

Waargenomen continuïteit in spraak : het belang van toonhoogte

Citation for published version (APA):

Brokx, J. P. L. (1979). *Waargenomen continuïteit in spraak : het belang van toonhoogte*. [Dissertatie 1 (Onderzoek TU/e / Promotie TU/e), Institute for Perception Research, Eindhoven]. Technische Hogeschool Eindhoven. <https://doi.org/10.6100/IR171313>

DOI:

[10.6100/IR171313](https://doi.org/10.6100/IR171313)

Document status and date:

Gepubliceerd: 01/01/1979

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

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
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

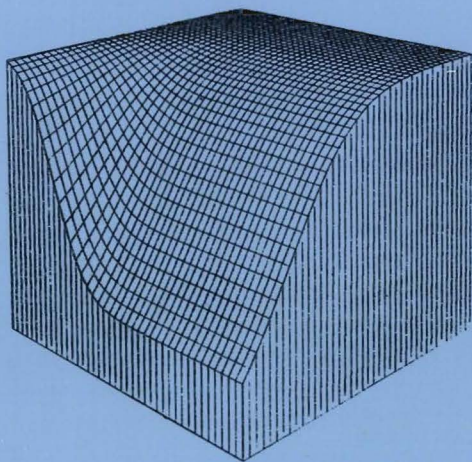
Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

WAARGENOMEN CONTINUITEIT IN SPRAAK:
HET BELANG VAN TOONHOOGTE

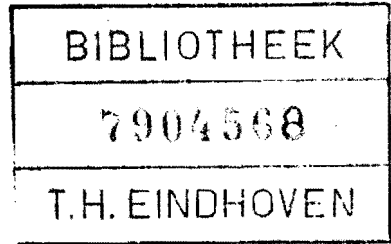


J.P.L. BROKX

**WAARGENOMEN CONTINUÏTEIT IN SPRAAK:
HET BELANG VAN TOONHOOGTE**

Perceived continuity in speech: The role of pitch

(with summary in English)



PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE
TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE
HOOGESCHOOL EINDHOVEN, OP GEZAG VAN DE RECTOR
MAGNIFICUS, PROF. DR. P. VAN DER LEEDEN, VOOR
EEN COMMISSIE AANGEWEEZEN DOOR HET COLLEGE
VAN DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP
DINSDAG 15 MEI 1979 TE 16.00 UUR

DOOR

JOHANNUS PETRUS LEONARDUS BROKX

GEBOREN TE ZUTPHEN

DIT PROEFSCHRIFT IS GOEDGEKEURD

DOOR DE PROMOTOREN:

PROF. DR. A. COHEN

PROF. DR. H. BOUMA

Dit onderzoek is tot stand gekomen onder de dagelijkse leiding van Dr. S.G. Nooteboom.

Het is uitgevoerd op het Instituut voor Perceptie Onderzoek te Eindhoven met financiële ondersteuning van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (Z.W.O.).

Mijn dank gaat uit naar iedereen die betrokken is geweest bij het tot stand komen van dit werk. In het bijzonder wil ik Prof.Dr. A. Cohen noemen voor zijn steun gedurende praktisch geheel het onderzoek. Verder wil ik met nadruk Dr. S.G. Nooteboom noemen, die in het dagelijks contact steeds bereid is geweest tot discussiëren en tot het geven van aanwijzingen om het onderzoek in goede banen te leiden. Zijn constructieve bijdragen aan dit onderzoek zijn onmisbaar geweest. Veel hulp heb ik gehad van Theo de Jong, Han Timmers en vooral ook van mijn collega Ko de Rooij. Tot slot wil ik de heer Basten bedanken voor de tekeningen en Jeanneke van Esch voor het razendsnelle typewerk.

INHOUD

1.	AUDITIEVE SAMENHANG EN SPRAAKVERSTAAN	7
1.1	Inleiding	7
1.2	Auditieve samenhang van betekenisloze reeksen geluiden	9
1.2.1	Waarneming van volgordes in reeksen geluiden	10
1.2.2	Auditieve splitsing	12
1.2.3	Auditieve splitsing en de waarneming van volgordes	14
1.3	Verstaanbaarheid van verbonden spraak	16
1.3.1	Factoren die een rol spelen bij het scheiden van stemmen	17
1.3.2	Onderbroken en temporeel gesegmenteerde spraak	21
1.4	Discussie en conclusie	32
2.	AUDITIEVE SAMENHANG VAN REEKSEN SPRAAKFRAGMENTEN MET ALTERNERENDE TOONHOOGTE	38
2.1	Inleiding	38
2.2	Exploratieve metingen	39
2.3	Klinkerreeksen en auditieve samenhang	42
2.4	Klinkerduren, pauzeduren en inzet-tot-inzettijden	48
2.5	Formantbuigingen en auditieve samenhang	50
2.6	Toonhoogtebewegingen en auditieve samenhang: aanvullende waarnemingen	58
2.7	Discussie en conclusie	60
3.	PERCEPTIEVE SCHEIDING VAN SIMULTANE SPRAAKUITINGEN	64
3.1	Inleiding	64
3.2	Natuurlijke spraak	66
3.2.1	Normaal geïntoneerde spraak	66
3.2.2	Quasi-monotone spraak	77
3.2.3	Bespreking van de resultaten	84
3.3	Volstrekt monotone spraak	86
3.3.1	Inleiding	86
3.3.2	Methode	87
3.3.3	Resultaten	90

3.3.4	Bespreking van de resultaten	93
3.4	Discussie en conclusie	95
4.	CONCLUSIE	102
4.1	Drie stappen op weg naar auditieve samenhang	102
4.2	Enkele kritische kanttekeningen en suggesties voor verder onderzoek	109
	SAMENVATTING	115
	SUMMARY	118
	LITERATUUR	121
	CURRICULUM VITAE	124

1. AUDITIEVE SAMENHANG EN SPRAAKVERSTAAN

1.1 Inleiding

Spraak bestaat uit reeksen elkaar snel opvolgende geluiden die voor een luisteraar herkenbare patronen moeten vormen en waaraan hij betekenissen kan hechten. De luisteraar moet in staat zijn om de verschillende geluiden die de spreker voortbrengt, als bij elkaar horend waar te nemen zonder dat er onzekerheid bestaat over de onderlinge tijdsrelaties tussen de verschillende akoestische gebeurtenissen. Verder moet de luisteraar de reeksen geluiden kunnen verwerken tot herkenbare taalpatronen.

Dit als bij elkaar horend waarnemen van akoestische gebeurtenissen zonder dat er onzekerheid bestaat over de onderlinge tijdsrelaties, zullen wij in het vervolg *auditieve samenhang* noemen.

Wij nemen aan dat auditieve samenhang een positieve bijdrage levert aan de herkenbaarheid van het signaal. Voor spraak houdt dit in dat auditieve samenhang bijdraagt tot de spraakverstaanbaarheid. Dit kan als volgt geïllustreerd worden.

Het is mogelijk opnames van duidelijk door één spreker geïsoleerd ingesproken woorden zonder pauzes samen te voegen zodat deze een betekenisvolle spraakuiting opleveren. Dit noemt men geassembleerde spraak.

Bij het luisteren hiernaar treden een aantal verschijnselen op die duiden op een verstoring van de auditieve samenhang. Luisteraars die geconfronteerd worden met deze vaak slecht verstaanbare geassembleerde spraakuitingen, hebben het idee dat hoewel de woorden door één spreker ingelezen zijn verschillende sprekers aan het woord zijn en dat de woorden elkaar in de tijd overlappen. Het kan zelfs voorkomen dat de volgorde van de woorden niet correct waargenomen worden. De auditieve samenhang van het signaal is verstoord, wat leidt tot een bemoeilijking van de herkenning. De oorzaak hiervan zal in dit geval gezocht kunnen worden in de dis-

continuïteiten op de woordgrenzen in het verloop van toonhoogte*, amplitude, spectrale samenstelling en in het ontbreken van normale fonologische overgangen zoals assimilatie. Ook zou de incorrecte temporele structuur van het geheel een bijdrage tot de verstoring kunnen opleveren.

Geassembleerde spraak is door het gebrek aan auditieve samenhang vaak slecht verstaanbaar. Maar vooral wanneer de grammaticale vormen en de betekenisstructuren van de spraakuitingen niet te sterk afwijken van wat de hoorder verwacht, kan het gebeuren dat spraakuitingen die duidelijk auditief niet samenhangend zijn, toch correct herkend worden. Kennelijk is auditieve samenhang wél bevorderlijk voor de verstaanbaarheid, maar niet een noodzakelijke voorwaarde.

Wanneer contextuele factoren zoals syntactische en semantische overgangswaarschijnlijkheden voldoende "overtollige" informatie in de spraakuitingen brengen, kan de luisteraar ook zonder auditieve samenhang reconstrueren wat de spreker bedoeld heeft.

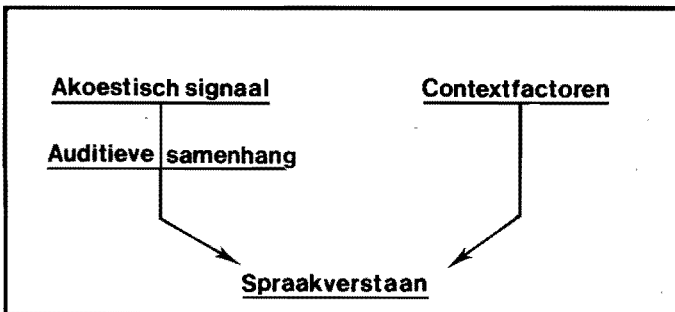


Fig. 1.1 Schematische voorstelling van de beïnvloeding van het spraakverstaan door auditieve samenhang en contextfactoren.

* Onder toonhoogte verstaan we in deze studie hetzij fysieke toonhoogte hetzij waargenomen toonhoogte. Waar onduidelijkheid zou kunnen ontstaan is steeds aangegeven wat bedoeld is.

Het doel van dit onderzoek is de relatie te onderzoeken tussen enkele stimulouseigenschappen van het akoestische signaal enerzijds en de auditieve samenhang anderzijds. Auditieve samenhang kan zich manifesteren in voldoende verstaanbaarheid, het correct waarnemen van volgordes van de spraakklanken en het subjectief waarnemen van de identiteit van de bron waaruit het signaal afkomstig lijkt te zijn. Dit is met name van belang om juist in alledaagse, niet optimale luisteromstandigheden, waarbij vaak sprake is van concurrerende geluiden, zoals omgevingslawaai of andere sprekers, in staat te zijn om de akoestische signalen die afkomstig zijn van één spreker te onderscheiden van interfererende geluiden en de klanken te integreren tot een geheel van herkenbare taalpatronen.

Herkenning van bijvoorbeeld woorden en syntactische verbanden valt buiten het onderwerp van deze studie. Bij het opzetten van het onderzoek zijn we ervan uitgegaan dat het mogelijk is de relatie tussen fysische eigenschappen van spraak en auditieve samenhang te onderzoeken en te beschrijven zonder tevens uitvoerig verantwoording af te leggen van alle processen die samen leiden tot het verstaan van spraak.

1.2 Auditieve Samenhang van Betekenisloze Reeksen Geluiden

Auditieve samenhang is moeilijk rechtstreeks te meten. Verstoring van de auditieve samenhang resulteert in verslechterde verstaanbaarheid. Deze verslechterde verstaanbaarheid kan een meetbaar resultaat zijn van verstoring van de auditieve samenhang. Maar zoals eerder betoogd, is de spraakverstaanbaarheid niet alleen afhankelijk van auditieve samenhang maar ook van bijdragen van contextfactoren. Indien men de contextfactoren constant weet te houden of te elimineren, dan kan verslechtering van het spraakverstaan alleen nog maar veroorzaakt worden door verstoring van de auditieve samenhang. Daarmee heeft men dan de mogelijkheid gekregen om de relatie tussen de fysische aspecten van het akoestische signaal en de auditieve samenhang nader te bestuderen. Het elimineren van contextfactoren in luisterproeven kan gebeuren

door gebruik te maken van betekenisloze reeksen geluiden zoals zuivere tonen, klinkers of medeklinker-klinker combinaties (CV-combinaties).

We zullen nu eerst een aantal literatuurgegevens bespreken die ons inziens onmiddellijk te maken hebben met het begrip auditieve samenhang.

1.2.1 Waarneming van volgordes in reeksen geluiden

Een aspect van auditieve samenhang is dat een luisteraar in staat is volgordes in reeksen geluiden correct waar te nemen. Hoewel de correcte waarneming van volgordes een noodzakelijke voorwaarde is voor de perceptie van verbonden spraak, is het geen triviale voorwaarde waaraan altijd wordt voldaan. Dit kan geïllustreerd worden met het volgende voorbeeld.

Thomas, Hill, Carrol en Garcia (1970) boden luisteraars herhaalde reeksen aan, bestaande uit vier klinkers (/i/, /ε/, /a/ en /u/). Deze klinkers waren zonder pauzes geassembleerd en toonhoogte, intensiteit en duur van elk van de klinkers waren zo goed mogelijk gelijk gemaakt. Luisteraars bleken nauwelijks of niet in staat te zijn om de volgorde van de klinkers te reproduceren indien de klinkerduren korter waren dan 125 ms. De gemiddelde duur van fonemen in verbonden spraak waarbij geen onzekerheid optreedt over de volgordes, is aanzienlijk korter.

In de vorige paragraaf hebben we gemeld dat ook bij de assemblage van grotere spraaksegmenten, zoals woorden, luisteraars soms slecht in staat zijn om volgordes correct waar te nemen. Dit gebrek aan auditieve samenhang lijkt te worden veroorzaakt door de discontinuïteiten van het akoestische signaal op de lussen, die bij assemblage ontstaan.

In de klinkerreeksen die door Thomas et al. gebruikt zijn, zijn discontinuïteiten van toonhoogte en amplitude op de grenzen van de klinkers vermeden door de toonhoogtes en amplitudes zo goed mogelijk constant te houden. Daarom moet in dit geval de oorzaak van de auditieve incoherentie in het optreden van spectrale discontinuïteiten gezocht worden,

dat wil zeggen discontinuïteiten in de formantstructuren op de lassen tussen de klinkers. Dit wordt bevestigd door de resultaten van Cole en Scott (1973) en Dorman, Cutting en Raphael (1975).

Cole en Scott lieten luisteraars vier medeklinker-klinker-combinaties horen, bijv. sa, sha, va en ga.

Deze waren opgenomen op een ronddraaiende lus van magnetische band (bandlus) waardoor zij met korte tussenpauzes een groot aantal malen hoorbaar gemaakt konden worden. De volgordes werden door de luisteraars het slechtst gereproduceerd bij de aanbiedingen van CV-syllabes zonder spectrale overgangen.

Deze waarneming vormde voor Dorman et al. (1975) aanleiding tot een serie experimenten waarin gekeken werd naar de waarneming van volgordes bij reeksen geassembleerde klinkers met en zonder spectrale overgangen. Zij prepareerden reeksen bestaande uit 4 geassembleerde klinkers met elk een duur van 120 ms en met een toonhoogte van 110 Hz. De reeksen bestonden uit (a) homogene klinkers, (b) tweeklankachtige klinkers, (c) /b/-klinker-/b/-lettergrepen en (d) pseudo-lettergrepen met fonetisch niet mogelijke spectrale overgangen.

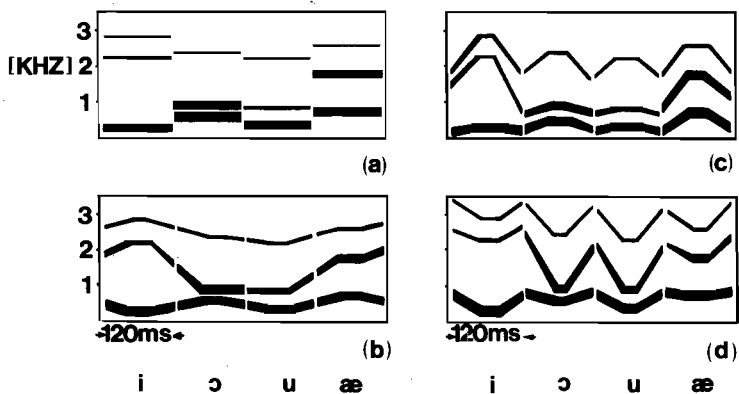


Fig. 1.2 Spectrografische weergave van vier soorten stimuli die door Dorman et al. gebruikt zijn.

- a) homogene klinkers
- b) tweeklankachtige klinkers
- c) klinkers verbonden met /b/-achtige formantbuigingen
- d) klinkers verbonden met fonetisch niet mogelijke spectrale overgangen.

De luisteraars moesten de volgordes van de reeksen klinkers vaststellen. Bepaalde combinaties van homogene klinkers leverden de luisteraars moeilijkheden op bij het vaststellen van de volgordes. Bij deze combinaties bleken de tweeklankachtige verglijdingen en /b/achtige formantbuigingen tussen de klinkers aanzienlijke verbeteringen van de resultaten te geven. De reeksen met fonetisch niet mogelijke formantbuigingen leverden in het algemeen de slechtste resultaten op. De lage scores voor dit laatste geval zouden te wijten kunnen zijn aan de moeilijkheid deze klanken te herkennen. De verschillen tussen de resultaten van de reeksen met homogene klinkers en de twee andere reeksen met formantbuigingen, suggereren dat formantovergangen van de ene klank naar de andere een bijdrage leveren aan het vermogen van de luisteraar om de volgordes van spraakklanken correct waar te nemen.

In vergelijking met de volgordewaarneming in verbonden spraak is deze zowel bij assemblages van korte geïsoleerde spraakklanken als van geïsoleerde woorden slecht. De resultaten van de experimenten van Dorman et al. en van Cole en Scott wijzen erop dat spectrale overgangen of formantbuigingen in reeksen korte spraakklanken met dezelfde toonhoogte de luisteraars in staat stellen volgordes van de fonetische segmenten beter te identificeren. Waarschijnlijk hebben formantbuigingen in gewone spraak een dubbele functie. Ten eerste dragen zij belangrijke fonetische informatie, nodig voor de herkenning van fonemen, ten tweede bevorderen ze de auditieve samenhang van de spraakstroom.

1.2.2 Auditieve splitsing

Bij het luisteren naar geassembleerde spraak hebben luisteraars vaak de indruk dat de woorden van de geassembleerde spraakuitingen ingesproken zijn door verschillende sprekers, ook als de woorden in feite door één spreker ingesproken zijn. Het lijkt erop dat de discontinuïteiten in het

akoestische signaal op de woordgrenzen de luisteraar in onzekerheid brengen over de identiteit van de geluidsbron. Dit verschijnsel zou verband kunnen houden met een perceptief verschijnsel dat in het Engels "auditory stream segregation" genoemd wordt en dat wij in het Nederlands zullen aanduiden met "auditieve splitsing". Dit treedt bijvoorbeeld op bij de aanbieding van een snelle opeenvolging van hoge en lage tonen. Perceptief lijkt zo'n reeks van opeenvolgende tonen zich te splitsen in twee onafhankelijke reeksen: een reeks met hoge tonen en een reeks met lage tonen. Het zich splitsen van een reeks tonen in meerdere perceptieve subreeksen is afhankelijk van zowel de snelheid waarmee de verschillende tonen elkaar afwisselen als van het verschil in frequentie tussen de hoge en lage tonen (Miller and Heise, 1950; van Noorden, 1975). Naarmate het verschil in frequentie tussen de tonen en het tempo waarmee de tonen worden aangeboden toenemen, wordt de reeks hoge en lage tonen gemakkelijker opgesplitst in twee perceptief gescheiden stromen.

Behalve dat de tonen zich afhankelijk van de frequentie perceptief gaan hergroeperen in gescheiden stromen, heeft auditieve splitsing ook consequenties voor de waarneming van volgordes.

Binnen elke perceptieve stroom blijken volgordes correct waargenomen te worden, maar de correcte waarneming van volgordes tussen nabijgelegen tonen die tot verschillende perceptieve stromen behoren blijkt niet goed mogelijk te zijn (Bregman and Campbell, 1971).

Het aanbrengen van frequentieverglijdingen van de ene toon naar de andere gaat het optreden van auditieve splitsing tegen. Dientengevolge wordt ook de waarneming van volgordes beter (Bregman en Dannenbring, 1973).

Het optreden van het verschijnsel van auditieve splitsing bij het luisteren naar reeksen geluiden is niet alleen beperkt tot reeksen zuivere tonen, maar treedt ook op bij reeksen spraakachtige klanken. Lackner en Goldstein (1974) namen waar dat bij reeksen van 200 ms durende, aaneengesloten

klinkers en medeklinker-klinkercombinaties afwisselend ingesproken door een man en een vrouw, auditieve splitsing optrad en eveneens dat dit interfereerde met de correcte waarneming van de volgordes van de items in de aangeboden reeksen.

1.2.3 Auditieve splitsing en de waarneming van volgordes

Auditieve samenhang heeft een tweetal voor de perceptie van spraak belangrijke aspecten. Ten eerste geeft het de luisteraar aanwijzingen over welke akoestische signalen van dezelfde geluidsbron (d.w.z. spreker) afkomstig zijn en dus perceptief bij elkaar moeten horen. Ten tweede waarborgt auditieve samenhang een correcte waarneming van de temporele relaties tussen de akoestische gebeurtenissen onderling.

Auditieve samenhang is een perceptieve eigenschap van een reeks geluiden. In het akoestisch spraaksignaal zijn kennelijk factoren aanwezig die bijdragen tot de auditieve samenhang. Continuïteit van de formantstructuur tussen spraakklanken bevordert duidelijk de correcte waarneming van volgordes (zie Dorman et al., 1975; Cole and Scott, 1973). De toonhoogte van spraakklanken lijkt ook een belangrijke rol te spelen bij het vaststellen van welke klanken bij elkaar horen.

Deze eenvoudige indeling naar oorzaak en gevolg zoals:

- *spectrale continuïteit* levert een bijdrage aan een correcte waarneming van volgorde in een reeks opeenvolgende spraakachtige klanken en
- *toonhoogte* levert de luisteraar een indicatie welke klanken van dezelfde bron afkomstig lijken te zijn en daardoor bij elkaar horen,

is goed bruikbaar om verschillende verschijnselen bij de perceptie van spraak te kunnen plaatsen. Een dergelijke indeling heeft echter wel het gevaar in zich om de zaken te simpel voor te stellen. Daarom lijkt het zinvol om een paar discussiepunten naar voren te halen.

Bij experimenten aan auditieve splitsing wordt de luisteraar meestal een zich steeds herhalend patroon van stimuli aangeboden. Deze steeds voorspelbare opbouw van de stimulus zou bevorderlijk kunnen zijn voor de perceptie van auditieve splitsing in reeksen stimuli. De kern van de zaak is echter niet dat splitsing op gaat treden, maar dat door de toonhoogteverschillen de samenhang van het signaal verstoord wordt. Een aanwijzing dat het repeterende karakter van de stimulus niet de enige oorzaak is van de verstoring van de auditieve samenhang wordt geleverd door een experiment van Van Noorden (1975). Bij reeksen van zuivere tonen met een toevalsverdeling van frequenties, nemen luisteraars evenals bij periodiek voortgezette reeksen een verstoring van de samenhang van het signaal waar.

Een ander discussiepunt vormt de rol van de toonhoogteverschillen bij het tot stand komen van splitsing. Van Noorden (1975) heeft gevonden dat bij snelle reeksen van korte complexe tonen met een duur van 40 ms, de samenhang van het signaal afhangt van het al of niet aaneensluiten van de spectrale componenten van de stimuli. Reeksen complexe tonen met verschillende toonhoogtes, maar met harmonische componenten in hetzelfde spectrale gebied, geven bij tijden van 100 ms, gemeten van inzet tot inzet, een samenhangend signaal. (We zullen voortaan zulke tijden "inzet-tot-inzet" tijden noemen). Dit, terwijl reeksen tonen met dezelfde toonhoogte maar samengesteld uit harmonische componenten in verschillende niet aaneensluitende spectrale gebieden, aanleiding geven tot een niet samenhangend signaal.

Tot nu toe zijn een aantal aspecten besproken waaraan het akoestisch signaal moet voldoen wil het als een samenhangend signaal door de luisteraar opgevat worden. Bij normale verbonden spraak levert samenhang in het geheel geen probleem op. Aan alle eisen voor auditieve samenhang voldoet het spraaksignaal vanzelf. Dit komt doordat de bewegingen van de spraakorganen die de geluidstrillingen voortbrengen, niet discontinu verlopen. Daarom zal bij het voortbrengen van spraak toonhoogte noch spectrale samenstelling discontinu

veranderen.

In het voorafgaande hebben we gezien dat de continuïteit van de toonhoogte en van de spectrale samenstelling een belangrijke functie vervullen bij de waarneming. Juist deze fysische continuïteit van het spraaksignaal is verantwoordelijk voor de auditieve samenhang. Bij de waarneming zorgt de auditieve samenhang ervoor dat de luisteraar de klanken die één enkele spreker voortbrengt, ook aan één enkele spreker toeschrijft en zonder onzekerheid over de volgorde van de fonemen kan verwerken. Bij het horen van meerdere gelijktijdig sprekende sprekers geeft het de luisteraar de mogelijkheid om een perceptieve selectie te maken van de geluiden die van dezelfde bron afkomstig zijn. Hij kan op grond van de auditieve samenhang van de spraak, die elk van de sprekers voortbrengt, de reeks akoestische gebeurtenissen perceptief opdelen in parallelle stromen auditieve informatie. Elk van deze stromen voldoet in principe aan de eisen voor auditieve samenhang. Door deze opdeling van het geluidssignaal is de luisteraar in staat om in dergelijke complexe situaties tot spraakperceptie te komen.

1.3 Verstaanbaarheid van Verbonden Spraak

In het dagelijkse leven hebben luisteraars vaak te maken met spraaksignalen die niet optimaal aangeboden worden en aan verstoringen van velerlei aard onderhevig zijn. Verstoring van het aangeboden spraaksignaal door interferentie met andere geluiden zal vermoedelijk het meest frequent voorkomen. Een voorbeeld van een dergelijke situatie is die waarin verschillende sprekers door elkaar heen spreken. Hoe is een luisteraar in staat om de geluiden die afkomstig zijn van dezelfde spreker te onderscheiden en te selecteren uit de andere gelijktijdige geluiden en hoe is de luisteraar in staat om de geselecteerde geluiden te verwerken tot herkenbare taalpatronen?

Een andere situatie is die waarin het spraaksignaal onderbroken wordt of onderbroken lijkt, doordat delen van het spraaksignaal gemaskeerd worden door veel luidere signalen.

Ook hierbij moet de luisteraar in staat zijn de stukjes spraak te selecteren en te verwerken tot herkenbare taalpatronen. In een dergelijke situatie is het niet een probleem om het te volgen spraaksignaal uit een verzameling van andere auditieve signalen te selecteren, maar meer een probleem om de fragmenten spraak met elkaar in verband te brengen en de boodschap te reconstrueren.

1.3.1 Factoren die een rol spelen bij het scheiden van stemmen

In een situatie waar meerdere sprekers tegelijk aan het woord zijn, moeten luisteraars in staat zijn een onderscheid te maken tussen de verschillende stemmen en vervolgens hun aandacht richten op één van die stemmen. In 1953 heeft Cherry een drietal experimenten gerapporteerd waarin hij probeerde de vraag te beantwoorden hoe wij kunnen herkennen wat iemand zegt als anderen tegelijkertijd aan het spreken zijn. Als factoren die een luisteraar zouden kunnen helpen bij het scheiden van twee simultane boodschappen, noemt hij het richting horen, visuele informatie, overgangswaarschijnlijkheden en verschillen in accent en in stemidentiteit. In het eerste experiment krijgen luisteraars twee simultane boodschappen, ingesproken door dezelfde spreker, binauraal aangeboden. Met deze opzet van het experiment blijven volgens Cherry de overgangswaarschijnlijkheden over als mogelijke helpende factor om beide stemmen te kunnen scheiden. De luisteraar mag zo vaak als nodig de opname beluisteren en moet dan verbaal een van de boodschappen reproduceren. Hoewel de luisteraars deze opdracht erg moeilijk vinden, blijken ze elke boodschap correct weer te kunnen geven. De overgangswaarschijnlijkheden zoals bepaald door de grammaticale structuur en de betekenis van de boodschap bieden de luisteraars kennelijk voldoende houvast om beide boodschappen te scheiden en één van deze te reproduceren.

In het tweede experiment laat Cherry de luisteraars hetzelfde doen, maar het stimulusmateriaal is hierbij iets

anders georganiseerd. De simultane boodschappen die ingesproken zijn door dezelfde spreker, zijn opgebouwd uit clichés. Nu blijkt dat het scheiden van beide boodschappen moeilijk is geworden. De antwoorden van de luisteraars zijn vaak opgebouwd uit clichés afkomstig uit de twee verschillende boodschappen. Het is duidelijk, dat de overgangswaarschijnlijkheden een heel belangrijke rol spelen.

Bij het derde experiment biedt Cherry de twee boodschappen dichotisch aan: het ene bericht op het ene oor, het andere bericht op het andere oor. Beide boodschappen zijn eveneens ingesproken door dezelfde spreker. Bij dit experiment introduceert Cherry de techniek van het schaduwen: de luisteraar krijgt als opdracht de boodschap op het ene oor te negeren en die op het andere te herhalen terwijl hij aan het luisteren is. Het blijkt dat deze opdracht voor de luisteraar betrekkelijk makkelijk is. Het dichotisch aanbieden van de boodschappen blijkt voor de luisteraars een goede hulp om deze boodschappen van elkaar te scheiden.

Cherry's experimenten laten zien dat zowel contextfactoren, zoals overgangswaarschijnlijkheden, als richting horen een invloed hebben op het perceptief kunnen scheiden en volgen van twee simultane boodschappen.

Anne Treisman (1964) beschrijft in haar artikel "Verbal cues, language and meaning in selective attention" een aantal experimenten waarin zij onderzoekt in hoeverre fysische verschillen een bijdrage kunnen leveren aan het vermogen van een luisteraar om verschillende simultane stemmen bij binaurale aanbieding perceptief te scheiden. Zij rapporteert de volgende bevindingen:

- verschillen tussen een mannelijke en vrouwelijke stem zorgen voor een zeer goede en efficiënte perceptieve scheiding;
- bij aanbieding van twee boodschappen, elk in een verschillende taal, maar ingesproken door dezelfde spreker blijkt complete scheiding van beide stemmen onmogelijk te zijn;

- hoe meer de fonetische kenmerken tussen beide talen verschillen, des te gemakkelijker kan de luisteraar beide stemmen scheiden;
- kennis van de gebruikte talen blijkt te interfereren met het scheidingsvermogen van de luisteraar;
- bij aanbieding van twee boodschappen in dezelfde taal en ingesproken door dezelfde spreker blijkt volledige scheiding onmogelijk te zijn.

Egan, Carterette en Thwing (1954) beschrijven een onderzoek waarin de effecten van band-doorlaat filteren en de effecten van relatieve verschillen in geluidsterkte van twee simultane boodschappen op het scheidend vermogen bestudeerd worden. Ook hierbij zijn beide boodschappen ingesproken door dezelfde spreker. De aanbieding is monoraal. Hoogdoorlaat filteren van óf de maskeertekst óf van de tekst waarop gelet moet worden heeft een positief effect op de verstaanbaarheid. In het tweede gedeelte van het experiment zijn de relatieve geluidsterktes gevariëerd.

Als de testzinnen luider worden dan de maskeertekst verbetert de verstaanbaarheid snel. In het gebied waarin beide boodschappen ongeveer even luid zijn, is er nauwelijks sprake van een verandering in de verstaanbaarheid. Buiten dit gebied neemt de verstaanbaarheid ongeveer 4% per dB verzwakking van de testzinnen af.

Behalve contextuele factoren en stimulusfactoren zoals verschillen in toonhoogte, stemkwaliteit en luidheid, zou ook prosodische continuïteit een belangrijke factor kunnen zijn bij het perceptief scheiden van simultane stemmen. In een experiment dat lijkt op dat van Treisman (1960) variëerde Darwin (1975) prosodische en betekenisfactoren onafhankelijk van elkaar. Luisteraars kregen in een schaduwexperiment paarsgewijze twee verschillende spraakpassages van ongeveer 50 woorden dichotisch aangeboden. Er werden een viertal verschillende condities geconstrueerd.

- Een normale conditie, waarin de twee originele passages, ieder op één oor aangeboden werden;

- Een tweede conditie met een breuk in de betekenisstructuur. Hierbij werd de eerste helft van de ene passage gelast aan de tweede helft van de andere passage. De zo verkregen combinaties werden ieder als één geheel ingelezen. Beide combinaties werden gelijktijdig ieder op één oor aangeboden. Prosodisch gezien vormt elk van deze combinaties één geheel terwijl bij de aanbidding de betekenisstructuur van het ene oor naar het andere overspringt.
- Een derde conditie met een prosodische breuk. Hierbij werden de als één geheel ingelezen combinaties uit de vorige conditie weer in twee helften gedeeld en de eerste helft van de ene combinatie aan de tweede helft van de andere combinatie gelast. De zo verkregen passages die ieder qua betekenis één geheel vormden, maar prosodisch discontinu waren, werden eveneens dichotisch aangeboden.
- Een vierde conditie waarin gelijktijdig zowel een prosodische breuk als een breuk in de betekenisstructuur voorkomt. Hierbij werden de twee originele passages uit de eerste conditie in twee stukken verdeeld waarna het eerste gedeelte van de ene passage aan het tweede gedeelte van de andere passage gelast werd. De twee zo verkregen combinaties werden eveneens dichotisch aangeboden en zijn elk zowel in prosodisch opzicht als in betekenisstructuur discontinu.

De opdracht van de luisteraar was het signaal op één van de oren te schaduwen en het signaal op het andere oor te negeren. De belangrijkste resultaten van zijn experimenten waren dat luisteraars bij de conditie waarin de prosodie van de zin op het ene oor overgaat naar de zin op het andere oor relatief vaak de aandacht mee verwisselden naar de boodschap op het oor dat genegeerd diende te worden. Dit kwam tot uiting in de aantallen woorden die na het optreden van de breuk geschaduwd werden en afkomstig waren van de boodschap op het te negeren oor. Wanneer de betekenisstructuur van de zin op het ene oor voortgezet werd in de zin van het andere oor, resulteerde dit in het weglaten van woorden.

Het lijkt erop dat luisteraars alle informatie die beschikbaar is, gebruiken om hun doel, het scheiden van twee simultane stemmen, zo goed mogelijk te bereiken. Zij doen dit door gebruik te maken van stimulouseigenschappen zoals verschillen tussen de stemmen, de richting van waaruit de stemmen komen, intensiteitsverschillen en wat de luisteraar nog meer kan vinden om de stemmen van elkaar te scheiden. Als het scheiden van de stemmen op grond van alleen stimulouseigenschappen niet lukt, kunnen de luisteraars gebruik maken van contextgegevens van de gesproken boodschappen.

Hoewel het niet makkelijk is een strategie te geven die een luisteraar in een bepaalde situatie volgt, sluit dit wel aan bij ideeën uit de literatuur over selectieve aandacht. Men neemt aan dat in de waarneming de aandacht selectief gericht wordt door twee simultaan uitgevoerde analyses: een "bottom-up" analyse van het binnenkomende signaal en een "top-down" analyse die gestuurd wordt door contextafhankelijke factoren (vgl. Norman, 1976). Tot nu toe hebben in de literatuur die gericht is op de selectieve aandacht, de contextfactoren aanzienlijk meer aandacht gekregen dan de stimulusfactoren.

1.3.2 Onderbroken en temporeel gesegmenteerde spraak

Naast verstoring van spraak door interfererende geluiden of stemmen levert onderbreking van het spraaksignaal een frequent voorkomende moeilijke luistersituatie op. Het onderbreken of wegvallen van het geluidssignaal kan in de praktijk voorkomen door een slecht communicatiekanaal of door maskering tengevolge van veel luidere signalen. Het probleem voor de luisteraar ligt hierin dat hij voor het verstaan in staat moet zijn om de fragmenten spraak met elkaar in verband te brengen. Deze fragmenten moet hij met elkaar integreren tot een percept van de boodschap.

Ingrepen in het tijddomein van het spraaksignaal hebben geleid tot een hele klasse van experimenten. Om een indruk te krijgen van de verschillende temporele ingrepen die in

de loop van de tijd onderzocht zijn, is het nuttig om de meest bekende hiervan de revue te laten passeren. Dit zijn onderbroken spraak, temporeel gesegmenteerde spraak en gealterneerde spraak. Bij *onderbroken spraak* wordt het signaal veelal onderbroken met een vaste tijdsteek, zie fig. 1.3 (Miller and Licklider, 1950; Cherry and Taylor, 1954; Huggins, 1964; Powers and Speaks, 1973; Wingfield and Wheale, 1975a, 1975b). Als variabele wordt meestal de onderbrekingsfrequentie gekozen. De verstaanbaarheid van de onderbroken spraak blijkt afhankelijk te zijn van de onderbrekingsfrequentie.

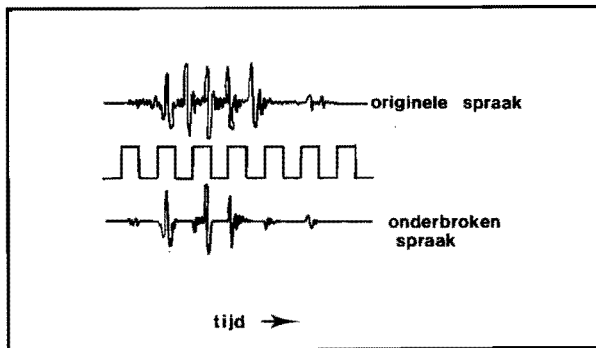


Fig. 1.3 Het effect van het onderbreken van een spraaksignaal.

Bij lage onderbrekingsfrequenties echter wordt de verstaanbaarheid voornamelijk bepaald door de verhouding van de duur van de spraaksegmenten en de periodeduur. Dat komt doordat grote delen van het spraaksignaal de luisteraar niet ter beschikking staan.

Temporeel gesegmenteerde spraak bezit dit nadeel niet. Het volledige spraaksignaal blijft de luisteraar ter beschikking staan. Hierbij worden met een vaste tijdsteek pauzes met een vaste lengte in het spraaksignaal ingelast (zie fig. 1.4).

Het geluidssignaal vertoont hierdoor grote gelijkenis met onderbroken spraak. Spraaksegmenten worden steeds afgewisseld met stille intervallen. De verstaanbaarheid vertoont,

behalve bij lage onderbrekingsfrequenties, grotendeels eenzelfde gedrag als bij onderbroken spraak.

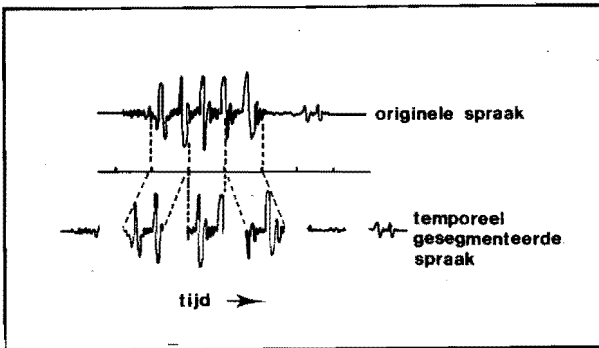


Fig. 1.4 Het effect van het met een vaste tijdsteek invoegen van stille intervallen in het spraaksignaal.

Gealterneerde spraak is eigenlijk een bijzonder geval. Hierbij is het geluidssignaal niet voor beide oren identiek. Met een vaste tijdsteek wordt het signaal van het linker naar het rechter oor en vice versa overgeschakeld. Voor elk van beide oren lijkt het signaal dus op onderbroken spraak (zie fig. 1.5).

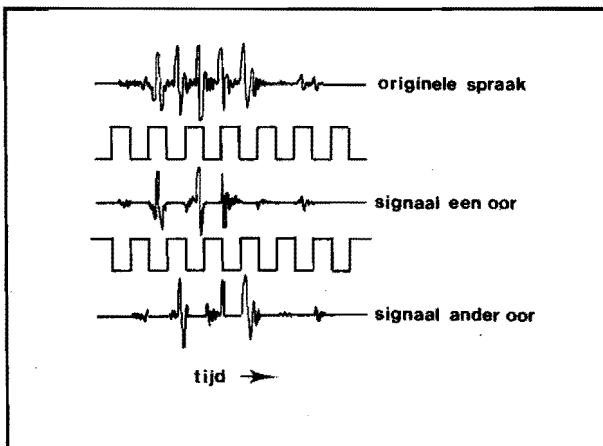


Fig. 1.5 Het gealterneerde spraaksignaal voor beide oren.

De beide oren krijgen elkaars complementaire deel van het geluidssignaal aangeboden. De verstaanbaarheid vertoont een gedrag dat vergelijkbaar is met onderbroken en temporeel gesegmenteerde spraak.

Van de soorten temporeel gemanipuleerde spraak die hier opgesomd zijn, zijn voor ons doel onderbroken spraak en temporeel gesegmenteerde spraak het belangrijkste. Onderbroken spraak staat het dichtst bij de in het dagelijks leven voorkomende verstoringen en temporeel gesegmenteerde spraak vertoont veel gelijkenis met onderbroken spraak, maar is eenvoudiger interpreteerbaar. Beide types worden nader besproken.

Onderbroken spraak

Bij onderbroken spraak zijn de voornaamste variabelen de duur van de spraaksegmenten en de duur van de intervallen waarin de spraak weggevallen is. Hoe meer van een woord gehoord wordt, hoe beter de verstaanbaarheid. Hoe langer de stille intervallen hoe slechter de verstaanbaarheid. In 1950 hebben Miller en Licklider in het artikel "The intelligibility of interrupted speech" een aantal experimenten met onderbroken spraak beschreven. Zij bestudeerden het effect op de verstaanbaarheid van het met verschillende frequenties onderbreken van spraak. Bij onderbrekingsfrequenties tussen de 1 en 5 Hz en met een gelijke duur van spraaksegmenten en van stille intervallen, vinden zij evenals latere onderzoekers, bijvoorbeeld Huggins (1964) (zie fig. 1.6) steeds een lokaal minimum in de verstaanbaarheid. Bij lagere onderbrekingsfrequenties wordt ongeveer de helft van de woorden correct verstaan. Indien de onderbrekingsfrequentie hoog is, dan ondervindt de luisteraar weinig hinder en verstaat hij praktisch alle woorden correct. Een minimum in de verstaanbaarheid rond een onderbrekingsfrequentie van 1 à 5 Hz wordt behalve bij onderbroken spraak ook bij gealterneerde en temporeel gesegmenteerde spraak aangetroffen.

Miller en Licklider suggereren dat de verstaanbaarheid rond deze frequenties juist minimaal is omdat de

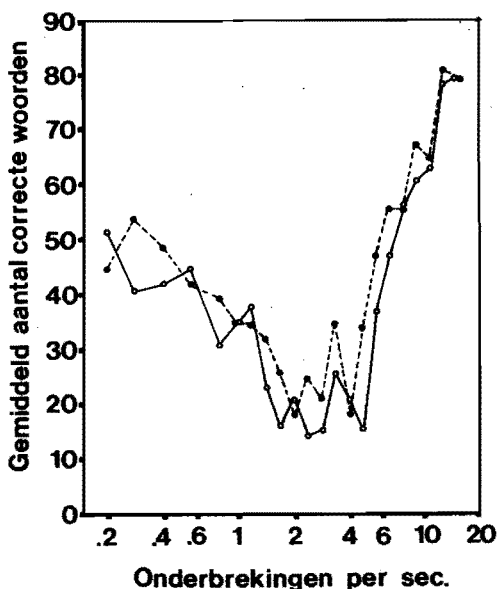


Fig. 1.6 De verstaanbaarheid van periodiek onderbroken spraak. De correcte scores zijn uitgezet als een functie van de onderbrekingsfrequentie. De gegeven resultaten zijn afkomstig van twee groepen van 10 luisteraars die elk één van de twee complementaire onderbroken signalen beluisterd hebben. Het totaal aantal gebruikte woorden was 100. (Huggins, 1964).

duur van de spraaksegmenten dan ongeveer gelijk is aan de duur van een woord. Zij redeneren als volgt: het gehele woord kan slechts correct gehoord worden als het woord tamelijk precies in een spraaksegment valt. Er is een vrij grote kans dat het woord niet verstaan wordt als het begin of het einde ervan in een stil interval terecht komt. Daarmee is dan de duur van de spraaksegmenten de bepalende factor voor de verstaanbaarheid. De gemiddelde woordduur is dan bepalend voor de onderbrekingsfrequentie waarbij het minimum in de verstaanbaarheid ontstaat.

Powers en Speaks (1973) vinden bij experimenten met onderbroken spraak dat behalve de duur van de spraaksegmenten en de temporele karakteristieken van de boodschap, de duur van de

stille intervallen mede bepalend is voor de verstaanbaarheid. Door namelijk de duur van de stille intervallen onafhankelijk van de duur van de spraaksegmenten te verkleinen wordt de verstaanbaarheid groter. Zij concluderen dat voor lage onderbrekingsfrequenties, waarbij de duren van de spraaksegmenten en van de stiltes groter zijn dan de gemiddelde woordduur, de verstaanbaarheid bijna uitsluitend afhangt van de verhouding tussen de duur van de spraaksegmenten en de periodeduur. Indien de onderbrekingsfrequenties hoger worden zodat de spraaksegmenten en de stille intervallen kleiner worden dan de gemiddelde woordduur, is de duur van de spraaksegmenten en de duur van de stille intervallen bepalend voor de verstaanbaarheid van onderbroken spraak.

Temporeel gesegmenteerde spraak

Bij onderbroken spraak lijkt het erop dat zowel de duur van de spraaksegmenten als de duur van de stille intervallen de verstaanbaarheid beïnvloeden. Een probleem met onderbroken spraak bij het bestuderen van de relatie tussen de verstaanbaarheid, de onderbrekingsfrequentie en de verhouding tussen de duren van de spraaksegmenten en stille segmenten, is dat steeds stukken auditieve informatie gedurende de stille intervallen wegvallen. Dit wordt vooral ernstig bij lage onderbrekingsfrequenties. Hierbij kunnen grote delen van woorden, gehele woorden of zelfs meerdere woorden in een stil interval terecht komen, waardoor de verstaanbaarheid bij lage frequenties nooit hoger kan worden dan de verhouding tussen de duren van de spraaksegmenten en de stille intervallen vermenigvuldigd met 100%.

Huggins (1972a, 1972b, 1974a, 1975a, 1975b) heeft daarom gebruik gemaakt van temporeel gesegmenteerde spraak. Het grote voordeel hiervan is dat alle akoestische informatie beschikbaar blijft, waardoor de verstaanbaarheid van temporeel gesegmenteerde spraak voornamelijk bepaald wordt door de duur van de spraaksegmenten en van de ingelaste pauzes. Huggins (1972a) stelt de hypothese, dat het V-vormige minimum in de verstaanbaarheid bij stimuli met een constante

verhouding spraak/stilte veroorzaakt wordt door twee van elkaar onafhankelijke effecten.

- (a) De kans dat een geïsoleerd spraaksegment herkend wordt vermindert als de duur van het spraaksegment kleiner wordt.
- (b) De kans dat de auditieve informatie van het ene spraaksegment met die in het volgende in verband gebracht kan worden, neemt toe naarmate de pauzes hiertussen kleiner worden.

Zijn meetgegevens, weergegeven in de volgende figuren, lijken zijn hypothese te bevestigen.

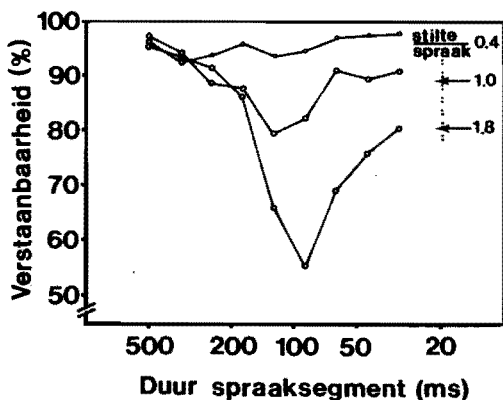


Fig. 1.7 Verstaanbaarheid uitgezet als functie van de duur van de spraaksegmenten. Er is sprake van drie versies van temporeel gesegmenteerde spraak, waarbij de duren van de stille intervallen zich verhouden als 0.4, 1.0 en 1.8 ten opzichte van de duur van de spraaksegmenten. (Huggins, 1972).

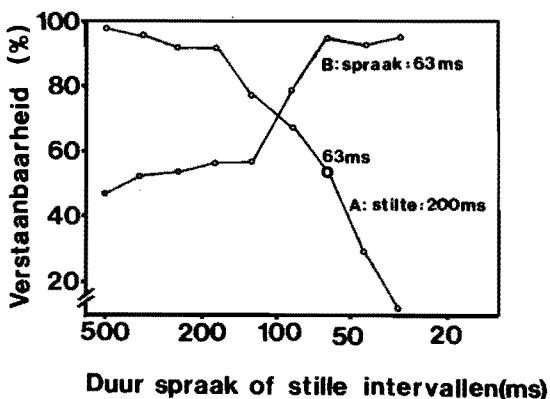


Fig. 1.8 Verstaanbaarheid voor temporeel gesegmenteerde spraak waarbij de duren van de stille intervallen en van de spraaksegmenten onafhankelijk gevarieerd zijn. Bij curve A zijn de stille intervallen constant gehouden (200 ms). Bij curve B zijn de duren van de spraakfragmenten constant gehouden (63 ms). (Huggins, 1972).

Model voor de beschrijving van de verstaanbaarheid van temporeel gesegmenteerde spraak

Uitgaande van de hypothese dat het effect op de verstaanbaarheid van de duren van de spraaksegmenten onafhankelijk is van het effect van de ingelaste pauzes, zullen we proberen om vanuit de meetgegevens van Huggins een beschrijving te geven van de verstaanbaarheid van temporeel gesegmenteerde spraak als functie van de duren van de spraaksegmenten en van de ingelaste pauzes.

Daartoe dient men de beide hypothesen van Huggins iets nauwkeuriger te specificeren:

- (a) Bij relatief lange spraaksegmenten, gescheiden door pauzes langer dan 200 ms, kan elk geïsoleerd spraaksegment goed verstaan worden. De kans op fouten zal dan praktisch nihil zijn. Bij zeer korte spraaksegmenten, die door pauzes van meer dan 200 ms gescheiden worden, is de verstaanbaarheid zeer klein en de kans op fouten bijna 1. Dus de kans op fouten neemt bij pauzes langer dan 200 ms toe naarmate de spraaksegmenten korter worden.
- (b) De kans op fouten bij spraaksegmenten korter dan 20 ms en gescheiden door pauzes langer dan 200 ms is ≈ 1 . Bij zeer korte pauzes tussen deze spraaksegmenten wordt de verstaanbaarheid nauwelijks aangetast. De kans op fouten is dan ≈ 0 . Dus naarmate de pauzes tussen de korte spraaksegmenten langer worden, neemt de kans op fouten toe.

Stelt men dat de kans op fouten bij lange pauzes als functie van de duur van de spraaksegmenten (P_{E_1}) en de kans op fouten bij korte spraaksegmenten als functie van de lengte van de pauzes (P_{E_2}) onafhankelijk van elkaar zijn en beschreven kunnen worden² door cumulatieve normale verdelingen, dan is het mogelijk om uit de data van Huggins' experimenten (Huggins 1972a, 1972b, 1975a) een schatting te maken van bij de verdelingsfuncties behorende parameters, te weten het gemiddelde μ en de standaarddeviatie σ . Deze zijn:

$$\mu_1 = 75 \text{ (ms)}, \sigma_1 = 89 \text{ (ms)}, \mu_2 = 99 \text{ (ms)}, \sigma_2 = 45 \text{ (ms)}$$

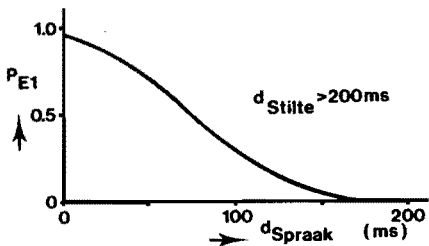


Fig. 1.9 Kans op fouten P_{E1} als een functie van de duur van de spraaksegmenten bij lange stille intervallen.

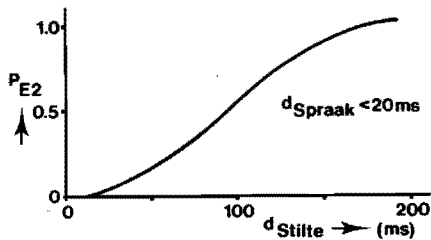


Fig. 1.10 Kans op fouten P_{E2} als een functie van de duur van de stille intervallen bij korte spraak segmenten.

Het is nu mogelijk om de totale kans op fouten (P_T), voor alle combinaties van de duren van de spraaksegmenten en van de ingelaste pauzes te beschrijven als een product van P_{E1} en P_{E2} .

$$P_T(d_{\text{spraak}}, d_{\text{stilte}}) = P_{E1}(d_{\text{spraak}}) \cdot P_{E2}(d_{\text{stilte}}) \quad (1)$$

De gemiddelde verstaanbaarheid V wordt dan:

$$V = (1 - P_T) \cdot 100\% \quad (2)$$

Hiermee wordt dan een goede beschrijving gegeven van de verstaanbaarheid van temporeel gesegmenteerde spraak. Dit blijkt uit de figuren 1.11 en 1.12. De onderbroken lijnen geven de waarden voor de verstaanbaarheid weer zoals deze op de bovenstaande manier berekend zijn.

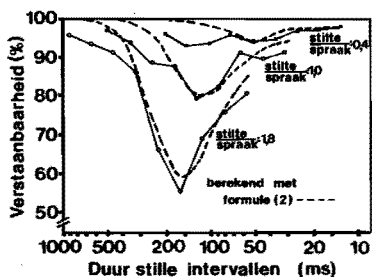


Fig. 1.11 Verstaanbaarheid voor drie versies van temporeel gesegmenteerde spraak, vergeleken met de met formule (2) berekende resultaten.

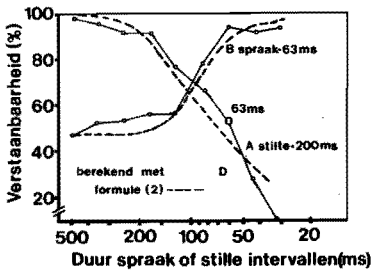


Fig. 1.12 Verstaanbaarheid voor temporeel gesegmenteerde spraak waarbij de duren van de stille intervallen en spraaksegmenten onafhankelijk gevarieerd zijn, vergeleken met de met formule (2) berekende resultaten.

In figuur 1.13 worden de berekende iso-verstaanbaarheidscurves weergegeven als functie van de duur van de spraaksegmenten en van de pauzes. Aan de hand van deze figuur wordt het duidelijk waarom de ligging van de minima in figuur 1.11 bij de verschillende verhoudingen tussen de duur van de spraaksegmenten en de duur van de ingelaste pauzes niet hetzelfde zijn. Zie ook figuur 1.14.

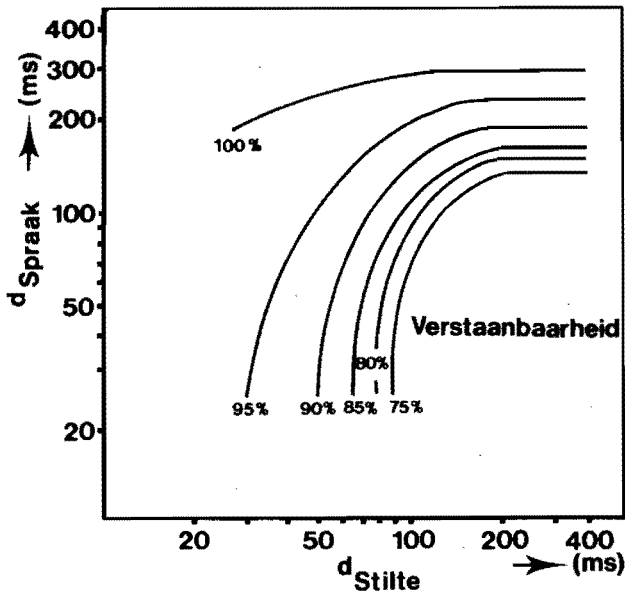


Fig. 1.13 Verstaanbaarheid als een functie van de duren van de spraaksegmenten en van de stille intervallen berekend met behulp van formule (2).

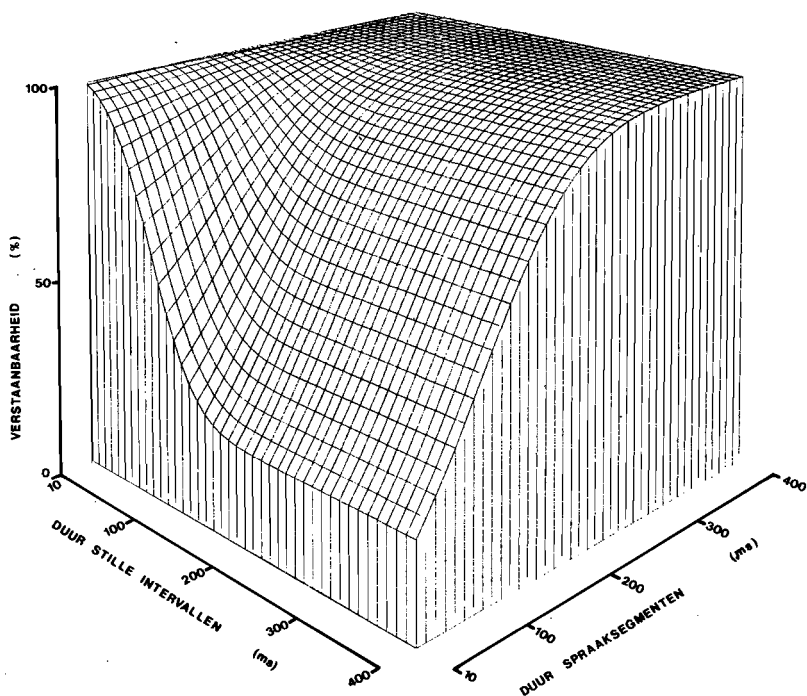


Fig. 1.14 Drie dimensionele weergave van de verstaanbaarheid als een functie van de duren van de spraaksegmenten en van de stille intervallen berekend met formule (2).

Ook wordt het optreden van een asymptoot in de verstaanbaarheidsscores bij groter wordende stille intervallen in figuur 1.12 begrijpelijker.

De goede beschrijving van de verstaanbaarheid van temporeel gesegmenteerde spraak met behulp van het hierboven gepresenteerde model, rechtvaardigt de conclusie dat de duren van de spraaksegmenten en de duren van de stille intervallen beide onafhankelijk van elkaar de verstaanbaarheid beïnvloeden. We kunnen proberen de resultaten van Huggins in termen van auditieve samenhang te verklaren door aan te nemen dat de auditieve samenhang van de spraak afneemt naarmate de stille intervallen langer worden. Wanneer het effect van de duur van

de stille intervallen op de verstaanbaarheid een bruikbare maat zou zijn voor de auditieve samenhang, zouden we moeten concluderen dat de auditieve samenhang minimaal is bij stille intervallen groter dan zo'n 100 à 200 ms. Wanneer namelijk de stille intervallen groter worden, neemt de verstaanbaarheid nauwelijks verder af. We moeten echter daarbij aantekenen dat ook bij grotere stille intervallen de zinsmelodie heel goed herkenbaar blijft. Dit wijst erop dat de notie auditieve samenhang zoals we die tot nu toe hebben gehanteerd te simpel is. In de volgende paragraaf zullen we hierop terugkomen.

1.4 Discussie en Conclusie

In dit eerste hoofdstuk zijn een aantal literatuurgegevens besproken. In deze paragraaf zal geprobeerd worden om de notie auditieve samenhang aan de hand van de literatuurgegevens nader in te vullen.

De gegevens die we in de literatuur gevonden hebben, geven aanleiding om te veronderstellen dat auditieve samenhang het resultaat is van een drietal sorteerprocessen:

1. Het perceptief al of niet scheiden van gelijktijdige klanken,
2. het al of niet integreren van op elkaar volgende akoestische gebeurtenissen tot één doorgaande auditieve gebeurtenis,
3. het al of niet inpassen van doorgaande auditieve gebeurtenissen in de toonhoogtecontour.

De noties "akoestische gebeurtenis" en "doorgaande auditieve gebeurtenis" zullen later duidelijker worden.

Gelijktijdige klanken

Twee gelijktijdig klinkende stemmen of tonen worden in het algemeen gescheiden waargenomen, maar wanneer beide exact

dezelfde toonhoogte hebben, kunnen ze versmelten tot één perceptief geheel. Dit gebeurt ook als de grondtonen een geheel veelvoud van elkaar zijn (Stumpf, 1890). Zelfs wanneer op het linker en het rechter oor afzonderlijk klanken gelijktijdig aangeboden worden die harmonisch samenhangen, versmelten beide signalen perceptief tot een enkelvoudige auditieve gebeurtenis (Stumpf, 1890; Broadbent and Ladefoged, 1957).

Op grond van het al of niet aanwezig zijn van verschillen tussen de toonhoogtes van beide gelijktijdige klanken, kan de luisteraar besluiten of hij te maken heeft met één enkele klank of met meerdere gelijktijdige klanken. Als de luisteraar gelijktijdige klanken hoort met verschillende toonhoogtes, kan hij ze zeer waarschijnlijk scheiden in afzonderlijke klanken op grond van de harmonische structuur van beide klanken (vgl. Goldstein, 1973). Langs apparatuurlijke weg is reeds geprobeerd om op deze manier stemmen te scheiden (Everton, 1975; Parsons, 1976; Frazier, Samsam, Braida and Oppenheim, 1976).

Opeenvolgende klanken of klanken die van elkaar gescheiden zijn door korte pauzes

De volgende stap wordt voor de luisteraar het inpassen van de afzonderlijke auditieve gebeurtenissen in de stroom van geluiden waar de aandacht op gericht is.

Bij niet spraaksignalen, b.v. reeksen sinustonen, wordt in de literatuur gevonden dat de verschillende geluiden niet meer als bij elkaar horend worden waargenomen als er voldoende grote frequentieverschillen tussen die geluiden bestaan. De geluiden worden perceptief verdeeld over klassen van geluiden die ongeveer dezelfde frequentie hebben. Dit effect wordt auditieve splitsing genoemd.

Frequentieverglijdingen tussen opeenvolgende tonen blijken dit effect van auditieve splitsing tot op zekere hoogte te kunnen verhinderen. Daaruit volgt dat de continuïteit van frequentie een belangrijke factor kan zijn voor de beslissing welke akoestische gebeurtenissen perceptief bij elkaar horen.

Als het akoestisch signaal gedurende enige tijd niet of slechts geleidelijk verandert, wordt het signaal waargenomen als één doorgaande auditieve gebeurtenis.

Voor spraakklanken hebben we in de literatuur gezien dat spectrale continuïteit, (d.i. het op elkaar aansluiten van de formantstructuur) in reeksen klinkers de volgorde-waarneming gunstig beïnvloedt. We kunnen dit zo interpreteren dat de spectrale continuïteit tussen de opeenvolgende klinkers de samenhang bevordert. Elke klinker kan op grond van spectrale continuïteit gemakkelijk ingepast worden in de stroom van klinkers. Dit levert een doorgaande auditieve gebeurtenis op, die volledig samenhangend is. Een gemakkelijke herkenning en een goede waarneming van de volgordes van de klanken binnen de doorgaande auditieve gebeurtenis is het resultaat.

Als er tussen de opeenvolgende klanken geen spectrale continuïteit is, is de samenhang niet volledig en hebben we geen doorgaande auditieve gebeurtenis.

Als we kijken naar temporeel gesegmenteerde spraak, zien we dat bij relatief korte spraaksegmenten de verstaanbaarheid snel afneemt als de duur van de pauzes tussen de spraaksegmenten toeneemt. Deze afname van de verstaanbaarheid is duidelijk te zien bij pauzes tussen de 0 en 200 ms. De vermindering van de verstaanbaarheid bij het verkorten van de spraaksegmenten zou een gevolg kunnen zijn van een vermindering van de mogelijkheden om foneemopeenvolgingen binnen zo'n spraaksegment te herkennen.

Het akoestisch signaal is bij temporeel gesegmenteerde spraak helemaal compleet. De afname van de verstaanbaarheid bij het langer worden van de pauzes duidt op een verslechterde herkenbaarheid, die waarschijnlijk wordt veroorzaakt door een vermindering van de auditieve samenhang. Klaarblijkelijk zijn de pauzes de oorzaak dat akoestische gebeurtenissen minder goed ingepast kunnen worden in de stroom van klanken waar de aandacht op gericht is. Dit kan verklaard worden door aan te nemen dat de afbeelding van spectrale

informatie in een sensorisch geheugen zo snel vervaagt, dat na zo'n 100 à 200ms weinig informatie overblijft. Daardoor wordt het voor de luisteraar moeilijker om spectrale informatie van opeenvolgende akoestische gebeurtenissen met elkaar in verband te brengen. De beschikbare evidentie wijst erop dat het inpassen van kort op elkaar volgende auditieve gebeurtenissen in één doorgaande auditieve gebeurtenis vooral afhankelijk is van de spectrale continuïteit.

Klanken, gescheiden door lange pauzes

Als men luistert naar temporeel gesegmenteerde spraak, dan valt op dat de toonhoogtecontour (zinsmelodie) ondanks lange pauzes goed herkenbaar blijft, zodat de spraakfragmenten duidelijk tot één stroom van auditieve gebeurtenissen behoren.

Hieruit leiden we af dat de luisteraar nog een derde beslissing moet nemen voordat het akoestische signaal als auditief samenhangend waargenomen wordt. Dit is de beslissing of doorgaande auditieve gebeurtenissen al of niet in de toonhoogtecontour of de zinsmelodie passen.

De experimentele gegevens over temporeel gesegmenteerde spraak suggereren dat spectrale continuïteit werkzaam is over intervallen tot ongeveer 100 à 200 ms lengte. De toonhoogte lijkt een veel groter bereik in het tijddomein te hebben. Zelfs wanneer stille intervallen tussen opeenvolgende spraaksegmenten enige honderden ms lang zijn, is de toonhoogtecontour of de zinsmelodie nog goed te volgen. Met het oog op de verschillen in tijd waarover deze factoren werkzaam zijn, lijkt de luisteraar de behoefte te hebben aan twee soorten continuïteit:

- spectrale continuïteit binnen doorgaande auditieve gebeurtenissen en
- continuïteit van toonhoogte om de doorgaande auditieve gebeurtenissen in te kunnen passen in de toonhoogtecontour of de zinsmelodie.

Een term die bij de formulering van de bovenstaande hypothese enige malen is genoemd, is "doorgaande auditieve

gebeurtenis". Als direkt op elkaar volgende akoestische gebeurtenissen in de waarneming met elkaar verbonden worden, zodanig dat zij perceptief één geheel vormen, dan hebben we te maken met een doorgaande auditieve gebeurtenis. Het woord "nieuwe" vormt bijvoorbeeld zo'n doorgaande auditieve gebeurtenis.

Een niet doorgaande auditieve gebeurtenis vormt het woord "bieten". Bij de /t/ treedt spectrale discontinuïteit en een stil interval van 80 tot 120 ms op. Huggins' resultaten met temporeel gesegmenteerde spraak laten zien dat bij pauzes van een dergelijke duur de verstaanbaarheid sterk terug zou moeten lopen. Toch wordt dit woord perceptief als één samenhangend geheel waargenomen. Wij suggereren dat in zulke gevallen de continuïteit van de zinsmelodie over de stille intervallen heen de luisteraar helpt de spraakstroom te volgen.

Bij doorgaande auditieve gebeurtenissen is de spectrale continuïteit, en over opeenvolgende doorgaande auditieve gebeurtenissen heen is de continuïteit van de toonhoogte de belangrijkste factor voor de auditieve samenhang.

De hier geschetste gedachtengang vormt het uitgangspunt voor een serie experimenten over de relatie tussen akoestische eigenschappen van spraakgeluiden en de auditieve samenhang van de klankstroom.

Deze experimenten zijn beschreven in de hoofdstukken 2 en 3 van deze studie. De experimenten beschreven in hoofdstuk 2 proberen een antwoord te geven op de vraag wat het relatief belang is van spectrale continuïteit en van de continuïteit in het toonhoogteverloop voor de waargenomen samenhang van reeksen spraakachtige geluiden en wat het tijdsinterval is waarover discontinuïteiten een belemmerend effect hebben op de waargenomen samenhang.

In hoofdstuk 3 is een aantal experimenten beschreven waarin onderzocht wordt hoe de perceptieve scheiding van gelijktijdige stemmen afhangt van het verschil in toonhoogteligging en van het verloop van de toonhoogtecontour. Een meer gedetailleerde operationalisering van de onderzoeksvragen wordt in de hoofdstukken 2 en 3 gegeven. Hoofdstuk 4

geeft een samenvattende interpretatie en een kritische evaluatie van de experimentele resultaten.

2. AUDITIEVE SAMENHANG VAN REEKSEN SPRAAKSEGMENTEN MET ALTERNERENDE TOONHOOGTE

2.1 Inleiding

In de praktijk is auditieve samenhang van het spraaksig-
naal voor een luisteraar belangrijk wanneer hij luistert
naar verstoorde spraak. Dit kan zijn: onderbroken spraak,
spraak gedeeltelijk gemaskeerd door veel luidere signalen, of
spraak die gehinderd wordt door andere spraaksignalen
(cocktail-party-probleem).

Om tot spraakperceptie te kunnen komen moet de luisteraar
in het ene geval in staat zijn om aparte stukjes seriëel
geordende akoestische informatie met elkaar in verband te
brengen. In het andere geval is de luisteraar genooddaakt
om het complexe geluidssignaal te ontrafelen in meerdere
parallele stromen akoestische informatie. Auditieve samen-
hang kan de luisteraar hierbij hulp bieden.

Hoe kan de auditieve samenhang van het spraaksignaal meet-
baar gemaakt worden, en wat zijn nu de fysische eigenschap-
pen van het spraaksignaal die verantwoordelijk zijn voor
de auditieve samenhang?

De reactie van een luisteraar op een gegeven stimulus is
de enige manier om iets te weten te komen over auditieve
samenhang. In het vorige hoofdstuk hebben wij gezien dat
wat een luisteraar waarneemt niet uitsluitend afhangt van
het akoestische signaal dat hem aangeboden wordt, maar ook
van contextfactoren.

Daarmee moet rekening worden gehouden bij het kiezen van
een methode om auditieve samenhang meetbaar te maken. Het
reeds genoemde voorbeeld van geassembleerde spraak zou
een ingang kunnen geven tot het bepalen van de akoestische
factoren die invloed hebben op de auditieve samenhang.

Als het assembleren van op zich goed verstaanbare woorden
tot een betekenisvolle spraakuiting het gevolg heeft dat
de resulterende spraakuiting auditief niet samenhangend is,
dan kunnen daar potentiëel vier factoren debet aan zijn,

te weten discontinu amplitudeverloop, discontinu toonhoogteverloop, discontinue formantstructuur en afwijkende temporele organisatie van spraakuitingen.

De in het vorige hoofdstuk besproken literatuurgegevens geven aanleiding te veronderstellen dat vooral de verstoring van het toonhoogteverloop en van de spectrale continuïteit van belang zijn.

2.2 Exploratieve Metingen

Voor een experimentele benadering van de gestelde vragen, zal gezocht moeten worden naar een opzet die voldoet aan de volgende eisen:

- verschillende fysische eigenschappen van het spraaksignaal moeten te variëren zijn,
- contextfactoren moeten niet als ongecontroleerde variabelen de resultaten beïnvloeden,
- de taak van de luisteraar moet niet moeilijk zijn.

Bij het zoeken naar een meetmethode leek het verstandig om te beginnen met stimuli die al auditief niet samenhangend zijn. Daarvoor hebben we in eerste aanleg stimuli genomen die verkregen zijn door assemblage van geïsoleerd ingesproken woorden.

Geassembleerde nonsens zinnen

Op het Instituut voor Perceptie Onderzoek is een systeem ontwikkeld waarmee op vrij snelle wijze eenvoudig woorden uitgepoort, gedigitaliseerd opgeslagen en weer uitgevoerd kunnen worden (Willems and de Jong, 1974). Dit systeem is o.a. geschikt om woorden te assembleren tot grotere spraakuitingen. Uitgaande van een klein aantal één-lettergrepige woorden (15 items) zijn met behulp van dit systeem een aantal nonsens zinnengeassembleerd in de geest van: "de non nam de leeuw mee naar zee". Deze zinnen hebben een correcte syntactische structuur, maar zijn qua betekenis erg ongebruikelijk. Door de spreker bij het inlezen van de woorden opdracht te geven deze monotoon en niet-monotoon en al of niet met een hoge of lage toonhoogte in te spreken, is het op vrij eenvoudige wijze mogelijk om binnen zekere

grenzen de toonhoogte van elk individueel woord te variëren. Door nu alleen monotone of alleen niet-monotone woorden te assembleren, kunnen de toonhoogtediscontinuïteiten op de woordgrenzen grofweg klein of groot gemaakt worden. De luisteraar kreeg als opdracht de aangeboden stimuli na te zeggen. De verstaanbaarheid van de spraakuitingen zou een bruikbare maat kunnen opleveren voor het effect van de toonhoogtediscontinuïteiten op de auditieve samenhang. Als meetmethode bleek de gevolgde procedure echter weinig geschikt. In eerste instantie waren de zinnen voor de luisteraars volstrekt onverstaanbaar, maar toen de woorden eenmaal na aanbieding van een aantal van dergelijke geassembleerde zinnen bekend waren, vormden nieuwe stimuli geen enkel probleem meer voor de luisteraars. Dit werd o.a. veroorzaakt door het feit dat het aantal gebruikte woorden zeer beperkt was.

Voor het nauwkeurig manipuleren van fysische eigenschappen van het spraaksignaal is het gebruikte systeem eigenlijk weinig geschikt. Als men de toonhoogte, de intensiteit of de temporele opbouw onafhankelijk van elkaar en goed gedefiniëerd zou willen variëren, dan lijkt het gebruik van een vocodersysteem voor de hand te liggen.

Uit pilootexperimenten bleek echter dat de vocoder die bij de aanvang van dit onderzoek beschikbaar was, spraak leverde van onvoldoende kwaliteit voor de experimenten die wij ons voorgenomen hadden te doen. (De LPC vocoder die gebruikt is voor de experimenten beschreven in paragraaf 3.3, kwam pas veel later ter beschikking). Om deze reden is besloten om gebruik te blijven maken van het reeds besproken systeem om woorden uit te poorten en te assembleren tot grotere spraakuitingen.

Geassembleerde reeksen betekenisloze syllaben

Stimulusmateriaal dat bestaat uit reeksen betekenisloze lettergrepen, biedt voldoende zekerheid dat linguïstische factoren de metingen niet verstoren.

Als men een reeks lettergrepen aanbiedt die afwisselend een hoge en een lage toonhoogte hebben en gescheiden worden door pauzes met identieke duren, dan valt zo'n reeks bij heel korte pauzes voor de hoorder uiteen in twee subreeksen die ieder door een verschillende spreker gesproken of gezongen lijken te zijn, terwijl bij relatief lange pauzes de hele reeks wordt gehoord als voortgebracht door één spreker. Een dergelijke stimulus met alternerende toonhoogte lijkt een interessant uitgangspunt te zijn voor het verdere onderzoek. Bij deze stimuli is het mogelijk de bijdrage van de verschillende prosodische factoren aan de auditieve samenhang uit te drukken in dezelfde eenheid, namelijk de pauzeduur waarbij de luisteraar in 50% van de gevallen één spreker en in 50% van de gevallen twee sprekers meent te horen. Ook is het criterium voor samenhangend/niet samenhangend, het waarnemen van één of twee sprekers niet te moeilijk en nauw verwant met auditieve samenhang.

Bij dit stimulusmateriaal met alternerende toonhoogte, treden verschijnselen op die goed aansluiten bij de literatuur over volgorde waarneming en auditieve splitsing. In het geval dat de auditieve samenhang van de stimulus door de aanwezigheid van een alternerend toonhoogtepatroon zodanig verstoord is dat een luisteraar twee sprekers in plaats van één spreker waarneemt, kan dit opgevat worden als een gevolg van auditieve splitsing.

Op grond van de toonhoogte worden de stimuli perceptief gegroepeerd in twee afzonderlijke klassen geluiden die elk aan één spreker worden toegekend. Binnen zo'n stroom blijft de waarneming van volgordes intact, terwijl de volgorde waarneming van akoestische gebeurtenissen die behoren tot twee verschillende perceptieve reeksen erg moeilijk, zo niet onmogelijk is.

Op grond van de beschreven exploratieve experimenten zijn enkele ervaringen verkregen zoals:

- effecten van contextafhankelijke factoren kunnen voldoende geëlimineerd worden door gebruik te maken van betekenisloze spraakuitingen,
- stimuli met alternerend toonhoogteverloop geven goed definieerbare en eenvoudig te gebruiken criteria voor het al of niet samenhangend zijn van het akoestisch signaal,
- de mate van samenhang kan uitgedrukt worden in de pauzeduur waarbij de luisteraars in 50% van de gevallen één spreker en in 50% van de gevallen twee sprekers menen te horen.

Deze ervaringen maken het mogelijk om tot nieuwe, nauwkeurigere gespecificeerde vragen te komen.

In hoeverre wordt de auditieve samenhang van het akoestische signaal verstoord door de aanwezigheid van sprongen in de toonhoogte?

Wat is het effect van temporele organisatie zoals pauzen en klinkerduren van de stimuli op de auditieve samenhang?

Wat is het belang van spectrale overgangen voor de auditieve samenhang van stimuli met alternerende toonhoogte?

De relatie tussen amplitude en auditieve samenhang leek in dit stadium van het onderzoek niet interessant te zijn en is niet nader onderzocht.

2.3 Klinkerreksen en Auditieve Samenhang

De eerste vraag is in hoeverre de auditieve samenhang van het akoestische signaal verstoord wordt door de aanwezigheid van discontinuïteiten in de toonhoogte. Om hierop een antwoord te krijgen is een experiment gedaan waarbij reeksen van negen geassembleerde klinkers gebruikt zijn. Gekozen is voor een opzet waarbij de toonhoogte in de stimulus van klinker tot klinker afwisselend hoog en laag is en waarbij de toonhoogteverschillen en de duren van de stille intervallen tussen de klinkers de experimentele variabelen zijn. De mate van samenhang zou dan, gegeven een zeker toonhoogteverschil tussen de klinkers, uitgedrukt kunnen worden in de maximale duur van de stille intervallen waarbij de luisteraar nog juist het akoestische signaal als niet samenhangend beoordeelt.

Om spectraal gezien zo groot mogelijke verschillen te krijgen zijn de klinkers /a/, /u/ en /i/ gekozen, te weten de hoekpunten van de klinkerdriehoek voor het Nederlands.

De klinkers in de geassembleerde reeksen hebben steeds een duur van 100 ms. De reeksen van 9 klinkers zijn steeds geassembleerd in dezelfde volgorde: /a, u, i, a, u, i, a, u, i/ met een afwisselend hoge en lage toonhoogte;

Hoge toonhoogte / u a i u /,

lage toonhoogte /a i u a i/.

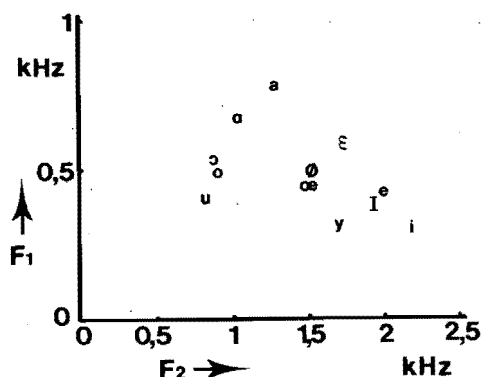


Fig. 2.1 Klinker-driehoek voor het Nederlands

Deze structuur van de stimulus levert als voordeel op dat in het geval van auditieve splitsing een andere volgorde waargenomen wordt dan in het geval van auditieve samenhang. Bij een auditief samenhangende reeks wordt de stimulus als a-u-i-a-u-i-a-u-i waargenomen, terwijl bij een niet samenhangende reeks óf a-i-u-a-i óf u-a-i-u waargenomen wordt.

Het toonhoogteverschil tussen de opeenvolgende klinkers kan per reeks van 9 klinkers de volgende waarden aannemen:

0, 2.9, 5.1, 7.8, 10.3, 13.2, 14.0, 15.4 of 17.8 halve tonen.

Deze getallen vormen een niet regelmatige reeks die ontstaan is doordat bij de synthese van deze klinkers het aantal posities voor de toonhoogte beperkt was.

De dueren van de stille intervallen tussen de klinkers zijn

binnen een reeks steeds identiek en kunnen waarden hebben van 0, 25, 50, 75, 100, 150, 200, 300 en 400 ms.

Een van de eisen die aan de stimulus gesteld zijn, is dat er een mogelijkheid moet bestaan om bepaalde fysische eigenschappen zoals toonhoogte, amplitude en spectrale samenstelling nauwkeurig te specificeren en zo mogelijk onafhankelijk te variëren. Het ligt voor de hand om evenals anderen dat gedaan hebben met gesynthetiseerde stimuli te werken. Dan zijn toonhoogte, intensiteit en spectrale samenstelling goed specificeerbaar, en kan men binnen een gegeven duur van de klinkers spectrale overgangen maken.

De synthese van de klinkers is gebeurd met een software synthetisator. Deze synthetisator bestaat uit een computerprogramma voor het uitrekenen van golfvormen en een 10 bits digitaal-analoog omzetter die de golfvormen hoorbaar maakt. In het programma wordt aan een digitale inputpuls met behulp van een filter een spectrum gegeven dat ongeveer 6 dB per octaaf afloopt. Vijf in serie geschakelde bandfilters simuleren de overdrachtsfunctie van het menselijk spraakkanaal. De samplefrequentie van dit systeem is 10 kHz. Om kwantiseringsruis te verminderen wordt het uitgegeven signaal gefilterd met een 36 dB/oct. Allison laagdoorlaatfilter dat ingesteld staat op 4500 Hz.

Voor de vijf bandfilters dienen opgegeven te worden: de frequenties, de bandbreedtes, het moment waarop zij moeten veranderen en de snelheid waarmee een nieuwe doelpositie moet worden bereikt. Bij verandering van de doelposities van de filters verloopt de frequentie van de ene doelpositie naar de andere doelpositie volgens een e-macht.

Aan de hand van de volgende specificaties zijn de klinkers met deze synthetisator gesynthetiseerd. De bandbreedtes zijn min of meer arbitrair. De vierde en de vijfde formant zijn constant gehouden op respectievelijk 3500 en 4500 Hz met bandbreedtes van 200 en 300 Hz en dienen uitsluitend om de klanken een natuurlijk karakter te geven.

klinker	formanten (Hz)	bandbreedtes (Hz)
/a/	F ₁ 800	60
	F ₂ 1250	100
	F ₃ 2500	150
/u/	F ₁ 320	60
	F ₂ 660	100
	F ₃ 2660	150
/i/	F ₁ 220	60
	F ₂ 2300	100
	F ₃ 3100	150

De 81 verschillende reeksen klinkers (9 toonhoogteverschillen, 9 pauzeduren) zijn in toevalsvolgorde ieder driemaal met een onderling tijdsverschil van 2 seconden op een stimulusband opgenomen.

Twee luisteraars, waaronder de auteur, kregen de opdracht driemaal en in het geval dat er moeilijkheden waren, vier maal de band op een luidheidsniveau van 55 dB SL af te luisteren met behulp van een koptelefoon. Zij hadden de gelegenheid om op elk willekeurig moment de band te stoppen, maar bij de instructie was afgeraden dit vaak te doen. De opdracht was te noteren of zij één of twee sprekers waarnamen. Dit onderscheid moest gemaakt worden door erop te letten of de stimulus als één reeks /a, u, i, a, u, i, a, u, i/ of als twee deelreeksen /a, i, u, a, i/ en /u, a, i, u/ waargenomen werd.

De luisteraar was geen instructie gegeven te trachten zo lang mogelijk één spreker of zo lang mogelijk twee sprekers waar te nemen. Ook bestaat er geen inzicht over de beïnvloeding van de luisterinstelling door voorafgaande stimuli.

Reeksen klinkers zonder toonhoogteverschillen en met een stille tijd tussen de klinkers van minder dan 50 ms zijn perceptief moeilijk te volgen. Men hoort één enkele stroom spraakachtige klanken, maar het blijkt moeilijk om deze geluiden te benoemen. Voor intervalduren die groter zijn dan 50 ms hoort men duidelijk één reeks klinkers, die uit-

sluitend aan één spreker toegeschreven kan worden.

Wanneer van klinker tot klinker toonhoogteverschillen van voldoende grootte aangebracht worden, dan hoort men afhankelijk van de intervalduur twee perceptief gescheiden reeksen klinkers, elk gezongen door een verschillende spreker. Wanneer men twee reeksen klinkers waarneemt, dan kan men naar believen de aandacht richten op de reeks klinkers met een lage toonhoogte /a-i-u-a-i/ of op de reeks klinkers met een hoge toonhoogte /u-a-i-u/.

Tussen de gebieden waarin men uitsluitend één of twee sprekers waarneemt, ligt een gebied waarin men afhankelijk van de gerichtheid van de luisteraar één of twee sprekers kan waarnemen.

Slechts twee luisteraars, waaronder de auteur, zijn bereid gevonden zich te onderwerpen aan het langdurige en het vermoeiende luisteren naar deze klinkerreksen. Voor elke aangeboden reeks klinkers hebben zij aangegeven of zij één reeks klinkers (één spreker) of twee reeksen klinkers (twee sprekers) waarnamen.

Tussen de resultaten van de twee verschillende luisteraars waren geen systematische verschillen. De gecombineerde resultaten staan weergegeven in fig. 2.2.

De experimentele variabelen zijn de duren van de stille intervallen en de toonhoogteverschillen tussen de opeenvolgende klinkers.

De gemeten variabele is de ligging van de omslagpunten, gedefiniëerd als die combinaties van toonhoogteverschillen en pauzeduren, waarbij de luisteraars in 50% van de gevallen één spreker en in 50% van de gevallen twee sprekers menen te horen.

Deze 50% punten zijn als volgt berekend. Bij de klinkerreksen is voor elke combinatie van toonhoogtesprongen en intervalduren de mate van de waargenomen samenhang bepaald aan de hand van de oordelen van de luisteraars. Vervolgens is aangenomen dat de mate van de waargenomen samenhang zowel horizontaal (constante toonhoogteverschillen, variabele

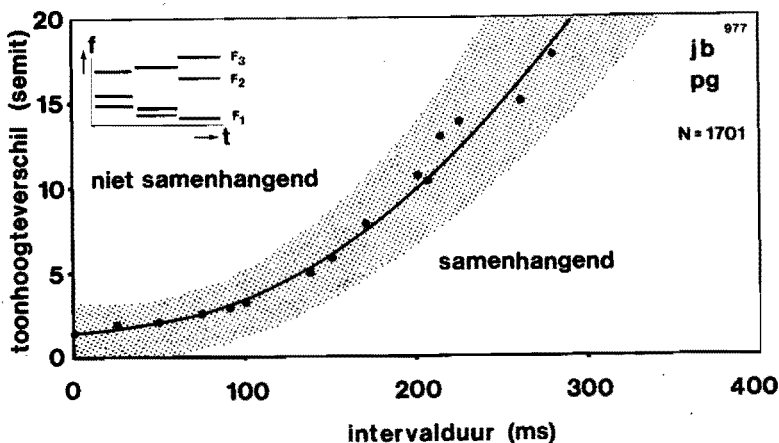


Fig. 2.2 Berekende 50% omslagpunten voor het horen van één of twee stromen klinkers in reeksen van 9 homogene klinkers met alternerende hoge en lage toonhoogte. Het gearceerde gebied geeft de standaardbreedtes aan van de cumulatieve normale verdelingen die de overgang van volledig samenhangend tot volledig niet samenhangend beschrijven. De resultaten zijn gemiddeld over twee luisteraars.

intervalduren) als vertikaal (constante intervalduren, variabele toonhoogteverschillen) volgens een cumulatieve normale verdeling verloopt van volledig samenhangend tot volledig niet samenhangend. Vervolgens zijn door de meetgegevens in horizontale en verticale richting cumulatieve normale verdelingen geschat, waarvan de 50% punten en de standaardbreedtes in fig. 2.2 zijn weergegeven.

Voor de perceptie van verbonden spraak zijn intervallen tot 150 ms wellicht het belangrijkste. De stille intervallen tussen de klinkers zijn min of meer vergelijkbaar met stemloze consonanten of consonantklusters. Deze zijn zelden langer dan 150 ms. Uit de gepresenteerde gegevens blijkt dat bij deze duren een relatief kleine toonhoogtesprong van 2 à 5 halve tonen reeds voldoende is om de auditieve samenhang van het signaal dermate te verstoren dat men in plaats van één spreker twee sprekers hoort.

Een van de vragen die nu naar boven komt is, wat het effect is van de temporele organisatie van de stimulus. Door de pauzeduren te variëren bij constante klinkerduren hebben we tegelijkertijd met die pauzeduren ook de inzet-tot-inzettijden gevariëerd. Daarom laten deze resultaten niet toe te beslissen wat de voornaamste temporele factor is: de pauzeduren of de inzet-tot-inzettijden. Ook is het mogelijk dat de klinkerduren, die in dit experiment niet gevariëerd zijn, de auditieve samenhang mede bepalen. Om meer inzicht te krijgen in het relatieve belang van deze verschillende temporele factoren is een experiment gedaan, dat beschreven wordt in de volgende paragraaf.

2.4 Klinkerduren, Pauzeduren en Inzet-tot-inzettijden

Wat is in metingen zoals beschreven in paragraaf 2.3 de belangrijkste temporele factor: de pauzeduur, de inzet-tot-inzettijd of de klinkerduur?

Om deze vraag te beantwoorden hebben we een stimulusband geprepareerd met evenals in het vorige experiment reeksen van 9 klinkers, weer in dezelfde volgorde a, i, u, a, i, u, a, i, u, maar nu met verschillende temporele condities. Een constante klinkerduur van 100 ms werd gecombineerd met pauzeduren van 50, 100 en 150 ms en een constante inzet-tot-inzettijd van 250 ms werd verkregen met klinkerduren van 100, 150 en 200 ms te combineren met pauzeduren van respectievelijk 150, 100 en 50 ms.

Alle zo verkregen temporele condities werden factoriëel gecombineerd met 9 toonhoogteverschillen tussen de alternerende toonhoogtes, op dezelfde wijze als in paragraaf 2.3 beschreven.

Ieder stimulustype dat zo ontstaat, komt driemaal op de band voor. De luisteraars zijn dezelfde als uit paragraaf 2.3 en krijgen via een hoofdtelefoon deze band viermaal te horen met een luidheidsniveau van 55 dB SL. De opdracht en de instructie zijn hetzelfde als in paragraaf 2.3.

De resultaten zijn weergegeven in fig. 2.3. Hierin staan als functie van de in de stimuli voorkomende toonhoogte-

sprongen en van de intervalduren, de geschatte 50% omslagpunten voor het waarnemen van één spreker of van twee sprekers voor de stimuli met constante inzet-tot-inzettijden en voor de stimuli met dezelfde klinkerduur.

Evenals in de vorige paragraaf zijn de 50% omslagpunten berekend door aan te nemen dat in verticale richting (constante intervalduur, variabel toonhoogteverschil) de mate van de waargenomen samenhang volgens een cumulatieve normale verdeling verloopt van volledig samenhangend tot volledig niet samenhangend. In de figuur zijn de standaardbreedtes van deze verdelingen aangegeven.

* constante inzet-tot-inzet tijd : 250ms

o constante klinkerduur : 100 ms

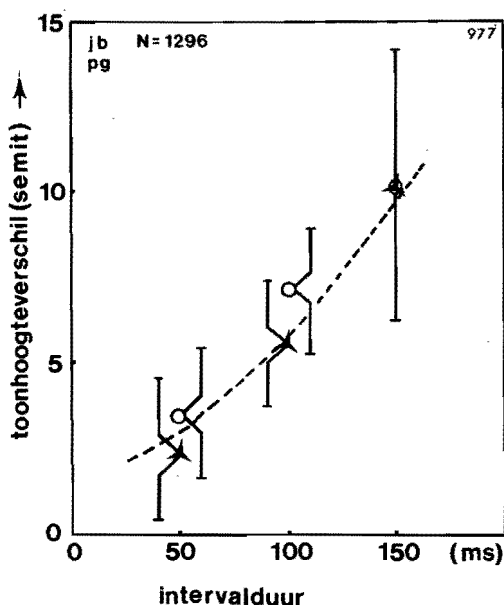


Fig. 2.3 50% Omslagpunten en standaardbreedtes van de cumulatieve normale verdelingen voor het horen van één of twee stromen klinkers in reeksen van 9 klinkers met alternerend hoge en lage toonhoogte. De stille intervallen tussen de klinkers zijn 50, 100 of 150 ms. In één serie zijn de inzet-tot-inzettijden van de ene klinker tot de andere constant 250 ms gehouden. De bijbehorende klinkerduren zijn 200, 150 of 100 ms. In de andere serie zijn de klinkerduren constant gehouden op 100 ms en de inzet-tot-inzettijden zijn dus 150, 200 en 250 ms. De resultaten zijn gemiddeld over twee luisteraars.

De resultaten die weergegeven staan in figuur 2.3 laten zien dat er geen duidelijke verschillen bestaan in de effecten van stimuli met constante klinkerduren en stimuli met constante inzet-tot-inzettijden. Daarom mogen we zeggen dat de duur van de klinkers weinig effect heeft op de waarneming van één of twee sprekers, tenminste voor klinkerduren tussen de 100 en de 200 ms. De duur van de pauzes tussen de klinkers is de belangrijkste temporele factor die samen met de grootte van de toonhoogteverschillen in hoge mate de auditieve samenhang bepaalt.

2.5 Formantbuigingen en Auditieve Samenhang

We zullen nu enkele metingen beschrijven waarin geprobeerd is de bijdrage van formantbuigingen aan de auditieve samenhang te bepalen.

In paragraaf 1.2.1 is op grond van de resultaten van Dorman et al. (1975) gesuggereerd dat de spectrale continuïteit van spraak van belang is voor de correcte waarneming van volgordes. Indien klinkers met elkaar verbonden worden door /b/-achtige formantovergangen of door formantverglidingen van de ene klinker naar de andere, blijkt dat de waarneming van volgordes in het algemeen beter is dan bij reeksen met uitsluitend homogene klinkers. Deze experimentele gegevens suggereren dat de aanwezigheid van formantovergangen van de ene klinker naar de andere de auditieve samenhang van het signaal bevordert.

Als men dit gegeven plaatst in het kader van de reeds besproken experimenten met klinkerreeksen, waarin de toonhoogte van klinker tot klinker alterneerde, dan rijst de vraag of medeklinkerachtige formantbuigingen van de ene klinker naar de andere het horen van twee sprekers kunnen beïnvloeden of zelfs tegengaan.

Als meetmethode hebben we ook hier voor de methode uit de vorige experimenten gekozen. Luisteraars krijgen weer reeksen klinkers aangeboden in de volgorde /a-u-i-a-u-i-a-u-i/. In elke reeks is de toonhoogte van klinker tot klinker afwisselend hoog en laag. Ook de opdracht aan de luisteraars

is weer dezelfde. Elke stimulus moet op samenhang beoordeeld worden aan de hand van het criterium of de luisteraars één of twee sprekers waarnemen.

De duur van de stille intervallen tussen de klinkers kan de volgende waardes bedragen: 0, 25, 50, 75, 100, 150, 200, 300 en 400 ms.

In plaats van homogene klinkers zijn lettergrepen gebruikt met

- /b/-achtige formantbuigingen aan het begin van de klinkers
- /b/-achtige formantbuigingen aan het begin en aan het einde van de klinkers
- /b/-achtige formantbuigingen aan het begin van elke klinker, gecombineerd met niet gebruikelijke spectrale buigingen aan het eind.

Deze stimuli zijn ook weer gesynthetiseerd met de software synthetisator. De synthesegegevens voor de homogene gedeelten van de klinkers zijn hetzelfde als die in paragraaf 2.3. De dueren van de klinkers zijn, inclusief de formantbuigingen 100 ms. De formantbuigingen aan het begin en het einde van de klinkers worden gerealiseerd in 25 ms. De /b/-achtige formantbuigingen worden verkregen door de formanten aan het begin en het einde van de klinkers te laten verglijden naar de volgende frequentiewaarden:

F_1 : 50 (Hz)

F_2 : 600 (Hz)

F_3 : 2100 (Hz)

De anomale formantbuigingen die aan het eind van de klinkers uit de laatste conditie zijn gerealiseerd, zijn zodanig gemaakt dat de formanten niet naar lagere, maar naar hogere frequenties verglijden. Deze formantbuigingen komen niet overeen met wat binnen de mogelijkheden van de menselijke spraakvorming ligt.

Beoogd is om met de /b/-/klinker/-/b/ combinaties spectrale continuïteit te verkrijgen op een wijze die aansluit bij de normale afwisseling tussen klinkers en medeklinkers in lopende spraak.

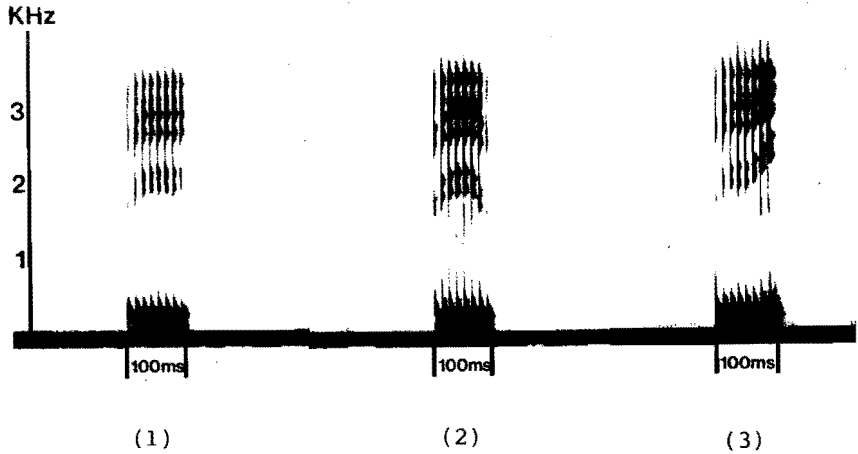


Fig. 2.4 Spectrografische weergave van een specimen van de gebruikte syllaben met de klinker /i/.
 (1) /b-/klinker/combinatie, (2) /b-/klinker-/b/combinatie, (3) /b-/klinker/combinatie met niet gebruikelijke spectrale buigingen aan het einde. Als functie van frequentie en tijd is de intensiteit aangegeven door de mate van zwarting.

Bij de /b-/klinker/combinaties is de spectrale continuïteit minder, d.w.z. de formanten sluiten minder goed op elkaar aan, maar de syllabeachtige structuur blijft gehandhaafd. De reeksen homogene klinkers leveren duidelijke spectrale discontinuïteiten aan de klinkergrenzen op. De spectrale continuïteit wordt tenslotte optimaal verstoord door de realisering van niet gebruikelijke formantbuigingen aan het einde van elke klinker, te laten volgen door /b/-achtige buigingen aan het begin van elke volgende klinker. De toonhoogteverschillen die binnen de reeksen gemaakt zijn kunnen de volgende waardes aannemen:

- voor de /b-/klinker/combinaties:

0, 2.8, 4.9, 7.7, 10.1, 12.9, 13.8, 15.0 en 17.5 halve tonen

- voor de /b/-/klinker/-/b/combinaties:

0, 2.4, 4.9, 7.3, 10.1, 12.9, 13.6, 15.2 en 17.9 halve tonen

- voor de /b/-/klinker/combinaties met anomale eindbuigingen

0, 2.4, 4.9, 7.5, 10.0, 12.7, 13.6, 15.2 en 17.5 halve tonen.

Deze waarden zijn bepaald door het beperkte aantal posities voor de toonhoogte bij de synthese.

Voor elk type lettergreep is een afzonderlijke stimulusband gemaakt. Elke stimulusband bevat 81 verschillende reeksen (9 toonhoogteverschillen, 9 pauzeduren). Deze reeksen zijn in toevalsvolgorde met een onderling tijdsverschil van 2 seconden op de stimulusband opgenomen. Elke conditie komt driemaal op de band voor.

De twee luisteraars, die ook bereid waren geweest de stimulusband met homogene klinkers uit paragraaf 2.3 te beluisteren, hadden de opdracht deze stimulusbanden driemaal en in geval van moeilijkheden vier maal op een luidheidsniveau van 55 dB SL af te luisteren en aan te geven of één of twee sprekers waargenomen werden.

De resultaten hiervan staan weergegeven in de figuren 2.5, 2.6 en 2.7. De geschatte 50% omslagpunten voor het horen van één of van twee sprekers staan weergegeven als functie van de toonhoogteverschillen en de duur van de stille intervallen tussen de lettergrepen.

Evenals in paragraaf 2.3 zijn deze punten berekend door aan te nemen dat de mate van waargenomen samenhang zowel in horizontale richting (constante toonhoogteverschillen, variabele intervalduren) als in verticale richting (constante intervalduren, variabele toonhoogteverschillen) volgens een cumulatieve normale verdeling verloopt van volledig samenhangend tot volledig niet samenhangend.

In fig. 2.8 zijn de geschatte 50% punten van deze drie soorten lettergrepen tesamen met die voor homogene klinkers (zie par. 2.3) weergegeven.

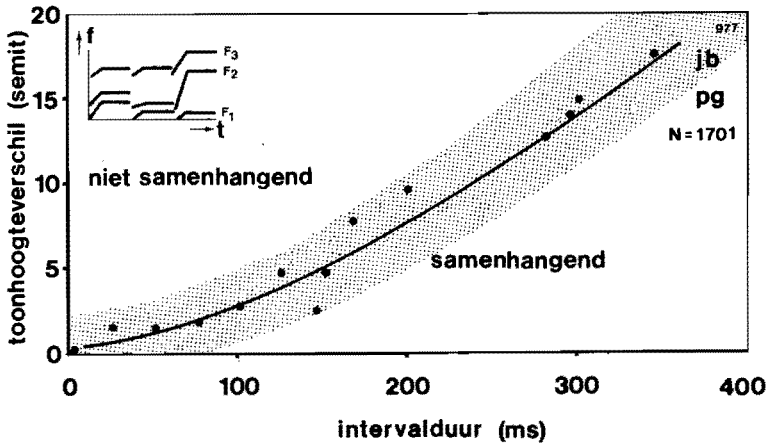


Fig. 2.5 Berekende 50% omslagpunten en standaardbreedtes van de cumulatieve normale verdelingen die het horen van één of twee stromen lettergrepen in reeksen van 9 /b-/klinker/combinaties met alternerend hoge en lage toonhoogte beschrijven. De resultaten zijn gemiddeld over twee luisteraars.

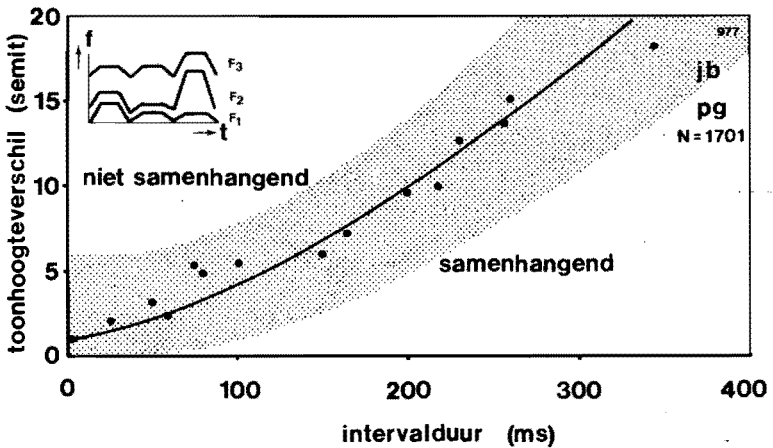


Fig. 2.6 Berekende 50% omslagpunten en standaardbreedtes van de cumulatieve normale verdelingen die het horen van één of twee stromen lettergrepen in reeksen van 9 /b-/klinker-/b/combinaties met alternerend hoge en lage toonhoogte beschrijven. De resultaten zijn gemiddeld over twee luisteraars.

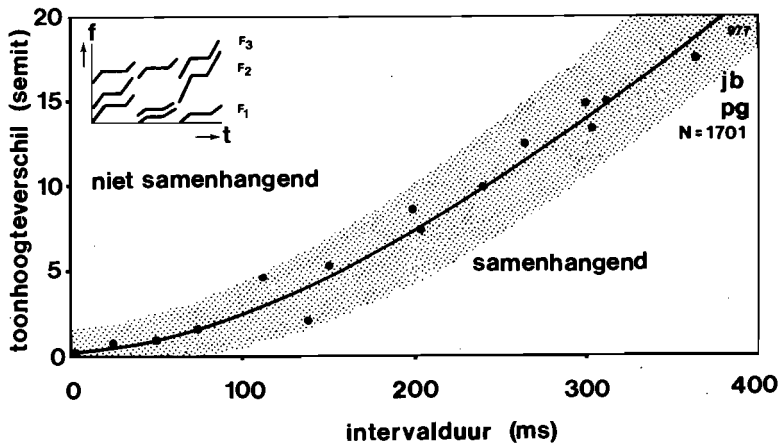


Fig. 2.7 Berekende 50% omslagpunten en standaardbreedtes van de cumulatieve normale verdelingen die het horen van één of twee stromen lettergrepen in reeksen van 9 /b/-/klinker/combinaties met niet gebruikelijke spectrale eindbuigingen beschrijven. De resultaten zijn gemiddeld over twee luisteraars.

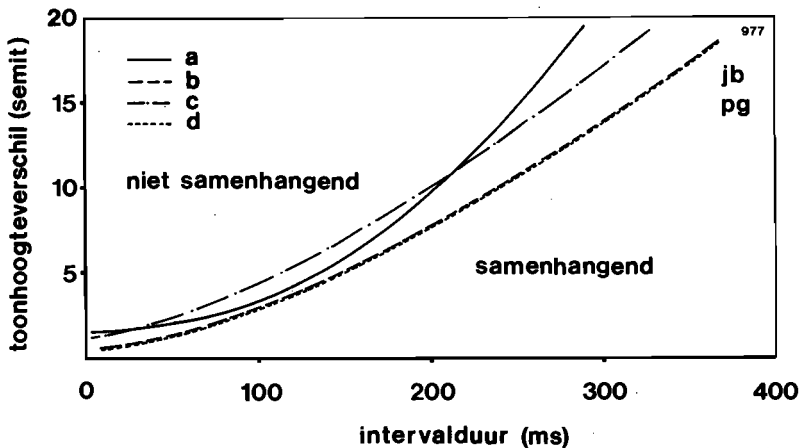


Fig. 2.8 Berekende 50% omslagpunten voor het horen van één of twee stromen klinkers of lettergrepen in reeksen van klanken met alternerend hoge en lage toonhoogte.

- (a) homogene klinkers
- (b) /b/-/klinker/combinaties
- (c) /b/-/klinker-/b/combinaties
- (d) /b/-/klinker/combinaties met niet gebruikelijke spectrale eindbuigingen.

De resultaten zijn gemiddeld over twee luisteraars.

De vier verschillende krommen zijn nauwelijks verschillend van elkaar. De resultaten van deze metingen laten zien dat formantbuigingen van het type zoals hier gebruikt is, weinig of geen effect hebben op het horen van één of twee sprekers. De geringe verschillen die er mogelijk toch nog bestaan tussen de condities a, b, c en d in fig. 2.8 zijn lastig te interpreteren omdat deze condities niet binnen één luistersessie gebracht konden worden. Tengevolge van het grote aantal luisteruren dat zelfs per conditie nodig was voor ieder van de luisteraars, kwamen er soms lange tijden te liggen tussen het opmeten van de verschillende krommen in fig. 2.8.

Om nu te onderzoeken of er wellicht toch een consistent secundair effect van de formantbuigingen op het horen van twee sprekers bestaat, is nog een andere luisterproef gedaan. Hierin werden de vier tot nu toe gebruikte typen formantovergangen binnen één stimulusband gebracht, terwijl het totale aantal stimuli beperkt werd gehouden door maar één pauzeduur toe te laten. Daarvoor werd een pauzeduur van 75 ms gekozen, omdat in verbonden spraak deze waarde redelijk overeenkomt met de duren van medeklinkers die tussen klinkers instaan.

De vier condities waren:

- 1: homogene klinkers
- 2: /b-/klinker/combinaties
- 3: /b-/klinker-/b/combinaties
- 4: /b-/klinker/combinaties met niet gebruikelijke spectrale buigingen aan het einde.

De stimuli zijn identiek aan die van de vorige experimenten. Op de stimulusband komen de vier condities, elk met negen mogelijke toonhoogtesprongen, drie keer in toevalsvolgorde voor. Elke luisteraar heeft de opdracht de band vier maal te beluisteren op een luidheidsniveau van 55 dB SL en moet aangeven of hij één of twee sprekers waarneemt. Er zijn vier luisteraars bereid gevonden om deze band te beluisteren en hun oordeel uit te brengen. Twee van deze luisteraars hadden ook meegedaan aan de vorige experimenten.

Evenals in paragraaf 2.3 zijn 50% omslagpunten voor het horen van één of twee stromen klinkers of lettergrepen berekend door aan te nemen dat in verticale richting (variabele toonhoogteverschillen) de mate van waargenomen samenhang beschreven kan worden door cumulatieve normale verdelingen die verlopen van volledig samenhangend tot volledig niet samenhangend. De 50% omslagpunten en de standaardbreedtes van de verdelingen berekend over het gemiddelde van vier luisteraars staan weergegeven in fig. 2.9. Toetsing met de Friedman-tweeweg-variantieanalyse leverde geen significante verschillen op.

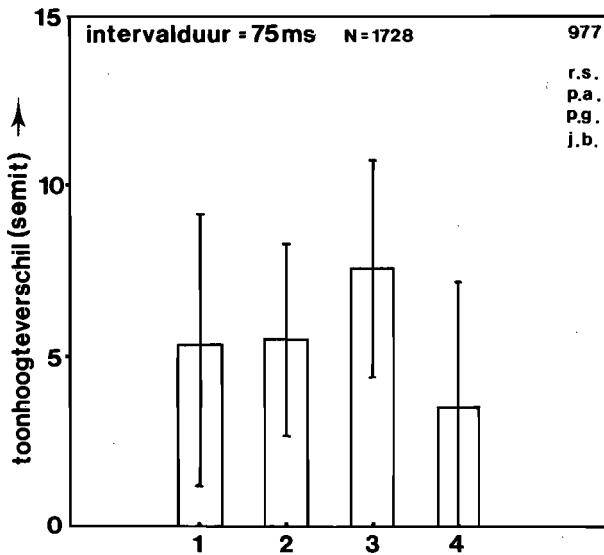


Fig. 2.9 50% Omslagpunten en standaardbreedtes van de cumulatieve normale verdelingen voor het horen van één of van twee stromen klinkers of lettergrepen in reeksen klanken met alternerend hoge en lage toonhoogte, bij een constante intervalduur van 75 ms, gemiddeld over vier luisteraars.

- (1) homogene klinkers
- (2) /b-/klinker/combinaties
- (3) /b-/klinker-/b/combinaties
- (4) /b-/klinker/combinaties met niet gebruikelijke spectrale eindbuigingen.

2.6 Toonhoogtebewegingen en Auditieve Samenhang: Aanvullende Waarnemingen

De vraag rijst wat de oorzaak is van het opdelen van het spraaksignaal in twee perceptieve stromen, zodat in reeksen klinkers met alternerende toonhoogte twee verschillende sprekers waargenomen kunnen worden.

Twee experimentele variabelen blijken direct invloed te hebben op de waarneming van één of twee sprekers: het toonhoogteverschil tussen de klinkers en de duur van de stille intervallen tussen de klinkers. Deze combinatie suggereert een virtuele toonhoogtebeweging. Indien de stille intervallen kort zijn, impliceert dit een snelle virtuele toonhoogtebeweging en als de intervallen lang zijn impliceert dit een langzame virtuele toonhoogtebeweging. Om iets meer duidelijkheid te krijgen of nu de toonhoogteverschillen of de snelheid van de virtuele toonhoogtebewegingen een rol speelt, is een aanvullende meting gedaan. Met behulp van de software synthetisator, besproken in paragraaf 2.3, is een lange homogene klinker /a/ gesynthetiseerd. De toonhoogtecontour van deze klinker heeft een continu verloop. Deze wordt gekarakteriseerd door niveaus met hoge en lage toonhoogte waartussen de toonhoogte continu, binnen een bepaalde tijd van het ene constante niveau naar het andere verglijdt. De toonhoogtebewegingen zijn dus niet meer virtueel maar reëel.

Drie toonhoogteverschillen (5.1, 10.4 en 12.6 halve tonen) zijn bij de verschillende stimuli gerealiseerd in 22 en in 200 ms. De toonhoogteverschillen, de snelheden van de toonhoogtebewegingen en de duren hiervan staan weergegeven in onderstaande tabel.

stimulus	toonhoogteverschil (semit)	duur van de toonhoogtebeweging (ms)	snelheid toonhoogte (semit/s) beweging
1	12.6	22	572
2	10.4	22	472
3	5.1	22	232
4	12.6	200	63
5	10.4	200	52
6	5.1	200	25.5

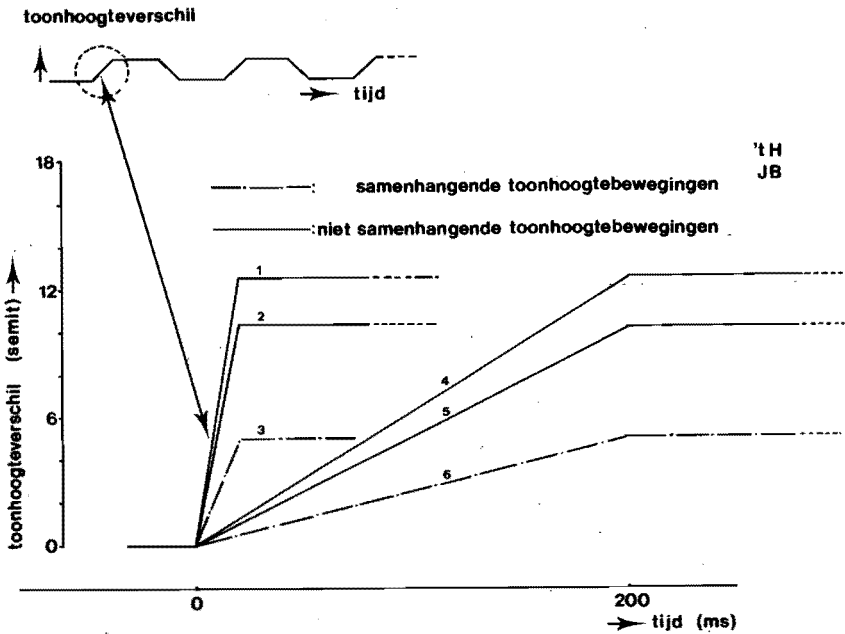


Fig. 2.10 Schematische voorstelling van 6 toonhoogtebewegingen in lange homogene klinkers met verschillende toonhoogtebewegingen en verschillende eindwaarden van de toonhoogte. De twee waarnemers namen stimuli (3) en (6) als auditief samenhangend en stimuli (1), (2), (4) en (5) als auditief niet samenhangend waar.

Twee luisteraars, waaronder de auteur, hebben naar deze stimuli geluisterd. De opdracht was aan te geven of deze stimuli auditief samenhangend of niet samenhangend waren. Als bij de stimulus auditieve splitsing werd waargenomen, dan werd de stimulus als auditief niet samenhangend beoordeeld. De twee luisteraars namen bij de stimuli met de toonhoogtebewegingen 1, 2, 4 en 5 (zie fig. 2.10) auditieve splitsing waar. Alleen de stimuli met de kleinste toonhoogteverschillen (stimulus 3 en 6) werden waargenomen als auditief samenhangend. Deze resultaten suggereren dat voornamelijk

het toonhoogteverschil en niet de snelheid van de toonhoogtebeweging in de stimulus verantwoordelijk is voor het waarnemen van één of twee sprekers.

Dit is in overeenstemming met de resultaten die verkegen zijn met de klinkerreeksen uit paragraaf 2.3. Voor intervallen kleiner dan 100 ms zijn de 50% omslagpunten voor het waarnemen van één of twee sprekers slechts in geringe mate afhankelijk van de duur van de stille intervallen waarin dan de virtuele toonhoogtebewegingen zouden moeten plaatsvinden.

2.7 Discussie en Conclusie

Hoewel auditieve samenhang een experimenteel moeilijk grijpbare notie is, levert de meetmethode die gebruikt is bij de in dit hoofdstuk besproken experimenten, toch bruikbare inzichten op die betrekking hebben op auditieve samenhang. In hoofdstuk 1 zijn in verband met auditieve samenhang een tweetal aspecten van de perceptie van spraak besproken, te weten de waarneming van volgordes en auditieve splitsing. Door aan de luisteraars de opdracht te geven te beoordelen of een stimulus afkomstig lijkt te zijn van één of twee sprekers, wordt in de experimenten de nadruk het meest gelegd op het perceptieve verschijnsel van auditieve splitsing. In hoeverre auditieve splitsing in dit geval samen gaat met "mispercepties" van volgorde is niet onderzocht. Uit de experimenten hebben we een aantal verbanden tussen fysische eigenschappen van het spraaksignaal en de waarneming van auditieve splitsing leren kennen. Allereerst blijkt het door het aanbrenge van toonhoogteverschillen in het spraaksignaal mogelijk de luisteraar de indruk te geven dat er meer dan één spreker aan het woord is. Bij stille intervallen tussen de spraaksegmenten korter dan 150 ms is een toonhoogteverschil van 2 à 5 halve tonen tussen de opeenvolgende spraaksegmenten al voldoende voor de luisteraar om de reeks spraakklanken niet meer als bij elkaar horend waar te nemen. Het toonhoogteverloop is kennelijk een belangrijk middel voor de luisteraar om een

stroom spraakklanken te kunnen interpreteren als geluiden die bij elkaar horen.

Ten tweede blijkt dat de duur van de spraaksegmenten, althans wanneer deze tussen de 100 en 200 ms ligt, geen bepalende factor is voor de auditieve samenhang van het spraaksignaal. Van de onderzochte temporele factoren die een effect zouden kunnen hebben op het horen van één of twee sprekers, bleek alleen de duur van de pauzes tussen opeenvolgende stemhebbende spraaksegmenten een effect te hebben. Ten derde blijkt ons vermoeden dat aanwezigheid van natuurlijke formantbuigingen de auditieve splitsing, veroorzaakt door de verschillen in toonhoogte, kan tegengaan, experimenteel niet aangetoond is.

In het kader van dit onderzoek zijn we er niet aan toegekomen om een eventuele bijdrage van verschillen in geluidssterkte tussen opeenvolgende spraakfragmenten te onderzoeken. Het is a priori zeer aannemelijk dat auditieve samenhang niet onafhankelijk is van zulke verschillen in geluidssterkte, maar hoe groot deze verschillen moeten zijn om een meetbaar effect te krijgen blijft binnen deze studie een open vraag.

De gebieden in fig. 2.2, 2.5, 2.6 en 2.7 waar luisteraars bij klinkerreeksen met alternerende toonhoogte één of twee luisteraars waarnemen vertonen grote gelijkens met wat Van Noorden (1975) vindt voor temporele samenhang respectievelijk voor splitsing bij reeksen zuivere tonen.

Het betrekkelijk kleine onzekerheidsgebied in onze data vergeleken met dat in Van Noordens data kan geheel verklaard worden uit de verschillende meetmethode. De door Van Noorden gebruikte "sweep" methode was erop gericht het gebied waarin zowel samenhang als geen samenhang kan optreden zo groot mogelijk te maken, terwijl onze resultaten verkregen zijn met een binaire gedwongen keuzeproef, die leidt tot een reductie van het onzekerheidsgebied.

De conclusie uit ons experiment dat vooral de toonhoogte van belang is voor de auditieve samenhang is in lijn met resultaten van Darwin en Bethell-Fox (1977).

In hun experimenten onderzoeken zij of klinkers die door continue formantverglidingen met elkaar verbonden zijn, door een discontinu toonhoogtepatroon perceptief aan verschillende sprekers toegewezen kunnen worden. Met andere woorden, zij proberen te onderzoeken of prosodische continuïteit die gesuggereerd wordt door spectrale continuïteit, verbroken kan worden door een discontinu toonhoogtepatroon. Met een formantsynthetisator hebben zij reeksen van drie homogene klinkers gesynthetiseerd waarbij de formantpatronen zonder discontinuïteiten van de ene homogene klinker naar de andere verglijden. De uren van de spectrale overgangen en van de homogene gedeeltes zijn elk 60 ms. Er zijn een tweetal toonhoogtepatronen gebruikt: een monotone patroon en een discontinu patroon waarbij de toonhoogte halverwege de spectrale overgangen abrupt verandert.

In tegenstelling tot de monotone toonhoogtecontouren geven de discontinue contouren aanleiding tot de perceptie van stopconsonanten. In gewone spraak worden dergelijke consonanten gekenmerkt door een stilte, geflankeerd door snelle spectrale veranderingen. Dit geeft een aanwijzing dat de continue formantpatronen door het opgelegde toonhoogtepatroon perceptief verdeeld worden in twee gescheiden reeksen. Perceptief lijkt het dan alsof binnen een reeks een stilte aanwezig is op het moment dat een stuk spraaksignaal op grond van de toonhoogte toebehoort aan een andere reeks. Dit is in overeenstemming met de subjectieve indruk dat er twee sprekers zijn: één sprekend met een hoge toonhoogte en een ander sprekend met een lage toonhoogte.

Auditieve samenhang is van groot belang om tot perceptieve integratie te komen van seriëel in de tijd geordende spraakfragmenten. De luisteraar moet in staat zijn om stukjes akoestische informatie die van dezelfde geluidsbron afkomstig zijn als zodanig te herkennen en met elkaar in verband te brengen.

In dit hoofdstuk is gekeken naar enkele fysische eigenschappen van het spraaksignaal die de luisteraar bij deze taak helpen en een bijdrage geven aan de auditieve samenhang. De toonhoogte blijkt een grote rol hierin te spelen. Deze geeft de luisteraar als het ware een indicatie over de bron waarvan de geluiden afkomstig zijn of lijken te zijn. Auditieve samenhang is niet alleen van belang bij de perceptie van seriëel geordende spraaksegmenten. Het speelt ook een grote rol bij de perceptie van parallelle akoestische informatie. In situaties waarin meerdere mensen tegelijkertijd aan het woord zijn, wordt vaak van de luisteraar verlangd dat hij in staat is om één enkele stem temidden van andere stemmen te blijven volgen en te verstaan wat er gezegd wordt. De toonhoogte van de stem zou een belangrijk hulpmiddel hiervoor kunnen zijn.

De potentiële bijdrage van toonhoogteverschillen en van het verloop van de toonhoogte aan de perceptieve scheiding van simultane spraakuitingen is onderzocht in een aantal experimenten die in het volgende hoofdstuk gerapporteerd zullen worden.

3. PERCEPTIEVE SCHEIDING VAN SIMULTANE SPRAAKUITINGEN

3.1 Inleiding

Indien twee ongeveer even luide opnames van de stem van één spreker tegelijkertijd door één luidspreker hoorbaar gemaakt worden, dan is het voor een luisteraar nog steeds redelijk goed mogelijk deze stemmen perceptief te scheiden en één van beide te volgen en te verstaan. In deze situatie kan de perceptieve scheiding niet verklaard worden door de stimulusfactoren die Cherry (1953) noemde; richting van het geluid, sprekergebonden eigenschappen van spraak en dialect, zijn alle identiek. Wat aan factoren overblijft zijn momentane verschillen in geluidssterkte en toonhoogte tussen beide spraaksignalen, de continuïteit in het toonhoogteverloop en in de spectrale samenstelling binnen ieder van de spraakuitingen.

In dit hoofdstuk zullen we vooral de bijdrage van toonhoogteverschillen tussen spraakuitingen aan de perceptieve scheiding onderzoeken. Er zijn een aantal redenen om in dit onderzoek naar de stimulusfactoren die perceptieve scheiding bevorderen, in eerste instantie de toonhoogte centraal te stellen.

Ten eerste is in het vorige hoofdstuk de toonhoogte de belangrijkste factor gebleken in het perceptief scheiden van niet-simultane, op elkaar volgende spraakklanken.

Ten tweede zijn er enkele informele observaties, die bij de eerste verkenning in het laboratorium in dezelfde richting wezen.

Zo bleek bij het optellen van de golfvorm van twee verschillende homogene klinkers, dat beide klinkers in het algemeen goed afzonderlijk herkenbaar bleven, behalve wanneer de toonhoogte van beide klinkers niet of weinig van elkaar verschilde. In dat geval versmolten ze perceptief tot één geheel waarin de klinkers niet meer afzonderlijk herkenbaar waren.

Een soortgelijk effect werd gevonden bij spraakuitingen. Als twee spraakuitingen, ingesproken door één of twee sprekers, met behulp van een vocoder een identieke toonhoogte

gegeven wordt, dan versmelten beide spraakuitingen perceptief tot één geheel. Als ook maar een gering toonhoogteverschil, bijvoorbeeld van 1 Hz wordt aangebracht, hoort men alweer twee afzonderlijke stemmen.

In 1890 heeft Stumpf dit perceptieve fenomeen voor niet-spraakklanken beschreven. Indien, volgens Stumpf, de toonhoogtes van twee gelijktijdige geluiden identiek zijn of gehele octaven van elkaar verschillen dan blijken beide geluiden te versmelten tot één klank. Als de toonhoogtes van beide geluiden verschillen zonder dat het verschil in frequentie een geheel veelvoud van een octaaf bedraagt, dan zijn beide klanken afzonderlijk goed waarneembaar.

Ook verschillen in geluidsterkte kunnen van invloed zijn op het scheiden van simultane spraakuitingen, ook als beide stemmen ongeveer even luid zijn. De luisteraar kan dan zijn aandacht willekeurig op een van beide stemmen richten. Een gering verschil in geluidsterkte kan dan resulteren in een betere verstaanbaarheid voor beide stemmen (Egan, Carterette and Thwing, 1954).

Naast het perceptieve scheiden van gelijktijdige geluiden is de tweede taak van de luisteraar het blijven volgen van de spraakuiting waarin hij geïnteresseerd is. Afgezien van de geluidsterkte, lijken de continuïteit in het toonhoogteverloop en in de spectrale samenstelling het meest in aanmerking te komen om de luisteraar te helpen één van twee simultane spraakuitingen te volgen. Spectrale continuïteit komt tot uiting in spectrale overgangen tussen opeenvolgende klanken. Op grond van de in het eerste hoofdstuk besproken literatuurgegevens mogen we veronderstellen dat spectrale continuïteit hoofdzakelijk werkzaam is bij spraakfragmenten die in de tijd niet te ver uit elkaar liggen. In verbonden spraak is dit vermoedelijk beperkt tot aansluitende spraakfragmenten. Voor niet in de tijd aansluitende spraakfragmenten komt de continuïteit in het verloop van de toonhoogte in aanmerking als helpende factor.

3.2 Natuurlijke Spraak

3.2.1 Normaal geïntoneerde spraak

In deze paragraaf zal gepoogd worden om experimenteel meer inzicht te krijgen in de factoren die voor een luisteraar van belang zijn om één stem te volgen in de aanwezigheid van andere stemmen. Daarvoor moet hij in staat zijn om ten eerste simultane spraakgeluiden perceptief te scheiden en ten tweede om één van beide stemmen te volgen en te verstaan. Het vermogen van de luisteraar om stemmen te scheiden is vermoedelijk van primair belang.

In de inleiding is het vermoeden geuit dat momentane toonhoogteverschillen tussen twee gelijktijdige en stemhebbende stukken spraak een belangrijke factor zijn in de perceptieve scheiding. Daarnaast zouden ook globale luidheidsverschillen een rol kunnen spelen.

Na enige exploratieve metingen kwam de volgende meetmethode, gebaseerd op Nakatani and Dukas (1973) in aanmerking om dit nader te onderzoeken.

Zinnen van zeven éénlettergrepige woorden met dezelfde grammaticale structuur en met een niet normale betekenisstructuur van het type "de vis vliegt kwaad over de brug", worden gemaskeerd door continu aanwezige spraak en worden zo de luisteraar aangeboden. De aard van de maskering die het meest effectief lijkt voor het doel dat wij ons gesteld hebben is die, waarbij de spraakstimuli gemaskeerd worden door andere continue spraak van dezelfde spreker. De stimuli bestaan uit eenvoudige zinnnetjes die één voor één gereproduceerd moeten worden door de luisteraar.

Door intonatieve verschillen aan te brengen tussen de zinnen en de maskeerder worden verschillende condities gecreëerd. Verder worden tijdens het experiment een aantal verschillende verhoudingen tussen de geluidsstrektes van het signaal en de maskeerder gebruikt. Door deze verschillende verhoudingen van de geluidsstrekte worden plafond- en vloer-effecten minder storend voor het vinden van verschillen in de verstaanbaarheid tussen verschillende condities. Volgens Nakatani and Dukas (1973) heeft een dergelijke meetmethode een groot oplossend vermogen.

Hoewel bij het opzetten van de hier gerapporteerde experimenten overwogen was om de gesproken testzinnen en de maskeerende tekst te manipuleren met een vocodersysteem, zodat de fysische eigenschappen van de spraak, met name toonhoogte en geluidsstrekte heel precies gevariëerd zouden kunnen worden, is daar in eerste instantie van afgezien omdat nog te weinig bekend was van de te onderzoeken verbanden tussen fysische eigenschappen van spraak en de verstaanbaarheid van door spraak gemaskeerde spraak. Daarom leek het beter eerst dit gebied op een wat ruwere wijze af te tasten met behulp van onbewerkte natuurlijke spraak, waarvan de eigenschappen gevariëerd werden door instructies aan de spreker. In paragraaf 3.3 zal een aansluitend experiment worden beschreven waarin met behulp van gevocoderiseerde spraak meer precies gekeken is naar het effect van toonhoogteverschillen tussen simultane spraakuitingen op de verstaanbaarheid.

Voor dit experiment zijn de volgende variabelen gekozen:

- de globale toonhoogteligging: volgens de verwachting zou een verschil in toonhoogteligging tussen de testzinnen en de maskeerder resulteren in een betere scheiding tussen de beide spraakuitingen en dus in betere verstaanbaarheid,
- de toonhoogtecontour: verwacht wordt dat een monotone toonhoogte de luisteraar minder houvast biedt dan een normale zinsmelodie, die binnen zijn verwachtingspatroon valt en daardoor het volgen van een spraakuiting vergemakkelijkt,
- de luidheid: deze variabele is inherent aan de gebruikte meetmethode. Het effect van een verschil in de luidheid tussen de testzinnen en de maskeerder op het scheidend vermogen zou op deze manier aan het licht kunnen komen.

De maskeerder en de testzinnen zijn ingesproken door dezelfde mannelijke spreker. Gekozen is voor vier intonatief verschillende soorten testzinnen.

- (1) zinnen met een normale toonhoogtecontour, ingesproken op een hoge toonhoogte;

- (2) zinnen met een normale toonhoogtecontour, ingesproken op een voor de spreker lage toonhoogte;
- (3) zinnen met een zo goed mogelijk monotoon toonhoogtepatroon, ingesproken op een hoge toonhoogte;
- (4) zinnen met een zo goed mogelijk monotoon toonhoogtepatroon, ingesproken op een lage toonhoogte.

Deze testzinnen worden in alle condities gemaskeerd door dezelfde maskeertekst, die ingesproken is met een normale intonatie en op een voor de spreker normale (lage) toonhoogte. De vier verkregen condities komen gedurende het experiment voor met zes verschillende geluidssterkteverhoudingen tussen het signaal en de maskeertekst:

$$S/M = 0, -3, -6, -9, -12, -15 \text{ dB}$$

Deze verhoudingen bleken in pilootexperimenten een goed oplossend vermogen te bieden.

De stimuli zijn op de volgende manier gemaakt. De maskeerder bestaat uit een tekst uit "Een stoet van dwergen" van S. Carmiggelt. De spreker heeft deze tekst ingesproken met de opdracht om normaal te intoneren en pauzes te vermijden of, indien onvermijdelijk, zo kort mogelijk te maken.

De testzinnen zijn opgebouwd uit zeven éénlettergrepige woorden en hebben een correcte grammaticale structuur: lidwoord, zelfstandig naamwoord, werkwoord, bijwoord, voorzetsel, lidwoord, zelfstandig naamwoord.

De zinnen zijn gegenereerd door voor iedere positie in de zin een woord te kiezen uit een verzameling woorden van dezelfde woordklasse. Alle gebruikte woorden behoorden tot het kernvocabulaire van éénlettergrepige woorden dat gebruikt wordt voor de Franse les in de eerste klas van de middelbare school. Uit de lijst van zinnen die zo ontstond zijn alle zinnen verwijderd die een normale, te verwachten betekenis hadden, zodat alleen zinnen overbleven met een niet normale, onverwachte betekenis. De tijd tussen de aanbidding van twee opeenvolgende testzinnen bedroeg steeds 7 sec. De spreesnelheid was ongeveer vier lettergrepen per seconde.

De stimuli voor conditie (1) zijn op de volgende wijze ver-

kregen. Een vrouwelijke spreker met een toonhoogteligging van gemiddeld ongeveer 210 Hz sprak iedere zin in op een band-lus. De mannelijke spreker die ook de maskeerder had ingesproken kreeg de opdracht deze zin met dezelfde toonhoogte te imiteren. Een jury van drie ervaren luisteraars controleerde tijdens het inspreken of de toonhoogtes van de mannelijke en de vrouwelijke spreker goed overeenkwamen. De zinnen voor conditie (2) zijn direct door de mannelijke spreker met een voor hem normale toonhoogteligging ingesproken. Deze toonhoogteligging is gemiddeld ongeveer 110 Hz.

De monotone zinnen voor condities (3) en (4) zijn verkregen door de spreker de opdracht te geven zo goed mogelijk monotoon te spreken. Als auditieve hulp kreeg de spreker via een koptelefoon een sinusvormig signaal van 214 Hz respectievelijk 112 Hz aangeboden. Er is niet veel aandacht geschonken aan het spraaktempo. In de praktijk heeft dat ertoe geleid dat het tempo in de testzinnen wat lager was dat in de maskeertekst.

De stimuli zijn opgenomen met een twee-sporen Revox A77. Met behulp van een Bruel & Kjaer dBA meter zijn bij de opnames de maximale geluidssterktes van elke testzin zo mogelijk binnen 2 dBA aan elkaar gelijk gemaakt. De geluidssterktes van de maskeertekst en van de testzinnen zijn op het gehoor zo goed mogelijk aan elkaar gelijk gemaakt. De aard en de grootte van de fluctuaties in de geluidssterkte van de maskeertekst, vergeleken met die van de kortere testzinnen die bovendien allemaal dezelfde taalstructuur en nagenoeg dezelfde duur hadden, maakten het niet erg zinnig om de geluidssterktes van de testzinnen en de maskeertekst objectief gelijk te stellen.

De maskeertekst en de testzinnen zijn op verschillende sporen opgenomen.

Om eventuele interacties tussen de zinnen en de condities te voorkomen zijn twee gelijkwaardige stimulusbanden gemaakt. De tweede stimulusband bevat dezelfde zinnen als de eerste band, maar de zinnen zijn zodanig ingelezen dat zij in een andere conditie terecht komen. Een andere reden om twee gelijkwaardige stimulusbanden te gebruiken is dat de geluids-

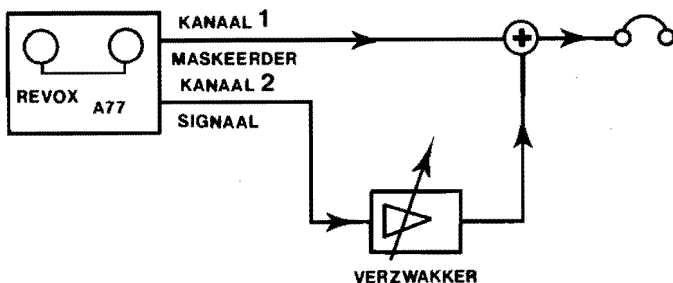


Fig. 3.1 Schematische weergave van de luisteropstelling.

sterkte van de maskeerder moeilijk onder controle te houden is. Door de fluctuaties in de geluidsterkte van de maskeer-tekst zouden bepaalde zinnen beter verstaan kunnen worden dan andere. Bij een beperkte verzameling zinnen zoals hier gebruikt is, kan dit tot gevolg kunnen hebben dat er vertekening in de resultaten optreedt.

Iedere stimulusband bestaat uit 96 zinnen, opgedeeld in zes blokken.

De blokken zijn ingedeeld naar verschillende klassen van de geluidsterkteverhouding tussen het signaal en de maskeerder. Ieder blok telt 16 zinnen, te weten vier zinnen van ieder van de vier condities.

Blok 1	S/M = 0 dB
Blok 2	S/M = - 3 dB
Blok 3	S/M = - 6 dB
Blok 4	S/M = - 9 dB
Blok 5	S/M = -12 dB
Blok 6	S/M = -15 dB

Iedere band is beluisterd door 10 verschillende proefperso-
nen. Zij kregen het signaal van de maskeertekst en van de
testzinnen gemengd via een koptelefoon aangeboden met als
opdracht iedere zin zo correct mogelijk na te zeggen. Om
inleereffecten te voorkomen, kreeg iedere luisteraar voor-
afgaande aan het experiment een oefenband met 32 zinnen te
beluisteren. Bij de scoring van de fouten bij het repro-

duceren zijn uitsluitend de vier belangrijkste woorden van iedere zin bekeken. Dit zijn de twee zelfstandige naamwoorden, het werkwoord en het bijwoord. De fouten zijn onderverdeeld in twee verschillende soorten: vervangingen en weglatingen.

Vervangingen door de luisteraars die slechts één foneem verkeerd zijn, zijn niet in de resultaten opgenomen. Dit is gedaan omdat dergelijke fouten niet noodzakelijk te maken hebben met de luistercondities. Zij traden ook op bij het reproduceren van stimuli die in een ijkingsproef aangeboden werden zonder de aanwezigheid van een maskeertekst.

Toevoegingen kwamen slechts bij hoge uitzondering voor en zijn daarom ook niet in de data opgenomen.

Met behulp van de Newman-Keulstest volgend op een klassieke vierweg-variantieanalyse, is onderzocht of tussen de intonatieve condities significante verschillen bestaan. Daarbij is steeds een betrouwbaarheidsniveau van 5% aangehouden.

Alléen de conditie met normaal geïntoneerde testzinnen met een hoge toonhoogte is significant verschillend van de rest. Tussen de twee groepen proefpersonen die ieder naar een verschillende band luisterden, waren geen verschillen.

In fig. 3.2 staan de percentages vervangingen plus weglatingen, gemiddeld over 20 luisteraars, voor de vier condities weergegeven als functie van de signaal-maskeerderverhouding.

Om een indruk te krijgen wat het effect is van een ten opzichte van de maskeertekst verschillende ligging van de toonhoogte moeten de resultaten van condities (1) en (2) en van de condities (3) en (4) (zie onderschrift fig. 3.2) paarsgewijze vergeleken worden. In fig. 3.3 staan de resultaten voor de zinnen met een normale intonatiecontour en met een hoge en een lage toonhoogteligging gemiddeld over alle signaal-maskeerder geluidsterkteverhoudingen weergegeven.

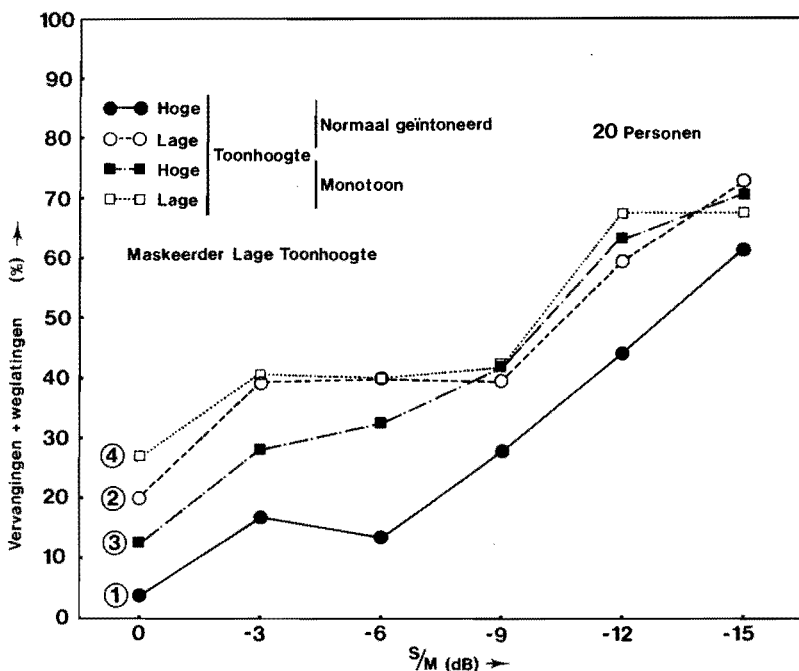
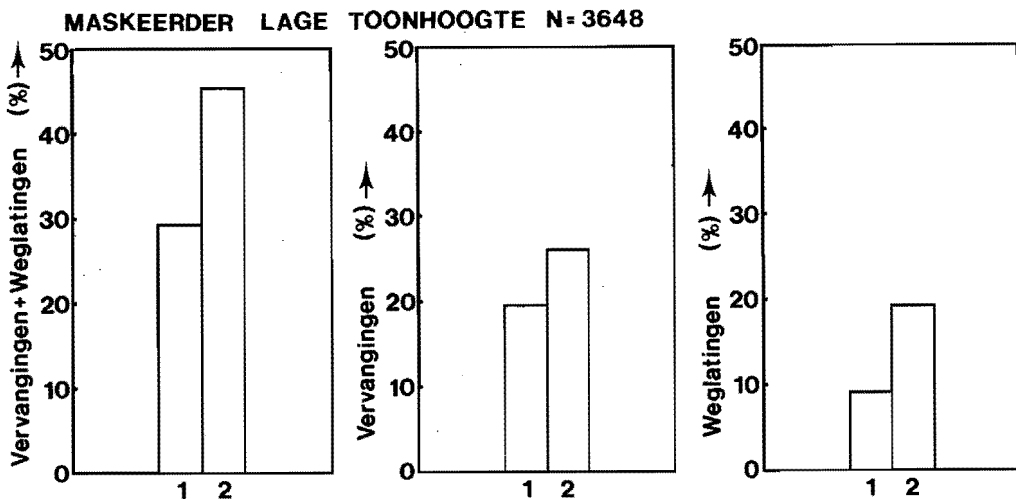


Fig. 3.2 Relatief aantal fouten voor vier condities uitgezet als functie van het relatieve geluidsniveau van testzinnen t.o.v. maskeertekst. De resultaten zijn gemiddeld over 20 luisteraars. De vier condities waren testzinnen, normaal geïntoneerd met hoge (1) en lage toonhoogtes (2), en monotone testzinnen met hoge (3) en lage toonhoogtes (4), gemaskeerd met een normaal geïntoneerde maskeertekst met een lage toonhoogte.

Bij zinnen met dezelfde toonhoogteligging als de maskeertekst (conditie 2) treden significant meer vervangingen en weglatingen op dan bij zinnen met een hogere toonhoogteligging (Newman-Keuls). Voor de monotone testzinnen met een hoge en lage toonhoogteligging staan de resultaten gemiddeld over alle signaal-maskeer geluidsstrekteverhoudingen weergegeven in de fig. 3.4.

De monotone testzinnen met dezelfde toonhoogteligging als de maskeerder leveren significant meer weglatingen op dan de monotone zinnen met een hoge toonhoogte. (Newman-Keuls). Het aantal vervangingen levert echter geen duidelijke verschillen op.



Resultaten gemiddeld over alle geluidssterkte verhoudingen

1: Hoge |
 2: Lage | **Toonhoogte, normaal geïntoneerde zinnen**

Fig. 3.3 Relatief aantal fouten, gemiddeld over 20 luisteraars en over alle relatieve geluidsniveaus voor de condities met normaal geïntoneerde testzinnen met hoge en met lage toonhoogte. De maskeertekst is normaal geïntoneerd en heeft een lage toonhoogte.

In fig. 3.2 is duidelijk de globale invloed te zien van de verschillen in geluidssterkte tussen de maskeerder en de testzinnen op de verstaanbaarheid. Opvallend is dat in de condities waarbij de testzinnen eenzelfde toonhoogte hebben als de maskeerder (condities 2 en 4), de verstaanbaarheid in het gebied van -3 t/m -9 dB signaal-maskeerder geluidssterkteverhouding nauwelijks beïnvloed lijkt te worden door de relatieve geluidssterkte van de stimuli. Voor een mogelijke verklaring, zie paragraaf 3.4.

Laten we de relatie tussen de signaal-maskeerder geluidssterkteverhouding en het optreden van de verschillende soorten fouten bij de vier verschillende condities nader bekijken.

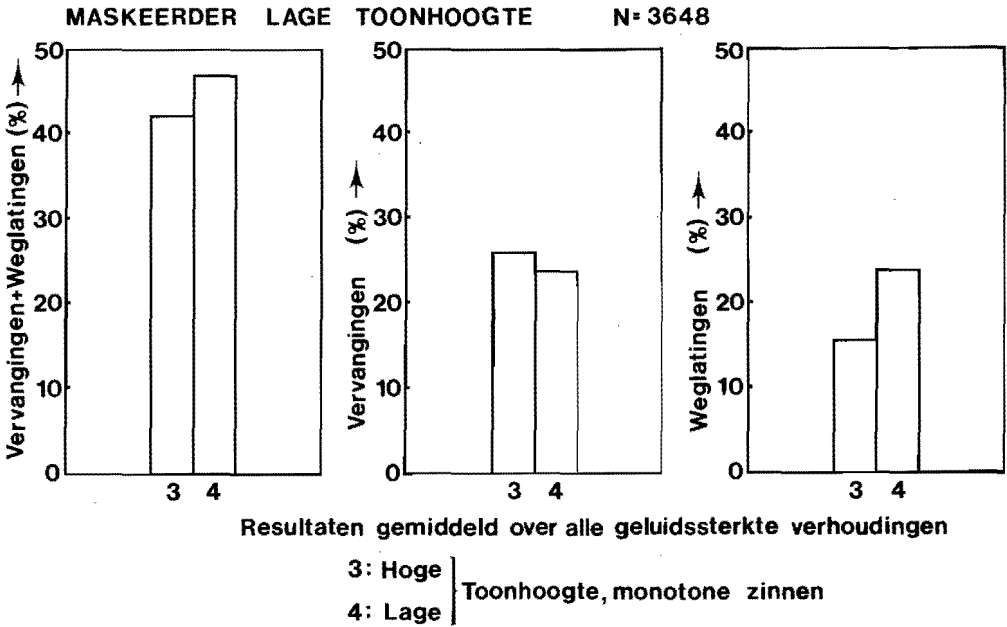


Fig. 3.4 Relatief aantal fouten, gemiddeld over 20 luisteraars en over alle relatieve geluidsniveaus, voor monotone testzinnen met een hoge en een lage toonhoogte. De maskeertekst is normaal geïntoneerd en heeft een lage toonhoogte.

Uit fig. 3.5 blijkt dat het aantal vervangingen door de luisteraar een ander gedrag vertoont dan het aantal weglatingen. Het aantal vervangingen neemt praktisch voortdurend toe bij afname van de signaal-maskeerder geluidssterkteverhouding.

Het aantal weglatingen daarentegen blijft bij een afname van de signaal-maskeerder geluidssterkteverhouding tot -6 à -9 dB vrij constant op een laag niveau. Bij een nog grotere afname van de geluidssterkte neemt het aantal weglatingen plotseling relatief zeer sterk toe (4 à 7%/dB).

Om de invloed van de aanwezigheid van een normale toonhoogtecontour, vergeleken met een monotone contour, op de verstaanbaarheid te vinden, dienen de resultaten van condities (1) en (3) en van (2) en (4) paarsgewijze vergeleken te worden.

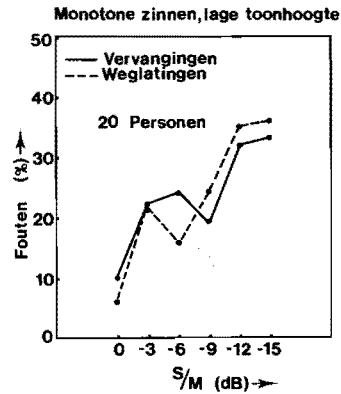
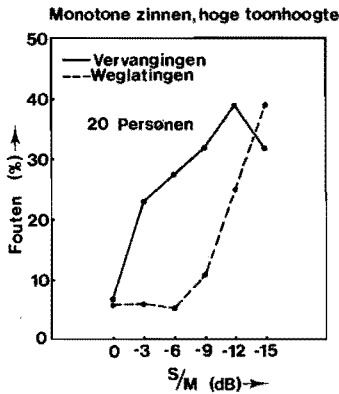
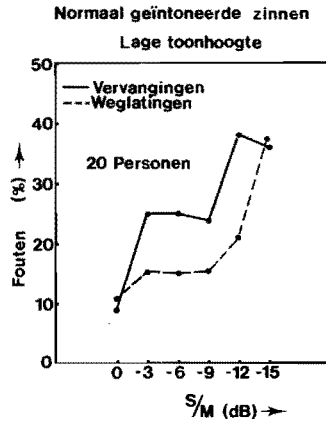
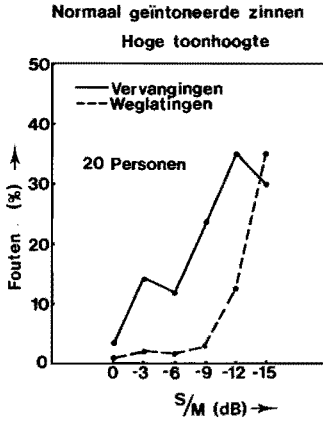


Fig. 3.5 De relatieve aantallen vervangingen en weglatingen, gemiddeld over 20 luisteraars, als functie van het relatieve geluidsniveau voor normaal geïntoneerde testzinnen met hoge toonhoogte (linksboven) en met lage toonhoogte (rechtsboven) en voor quasimonotone testzinnen met hoge toonhoogte (linksonder) en met lage toonhoogte (rechtsonder). De maskeertekst is normaal geïntoneerd met een lage toonhoogte.

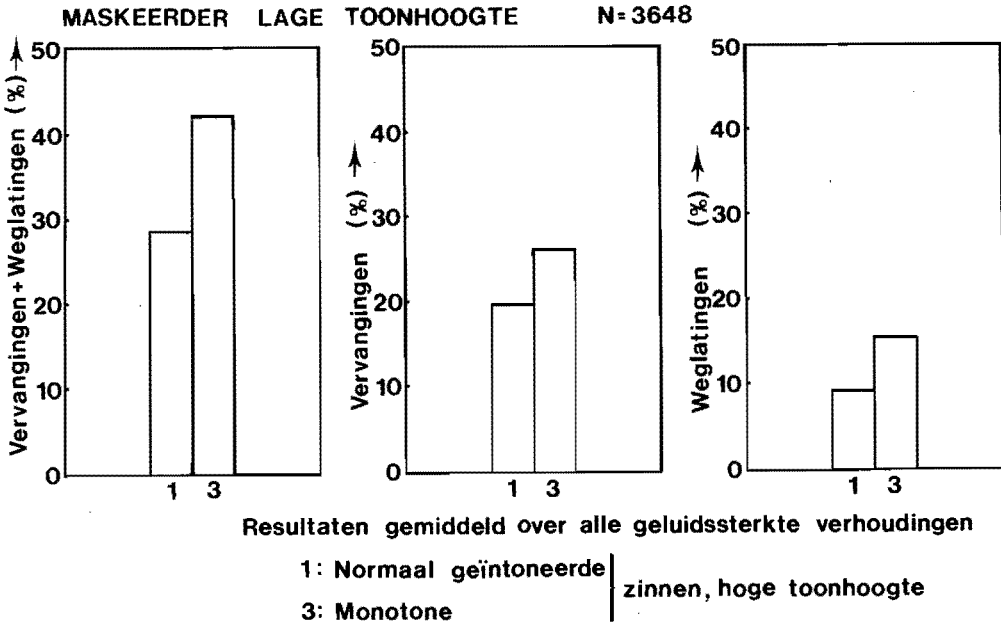


Fig. 3.6 Relatief aantal fouten, gemiddeld over 20 luisteraars en over alle relatieve geluidsniveaus, voor condities met normaal geïntoneerde testzinnen met hoge toonhoogte en monotone testzinnen met hoge toonhoogte. De maskeertekst is normaal geïntoneerd met een lage toonhoogte.

Bij de condities (1) en (3), dit zijn testzinnen met een hoge toonhoogteligging, zijn de resultaten tussen de monotone testzinnen en de testzinnen met een normale toonhoogtecontour met uitzondering van de weglatingen, significant verschillend. Bij vergelijking van de resultaten van condities (2) en (4), testzinnen met eenzelfde ligging van de toonhoogte als bij de maskeertekst, zijn de verschillen niet significant verschillend (Newman-Keuls).

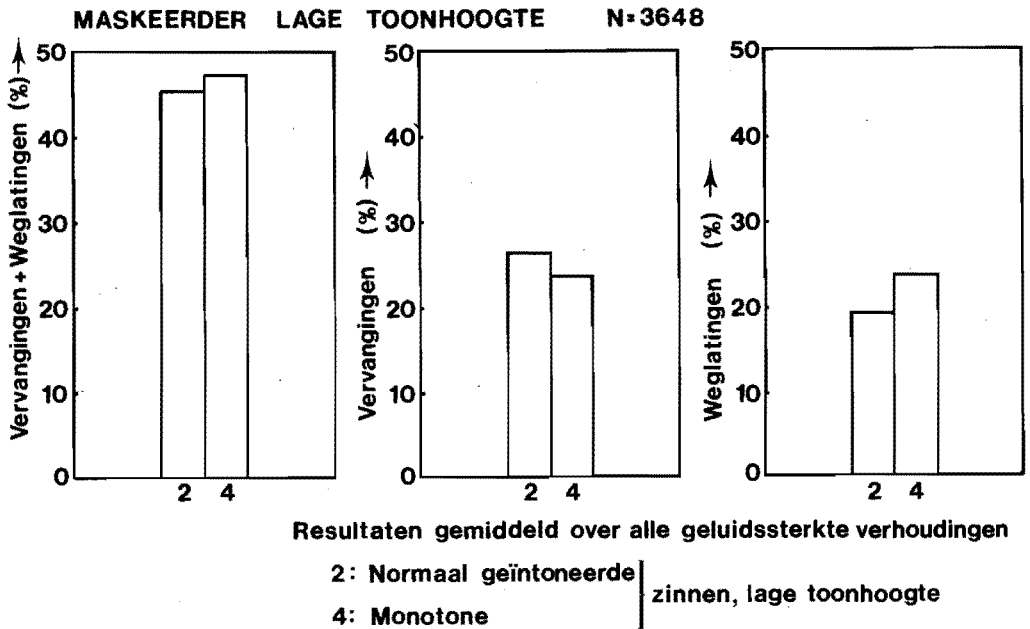


Fig. 3.7 Relatief aantal fouten, gemiddeld over 20 luisteraars en over alle relatieve geluidsniveaus voor condities met normaal geïntoneerde testzinnen met lage toonhoogte en monotone testzinnen met lage toonhoogte. De maskeertekst is normaal geïntoneerd met een lage toonhoogte.

3.2.2 Quasi-monotone spraak

Het cruciale punt bij deze experimenten is dat, wanneer de toonhoogtes van de maskeertekst en het signaal op hetzelfde moment identiek zijn, maximale versmelting verwacht mag worden. Vanuit dat oogpunt moeten de effecten, die teweeg gebracht worden door een verschil in toonhoogte tussen signaal en maskeertekst en de aanwezigheid van een al of niet monotone toonhoogtecontour, beoordeeld worden aan de hand van de toonhoogteverdelingen van de maskeertekst en de gebruikte testzinnen.

In de volgende figuren staan de toonhoogteverdelingen voor de gebruikte maskeertekst en de testzinnen weergegeven. De toonhoogtes zijn gemeten met een Frøkjær-Jensen Trans Pitch Meter. Aan de hand van deze verdelingsfuncties wordt het

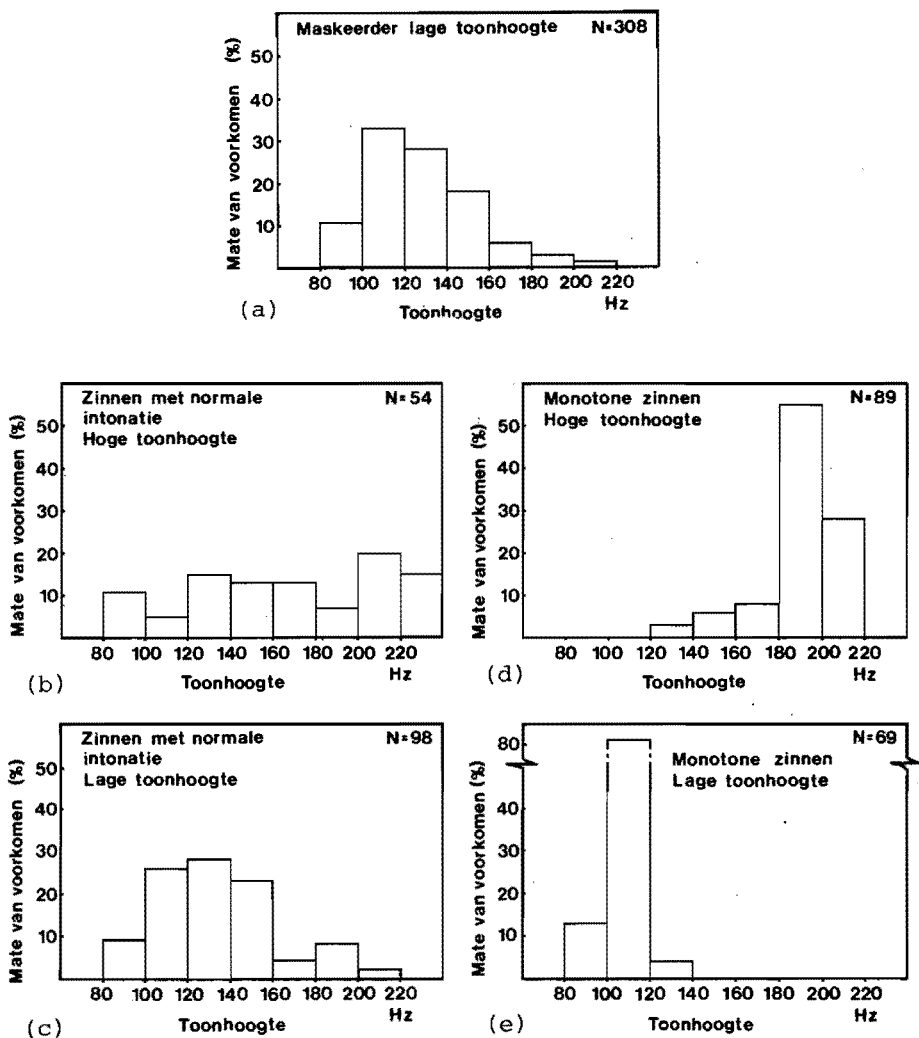


Fig. 3.8 Toonhoogteverdelingen voor:

- (a) normaal geïntoneerde maskeerder met lage toonhoogte
- (b) normaal geïntoneerde testzinnen ingesproken met hoge toonhoogte
- (c) normaal geïntoneerde testzinnen ingesproken met lage toonhoogte
- (d) monotone testzinnen ingesproken met hoge toonhoogte
- (e) monotone testzinnen ingesproken met lage toonhoogte.

duidelijk dat de oorspronkelijke keuze van een maskeertekst met een normaal intonatiepatroon ongelukkig is. Door de grote spreiding in de toonhoogtes van de maskeertekst zal zelden maximale versmelting ten gevolge van identieke toonhoogtes van maskeertekst en testzinnen op kunnen treden. De keuze van een monotone maskeertekst lijkt dan ook zeer voor de hand te liggen. Vanwege de geringere spreiding van de toonhoogte zullen de resultaten ook beter interpreterbaar zijn.

De vragen uit de vorige paragraaf, naar de bijdrage van verschillen in toonhoogteligging aan het scheidend vermogen van de luisteraar, en naar de bijdrage van een al of niet monotone toonhoogtecontour aan het beter kunnen volgen van één van de stemmen, kunnen opnieuw gesteld worden bij het gebruik van een monotone maskeertekst.

De opzet van dit experiment is bescheidener dan dat van het vorige. Er worden geen verschillende geluidssterkteverhoudingen tussen signaal en maskeertekst genomen. De stimuli hebben globaal dezelfde geluidssterkte als de maskeertekst. De stimuli zijn verder identiek aan die in het vorige experiment. Dit wil zeggen dat zinnen zijn gebruikt van zeven éénlettergrepige woorden met een correcte grammaticale structuur maar vrijwel zonder redundantie in de betekenis. Deze testzinnen worden gemaskeerd door een tekst die door dezelfde spreker zo goed mogelijk monotoon is ingelezen.

Experimentele variabelen zijn de globale ligging van de toonhoogte en de aan- of afwezigheid van een normale toonhoogtecontour. Er zijn vier verschillende condities verkregen door gebruik te maken van intonatief dezelfde soorten testzinnen als in het vorige experiment:

- (1) zinnen met een normale toonhoogtecontour, ingesproken op een hoge toonhoogte,
- (2) zinnen met een normale toonhoogtecontour, ingesproken op een lage toonhoogte,

- (3) zinnen die zo goed mogelijk monotoon op een hoge toonhoogte zijn ingesproken, en
- (4) zinnen die monotoon op een lage toonhoogte zijn ingesproken.

Zowel de testzinnen als de maskeertekst zijn ingesproken door dezelfde mannelijke spreker op dezelfde wijze als in het vorige experiment. De maskeertekst en de testzinnen, opgenomen op een verschillend spoor van een Revox A77, worden tegelijkertijd binauraal aan de luisteraars ten gehore gebracht. De geluidssterktes van de maskeertekst en van de testzinnen zijn zo goed mogelijk aan elkaar gelijk gemaakt. De stimulusband bevat van elke conditie vier zinnen.

In dit experiment is dus de geluidssterkte niet als experimentele variabele gebruikt. Hierdoor werd een aanzienlijke bekorting in de experimentele procedure bereikt. De luisteraars, 20 personen, kregen de stimuli aangeboden met de opdracht elke testzin zo goed mogelijk verbaal te reproduceren. De scoringsmethode is identiek aan die van het vorige experiment. Dit wil zeggen dat alleen fouten die voorkomen in de twee zelfstandige naamwoorden, het werkwoord en het bijwoord gescoord worden. Eveneens is er een onderscheid gemaakt tussen twee soorten fouten: vervangingen en weglatingen.

In fig. 3.9 staan voor elk van de vier condities de relatieve aantallen vervangingen plus de aantallen weglatingen uitgezet. Opvallend is dat bij het gebruik van een monotone maskeertekst de verschillen tussen de condities groter zijn dan bij het gebruik van een niet-monotone maskeertekst.

Met behulp van de Newman-Keulstest, volgend op een 2-weg-variantie analyse, is onderzocht of tussen de intonatieve condities significante verschillen bestaan. Daarbij is steeds een betrouwbaarheidsniveau van 5% aangehouden. In fig. 3.9 zijn condities 1 en 3 significant verschillend van condities 2 en 4.

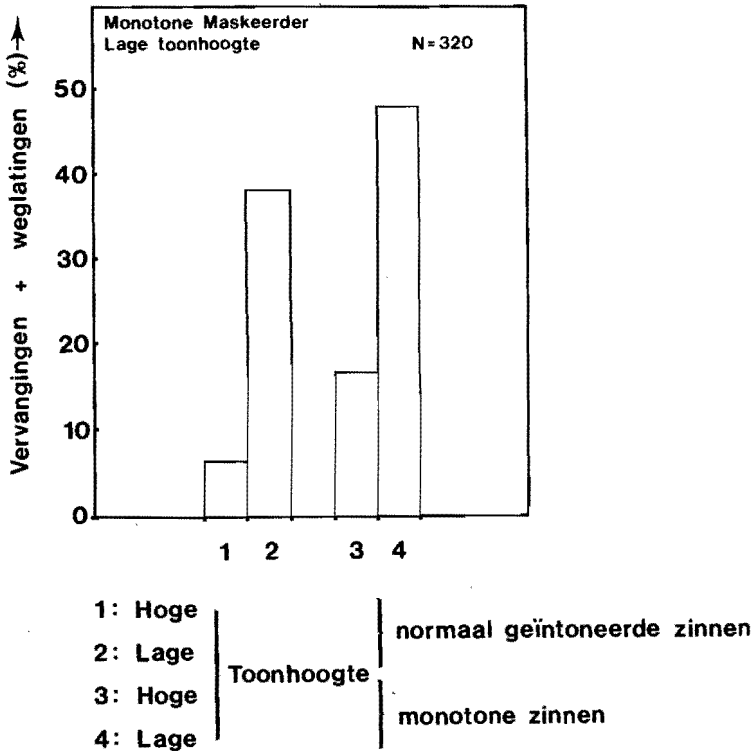


Fig. 3.9 Relatief aantal fouten, gemiddeld over 20 luisteraars voor normaal geïntoneerde testzinnen met hoge en lage toonhoogte en voor monotone testzinnen met hoge en lage toonhoogte. De maskeerder is monotoon ingesproken met een lage toonhoogte.

Het effect van verschillen in de ligging van de toonhoogte tussen de maskeertekst en de testzinnen komt goed naar voren in de fig. 3.10 en 3.11, waarin de resultaten van fig. 3.9 nader uitgesplitst worden in vervangingen en weglatingen. In de figuren 3.10 en 3.11 zijn alle verschillen tussen condities 1 en 2 en tussen 3 en 4 significant (Newman-Keuls).

De effecten van het al of niet aanwezig zijn van een normale toonhoogtecontour worden duidelijk bij vergelijking van de

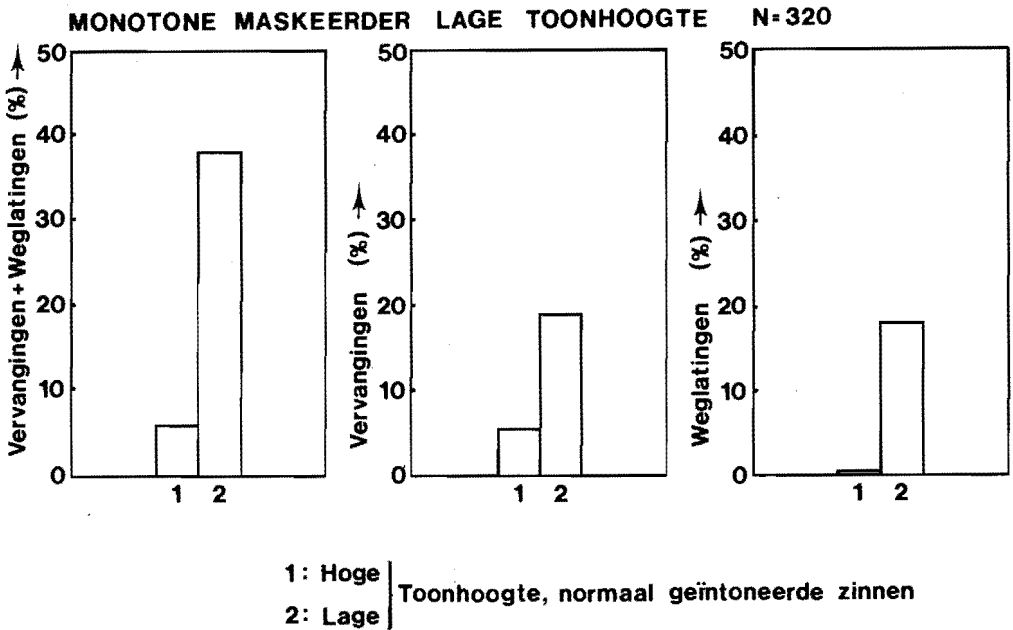


Fig. 3.10 Relatief aantal fouten, gemiddeld over 20 luisteraars voor normaal geïntoneerde testzinnen met hoge en lage toonhoogtes. De maskeerder is monotoon ingesproken met lage toonhoogte.

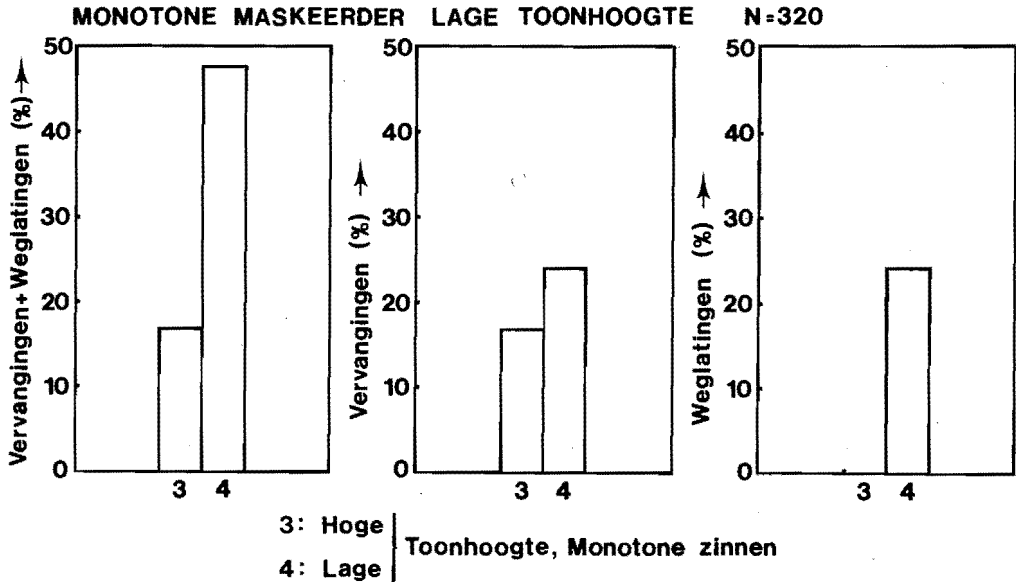
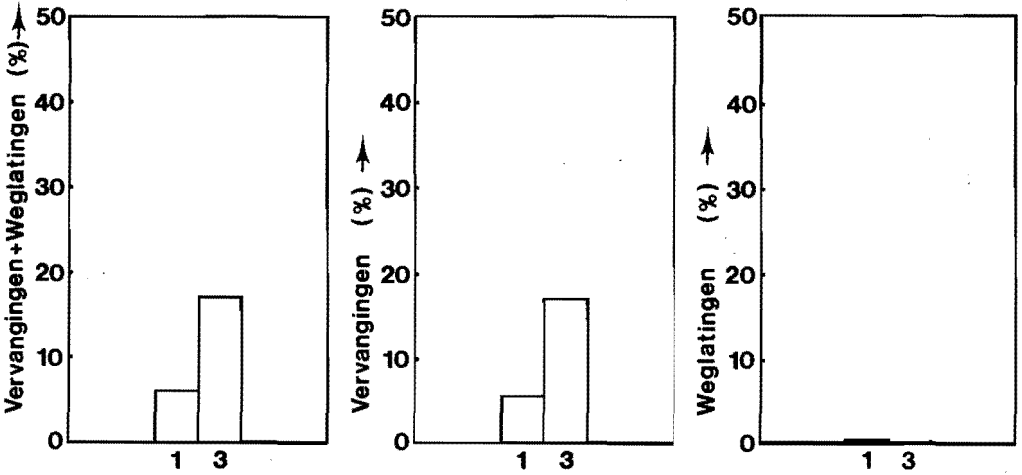


Fig. 3.11 Relatief aantal fouten, gemiddeld over 20 luisteraars voor monotone testzinnen met hoge en lage toonhoogtes. De maskeerder is monotoon ingesproken met een lage toonhoogte.

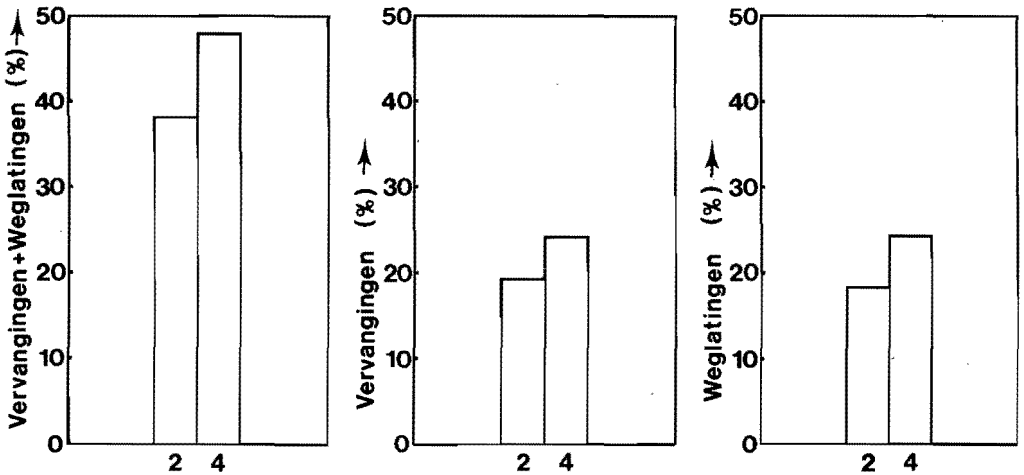
MONOTONE MASKEERDER LAGE TOONHOOGTE N=320



1: Normaal geïntoneerde zinnen, hoge toonhoogte
 3: Monotone zinnen, hoge toonhoogte

Fig. 3.12 Relatief aantal fouten, gemiddeld over 20 luisteraars voor normaal geïntoneerde en monotone testzinnen met een hoge toonhoogte. De maskeerder is monotoon ingesproken op een lage toonhoogte.

MONOTONE MASKEERDER LAGE TOONHOOGTE N=320



2: Normaal geïntoneerde zinnen, lage toonhoogte
 4: Monotone zinnen, lage toonhoogte

Fig. 3.13 Relatief aantal fouten, gemiddeld over 20 luisteraars voor normaal geïntoneerde en monotone testzinnen met lage toonhoogte. De maskeerder is monotoon ingesproken op een lage toonhoogte.

resultaten voor condities 1 en 3 en van condities 2 en 4 (zie fig. 3.12 respectievelijk fig. 3.13).

In fig. 3.12 is alleen het verschil in het aantal vervangingen tussen de condities 1 en 3 significant. In fig. 3.13 zijn geen van de verschillen significant (Newman-Keuls).

3.2.3 Bespreking van de resultaten

De oorspronkelijke vraagstelling aan het begin van dit hoofdstuk is de vraag naar de fysische factoren die de luisteraar helpen bij het scheiden en volgen van gelijktijdige spraakuitingen.

In de taak van de luisteraars in deze experimenten, het verbaal reproduceren van één stem terwijl twee stemmen tegelijkertijd aanwezig zijn, zijn beide aspecten nauw met elkaar verbonden en bepalend voor het resultaat.

In de beschreven experimenten zijn de bijdragen van de volgende factoren nader onderzocht:

- geluidssterkteverschillen tussen simultane stemmen
- toonhoogteverschillen tussen simultane stemmen
- aanwezigheid van een monotone of een normale toonhoogtecontour.

De resultaten laten weinig twijfel over het belang van de verhouding tussen de geluidsstrektes van testzinnen en maskeertekst voor de verstaanbaarheid. De geringe verbetering van de verstaanbaarheid van de testzinnen wanneer de geluidsstrekte van de testzinnen iets minder is dan die van de maskeertekst, ten opzichte van de situatie met gelijke geluidsstrekte, zoals gevonden door Egan, Carterette en Thwing (1954), komt in onze resultaten niet voor. Wel vonden we dat de geluidsstrekteverhouding een ander effect heeft op vervangingen dan op weglatingen.

De gepresenteerde data suggereren dat de toename van het aantal fouten bij een vermindering van de geluidsstrekte van het signaal zich in twee stadia voltrekt.

Bij het afnemen van de geluidsstrekte van de testzinnen ten opzichte van de maskeertekst treden al snel fouten op in de herkenning. Dit treft vooral medeklinkers en medeklinkergroepen. De perceptie van de klinkers blijkt in het algemeen beter bestand tegen maskering bij verdere afname van de geluidsstrekte van de testzinnen. Indien de geluidsstrekte zo ver is afgenomen dat ook de klinkers slecht verstaan worden, blijken de luisteraars meer geneigd het hele woord weg te laten.

Ook het verschil in globale toonhoogteligging tussen testzin en maskeertekst heeft een betrekkelijk groot effect op de verstaanbaarheid. Wanneer een niet-monotone maskeertekst wordt gebruikt, wordt het effect tegengegaan door de grote toonhoogtefluctuaties in de maskeertekst.

Bij het gebruik van een quasi-monotone maskeertekst worden de effecten veel duidelijker. Wanneer de fouten worden uitgesplitst in weglatingen en vervangingen, blijkt dat als er een verschil bestaat tussen maskeertekst en testzinnen in de ligging van de toonhoogte, geen of bijna geen weglatingen voorkomen. Indien de ligging van de toonhoogte van de maskeerder en het signaal hetzelfde zijn, dan treden er juist veel weglatingen op.

In beschouwingen over de rol die de toonhoogte van de testzinnen speelt bij het volgen, moet men rekening houden met het effect van de globale toonhoogteligging en het effect van de toonhoogtecontour van de maskeertekst.

Bij combinatie van een niet-monotone maskeertekst met testzinnen met een hoge toonhoogteligging is perceptieve scheiding in het algemeen relatief goed. De aanwezigheid van een normale toonhoogtecontour blijkt het aantal foutieve responsies en weglatingen significant te verminderen.

Als de testzinnen dezelfde toonhoogteligging hebben als de maskeerder, is de perceptieve scheiding ten gevolge van momentane toonhoogteverschillen slechter. Bij stimuli met een niet-monotone, normale toonhoogtecontour komen minder weglatingen voor dan bij testzinnen met een quasi-monotone

toonhoogtecontour, terwijl het aantal andere fouten niet minder is.

Wanneer een monotone maskeertekst met lage toonhoogteligging gecombineerd wordt met testzinnen met een hoge toonhoogteligging en een normale toonhoogtecontour, is de verstaanbaarheid goed. Dit kan verklaard worden door de geringe kans op samenvallen van de twee toonhoogtes, waardoor de perceptieve scheiding goed blijft. De kans op het samenvallen van de toonhoogtes is echter nog geringer bij het combineren van een monotone, op een lage toonhoogte ingelezen, maskeertekst met monotone testzinnen die een hoge toonhoogteligging hebben. Kennelijk is er een positief effect van de aanwezigheid van normale toonhoogtecontouren op de verstaanbaarheid, dat niet verklaard kan worden uit de kans op samenvallen van de toonhoogtes van testzinnen en maskeertekst.

Als de toonhoogteligging van de maskeerder en de stimuli ongeveer gelijk zijn, zijn de verschillen tussen de monotone en niet-monotone stimuli niet significant. De scheiding tussen maskeerder en stimulus is slecht en de aanwezigheid van een normale toonhoogtecontour geeft slechts een geringe verbetering die in dat geval verklaard kan worden door een iets kleinere kans op samenvallen van de toonhoogtes, waardoor de perceptieve scheiding beter kan zijn.

Deze resultaten leiden tot de conclusie dat een normale toonhoogtecontour een gering positief effect heeft op het kunnen volgen van een spraakuiting. Voorwaarde is echter dat de luisteraar eerst in staat moet zijn om simultane stemmen van elkaar te scheiden op grond van verschil in toonhoogteligging.

3.3 Volstrekt Monotone Spraak

3.3.1 Inleiding

In de inleiding van dit hoofdstuk is reeds gesproken over het verschijnsel dat wanneer twee spraakuitingen die met

een kanaalvocoder voorzien zijn van dezelfde toonhoogte, tegelijkertijd hoorbaar gemaakt worden, beide spraakuitingen perceptief versmelten tot één geheel. Al bij een gering verschil in toonhoogte tussen beide spraakuitingen hoort een luisteraar twee stemmen, maar hij kan geen van beide goed verstaan. Bij een groter toonhoogteverschil wordt het gemakkelijker de spraakuitingen te verstaan.

Dit verschijnsel geeft de indruk dat perceptieve scheiding van simultane spraakuitingen afhankelijk is van de grootte van het toonhoogteverschil tussen beide spraakuitingen en dat scheiding minimaal is wanneer sprake is van identieke toonhoogtes.

In de vorige experimenten is aangetoond dat globale toonhoogteverschillen een belangrijke rol spelen bij het scheiden van gelijktijdige spraakuitingen.

In deze paragraaf zullen we proberen nauwkeuriger na te gaan wat het effect is van de grootte van toonhoogteverschillen tussen maskeertekst en testzinnen op de verstaanbaarheid.

3.3.2 Methode

Omdat wij in deze paragraaf voornamelijk geïnteresseerd zijn in het effect van de grootte van toonhoogteverschillen, ligt het voor de hand om gebruik te maken van volstrekt monotone spraak. Volstrekt monotone spraak kan gemaakt worden met behulp van een vocoder.

De meetmethode voor de verstaanbaarheid van de gemaskeerde spraak moet voldoende gevoelig zijn en een voldoende groot oplossend vermogen hebben om ook geringe verschillen tussen de condities aan te kunnen tonen. Daarom is ook hier de keuze gevallen op de meetmethode van Nakatani, eerder beschreven in paragraaf 3.2.1.

Het experiment is als volgt opgebouwd. De testzinnen worden gemaskeerd door een lange doorlopende tekst. De luisteraar heeft als opdracht de testzinnen zo correct mogelijk na te

zeggen, waarbij de antwoorden op magnetische band worden vastgelegd om later geanalyseerd te worden. Als variabele is het toonhoogteverschil tussen de testzinnen en de maskeertekst genomen. Gedurende het experiment wordt ook de geluidssterkteverhouding tussen de testzinnen en de maskeertekst gevariëerd.

Als maskeerder is een lange tekst genomen uit "Een stoet van dwergen" van S. Carmiggelt. Deze tekst, ingelezen door een mannelijke spreker, is met behulp van een LPC (Linear Predictive Coding) vocoder geanalyseerd en met een Rockland digitale synthetisator geresynthetiseerd met een constante toonhoogte van 100 Hz (vgl. Vogten and Willems, 1977). Voor de testzinnen is eenzelfde procedure gevolgd. Evenals bij de vorige experimenten is gebruik gemaakt van syntactisch identieke en grammaticaal correcte zinnen ter lengte van zeven éénlettergrepige woorden. Deze zinnen zijn door een aselechte keuze van de woorden wat betekenis betreft onwaarschijnlijk en niet voorspelbaar. Deze zinnen, die ingesproken zijn door dezelfde spreker die ook de maskeertekst heeft ingelezen, zijn na analyse met de LPC vocoder geresynthetiseerd met een constante toonhoogte van 100, 103, 106, 109, 120 of 200 Hz. In pilootexperimenten bleken deze waarden voor de toonhoogtes een voldoende groot oplossend vermogen te geven.

Alle geresynthetiseerde spraakuitingen zijn opgenomen met een twee-sporen Revox A77 bandrecorder, waarbij de grootste zorg is besteed aan de constantheid van de bandsnelheid. De opnamen zijn gemaakt in een airconditioned ruimte waarbij de recorder enige uren voor de opnames aangezet werd om op werkteemperatuur te komen.

Steeds is de recorder ruim voor elke opname gestart om inslingerverschijnselen te voorkomen. Deze voorzorgen zijn genomen om variaties in de toonhoogte van de geresynthetiseerde opnamen zo klein mogelijk te houden.

Met behulp van een Bruel & Kjaer dBA meter zijn bij de opnames de maximale geluidssterktes van elke testzin zo

mogelijk binnen 2 dBA aan elkaar gelijk gemaakt. De geluidssterktes van de maskeertekst en van de testzinnen zijn op het gehoor zo goed mogelijk aan elkaar aangepast.

De stimulusband bestaat uit 96 zinnen die verdeeld zijn in vier blokken van 24 zinnen. Binnen elk blok komen in willekeurige volgorde vier zinnen met een identieke toonhoogte voor. Ten opzichte van de maskeertekst wordt het geluidsniveau van elk blok testzinnen gevariëerd en wel zodanig dat de volgende geluidssterkteverhoudingen tussen de testzinnen en de maskeertekst verkregen worden:

$$S/M = 0, -5, -10 \text{ en } -15 \text{ dB}$$

De maskeertekst en de testzinnen zijn op een verschillend spoor opgenomen en de beide geluidssignalen worden bij de presentatie met elkaar gemengd en met een Sennheiser HD424 hoofdtelefoon hoorbaar gemaakt.

De stimulusband is door 10 verschillende luisteraars, die alle enige ervaring hadden met auditieve experimenten, beluisterd met de opdracht elke testzin zo correct mogelijk te reproduceren. Voorafgaande aan deze stimulusband kregen de luisteraars een korte oefenband van 24 zinnen te horen om ervaring op te doen met het luisteren naar volstrekt monotone vocoderspraak en te wennen aan de experimentele omstandigheden.

De fouten die de luisteraars maakten bij het nazeggen van de testzinnen zijn op dezelfde manier gescoord als bij de andere experimenten. Er is uitsluitend gekeken naar de fouten die de luisteraars maakten in de vier inhoudswoorden van elke stimuluszin. Dit zijn de twee zelfstandige naamwoorden, het werkwoord en het bijwoord.

De fouten zijn onderverdeeld in twee verschillende soorten: vervangingen en weglatingen.

Met behulp van de Newman-Keuls test volgend op een 3-weg-variantieanalyse, is onderzocht of tussen de toonhoogte-

condities significante verschillen bestaan. Daarbij is een betrouwbaarheidsniveau van 5% aangehouden.

3.3.3 Resultaten

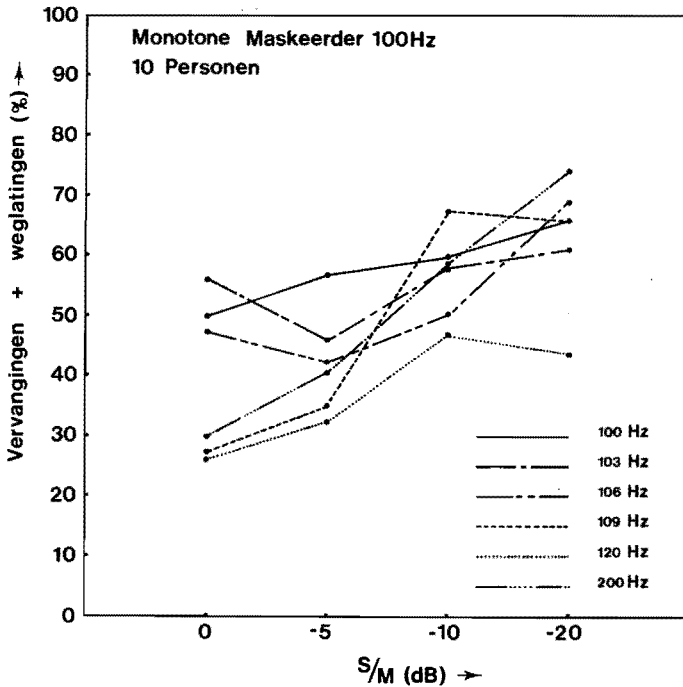


Fig. 3.14 Relatief aantal fouten, gemiddeld over 10 luisteraars als functie van het relatieve geluidsniveau, voor volstrekt monotone testzinnen van 100, 103, 106, 109, 120 en 200 Hz. De maskeerder is volstrekt monotoon 100 Hz.

In fig. 3.14 staat voor de verschillende stimuli het totale aantal fouten als functie van het relatieve geluidsniveau van de testzinnen t.o.v. de maskeertekst weergegeven. Alle toonhoogtecondities zijn significant van elkaar verschillend (Newman-Keuls). In de fig. 3.15 en 3.16 staan de vervangingen en de weglatingen apart weergegeven. Alle verschillen tussen de toonhoogtecondities zijn in fig. 3.15 niet en in fig. 3.16 wel significant (Newman-Keuls).

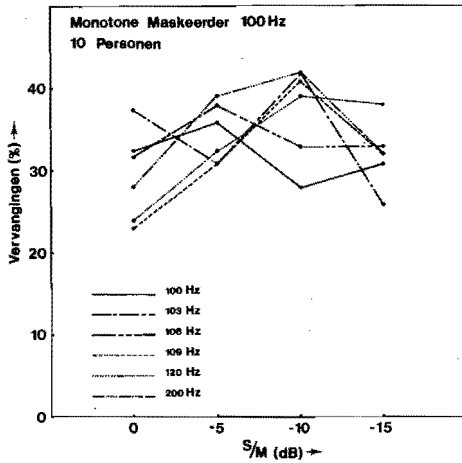


Fig. 3.15 Relatief aantal vervangingen, gemiddeld over 10 luisteraars, als functie van het relatieve geluidsniveau, voor volstrekt monotone testzinnen van 100, 103, 106, 109, 120 en 200 Hz. De maskeerder is volstrekt monotoon 100 Hz.

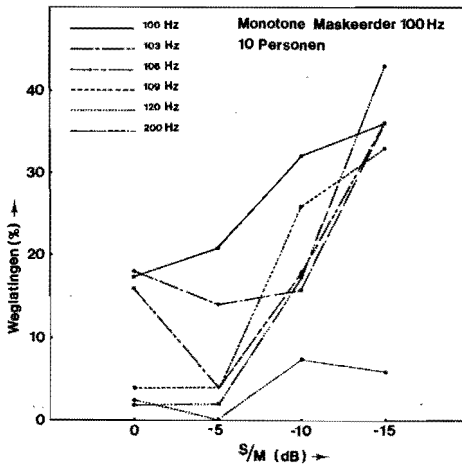


Fig. 3.16 Relatief aantal weglatingen, gemiddeld over 10 luisteraars, als functie van het relatieve geluidsniveau, voor volstrekt monotone testzinnen van 100, 103, 106, 109, 120 en 200 Hz. De maskeerder is volstrekt monotoon 100 Hz.

Direct opvallend is dat de weglatingen een functie zijn van het relatieve geluidsniveau van de testzinnen. Naarmate het geluidsniveau van de testzinnen t.o.v. de maskeertekst kleiner wordt neemt het aantal weglatingen toe. Het aantal vervangingen daarentegen laat geen significante verschillen zien over het gebied waarin het geluidsniveau gevarieerd is.

Extra aandacht moet worden geschonken aan de interpretatie van de aantallen weglatingen bij testzinnen van 100 Hz, 120 Hz en 200 Hz. Bij het gebruik van een maskeertekst van 100 Hz mag verwacht worden dat testzinnen met een toonhoogte van eveneens 100 Hz optimaal met de maskeertekst versmelten tot één perceptief geheel. Hoewel het aantal weglatingen voor testzinnen van 100 Hz bij de verschillende geluidsniveaus hoger is dan voor testzinnen met een andere toonhoogte, lijkt toch geen volledige versmelting op te treden. Daarvoor is het totale aantal fouten (50 à 65%) te gering. Waarschijnlijk zijn de momentane geluidsstrekteverschillen tussen de maskeertekst en de testzinnen zo groot geweest, dat op veel momenten de akoestische informatie van de testzin weinig of niet gemaskeerd werd door de op die momenten relatief geringe akoestische informatie van de maskeertekst. Het geringe aantal weglatingen bij de testzinnen met een toonhoogte van 120 Hz (0 à 7,5%) betekent waarschijnlijk dat het verschil in toonhoogte tussen maskeertekst en testzinnen een positief effect heeft op het scheiden van twee stemmen.

Bij testzinnen met een toonhoogte van 200 Hz is het opvallend dat bij vrijwel alle geluidsniveaus een veel betere scheiding verkregen wordt dan bij stimuli met een toonhoogte van 100 Hz. Op grond van Stumpf (1890, vgl. par. 1.4) hadden we verwacht dat in alle omstandigheden bij een octaaf verschil in toonhoogte, de scheiding even slecht zou zijn als bij gelijke toonhoogtes van maskeertekst en testzinnen. Voor een mogelijke verklaring zie par. 3.3.4.

Wanneer we de resultaten voor de condities met verschillende toonhoogtes middelen over de verschillende geluidsniveaus, krijgen we een indruk over het verband tussen de toonhoogte van de testzinnen en het percentage vervangingen en weglatingen.

In fig. 3.17 is goed te zien dat uitsluitend de weglatingen een duidelijke functie zijn van de toonhoogte van de gebruikte testzinnen. Het gemiddeld aantal vervangingen is geheel onafhankelijk van het toonhoogteverschil tussen testzinnen en maskeertekst.

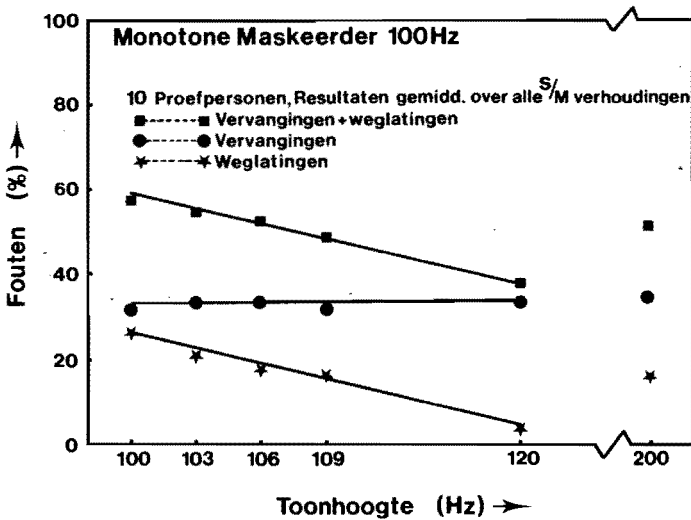


Fig. 3.17 Relatief aantal fouten, uitgesplitst in weglatingen, vervangingen en de som van beide, gemiddeld over 10 luisteraars en over alle relatieve geluidsniveaus, als functie van de toonhoogte. De maskeerder is volstrekt monotoon 100 Hz.

3.3.4 Bespreking van de resultaten

Aan het begin van deze paragraaf is de vraag gesteld wat de rol is die de grootte van het verschil in toonhoogte tussen testzin en maskeertekst speelt bij het perceptief scheiden van gelijktijdige spraakuitingen. Uit het laatste experiment, waarbij gebruik is gemaakt van volstrekt monotone

spraak komt als resultaat naar voren dat althans voor de gebruikte vocoderspraak het toonhoogteverschil uitsluitend effect heeft op een bepaald soort fouten dat de luisteraar maakt bij het nazeggen van de testzinnen, namelijk de weglatingen.

Het is redelijk om aan te nemen dat weglatingen het gevolg zijn van het volstrekt onverstaanbaar worden van de betreffende passages. De kans op volstrekt onverstaanbaar worden is groter bij volledige versmelting dan bij perceptieve scheiding van testzin en maskeertekst.

Daarom nemen we aan dat het percentage weglatingen een maat is voor de mate van versmelting.

De resultaten die verkregen zijn voor stimuli met een toonhoogte van 200 Hz gedragen zich anders dan voor de andere stimuli. Bij geringe geluidssterkteverschillen t.o.v. de maskeertekst zijn de aantallen weglatingen zeer gering, maar bij lagere geluidsniveaus neemt het aantal weglatingen sneller toe dan het geval is bij testzinnen van een andere toonhoogte. Een verklaring hiervoor kan liggen in de spectrale opbouw van deze stimuli. De stimuli van 200 Hz hebben alle harmonische componenten gemeenschappelijk met de maskeerder. Bij een gelijk geluidsniveau van de maskeerder en deze stimuli zijn de harmonische componenten van de laatste gemiddeld 6 dB groter dan de overeenkomstige spectrale componenten van de maskeerder. Daarom zullen deze stimuli bij een relatief geluidsniveau van 0 dB goed te onderscheiden zijn van de maskeerder. Wanneer het geluidsniveau van de stimuli van 200 Hz 6 dB of meer lager wordt t.o.v. de maskeerder, betekent dit dat de spectrale componenten van deze stimuli flink gemaskeerd gaan worden door de overeenkomstige componenten van de maskeerder.

De meetresultaten laten duidelijk zien dat het verschil in toonhoogte tussen twee stemmen een rol speelt bij het perceptief kunnen scheiden van beide spraakuitingen. Naarmate het toonhoogteverschil groter wordt, kunnen beide spraakuitingen beter gescheiden worden. Ook is gebleken dat de

toonhoogte geen onafhankelijke rol speelt bij het scheiden van gelijktijdige spraakuitingen. Er is sprake van een interactie met het relatieve geluidsniveau.

3.4 Discussie en Conclusie

De in dit hoofdstuk besproken experimenten waren erop gericht om de bijdrage te onderzoeken van enkele specifieke factoren op het vermogen van luisteraars om spraakuitingen te verstaan die aangeboden worden tegen een achtergrond van gelijktijdige en interfererende spraak.

Om de verkregen resultaten te kunnen interpreteren zijn enkele veronderstellingen nodig over het verband tussen de verschillende soorten fouten die luisteraars maken bij het reproduceren van de testzinnen en de mogelijke oorzaken van die fouten. De twee soorten fouten die we onderscheiden zijn vervangingen en weglatingen. We zullen ook twee soorten oorzaken onderscheiden, te weten maskering en versmelting. Onder maskering verstaan we de situatie dat het momentane geluid van de testzinnen relatief t.o.v. het momentane geluid van de maskeertekst zo zwak is dat het niet meer gehoord wordt, of althans niet meer tot herkenning van de testzin bijdraagt. Onder versmelting verstaan we de situatie dat het momentane geluid van de testzin en het momentane geluid van de maskeertekst perceptief versmelten tot één nieuw geluid. Dit nieuwe geluid is zeer waarschijnlijk onherkenbaar.

Uitgaande van deze twee soorten fouten en oorzaken zullen we nu achtereenvolgens bespreken het effect van de relatieve geluidsterkte, het effect van de toonhoogteverschillen en het effect van normale toonhoogtecontouren versus quasi-monotone toonhoogtecontouren. Tenslotte zullen we enkele conclusies formuleren.

Het effect van relatieve geluidsterkte

Wanneer we spreken over het relatieve geluidsniveau moeten we steeds in het oog houden dat het geluidsniveau van zowel de testzinnen als van de maskeertekst constant fluctueert tussen stilte, b.v. in de stille intervallen voor plofklanken, en betrekkelijk luide delen van het signaal, b.v.

de klinkers in beklemtoonde lettergrepen. Naarmate het relatieve geluidsniveau van de testzinnen afneemt, neemt de kans op maskering toe. Bij een betrekkelijk geringe kans op maskering zal er van de meeste woorden in de testzinnen nog voldoende ongemaskeerde informatie overblijven om de luisteraar tot herkenning te laten komen, ook al zijn wellicht delen van die woorden zodanig gemaskeerd dat de herkenning niet foutloos verloopt. Wanneer echter de kans op maskering betrekkelijk groot wordt, zal het vaker gebeuren dat een woord geheel of vrijwel geheel gemaskeerd wordt. Dan zal er zo weinig informatie over dat woord de luisteraar bereiken dat hij niet tot herkenning geprikkeld wordt en het betreffende woord weglaat. Op grond van deze overwegingen mogen we een foutengedrag verwachten, dat er ongeveer uitziet als in fig. 3.18.

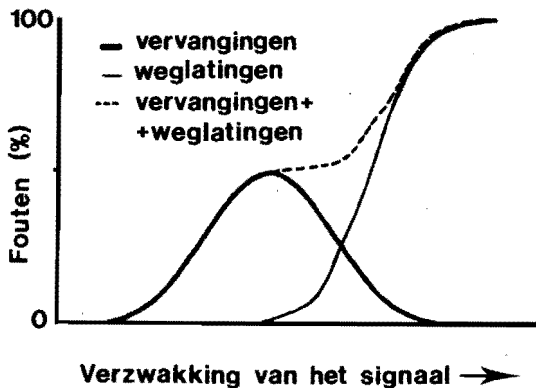


Fig. 3.18 Schematische voorstelling van het vermoedelijk gedrag van de aantallen vervangingen en weglatingen als functie van de afname van het relatieve geluidsniveau van de testzinnen.

Het aantal vervangingen moet na aanvankelijke toename bij afnemende geluidsterkte, noodzakelijk weer afnemen omdat bij zeer lage geluidsniveaus het aantal weglatingen de honderd procent zal naderen. De verdeling van het totaal aantal fouten is, volgens deze beschouwing de som van twee verdelingen die nogal verschillend van vorm zijn. Hiermee is het begrijpelijk dat de verdeling van het totaal

aantal fouten niet een mooie psychometrische kromme volgt, die goed te beschrijven zou zijn met b.v. een cumulatieve normale verdeling, of een andere S-vormige kromme, maar eerder een plateau lijkt te hebben (vgl. fig. 3.2). Zo'n plateau is ook gevonden maar niet verklaard door Egan, Carterette en Thwing (1954).

In fig. 3.5 hebben we gezien dat, althans wanneer er een relatief groot globaal toonhoogteverschil is tussen testzinnen en maskeertekst, het aantal weglatingen vrijwel verwaarloosbaar is bij signaal-maskeerderverhoudingen groter dan -9 dB. Daar beneden neemt het aantal weglatingen zeer sterk toe. We nemen aan dat deze weglatingen vrijwel geheel toe te schrijven zijn aan maskering, en niet aan versmelting.

Het effect van toonhoogteverschillen

In fig. 3.5 hebben we ook gezien dat, wanneer er relatief weinig globaal toonhoogteverschil is tussen testzinnen en maskeertekst, het percentage weglatingen bij signaal-maskeerderverhoudingen groter dan -9 dB verre van verwaarloosbaar is, en zelfs vrijwel even groot kan zijn als het percentage vervangingen.

Deze weglatingen hebben kennelijk iets te maken met het relatief kleine verschil in toonhoogteligging, waarbij de kans op versmelting van gelijktijdige geluiden van testzinnen en maskeertekst relatief groot wordt. We interpreteren daarom deze weglatingen als het gevolg van versmelting van vooral stemhebbende, gelijktijdige fragmenten van testzinnen en maskeertekst. Dus de curves voor weglatingen in de rechter helft van fig. 3.5 zijn, in deze interpretatie, terug te voeren op twee verschillende oorzaken, te weten kans op versmelting voor signaal-maskeerderverhoudingen boven de -9 dB, plus een kans op maskering voor signaal-maskeerder verhoudingen onder de -9 dB.

Overigens lijkt het aannemelijk dat wanneer de toonhoogtes van zowel testzinnen als maskeertekst beide fluctueren, versmelting ten gevolge van momentaan identiek-zijn van de twee toonhoogtes op kan treden bij relatief korte spraak-

fragmenten, zodat het ook kan voorkomen dat een betrekkelijk klein deel van een woord door versmelting onherkenbaar wordt, bij het herkenbaar blijven van de rest. In dat geval zou er voldoende herkenbaars over kunnen blijven om de luisteraar tot herkenning te prikkelen, maar er zou dan gemakkelijk een foute herkenning, dus een vervanging kunnen ontstaan. Dit zou verklaren dat niet alleen het totaal percentage weglatingen, maar ook, zij het in mindere mate, het totaal percentage vervangingen, groter is bij gering globaal toonhoogteverschil (vgl. fig. 3.3 en 3.4). Wanneer de maskeertekst quasi-monotoon gesproken is, zullen de wisselingen in het toonhoogteverschil tussen testzinnen en maskeertekst minder snel verlopen, waardoor de fragmenten waarbij versmelting optreedt, langer zullen zijn dan bij normaal geïntoneerde maskeertekst. Hieruit kunnen we verklaren dat in het experiment met quasi-monotone maskeerder, en vooral wanneer ook de testzinnen quasi-monotoon zijn, het percentage weglatingen aanzienlijk beter discrimineert tussen de toonhoogtecondities dan het percentage vervangingen (vgl. fig. 3.10 en 3.4).

Als we deze redenering, te weten dat de kans op weglatingen toeneemt ten koste van de kans op vervangingen naarmate de duur van de fragmenten waarover versmelting optreedt groter wordt, doortrekken naar het experiment waarbij langs kunstmatige weg de toonhoogtes van testzinnen en maskeertekst constant zijn gemaakt, blijken de daar verkregen resultaten geheel in het te verwachten patroon te vallen.

Immers, wanneer beide toonhoogtes onveranderlijk zijn, dan zal, wanneer er versmelting optreedt, deze zich uitstrekken over de hele spraakuiting. Alleen die delen van de spraakuiting zullen herkenbaar blijven die relatief t.o.v. de maskeertekst, zó luid zijn dat het geluid van de maskeerder, hoewel versmolten met dat van de testzin, niet wezenlijk iets aan het geluid van de testzin verandert. Dit is bijvoorbeeld het geval in de stille intervallen van de maskeertekst, of bij weinig intensieve fricatieven.

De resultaten in fig. 3.15, 3.16 en 3.17 laten dan ook, geheel in de lijn van dit betoog, zien dat bij volstrekt

monotone maskeerder en testzinnen, het percentage weglatingen, met name bij signaal-maskeerderverhoudingen boven de -9 dB afhankelijk is van het toonhoogteverschil tussen maskeerder en testzinnen, en daarmee in onze interpretatie, van de mate van versmelting.

Het percentage vervangingen is ongevoelig voor het toonhoogteverschil, wat verklaard wordt doordat versmelting, wanneer deze optreedt, steeds zo lang zal duren dat er onvoldoende herkenbaars van de woorden overblijft om de luisteraar tot herkenning te prikkelen en dus vrijwel steeds tot weglatingen zal leiden.

Dat het percentage vervangingen, tegen de verwachting in ook betrekkelijk ongevoelig lijkt voor de relatieve geluidssterkte is zeer waarschijnlijk een artefact van de gebruikte vocoderspraak. Ook bij gunstige signaal-maskeerderverhouding treden door de slechte kwaliteit relatief veel fouten in de herkenning op, waardoor het effect van afnemende relatieve geluidssterkte sterk versmeerd wordt. Opmerkelijk is dat bij een constant toonhoogteverschil van 20 Hz het aantal weglatingen nauwelijks toeneemt bij afnemende relatieve geluidssterkte. Dit wijst erop dat bij dit toonhoogteverschil (ca. 3 halve tonen) perceptieve versmelting van simultane geluiden niet of nauwelijks optreedt.

Normale toonhoogtecontour versus quasi-monotone spraak

Wat we nog niet besproken hebben is de positieve bijdrage van de aanwezigheid van normale toonhoogtecontouren in vergelijking met quasi-monotone toonhoogte. Deze kan niet verklaard worden uit verschillen in de kans op versmelting tussen de twee condities.

Twee mogelijke verklaringen dienen zich aan, één die past binnen de lijn van het eerder gevolgde betoog en één van een andere orde. De eerste verklaring gaat uit van de observatie dat in de normaal geïntoneerde testzinnen de vier inhoudswoorden steeds met een accent zijn ingesproken. Een accent komt niet alleen tot uiting in de aanwezigheid van een stijging of daling van de toonhoogte, maar ook in een grotere geluidssterkte t.o.v. overige, niet geaccentu-

eerde woorden. Dit betekent niet dat de geluidsterktes van de inhoudswoorden in normaal geïntoneerde zinnen, waarin deze woorden geaccentueerd zijn, groter zijn dan de geluidsterktes van de inhoudswoorden in de quasi-monotoon gesproken testzinnen, waarin de inhoudswoorden niet geaccentueerd zijn. Dit is niet zo, omdat de dBA waardes van de toppen, en dus in het algemeen van de inhoudswoorden van beide zinstypen, per geluidsterkteconditie gelijk waren. Het betekent wel dat in de normaal geïntoneerde zinnen de inhoudswoorden scherper contrasteerden wat de geluidsterkte betreft met hun onmiddellijke omgeving dan het geval was in de quasi-monotone zinnen. Hierdoor en door de aanwezigheid van toonhoogtebewegingen werd wellicht hun opvallendheid vergroot en de verstaanbaarheid verbeterd. Een verklaring van een andere orde is de volgende. De toonhoogtecontouren van de normaal geïntoneerde testzinnen waren voor al deze testzinnen vrijwel identiek en daardoor ook spoedig aan de luisteraars bekend. Ze konden ook in principe op grond van de gehoorde toonhoogtecontour weten waar de functiewoorden en waar de inhoudswoorden zaten, en tevens waar het begin en waar het eind van iedere zin was. Dit kan de luisteraar aanzienlijke hulp geboden hebben bij het reproduceren van de testzinnen.

Voornaamste conclusies

- Onze voornaamste conclusies uit deze resultaten zijn dat
- (a) een verschil in globale toonhoogteligging de luisteraar in belangrijke mate helpt bij het perceptief scheiden van gelijktijdige stemmen en bij het volgen en verstaan van één van die stemmen,
 - (b) het precieze verschil in toonhoogte tussen gelijktijdige spraakuitingen de luisteraar helpt bij het uit elkaar houden van die geluiden. Daarbij levert een toonhoogteverschil van ca. 3 halve tonen een goede scheiding op,
 - (c) door spraak gemaskeerde spraakuitingen met een normale toonhoogtecontour gemakkelijker te verstaan zijn dan

quasi-monotone spraakuitingen, althans voor zover het geaccentueerde inhoudswoorden betreft.

4. CONCLUSIE

In dit hoofdstuk willen we proberen onze bevindingen nog een keer op een rij te zetten, met elkaar in verband te brengen en voorzover dat niet al in de experimentele hoofdstukken is gedaan, proberen verklaringen te vinden voor bepaalde aspecten van die bevindingen, en tevens enkele kritische kanttekeningen te plaatsen bij de experimenten die gedaan zijn.

4.1 Drie Stappen op Weg naar Auditieve Samenhang

In paragraaf 1.3 hebben we een werkhypothese opgesteld die tot doel had om een leidraad te vormen voor de experimenten. In antwoord op de vraag wanneer het spraaksignaal als auditief samenhangend door de luisteraar waargenomen wordt, hebben we verondersteld dat de luisteraar een drietal taken moet uitvoeren.

Dit zijn:

1. het perceptief al of niet scheiden van gelijktijdige klanken,
2. het al of niet integreren van elkaar opvolgende akoestische gebeurtenissen tot één doorgaande auditieve gebeurtenis,
3. het al of niet inpassen van doorgaande auditieve gebeurtenissen in de toonhoogtecontour.

Daarbij is voorgesteld om te onderzoeken welke factoren van belang zijn om gelijktijdige stemmen van elkaar te kunnen scheiden, en met name wat de rol van de toonhoogte daarbij is. Verder wilden we onderzoeken wat het belang van toonhoogte en spectrale continuïteit is voor het waarnemen van een perceptieve samenhang tussen opeenvolgende akoestische gebeurtenissen en wat de grootte is van de tijdsintervallen waarover deze factoren werkzaam zijn.

We zullen nu proberen onze bevindingen in samenhang te bespreken aan de hand van het denkraam in fig. 4.1, waarin we de bovengenoemde taken die een luisteraar moet uitvoeren om tot auditieve samenhang te komen, nader uitgewerkt hebben. Daarbij zullen we de structuur van het schema volgen, en niet de volgorde waarin de experimenten beschreven zijn

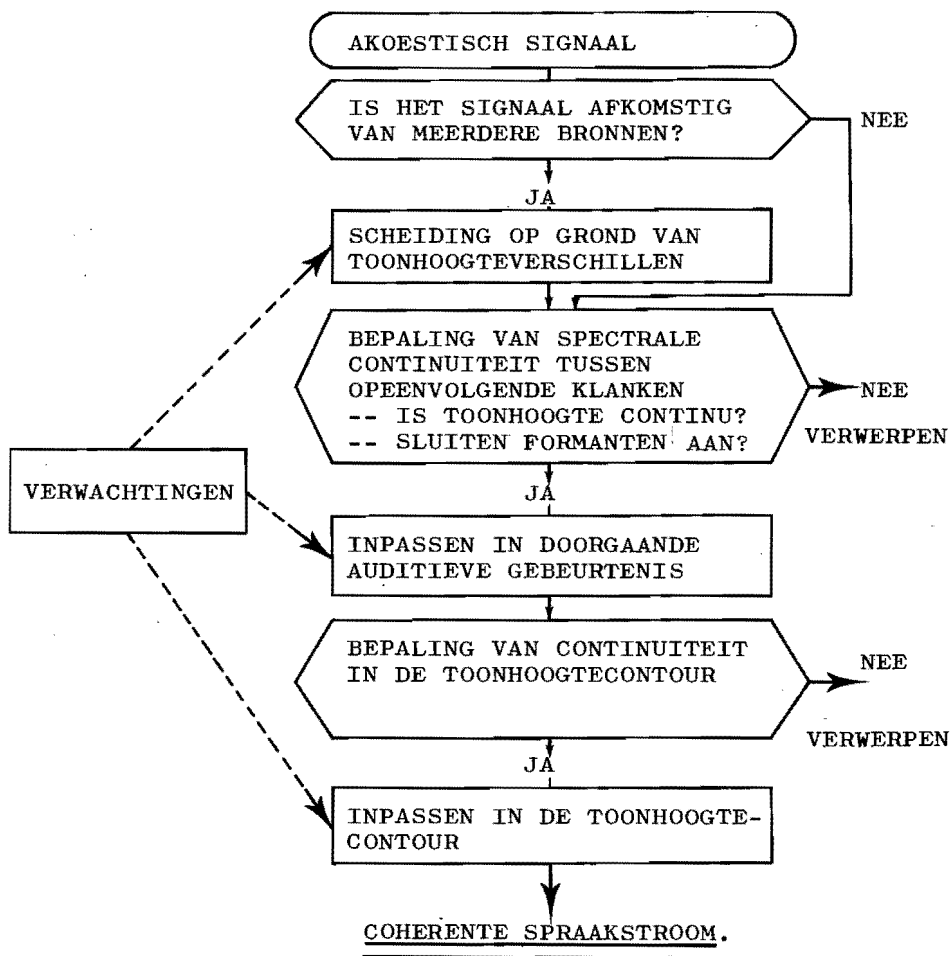


Fig. 4.1 Schematische weergave van beslissingen die een luisteraar moet nemen om het spraaksignaal als auditief samenhangend waar te kunnen nemen.

in de eerdere hoofdstukken.

Het denkraam dat in fig. 4.1 staat weergegeven heeft een opbouw die min of meer opgedrongen is door de werkhypothese en door de vorm van de experimenten uit deze studie. Het heeft niet de pretentie overeenkomstig de werkelijkheid te zijn, maar is bedoeld om een houvast te geven bij de bespreking van de resultaten. Het is volstrekt niet onmogelijk dat de beslissingen die gemakshalve in het betoog afzonderlijk besproken worden, in werkelijkheid afhankelijk en zo met elkaar verweven zijn dat de chronologische structuur die in het schema gesuggereerd wordt, onrealistisch is. De eerste beslissing die de luisteraars moeten nemen is of er sprake is van slechts één spraaksignaal of dat er sprake is van meerdere, gelijktijdige spraaksignalen. Bij de aanwezigheid van meerdere, gelijktijdige spraaksignalen, zal de luisteraar deze perceptief moeten scheiden. Dit is theoretisch mogelijk door gebruik te maken van de harmonische structuur van de spraakklanken.

In overeenstemming met Goldstein (1973) kunnen we aannemen dat in de waarneming de toonhoogte van een geluid bepaald wordt door het vergelijken van de harmonische structuur van dat geluid met die van een intern gegenereerde "harmonische kam". Deze laatste wordt zo gekozen dat een zo goed mogelijke aanpassing verkregen wordt in termen van een "maximale waarschijnlijkheidsschatting". De grondtoon van de harmonische kam is dan de waargenomen toonhoogte. Wanneer er twee of meer geluiden met verschillende toonhoogtes tegelijk aanwezig zijn, zullen na de bepaling van een eerste toonhoogte op grond van de best passende harmonische kam, nog een groot aantal "verworpen" harmonischen overblijven. Wij kunnen ons voorstellen dat op deze verworpen componenten dezelfde procedure nogmaals wordt toegepast, zodat ook van een tweede geluid de toonhoogte wordt bepaald door aanpassing van een nieuwe harmonische kam. Het is denkbaar dat deze procedures voor de bepaling van een of meer toonhoogtes beïnvloed kunnen worden door de verwachtingen van de luisteraars op grond van eerdere waarnemingen (vgl. Gerson and Goldstein, 1978). Dan is

het niet strikt noodzakelijk te veronderstellen dat per kortdurend segment steeds meer dan één toonhoogte bepaald wordt. Hoe dit ook zij, wanneer de gelijktijdige geluiden identieke toonhoogtes hebben, zullen zij versmelten tot één geluid, doordat alle harmonischen van beide geluiden zullen samenvallen met componenten van één harmonische kam. Deze geluiden zijn waarschijnlijk niet herkenbaar, of kunnen niet veel bijdragen aan de herkenning. In de experimenten aan het scheiden van gelijktijdige spraakuitingen (hoofdstuk 3) blijken, overeenkomstig deze verwachting, toonhoogteverschillen tussen gelijktijdige spraakfragmenten een positief effect te hebben op het vermogen van de luisteraars om één van de spraakuitingen te verstaan. De resultaten van deze, in zekere zin exploratieve experimenten laten niet toe de "harmonische kam"-theorie meer in detail te toetsen. De experimenten met volstrekt onveranderlijke toonhoogtes (vgl. par. 3.3) laten wel zien dat het scheiden van stemmen niet een alles of niets gebeuren is, maar in eerste benadering beter gaat naarmate het toonhoogteverschil groter is. Dat bij een toonhoogteverschil van precies een octaaf de scheiding weer slecht gaat, is kwalitatief in overeenstemming met de "harmonische kam"-theorie.

De tweede beslissing die een luisteraar moet nemen volgens het schema in fig. 4.1 is of onmiddellijk op elkaar volgende akoestische gebeurtenissen een doorgaande auditieve gebeurtenis vormen. Het voornaamste argument om deze stap in onze hypothesen op te nemen (vgl. par. 1.4) is ontleend aan de experimenten van Huggins met temporeel gesegmenteerde spraak (vgl. par. 1.3) en het effect van de continuïteit in de formantstructuur in de volgorde-perceptie experimenten van Dorman et al. (par. 1.2). In de experimenten beschreven in hoofdstuk 2, die in hoofdzaak gericht waren op het belang van de continuïteit van de toonhoogtecontour voor de auditieve samenhang, hebben we tevens een mogelijke bijdrage van formantbuigingen op de waargenomen samenhang onderzocht. We hebben in strijd met de verwachtingen, geen duidelijk effect van de aan- of afwezigheid van formant-

buigingen op het horen van één of twee spraakstromen kunnen vinden. Indien de aanwezig lijkende tendens tot betere samenhang bij de aanwezigheid van fonetisch mogelijke formantbuigingen echt is, dan is dit effect toch marginaal t.o.v. dat van de toonhoogte. De conclusie lijkt gerechtvaardigd dat formantbuigingen géén of hooguit een marginale rol spelen bij het horen van één of twee spraakstromen. Daarbij moeten echter twee kanttekeningen gemaakt worden. Ten eerste kan spectrale continuïteit anders opgevat worden dan als continuïteit in de formantstructuur. Ook de harmonische structuur van de opeenvolgende spraakgeluiden bepaalt de spectrale continuïteit en het is denkbaar dat in onze stimuli deze vorm van spectrale continuïteit, t.w. het opelkaar aansluiten in frequentie en amplitude van de harmonische componenten in opeenvolgende auditieve gebeurtenissen, voor de perceptie veel belangrijker was dan de continuïteit van de formanten. We hebben echter in onze experimenten niet de continuïteit van de harmonische structuur, zoals wel gedaan is voor niet-spraakgeluiden door Van Noorden (1975) (vgl. ook par. 1.3), maar de continuïteit van de toonhoogte als aparte experimentele variabele gebruikt. Het is redelijk om de mogelijkheid open te laten dat in onze experimenten van hoofdstuk 2, met name bij afwezigheid van stille intervallen of bij zeer korte stille intervallen, de spectrale continuïteit van de harmonische structuur belangrijker was dan de continuïteit in de toonhoogte. We zullen hieronder bij de bespreking van de derde stap in de richting van auditieve samenhang, nog terugkomen op de rol van de continuïteit in de harmonische structuur. Een tweede kanttekening bij de afwezigheid van een effect van formantbuigingen op de auditieve samenhang is dat dit samen kan hangen met de specifieke taak van de luisteraars in deze experimenten, te weten aan te geven of zij één of twee sprekers hoorden. Wellicht is auditieve samenhang minder een primair en homogeen verschijnsel dan wij gedacht hebben en zijn verschillende aspecten van auditieve samenhang van belang voor verschillende herkenningstaken. Dit zou kunnen verklaren waarom wél een effect van formantstructuur

is gevonden op de waarneming van volgordes (Dorman et al. 1975) en op de splitsing van de spraakstroom in klinkers en medeklinkers (Cole and Scott, 1973) en niet, in onze experimenten, op het horen van één of twee sprekers. In ieder geval lijkt het ons dat de negatieve resultaten in hoofdstuk 2 met betrekking tot formantbuigingen onvoldoende reden zijn om de notie "doorgaande auditieve gebeurtenis" uit het schema in fig. 4.1 te schrappen. Ook bij deze tweede beslissing hebben we weer in het schema tot uiting gebracht dat het redelijk is om aan te nemen dat het inpassen in een doorgaande auditieve gebeurtenis niet uitsluitend door de stimulus bepaald wordt, maar ook beïnvloed kan worden door de verwachtingen van de luisteraar.

De derde beslissing die de luisteraar volgens het schema in fig. 4.1 moet nemen is of de doorgaande auditieve gebeurtenissen wel of niet passen in de toonhoogtecontour. De experimenten in hoofdstuk 2 hebben in eerste instantie hier betrekking op. De resultaten van deze experimenten laten zien dat, wanneer opeenvolgende auditieve gebeurtenissen gescheiden zijn door pauzes korter dan 100 ms, d.w.z. in de orde van grootte van wat normaal is voor de stille periodes van plofklanken in spraak, het wel of niet passen in de toonhoogtecontour vrijwel uitsluitend wordt bepaald door het verschil in toonhoogte. Er is dan nauwelijks invloed van de duur van het stille interval. Dit is voor een situatie zonder stille intervallen maar met hoorbare toonhoogteverglijdingen nog eens bevestigd in een afzonderlijk experiment (paragraaf 2.6). De helling van de toonhoogteverglijdingen had geen effect, het verschil tussen de opeenvolgende toonhoogteniveaus wél. Dit is anders voor pauzuren groter dan 100 ms. In dat geval is er een duidelijk effect van de duur van de pauzes, in die zin dat naarmate de pauze langer wordt een grotere toonhoogtesprong noodzakelijk is om de reeks auditieve gebeurtenissen op te splitsen in twee subreeksen die van twee verschillende sprekers afkomstig lijken te zijn. Het is in het licht van wat eerder gezegd is over de mogelijke rol van de

continuïteit van de harmonische structuur, verleidelijk om aan te nemen dat deze resultaten het gevolg zijn van twee verschillende auditieve processen: één voor pauzeduren korter dan 100 ms, en één voor pauzeduren langer dan 100 ms. Het eerste auditieve proces zou dan leiden tot waargenomen samenhang op grond van de spectrale continuïteit, het tweede tot inpassen in de toonhoogtecontour. De werking van de spectrale continuïteit zou zich uitstrekken tot zo'n 100 ms, wat in overeenstemming is met de eerder aangehaalde gegevens van Huggins en Van Noorden. De werking van de toonhoogtecontour zou zich uitstrekken over op zijn minst enkele honderden millisecondes.

Buiten deze verklaring van de data is het begrijpelijk dat zolang de spectrale continuïteit invloed heeft, een geringe verschuiving in toonhoogte, ongeacht de pauzeduur, al leidt tot splitsing, omdat immers een geringe verschuiving in de grondtoon snel tot aanzienlijke verschuivingen in de harmonische structuur leidt. Dat desondanks een verschuiving van enkele halve tonen toelaatbaar is, zou er misschien op kunnen duiden dat vooral de grondtoon en de laagste harmonische componenten, waar de verschuivingen in het frequentiedomein het kleinst zijn, zwaar meetellen. Boven de 100 ms is het niet het verschil in toonhoogte dat telt, maar de helling van de toonhoogtebeweging, gerekend vanaf het moment dat de spectrale afbeelding van het vorige geluid, de spectrale afbeelding dus waarop de toonhoogtegebeurtenis berustte, uitgestorven is, en alleen informatie over die toonhoogte bewaard is gebleven. Dat wil zeggen gerekend vanaf zo'n 100 ms na het einde van dit geluid. De waargenomen toonhoogte heeft kennelijk een veel langere overlevingstijd dan de spectrale informatie en kan daardoor perceptief duidelijke contouren, melodieën vormen van aanzienlijke lengte. Dat het wel of niet passen van een auditieve gebeurtenis in zo'n contour afhangt van de helling van de virtuele toonhoogtebeweging is in ieder geval intuïtief bevredigend.

Weer hebben we, bij deze derde beslissing, in het schema aangegeven dat de verwachtingen van de luisteraar een invloed kunnen uitoefenen op de beslissing of een bepaalde auditieve gebeurtenis wel of niet past in de toonhoogtecontour. In de experimenten van hoofdstuk 2 hebben we de verwachtingen niet als experimentele variabelen gehanteerd. In zekere zin hebben we dit wél gedaan in de experimenten in hoofdstuk 3 aan perceptieve scheiding van stemmen, daar waar we normale en voor alle betreffende spraakuitingen vrijwel identieke toonhoogtecontouren hebben geplaatst tegenover abnormale, quasi-monotone toonhoogtecontouren. We hebben echter gezien dat de positieve bijdrage van de normale toonhoogtecontouren aan de verstaanbaarheid niet noodzakelijk in verband hoeft te worden gebracht met het waargenomen toonhoogteverloop, maar ook verklaard kan worden uit de grotere variaties in geluidssterkte. Ook in dit geval moeten we dus concluderen dat onze experimenten over de rol die de verwachtingen van de luisteraar spelen in de auditieve samenhang geen uitsluitel geven. Ons onderzoek is beperkt gebleven tot de bijdrage van enkele stimulusfactoren aan de auditieve samenhang van spraak.

4.2 Enkele Kritische Kanttekeningen en Suggesties voor Verder Onderzoek

In de loop van het onderzoek dat in deze studie is beschreven, is nogal wat tijd gaan zitten in het zoeken naar experimentele paradigma's die geschikt waren om de uitgangsvragen te beantwoorden. Achteraf gezien is er alleszins ruimte om enkele kritische kanttekeningen te plaatsen bij de gekozen oplossingen. Deze kanttekeningen zijn van tweeërlei aard, ten eerste betrekking hebbend op de gebruikte methodes, ten tweede betrekking hebbend op de relatie tussen de verkregen resultaten en de uitgangsvraag van het onderzoek.

Deze kritische kanttekeningen leiden als vanzelf tot suggesties voor verder onderzoek. We zullen eerst de experimenten in hoofdstuk 2 en dan die in hoofdstuk 3 op deze wijze bespreken.

De methode die we gevolgd hebben in de experimenten van hoofdstuk 2, gericht op de akoestische factoren die bijdragen tot auditieve samenhang van reeksen spraakachtige gebeurtenissen, bracht met zich mee dat binnen iedere stimulus een reeks van drie opeenvolgende, herkenbare, klinkersegmenten (al of niet met formantbuigingen die de aanwezigheid van medeklinkers suggereren) tot driemaal toe herhaald werd en dat binnen die reeks een strikte afwisseling van twee toonhoogtes gebruikt werd. Een voordeel van deze methode is dat bij auditieve splitsing twee verschillende subreeksen van herkende klinkers ontstaan. Men kan zich echter afvragen of de splitsing die verkregen wordt niet een artefact is van de strikte herhaling van de toonhoogtes in iedere subreeks. In eerste verkenningen bleek dat bij kortere reeksen veel moeilijker auditieve splitsing te krijgen is. In natuurlijke luistercondities komt althans in spraak een dergelijke strikte afwisseling van toonhoogtes niet voor. Bovendien is niet bekend hoe lang de reeks moet zijn om een verzadiging te krijgen van het effect van strikte herhaling van toonhoogtes in iedere subreeks. Het is dus mogelijk dat de kwantitatieve eigenschappen van onze gegevens in strikte zin alleen gelden voor dit soort reeksen van negen klinkers.

Een tweede bezwaar dat aangevoerd kan worden tegen deze methode is dat de gebruikte stimuli, ook in andere opzichten dan hierboven beschreven, wel erg ver af staan van natuurlijke spraak. Er is geen sprake van taalstructuur in de stimuli anders dan dat er klinkers en eventueel medeklinkers in herkend kunnen worden. Het is niet ondenkbaar dat in stimuli met een meer talige structuur en daardoor een grotere redundantie in informatie, auditieve samenhang veel minder gewicht krijgt bij het beoordelen of de spraak die men hoort van één of van twee sprekers afkomstig is. De linguïstische samenhang kan wel eens minstens zo belangrijk of zelfs belangrijker zijn dan de door ons onderzochte stimulusfactoren.

Ten derde hebben wij niet alle daarvoor in aanmerking komende stimulusfactoren onderzocht. Het lijkt niet onaannemelijk,

gezien de resultaten verkregen met zuivere tonen in experimenten van Van Noorden (1975), dat ook afwisselingen in geluidsterkte, bepalend zijn voor de auditieve samenhang. Wat betreft de formantbuigingen was het, gezien de resultaten, wellicht verstandiger geweest om ook een stimulusconditie te bestuderen waarin de continuïteit van de formanten optimaal was, dus continue formantverglidingen van klinker naar klinker. Als we in zo'n conditie geen positieve bijdrage van de continuïteit in de formantstructuur hadden gemeten, zou onze conclusie dat de formantstructuur geen of weinig effect heeft op het horen van één of twee sprekers onder onze experimentele condities, sterker zijn geweest.

Wanneer men in de toekomst verder zou willen gaan met experimenten van het type dat in hoofdstuk 2 beschreven is, bijvoorbeeld om luidheid en formantverglidingen nader te onderzoeken, zou het ook aanbeveling verdienen om te onderzoeken of de hier behaalde resultaten overeind blijven, wanneer strikte herhaling van toonhoogtes in iedere subreeks wordt vermeden. Dit kan op ten minste twee manieren: ten eerste door een afwisseling aan te brengen tussen toonhoogtegebieden i.p.v. een afwisseling tussen strikte toonhoogtes, ten tweede door de afwisseling plaats te doen vinden tussen een lage en een hoge, langzaam aflopende toonhoogtelijn, bij voorkeur met een helling die overeenkomt met de zogenaamde "declinatielij", die de basis vormt voor veel natuurlijke intonatiepatronen. Verder zou het ons inziens de moeite waard zijn om toonhoogte en spectrale aansluiting van de harmonische structuur los van elkaar te variëren, al zal dat onvermijdelijk met zich meebrengen dat de stimuli nog verder van spraak af komen te liggen. Het moet mogelijk zijn om met behoud van de herkenbaarheid van klinkers, de stimuli zo te kiezen dat opeenvolgende klinkers geen harmonische componenten gemeenschappelijk hebben, maar wel dezelfde toonhoogte.

Ook bij de experimenten in hoofdstuk 3 aan het scheiden van stemmen zijn kritische kanttekeningen te plaatsen. Hoewel

de door ons gekozen, op Nakatani et al. (1973) gebaseerde meetmethode, goed aan zijn doel lijkt te beantwoorden, zouden we wellicht deze meetmethode nog wel iets beter hebben kunnen toepassen. Zo hebben wij er niet erg op gelet of het spraaktempo in de maskerende tekst en in de testzinnen wel hetzelfde was. In de praktijk heeft dat ertoe geleid dat het spraaktempo in de testzinnen wat lager was dan in de maskeertekst. Het is onbekend wat het effect is van spraaktempo op de perceptieve scheiding van stemmen en in de toekomstige experimenten zou men er goed aan doen dit nader te onderzoeken. Verder is het te beredeneren in het licht van onze resultaten (vgl. par. 3.2.1), dat hoe sterker de fluctuaties in toonhoogte en geluidssterkte in de maskeertekst zijn, hoe groter de spreiding in de verstaanbaarheidsscores is. Om de meetmethode nauwkeuriger te maken verdient het aanbeveling de maskeertekst wat betreft intonatie en geluidssterkte zo gelijkmatig mogelijk te laten inspreken. Wij zijn pas tijdens de experimenten op het idee gekomen de fouten tegen de herkenning op te splitsen in vervangingen en weglatingen. Hierdoor waren we op een kwalitatieve wijze in staat een mogelijke verklaring te vinden voor de door ons zelf en door anderen gevonden plateaus in het verloop van het totaal aantal fouten als functie van de relatieve geluidssterkte tussen de testzinnen en de maskeertekst. Een nauwkeurigerkwantitatieve toetsing van deze veronderstelling was niet mogelijk door de grote spreiding van de data en door het ontbreken van een model voor de precieze relatie tussen de verzameling vervangingen en de verzameling weglatingen. Ook dit wacht op verder onderzoek. Het lijkt denkbaar dat zo'n onderzoek een natuurlijke verklaring zou bieden voor de lichte verbetering van de verstaanbaarheid die Egan, Carterette en Thwing (1954) hebben gevonden voor geringe signaal-maskeerder geluidssterkteverhoudingen, kleiner dan 0 dB t.o.v. relatieve geluidssterktes van precies 0 dB. In onze experimenten aan perceptieve scheiding van stemmen zijn een aantal factoren buiten beschouwing gebleven. Wij willen het daarbij niet hebben over richting horen en linguïstische structuur, omdat deze van begin af aan met opzet

buiten onze vraagstelling zijn gehouden. Maar verschillen in timbre, zoals overeenkomend met verschillen tussen sprekers, de eerder genoemde verschillen in spreektempo en andere verschillen in temporele organisatie van spraak, zoals die tussen sprekers kunnen bestaan, hadden zeker in aanmerking kunnen komen om in het kader van dit onderzoek op hun bijdrage aan perceptieve scheiding onderzocht te worden. Wij zijn daar niet aan toegekomen en hebben deze vragen overgelaten voor toekomstig onderzoek.

Ons onderzoek heeft ook de vraag opgeroepen, of beter, opnieuw de aandacht gericht op de vraag welke rol de toonhoogteperceptie speelt in het scheiden van volstrekt gelijktijdige spraakgeluiden. In dit opzicht zijn onze metingen te grof en te veel exploratief om precieze antwoorden te kunnen geven. Een psychoakoestisch onderzoek zou hier op zijn plaats zijn om na te gaan in hoeverre bestaande theorieën van toonhoogteperceptie kunnen helpen het vermogen van luisteraars om gelijktijdige geluiden afzonderlijk waar te nemen, te begrijpen en te verklaren.

Tenslotte wellicht de belangrijkste omissie van ons onderzoek. Wij zijn er niet aan toe gekomen het belang van auditieve samenhang voor de perceptieve scheiding van stemmen te onderzoeken. Uit enkele verkennende metingen (par. 2.2) is ons gebleken dat wanneer er geen interfererende spraak aanwezig is, auditieve samenhang niet strikt noodzakelijk is voor de verstaanbaarheid van spraak, althans wanneer die spraak voldoende linguïstisch redundant is. Het lijkt zeer waarschijnlijk dat dit drastisch verandert wanneer wel interfererende spraak aanwezig is. Een logisch vervolg op onze experimenten, noodzakelijk om de brug te slaan tussen de experimenten uit hoofdstuk 2 en die uit hoofdstuk 3, zou zijn om die laatste experimenten uit te breiden met stimuluscondities waarin de auditieve samenhang langs kunstmatige weg verstoord is, hetzij door spraakuitingen samen te stellen door woorden of lettergrepen van verschillende sprekers aan één te rijgen, hetzij door met behulp van een analyse-resynthesesysteem in spraakuitingen van één spreker kunstmatig discontinuïteiten

in toonhoogte, formantstructuur en geluidssterkte aan te brengen. Onze verwachting is dat in zulke experimenten de auditieve samenhang van wezenlijk belang zal blijken voor het spraakverstaan.

SAMENVATTING

Spraak bestaat uit reeksen elkaar snel opvolgende geluiden die voor de luisteraars herkenbare patronen vormen waaraan hij betekenissen kan hechten. Een eerste voorwaarde waaraan spraak moet voldoen, is dat de luisteraar in staat is om de geluiden die de spreker voortbrengt als bij elkaar horend waar te nemen, zonder dat er onzekerheid bestaat over de onderlinge tijdrelaties tussen de verschillende akoestische gebeurtenissen. Deze eigenschap van spraak wordt aangeduid als auditieve samenhang. Het doel van deze studie is de relatie te onderzoeken tussen de fysische eigenschappen van spraak en de auditieve samenhang.

Als een signaal als auditief niet-samenhangend wordt waargenomen, ontstaat de indruk dat de klanken uit verschillende bronnen afkomstig zijn. Vaak zal dat ertoe leiden dat de volgorde van opeenvolgende akoestische gebeurtenissen slecht wordt waargenomen. Ook kunnen de luisteraars vaak niet horen of akoestische gebeurtenissen die elkaar in werkelijkheid opvolgen, wel of niet gelijktijdig zijn. In het geval dat akoestische gebeurtenissen gelijktijdig zijn en als één klank waargenomen worden, spreekt men van fusie of versmelting. Wanneer zij gehoord worden als twee of meer geluiden spreekt men van splitsing.

In het eerste hoofdstuk wordt een overzicht gegeven van gegevens uit de literatuur over de waarneming van volgordes, auditieve splitsing, de verstaanbaarheid van onderbroken en temporeel gesegmenteerde spraak en factoren die van belang zijn voor het perceptief scheiden van gelijktijdige stemmen.

Aan de hand van deze gegevens is een werkhypothese opgesteld die inhoudt dat de notie auditieve samenhang het resultaat is van een drietal sorteerprocessen.

1. Het perceptief al of niet scheiden van gelijktijdige klanken,

2. Het al of niet integreren van op elkaar volgende akoestische gebeurtenissen tot één doorgaande auditieve gebeurtenis,
3. Het al of niet inpassen van doorgaande auditieve gebeurtenissen in de toonhoogtecontour.

Het experimentele gedeelte van deze studie bestaat uit twee reeksen experimenten waarin de auditieve samenhang van reeksen spraakfragmenten met alternerende toonhoogte bestudeerd wordt en uit experimenten over de perceptieve scheiding van gelijktijdige spraakuitingen. Deze experimenten staan respectievelijk beschreven in hoofdstuk 2 en hoofdstuk 3.

In hoofdstuk 2 zijn de continuïteit van de toonhoogte, temporele organisatie en spectrale verglijdingen onderzocht op hun belang voor het waarnemen van stemcontinuïteit in reeksen klinkers en reeksen betekenisloze lettergrepen. De opdracht van de luisteraars was om aan te geven of de reeks klanken die zij hoorden door één of door twee sprekers was voortgebracht. Het toonhoogteverloop blijkt de voornaamste factor te zijn. Geringe toonhoogtesprongen in de orde van 5 halve tonen kunnen een reeks klanken doen uiteen vallen in twee perceptief gescheiden stromen van klanken die aan verschillende sprekers kunnen worden toegewezen. Bij duren van de stille intervallen tussen de klanken groter dan 100 ms, bleek toename ervan een duidelijk positief effect te hebben op de waargenomen stemcontinuïteit. De duur van de klanken evenals de aanwezigheid van spectrale verglijdingen bleken in deze experimenten geen effecten te hebben.

In hoofdstuk 3 is onderzocht wat het belang is van de relatieve ligging van de toonhoogte en de aan- of afwezigheid van een normale toonhoogtecontour, voor de mogelijkheid om spraakuitingen te volgen en te verstaan wanneer deze spraakuitingen gemaskeerd worden door spraak.

Dit is gedaan in een serie verstaanbaarheidsmetingen, waarbij het spraakmateriaal bestond uit zinnen die qua woordinhoud hoogst onwaarschijnlijk waren, maar alle eenzelfde eenvoudige grammaticale structuur vertoonden.

De experimentele variabelen waren relatieve toonhoogte en normale intonatie versus quasi-monotone toonhoogtecontouren. De afhankelijke variabele was het percentage fout herkende woorden. De resultaten laten zien dat zowel een verschil in toonhoogte tussen de testzinnen en de maskeertekst, als de aanwezigheid van een normale toonhoogtecontour een positief effect heeft op de verstaanbaarheid. Het effect van het verschil in toonhoogte tussen de testzinnen en de maskeertekst is preciezer onderzocht in een experiment waarbij zowel de testzinnen als de maskeertekst met behulp van een LPC vocodersysteem volstrekt monotoon zijn gesynthetiseerd. Geringe verschillen in toonhoogte, in de orde van 3 halve tonen, leiden al tot een aanzienlijke verhoging van de verstaanbaarheid.

In hoofdstuk 4 is geprobeerd om de gegevens uit de literatuur en de resultaten uit de experimenten met elkaar in verband te brengen. Dit is gedaan aan de hand van de in het eerste hoofdstuk opgestelde werkhypothese.

Tot slot bevat dit hoofdstuk enige kritische kanttekeningen en suggesties voor verder onderzoek.

SUMMARY

Speech consists of sequences of sounds which must form recognizable patterns to the listener and to which he can attach meanings.

A first condition speech has to fulfil is that a listener should be able to perceive the different sounds produced by the speaker as belonging together. There should be no uncertainty about the temporal order of various acoustic events. This property of the speech signal is called auditory coherence.

The purpose of this study is to examine the relationship between the physical properties of the speech signal and auditory coherence. If a signal is perceived as incoherent, a listener receives the impression that several speakers are involved, even though the words have been read by just one speaker. It may even happen that the order of the words cannot be distinguished correctly. Furthermore, listeners are often unable to hear whether successive acoustic events are simultaneous or not. The case in which simultaneous acoustic events are perceived as one sound will be called fusion. When the acoustic events are heard as two or more sounds, the term auditory segregation is used.

Chapter 1 reviews the literature on the perception of temporal order, auditory stream segregation, the intelligibility of temporally interrupted and temporally segmented speech and factors influencing perceptual separation of simultaneous voices.

On these data a working hypothesis is based to the effect that the notion of auditory coherence is supposed to be the result of three selective processes.

1. Whether or not to separate perceptually simultaneous sounds;
2. Whether or not to integrate successive acoustic events as one continuous auditory event;
3. Whether or not to fit the continuous auditory events into the pitchcontour.

The experimental part of this study is divided into two sets of experiments: one in which the auditory coherence of series of vowels or syllables with an alternating pitch is studied (chapter 2), and one dealing with the perceptual separation of simultaneous utterances (chapter 3).

In chapter 2 various acoustic attributes of the speech signal are investigated: the continuity of pitch, temporal organisation and spectral transitions in the light of their possible contribution to the perception of voice continuity in series of vowels and pseudo syllables. The listeners' instruction was to state whether series of sounds seemed to be produced by one or by two speakers. The pitch contour appeared to be the most important factor. Small pitch differences of about 5 semitones may lead to segregating a series of sounds into two perceptual sub-sequences of sounds, which may be assigned to different speakers. For durations of the silent intervals between the sounds longer than 100 ms, a clearly positive effect of silent interval-duration was found on the perceived voice continuity. Neither the duration of the sounds nor the presence of spectral transitions appeared to have any effect in these experiments.

Chapter 3 deals with the importance of the relative level of pitch as opposed to the pitch of the masker and the presence or absence of a normal pitch contour for following and understanding utterances masked by speech. A series of intelligibility measurements were made with speech material consisting of sequences with highly improbable contents within one and the same grammatical framework. The results in percentages of words correct show that a difference in pitch between the target sentences and the masker as well as the presence of a normal pitch contour increases intelligibility.

The effect of the difference in pitch level between the target sentences and the masker was investigated more precisely in an experiment in which the target sentences as well as the masker were given a completely monotonous pitch

through an LPC vocoder system. Small differences in pitch of about 3 semitones were sufficient for a considerable increase of intelligibility.

Chapter 4 attempts to connect data from the literature with the experimental results obtained on the basis of the hypothesis set up in chapter 1.

Finally, chapter 4 also contains some critical remarks and suggestions for further research.

LITERATUUR

- Bregman, A.S. and J. Campbell: Primary auditory stream segregation and perception of order in rapid sequences of tones, *Journal of Exp. Psychology*, 89(2), 1971, 244-249.
- Bregman, A.S. and G.L. Dannenbring: The effect of continuity on auditory stream segregation, *Perception & Psychophysics*, 13(2), 1973, 308-312.
- Broadbent, D.E. and P. Ladefoged: On the fusion of sounds reaching different sense organs, *J. Acoust. Soc. Amer.*, 29, 1957, 708-710.
- Cherry, E.C.: Some experiments on the recognition of speech, with one and with two ears, *J. Acoust. Soc. Amer.*, 25, 1953, 975-979.
- Cherry, E.C. and W.K. Taylor: Some further experiments upon the recognition of speech, with one and with two ears, *J. Acoust. Soc. Amer.*, 26, 1954, 554-559.
- Cole, R.A. and B. Scott: Perception of temporal order in speech: the role of vowel transitions, *Canad. J. Psychol. Rev. Canad. Psychol.* 27(4), 1973, 441-449.
- Darwin, C.J.: On the dynamic use of prosody in speech perception, In: A. Cohen and S.G. Nooteboom, *Structure and Process in Speech Perception*, Springer-Verlag, Berlin-Heidelberg-N.Y., 1975, 178-193.
- Darwin, C.J. and C.E. Bethell-Fox: Pitch continuity and speech source attribution, *J. of Exp. Psych.: Human Perception and Performance*, 3(4), 1977, 665-672.
- Dorman, M.F., J.E. Cutting and L.J. Raphael: Perception of temporal order in vowel sequences with and without formant transitions, *J. of Exp. Psych.: Human Perception and Performance*, 104(2), 1975, 121-129.
- Egan, J.P., E.C. Carterette and E.J. Thwing: Some factors affecting multi-channel listening, *J. Acoust. Soc. Amer.*, 26(5), 1954, 774-782.
- Everton, J.: The separation of the voice signals of simultaneous speakers, 1975, *Proefschrift Univ. of Utah* (niet gepubliceerd).
- Frazier, R.H., S. Samsam, L.D. Braida and A.V. Oppenheim: Enhancement of speech by adaptive filtering, in: *Procedures of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, N.Y., 1976, 251-253.

- Gerson, A. and J.L. Goldstein: Evidence for a general template in central optimal processing for pitch of complex tones, *J. Acoust. Soc. Amer.*, 63(2), 1978, 498-510.
- Goldstein, J.L.: An optimum processor theory for the central formation of the pitch of complex tones, *J. Acoust. Soc. Amer.*, 54, 1973, 1496-1516.
- Huggins, A.W.F.: Distortion of the temporal pattern of speech: interruption and alternation, *J. Acoust. Soc. Amer.*, 36, 1964, 1055-1064.
- Huggins, A.W.F.: Second experiment on temporally segmented speech, Research Laboratory of Electronics, Massachusetts Institute of Technology, QPR 106, 1972a, 137-140.
- Huggins, A.W.F.: Perception of temporally segmented speech, Proceedings of the 7th International Congress of Phonetic Sciences, Montreal, 1971, A. Rigault and R. Charbonneau, Den Haag, Mouton, 1972b.
- Huggins, A.W.F.: More temporally segmented speech: is duration or speech content the critical variable in its loss of intelligibility, Research Laboratory of Electronics, Massachusetts Institute of Technology, QPR 114, 1974a, 185-193.
- Huggins, A.W.F.: On perceptual integration of dichotically alternated pulse trains, *J. Acoust. Soc. Amer.*, 56, 1974b, 939-943.
- Huggins, A.W.F.: Temporally segmented speech, *Perception & Psychophysics*, 18(2), 1975a, 149-157.
- Huggins, A.W.F.: Temporally segmented speech and echoic storage, In: A. Cohen and S.G. Nootboom, Structure and Process in Speech Perception, Springer-Verlag, Berlin-Heidelberg-N.Y., 1975b, 209-225.
- Lackner, J.R. and L.H. Goldstein: Primary auditory stream segregation of repeated consonant-vowel sequences, *J. Acoust. Soc. Amer.*, 56(5), 1974, 1651-1652.
- Miller, G.A. and G.A. Heise: The trill threshold, *J. Acoust. Soc. Amer.*, 22, 1950, 637-638.
- Miller, G.A. and J.C.R. Licklider: The intelligibility of interrupted speech, *J. Acoust. Soc. Amer.*, 22(2), 1950, 167-173.
- Nakatani, L.H. and K.D. Dukes: A sensitive test of speech-communication quality, *J. Acoust. Soc. Amer.*, 53(4), 1973, 1083-1091.
- Noorden, L.P.A.S. van: Temporal coherence in the perception of tone sequences, Proefschrift Technische Hogeschool Eindhoven, 1975.

- Norman, D.A.: Memory and attention, 2nd ed. New York etc., Wiley, 1976.
- Parsons, T.W.: Separation of speech from interfering speech by means of harmonic selection, *J. Acoust. Soc. Amer.*, 60, 1976, 911-918.
- Powers, G.L. and Ch. Speaks: Intelligibility of temporally interrupted speech, *J. Acoust. Soc. Amer.*, 54, 1973, 661-667.
- Stumpf, C.: *Tonpsychologie*, Lizenzausgabe des S. Hirzel-Verlages Leipzig 1890, Knuf-Bonset, Hilversum-Amsterdam, 1965.
- Thomas, I.B., P.B. Hill, E.S. Carroll and B. Garcia: Temporal order in the perception of vowels, *J. Acoust. Soc. Amer.*, 48, 1970, 1010-1013.
- Treisman, A.M.: Contextual cues in selective listening, *Quarterly Journal of Experimental Psychology*, 12, 1960, 242-248.
- Treisman, A.M.: Verbal cues, language and meaning in selective attention, *Am. J. Psychol.* 77, 1964, 206-219.
- Vogten, L.L.M. and L.F. Willems: The formator: a speech analysis-synthesis system based on formant extraction from linear prediction coefficients, *Instituut voor Perceptie Onderzoek, Eindhoven, Annual Progress Report*, 12, 1977, 47-54.
- Willems, L.F. and Th.A. de Jong: Research tools for speech perception studies, *Instituut voor Perceptie Onderzoek, Eindhoven, Annual Progress Report*, 9, 1974, 77-81.
- Wingfield, A. and J.L. Wheale: Word rate and intelligibility of alternated speech, *Perception & Psychophysics*, 18(5), 1975a, 317-320.
- Wingfield, A. and J.L. Wheale: Interaural alternation, information load, and speech intelligibility, *J. Acoust. Soc. Amer.*, 57, 1975b, 1219-1220.

CURRICULUM VITAE

- 11-09-1950 geboren te Zutphen
- 07-06-1968 eindexamen H.B.S.-B, Hertog Jan College Valkenswaard
- 20-09-1974 diploma natuurkundig ingenieur, Technische Hogeschool Eindhoven. Afstudeeronderzoek: Autonome Reacties bij het Stotteren, o.l.v. ing. M. Verduin, Dr.Ir. C.J. Snijders.
- 01-10-1974 t/m 31-01-1975 fellowship Karolinska Instituut, Stockholm
- 01-02-1975 t/m 30-06-1978 werkzaam op het Instituut voor Perceptie Onderzoek, Eindhoven in dienst van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek.

STELLINGEN

1. De gevoeligheid van de methode van Nakatani en Dukes voor het meten van kwaliteitsverschillen tussen spraak-communicatiekanalen, kan vergroot worden door gebruik te maken van monotone spraak.
vgl. Nakatani, L.H. and K.D. Dukes; a sensitive test of speech communication quality, J.Acoust. Soc.Amer. 53, (4), 1973, 1085-1092.
2. Waar de grenzen liggen voor vrijheid van wetenschappelijk onderzoek mag niet uitsluitend aan de onderzoekers zelf ter beoordeling worden overgelaten.
vgl. K. Anér; Att mixtra med livets grundvalar, Framsteg och Forskning 1976(8), 26-28.
3. Steeds sterker zal gelden dat de waardering van technische gebruiksartikelen bepaald wordt door de mate waarin die producten aangepast zijn aan de gebruiker.
vgl. N. Negroponte; the metaphysics of television, submitted to IFIPS W.G. 5.2. Conference, Seillac, 1979.
4. Ondanks het toenemend gebruik van beeldschermen wordt nog nauwelijks experimenteel onderzoek gedaan aan de leesbaarheid ervan.
zie bijvoorbeeld: Visual display units and their applications, ed. D. Grover, IPC Business Press Ltd. 1976.
5. Een energiecrisis kan een middel zijn om het bedrijfsleven aan te sporen tot innovatie.
6. Gezien het grote aantal medicijnen dat voor zelfmedicatie wordt gebruikt, is het vermelden van een uiterste gebruiksdatum wenselijk op de aan patiënten uitgegeven verpakking.
Medische Consumptie, Sociologisch Instituut R.U. Groningen, 1974.
7. Zowel mensen die principieel geen medicijnen gebruiken als mensen die zonder medische indicatie altijd pillen slikken, laten zich niet leiden door rationele overwegingen.
8. Niet alleen in de perceptie stellen "contextfactoren" mensen in staat verbazingwekkende prestaties te leveren.

9. Het verdient aanbeveling vruchtensappen of andere dranken die voor het gebruik geschud moeten worden, niet in een slecht afsluitbare kartonnen verpakking te leveren.
10. Bij toepassing van geavanceerde electronica in de keuken, kan de eerste betekenis van het woord "aangebrand" uit het woordenboek worden geschrapt.
 - v. Dale Groot Woordenboek der Nederlandse Taal, 1970.

Eindhoven, 15 mei 1979

J.P.L. Brokx

ERRATUM

Op pagina 90 dient de laatste alinea vervangen te worden door de tekst:

In Fig. 3.14 staat voor de verschillende stimuli het relatieve totale aantal fouten als functie van het relatieve geluidsniveau van de testzinnen t.o.v. de maskeertekst weergegeven. Alle toonhoogtecondities verschillen significant van de toonhoogteconditie met testzinnen van 120 Hz (Newman-Keuls). In de Fign. 3.15 en 3.16 staan de relatieve aantallen vervangingen en weglatingen apart weergegeven. Alle toonhoogtecondities zijn in Fig. 3.15 niet en in Fig. 3.16 wel significant verschillend van de toonhoogteconditie met testzinnen van 120 Hz (Newman-Keuls).

Waargenomen continuïteit in spraak: het belang van toonhoogte.

J.P.L. Brokx