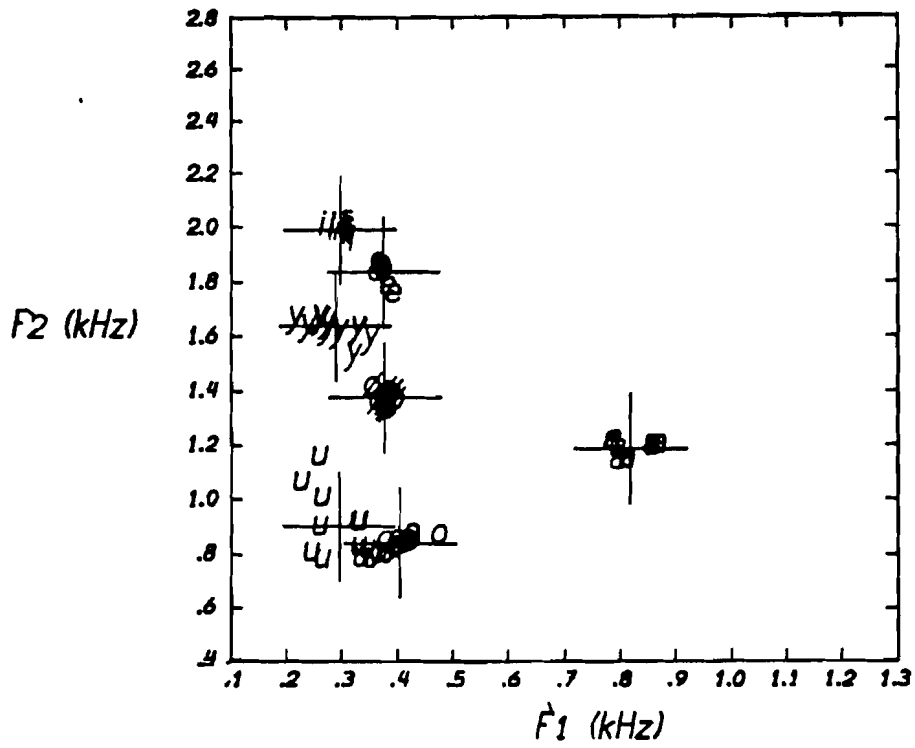
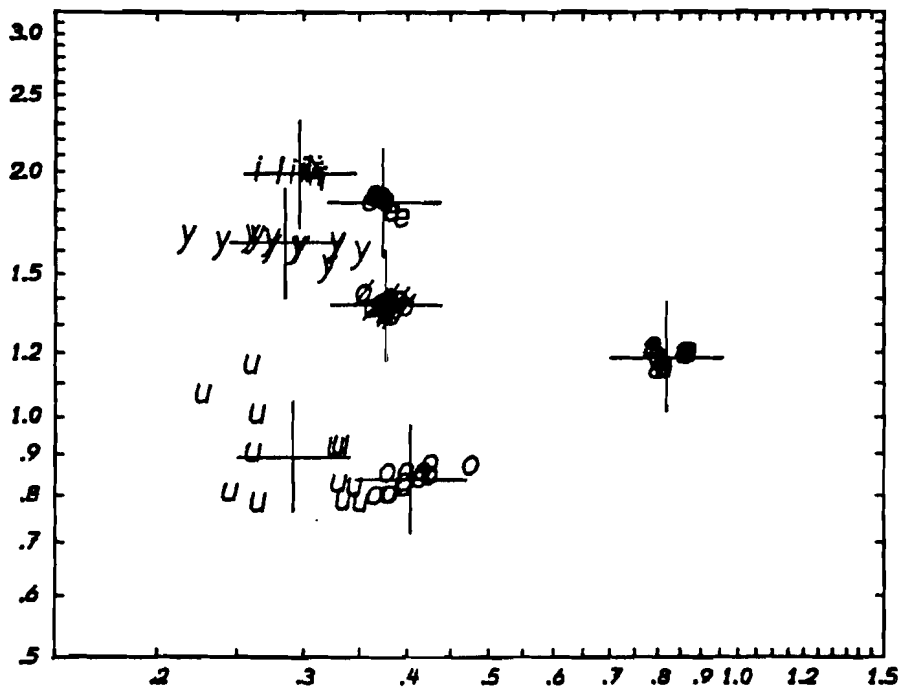


Fig. 9 b.

Klinkerdriehoek van spreker 2 (m.)

Fig.10 b.



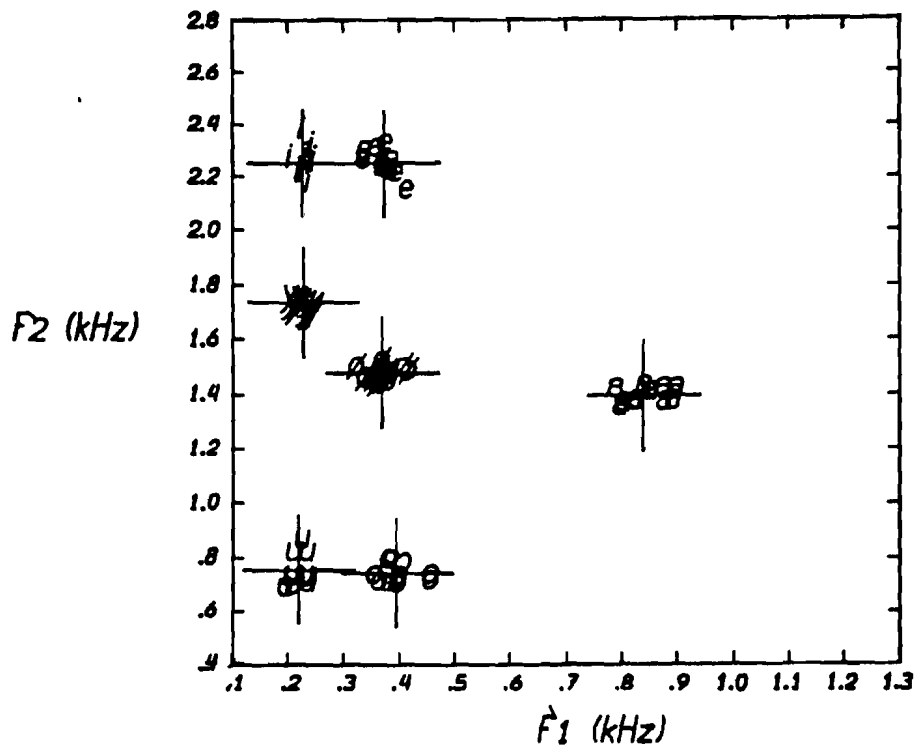
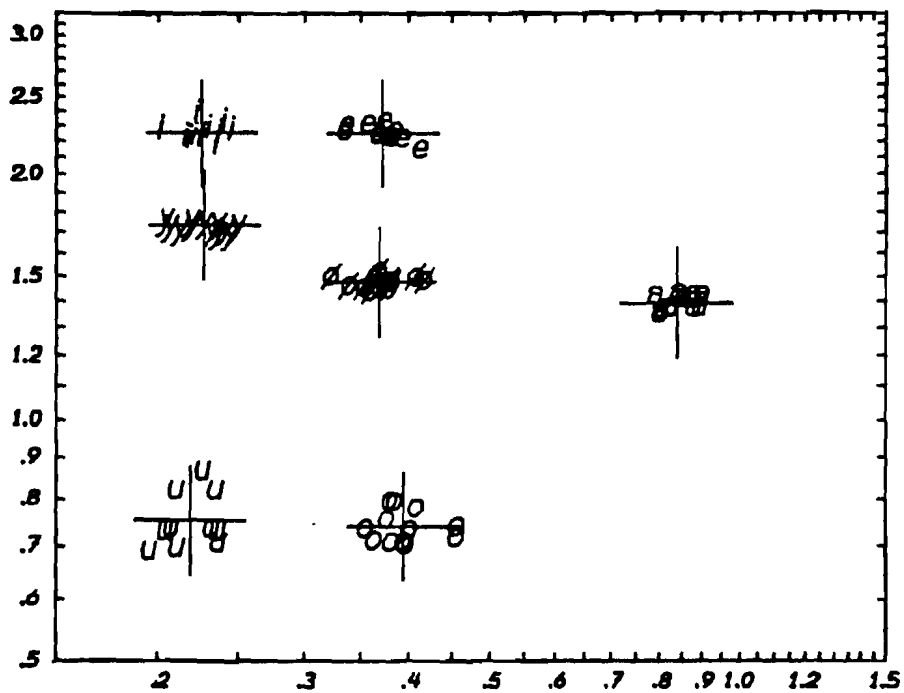


Fig. 9 c. Klinkerdriehoek van spreker 3 (m.)  
 Fig.10 c.



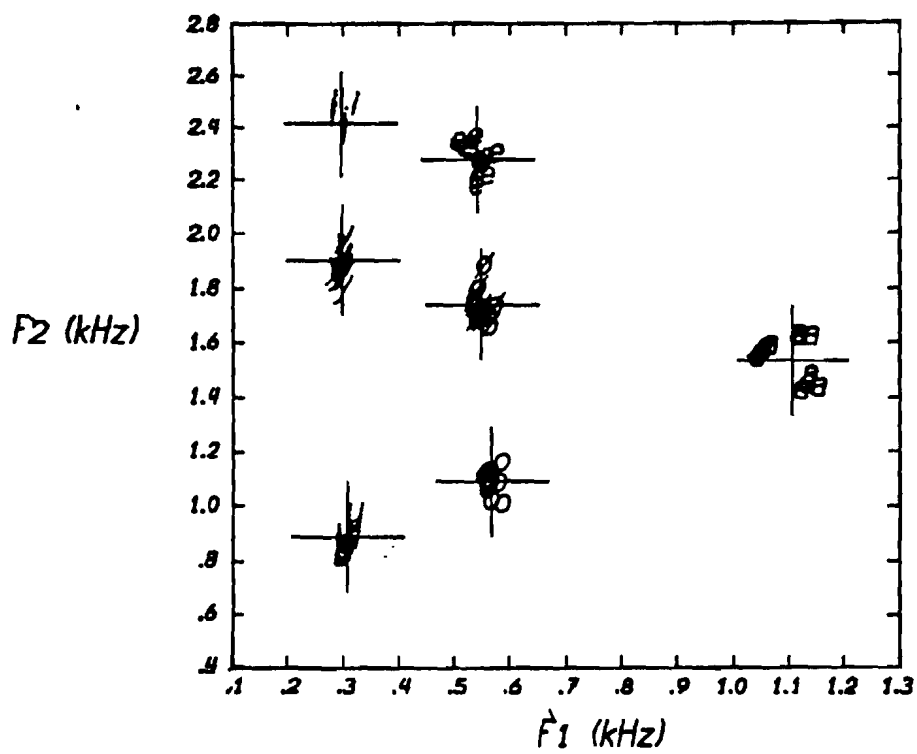
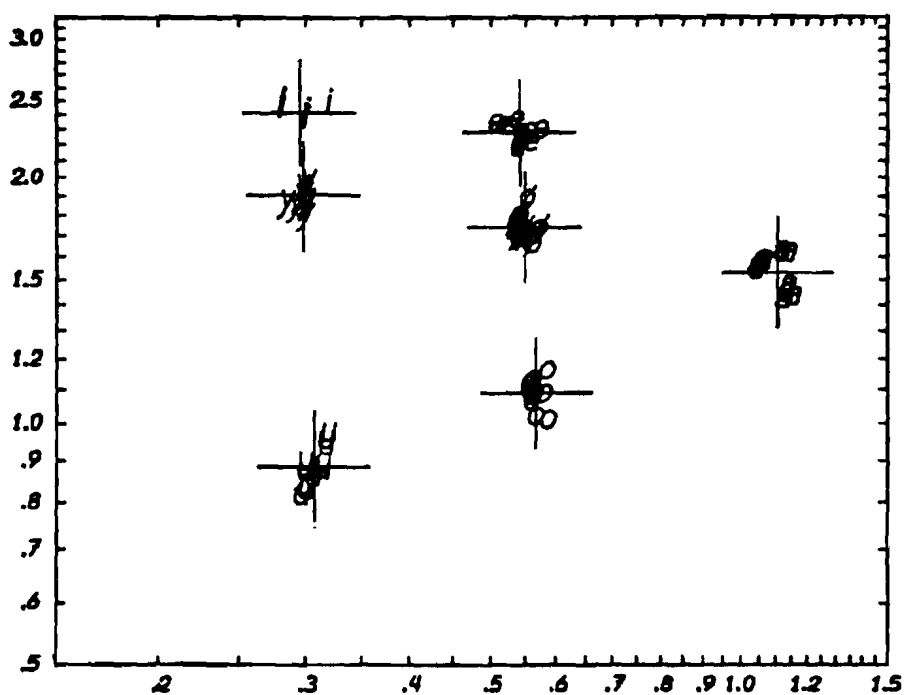


Fig. 9 d.

Klinkerdriehoek van spreker 4 (v.)

Fig. 10 d.



### Transformaties van de klinkerdriehoeken

De gebieden van de klinkers zijn goed van elkaar gescheiden, voor één spreker. Wanneer de klinkerdriehoeken van de vier sprekers in één figuur afgebeeld worden, wordt het beeld ineens zeer troebel (figuur 11 a en 12 a). De puntenwolken van de klinkers overlappen en penetreren elkaar. Duidelijk is, dat een automatisch herkenningssysteem, gebaseerd op F1 en F2, hier grote moeite mee zal hebben. De klinkerdriehoeken van de vier sprekers hebben ruwweg dezelfde vorm en dezelfde indeling, zodat het mogelijk lijkt, om door schuiven, draaien, rekken en vervormen de klinkerdriehoeken van twee sprekers in elkaar over te laten gaan. Ook zouden de klinkerdriehoeken van alle vier sprekers afgebeeld kunnen worden op de klinkerdriehoek van één standaardspreker.

In eerste instantie zal men aan lineaire transformaties denken. Draaiingen lijken op het eerste gezicht niet erg nuttig, omdat de lijnen u - y - i en o - ø - e bij alle vier sprekers vrijwel dezelfde richting hebben.

In dit onderzoek zijn een aantal tamelijk eenvoudige lineaire transformaties gekozen, die alle bestaan uit verschuiven en uit één- of tweezijdig rekken in de richting van de coördinaat-assen.

In de hierna volgende beschrijving zal het formantvlak F1, F2 aangeduid worden met x, y. Bij logaritmische assen staat x, y voor log F1, log F2.

Verder zal gebruik gemaakt worden van het zwaartepunt  $m_x, m_y$  en de afmetingen  $s_x, s_y, s$  van de klinkerdriehoek van een spreker.

Deze parameters worden berekend volgens

$$m_x = \bar{\sum}_i x_i \quad ; \quad s_{xx} = \bar{\sum}_i (x_i - m_x)^2 = \bar{\sum}_i x_i^2 - m_x^2$$

$$m_y = \bar{\sum}_i y_i \quad ; \quad s_{yy} = \bar{\sum}_i (y_i - m_y)^2 = \bar{\sum}_i y_i^2 - m_y^2$$

$$s_x = \sqrt{s_{xx}} \quad ; \quad s_y = \sqrt{s_{yy}} \quad ; \quad s = \sqrt{s_{xx} + s_{yy}} \quad .$$

Het symbool  $\bar{\sum}_i$  is een afkorting voor:  $\frac{1}{n} \sum_{i=1}^n$  .

De getallen  $x_i, y_i$  vertegenwoordigen hier de (zeven) gemiddelde klinkers van een spreker. Voor de bepaling van de parameters van een klinkerdriehoek zijn dus de klinkers geconcentreerd gedacht in hun gemiddelde, dus in de kruispunten.

Met de transformaties worden punten  $x, y$  verplaatst naar  $x', y'$ .

Parameters zónder accent horen bij de originele ongetransformeerde klinkerdriehoek van een spreker. Parameters mét accent (') horen bij de klinkerdriehoek, die ontstaat na de toepassing van een transformatie.

De onderzochte transformaties zijn als volgt gedefinieerd.

1 - Een verschuiving en een tweezijdige rek, zodanig dat de frequenties  $x_i, y_j$  van bepaalde klinkers een voorgeschreven vaste waarde krijgen:

$$x' = [(x - x_1) x_2' - (x - x_2) x_1'] / (x_2 - x_1)$$

$$y' = [(y - y_1) y_2' - (y - y_2) y_1'] / (y_2 - y_1)$$

waarin  $x_1, x_2$  de gemiddelde  $x$  van klinker /y/, /a/

en  $y_1, y_2$  de gemiddelde  $y$  van klinker /u/, /i/

2 - Een verschuiving en alzijdige rek, zó dat het zwaartepunt  $m_x, m_y$  in een vast punt  $m_x', m_y'$  komt te liggen en de afmeting  $s$  een vaste waarde  $s'$  krijgt:

$$x' = (x - m_x) \frac{s'}{s} + m_x'$$

$$y' = (y - m_y) \frac{s'}{s} + m_y'$$

3 - Een verschuiving en tweezijdige rek, zó dat het zwaartepunt  $m_x, m_y$  in een vast punt  $m_x', m_y'$  komt te liggen en de afmetingen  $s_x, s_y$  een vaste waarde  $s_x', s_y'$  krijgen:

$$x' = (x - m_x) \frac{s_x'}{s_x} + m_x'$$

$$y' = (y - m_y) \frac{s_y'}{s_y} + m_y'$$



4 - Verschuiving en rek alleen voor  $F_1$ , zó dat  $m_x$  en  $s_x$  vaste waarden  $m'_x$  en  $s'_x$  krijgen:

$$x' = (x - m_x) \frac{s'_x}{s_x} + m'_x$$

$$y' = y$$

5 - Een verschuiving, zó dat het zwaartepunt  $m_x, m_y$  op een voorgeschreven plaats  $m'_x, m'_y$  komt te liggen:

$$x' = x - m_x + m'_x$$

$$y' = y - m_y + m'_y$$

6 - Verschuiving en eenzijdige rek, zó dat  $m_x, m_y$  en  $s_x$  vaste waarden  $m'_x, m'_y$  en  $s'_x$  krijgen:

$$x' = (x - m_x) \frac{s'_x}{s_x} + m'_x$$

$$y' = y - m_y + m'_y$$

7 - Een alzijdige rek vanuit de oorsprong, zó dat de afstand van het zwaartepunt tot de oorsprong een vaste waarde krijgt:

$$x' = x \frac{r'}{r} \quad \text{met } r' = \sqrt{m_x'^2 + m_y'^2}$$

$$y' = y \frac{r'}{r} \quad r = \sqrt{m_x^2 + m_y^2}$$

8 - Een verschuiving evenwijdig aan de lijn  $x = y$ , zodanig dat het zwaartepunt op een vaste lijn met vergelijking  $x + y = 2c'$  komt te liggen, die loodrecht staat op de eerste lijn:

$$x' = x - c + c' \quad \text{met } c' = (m'_x + m'_y)/2$$

$$y' = y - c + c' \quad c = (m_x + m_y)/2$$

Deze transformaties zijn bij iedere spreker afzonderlijk toegepast. De hiertoe benodigde parameters  $m_x, m_y$  enz. (zonder accent) zijn voor iedere spreker apart bepaald, met behulp van de originele, ongetransformeerde klinkerdriehoek.

De overige parameters  $m'_x, m'_y$  enz. (met accent) zijn steeds vast gekozen. Deze parameters kunnen beschouwd worden als de parameters van een standaardspreker.

In de figuren 11 en 12 b - i zijn de klinkerdriehoeken van de vier sprekers gelijktijdig weergegeven, nadat een van de transformaties is toegepast. Bij logaritmische assen is eerst de logaritme genomen, en daarna de transformatie uitgevoerd in het  $\log F1 - \log F2$  vlak.

Transformatie 3 lijkt op het oog de beste resultaten te geven: de gebieden van de klinkers zijn hier het meest geconcentreerd en goed van elkaar gescheiden (het contact u/o is afkomstig van spreker 2) .

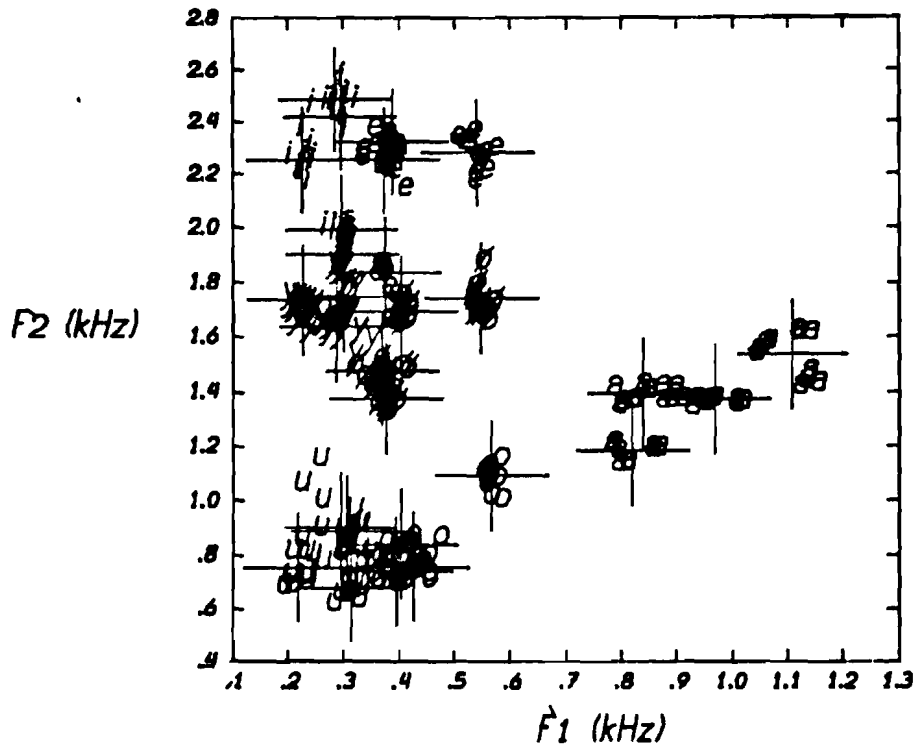
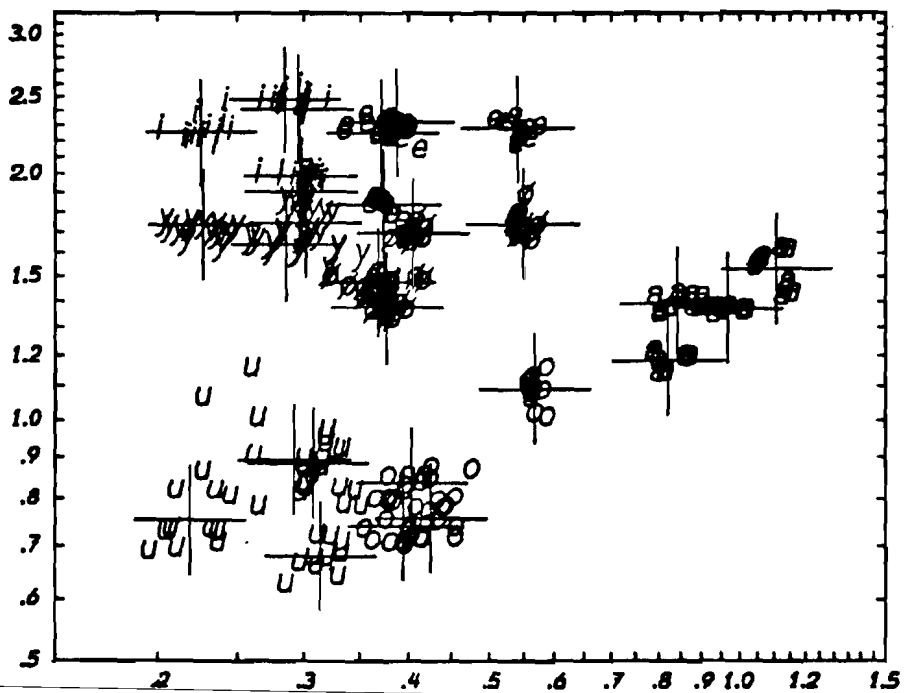


Fig. 11 a.

Klinkerdriehoeken van de vier sprekers

Fig. 12 a. weergegeven in één figuur.



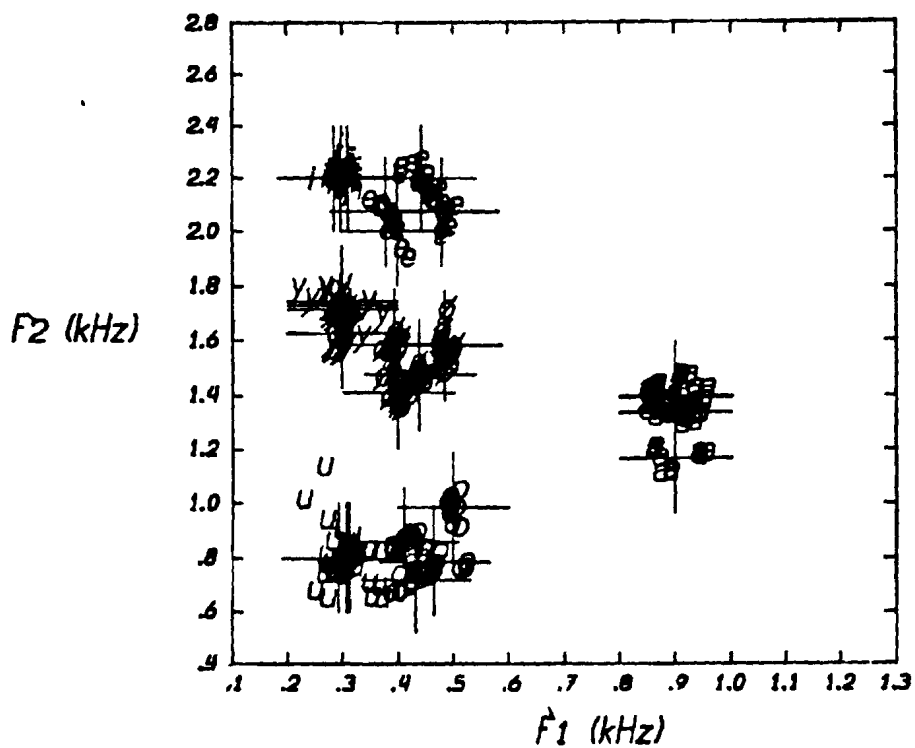
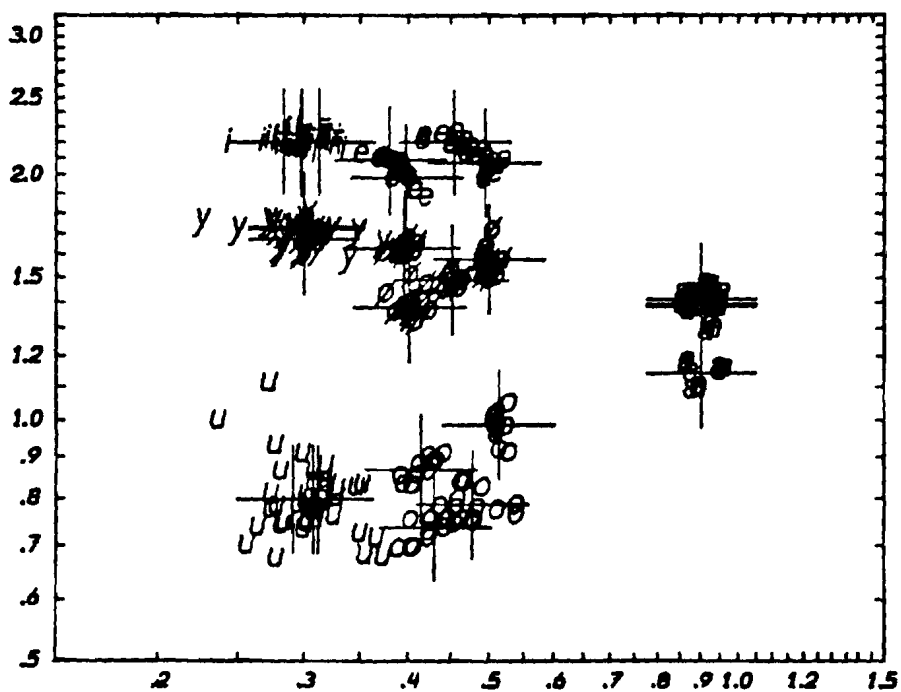


Fig. 11 b.

Klinkerdriehoeken van de vier sprekers

Fig. 12 b. na toepassing van transformatie 1 .



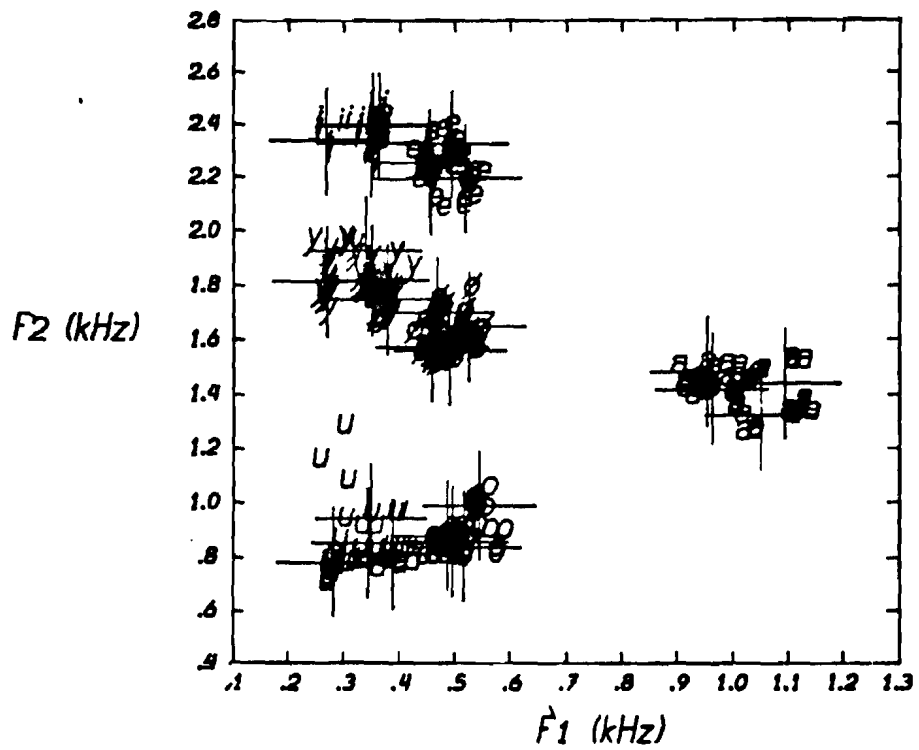
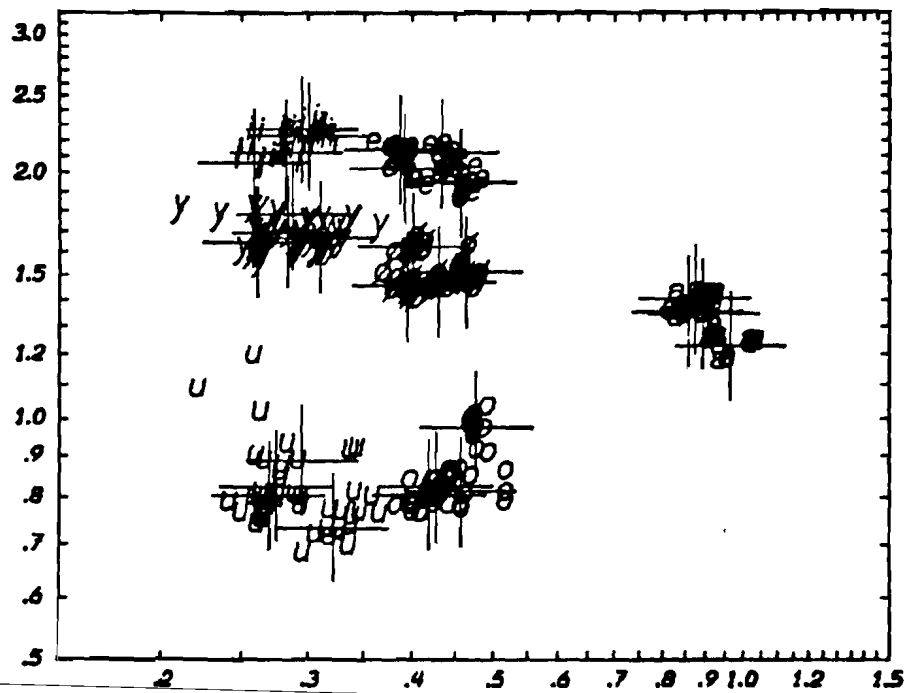


Fig. 11 c.

Klinkerdriehoeken van de vier sprekers

Fig. 12 c. na toepassing van transformatie 2 .



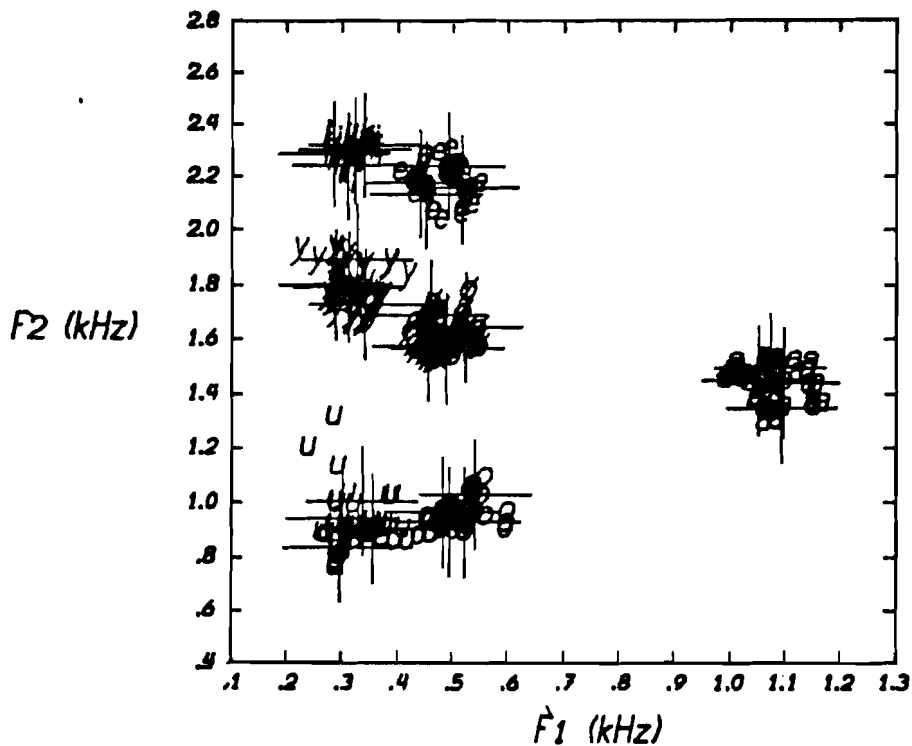
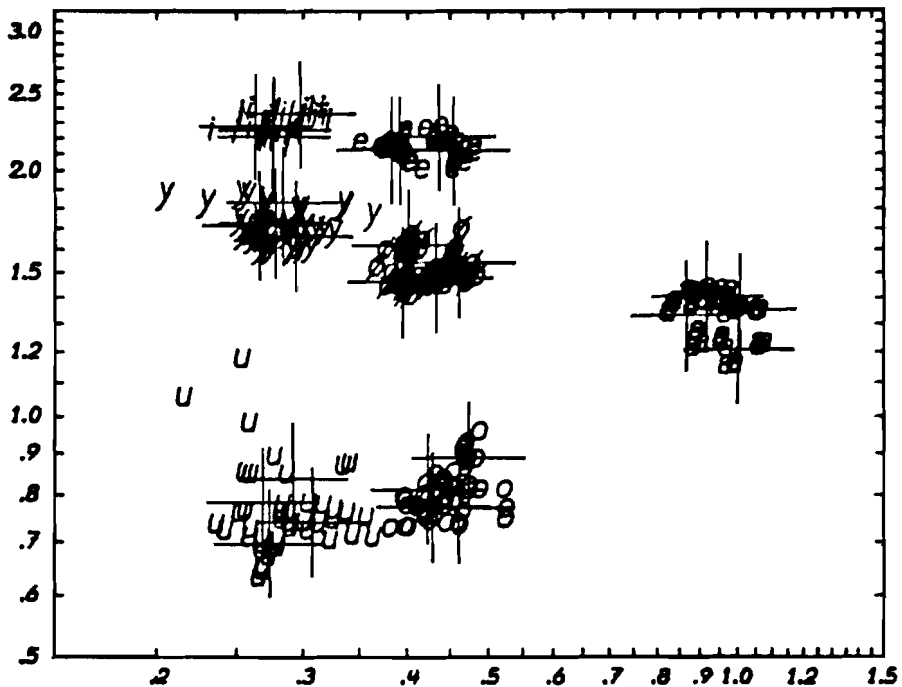


Fig. 11 d.

Klinkerdriehoeken van de vier sprekers

Fig. 12 d.

na toepassing van transformatie 3 .



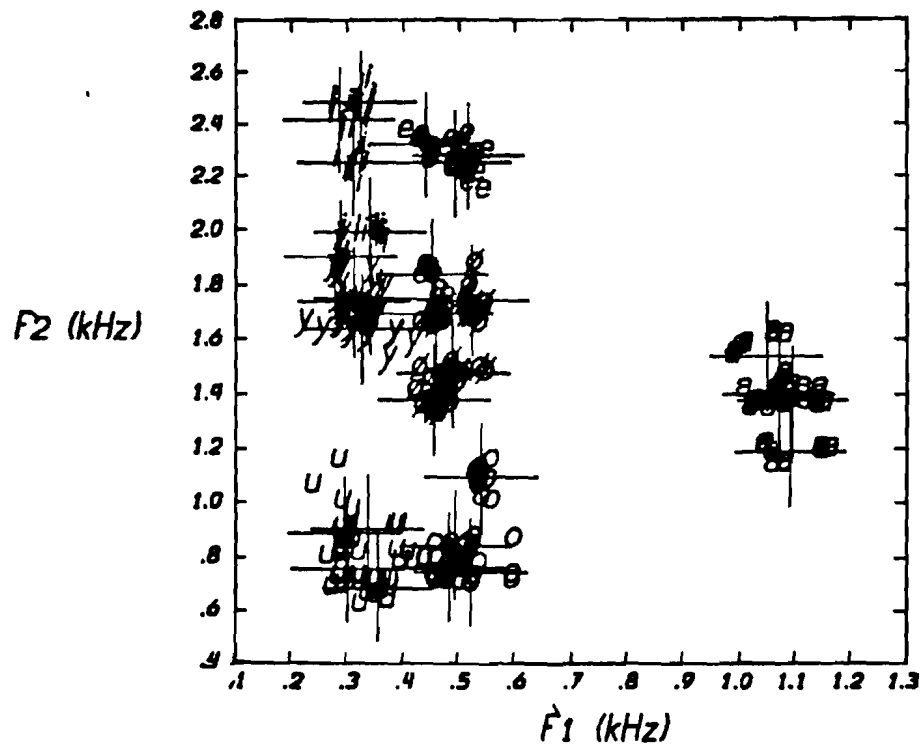
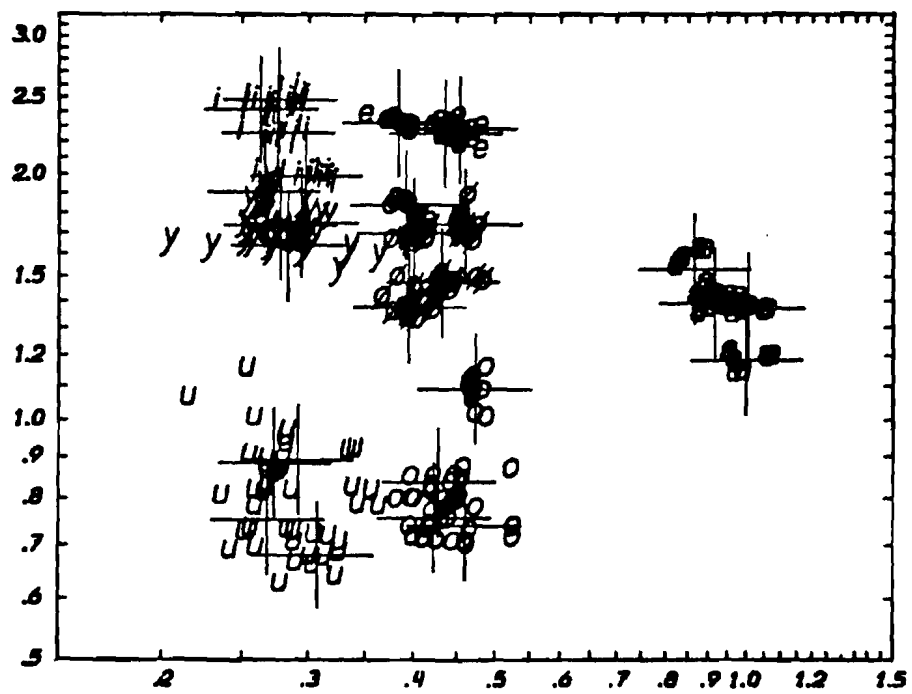


Fig. 11 e.

Klinkerdriehoeken van de vier sprekers

Fig. 12 e. na toepassing van transformatie 4 .



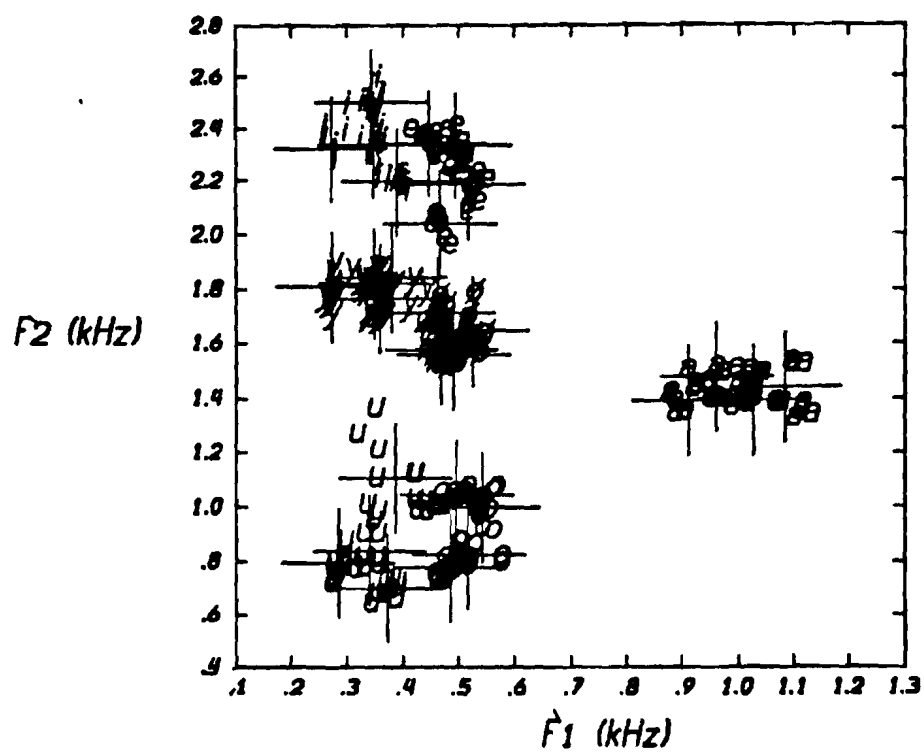
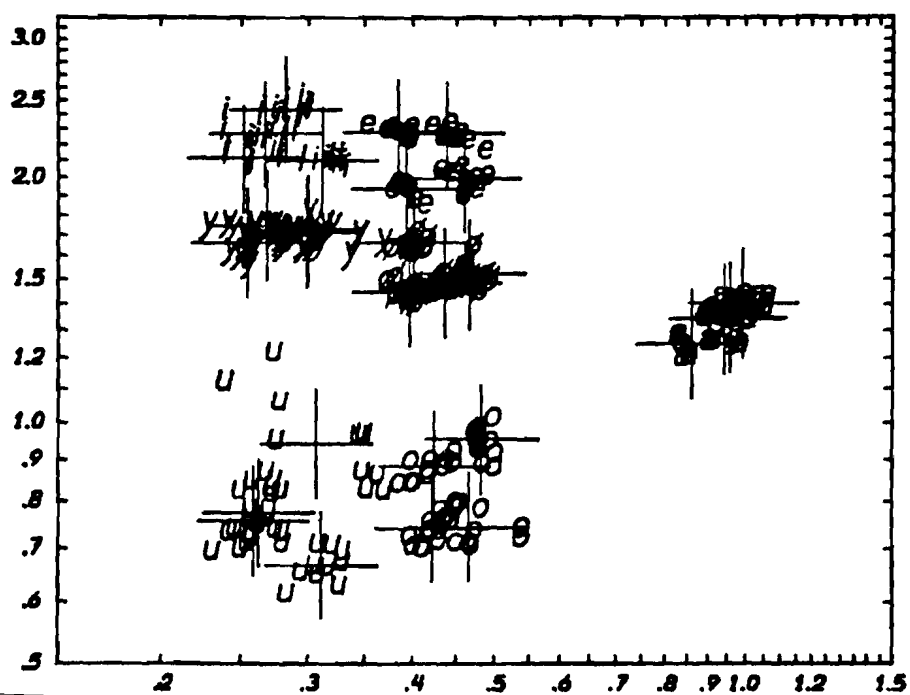


Fig. 11 f.

Klinkerdriehoeken van de vier sprekers

Fig. 12 f. na toepassing van transformatie 5 .





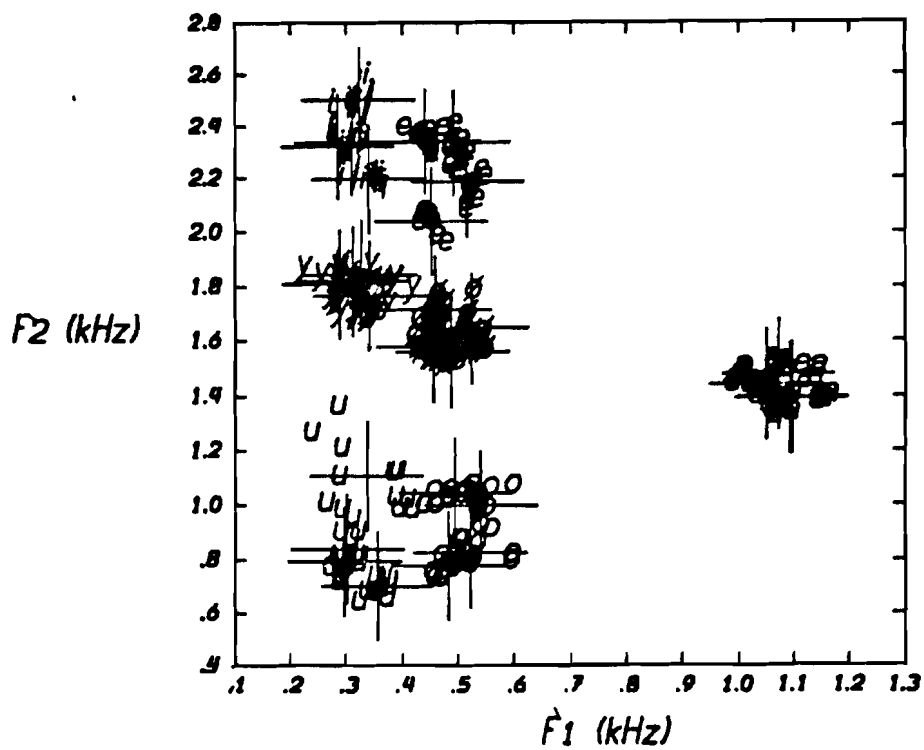
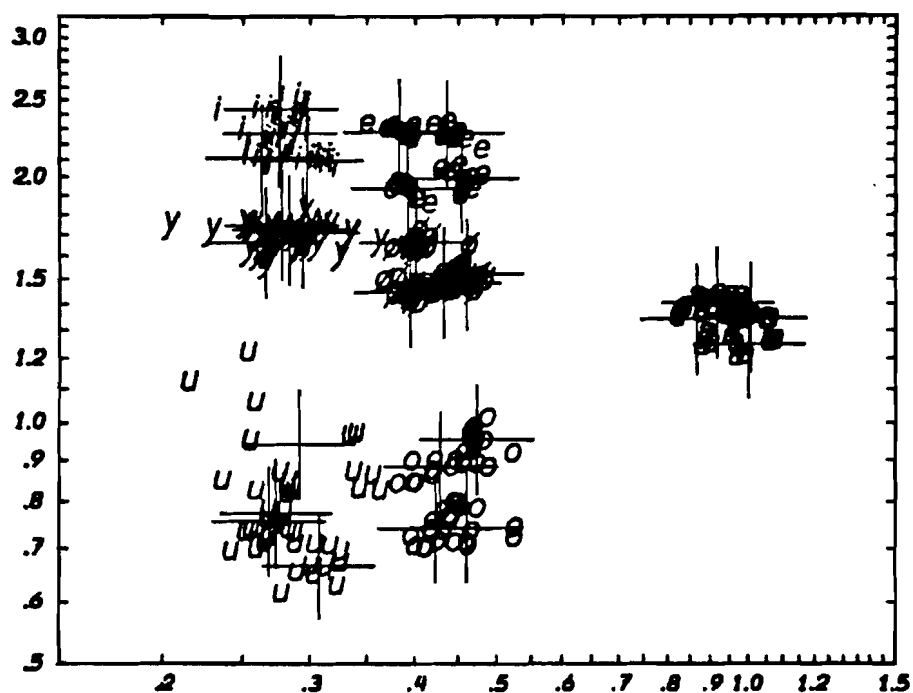


Fig. 11 g.

Klinkerdriehoeken van de vier sprekers

Fig. 12 g. na toepassing van transformatie 6 .



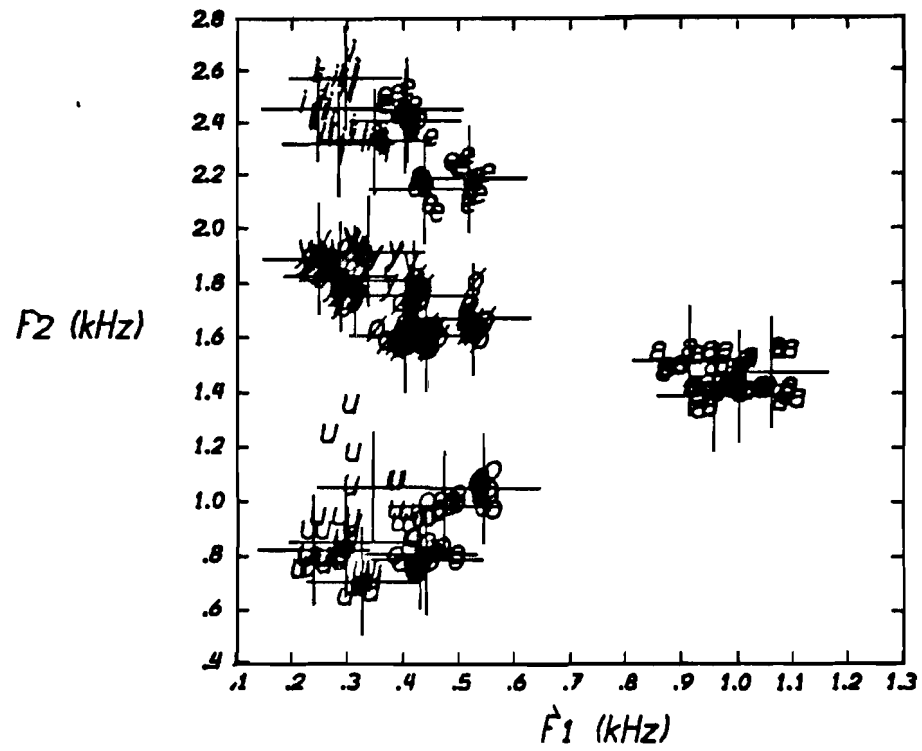


Fig. 11 h.

Klinkerdriehoeken van de vier sprekers

Fig. 12 h.  
n.v.t.

na toepassing van transformatie 7 .

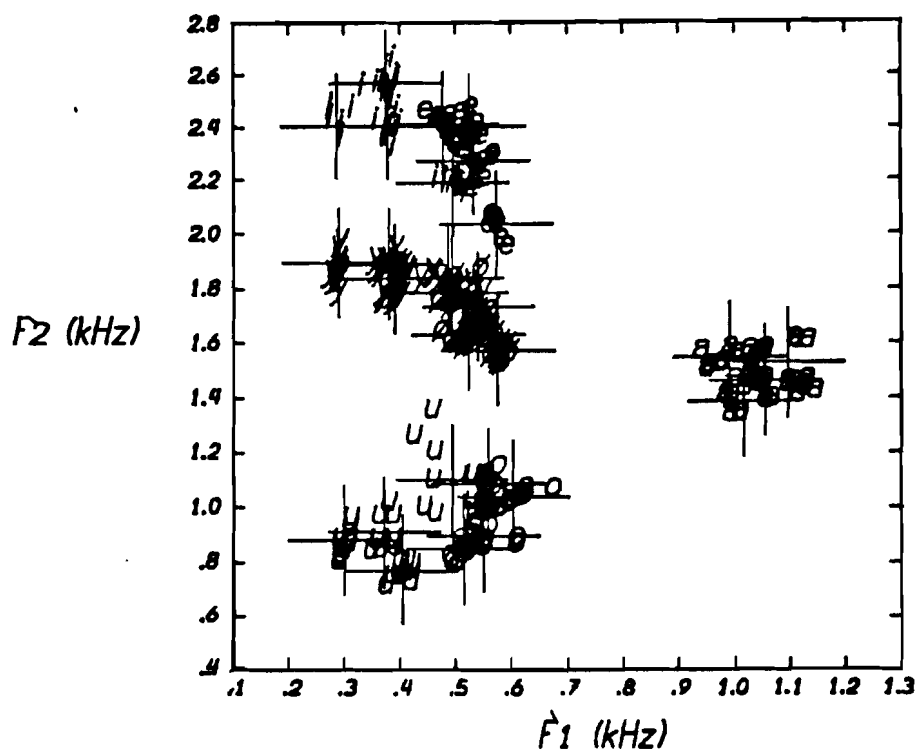
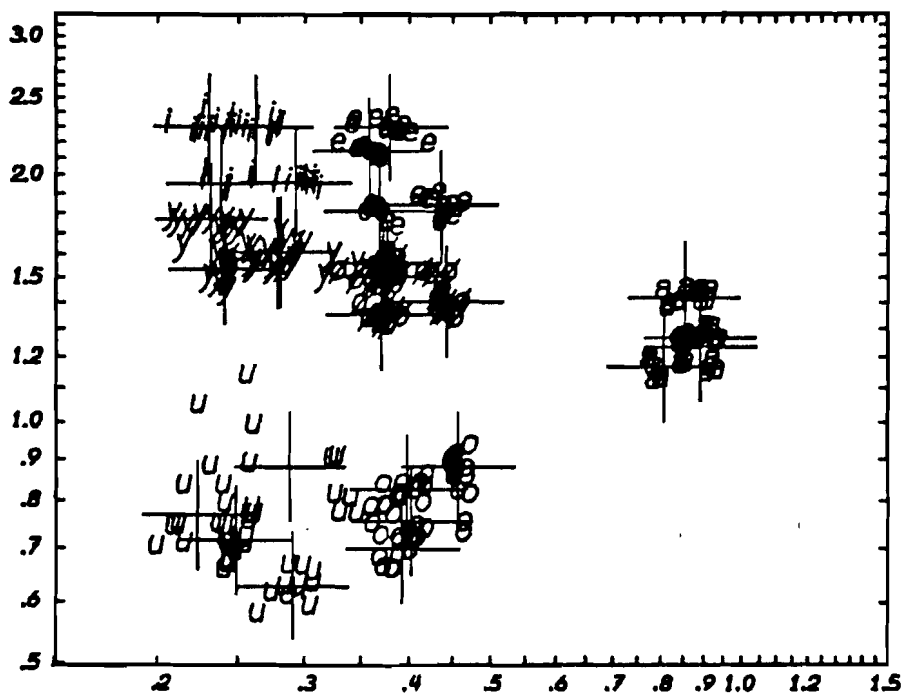


Fig. 11 i.

Klinkerdriehoeken van de vier sprekers

Fig. 12 i. na toepassing van transformatie 8 .



Een maat voor de afwijkingen tussen sprekers

Om de beoordeling van de transformaties te ondersteunen met een getal, werd een maat gedefinieerd en berekend. Deze maat geeft aan, hoe groot de spreiding van de gemiddelde klinkers voor de sprekers onderling is, in verhouding tot de grootte van de gemiddelde klinker-driehoek.

Zij  $x_{ij}$ ,  $y_{ij}$  het gemiddelde van klinker nr.  $i$  en van spreker nr.  $j$ . Dit zijn dus weer de kruispunten in de figuren.

De spreiding  $S_k$  van de gemiddelde klinkers voor de sprekers onderling is gedefinieerd door:

$$\begin{aligned} S_k^2 &= \sum_i \left[ \sum_j x_{ij}^2 - \left( \sum_j x_{ij} \right)^2 \right] + \sum_i \left[ \sum_j y_{ij}^2 - \left( \sum_j y_{ij} \right)^2 \right] \\ &= \sum_i \sum_j \left( x_{ij} - \sum_l x_{il} \right)^2 + \sum_i \sum_j \left( y_{ij} - \sum_l y_{il} \right)^2 \end{aligned}$$

Het symbool  $\sum_i$  is een afkorting voor:  $\frac{1}{n} \sum_{i=1}^n$ .

Daarbij is  $n$  het aantal doorlopen indices, dus het aantal klinkers resp. sprekers waarover gesommeerd wordt.

De grootte  $S_d$  van de gemiddelde klinkerdriehoek is gedefinieerd door:

$$\begin{aligned} S_d^2 &= \left[ \sum_i \left( \sum_j x_{ij} \right)^2 - \left( \sum_i \sum_j x_{ij} \right)^2 \right] + \left[ \sum_i \left( \sum_j y_{ij} \right)^2 - \left( \sum_i \sum_j y_{ij} \right)^2 \right] \\ &= \sum_i \left( \sum_j x_{ij} - \sum_k \sum_l x_{kl} \right)^2 + \sum_i \left( \sum_j y_{ij} - \sum_k \sum_l y_{kl} \right)^2 \end{aligned}$$

Als maat voor de afwijkingen tussen sprekers is dus gekozen de verhouding:

$$M = S_k / S_d$$

Deze maat is niet gevoelig voor verschuivingen en alzijdige rek in het  $x - y$  vlak, in de vorm:  $x' = a x + b$ ,  $y' = a y + c$ .

De maat M is berekend zowel vóór als na toepassing van de transformaties:

toegepaste transformatie	waarde van de maat M			transformatie- parameters
	lineaire assen	logarithm. assen		
geen	.278	.308		
1	.130	.168		$x_y, x_a, y_u, y_i$
2	.113	.151		$m_x, m_y, s$
3	.097	.126		$m_x, m_y, s_x, s_y$
4	.250	.209		$m_x, s_x$
5	.183	.187		$m_x, m_y$
6	.171	.168		$m_x, m_y, s_x$
7	.173			$r = (m_x^2 + m_y^2)^{\frac{1}{2}}$
8	.203	.209		$c = (m_x + m_y)/2$

Hoe kleiner M , hoe dichter de sprekers bij elkaar liggen.

Transformatie 3 komt dus ook hier als beste uit de bus, gevolgd door transformatie 2 .

### Discussie

In het voorgaande is transformatie 3 als beste te voorschijn gekomen. Bij deze transformatie worden F1 en F2 onafhankelijk van elkaar getransformeerd. Als transformatie-parameters worden benodigd: gemiddelde en spreiding van F1 en van F2 van één spreker.

Men zou de transformatie kunnen toepassen in een automatisch herkenningssysteem, dat de klinkers op basis van F1 en F2 onderscheidt, en daarmee een aanpassing aan verschillende sprekers kunnen bereiken. Hiervoor zal dus informatie over de parameters van een bepaalde spreker verkregen moeten worden. Het systeem zou dit bij voorkeur zelf moeten doen.

Dit zou als volgt kunnen. In het begin van een sessie kan men een spreker een voorgeschreven standaardzin laten uitspreken, waarin bijvoorbeeld alle klinkers één keer voorkomen. Het systeem kan dan een schatting maken.

In de loop van een sessie kan men het gemiddelde en de spreiding over een langere periode bepalen, en een nauwkeuriger schatting maken. Hiervoor zou wél enige kennis nodig zijn over de frequentie van klinkers in gewone spraak.

Literatuurlijst

M. Abramowitz, I.A. Stegun  
Handbook of mathematical functions  
Dover Publications, Inc., New York (1972)

S. Barth, R. Chulliat  
Modifications inter-locuteurs de l'échelle formantique  
9 èmes Journées d'étude sur la parole, Lannion, p.149-156 (1978)

F.J. Harris  
On the use of windows for harmonic analysis with the discrete  
Fourier transform  
Proceedings of the IEEE, vol. 66, no. 1, p.51-83 (1978)

D.H. Klatt  
Review of the ARPA speech understanding project  
J. Acoust. Soc. Am., vol. 62, no.6, p. 1345-1366 (1977)

F.J. Koopmans-van Beinum  
Vowel contrast reduction - an acoustic and perceptual study of  
dutch vowels in various speech conditions  
Academische Pers B.V., Amsterdam (1980)

J. Makhoul  
Linear prediction: a tutorial review  
Proceeding of the IEEE, vol. 63, no. 4, p. 561-580 (1975)

J.D. Markel, A.H. Gray, Jr.  
Linear prediction of speech  
Springer-Verlag, Berlin Heidelberg New York (1976)

H. Matsumoto, H. Wakita  
Frequency warping for nonuniform talker normalization  
1979 IEEE International Conference on Acoustics, Speech &  
Signal Processing, p. 566-569 (1979)

D.J.P.J. van Nierop, L.C.W. Pols, R. Plomp  
Frequency analysis of dutch vowels from 25 female speakers  
Acustica, vol. 29, p. 110-118 (1973)

S.G. Nooteboom, A. Cohen  
Spreken en verstaan - een inleiding tot de experimentele fonetiek  
Van Gorcum, Assen/Amsterdam (1976)

A.V. Oppenheim, R.W. Schaffer  
Digital signal processing  
Prentice-Hall, Inc., New Jersey (1975)

G.E. Peterson, H.L. Barney  
Control methods used in a study of the vowels  
J. Acoust. Soc. Am., vol. 24, no. 2, p. 175-184 (1952)

R. Plomp, L.C.W. Pols, J.P. van de Geer  
Dimensional analysis of vowel spectra  
J. Acoust. Soc. Am., vol. 41, no. 3, p. 707-712 (1967)

L.C.W. Pols, H.R.C. Tromp, R. Plomp  
Frequency analysis of dutch vowels from 50 male speakers  
J. Acoust. Soc. Am., vol. 53, no. 4, p. 1093-1101 (1973)

Y. Tohkura, F. Itakura, S. Hashimoto  
Spectral smoothing technique in PARCOR speech analysis-synthesis  
IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, no. 6,  
p. 587-596 (1978)

H. Wakita  
Normalization of vowels by vocal-tract length and its application  
to vowel identification  
IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25, no. 2,  
p. 183-192 (1977)



## Appendix

### Het Newton-venster

Bij discrete orthogonale polynomen komt een discrete gewichtsfunctie voor. Bij bestudering van deze materie ontstond het idee, dat sommige van die gewichtsfuncties ook bruikbaar kunnen zijn als tijdvenster. Bij de Krawtchouk-functies wordt een binomiale verdeling gebruikt als gewichtsfunctie (zie bijv. Abramowitz & Stegun, par. 22.17). Als venster wordt dit het Newton (binomiaal) venster

$$w_n = \binom{N}{n} = \frac{N!}{n! (N-n)!}, \quad n \in [0, N], \quad N \text{ vast}$$

Dit venster heeft een aantal interessante eigenschappen:

- Het is een echt discreet venster, terwijl vele andere vensters gedefinieerd zijn als product van een continue functie en een discreet rechthoekig venster.
- De getallen  $w_n$  kunnen gemakkelijk berekend worden met behulp van de recursieve betrekking:

$$\frac{w_{n+1}}{w_n} = \frac{N-n}{n+1}$$

Dit kost één vermenigvuldiging en deling per stap, en is dus qua rekentijd zeer goedkoop.

- Buiten het interval  $[0, N]$  zijn de getallen  $w_n$  "vanzelf" nul, ten gevolge van de recursieve betrekking.
- Het spectrum (periodogram) kan met behulp van het binomium van Newton verrassend eenvoudig berekend worden:

$$\begin{aligned} W(\omega) &= \sum_n w_n e^{-j\omega n} = \sum_{n=0}^N \binom{N}{n} e^{-j\omega n} = (1 + e^{-j\omega})^N = \\ &= \left( e^{j\frac{\omega}{2}} + e^{-j\frac{\omega}{2}} \right)^N e^{-j\frac{\omega}{2}N} = \left( 2 \cos\left(\frac{\omega}{2}\right) \right)^N e^{-j\frac{\omega}{2}N} \end{aligned}$$

Hieruit volgt direct, dat de modulus van het spectrum géén zijlobben heeft, dit in tegenstelling tot vele andere vensters (cf. Harris, 1978). Er is slechts één centrale hoofdlob, en bij toenemende  $N$  wordt deze steeds scherper.

- Het Newton-venster kan beschouwd worden als het discrete analogon van de continue Gauss-functie. Dit blijkt in de theorie van orthogonale polynomen, bij vergelijking van continue en discrete gevallen.

- Het Newton-venster kan ook beschouwd worden als binomiale verdeling, en uit de statistiek is bekend, dat voor grote  $N$  een Gauss-functie benaderd wordt.

- De autocorrelaties van een Newton-venster vormen weer een (verschoven) Newton-venster:

$$w'_k = \sum_n w_n w_{n-k} = \sum_n \binom{N}{n} \binom{N}{n-k} = \binom{2N}{k+N}$$

In deze vorm is het venster bruikbaar als venster voor de autocorrelaties van een signaal, en als zodanig reeds toegepast door Tohkura, Itakura & Hashimoto (1978).

Figuur 13 a toont tijdpatroon en log-spectrum van een Newton-venster met  $N = 50$ . Aan de uiteinden van het tijdpatroon worden de functiewaarden al gauw zeer klein. Men kan natuurlijk de uiteinden afkappen, maar dan verschijnen er wel zijlobben in het spectrum (figuur 13 b). De mate van kappen beïnvloedt de hoogte ervan, en men kan dit bijv. zó instellen, dat het venster op een Hamming-window gaat lijken (figuur 13 c - e).

Met behulp van het Newton-venster kunnen ook andere vensters gedefinieerd worden, bijvoorbeeld een quotient van twee vensters:

$$w_n = \binom{N}{n} / \binom{N+2D}{n+D} \quad \text{met } D > 0$$

Deze vensters (figuur 14 a - c) lijken op bekende  $\cos^a$ -vensters (cf. Harris, 1978).

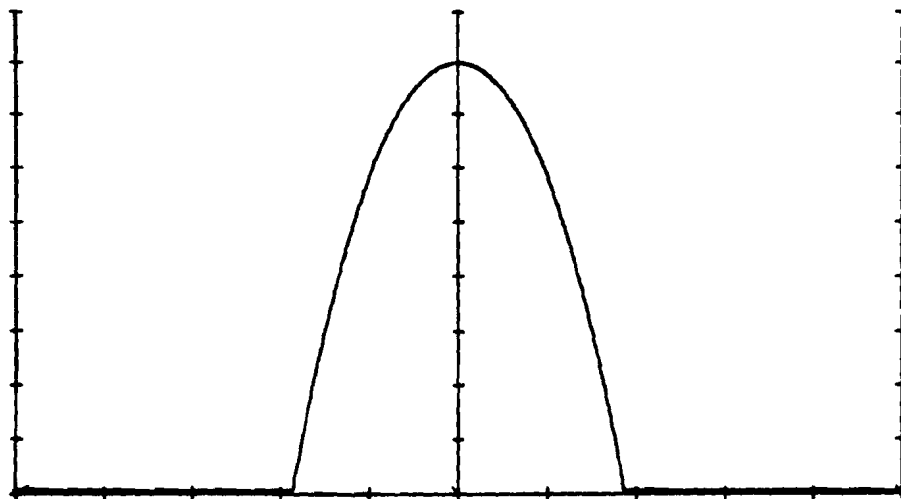
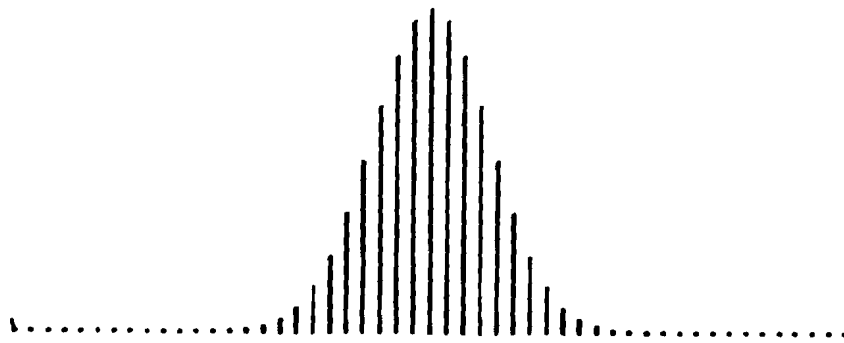


Fig. 13 a. Tijdpatroon en log-spectrum van een Newton (binomiaal) venster met  $N = 50$  (51 punten). Er zijn géén zijlobben in het spectrum. Assen boven: lineair. Assen onder: horizontaal lineair, interval  $(-\pi, \pi)$ ; verticaal logaritmisch, 10 dB per schaaldeel.

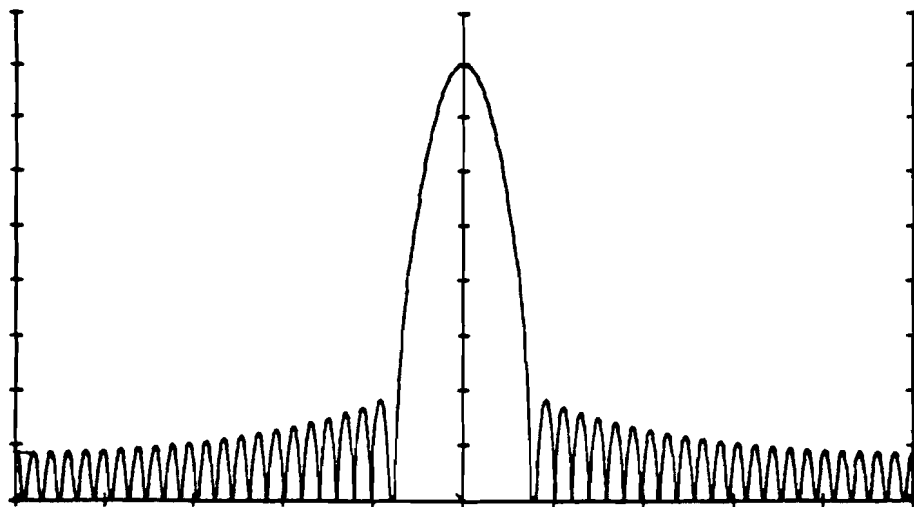
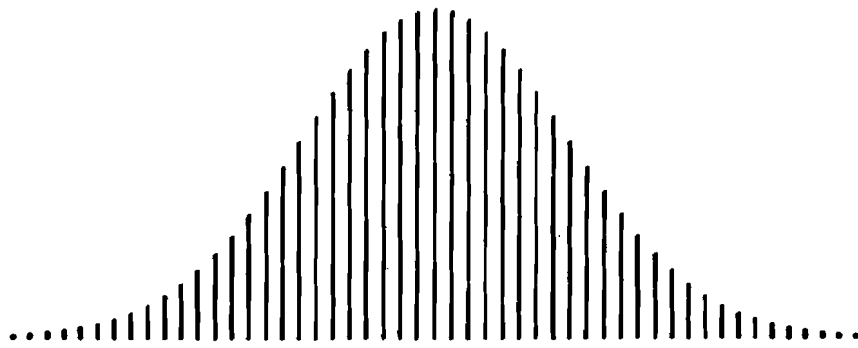


Fig. 13 b. Tijdpatroon en log-spectrum van een afgekapt Newtonvenster met  $N = 50 + 2 \times 100$  (51 punten) . Aan ieder uiteinde zijn 100 punten weggelaten.

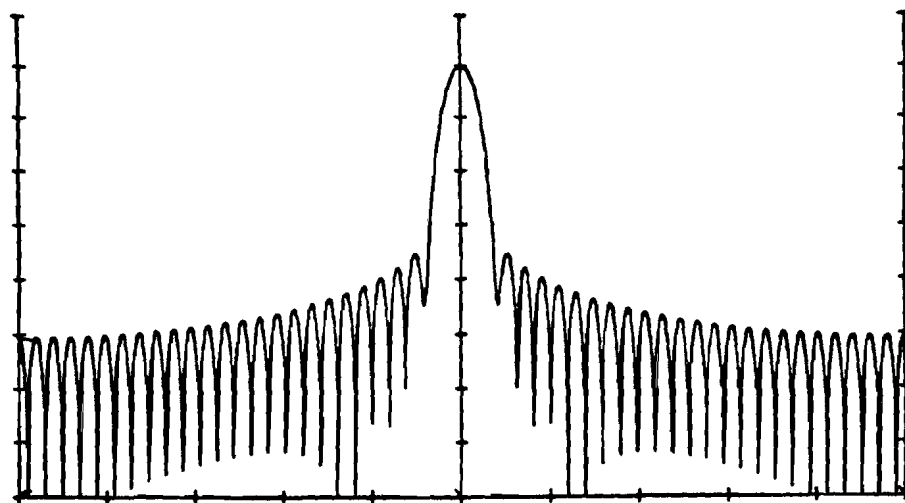
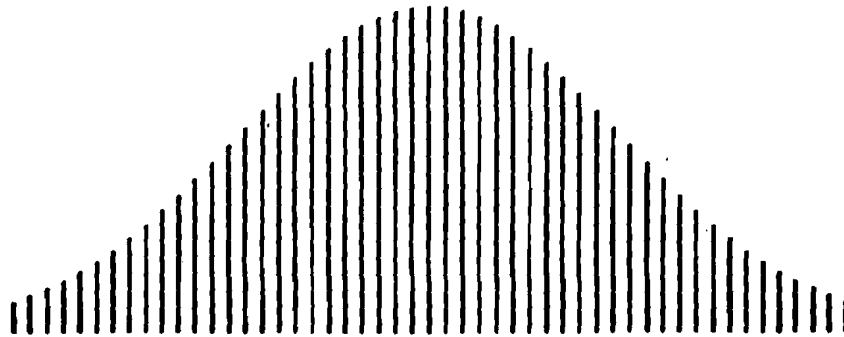


Fig. 13 c. Tijdpatroon en log-spectrum van een afgekapt Newtonvenster met  $N = 50 + 2 \times 235$  (51 punten) . Aan ieder uiteinde zijn 235 punten weggelaten.

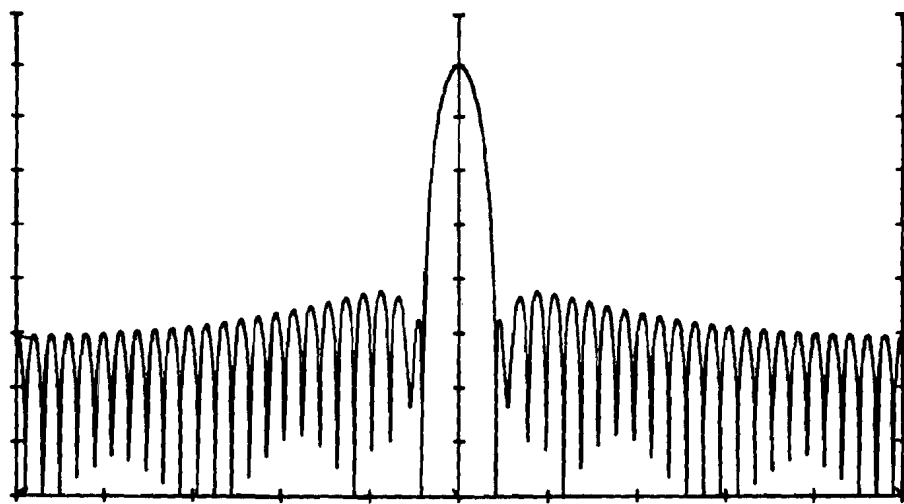
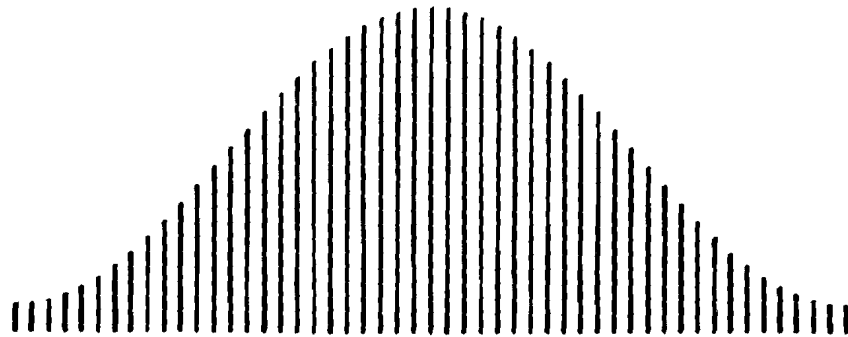


Fig. 13 d. Hamming-window met  $N = 50$  (51 punten) , ter vergelijking.

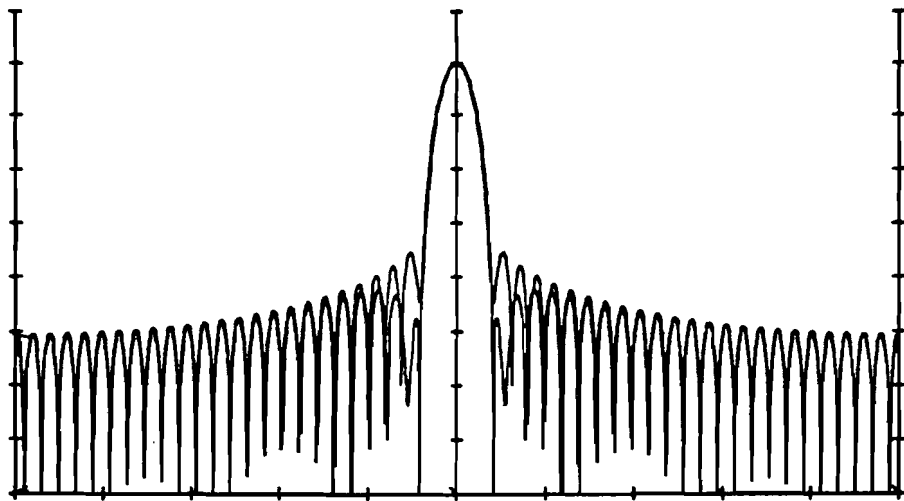


Fig. 13 e. Spectra van Hamming- en afgekapt Newton-venster  
(als in fig. 13 d resp. c) .

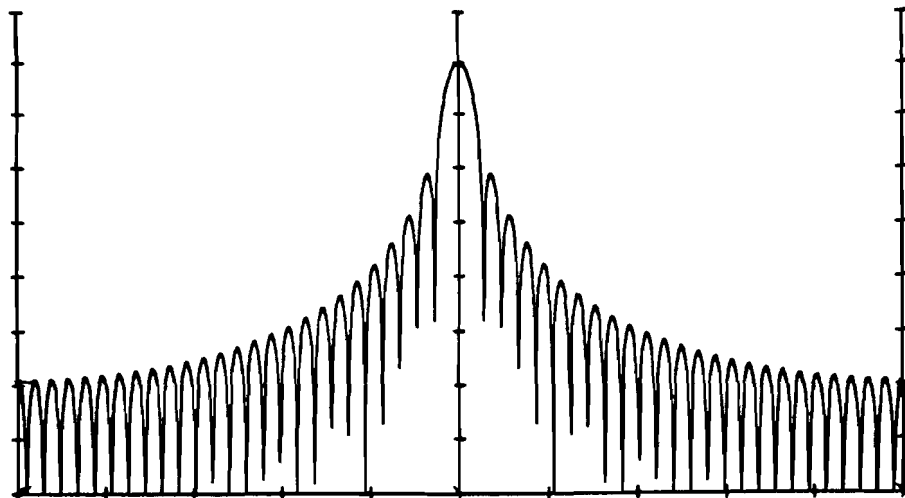
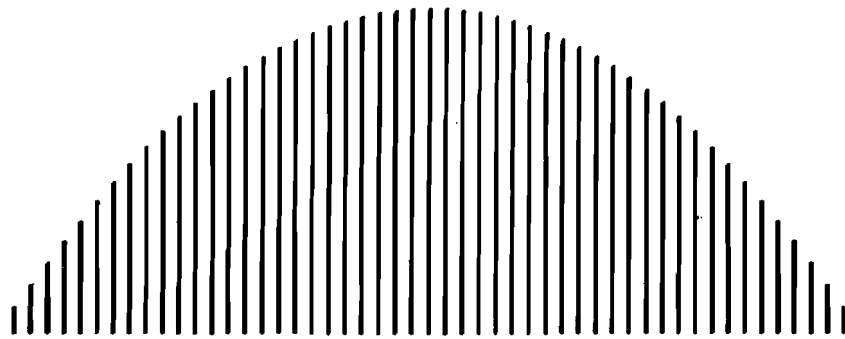


Fig. 14 a. Quotient van twee Newton-vensters (zie tekst) met  $N = 50$  en  $D = 1$  (51 punten) .



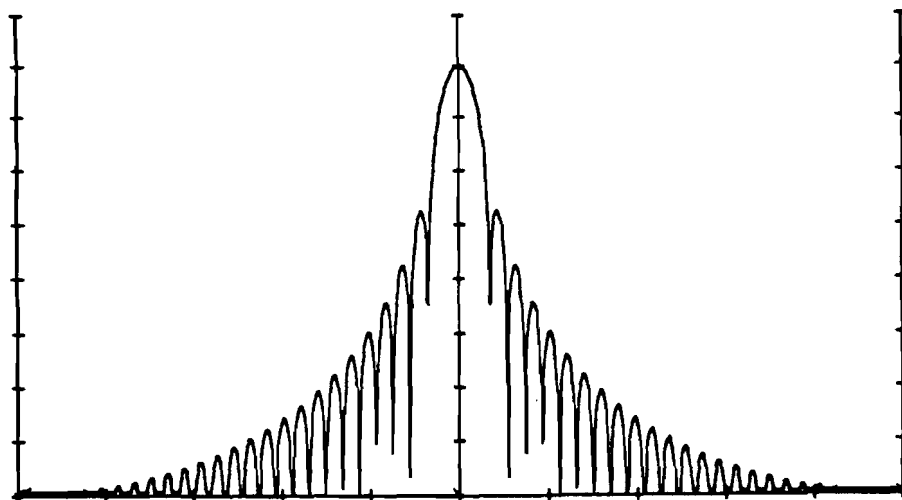
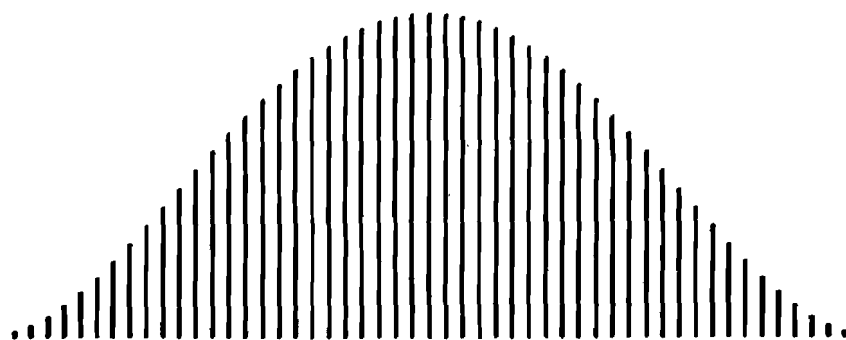


Fig. 14 b. Quotient van twee Newton-vensters (zie tekst) met  $N = 50$  en  $D = 2$  (51 punten) .

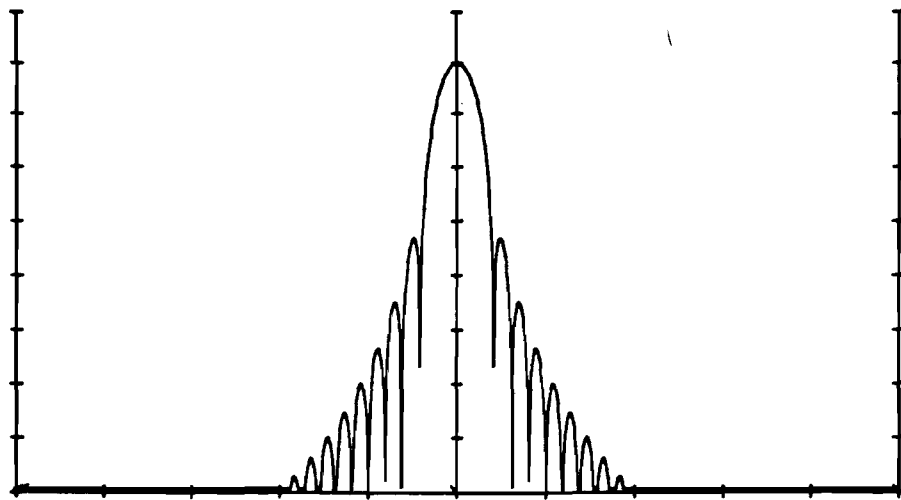
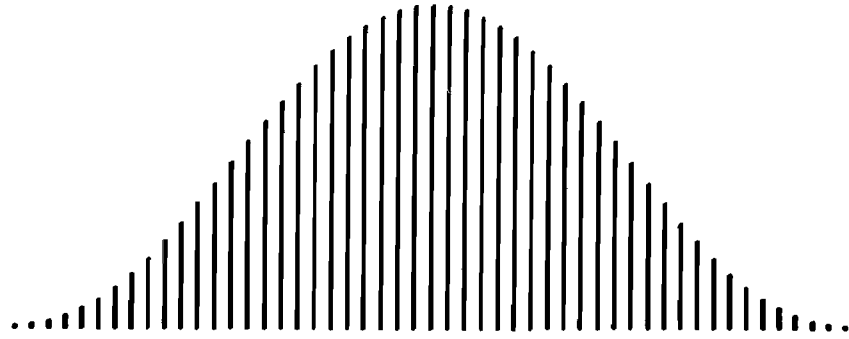


Fig. 14 c. Quotient van twee Newton-vensters (zie tekst) met  $N = 50$  en  $D = 3$  (51 punten) .

## Beschouwingen over LPC-analyse

### Autocorrelaties en spectrum van het restsignaal

In de LPC-analyse is het restsignaal (residual)  $e_n$  gedefinieerd door (cf. Makhoul, formule 7)

$$e_n = \sum_{k=0}^p a_k s_{n-k} \quad \text{met } a_0 = 1 \quad (1)$$

Beschouw de autocorrelaties van het restsignaal:

$$E_k = \sum_n e_n e_{n-k} = \sum_{m=0}^p \sum_{n=0}^p a_m a_n R_{m-n+k} \quad (2)$$

waarin

$$R_k = \sum_n s_n s_{n-k}$$

de autocorrelaties van het signaal  $s_n$  zijn.

Bij berekeningen is het handig, om te definiëren:  $a_k = 0$  voor  $k \notin [0, p]$

Na enig manipuleren met de indices volgt uit (2)

$$E_k = \sum_{l=-p}^p r_l R_{k-l} \quad (3)$$

waarin

$$r_l = \sum_k a_k a_{k-l}$$

de autocorrelaties van het (inverse) filter  $a_k$ . De getallen  $r_l$  kunnen ook beschouwd worden als de coëfficiënten van een filter voor de autocorrelaties, immers (3) heeft dezelfde vorm als (1) en is een convolutie.

Beschouw vervolgens het spectrum van het restsignaal:

$$E(\omega) = \sum_n e_n e^{-j\omega n}$$

( $E(\omega)$  te onderscheiden van  $E_k$  !) en het bijbehorende energiespectrum:

$$Q(\omega) = E(\omega) \bar{E}(\omega) = |E(\omega)|^2$$

Het energiespectrum wordt ook verkregen, door de autocorrelaties te transformeren:

$$Q(\omega) = \sum_k E_k e^{-j\omega k} \quad (4)$$

en omdat  $E_{-k} = E_k$  :

$$Q(\omega) = E_0 + 2 \sum_{k=1}^{\infty} E_k \cos(\omega k) \quad (5)$$

De term  $E_0$  vertegenwoordigt een constant, vlak, "wit" energiespectrum, en de overige termen vormen een fluctuatieromheen.

#### Een paradox bij LPC-analyse

Per definitie is  $E_0 = \sum_n e_n^2$ , en dit is de "total squared error"  $E$

(cf. Makhoul, formule 8), dus  $E_0 = E$ .

In de LPC-analyse wordt  $E$ , dus  $E_0$ , geminimaliseerd door de filtercoëfficiënten  $a_k$  te variëren, bij een gegeven signaal  $s_n$ .

Dit betekent volgens het voorgaande, dat de vlakke, witte component van het energiespectrum zo klein mogelijk wordt gemaakt.

Tegelijk wordt gesteld, dat het spectrum zo vlak mogelijk wordt, en dat het optimale filter  $a_k$  een "whitening filter" is (cf. Makhoul, na formule 61; Markel & Gray, par. 6.3).

Dan zou men echter verwachten, dat de vlakke component van het spectrum juist zo groot mogelijk gemaakt zou moeten worden, althans in verhouding tot de fluctuaties, dus  $E_0$  zo groot mogelijk in verhouding tot de overige  $E_k$ 's.

Dit is een paradox.

Spiegeling van polen en nulpunten aan de eenheidskring

De overdrachtsfunctie van het inverse filter  $a_k$  is

$$A(\omega) = \sum_{k=0}^p a_k e^{-j\omega k} \quad \text{met } a_0 = 1$$

$$= \sum_{k=0}^p a_k z^k \quad \text{met } z = e^{-j\omega}$$

Zij  $z_0$  een nulpunt van het polynoom. Omdat  $a_0 \neq 0$  is  $z_0 \neq 0$ .

Dan is  $z_0$  te schrijven als:  $z_0 = r e^{j\varphi}$ . Spiegeling van  $z_0$  aan de eenheidskring geeft het punt:  $z'_0 = \bar{z}_0^{-1} = \frac{1}{r} e^{j\varphi}$ .

Het polynoom kan geschreven worden als een product, waarin de factor  $(z - z_0)$  voorkomt. Vervang  $(z - z_0)$  door:  $|z_0|^2 (z - z'_0)$ .

Dan ontstaat een ander filter. Echter,  $a_0$  blijft hetzelfde, omdat

$|z_0|^2 z'_0 = z_0 \bar{z}_0 \bar{z}_0^{-1} = z_0$ . Belangrijker is, dat de frequentie-karakteristiek van het filter,  $|A(\omega)|$ , niet van vorm verandert en slechts met een factor  $|z_0|$  wordt vermenigvuldigd.

Met  $z = e^{-j\omega}$  geldt namelijk:  $|z - z_0| = |z_0| |z - z'_0|$ .

Dit laatste is gemakkelijk te bewijzen, door het kwadraat van de modulus uit te schrijven.

Wanneer  $z_0$  niet op de reële as ligt, moet tegelijk met  $z_0$  ook het toegevoegde nulpunt  $\bar{z}_0$  gespiegeld worden, om weer een filter met reële coëfficiënten te verkrijgen.

De hierboven beschreven spiegeling kan voor ieder reëel nulpunt en voor ieder toegevoegd nulpuntenpaar afzonderlijk uitgevoerd worden. Aldus worden  $2^q$  verschillende filters verkregen, met

$$q = (\text{aantal reële nulpunten}) + (\text{aantal toegevoegde paren})$$

en nulpunten op de eenheidskring niet meegeteld.

Deze filters hebben dus alle, afgezien van een evenredigheidsfactor, dezelfde frequentie-karakteristiek  $|A(\omega)|$ .

Voor het spectrum van het restsignaal geldt op grond van (1) :

$$E(\omega) = A(\omega) S(\omega)$$

en bijgevolg voor het energiespectrum van het restsignaal:

$$Q(\omega) = |E(\omega)|^2 = |A(\omega)|^2 |S(\omega)|^2$$

Stel, dat  $a_k$ , dus  $A(\omega)$ , het optimale "whitening" filter is, verkregen met behulp van LPC-analyse. Bij spiegeling van de nulpunten van  $a_k$ , volgens het voorgaande, wordt  $|A(\omega)|$  vermenigvuldigd met een constante factor, en derhalve  $Q(\omega)$  eveneens.

De vorm van  $Q(\omega)$  blijft behouden.

Er is dus niet slechts één "whitening" filter, maar een hele familie.

#### Een andere normalisering van het filter

Beschouw alle filters  $a_k$  zodanig dat:

$$e_n = \sum_{k=0}^p a_k s_{n-k}$$

dus zonder de voorwaarde  $a_0 = 1$ .

De "total squared error"  $E$  is gedefinieerd door:  $E = \sum_n e_n^2$ .

Twee filters  $a_k$  en  $c a_k$  ( $c \neq 0$ ) mogen als gelijkwaardig beschouwd worden, immers  $c$  is slechts een versterkingsfactor. Hierdoor wordt een equivalentie-relatie voor filters gedefinieerd. Helaas zijn  $e_n$  en  $E$  niet invariant bij de overgang  $a_k \rightarrow c a_k$ . Door de filters  $a_k$  te normaliseren kan wel invariantie bereikt worden.

Een mogelijkheid hiertoe is:

$$e_n = \sum_{k=0}^p \frac{a_k}{a_0} s_{n-k}$$

Nu is  $e_n$ , en dus ook  $E$ , invariant bij de overgang  $a_k \rightarrow c a_k$ .

Het filter met coëfficiënten  $\frac{a_k}{a_0}$  heeft de vorm, die bij LPC-analyse gebruikelijk is, immers voor  $k = 0$  is  $\frac{a_k}{a_0} = 1$ .

Een andere mogelijkheid is:

$$e_n = \sum_{k=0}^p \frac{a_k}{\sqrt{\sum_{l=0}^p a_l^2}} s_{n-k}$$

Hiervoor is:

$$E = \sum_n e_n^2 = \frac{\sum_{m=0}^p \sum_{n=0}^p a_m a_n R_{m-n}}{\sum_{l=0}^p a_l^2}$$

Bij variatie der  $a_i$  is:

$$\frac{\partial E}{\partial a_i} = 2 \frac{\sum_{k=0}^p a_k R_{k-i} - E a_i}{\sum_{l=0}^p a_l^2}$$

Het minimaliseren van  $E$ , zoals gebruikelijk bij LPC-analyse, leidt nu tot een eigenwaarde-probleem.

Blijft natuurlijk de vraag, of deze andere aanpak betere resultaten oplevert bij de analyse van signalen.

### Tijdsomkering

Wanneer het signaal  $s_n$  verschoven wordt in de tijd, blijven de autocorrelaties gelijk:

$$\begin{aligned}
 s'_n &= s_{n-d} \\
 R'_k &= \sum_n s'_n s'_{n-k} = \sum_n s_{n-d} s_{n-d-k} = \\
 &= \sum_n s_n s_{n-k} = R_k
 \end{aligned}$$

Maar ook wanneer het signaal  $s_n$  wordt omgekeerd in de tijd, veranderen de autocorrelaties niet:

$$\begin{aligned}
 s'_n &= s_{d-n} \\
 R'_k &= \sum_n s'_n s'_{n-k} = \sum_n s_{d-n} s_{d-n+k} = \\
 &= \sum_n s_n s_{n-k} = R_k
 \end{aligned}$$

Bij de analyse van een signaal kan dus op basis van de autocorrelaties geen onderscheid gemaakt worden tussen het signaal zelf en het in de tijd omgekeerde signaal. Er kan niet worden onderscheiden tussen een toenemend en een afnemend signaal.

Het LPC-filter wordt geheel bepaald door de autocorrelaties. Het LPC-filter blijft dus ook hetzelfde bij tijdsomkering.

De "predictor" is tegelijk een "postdictor".