

Spreker identificatie experiment

Citation for published version (APA):

Briennesse, P., Deriks, J. G. F., Gunter, P. L. J., Janssen, G. M., Peeters, A. G., & Tas, R. D. (1987). *Spreker identificatie experiment*. (IPO-Rapport; Vol. 585). Instituut voor Perceptie Onderzoek (IPO).

Document status and date:

Gepubliceerd: 02/06/1987

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.

Instituut voor Perceptie Onderzoek
Postbus 513, 5600 MB Eindhoven

Rapport no. 585

Spreker identificatie experiment

P. Briennesse, J. Deriks, P. Gunter,
G. Janssen, A. Peeters, R. Tas

Spreker identificatie experiment

door:

Patrick Briennesse

Joost Deriks

Pieter Gunter

Ger Janssen

Arthur Peeters

Robert Tas

2 juni 1987

Samenvatting

Om verschillende spraakreproducerende systemen onderling te kunnen vergelijken is een spraakkwaliteitsbepaling nodig. Een aanwijzing voor de kwaliteit is de mate waarin sprekerspecifieke kenmerken worden aangetast. Doel van dit project is deze kwaliteitsmaat te testen met een sprekeridentificatie-experiment. Er worden drie spraakreproductietechnieken gebruikt, te weten PCM, LPC en MEA. Deze technieken verschillen onderling in de zuinigheid waarmee het spraaksignaal gecodeerd is (respectievelijk 120, 12 en 2.4 kbit/s.) De spraakfragmenten die gebruikt worden zijn geselecteerd uit opnamemateriaal van een door proefpersonen uitgevoerde variant op het spel "zeeslag". Op deze wijze wordt spontane spraak met een beperkte woordenschat verkregen. Van 16 sprekers zijn in totaal 159 spraakfragmenten verzameld, variërend van een enkel woord tot korte zinnen. Na bewerking met genoemde reproductietechnieken zijn die fragmenten ter identificatie aangeboden aan 20 luisteraars.

Uit de identificatietest blijkt dat de sprekerkenmerken inderdaad in verschillende mate aangetast worden. Met een betrouwbaarheid van 95% ligt het percentage van herkende sprekers in de volgende intervallen:

- PCM: $80.3\% < m < 91.1\%$
- LPC: $69.0\% < m < 78.8\%$
- MEA: $35.1\% < m < 56.3\%$

De bekendheid van de luisteraars met de sprekers is geschaald. Het blijkt dat bij MEA, gerelateerd aan PCM, onbekende sprekers slechter herkend worden dan bekende sprekers. Bij LPC is dit effect niet aantoonbaar. Verder hebben leereffecten in dit experiment een grote rol gespeeld, omdat dezelfde fragmenten gebruikt werden voor zowel PCM, LPC als MEA-spraak.

Inhoud

1	Inleiding	1
2	Theorie Spraak	3
2.1	Natuurlijke spraak	3
2.2	Bron-filter model	3
2.3	Een systeem voor analyse en resynthese van spraak	4
2.3.1	De Analyse	5
2.3.2	De correctie	6
2.3.3	Zuinige codering	6
2.3.4	De Resynthese	6
3	Het Experiment	8
3.1	Inleiding	8
3.2	Opmames	9
3.2.1	Het Spel	9
3.2.2	De Spelers	10
3.2.3	De Studio	11
3.3	Verwerking tot Stimulusbanden	12
3.3.1	Selectie, inname en segmentatie van de fragmenten	12
3.3.2	De Analyse	12
3.3.3	De Correctie	13
3.3.4	Zuinige codering	13
3.3.5	Het maken van de banden	13
3.4	Luistertests	14
3.4.1	De Luisteraars	14
3.4.2	De Luisterruimte	14
3.4.3	De Formulieren	14
3.4.4	De Procedure	15
4	Resultaten	16
5	Discussie/Conclusies	21
	Literatuur	24
A	Resultaten	25
B	Formulieren	27
C	Lijst van alle op de stimulusbanden gebruikte fragmenten	34

Hoofdstuk 1

Inleiding

Dit is een verslag van een projekt van tweedejaars natuurkundestudenten. Projektwerk is een van de mogelijkheden van het natuurkundepraktikum in het zesde trimester van de studie. Het praktikum kan dan buiten de faculteit plaatsvinden, zoals in dit geval bij het IPO.

Een van de onderzoeksgebieden van het IPO is spraak, in het bijzonder kunstmatige spraak. Met behulp van de computer kan een oorspronkelijk spraaksignaal gereproduceerd worden. Een eerste manier waarop dit gebeuren kan is de zuivere digitale omzetting van het spraaksignaal, de PCM spraak (Pulse Coded Modulation). Voor toepassing is het echter van belang de digitale informatiestroom die de spraak representeert zoveel mogelijk te reduceren, onder meer in verband met de geheugencapaciteit. Gezocht wordt naar een model dat spraakgeluid beschrijft in een aantal perceptief belangrijke parameters. De informatie over de spraak kan dan beperkt worden tot informatie over parameters. Men spreekt van meer of minder zuinige codering naar gelang het aantal bits dat een seconde spraak beschrijft. De codering wordt zuiniger door het aantal parameters te beperken of het aantal bits waarmee de parameters gekwantiseerd zijn. Te verwachten is dat met het zuiniger maken van de codering de kwaliteit van de spraak terugloopt, omdat er informatie verloren gaat. De vraag is alleen in welke mate dit gebeurt.

Nu is spraakkwaliteit moeilijk te meten. Als objectieve maat wordt vaak de spraakverstaanbaarheid aangewezen. Toch moet hier opgepast worden. Verstaanbaarheid is immers wel een vereiste voor goede kwaliteit, maar nog geen garantie. Een mogelijk andere aanwijzing voor de kwaliteit is de mate waarin sprekersspecifieke kenmerken door de codering worden aangetast.

Doel van dit projekt is om met een sprekeridentificatie experiment na te gaan of deze laatstgenoemde kwaliteitsbepaling bruikbaar is bij het vergelijken van een drietal spraakcoderingstechnieken (PCM, LPC en MEA, zie voor nadere uitleg de hoofdstukken "Theorie" en "Experiment"). Gekeken wordt dus of deze in "zuinigheid" sterk uiteenlopende technieken de sprekerkenmerken inderdaad in verschillende mate aantasten.

Met dit doel voor ogen is er eerst gezocht naar een geschikte manier om spraakmateriaal te verzamelen. Na bewerking van het materiaal wordt het in een identificatie experiment aan proefpersonen aangeboden.

In het hoofdstuk "Theorie" wordt ingegaan op de theoretische achtergronden van (kunstmatige) spraak en de coderingstechnieken. Het hoofdstuk "Experiment" geeft eerst een motivatie voor het soort experiment dat uitgevoerd is en daarna een

uiteenzetting over het praktische werk.

Op deze plaats willen wij onze begeleiders Berry Eggen, Jan-Roelof de Pijper en Leo Vogten bedanken voor de prettige samenwerking tijdens dit projekt.

Hoofdstuk 2

Theorie Spraak

2.1 Natuurlijke spraak

Bij de produktie van spraak zijn veel verschillende organen betrokken (zie figuur 2.1).

De bron van het spraakgeluid ligt in het systeem van de longen, luchtpijp, strottehoofd en stembanden. Wanneer de stembanden bij de spraakproduktie in trilling zijn, resulteert dit in een brongeluid met een bepaalde frequentie F_0 , en een reeks boventonen met frequenties $f_n = nF_0$ ($n=1,2,3,\dots$). Er is dan sprake van een stemhebbende bron.

Niet alle klanken in de spraak zijn echter stemhebbend, zoals bijvoorbeeld de f, g, s enzovoort. Bij deze klanken wordt het brongeluid gevormd door een geruis dat ontstaat door turbulenties van de luchtstroom uit de longen. Dit is de zogenaamde stemloze bron.

De omvorming van het brongeluid tot de uiteindelijke spraakklanken geschiedt in de keel, mond en neusholte. Hier ontstaan door de verschillende standen van de verschillende organen (zoals lippen, tong enzovoort, zie figuur 2.1) resonanties bij verschillende frequenties. Op deze manier worden bepaalde boventonen versterkt, waardoor het brongeluid tot het uiteindelijke spraakgeluid gefilterd wordt.

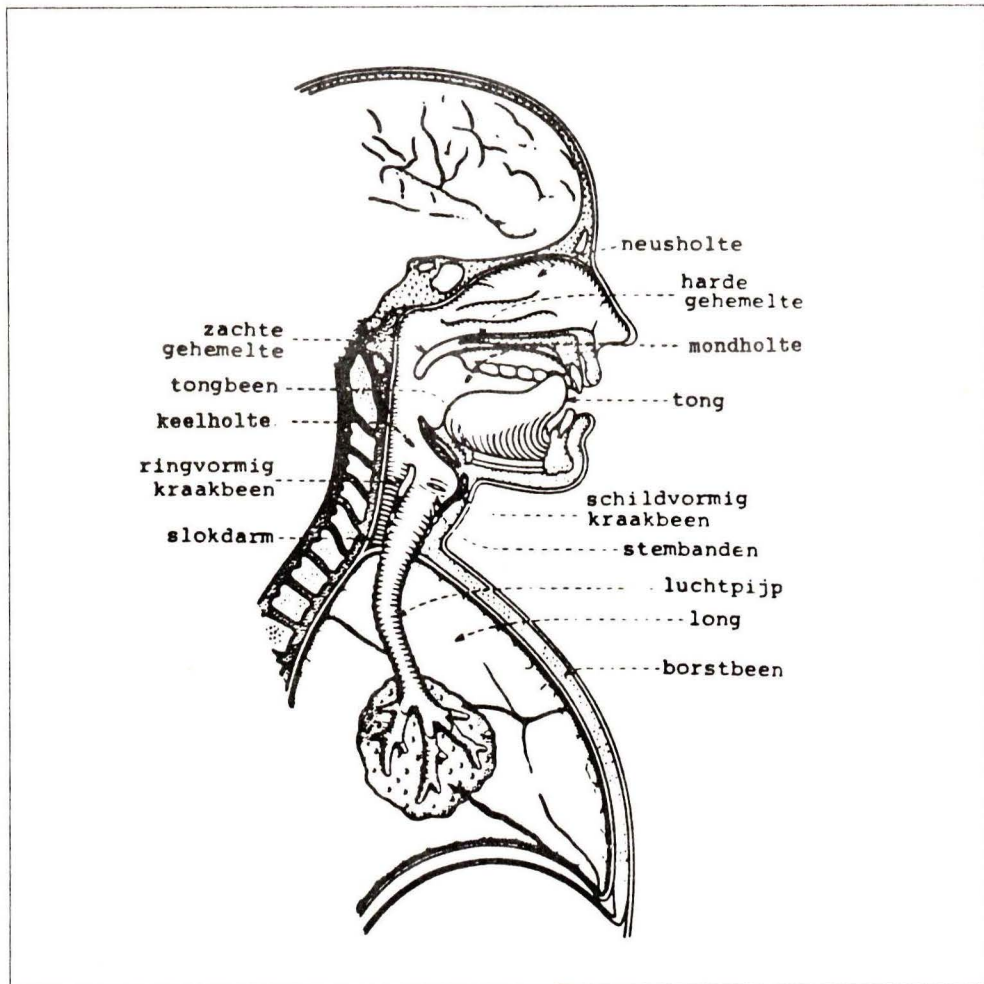
2.2 Bron-filter model

De spraakproduktie kan elektronisch nagebootst worden met behulp van het volgende model (zie figuur 2.2).

Voor de stemhebbende klanken bestaat de bron uit impulsen met een instelbare periode $1/F_0$. Het frequentiespectrum van de bron bestaat hierdoor uit impulsen die zich bevinden op de frequenties $F_0, 2F_0, 3F_0$ enzovoort.

De bron heeft dan een grondtoon met frequentie F_0 en een aantal boventonen ($2F_0, 3F_0$ enzovoort) van gelijke sterkte. De stemloze klanken worden verkregen door de bron witte ruis te laten produceren. Dit signaal wordt vervolgens versterkt met een variabele versterkingsfactor g . Het hierachter geplaatste filter O representeert de filtering die bij natuurlijke spraak door de neus, mond, en keelholte wordt gerepresenteerd. In figuur 2.3 is de overdrachtsfunctie $O(f)$ van het filter O grafisch weergegeven.

De resonantiefrequenties zijn terug te vinden als pieken in de overdrachtsfunctie

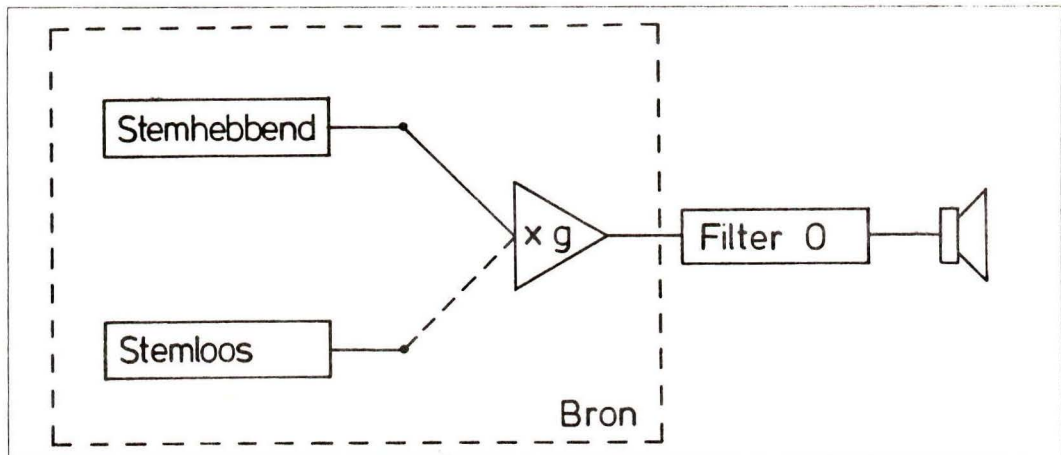


Figuur 2.1: Schematische voorstelling van de belangrijkste organen die bij het spreken betrokken zijn.

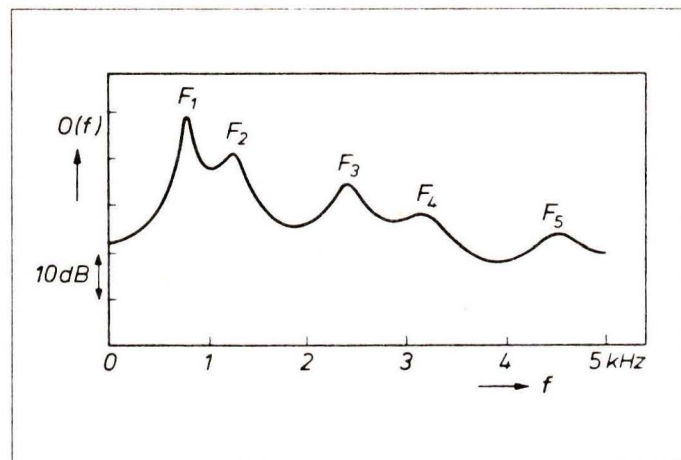
van het filter. Bij natuurlijke spraak neemt de sterkte van de boventonen af met 6 dB per octaaf. Omdat hiermee bij de bron van het elektronische systeem geen rekening is gehouden is deze daling verwerkt in het filter O.

2.3 Een systeem voor analyse en resynthese van spraak

Het bron-filter model uit §2.2 kan goed gebruikt worden voor het analyseren, resyntheseren en manipuleren van spraak. Om dit met de computer te kunnen doen is het noodzakelijk om de spraak in digitale vorm op te slaan. Een methode om de spraak digitaal te coderen is de golfvormcodering. Bij de golfvormcodering wordt de amplitude van de continue, analoge golfvorm van de spraak op discrete tijdstippen afgebeeld in getalvorm. Dit wordt ook wel pulscodemodulatie (PCM) genoemd. De hier verder toegepaste PCM gecodeerde spraak heeft een bitstroom van 120 kbit/s. Deze bitstroom is te reduceren met behulp van parametercodering, die spraak in termen van parameters van het bron-filter model codeert, tot 12 kbit/s. Een uitvoeriger beschrijving van deze parameters is te vinden in §2.3.1, "De Analyse". Een systeem voor analyse en resynthese van spraak ziet er als volgt uit (zie figuur 2.4).



Figuur 2.2: Schematische voorstelling van het bron-filter model.

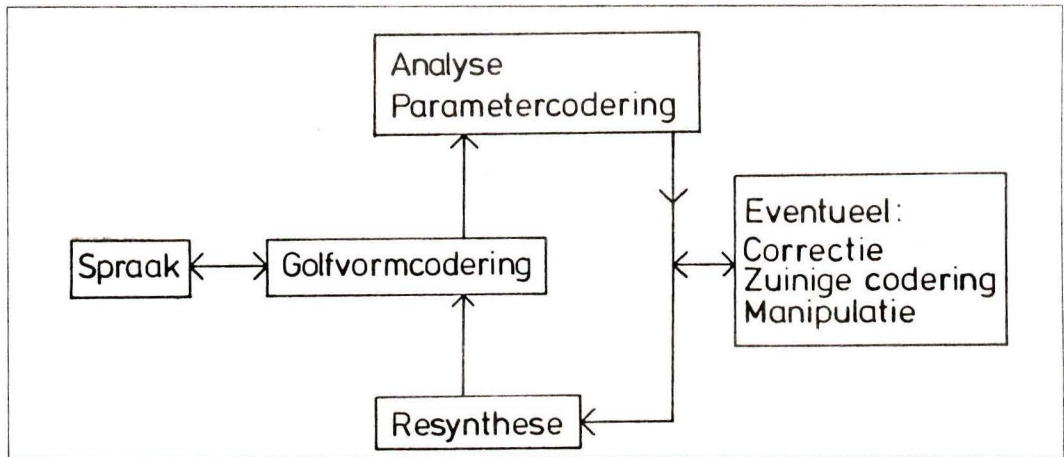


Figuur 2.3: Voorbeeld van een overdrachtsfunctie, gebruikt in het filter O .

Het een en ander wordt toegelicht in de volgende paragrafen.

2.3.1 De Analyse

Tijdens de analyse wordt de spraak in termen van 13 parameters van het bron-filter model gecodeerd. De 3 parameters voor de bron uit het model van figuur 2.2 zijn de toonhoogte (en daarmee ook de bronfrequentie F_0), de versterkingsfactor g , en de parameter die aangeeft of de bron stemhebbend of stemloos is. Het filter wordt beschreven door de overige 10 parameters. Met het overdrachtsfilter uit figuur 2.3 kan men resonantiepieken (formanten) beschrijven. Elke formant wordt op zijn beurt beschreven door zijn positie op de frequentieschaal en zijn bandbreedte. Bij de analyse wordt het spraakfragment in stukken van 10 ms gehakt. Voor elk van deze periodes van 10 ms (frames) worden de 13 parameters bepaald. De op deze manier gecodeerde spraak wordt ook wel LPC (Lineair Predictive Coding) spraak genoemd en vereist een bitstream van 12 kbit/s.



Figuur 2.4: Globale werking van een systeem voor analyse en resynthese van spraak.

2.3.2 De correctie

Wanneer met behulp van de computer de modelparameters automatisch zijn bepaald kan het nodig zijn om bepaalde fouten die bij de analyse zijn gemaakt te corrigeren. Twee veel voorkomende fouten zijn:

- Toonhoogte fouten. Het verloop van de toonhoogte (bronfrequentie) contour tegen de tijd kan soms uitschieters aangeven, die de menselijke stem nooit in zo'n korte tijd teweeg kan brengen.
- Stemhebbend/stemloos fouten. De analyse resultaten kunnen soms een zodanig stemhebbend/stemloos verloop aangeven, dat het na vergelijking met het originele materiaal duidelijk is dat er fouten gemaakt zijn.

2.3.3 Zuinige codering

De bitstroom van de LPC spraak is te reduceren door bijvoorbeeld het aantal analysestappen (frames) per seconde te verminderen. Verder kunnen de modelparameters grover gekwantiseerd worden. Het is bijvoorbeeld mogelijk om de vierde en vijfde formant vast te zetten op een bepaalde waarde, aangezien het menselijk oor minder gevoelig is voor veranderingen in deze formanten dan voor veranderingen in de eerste drie formanten. De bitstroom wordt dan gereduceerd tot bijvoorbeeld 2 kbit/s. Dergelijke zuinige coderingen worden ondermeer gebruikt bij spraakproductie met hardware spraaksyntheseschips.

2.3.4 De Resynthese

Uitgaande van een, al dan niet zuinig gecodeerde parameterbeschrijving wordt bij de resynthese de golfvorm van het spraakgeluid weer berekend en daarna ten gehore gebracht. Deze software synthese verschilt dus van de onbewerkte PCM gecodeerde spraak.

Er zijn dus drie "typen" spraak te onderscheiden, te weten PCM spraak, die nauwelijks van de originele opnamespraak te onderscheiden is, LPC spraak en een vorm van zuiniger gecodeerde spraak, de MEA spraak (De MEA spraak wordt verder toegelicht in §3.3.4 op blz. 12). In het volgende hoofdstuk worden deze drie soorten spraak, met bitstromen van respectievelijk 120, 12 en 2 kbit/s, gebruikt in het sprekeridentificatie experiment.

Hoofdstuk 3

Het Experiment

3.1 Inleiding

Het experiment is grofweg in drie delen te splitsen. Het eerste deel bestaat uit het verzamelen van spraakmateriaal. Daarna wordt dit materiaal bewerkt met de computer (deel twee), zodat het in de identificatietest, in het derde deel, aan proefpersonen kan worden aangeboden. Deze driedeling is in de rest van dit hoofdstuk aangehouden.

Er zal nu wat nader worden ingegaan op de methode die gebruikt is om het spraakmateriaal te verkrijgen. Een idee hiervoor, ontwikkeld door het Amerikaanse Naval Research Laboratory (NRL), (Schmidt-Nielsen & Stern, 1985), (Schmidt-Nielsen & Everett, 1982), is dat er opnames worden gemaakt van sprekers die met elkaar een variant van het spelletje "Zeeslag" spelen. De achterliggende gedachte is dat op deze manier spontane (in tegenstelling tot voorgelezen) spraak verkregen wordt. In de praktijk zal sprekerherkenning immers meestal geschieden bij conversatie en dus bij spontane uitingen. Het maken van onderscheid hier is belangrijk omdat mensen als ze spontaan praten woorden vaak minder zorgvuldig uitspreken, klanken inslikken etc.

Voordelen van het spel "Zeeslag" zijn verder dat het interessant genoeg is om de spelers (=sprekers) gemotiveerd te houden en dat het eenvoudig is om uit te voeren, zowel voor de experimentator als voor de spelers. Het verdient daarom de voorkeur boven andere verzamelmethodes, waarvan er enkele beschreven worden in Schmidt-Nielsen en Everett uit (1982).

In §3.2.1 zal nader op de spelregels van "Zeeslag" worden ingegaan. De condities waaronder de spelopnames gemaakt zijn worden beschreven in §3.2.2. De keuze van de spelers, dus sprekers, en hun aantal zal tenslotte worden toegelicht in §3.2.3.

Nadat de opnames gemaakt zijn moeten ze verwerkt worden tot het stimulusmateriaal dat aan de luisteraars aangeboden wordt ter identificatie. Onder verwerking moet dan allereerst vestaan worden het kiezen van de geschikte fragmenten uit het opnamemateriaal. Vervolgens worden deze fragmenten met behulp van de drie verschillende coderingstechnieken geanalyseerd en geresynthetiseerd. De drie soorten spraak worden tenslotte in een geschikte volgorde op band opgenomen. Deze banden vormen het stimulusmateriaal. Een en ander wordt nader uiteengezet in §3.3.1.

Tenslotte kunnen de luistertesten worden afgenomen. Proefpersonen trachten de sprekers van de fragmenten te identificeren voor de verschillende soorten kunstmatige spraak. Onder "Luistertesten" wordt van dit onderdeel van het experiment een

beschrijving gegeven. Ook de keuze van de proefpersonen komt hier aan de orde. In het volgende hoofdstuk: "Resultaten" zijn de uitkomsten van het experiment verder verwerkt.

3.2 Opnames

3.2.1 Het Spel

Om spontane spraak te verkrijgen was een vereenvoudigde vorm van het spelletje zeeslag ("battleship"), zie Appendix B voor gedetailleerde spelbeschrijving en gehanteerde regels. Hierbij moesten de spelers om de beurt door het vermelden van de coördinaten van een hokje van het veld van de tegenstander proberen een vijandelijk 'schip' te raken (bijvoorbeeld persoon A roept 'alfa-5', waarop de tegenstander 'raak', 'mis', of iets dergelijks antwoordt).

Om spontane spraak te verkrijgen is een vereenvoudigde vorm van de NRL Communicability Test (Schmidt-Nielsen & Evert, 1982) gebruikt. Als belangrijkste voordelen van dit spel worden in dit artikel genoemd:

- Het spel is motiverend voor een spontane dialoog.
- Omdat de spelers bij elk spel hun eigen 'schepen' plaatsen, behoeft deze methode weinig geprepareerd materiaal om de dialoog te bewerkstelligen.
- De eventueel optredende leereffecten zijn verwaarloosbaar (het maakt niet uit hoe vaak een speler het spel speelt).
- Er kan uit worden gegaan van zowel ervaren als onervaren sprekers, omdat de opzet van het spel zeer eenvoudig is.
- Het verkregen spraakmateriaal is qua fonetische samenstelling voor verschillende sprekers hetzelfde.

Wij hebben hieraan nog toe te voegen:

- Het spel vergt maximaal rond de zes minuten speeltijd, zodat volgens een strakke planning gewerkt kan worden, en de spelers niet teveel tijd aan de sessie kwijt zijn.

Uitvoering en opname van het spel:

Per opnamesessie waren vier mensen uitgenodigd, die allemaal een keer tegen elkaar moesten spelen. Dit leverde per opname zes spelletjes op, die opgenomen werden op een bandrecorder.

In totaal werden vier van zulke sessies gehouden, die in totaal 24 opgenomen spellen opleverden en waaraan in totaal 16 spelers meewerkten.

De vier sprekers die per sessie verschenen, werden ontvangen in de regiekamer van de IPO studio, waar hun de spel-instructies overhandigd werd. De spelregels zijn vermeld in het formulier "Spel-instructies", zie Appendix B. Bij het betreden van de studio werd een spelformulier overhandigd dat, wanneer de spelers plaats genomen hadden, ingevuld werd. Wanneer dit gebeurd was, werd dit via de per speler opgestelde microfoon kenbaar gemaakt aan de regiekamer. De regiekamer

bevestigde dit door via een interkom aan de spelers in de studio te vragen of ze konden beginnen met het inspreken van hun naam, waarna het spel van start kon gaan.

De opname van het spel werd gemaakt op een tweespoors revox B77 MK 2 recorder, waarbij werd getracht iedere speler zo goed mogelijk gescheiden van de andere op een kanaal op te nemen. Omdat er vaak een groot verschil was tussen het spraakniveau van de spelers, liepen de opnames via een mengpaneel, waarbij ieder kanaal naar believen kon worden bijgesteld. Deze bijregeling geschiedde grof gedurende de korte communicatie tussen regiekamer en spelers voor aanvang van het spel en fijn gedurende het spel.

Omdat het spraakvolume varieert met de afstand spreker-microfoon, werd de spelers voor aanvang verzocht deze afstand zo goed mogelijk op ongeveer 20 cm te houden gedurende de opname. Een afstand van 20 cm werd in IPO rapportnr 521 als optimale afstand gevonden voor opnames gemaakt in de IPO studio met de door ons gebruikte microfoons Bruel en Kjaer type 4003.

3.2.2 De Spelers

Aan de zeeslag-spelletjes werd door zestien personen deelgenomen. Dit aantal van zestien is gekozen om de volgende redenen.

Allereerst mag het aantal niet te klein zijn. De resultaten zouden anders niet in aanmerking komen voor statistische bewerkingen. Dit stelt een benedengrens aan het aantal spelers van ruwweg tien. Een bovengrens wordt bepaald door het aantal personen dat hun medewerking kan en wil verlenen.

Uit praktische overwegingen worden zowel de spelers als de luisteraars gekozen uit IPO-medewerkers. Hierbij wordt geprobeerd om de groep spelers en de groep luisteraars gescheiden te houden. Het is dus van belang, dat er genoeg personen overblijven om later aan de luistertests te kunnen deelnemen.

De opnames van de spelletjes worden steeds uitgevoerd met groepjes van vier spelers. Hierdoor is het handig om als aantal spelers een veelvoud van vier te nemen.

Samenvattend kan worden gezegd, dat het gewenst is om het aantal spelers zo groot mogelijk te kiezen. Uit praktische overwegingen was al besloten om de spelers en de luisteraars te kiezen uit IPO-medewerkers, zodat de luisteraars de spelers kennen. Het streven naar een groot aantal spelers, maakt het welhaast noodzakelijk om bekende spelers te kiezen (Schmidt-Nielsen en Stern, 1985). Wanneer onbekende spelers worden gebruikt dan moeten de luisteraars eerst de spelers leren kennen. Als het aantal onbekende spelers betrekkelijk groot is, zullen de luisteraars zich nog maar een klein gedeelte van de spelers herinneren bij het uitvoeren van de luistertest. Dit beperkt het aantal te gebruiken spelers tot vijf à zes. Dezelfde beperking geldt dan echter ook voor het aantal spelers waaruit de luisteraars kunnen kiezen bij het identificeren van een stem. Dit levert het gevaar op dat de luisteraars bij de identificatie namen gaan schrappen van personen waarvan zij zeker weten dat zij niet de spreker zijn van het desbetreffende stukje spraak. Indien bekende spelers worden gebruikt, dan treden deze problemen niet op en kan, afhankelijk van het beschikbaar aantal personen, een veel grotere groep spelers worden gebruikt.

Dit alles in overweging genomen, werd besloten om uit te gaan van zestien spelers. Om redenen die later worden besproken (zie §3.3.5), waren twee van deze spelers volkomen onbekend op het IPO. De overige veertien zijn allen mannelijke IPO-

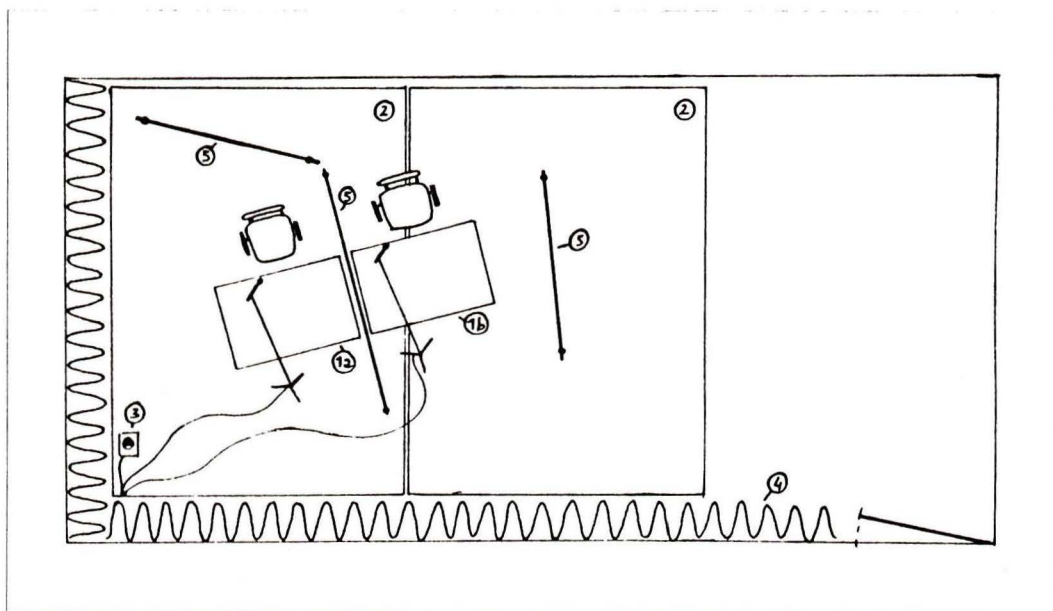
medewerkers. Verder werd van hun geëist dat ze reeds langere tijd zijn verbonden aan het IPO, zodat ze bekend zijn bij veel van hun collega's. Een laatste eis was dat ze van Nederlandse afkomst zijn, zodat ze niet met een buitenlands accent spreken.

3.2.3 De Studio

Bij de inrichting van de IPO studio is erop gelet dat zo min mogelijk geluidsreflecties op konden treden. Dit was nodig vanwege de volgende redenen:

- Geluidsreflecties kunnen een periodieke (in volume afnemende) herhaling van het door de microfoon opgepikte signaal veroorzaken. Dit kan problemen opleveren bij de toonhoogteanalyse van de opgenomen fragmenten.
- De twee spelers werden ieder op gescheiden kanalen opgenomen. Dit had als voordeel dat de door de ene spreker geproduceerde bijgeluiden (zuchten, smakken, tikken en bewegen) niet interfereerde met de spraak van de andere spreker. Dankzij de grote geluidsabsorbtie van de studio was de kanaalscheiding voldoende.

Reflecties werden zoveel mogelijk tegengegaan door het aanbrengen van een wollen vloerkleed, dikke gordijnen, absorberende schotten tussen en achter de spelers, het scheef zetten van het studiointerieur tov de wanden, en het aanbrengen van vilt op de tafels. Dit laatste had bovendien als voordeel dat het geluid van schuiven van de papieren op de tafels grotendeels geelimineerd werd. In figuur 3.1 is de opstelling op schaal weergegeven.



Figuur 3.1: De studio opstelling; 1a,b: tafel,stoel en microfoon voor spelers a en b; 2: tapijten; 3: intercom; 4: gordijn; 5: geluidsabsorberende schermen; schaal 1:43.

De hoogte van de studio is ongeveer 2.80 m. De schermen hebben alle dezelfde afmetingen, maar sommige zijn rechtop geplaatst.

3.3 Verwerking tot Stimulusbanden

3.3.1 Selectie, inname en segmentatie van de fragmenten

Bij de verwerking van de opnames tot de uiteindelijke stimulusbanden is begonnen met het selecteren van spraakfragmenten die geschikt zijn als stimulusmateriaal. Bij deze selectie zijn de volgende criteria gehanteerd:

Allereerst moet de spraak goed verstaanbaar en van een goede opname kwaliteit zijn. Passages die te zacht zijn uitgesproken of worden gestoord door hoesten, zuchten, lachen en meer van dergelijke bijgeluiden vallen af. In de praktijk is gebleken dat uit 5 minuten spraak makkelijk 10 goede fragmenten te selecteren zijn.

Ten tweede moet van elke spreker een zodanig gevarieerde serie fragmenten gevormd worden dat zijn specifieke sprekerkenmerken voldoende naar voren komen. Omdat van een spreker moeilijk te bepalen is welke factoren in zijn spraak voor hem of haar kenmerkend zijn, zijn zoveel mogelijk fragmenten verzameld. Daarom zijn van iedere spreker fragmenten opgenomen waarin de woorden "alfa", "beta", "charlie", "delta" en "echo" met een willekeurig getal voorkomen. Dit geheel is aangevuld met zinnen als "Die is mis", "Raak en gezonken" enzovoort.

Verder is erop gelet dat niet te veel voor de spreker karakteristieke zinnen of woorden zijn opgenomen. Het gevaar bestaat namelijk dat de herkenning niet meer geschiedt op grond van de sprekerkenmerken maar door herkenning van de spreekgewoontes van de spreker (zoals bijvoorbeeld bepaalde stopwoorden). Dit is voorkomen door alleen fragmenten op te nemen die functioneel zijn voor het verloop van het spel, zoals bijvoorbeeld "Die is raak", "Je hebt verloren" enzovoort. Een complete lijst van de gebruikte fragmenten is te vinden in Appendix C.

Nadat een fragment geschikt bevonden is, is het vanaf de analoge tape opgenomen in de computer en in digitale vorm opgeslagen in een dummy-file. Daarna is het fragment precies op de juiste plaatsen afgekapt en weggeschreven naar een zogenaamde N-file. Een N-file is een file waarin de spraak in golfvormcode is opgeslagen. Bij het nauwkeuriger segmenteren van de dummy-file is er op gelet dat er geen hinderlijke tikken aan het begin en eind van de N-file ontstaan. Dit is gedaan door het begin en eind steeds op een nuldoorgang van de dummy-file-golfvorm te kiezen.

3.3.2 De Analyse

Na het segmenteren is het fragment klaar voor de analyse. Deze is uitgevoerd in twee stappen. Als eerste zijn de versterkingsfactor, de stemhebbend/stemloos parameter en de parameters die de formanten beschrijven bepaald. Dit is gebeurd met het programma AAP (Vogten, 1985) dat hiervoor een nieuwe file, de A/P-file, in het geheugen creëert. Vervolgens is de toonhoogte gemeten met het programma PCT (Hermes, 1987) De verkregen toonhoogte is aan de A/P-file toegevoegd. Hierna bevat deze file alle dertien parameters die nodig zijn om het fragment te reproduceren. De resynthese is uitgevoerd met behulp van het programma SYN. Dit programma creëert weer een N-file die door de computer rechtstreeks kan worden uitgegeven. Na het doorlopen van deze stappen zijn er in het geheugen twee N-files, een PCM en een LPC versie. Omdat de analyse en resynthese nogal veel tijd in beslag nemen zijn deze 's nachts door de computer uitgevoerd met behulp van een batch-job.

3.3.3 De Correctie

Na de analyse is voor elk fragment de LPC spraak vergeleken met de PCM versie om op deze wijze eventuele analysefouten op te sporen. Elk duidelijk hoorbaar verschil is gecorrigeerd. In de praktijk bleek het alleen te gaan om correctie van de stemhebbend/stemloos parameter. Een veel voorkomende fout is bijvoorbeeld dat "Die is raak" wordt weergegeven als "Die iz raak". Bij de correctie van een ten onrechte stemloze klank is de bronfrequentie over het stemloze stuk geëxtrapoleerd. Van de klanken "v" en "z" is een klein stukje stemloos gemaakt zodat deze wat meer als stemhebbende wrijfklanken overkomen. In totaal is op deze manier ongeveer een derde van de fragmenten gecorrigeerd.

3.3.4 Zuinige codering

De geanalyseerde spraakfragmenten, opgeslagen in de A/P-files zijn nog zuiniger gecodeerd met het programma MEA (Vogten, 1985). MEA is een programma dat van een spraakfragment de frequentie en de bandbreedte van de vijfde formant op een vaste waarde zet. Verder zet MEA ook de frequentie van de vierde formant vast. Het resultaat van de bewerking wordt in een andere A/P-file geschreven. Deze A/P-file is weer gesynthetiseerd met het programma SYN.

3.3.5 Het maken van de banden

Het Fragment

De inhoud van de fragmenten is eerder in dit hoofdstuk al besproken. Elk fragment bevat negen à tien zinnen of uitdrukkingen. Dit houdt in dat de fragmenten zullen bestaan uit tien à vijftien seconden spraak.

In een onderzoek van Schmidt-Nielsen en Stern (1985) komen beiden tot de conclusie dat een fragmentlengte van twee à drie seconden al voldoende kan zijn om een spreker te herkennen. De fragmenten van tien à vijftien seconden zullen dus waarschijnlijk lang genoeg zijn.

Oorspronkelijk was het de bedoeling om de luisteraars de mogelijkheid te geven de fragmenten meerdere keren te horen. Twee proefluisteraars, die niet meer aan de echte tests mee mochten doen, hebben naar een band geluisterd en verklaard dat deze mogelijkheid overbodig is. Hierna werd besloten om deze mogelijkheid te schrappen. Dit heeft tevens als voordeel dat meerdere personen tegelijk kunnen luisteren. Ook bevestigden de proefluisteraars dat de fragmenten lang genoeg waren.

Uiteindelijk werd een band samengesteld met achttien fragmenten met een gemiddelde tijdsduur van ongeveer twintig seconden per fragment. Deze fragmenten bevatten tien à vijftien seconden spraak, met tussen elke zin één seconde pauze.

De Band

Het totale aantal fragmenten per band was achttien. Er waren veertien sprekers en dus ook veertien fragmenten. Twee van deze fragmenten kwamen twee keer voor. Om een spreker te identificeren moeten de luisteraars een naam kiezen uit een lijst van dertig. Om te voorkomen dat ze namen schrappen van sprekers die al zijn geweest, werd hun meegedeeld dat een spreker meerdere keren voor kon komen. Om te controleren of de luisteraars deze waarschuwing ook werkelijk in acht namen,

kwamen twee van de veertien fragmenten twee keer voor en werden hieraan nog twee fragmenten, van twee voor de luisteraars onbekende sprekers, toegevoegd.

Voorafgaand aan elk fragment werd door een mannenstem het nummer van het fragment genoemd. Elk fragment werd gevolgd door vijftien seconden pauze om de luisteraars in de gelegenheid te stellen het antwoordformulier in te vullen.

De volgorde van de zinnen was voor één fragment van een bepaalde spreker in de drie spraakversies verschillend. Ook was de volgorde van de fragmenten per band verschillend en kwam op elke band een ander tweetal fragmenten dubbel voor.

3.4 Luistertests

3.4.1 De Luisteraars

Aan de luisterproeven werd deelgenomen door zeven dames en dertien heren. Alle luisteraars zijn I.P.O.-medewerkers geselecteerd op de voorwaarde dat zij het merendeel van hun collega's kennen. Twee luisteraars waren tevens spreker, zodat hun stemmen ook op de banden voorkwamen. Er werd echter naar gestreefd om zo weinig mogelijk sprekers te laten luisteren.

De luisteraars werden verdeeld in drie groepen. De grootte van de groepen werd bepaald door de tijdstippen waarop de luisteraars beschikbaar waren. Aan elk van de drie groepen werden de banden in een andere volgorde aangeboden. In tabel 3.1 zijn de verdelingen weergegeven.

Tabel 3.1: *De verdelingen van het aantal luisteraars en de bandvolgordes over de drie luistergroepen.*

Groep	Luisteraars	Volgorde		
groep 1	4	1. PCM	2. LPC	3. MEA
groep 2	11	1. PCM	2. MEA	3. LPC
groep 3	5	1. MEA	2. LPC	3. PCM

3.4.2 De Luisterruimte

De luistertests werden afgenomen in een geluiddichte ruimte, voorzien van tafels, stoelen en hoofdtelefoons. De tafels en stoelen waren opgesteld in rijen, met tussen elke rij een scheidingswand of een open ruimte van ongeveer een meter. Hierdoor werd het voor de luisteraars onmogelijk om te zien wat de anderen opschreven.

De spraakfragmenten waren te horen door de hoofdtelefoons (Pioneer, type Monitor 10). Deze waren aangesloten op een bandrecorder (Revox, type B 77 MK II), die was opgesteld in een naast de luisterruimte gelegen kamer.

3.4.3 De Formulieren

Om een spreker te identificeren konden de luisteraars een persoon kiezen uit een lijst met dertig namen. Op deze lijst komen de namen voor van alle sprekers, met uitzondering van de twee voor de luisteraars onbekende sprekers. Deze veertien namen zijn nog aangevuld met de namen van een aantal willekeurig gekozen I.P.O.-medewerkers, niet zijnde dames of buitenlanders. Aan elk van de luisteraars werd een viertal formulieren, genummerd van 1 tot en met 4, uitgereikt.

In "Formulier 1" dienden de luisteraars op een zevenpuntsschaal de zogenaamde familiariteit aan te geven. Dit houdt in dat ze moesten aangeven hoe goed ze de stemmen van de eerder genoemde dertig personen kennen.

"Formulier 2" bevat de luisterinstructies. Hierin staat beschreven hoe de luisteraars de test dienden af te leggen. In dit formulier wordt ook vermeld dat een fragment van een bepaalde spreker meerdere keren voor kan komen.

"Formulier 3" is het antwoordformulier. Hierin vulden de luisteraars de namen van de sprekers in en gaven ze aan of ze het antwoord zeker wisten, of ze twijfelden, of dat ze gokten.

"Formulier 4" is een lijst met de dertig namen, waaruit gekozen kon worden. De vier formulieren zijn te vinden in Appendix B.

3.4.4 De Procedure

Tijdens elke luistertest kon steeds één groep naar een band luisteren. De tijd die verstreek - voor één groep - tussen twee luistersessies was ongeveer een dag.

Alvorens de luisteraars daartoe plaats namen in de luisterkamer werden voor iedere luisteraar een pen en een exemplaar van elk van de formulieren 2, 3 en 4 klaargelegd. Indien een groep naar de eerste band ging luisteren, werd hieraan nog een exemplaar van formulier 1 toegevoegd. In dit laatste geval werden de luisteraars verzocht om eerst formulier 1 in te vullen. Nadat deze formulieren waren opgehaald, werd begonnen met de eigenlijke luistertest. In het andere geval werd deze procedure overgeslagen en werd meteen met de test begonnen.

Zodra alle luisteraars een hoofdtelefoon hadden opgezet werd, zonder verdere mededelingen, de band gestart. De band werd één keer in zijn geheel afgespeeld. Na afloop van het laatste fragment werden alle formulieren opgehaald. Nadat de luisteraars alle drie de banden hadden gehoord, werden ze in de gelegenheid gesteld om hun antwoorden te vergelijken met die van de andere luisteraars en met de juiste antwoorden. Omdat op alle drie de banden dezelfde stemmen voorkomen, werd dat niet toegestaan na het horen van de eerste of de tweede band.

Het was in principe niet mogelijk voor de luisteraars om schriftelijk commentaar te leveren. Indien men hieraan echter toch behoefte had, dan kon men op het antwoordformulier enkele opmerkingen schrijven, of men kon na afloop mondeling commentaar geven. Het kwam voor dat luisteraars schriftelijk of mondeling commentaar gaven.

Hoofdstuk 4

Resultaten

Spreker identificatie

Voor PCM, LPC en MEA spraak is het gemiddeld aantal correct herkende sprekers (=goed antwoord) en het percentage goede antwoorden berekend. Hierin zijn alleen de antwoorden verwerkt die gegeven zijn bij de fragmenten van de 14 voor de luisteraars bekende sprekers. In tabel 4.1 zijn deze resultaten vermeld voor alle 20 luisteraars uit de 3 verschillende groepen samen. Tevens zijn in deze tabel de standaarddeviaties (std.dev.) en de middelbare afwijkingen (middelb.afw., standaarddeviaties van het gemiddelde) vermeld. De standaarddeviaties zijn volgens de volgende formule berekend:

$$\sigma_{n-1} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (4.1)$$

De middelbare afwijkingen zijn met behulp van de onderstaande formule berekend:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} \quad (4.2)$$

In de twee bovenstaande vergelijkingen is x_i het percentage goede antwoorden van luisteraar i ($0 \leq i \leq 20$), \bar{x} het gemiddelde percentage goede antwoorden, en n het aantal metingen waarover gemiddeld is.

De tabellen met bovenstaande resultaten van de 3 verschillende groepen afzonderlijk staan in de appendix A. Voor de 3 spraaksoorten afzonderlijk zijn de betrouwbaarheidsintervallen van het percentage goede antwoorden berekend met behulp van de volgende formule:

$$\bar{x} - t\sigma_x < m < \bar{x} + t\sigma_x \quad (4.3)$$

Tabel 4.1: *Gemiddeld aantal en percentage correct herkende sprekers, met standaarddeviatie.*

Goede Antwoorden	PCM		LPC		MEA	
	aantal	%	aantal	%	aantal	%
gemiddeld aantal	12	85.7	10.4	73.9	6.4	45.7
std.dev.	1.9	13.9	2.1	15.3	3.8	27.4
middelbare afwijking	0.4	3.1	0.5	3.4	0.8	6.1

Hierin is $\sigma_{\bar{x}}$ bovenstaande middelbare afwijking (zie vergelijking 4.2), en \bar{x} het gemiddelde percentage goede antwoorden. De waarde van t hangt af van het aantal vrijheidsgraden (in dit geval het aantal luisteraars verminderd met 1 ($n-1$)) en de zekerheid waarmee de uitspraak gedaan wordt (uitgedrukt in de kans dat een bepaalde waarde binnen een bepaald interval ligt). Voor de waarde van t is gebruik gemaakt van de tabel "Percentile values(t_p) for student's t distribution", (Spiegel, 1972). In vergelijking 4.3 is m de waarde waarnaar het gemiddeld percentage goede antwoorden nadert als het aantal luisteraars (oftewel n) naar oneindig gaat. Met behulp van vergelijking 4.3 is het 95% betrouwbaarheidsinterval bepaald voor de drie spraakversies ($\alpha = 0.1$).

- PCM: $80.3\% < m < 91.1\%$
- LPC: $69.0\% < m < 78.8\%$
- MEA: $35.1\% < m < 56.3\%$

Het verlies van identificatievermogen

Een manier om het verlies van identificatievermogen weer te geven is het relateren van het percentage goede antwoorden van de MEA en LPC spraak aan het percentage goede antwoorden van de PCM spraak. Dit kan met behulp van de index van het verlies in sprekeridentificatievermogen. Voor de LPC spraak wordt deze index:

$$Index(LPC) = \frac{\text{percentage goede antwoorden van de LPC spraak}}{\text{percentage goede antwoorden van de PCM spraak}} \quad (4.4)$$

Voor de MEA spraak wordt deze index:

$$Index(MEA) = \frac{\text{percentage goede antwoorden van de MEA spraak}}{\text{percentage goede antwoorden van de PCM spraak}} \quad (4.5)$$

In tabel 4.2 zijn deze indices als functie van de (vijfpunts)familiariteit (fam., zie ook de volgende paragraaf: "Familiariteit") voor LPC en MEA spraak weergegeven. In deze tabel is de zevenpuntsschaal gereduceert tot een vijfpuntsschaal door de resultaten van de 2 laagste schalen (1 en 2) en de 2 hoogste schalen (6 en 7) bij elkaar te nemen. Dit is gedaan omdat uit het verloop van het percentage goede antwoorden als functie van de familiariteit (zie tabel 4.3) blijkt dat de laagste schaal te weinig gekozen is en omdat voor schaal 7 een verzadigingseffect optreedt. Verder is in tabel 4.2 de gemiddelde index (gem.index) voor LPC en MEA spraak weergegeven.

Tabel 4.2: *Index van het verlies in sprekeridentificatie voor LPC en MEA spraak.*

Fam.	PCM	LPC		MEA	
	%goed	%goed	index	%goed	index
1+2	52.2	47.8	0.916	19.6	0.375
3	83.3	70.0	0.840	36.7	0.441
4	91.2	76.5	0.839	41.2	0.452
5	91.1	82.1	0.901	48.2	0.529
6+7	95.6	80.7	0.844	58.8	0.615
Gem.index			0.868		0.482

Familiariteit

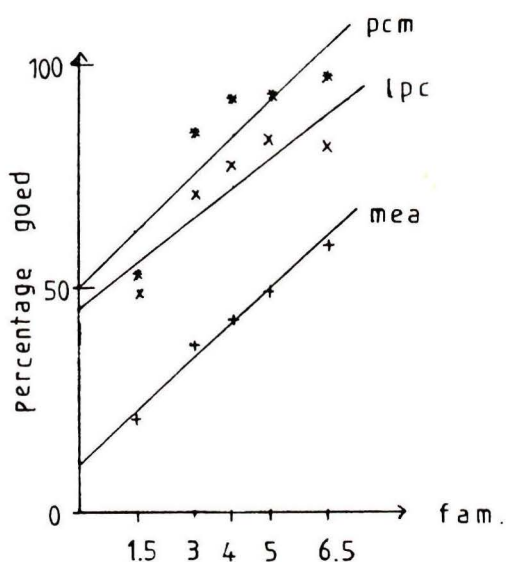
De luisteraars hebben voor de luistertesten formulier 1 (zie appendix B) ingevuld, waarin zij de familiariteit met de personen op het formulier schaalden in een zevenpuntsschaal van 1 tot 7, waarbij schaal 1 aangeeft dat de luisteraar de stem van de persoon niet kent en schaal 7 aangeeft dat de luisteraar de stem van de persoon zeer goed kent. De familiariteit met de 14 personen uit deze lijst die ook werkelijk op de stimulusband voorkwamen is gemiddeld over de 14 sprekers en over alle 20 luisteraars. Het resultaat hiervan is een gemiddelde familiariteit van 4.7.

Van de antwoorden, behorend bij de verschillende schalen van familiariteit, is het percentage goede antwoorden uitgerekend. Deze percentages, als functie van de bijbehorende familiariteitsschaal (fam.), zijn, samen met de aantallen antwoorden, voor de 3 verschillende spraaksoorten apart vermeld in tabel 4.3. Tevens is in deze tabel vermeld hoe vaak er gescoord werd op een bepaalde waarde van de fam.schaal (Voork. fam.).

Tabel 4.3: *Verdeling van het percentage goede antwoorden als functie van de familiariteit (vijfpuntsschaal).*

Fam.	Voork. Fam.	PCM		LPC		MEA	
		goed	%goed	goed	%goed	goed	%goed
1+2	46	24	52.2	22	47.8	9	19.6
3	30	25	83.3	21	70	11	36.7
4	34	31	91.2	26	76.5	14	41.2
5	56	51	91.1	46	82.1	27	48.2
6+7	114	109	95.6	92	80.7	67	58.8

Dezelfde resultaten als in tabel 4.3, maar dan voor de oorspronkelijke zevenpuntsschaal zijn vermeld in tabel A.4 (zie appendix A). Het percentage goede antwoorden als functie van de (vijfpunts)familiariteit uit tabel 4.3 is uitgezet in een grafiek (zie figuur 4.1), waarbij de resultaten van PCM, LPC en MEA spraak gescheiden zijn gehouden.



Figuur 4.1: Percentage goede antwoorden versus de familiariteit.

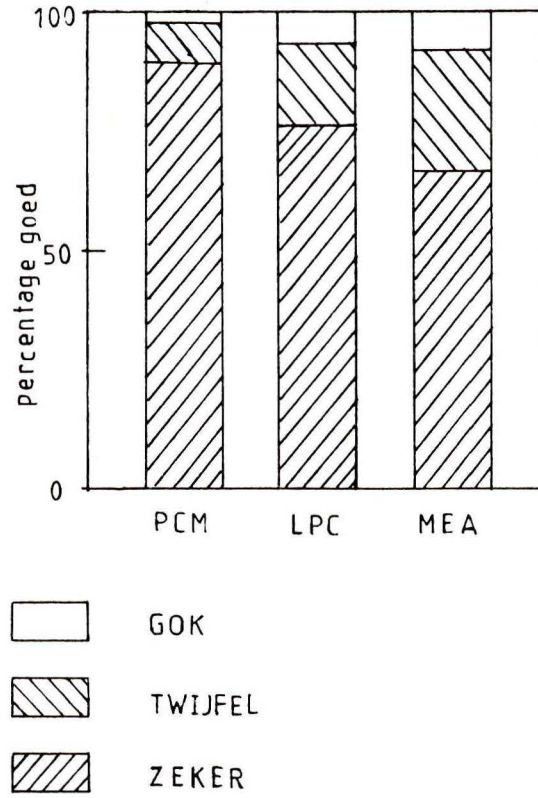
Met behulp van de kleinste kwadratenaanpassing zijn de beste rechte lijnen voor het verband tussen het percentage goede antwoorden en de mate van familiariteit van de 3 spraaksoorten berekend, met als resultaat:

- PCM: $Y=8.0X+50.6$
- LPC: $Y=6.5X+45.4$
- MEA: $Y=7.6X+10.7$

In figuur A.1 in appendix A zijn dezelfde resultaten weergegeven als in grafiek 4.1, maar dan op de oorspronkelijke zevenpuntsschaal.

De zekerheid waarmee de goede antwoorden gegeven zijn

Van de totale aantallen goede antwoorden van de PCM, LPC en MEA spraak is berekend welk percentage op een gok berust, bij welk percentage getwijfeld is, en welk percentage met zekerheid gegeven is. Deze percentages zijn voor de 3 spraaksoorten apart weergegeven in figuur 4.2. Hierbij is gebruik gemaakt van de resultaten van formulier 3 (zie appendix B), waarin de luisteraars de zekerheid waarmee ze het antwoord gaven, schaalden op een driepuntsschaal (gok, twijfel, zeker).



Figuur 4.2: *Percentage gok, twijfel en zeker voor de goede antwoorden.*

Hoofdstuk 5

Discussie/Conclusies

De in de resultaten genoemde betrouwbaarheidsintervallen voor het percentage goede antwoorden bij PCM, LPC en MEA overlappen elkaar niet. Omdat de zekerheid dat het percentage in het interval ligt 95% is, is de conclusie gerechtvaardigd dat de herkenbaarheid van de spreker afneemt van PCM naar LPC naar MEA en dus afneemt bij reductie van de bitstroom. Bovendien blijkt uit figuur 4.2 dat bij herkenning van een spreker het percentage twijfelgevallen toeneemt van PCM naar LPC naar MEA.

Een van de voorwaarden van het experiment is dat de luisteraars de sprekers redelijk goed kennen. Uit de gemiddelde familiariteit, deze is 4.7 op een zevenpuntschaal, blijkt dat deze doelstelling is bereikt. Uit tabel 4.2 blijkt dat in het geval van LPC-spraak, de afname van de sprekeridentificatie-score niet afhankelijk is van de familiariteit (index is ongeveer constant over de verschillende waarden van de familiariteit). In het geval van MEA-spraak wordt de herkenning van de onbekende sprekers (lage familiariteit) sterker aangetast dan de herkenning van bekende sprekers (hoge familiariteit). Dit kan als volgt aannemelijk gemaakt worden:

Bij goed bekende personen is een kleine aanwijzing vaak voldoende voor identificatie, dus als een groot deel van de kenmerken verdwenen zijn, kunnen de overgebleven kenmerken nog tot herkenning leiden. Bij minder bekende personen zijn (relatief) meer kenmerken nodig, voordat herkenning optreedt.

leereffect

Bij de luistertesten zijn voor de verschillende soorten spraak (PCM, LPC en MEA) dezelfde fragmenten gebruikt. Hierdoor is het mogelijk dat men een spreker identificeerd doordat men een karakteristiek fragment van deze spreker al een keer heeft gehoord. Om inzicht te krijgen in de grootte van dit leerproces zijn de luisteraars opgedeeld in drie groepen die de PCM, LPC en MEA spraak in verschillende volgorde kregen aangeboden. In tabel 5.1 is de aan de verschillende groepen aangeboden volgorde met het daarbij behorende percentage goede antwoorden voor de MEA versie vermeld.

Uit deze gegevens blijkt dat het leereffect groot is. Hoewel de spreiding in de percentages goede antwoorden voor de verschillende versies groot is vertonen alle

Tabel 5.1: *Bandvolgordes over de drie luistergroepen en percentage goede antwoorden voor de MEA versie.*

Groep	Volgorde			percentage goed MEA
groep 1	1. PCM	2. LPC	3. MEA	69.6
groep 2	1. PCM	2. MEA	3. LPC	52.6
groep 3	1. MEA	2. LPC	3. PCM	11.4

groepen de zelfde trend: de herkenbaarheid van de spreker neemt steeds af van PCM naar LPC naar MEA. Daarom is toch besloten bij de verwerking van de resultaten de verschillende groepen samen te voegen. Bij een eventuele herhaling van het experiment zou men het leerproces op de volgende manieren kunnen verkleinen:

- Bij de keuze van het stimulusmateriaal nog strenger te letten op het vocabulair van het spelletje. Men zou voor elke spreker slechts vijf fragmenten kunnen gebruiken waarin de woorden "alfa", "beta", "charlie", "delta" en "echo" met een willekeurig getal voorkomen. Omdat het voor iedere spreker mogelijk is een dergelijke serie samen te stellen zal het leereffect veel kleiner zijn.
- Door de tijd tussen het aanbieden van de verschillende versies te vergroten is de kans kleiner dat de luisteraars fragmenten onthouden en wordt het leereffect tegengegaan.
- Het leereffect kan helemaal omzeild worden door er voor te zorgen dat geen enkele persoon een fragment meerdere keren te horen krijgt. Dit kan door de hoeveelheid fragmenten zodanig te vergroten dat elke luisteraar of groep luisteraars een unieke stimulusband te horen krijgt. Het nadeel hiervan is dat het vergelijken van de resultaten voor de verschillende versies veel moeilijker wordt. Bovendien is het innemen, segmenteren, analyseren, en resynthetiseren van een dermate groot aantal fragmenten een tijdrovende bezigheid.
- Een andere methode om het leereffect te elimineren is het gebruik maken van een zo groot aantal luisteraars dat voor elke luistertest nieuwe luisteraars gebruikt kunnen worden, waardoor geen enkele luisteraar een fragment dubbel krijgt te horen. Het nadeel hiervan is dat het vergelijken van de resultaten voor de verschillende versies weer erg moeilijk wordt en dat men moet beschikken over een groot aantal luisteraars.

Het spel

De spelvorm blijkt een geslaagde manier te zijn om spontaan spraakmateriaal te verzamelen. Een halve competitie tussen vier spelers stimuleert het spontane element omdat de meesten elkaar kennen en er vaak inderdaad van competitie sprake is. Bovendien leverde de uitvoering van het spel geen problemen op voor de spelers (de meesten waren zelfs al bekend met "zeeslag"). De 24 opnames bevatten genoeg materiaal van de kwaliteit die voor verdere verwerking vereist was. De mogelijkheid bestaat om alleen de "alfa"'s, "beta"'s enzovoort te gebruiken, maar ook om meer uitgebreide zinsdelen te selecteren. In verband met het leereffect moet de keuze hier echter zorgvuldig overwogen worden.

Al met al is uit het identificatie experiment gebleken dat de verschillende reproductietechnieken inderdaad sprekerkenmerken in verschillende mate aantasten. Aantasting van sprekerkenmerken is dus een bruikbare kwaliteitsmaat voor gereproduceerde spraak. Verder kan het spel "Zeeslag" uitstekend dienst doen als leverancier van spontane spraak met een beperkte woordenschat.

Literatuur

- Schmidt-Nielsen, A. & Stern, K.R. (1985), Identification of known voices as a function of familiarity and narrow-band coding, *J.Acoust.Soc.Am.* 77 (2) 658-663.
- Schmidt-Nielsen, A. & Everett, S.S. (1982), A Conversational Test for Comparing Voice Systems Using Working Two-Way Communication Links, *IEEE transactions on acoustics, speech, and signal processing*, vol. *assp-30*,no.6,december 1982 853-863.
- Hart, 't, J & Nooteboom, S.G. & Vogten,L.L.M. & Willems, L.F. (1981/82), Manipulaties met spraakgeluid, *Philips techn.T.40*,1981/82,no.4 108-119.
- Spiegel, M.R. (1972), Theory and problems of statistics, *McGraw.Hill Book Company*,New York 344.
- Hermes, D.J. (1987), Measurement of pitch by subharmonic summation, *Journal of the Acoustical Society of America*.
- Vogten, L.L.M. (1985), LVS-speech processing programs on IPO-VAX 11/780, *Handleiding* 67.
- VAXVMS Primer, *Digital equipment corporation*, 1982.
- Goossens, J.J.M. (1986), De invloed van opnamecondities op analyse en resynthese van spraak, *IPO rapportnr.* 521.

Appendix A

Resultaten

Het percentage goede antwoorden voor groep1, met volgorde van aanbod PCM, LPC, MEA staat in Tabel A.1.

Tabel A.1: *Percentage goede antwoorden met mate van zekerheid voor groep1.*

Antwoorden	PCM	LPC	MEA
%goed	85.7	75.0	69.6
waarvan: gok	2.1	2.4	2.6
twijfel	4.2	9.5	17.9
zeker	93.8	88.1	79.5

Het percentage goede antwoorden voor groep2, met volgorde van aanbod: PCM, MEA, LPC staat in Tabel A.2.

Tabel A.2: *Percentage goede antwoorden met mate van zekerheid voor groep2.*

	PCM	LPC	MEA
%goed	83.8	79.2	52.6
waarvan: gok	1.6	3.3	7.4
twijfel	9.3	15.6	29.6
zeker	89.1	81.1	63

Het percentage goede antwoorden voor groep3 , met volgorde van aanbod: MEA, LPC, PCM staat in Tabel A.3.

Het aantal goede en percentage goede antwoorden voor de drie spraakversies als functie van de familiariteit staat weergegeven in Tabel A.4.

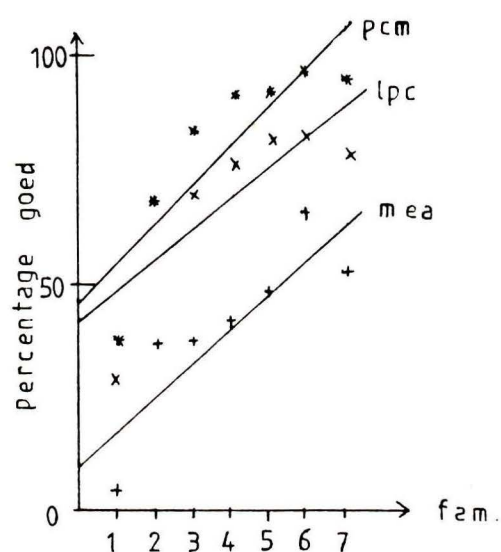
De familiariteit geeft aan hoe goed een luisteraar de sprekers kent. Ze is onderverdeeld in een zevenpuntsschaal van 1 tot 7, waarbij schaal 1 aangeeft dat de luisteraar de stem van de spreker niet kent en schaal 7 aangeeft dat de luisteraar de stem van de spreker zeer goed kent. Bovenstaande verdeling is uitgezet in een grafiek (zie figuur A.1).

Tabel A.3: *Percentage goede antwoorden met mate van zekerheid voor groep 3.*

	PCM	LPC	MEA
%goed	90	61.4	11.4
waarvan: gok	3.2	18.6	37.5
twijfel	9.5	30.2	25.0
zeker	87.3	51.2	37.5

Tabel A.4: *Verdeling van het percentage goede antwoorden als functie van de familiariteit.*

Fam.	Voork. Fam.	PCM		LPC		MEA	
		goed	%goed	goed	%goed	goed	%goed
1	24	9	37.5	7	29.2	1	4.2
2	22	15	68.2	15	68.2	8	36.4
3	30	25	83.3	21	70	11	36.7
4	34	31	91.2	26	76.5	14	41.2
5	56	51	91.1	46	82.1	27	48.2
6	57	55	96.5	47	82.5	37	64.9
7	57	54	94.7	45	78.9	30	52.6



Figuur A.1: *Percentage goede antwoorden versus de familiariteit*

Appendix B

Formulieren

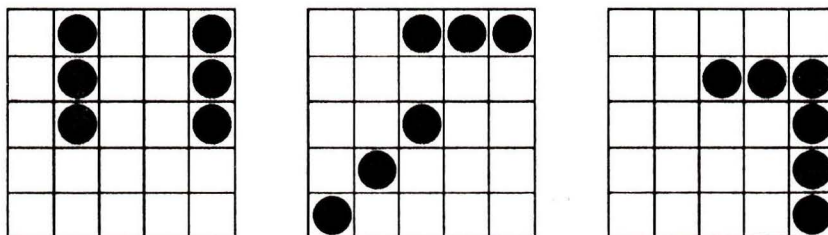
Spel-instructies : N SIDE ¹

De opzet van dit experiment is om uit de geluidsopname van het dadelijk te spelen spel fragmenten te gebruiken voor een spraakherkenningstest. Voor een goed verloop van het experiment is het van belang dat u zo min mogelijk tegelijkertijd praat, en dat de afstand tot de microfoon ongeveer 20 cm. bedraagt.

Spelregels

Het spel wordt gespeeld op de twee velden afgedrukt op uw spelformulier. U en uw tegenstander hebben beiden twee schepen die elk drie cellen verticaal, horizontaal of diagonaal beslaan.

Voorbeelden van schipposities



Doel van het spel is om als eerste de twee schepen van de tegenstander te doen zinken. U speelt om de beurt. Elke beurt bestaat uit een “*schot*”. Om te schieten moet u een cel in het veld van uw tegenstander aanduiden (bijv. Alfa-1, Charlie-5, Delta-3 etc.). Uw tegenstander markeert de door u aangeduide cel en vertelt u of het “*schot*” raak was of mis. Een schip zinkt indien alle drie de cellen waaruit het bestaat getroffen zijn. Noteer uw schoten in het rechterveld (om bij te houden welke schoten raak waren), en die van uw tegenstander in het linkerveld.

Plaats nu uw schepen in het linkerveld en laat uw tegenstander weten wanneer u gereed bent om te spelen.

Veel Succes !!!

¹Natuurkunde-practicum: Spreker-IDentificatie-Experiment

Spel-formulier : N SIDE ²

Uw naam _____
Naam tegenspeler _____
Datum _____
Spelnummer _____

	1	2	3	4	5
A					
B					
C					
D					
E					

Schoten *TEGENSPELER*

	1	2	3	4	5
A					
B					
C					
D					
E					

UW schoten

- A = Alpha
- B = Beta
- C = Charlie
- D = Delta
- E = Echo

²Natuurkunde-practicum: Spreker-IDentificatie-Experiment

FORMULIER 1

Naam:

U wordt verzocht in onderstaande tabel aan te geven hoe goed u de stem van de personen kent. Lees eerst de namenlijst door en kruis dan bij elke naam het hokje van uw keuze aan.

Nr	Naam	Ik ken de stem van de persoon:						
		niet					zeer goed	
1	Aarts (Twan)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Bruurs (Ad)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Deliege (Rene)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Eggen (Berry)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Ellermann (Henk)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Elsendoorn (Ben)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	Griendt, van der (Henk)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Hart, 't (Hans)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Hermes (Dik)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Hertog, den (Jacob)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Houtsma (Aad)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	Jong, de (Theo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Kerkhof (Theo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	Leeuwen, van (Hugo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	Moonen (Gerard)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Nes, van (Floris)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	Neve (Han)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	Nooteboom (Sieb)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	Pijper, de (Jan-Roelof)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20	Reitsma (Pieter)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21	Ridder, de (Huib)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22	Rijnsoever, van (Pieter)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23	Spaai (Gerard)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24	Terken (Jacques)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25	Theelen (Paul)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26	Tiesinga (Jan)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27	Wagenaars (Wil)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28	Waterham (Ronald)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29	Westhoff (Koos)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30	Willems (Lei)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

FORMULIER 2

Luisterinstructies

U gaat dadelijk luisteren naar een band met achttien spraakfragmentjes. Hierbij wordt u verzocht in formulier drie, op de juiste plaats, de naam van de spreker en de mate van uw zekerheid (gok, twijfel, zeker) in te vullen. Indien u absoluut niet weet wie de spreker is, zet u een streepje in plaats van de naam. Probeer echter zo veel mogelijk namen in te vullen.

Het is niet noodzakelijk dat alle op formulier vier vermelde personen op de band voorkomen. Het is mogelijk dat er meerdere fragmenten van een spreker zijn.

Na de aankondiging "fragment .." volgt het eigenlijke fragment, dat ongeveer twintig seconden zal duren. Na elk fragment volgt een korte pauze om u in de gelegenheid te stellen de naam van de spreker in te vullen.

Veel succes !

FORMULIER 3

Naam:

fragment nummer	spreker	zekerheid		
		gok	twijfel	zeker
1	[]	[]	[]
2	[]	[]	[]
3	[]	[]	[]
4	[]	[]	[]
5	[]	[]	[]
6	[]	[]	[]
7	[]	[]	[]
8	[]	[]	[]
9	[]	[]	[]
10	[]	[]	[]
11	[]	[]	[]
12	[]	[]	[]
13	[]	[]	[]
14	[]	[]	[]
15	[]	[]	[]
16	[]	[]	[]
17	[]	[]	[]
18	[]	[]	[]

Bedankt voor uw medewerking!

FORMULIER 4

Namen van de sprekers waaruit u kunt kiezen.

Nr	Naam
1	Aarts (Twan)
2	Bruurs (Ad)
3	Deliege (Rene)
4	Eggen (Berry)
5	Ellermann (Henk)
6	Elsendoorn (Ben)
7	Griendt, van der (Henk)
8	Hart, 't (Hans)
9	Hermes (Dik)
10	Hertog, den (Jacob)
11	Houtsma (Aad)
12	Jong, de (Theo)
13	Kerkhof (Theo)
14	Leeuwen, van (Hugo)
15	Moonen (Gerard)
16	Nes, van (Floris)
17	Neve (Han)
18	Nooteboom (Sieb)
19	Pijper, de (Jan-Roelof)
20	Reitsma (Pieter)
21	Ridder, de (Huib)
22	Rijnsoever, van (Pieter)
23	Spaai (Gerard)
24	Terken (Jacques)
25	Theelen (Paul)
26	Tiesinga (Jan)
27	Wagenaars (Wil)
28	Waterham (Ronald)
29	Westhoff (Koos)
30	Willems (Lei)

Appendix C

Lijst van alle op de stimulusbanden gebruikte fragmenten

Spreker: Pieter van Rijnsoever

- "Alfa 1"
- "Beta 2"
- "Charlie 3"
- "Delta 3"
- "Echo 5"
- "Die is ook raak"
- "Misgeschoten"
- "Da's nog raak ook"
- "Ook al raak"
- "Da's een voltreffer"
- "Die is mis"

Spreker: Jacob den Hertog

- "Ik begin met charlie 3"
- "Dat is ook een voltreffer"
- "Beta 4,... is mis"
- "Charlie 4"
- "Beta 3"
- "Echo 3 die is raak"
- "Alfa 3 is raak"
- "Delta 3 is raak"
- "Die is mis"
- "Dan krijg ik hier een charlie 5"

Spreker: Dik Hermes

- "Charlie 2"
- "Die is ook mis"
- "Alfa 3"
- "Die is raak"
- "Echo 3"

"Ja, jij bent"
"Delta 3"
"Dan ben jij nu"
"Beta 2"
"Nou dat kan alleen nog maar charlie 1 zijn"

Spreker: Jacques Terken

"Charlie 4"
"Beta 5, mis"
"Delta 2"
"Echo 5, die is ook raak"
"Ik voorspel dat charlie 3 dan ook raak is"
"Alfa 3, raak"
"Epsilon 2"
"Die is helaas mis"
"Charlie 5 is mis"

Spreker: Hans 't Hart

"Ik ben klaar"
"Nou mag jij beginnen"
"Charlie 3"
"Beta 2"
"Delta 3"
"Echo 1"
"Ja, ik ben"
"Alfa 3"
"Raak,... en gezonken"
"Sorry, charlie 2, heb je al geschoten"

Spreker: Jan Tiesinga

"Moment"
"Wat mij betreft kan het ook"
"Beta 2"
"Delta 3"
"Charlie 3"
"Echo 3"
"Alfa 3"
"Raak en gezonken"
"Dan zijn we klaar"
"We zijn er"

Spreker: Han Neve

"Beta 2"
"Charlie 3"
"Raak"
"Alfa 3"
"Delta 3"
"Ja, dan zal het wel afgelopen zijn"
"Raak en gezonken"

"Ja, euh, goed, ik schiet mis, dus..."
"Echo 3"
"Ja dan zit jij op echo 1 natuurlijk"

Spreker: Henk van der Griendt

"Charlie 3"
"Delta 4"
"Beta 4"
"Echo 2"
"Raak en gezonken"
"Mis,... alfa 1"
"Dit is het einde van het spel, hè"
"Raak en gezonken, je hebt gewonnen"
"Is er een gezonken?"
"Ook raak en allemaal gezonken, van jou ook"

Spreker: Ben Elsendoorn

"Beta 2"
"Charlie 2, mis"
"Delta 4 voor mij"
"Alfa 2"
"Mag ik van u, echo 1"
"U heeft gewonnen"
"Euh, charlie 1"
"Mag ik dan van u"
"Met een misser en zes raak"
"Die is mis"

Spreker: Ronald Waterham

"Doe mij, euh, beta 2"
"Charlie 3"
"Euh, 'n delta 4"
"Die was raak"
"Echo 4"
"Ook mis, goed zo"
"Ja, jij bent goed aan 't schieten"
"Alfa 3 van jou"
"Tuurlijk ook raak"
"Ik had je niet moeten laten beginnen"

Spreker: Henk Ellermann

"Charlie 3"
"Beta 2 is mis"
"Dan wou ik graag echo 1"
"Alfa 4, die is raak"
"Zou ik van jou dan delta 5 mogen?"
"Nou ik zal eens een te hoog schieten"
"Ja, ik durf niet langer"
"Begin jij dit keer maar, ik mocht de vorige keer beginnen"

"Dan neem ik van jou maar eens..."

"Ja, die 's raak"

Spreker: René Deliege

"Beta 2, is gelijk raak"

"Dan voor mij, charlie 2"

"Delta 3 is mis"

"Echo 4 is mis"

"Alfa 5 is ook raak"

"Nou goed, zal ik compenseren door een keer naast te schieten"

"Nou vuur maar"

"Zo, da's de eerste"

"Nou, 's een gokje wagen"

"Bij jou dan alfa 2"

Spreker: Arthur Peeters

"Euh, ik schiet op beta 2"

"Alfa 2"

"Die is raak "

"Charlie 5 is mis"

"Delta 2 is mis"

"Euh, nou ja, goh"

"Nou, wie mag t'er beginnen"

"Die is al gelijk raak"

"Echo 3"

"Ja, dan schiet ik maar op euh"

Spreker: Berry Eggen

"Ik schiet op euh,... beta 2"

"Alfa 5"

"Delta 3"

"Charlie 5 is raak"

"Echo 3 is... euh... mis"

"Charlie 2, jij weet ze wel te verzinnen, die is ook weer mis"

"Nou moet ik alleen nog tegen"

"Alfa 3, goed gedacht"

"Een stukje stilte opnemen"

"Geweldig jongen"

Spreker: Hugo van Leeuwen

"Charlie 3"

"Echo 1"

"Delta 3"

"Die is raak"

"Beta 4"

"Alfa 5, gelijk spel"

"Goed, ik zal wel effe blijven zitten dan"

"Mis"

"Die is mis en je hebt verloren"

"Charlie 5, einde spel"

Spreker: Patrick Briennesse

"Ik schiet, euh, beta 4"

"Die is raak"

"Dan schiet ik op, euh, charlie 5"

"Die is hartstikke mis"

"Ik schiet op echo 3"

"Alfa 4 is mis"

"Nou dan gok ik op delta 1"

"Goed dan begin ik wel, of nie...?"

"Nou dan doe ik beta 1"

"Jij was toch begonnen?"