

Samenspraak Fonetische kennis verwerven en implementeren

Afscheidscollege prof. dr. ir. Louis C.W. Pols

Universiteit van Amsterdam, vrijdag 3 juni 2005, 14:00u

1. Inleiding

Samenspraak.
Fonetische kennis verwerven
en implementeren

Speaking together.
Acquiring and implementing
phonetic knowledge

Louis C.W. Pols

Farewell Lecture, Univ. van Amsterdam
Friday, June 3rd, 2005, 14:00 hrs.

Fig. 338

Zoals gebruikelijk bij een afscheidscollege zal ik mijn tekst in het Nederlands presenteren. However, as a courtesy to some of the foreign guests of today, I will show keywords of my text in English. This also allows me to present some interesting data in the form of tables or figures. Zowel kinderen als leergierige volwassenen *verwerven* voortdurend veel *kennis* die ze bewust of onbewust ten nutte maken in hun alledaags handelen. Voor universitaire studenten en voor actieve academici, en hopelijk ook nog wel voor met-emeritaatgaande hoogleraren, is die kennisverwerving en het toepassen daarvan, bijna een dagtaak. Ik wil het onderwerp in dit afscheidscollege nader toespitsen op het verwerven van *fonetische* kennis en het implementeren van die specifieke kennis, zowel in het alledaags handelen alsook in specifieke applicaties. Op die manier meen ik enige structuur te kunnen aanbrengen in de onderwerpen die ik vanmiddag aan de orde wil stellen. Het ligt immers voor de hand dat ik zowel wil teruggrijpen op, alsook extrapoleren over, al die onderwerpen waarmee ik me sinds oktober 1982 (toen ik als hoogleraar Fonetische Wetenschappen op deze universiteit werd benoemd) heb bezig gehouden en die me tot op de dag van vandaag zijn blijven boeien.

Introduction

- what is phonetic sciences?
- what is speech?
- phoneticians vs. other linguists
- variability in speech
- acquiring phonetic knowledge by language learners, language users and speech technology systems

June 3, 2005 Farewell lecture 2

Omdat er hier vanmiddag in deze mooie aula zeker niet alleen maar fonetici aanwezig zijn, wil ik allereerst mijn *vakgebied* kort omschrijven en globaal omkaderen, en ik zal daarbij niet aarzelen dat vakgebied en de aanpalende periferie, ruim te interpreteren. Fonetiek is immers een bij uitstek *interdisciplinair vakgebied* dat gaat over het spraaksignaal, de gesproken taal en over spraakcommunicatie, maar uiteraard ook over fonemen en over prosodie. Het gaat over spreken en luisteren en over mentale opslag en verwerking. Het gaat over vroege spraakverwerving en over spraakpathologie, maar het gaat ook over (taal- en) spraaktechnologie en spraakdatabestanden (al dan niet multimodaal of audio-visueel). Het gaat over talen van de wereld en over dialecten, en overigens over allerlei andere zaken waar fonetici goed in blijken te zijn, zoals laboratorium fonologie, het evalueren van oorimplantaten, of het ontwerpen van sprekende web animaties.

Mijn leerstoel Fonetische Wetenschappen, die van 1959 tot zijn overlijden in 1980 was bezet door mijn voorganger prof. dr. ir. Hendrik Mol, is nu op deze Universiteit van Amsterdam een onderdeel van de opleiding Taalwetenschap binnen de afdeling Taal- en Letterkunde van de faculteit der Geesteswetenschappen. Maar

binnen een andere universitaire structuur, waarvan er elders in de wereld diverse voorbeelden zijn, hadden we ook bij Psychologie, of bij Informatica of bij Geneeskunde kunnen behoren. In feite zijn er ook in Amsterdam actieve contacten met al die faculteiten via onderwijs en/of onderzoek.

Spraak blijkt een fantastisch informatief en efficiënt akoestisch signaal te zijn om talige communicatie mogelijk te maken. Dit betreft zowel communicatie tussen mensen (kinderen en volwassenen) alsook tussen mens en machine. Spraak is daarbij niet slechts tot-geluid-omgetransformeerde tekst, maar bevat tevens allerlei informatie over de spreker zelf (sekse, leeftijd, gewicht), over de gezondheids- en gemoedstoestand waarin deze spreker verkeert (verkouden, zenuwachtig of boos), over zijn akoestische omgeving (galmende kamer of vrije veld) of over het communicatiekanaal dat hij gebruikt (mobieltje of rechtstreeks) [68]. Daarbij bevat het spraaksignaal ook informatie over het spreektempo, over de mate van beklemtoning of reductie van lettergrepen of woorden, over de gerealiseerde zinsaccenten, over het vragend of bevestigend zijn van een zin, over de mate waarin iemand dialect- of moedertaalspreker is, of dat het hier kennelijk om machinale spraak gaat, en over nog veel meer. Een wezenlijk verschil tussen fonetici en de meeste andere taalkundigen, is dat een foneticus al die variatie in het spraaksignaal juist boeiend en uitdagend vindt, terwijl de meeste andere taalkundigen zich hier zo snel en zo veel mogelijk van proberen te distantieren. Niettemin is het ook voor fonetici van groot belang te doorgronden hoe dat extreem variabele spraaksignaal kan worden getransformeerd in een normatieve waarneming in de vorm van fonemen, woorden, zinnen en dialogen, en hoe jonge kinderen en niet-moedertaalsprekers dat spreken en verstaan aanleren.

Hoe de variabiliteit binnen en tussen sprekers door *luisteraars* wordt aangepakt, is een researchprobleem dat ook nog steeds niet is doorgrond, en al helemaal niet binnen de *spraaktechnologie*. Ook is het fascinerend je te realiseren dat enerzijds hele subtiele details in het spraaksignaal essentieel zijn, b.v. om een onderscheid te kunnen maken tussen ‘hij spreekt de kerel soms’ (enkelvoud) of ‘de kerels soms’ (meervoud) [24], of b.v. om op het juiste moment in te kunnen breken in een dialoog met een zgn. beurtwisseling. Terwijl anderzijds *sterk gereduceerde* spraak geenszins hoeft te lijden tot misverstanden (b.v. [x@mEnt] i.p.v. ‘op een gegeven moment’). Al die onderliggende fonetische kennis over het proces van spreken en verstaan en over het spraaksignaal zelf, moet verworven worden door eerste en tweede taalleerders, taalgebruikers en taal- en spraaktechnologische systemen.

Hiermee keer ik terug naar mijn hoofdonderwerp, namelijk het verwerven van fonetische kennis en het implementeren daarvan.

fonetische kennis verwerven bij/via/uit:	implementaties:	zie sectie
vroege spraakontwikkeling	in efficiënte talige communicatie	2.1
spraakpathologie	bij afwijkende communicatie	2.1, 2.2
fonetisch onderzoek	in fundamentele kennis of in applicaties	2.3
spraakdatabestanden	in regels of anderszins	2.4
studie fonetiek	ten behoeve van onderwijsstoetsing	2.5

Overview	
acquiring phonetic knowledge in/by	implementing phonetic knowledge
early speech development	in efficient language communication
speech pathology	in deviant communication
phonetics research	in fundamental knowledge, or applications
speech databases	in rules or otherwise
students of phonetics	for teaching purposes

June 3, 2005 Farewell lecture 3

2.1 Vroege spraakontwikkeling en spraakpathologie

Laat ik beginnen met de vroege spraakontwikkeling en de spraakpathologie. Hoe verwerft een kind fonetische kennis? Ik zou zeggen dat je dat als kind vooral ‘al doende’ leert, alhoewel er natuurlijk ook sturing via het onderwijs en de opvoeding plaats vindt. In onze leerstoelgroep is er altijd veel belangstelling geweest voor de hele vroege spraakverwerving, zeg maar in het eerste levensjaar. Precies die periode waarvoor de meeste taalkundigen minder belangstelling hebben omdat grammaticale structuren, woordvormen en woordbetekenis dan nog niet zo herkenbaar zijn. Het is echter wel de periode waarin, via de moeder-kind interactie en exploratie, de fundamentele tot prosodie, tot woord- en zinsstructuren en tot talige communicatie worden gelegd.

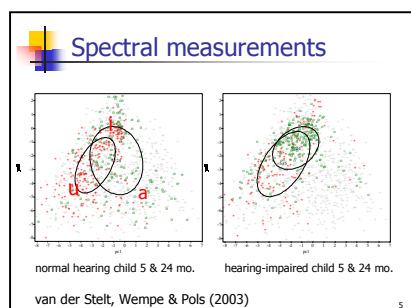
Early speech development	
<ul style="list-style-type: none"> ■ this work mainly supervised by Fl. v. Beinum ■ emphasis today on acoustic analyses of early vocalizations of deaf and normal-hearing young children (5 – 24 months) <ul style="list-style-type: none"> ■ early vocalizations difficult to label ■ high pitch; formants poorly defined ■ alternative: bandfilter analysis + PCA ■ changes in vowel space 5 -> 24 mo. ■ future: dynamic changes within one vocalization and comparison over languages 	

June 3, 2005 Farewell lecture 4

Behalve naar de vroege spraakverwerving van normale kinderen [19, 69] is er in onze groep ook veel onderzoek gedaan bij kinderen met een motorische en/of een auditieve handicap, zoals schiziskinderen [22], alsook dove of zwaar slechthorende kinderen [13]. Omdat bijna al dit onderzoek op uiterst adequate wijze begeleid is door Florian van Beinum, onze UHD die in juni vorig jaar met pensioen is gegaan met een Symposium ‘Spraak in

de maak’, richt ik mij hier nu liever op een ander aspect waarin *mijn eigen* specifieke kennis over signaalanalyse en dataverwerking goed aansluit bij de behoefte om op een inzichtelijke wijze de groei naar een volwaardige spraakproductie van deze kinderen te beschrijven en te illustreren. Vocalisaties van baby’s zijn niet eenvoudig als fonemen of woorden te benoemen, maar hoogstens in termen van respiratie, fonatie en articulatie. Dat akoestisch signaal vertoont lang niet altijd een duidelijk periodiek gedrag en evenmin een duidelijke formantstructuur (formanten representeren de resonanties van het mond-neus-keelkanaal). Die weinig geprononceerde formantstructuur is vooral een gevolg van de hoge toonhoogte van deze stemmetjes, die wel kan oplopen tot 400 Hz en meer. Een handmatige formantmeting van zo’n klanksegment wordt dan gemakkelijk beïnvloed door de vermeende klinkerkwaliteit die de onderzoeker in dat signaal denkt te horen. De door mij en anderen eind zestiger jaren bij TNO-IZF ontwikkelde spectrale analysemethode, die uit gaat van het hele spectrum zoals gemeten via een 20-tal bandfilters, gevolgd door een vorm van datareductie naar een beperkt aantal dimensies, is veel objectiever en beter reproduceerbaar, en blijkt niettemin grote gelijkensissen te vertonen met een formantrepresentatie en blijkt ook perceptieve gelijkensissen en verwarringen tussen

klanken goed te kunnen beschrijven. Dit is ook niet zo verwonderlijk als men bedenkt dat het oor veel meer een bandfilter analysator is dan een formantanalysator [14]. Ook deze bandfilter analysemethode heeft echter moeite met periodieke signalen met een hoge F0, maar daarvoor heeft Ton Wempe een truc bedacht die het toch goed mogelijk maakt het omhullende spectrum te meten [76].

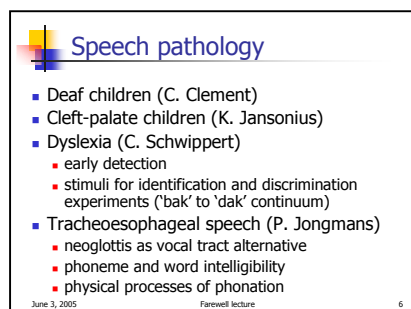


Aldus toegerust kunnen ongelabelde vocalisaties van jonge kinderen op een leeftijd vanaf enkele maanden tot 2 jaar en ouder, automatisch worden gemeten en worden weergegeven in een formantachtige representatie. Zo is goed te illustreren hoe een beperkte klinkerruimte bij 5 maanden, met voornamelijk /i/- en /u/-achtige klinkers, groeit naar een tamelijke volwaardige klinkerruimte bij 2 jaar. Omdat op 2-jarige leeftijd de woorden en klanken

meestal al goed te identificeren zijn, kunnen aan de verzameling van *ongelabelde* segmenten eenvoudig *gelabelde* klinkersegmenten worden toegevoegd als referentiekader. Zo is ook de ontwikkeling binnen en tussen *normaalhorende* kinderen te vergelijken met die van *zwaar slechthorende* kinderen [70]. Door de contacten die Jeannette van der Stelt op dit moment al heeft met onderzoeksgroepen in Amerika, Frankrijk en Rusland, en door het feit dat die beschikken over soortgelijke kinderspraakbestanden, zijn ook *verschillen* tussen talen te onderzoeken.

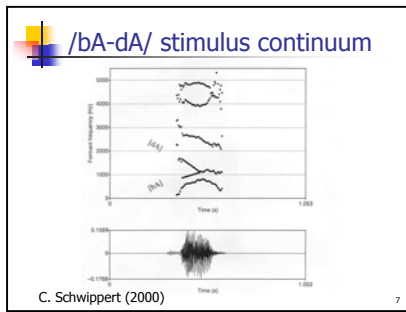
Tot nu toe hebben we ons beperkt tot maximaal 10 geïsoleerde meetpunten per uiting, zonder veel aandacht te besteden aan de temporele structuur *binnen* een uiting. Maar net zoals één enkele spectrale meting kan worden weergegeven als een punt in een ruimte, zo kunnen ook opeenvolgende metingen worden weergegeven als een spoor van punten in die ruimte, waardoor inzicht kan worden verworven in articulatorische dynamiek.

2.2 Spraakpathologie bij volwassenen (*dyslexie, laryngectomie*)



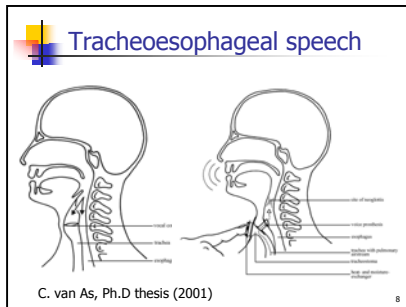
Naast de vroege ontwikkeling van kindertaal is er in onze groep ook altijd belangstelling geweest voor de productie en perceptie van pathologische en normale spraak van opgroeiende kinderen en volwassenen [4, 13, 22, 73, 74]. In deze sectie wil ik me beperken tot twee totaal verschillende vormen van pathologie, namelijk dyslexie en laryngectomie.

Vooraf door de participatie van Florian van Beinum in de NWO Stuurgroep Dyslexie, hebben wij ons bezonnen op adequate stimuli voor luisterexperimenten voor de vroegdetectie van dyslexie. Hierbij hebben we voortgebouwd op het promotie-onderzoek van Cecile Kuijpers [27]. Zij keek, bij kinderen van 4 tot 12 jaar, naar de *perceptie* van stemcontrast in gemanipuleerde stimuli (het kind hoort een gemanipuleerde stimulus die patto- of paddo-achtig klinkt en moet een van de twee poppen met de aangeleerde namen Patto of Paddo aanwijzen). Ook keek zij naar de *productie* van stemassimilatie in woorden zoals 'leesboek' of 'groot beest'.



Voor het Dyslexieprogramma werd door Caroline Schwippert [57] een continuüm gecreëerd van ‘bak’ naar ‘dak’, uitgaande van een natuurlijke uiting ‘bak’. Hiermee werden identificatie- en discriminatie-experimenten uitgevoerd bij volwassen dyslectici en bij een controlegroep [58]. Het bleek dat alleen onder kritieke condities van kleine stapgrootte en korte transitieduur de performance van de volwassen dyslectici minder

goed was. Niettemin lijken de signalen zeer geschikt voor vroegdetectie bij kinderen.



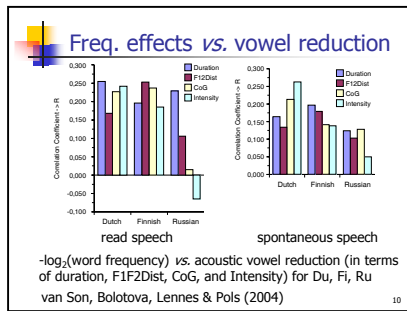
Als bij mensen met keelkanker de stembanden operatief moeten worden verwijderd, kunnen zij niet meer praten omdat enerzijds de luchtpijp niet meer uitmondt in de keel maar via een stoma in de hals, en omdat anderzijds de spraakbron in de vorm van stembanden niet meer beschikbaar is. Gelukkig kan bij de operatie al onmiddellijk een spraakventiel aangebracht worden tussen luchtpijp en slokdarm, waardoor de lucht uit de longen, bij afsluiting van

het stoma, alsnog via de slokdarm de mond-keelholte in kan stromen. Al dan niet na enige operatieve hulp kan met die lucht een vernauwing in de slokdarm, de zogenaamde neo-glottis, tot trilling worden gebracht. Deze eenvoudige neo-glottis, die niets heeft van de complexe regelstructuur van onze stembanden, fungeert dan als alternatieve stembron, waardoor vrijwel alle patiënten weer verstaanbaar kunnen communiceren. Na het baanbrekende onderzoek van Corina van As [3] over de stemkwaliteit van deze patiënten, is nu onze aio Petra Jongmans [23] een vervolgonderzoek gestart naar de foneem- en woordverstaanbaarheid en naar de natuurlijkheid van deze spraak, alsook naar de fysische processen van deze alternatieve stemgeving. Mijn collega Frans Hilgers [20] heeft hier tijdens zijn oratie op 23 september j.l. al heel veel meer over verteld. Ik wil er hier nu alleen nog op wijzen dat het eigenlijk een wonder is dat deze spraak nog zo goed klinkt als die klinkt, omdat immers grotendeels de flexibiliteit van foneren, toonhoogtevariaties en stemhebbend-stemlooswisselingen ontbreekt.

2.3 Specifiek fonetisch onderzoek

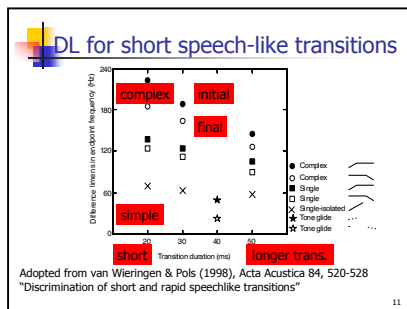
Sinds mijn aanstelling in 1982 ben ik in de gelegenheid geweest om in samenspraak met mijn promovendi en post-docs heel veel specifiek fonetisch onderzoek uit te voeren en daar op congressen en in de open literatuur over te rapporteren. Ik zal hieronder in een aantal rubrieken wat saillante punten uit dat onderzoek presenteren en wat aspecten voor vervolgonderzoek aangeven.

2.3.1 jnd, reductie, coarticulatie en prominentie



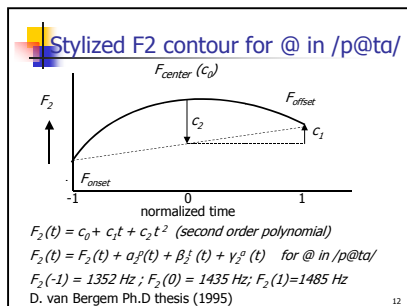
Een spreker spreekt zo slordig als de luisteraar hem toestaat te doen. Evenzo past de spreker zich waar nodig aan aan de omgevingsomstandigheden, b.v. in een galmende kerk of over de telefoon. Het blijkt daarnaast zo te zijn dat hoogfrequente en/of meervoorspelbare woorden significant slordiger worden uitgesproken dan laagfrequente en/of minder voorspelbare woorden [67]. Daarnaast is er een, grotendeels articulatorisch bepaalde, nabuur-

interactie oftewel coarticulatie die maakt dat b.v. de 'k' in 'koe' anders klinkt dan die in 'kie', en dat de 'ie' in 'dief' anders klinkt dan die in 'dier'. Er is veel discussie geweest in de literatuur of deze coarticulatie ontstaat door articulatorische beperkingen (d.w.z. een spreker kan zijn articulatoren nu eenmaal niet heel erg snel van de ene naar de andere plaats bewegen), of dat het meer een vorm is van sprekerluiheid en/of efficiency. Rob van Son en ik geloven sterk in het laatste, omdat de spreker tijdens normale en zelfs bij snelle conversaties eigenlijk nooit tegen de grenzen van de bewegelijkheid van zijn articulatoren aanloopt [50, 59, 62]. Dit betekent ook dat het Lindblom model [28], gebaseerd op articulatorische undershoot gecompenseerd door perceptieve overshoot, volgens ons geen steek houdt. Foneem- en woordherkenning kan dus ook geen eenvoudige template matching zijn, maar vereist dat de luisteraar rekening houdt met locale en globale context.



Bij analytische luistertaken met artificiële stimuli blijkt de luisteraar heel gevoelig te zijn voor kleine verschillen in het signaal, uitgedrukt in het zogenaamde juist waarneembare verschil, of jnd. Bij meer spraakachtige en dus complexere stimuli, neemt die jnd snel toe. Zo is het jnd voor een korte formanttransitie rond de 1000 Hz meer dan 200 Hz [78, 79], zodat de noodzaak voor een hele precieze waarneming van iedere formantfrequentie ook

minder groot lijkt. En dat is misschien maar goed ook, want hoe zouden we anders juist bij een goed gehoor ooit tot foneemcategorisatie kunnen komen bij zo'n variabele input?

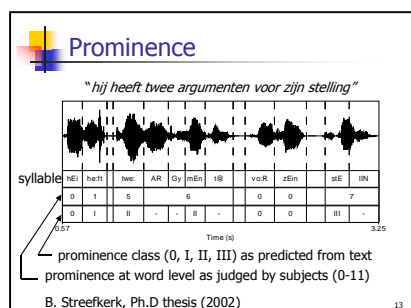


Van Bergem [6] heeft baanbrekend onderzoek gedaan naar *klinkerreductie*. Hij heeft laten zien dat die reductie niet simpelweg een meer centrale ligging van de klinker inhoudt, maar veeleer een grotere aanpassing aan de medeklinkeromgeving betekent. Zo wordt de gereduceerde /l/ in de eerste lettergreep van het woord 'miljoen' fysisch gezien niet zo zeer een /@/ maar veeleer een /u/ door de m-l omgeving. Ook heeft hij laten zien dat b.v. de

F2-transitie in de schwa van woordjes zoals 'patat' goed te modelleren, en dus te voorspellen, is door de effecten van /p/, /t/ en /a/ mee te nemen. Bijna 80% van de variantie in de F2-middenpositie kon hiermee worden verklaard. Ik vind dit een mooie

voorbeeld van *computationale fonetiek* en zou graag zien dat onze fonetische meetresultaten vaker zo zouden worden gepresenteerd.

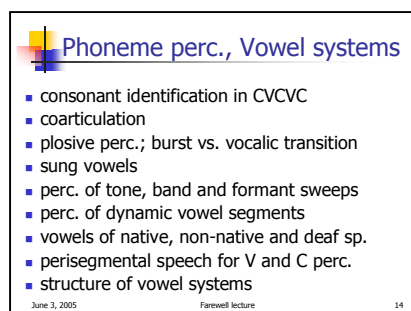
Naast *klinkerreductie* ligt het voor de hand dat er ook zoiets bestaat als *medeklinkerreductie*. In een post-docproject van Rob van Son hebben wij hier voor het eerst nadrukkelijk aandacht aan besteed [63, 64].



Het tegenovergestelde van reductie is *prominentie*, dit is het *benadrukken* van een lettergreep omdat de woorduitspraak dat vereist (je zegt niet *uitspraak*, maar *'uitspraak*), òf omdat b.v. de aanwezigheid van nieuwe informatie in de zin dat gewenst maakt. Barbartje Streefkerk [71, 72] heeft onderzocht in hoeverre die gewenste prominentie te voorspellen is uit een tekst op papier, te detecteren is in gesproken teksten, en te optimaliseren is in spraaksynthese.

Hier ligt nog veel ruimte voor nieuwe onderzoekvragen.

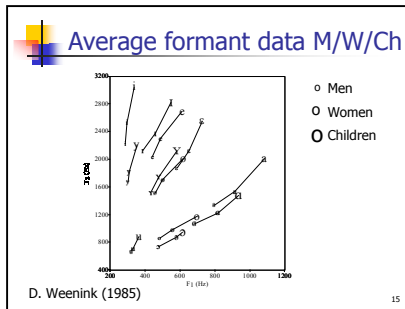
2.3.2 foneemperceptie en klinkersystemen



Tijdens de beginjaren van mijn aanstelling in Amsterdam heb ik, samen met Florien van Beinum en Bert Schouten en diverse promovendi, veel aandacht besteed aan klinker- en medeklinkerperceptie met natuurlijke en gemanipuleerde stimuli (o.a. [39, 25, 56, 8, 49, 50, 4, 65]) en aan de structuur van klinkersystemen [11]. Maar omdat ik in dit college niet alle onderwerpen uitgebreid kan behandelen verwijs ik voor meer informatie over deze interessante thematiek naar de referenties in de gedrukte versie van dit college.

2.3.3 sprekernormalisatie

Veel spraakherkenningssystemen ontbreekt het aan robuustheid, d.w.z. dat ze onder enigszins van de trainingsdata afwijkende omstandigheden al meteen veel slechter presteren. Één aspect van die afwijkende omstandigheden betreft variatie *tussen* sprekers. *Sprekeronafhankelijke* herkenningssystemen werken dan ook meestal slechts met een *kleine* vocabulaire, terwijl *sprekerafhankelijke* systemen die een *grote* vocabulaire aankunnen, langdurige training vereisen. Ook de mens is enigszins gevoelig voor dat fenomeen: Als men een serie klinkerstimuli *geblokt* aanbiedt, d.w.z. opeenvolgende klinkers van dezelfde spreker, dan is de herkenning daarvan significant beter dan bij *gemengde* presentatie, wanneer de spreker van stimulus tot stimulus steeds wisselt. Toch is de mens nog steeds fantastisch goed in het zich snel aanpassen aan steeds weer wisselende sprekersignalen, die fysisch gezien voor dezelfde fonemen en woorden toch enorm kunnen variëren over sprekers, met als meest extreme situatie de communicatie tussen een vader en een klein kind: De vader met een zware, lage stem en een groot hoofd, het kind met een hoog stemmetje en een klein koppie.



Speaker adaptive V identification

- TIMIT data, analyzed by David Weenink (2005)
- stressed vowel identification (13 categories)
- 438 males x 13 V x (1...25) rep. = 35,385 items
- LDA on one (static) frame: 59.3% correct
- from one static to 3 (dynamic) frames: 66.9%
- + *extrinsic* normal.: 59.3 -> 62.2; 66.9 -> 69.2
- one V-center/sp. (438 x 13 = 5,374 items):
59.3 -> 78.9; 66.9 -> 90.1
- + *extrinsic* normal.: 78.9 -> 87.9; 90.1 -> 94.5
- adaptive *intrinsic* normalization: 59.3 -> 60.3

June 3, 2005 Farewell lecture

Mogelijke modellen voor de manier waarop deze sprekernormalisatie zou kunnen werken, zijn systematisch onderzocht in het nu spoedig af te ronden promotieproject van David Weenink. Een groot probleem hierbij is een optimum te vinden tussen enerzijds snel leren, adapteren en generaliseren op grond van *weinig* nieuwe informatie en anderzijds opgedane kennis niet te snel weer vergeten.

2.3.4 discourse informatiestructuur en beurtwisselingsgedrag

Het fonetisch onderzoek heeft in de afgelopen 23 jaar van mijn aanstelling een zeer snelle ontwikkeling door gemaakt. Het meest duidelijk komt dit, naast de spraak-technologische ontwikkelingen en het gebruik van grote spraakdatabestanden (waarover zo dadelijk meer), naar mijn mening tot uiting in de veel grotere aandacht voor natuurlijke spraak en voor communicatieve processen zoals prosodische aspecten van discourse, het promotieproject van Monique van Donzel [16] en beurtwisselingsgedrag in dialogen, het nieuwe VIDI-project van Rob van Son [77].

Van Donzel onderzocht de informatiestructuur van door 4 mannen en 4 vrouwen navertelde teksten, zowel in termen van de akoestische middelen die de sprekers gebruiken (zoals toonhoogteaccenten en pauzes), alsook de interpretaties daarvan door de luisteraars (zoals waargenomen prominentie en discourse markers). Dit uitgebreide onderzoek bleef om praktische redenen beperkt tot pauzerings- en intonationale kenmerken. Andere belangrijke informatie zoals spectrale en segmentele informatie bleef dus buiten beschouwing, evenals de linguïstische structuur van de boodschap, in termen van b.v. woordvolgorde en grammaticale structuur. Verder bleef dit project beperkt tot monologen waarbij er geen gesprekspartner was die kon interrumperen. Op al deze punten zijn dus uitbreidingen mogelijk.

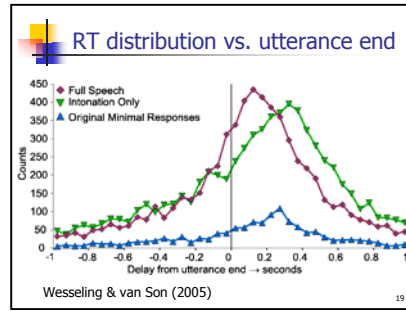
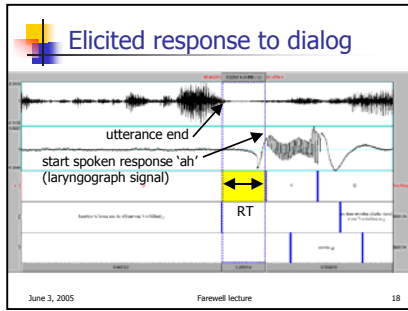
Dialogs and turntaking

- post doc Rob van Son; aio Wieneke Wesseling
- selected CGN dialogs
- say /a/ as if interrupting
- pilot with some 10 subjects
- use laryngograph signal to detect earliest point in time of planning an interrupt
- compare this with actual interrupts

June 3, 2005 Farewell lecture

Deze zullen tenminste gedeeltelijk aan de orde komen in het VIDI-project van van Son over beurtwisselingsgedrag in dialogen. De *fonetische* aanpak die ook in dit project wordt gekozen, blijkt het duidelijkst uit de gekozen experimentele procedure: teneinde een beter inzicht te krijgen in de akoestische en linguïstische aspecten die een mogelijke beurtwisseling aankondigen, wordt proefpersonen gevraagd om, al luisterend naar een

dialog, zo snel mogelijk aan te geven wanneer zij zouden willen interrumperen. Zij doen dit door 'aa' te zeggen, en een laryngograaf registreert daarbij hun stemband-activiteiten.



Eerste resultaten laten zien dat de vroegste stembandactie verrassend vroeg optreedt, hetgeen suggereert dat luisteraars kennelijk al vroeg aan het plannen en vooruitdenken zijn. Welke akoestische, prosodische en linguïstische aspecten van de dialoog hier het meest bepalend zijn, is uiteraard punt van verder onderzoek. Een van de manieren om dat te onderzoeken, is de gesproken tekst zodanig te manipuleren dat die onverstaaanbaar wordt maar daarbij wel de intonatie en de pauzes te handhaven. Ook zal gekeken worden wat additionele video-opnamen van de spreker aan informatie toevoegen. De resultaten uit dit onderzoek moeten er mede toe gaan leiden dat mens-computer dialogen efficiënter en natuurlijker gaan verlopen.

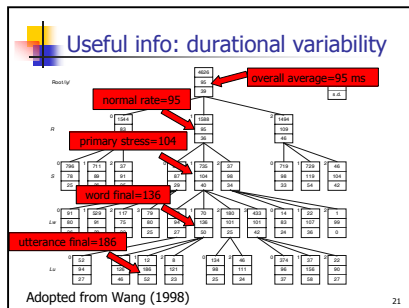
2.3.5 spraaktechnologie (herkenning, synthese evaluatie)

Aspects of Speech technology

- Faculty of Arts not very adequate to develop and operationalize complex systems
- still, specific phonetic contributions possible
- like modelling phoneme duration (Pols, Wang, ten Bosch, 1996; Wang, 1997)
- like modelling prominence (Streefkerk, 2002)
- desire to implement prosodic knowledge
- parametric representations
- synthesis evaluation

Binnen de huidige structuur van onze faculteit Geesteswetenschappen is het nauwelijks reëel te veronderstellen dat er op een fonetisch instituut voldoende middelen en deskundigheid aanwezig zullen zijn om een adequate technologische bijdrage te kunnen leveren aan de ontwikkeling en operationalisering van complexe taal- en spraak-technologische systemen. Dat laat onverlet dat ik meen dat de Fonetiek onontbeerlijk is om wezenlijke

bijdragen te kunnen leveren aan *specifieke* aspecten van zo'n systeem. De brute force probabilistische benaderingsmethode, die op veel en variabele trainingsdata is gebaseerd, is nog steeds erg populair en redelijk succesvol [1]. Toch blijven er aspecten van foneem- en woordmodellering en van specifieke kennis van het spraak-signaal, die een fonetische inbreng potentiaal zeer waardevol kunnen doen zijn.

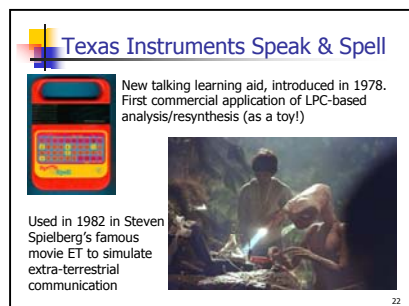


Zo kan specifieke informatie over de *duur* van de individuele fonemen een bijdrage leveren aan de herkenning, maar draagt die nu alleen maar bij in de vorm van een normale distributie van de foneemduur waarin alle mogelijke condities zijn samengevoegd [52]. Dat klinkers in een beklemtoonde lettergreep, of in woorden in zinsfinale positie, of bij een langzaam spreektempo, systematisch langer van duur zijn, blijft dus ongebruikt. Xue Wang [75]

heeft in zijn proefschrift laten zien dat hier wel degelijk enige winst is te behalen door in een post-processing fase, uit een kleine set van meest waarschijnlijke zinnen, *die* zin te kiezen waarbij de feitelijke duurinformatie het best past.

Zo is het ook nog steeds een ideaal van veel fonetici om ook andere *prosodische kennis* in te zetten in het herkenningproces. Een praktisch probleem hierbij is, dat de daarvoor benodigde specifieke informatie eenduidig en liefst foutloos meetbaar moet zijn in het spraaksignaal, en dat is nu nog niet goed mogelijk [54].

De parametrische representatie van het spraaksignaal in spraaktechnologische systemen gebeurt al meer dan 30 jaar vrijwel uitsluitend via linear predictive coding of varianten daarvan, zoals mel-frequency-based cepstral coefficients. Ik heb zelf destijds in mijn proefschrift [37] al aangegeven dat een principale componenten analyse op bandfilterdata vrijwel synoniem is aan MFCC, ook J. Cohen en R. Bakis van IBM hebben dit al vroeg gesignaleerd.



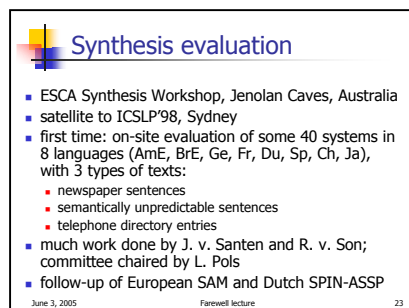
Texas Instruments Speak & Spell

New talking learning aid, introduced in 1978. First commercial application of LPC-based analysis/resynthesis (as a toy!)

Used in 1982 in Steven Spielberg's famous movie ET to simulate extra-terrestrial communication

Het is wellicht interessant hier te memoreren dat de eerste commerciële applicatie van Amerikaanse LPC analyse/resynthese door Texas Instruments niet in de computersfeer lag, maar als een modern vormgegeven speelattribuut Speak & Spell, dat in 1978 op de markt kwam en door mijn kinderen meteen werd omarmd, maar dat ook door Steven Spielberg werd ingezet in zijn beroemde film ET uit 1982 om kinderen te laten communiceren met buitenaardse wezens.

Mede in het kader van het Nederlandse SPIN-ASSP programma en de Europese programma's SAM en SPIN, heb ik mij samen met anderen zo'n 15 jaar bezig gehouden met de evaluatie van spraaksynthesesystemen (o.a [47, 7, 41, 48, 42, 43]).



Synthesis evaluation

- ESCA Synthesis Workshop, Jenolan Caves, Australia
- satellite to ICSLP'98, Sydney
- first time: on-site evaluation of some 40 systems in 8 languages (AmE, BrE, Ge, Fr, Du, Sp, Ch, Ja), with 3 types of texts:
 - newspaper sentences
 - semantically unpredictable sentences
 - telephone directory entries
- much work done by J. v. Santen and R. v. Son; committee chaired by L. Pols
- follow-up of European SAM and Dutch SPIN-ASSP

June 3, 2005 Farewell lecture 23

Als een afsluiting van al dat werk zie ik de Synthese Workshop in Jenolan Caves, Australië in 1998, waar voor de eerste keer een uitgebreide on-site evaluatie plaats vond [55] van een 40-tal systemen in 8 talen (Amerikaans en Brits Engels, Duits, Frans, Nederlands, Spaans, Chinees en Japans) met 3 representatieve typen tekst, namelijk krantenteksten, semantisch onvoorspelbare zinnen zoals 'de vloer sloot de vis die liep' [5], en informatie uit het

telefoonboek. De nu zeer populaire concatenatieve synthese zorgt voor een goede initiële spraakwaliteit, maar verlegt het probleem naar hoe, op een efficiënte wijze, meerdere stemmen en spreekstijlen zijn te verwerven, hoe emotie is te implementeren [29] en hoe de natuurlijkheid is te vergroten.

2.3.6 talen en dialecten



Languages and dialects

- phonetics research: mainly Dutch or English, but.....
- Russian intonation (post-doc Cecilia Odé)
- project proposal Tundra Yukagir (Odé)
- Intas 915 project (Russian, Finnish, Dutch)
- Franconian tone languages (Boersma c.s.)
- Frisian TTS (J. Dijkstra, MA-thesis, Fryske Ak. prize)
- Latin, Pintupi word stress (D. Apoussidou)
- length acquisition in Swedish (A. Öllander)
- word boundary assimilation in Czech (A. Rezhiegl)
- Croatian folk songs (G. Kovacic)
- South-African vowels (L. Botha)
- Pharyngeal coartic. in Egyptian Arabic (A. Elgendy)

24

Fonetische onderzoeksmethodes zijn meestal taalonafhankelijk en veel fonetici doen hun onderzoek gemakshalve aan hun moedertaal of aan het Engels. Toch lopen er bij ons ook diverse taalspecifieke projecten. Mede door de aanwezigheid van Cecilia Odé in onze groep sinds 2002, wordt er nu ook onderzoek gedaan aan Russische intonatie en de communicatieve functie

daarvan [31]. Door onze participatie in het Europese INTAS programma hebben we ook vergelijkend onderzoek uitgevoerd naar een aantal fonetische eigenschappen van het Fins, Russisch en Nederlands [61]. Binnen het NWO-programma 'Bedreigde Talen' heeft Odé een nieuw project aangevraagd over het Tundra Yukagir. Recentelijk is een ander NWO-project gehonoreerd, een zogenaamd klein programma, over Frankische toontalen, en meer specifiek over het Limburgs, dat was ingediend door Ben Hermans, Marc van Oostendorp en Paul Boersma. Talige variatie was ook aan de orde in een recent afstudeerproject van Jelske Dijkstra [15] over Friese tekst-naar-spraaksynthese dat de 2-jaarlijkse scriptieprijs van de Fryske Akademy heeft gekregen, en zo zouden er nog diverse andere projecten kunnen worden genoemd (o.a [12, 17, 26, 32, 53]).

2.3.7 fonologische aspecten van spraakverwerving

Phonological aspects

- of language acquisition
- NWO Vernieuwingsimpuls 2002 P. Boersma
- 'Adequacy and acquisition of functional constraint grammars'
- using Optimality Theory
- Gradual Learning Algorithm
 - second language acquisition
 - stress in Latin, Pintupi, Dutch (D. Apoussidou)

June 3, 2005 Farewell lecture 25

Zonder ook maar enigszins te willen pretenderen dat mijn huidige leeropdracht ook de Fonologie omvat, is het onmiskenbaar dat met het promotieonderzoek van Paul Boersma [9] over functionele fonologie, en meer nog met de honorering van zijn projectaanvraag 'Adequacy and acquisition of functional constraint grammars' in het kader van de NWO-Vernieuwingsimpuls in 2002, hoogstaand fonologisch onderzoek met gebruikmaking van

Optimality Theory zijn intrede heeft gedaan in onze leerstoelgroep. Het is verbluffend en inspirerend om te zien hoe adequaat het door hem ontwikkelde Gradual Learning Algoritme [10] in staat is de verwerving van fonemen door tweede-taalleerders [18], alsook de vroege verwerving van klemtoonstructuren in talen zoals het Latijn [2] en het Pintupi, en hopelijk binnenkort ook voor het Nederlands, adequaat kan worden gemodelleerd.

2.4 Spraakdatabestanden

Speech databases

- TNO Soesterberg: 10 sp. x 12 V (JASA 1967)
- then 50 males and 25 females, hVt (1973)
- then 270 CVC words (Ph.D thesis, Pols, 1977)
- normal and fast speech Ph. Bloemendal
- IFA corpus (5.5 hrs of speech)
- CGN (1000 hrs. of speech)
- VIDII project R. v. Son (multimodal dialogs)
- historical collection (> 100 hrs.)

June 3, 2005 Farewell lecture 26

Ik ben mijn wetenschappelijke carrière ooit begonnen bij het TNO Instituut voor Zintuigfysiologie in Soesterberg, op de afdeling Gehoor en Spraak onder de bezielende leiding van prof. dr. ir. Reinier Plomp. Later kreeg dat instituut de naam Human Factors, oftewel Technische Menskunde, en na de recente reorganisatie binnen TNO wordt het allemaal weer anders. Mijn eerste serieuze publicatie, in The Journal of the Acoustical Society

of America in 1967 [36], ging over een spectrale analyse van de 12 Nederlandse enkelvoudige klinkers gesproken door 10 mannen in woordjes van het type 'hot', 'haat' en 'hoot'. Dat setje van 120 woorden was mijn eerste spraakbestand. Later is dat bestand uitgebreid naar de klinkerset van 50 mannen en 25 vrouwen [51, 30]. Tot op de dag van vandaag wordt nog regelmatig verwezen naar die bandfilter- en formantanalyses van de Nederlandse klinkers en naar de daaraan gerelateerde perceptieve data [46]. In mijn proefschrift uit 1977 [37] was de woordset uitgebreid tot 270 CVC woorden van het type 'pan', 'heer' en 'nu', en werd manipuleerbare

synthese geïntroduceerd en later ook automatische woordherkenning [38]. Dit type analyse en datarepresentatie is nu weer heel actueel bij onze analyses van kindervocalisaties, van spraak van gelaryngectomeerden, van het Poldernederlands, alsook bij sprekernormalisatie.

In mijn jarenlange samenwerking met, de nu post-doc, Rob van Son zijn nog diverse andere geannoteerde spraakdatabestanden door hem aangelegd, zoals een normaal en snel voorgelezen tekst door de bekende radio- en filmcommentator Philip Bloemendal om de mogelijke processen van akoestische undershoot en perceptieve overshoot te bestuderen [62]. Of het zgn. IFA-corpus, waarin een achttal sprekers op allerlei manieren teksten inspreken: van los uitgesproken lettergrepen, woorden en zinnen, tot voorgelezen teksten en spontane vertellingen [66]. In zijn nieuwste VID-project zullen, naast aandacht voor een selectie van bestaande dialogen uit het CGN-corpus, ook nieuwe dialogen audiovisueel worden geregistreerd. In dit onderzoek samen met de aio Wieneke Wesseling wordt speciale aandacht besteed aan beurtwisselingsgedrag [77].

Zo is de overgang van sterk gestructureerd (maar vrij onnatuurlijk) spraakmateriaal naar ongestructureerde natuurlijke conversatie, onmiskenbaar. De uitdaging om al dit materiaal toegankelijk te maken via transcripties en om er zinvolle onderzoeksvragen mee te lijf te gaan, is groot.

Het bestaan van het IFA-corpus heeft het ons mogelijk gemaakt, om zonder grote extra inspanningen, toch zinvol te kunnen participeren in een internationaal samenwerkingsverband met Finnen en Russen, INTAS 915, dat bedoeld was om diverse fonetische eigenschappen te vergelijken over deze drie totaal verschillende talen [61]. Het IFA-corpus met ca. 5 uur spraakmateriaal is inmiddels ook opgenomen in het Corpus Gesproken Nederlands, dat uit een kleine 1000 uur spraak van Nederlandse en Vlaamse sprekers bestaat [33]. Als lid van de CGN Stuurgroep heb ik graag mijn steentje bijgedragen aan de totstandkoming van dit megaproject. Alle beschikbare annotaties en alle audio-files zijn nu samengebracht op 33 DVD's, en inmiddels zijn zij ook op diverse plaatsen, waaronder op ons eigen IFA, direct toegankelijk gemaakt door opslag op een grote schijf op een netwerkserver. Hiermee is dat corpus dan b.v. ook goed inzetbaar voor het project Poldernederlands waaraan de aio Irene Jacobi werkt, met Jan Stroop als haar eerste begeleider [21].

 'op een gegeven moment'

- >1200 occurrences in CGN (Dutch and Flemish)
- of which 790 in Dutch spontaneous speech

Op @n x@x'ev@(n) mom'Ent		
Op @ x@x'ev@ mOm'Ent(t)		
Op @ x'ev@ m'Ent		
Op @ x'em'Ent		
x@m'Ent		
Total	790	100

June 3, 2005 Farewell lecture 27

Ter illustratie voor deze middag, heb ik binnen het CGN-corpus gezocht naar eclatante voorbeelden van klank- en lettergreepreductie. Tijdens onderwijs-sessies heb ik tegenover onze studenten vaak beweerd dat de 7-lettergrepige uitdrukking 'op een gegeven moment' in spreektaal veelvuldig wordt gereduceerd tot zoiets als [x@m'Ent] met slechts 2 lettergrepen. Via het exploratieprogramma COREX is het CGN-corpus hierop eenvoudig te doorzoeken terwijl Rob van Son voor een goede toegankelijkheid van zijn IFA-corpus heeft gezorgd via SQL queries. Deze uitdrukking komt meer dan 1200 keer voor in het CGN-corpus, waarvan 790 keer uitgesproken door Nederlandse sprekers in vrije conversatie. In de geprojecteerde tabel is de distributie van voorkomen van de verschillende varianten weergegeven. Het zal duidelijk zijn dat de meest extreme vorm van reductie vrij zeldzaam is, maar dat geldt evenzo voor de volledige vorm.

Naast deze vrij recent verworven spraakcorpora is er nog een interessante ontwikkeling te melden, namelijk die van het beschikbaar komen van een historische collectie van allerlei spraakopnamen die op de zolder van ons instituut lagen opgeslagen. Deze opnamen, die meer dan 100 uur spraak omvatten, gaan terug tot de dertiger jaren van de vorige eeuw en zijn geregistreerd op zulke diverse dragers als wasrollen, glasplaten, aluminium-, vinyl- en kunststofplaten, alsook papierbanden en tapes. Met een subsidie, verworven door Marian Schilder van ons Universiteitsmuseum via de zogenaamde Nuisgelden, is een groot deel van deze opnamen inmiddels op het Meertens Instituut gedigitaliseerd. Momenteel zijn al die audio files op niet erg efficiënte wijze opgeslagen op vele duizenden CD's. Binnenkort zullen al deze files over worden gezet op één grote schijf op onze server. Zoals vaak met historisch materiaal, zijn helaas de metadata zeer incompleet, hetgeen systematisch zoeken bemoeilijkt, en zijn er ook nog geen fondsen gevonden om interessant geachte *onderdelen*, laat staan *alles*, te transcriberen. Niettemin zijn er al enkele interessante krenten uit deze brij tevoorschijn gekomen. Zo heeft mijn collega Frans Hilgers tijdens zijn oratie een unieke opname op wasrol kunnen laten horen van de spraak en zang van een gelarygectomeerde uit 1925. Deze in de literatuur goed gedocumenteerde case betreft vrijwel zeker de oudste opname in zijn soort. Tevens hebben Ton Wempe en Paul Boersma reeds enkele oude opnamen van sprekers uit Venlo kunnen traceren, die wellicht een interessante toevoeging zullen zijn aan het net gestarte onderzoek aan het hedendaagse Limburgs.

2.5 Onderwijs in de fonetiek

Teaching and Phonetics

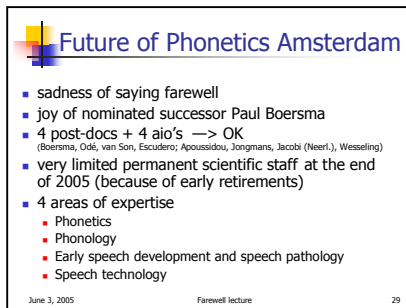
- Phonetics courses in BA Linguistics
- MA track 'Speech comm. and technology'
- also courses at Artificial Intelligence
- this may attract more MA-students than in our own faculty
- students in Phonetics have good perspectives
- potential for Cognition, Medicine, Psychology
- staff with proper expertise required

June 3, 2005 Farewell lecture 28

Wellicht meer nog dan in het *onderzoek*, komen specifieke aspecten van kennisverwerving en implementatie ter sprake in het *onderwijs*. Fonetische kennis is een onontbeerlijk aspect van taalkundige kennis, en terecht zijn er dan ook een drietal fonetische modules: 'Spraak', 'Fonetische transcriptie', en 'Spraakverwerking per computer' opgenomen in het huidige Amsterdamse bachelorprogramma Taalwetenschap. Alhoewel onze

leerstoelgroep een onderdeel is van de opleiding Taalwetenschap, is dat toch zeker niet het enige kader waarbinnen wij opereren. Door het interdisciplinaire karakter van de Fonetiek, zal het u niet verbazen dat er ook onderwijssamenwerking bestaat met Informatica en Geneeskunde en dat samenwerking goed mogelijk zou zijn met Psychologie en Cognitie. Bij de interfacultaire opleiding Kunstmatige Intelligentie verzorgen wij de bachelormodule 'Spraakherkenning en spraaksynthese', en in de masterspecialisatie 'Taal en Spraak', de modules 'Spoken Language Generation' en 'Speech Technology project'. De verwachting is dat hier wellicht *meer* studenten op af zullen komen dan op de specialisatie 'Speech Technology' binnen de MA Linguistics in onze eigen faculteit. Tot nu toe is het altijd zo geweest dat vrijwel alle studenten met een goede fonetische opleiding moeiteloos werk konden vinden in het onderzoek of in de maatschappij.

3. Toekomst



Future of Phonetics Amsterdam

- sadness of saying farewell
- joy of nominated successor Paul Boersma
- 4 post-docs + 4 aio's → OK
(Boersma, Ode, van Son, Escudero; Apoussidou, Jongmans, Jacobi (Neerl.), Wesseling)
- very limited permanent scientific staff at the end of 2005 (because of early retirements)
- 4 areas of expertise
 - Phonetics
 - Phonology
 - Early speech development and speech pathology
 - Speech technology

June 3, 2005 Farewell lecture 29

Nu een woord over de toekomst. Helaas is het de laatste tijd, hier in Amsterdam en elders in Nederland, gebruikelijk dat bij een afscheidscollege de ondertoon vaak wat somber is omdat er niet langer de uitdrukkelijk wens tot voortzetting van het betreffende specialisme bestaat bij het faculteitsbestuur, dat daarbij meestal ook nog geplaagd wordt door een chronisch geldgebrek. In mijn geval is er gelukkig veel minder reden tot somberheid omdat

ons faculteitsbestuur gesteund door het College van Bestuur heeft besloten tot de tijdige benoeming van Paul Boersma als mijn opvolger. Omdat het hier ook nog eens een eigen promovendus betreft die moeiteloos de internationale kwaliteitstoets kan doorstaan, kan ik over deze ontwikkelingen alleen maar heel gelukkig en dankbaar zijn. Omdat Boersma naast een gerenommeerd fonoloog, ook een bekwaam foneticus is en ook nog eens verantwoordelijk is, samen met Weenink, voor het onvolprezen en door meer dan 10.000 gebruikers over de hele wereld hogelijk gewaardeerde spraak-signaalverwerkingsprogramma *praat*, is er alle reden aan te nemen dat de leerstoel Fonetische Wetenschappen binnen de Opleiding Taalwetenschap onder zijn leiding een mooie en diverse toekomst tegemoet gaat.

Ik ben daarbij verheugd de leerstoelgroep te kunnen achterlaten met vier extern gefinancierde post-docs en vier aio's. Zorgelijker is het feit dat door vervroegde pensionering de vaste staf aan het eind van 2005 zeer sterk zal zijn gedecimeerd hetgeen zeker consequenties zal hebben voor het onderwijsaanbod. Voor die toekomstige ontwikkelingen ben ik echter niet meer verantwoordelijk, maar voor een adequate samenvatting van mijn college van vandaag natuurlijk wel.



Trends in past 25 years

- from simple/stationary to complex/dynamic
- from laboratory speech to natural speech
- from artificial tasks to communicative relevance
- data growth in size, diversity and complexity
- growth in capabilities of computer/software/storage
- speech & language & communicative & multimodal & multilingual & technology
- computational phonetics
- closer to industrial, clinical and educational applic.
- still basic problems of invariance and variability

June 3, 2005 Farewell lecture 30

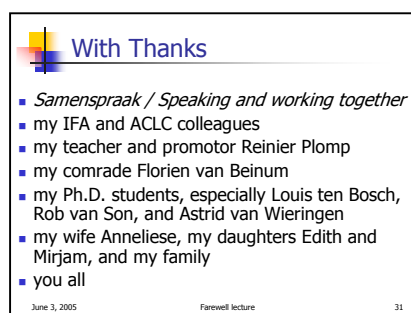
Ik heb gewezen op duidelijke trends in het fonetisch onderzoek van de afgelopen 25 jaar: de aandacht heeft zich meer en meer verlegd van eenvoudige stationaire signalen naar complexe dynamische signalen, van laboratoriumspraak naar natuurlijke spraak [35], van tamelijk gekunstelde taken voor luisteraars naar meer en meer communicatief relevante taken. Het spraakmateriaal waarop heden ten dage analyses kunnen worden uitgevoerd is,

door veel snellere computers en veel grotere opslagcapaciteit, exponentieel gestegen in omvang, diversiteit en complexiteit. De foneticus kan zich niet meer beperken tot het spraaksignaal alleen, maar moet zich ook bewust zijn van de talige, communicatieve en multimodale aspecten. De voor iedereen toegankelijke software is veel krachtiger en gebruikersvriendelijker geworden. En mede daardoor hoeft het nu niet meer bij incidentele observaties te blijven maar kunnen verschijnselen veel systematischer worden beschreven en kunnen we computationele fonetiek bedrijven [44]. Het huidige wetenschappelijk onderzoek staat veel dichterbij industriële, klinische en educatieve applicaties. Naast al deze nieuwe ontwikkelingen kan ik er ook niet omheen te constateren dat veel oude problemen nog steeds bestaan. Zo ben ik een jaar na mijn benoeming enorm geïnspireerd door een symposium op MIT met de titel 'Invariance and variability in speech processes' [34, 40]. Dit probleem van

variatie in de productie naast noodzakelijke vormen van invariantie voor perceptie en herkenning, houdt ons nog steeds bezig.

De brede bachelor en de korte gespecialiseerde master opleiding laten niet toe dat veel aandacht kan worden besteed aan de nieuwste ontwikkelingen in ons vakgebied. De huidige student met een BA- of MA-diploma zal dus nog heel veel additionele kennis en vaardigheid moeten verwerven alvorens van nut te kunnen zijn in een specifiek onderzoeksproject of in de maatschappij. In het nu voorgestelde traject van 3 jaar BA, plus 2 jaar onderzoeks-MA, plus 3 jaar voor de promotie, is het dan ook de vraag of de hoge eisen die *nu* nog steeds aan het proefschrift worden gesteld, in de toekomst gehandhaafd kunnen blijven, temeer omdat tijdige afronding een steeds dwingendere eis wordt. In de afgelopen 3 jaar waren er in heel Nederland nog steeds ca. 7 aan de fonetiek gerelateerde promoties per jaar [45], ik vrees dat dat aantal in de toekomst drastisch zal verminderen omdat het reservoir aan goed opgeleide fonetici uitgeput raakt.

4. Dankwoord



Ik kom toe aan mijn dankwoord. Het eerste woord in de titel van mijn voordracht was *samenspraak*, dat is een mooi woord in de Nederlandse taal, speciaal voor een foneticus. Het geeft ook goed weer hoe ik heb proberen te functioneren met mijn groep vaste en tijdelijke mensen, nu en vroeger. Ik kan niet goed *alleen* functioneren en werk graag *samen* met anderen aan een gemeenschappelijk doel. Ik hoop dat ik als begeleider mensen heb weten te stimuleren

en motiveren om het einddoel te bereiken. Ikzelf heb zeker heel veel geleerd van het enthousiasme, de daadkracht en de vindingrijkheid van de vaste staf, aio's, post-docs en gasten waarmee ik sinds 1982 heb mogen samenwerken.

Ook was het een genoegen te mogen functioneren in diverse samenwerkingsverbanden over de jaren heen, zoals de vroegere werkgemeenschap Fonetiek en de NWO Stichting Taalwetenschap, het spraakexpertisecentrum SPEX, de Stichting Spraaktechnologie, het onderzoekinstituut IFOTT (nu ACLC), de Stuurgroepen van zulke nationale programma's als SPIN-ASSP, CGN en IMIX, en de actieve participatie in Europese projecten zoals SAM, SPIN, Eagles en EuroCocosda. Extra interessant en tevens goed voor het netwerken, was mijn participatie in het NWO Gebiedsbestuur Geesteswetenschappen en in de Senaat van deze universiteit.

Het wetenschappelijk handwerk is mij bijgebracht door mijn leermeester Reinier Plomp, hoofd van de afdeling Audiologie toen ik in 1965 bij TNO-IZF kwam. Hij was tevens mijn latere promotor aan de VU. Hij heeft mij geleerd te redeneren, te experimenteren en te publiceren. Hij heeft mij al in 1970 meegesleurd naar Amerika om daar spraaklaboratoria en congressen te bezoeken, en hij heeft het leggen van internationale contacten altijd erg gestimuleerd. Dit heeft er o.a. toe geleid dat ik mij intensief heb bezig gehouden met de oprichting in 1987 en de latere groei van de European Speech Communication Association ESCA, later ISCA, en met de door die club georganiseerde workshops, waaronder die in Noordwijkerhout in 1989, en grote internationale congressen genaamd Eurospeech en nu sinds kort Interspeech.

Een tweede persoon waaraan ik veel dank verschuldigd ben is mijn maatje tijdens al de jaren op het IFA en tijdens vele gemeenschappelijk bezochte

wetenschappelijke bijeenkomsten, namelijk Florian van Beinum. Samen konden wij de diversiteit aan onderwerpen in onderwijs en onderzoek meestal wel aan.

Tenslotte ben ik verknocht aan al mijn promovendi, omdat ieder voltooid proefschrift een vaak lange periode van intensieve samenwerking afsluit met een tastbaar resultaat. Een drietal van hen wil ik speciaal noemen, omdat bij hen de samenwerking zich lang heeft voortgezet en de gemeenschappelijke inzet in experimenten, dataverwerking en publicaties extra groot was, namelijk Louis ten Bosch, Rob van Son en Astrid van Wieringen.

Een vakgroep, of zoals dat nu heet leerstoelgroep, runnen binnen een complexe universitaire structuur van opleidingen, afdelingen, en onderwijs- en onderzoeksinstituten, alsook binnen nationale en internationale verbanden, is boeiend en uitdagend, maar ook tijdrovend en vermoeiend. Het is dan belangrijk dat het thuisfront je zo veel mogelijk steunt, maar ook aandacht afdwingt voor heel andere zaken. Ik heb altijd erg veel steun gehad van mijn vrouw Anneliese en wij zijn de gelukkige ouders van twee lieve dochters Edith en Mirjam die nu al heel erg volwassen zijn, maar ons met hun gezinnetjes met inmiddels vier kleinzonen veel vreugde blijven bereiden. Uiteraard zijn zij hier vandaag aanwezig, maar ik kan er nog steeds nauwelijks over uit dat ons gezin ook compleet was toen ik in Aalborg tijdens Eurospeech 2001 de ISCA medal for scientific achievement mocht ontvangen. Kort en goed, ik ben een gelukkig mens en kijk graag vooruit naar al het moois dat nog in het verschiet ligt.

5. Referenties

- [1] Alphen, P. van (1992), *HMM-based continuous-speech recognition. Systematic evaluation of various system components*, Ph.D. thesis Univ. of Amsterdam: 216 pp.
- [2] Apoussidou, D. & Boersma, P. (2003), "The learnability of Latin stress", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 25:101-148.
- [3] As, C. van (2002), *Tracheoesophageal speech. A multidimensional assessment of voice quality*, Ph.D. thesis Univ. of Amsterdam: 211 pp.
- [4] Bakkum, M.J., Plomp, R. & Pols, L.C.W. (1993), "Objective analysis versus subjective assessment of vowels pronounced by native, non-native, and deaf male speakers of Dutch", *J. Acoust. Soc. Amer.*, 94(4): 1989-2004.
- [5] Benoît, C., Grice, M. & Hazan, V. (1996), "The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences", *Speech Communication* 18(4): 381-392.
- [6] Bergem, D. R. van (1995), *Acoustic and lexical vowel reduction*, Ph.D. thesis Univ. of Amsterdam: 195 pp.
- [7] Bezooijen, R. & Pols, L.C.W. (1991), "Evaluating text-to-speech systems: Some methodological aspects", *Speech Communication*, 9(4): 263-270.
- [8] Bloothoof, G. (1985), *Spectrum and timbre of sung vowels*, Ph.D thesis Free Univ. of Amsterdam: 169 pp.
- [9] Boersma, P. (1998), *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*, Ph.D. thesis Univ. of Amsterdam: 493 pp.
- [10] Boersma, P. & Hayes, B. (2001), "Empirical tests of the Gradual Learning Algorithm", *Linguistic Inquiry* 32: 45-86.
- [11] Bosch, L. F. M. ten (1991), *On the structure of vowel systems. Aspects of an extended vowel model using effort and contrast*, Ph.D. thesis Univ. of Amsterdam: 190 pp.
- [12] Botha, L. (1996), "Towards modelling acoustic differences between L1 and L2 speech: The short vowels of Afrikaans and South-African English", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 20:65-80.
- [13] Clement, C.J. (2004), *Development of vocalizations in deaf and normally hearing infants*, Ph.D. thesis Univ. of Amsterdam, LOT 100: 264 pp.
- [14] Dijk, J. van (2001), *Mechanical aspects of hearing*, Ph.D. thesis Univ. of Amsterdam: 211 pp.

- [15] Dijkstra, J., Pols, L.C.W. & Son, R.J.J.H. van (2004), "Frisian TTS, an example of bootstrapping TTS for minority languages", *Proc. 5th ISCA Synthesis Workshop*, Pittsburgh: 97-102.
- [16] Donzel, M. E. van (1999), *Prosodic aspects of information structure in discourse*, Ph.D. thesis Univ. of Amsterdam: 194 pp.
- [17] Elgendy, A. M. (2001), *Aspects of pharyngeal coarticulation*, Ph.D. thesis Univ. of Amsterdam: 312 pp.
- [18] Escudero, P. & Boersma, P. (2004), "Bridging the gap between L2 speech perception research and phonological theory", *Studies in Second Language Acquisition*, 26(4): 551-585.
- [19] Gelderen, A. J. S. van (1992), *De evaluatie van spreekvaardigheid in communicatieve situaties. Globale beoordeling en gedetailleerde analyse van spreekprestaties van 11- en 12-jarigen*, Ph.D. thesis Univ. of Amsterdam: 265 pp.
- [20] Hilgers, F. (2004), "Van verbazing naar vanzelfsprekendheid. From astonishing to predictable speech", *Oratie Univ. van Amsterdam, Vossiuspers UvA*, Amsterdam: 64 pp.
- [21] Jacobi, I., Pols, L.C.W. & Stroop, J. (presented for publication), "Polder Dutch: Aspects of the /ɛl/-lowering in standard Dutch", *Proc. Interspeech 2005*, Lisbon.
- [22] Jansoniuss-Schultheiss, K. (1999), *Twee jaar spraak en taal bij schisis*, Ph.D. thesis Univ. of Amsterdam: 277 pp.
- [23] Jongmans, P., As. C. van, Pols, L. & Hilgers, F. (2003), "An introduction to the assessment of intelligibility of tracheoesophageal speech", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 25: 185-196.
- [24] Kemps, R.J.J.K. (2004), *Morphology in auditory lexical processing. Sensitivity to fine phonetic detail and insensitivity to suffix reduction*, Ph.D. thesis Univ. of Nijmegen: 159 pp.
- [25] Klaassen-Don, L. E. O. (1983), *The influence of vowels on the perception of consonants*, Ph.D. thesis Univ. of Leiden: 153 pp.
- [26] Kovačić, G. Boersma, P. & Domitročić, H. (2003), "Long-term average spectra in professional folk singing voices: A comparison of the *klapa* and *dozivački* styles", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 25:53-64.
- [27] Kuijpers, C. T. L. (1993), *Temporal coordination in speech development. A study on voicing contrast and assimilation*, Ph.D. thesis Univ. of Amsterdam: 165 pp.
- [28] Lindblom, B. & Studdert-Kennedy, M. (1967), "On the role of formant transitions in vowel recognition", *J. Acoust. Soc. Amer.* 42: 830-843.
- [29] Mozziconacci, S. (1998), *Speech variability and emotion: Production and perception*, Ph.D. thesis Technological Univ. of Eindhoven: 210 pp.
- [30] Nierop, D.J.P.J. van, Pols, L.C.W. & Plomp, R. (1973), "Frequency analysis of Dutch vowels from 25 female speakers", *Acustica* 29(2): 110-118.
- [31] Odé, C. (2003), "Description and transcription of Russian Intonation (ToRI)", In: J. Schaeken, P. Houtzagers & J. Klasbeek (Eds.), *Dutch contributions to the thirteenth international congress of Slavists*, Studies in Slavic and General Linguistics, Rodopi: Amsterdam, 279-288.
- [32] Ölander, A, Perridon, H. & Pols, L.C.W. (1999), "Length acquisition of stressed syllables in Swedish by Dutch L2 learners", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 23: 67-76.
- [33] Oostdijk, N., Goedertier, W., Eynde, F. van, Boves, L., Martens, J.P., Moortgat, M. & Baayen, H. (2002), "Experiences from the Spoken Dutch Corpus project", *Proc. 3rd LREC*: 340-347.
- [34] Perkell, J.S. & Klatt, D.H. (Eds.) (1986), *Invariance and variability in speech processes*, LEA, Hillsdale NJ: 604 pp.
- [35] Plomp, R. (2002), *The intelligent ear. On the nature of sound perception*, LEA, Mahwah, NJ: 174 pp.
- [36] Plomp, R., Pols, L.C.W. & Geer, J.P. van der (1967), "Dimensional analysis of vowel spectra", *J. Acoust. Soc. Amer.* 41:707-712.
- [37] Pols, L.C.W. (1977), *Spectral analysis and identification of Dutch vowels in monosyllabic words*, Ph.D. Thesis Free Univ. of Amsterdam: 152 pp.
- [38] Pols, L.C.W. (1979), "Real-time recognition of spoken words", *IEEE Trans. Comp.* C20: 972-977.
- [39] Pols, L.C.W. (1983), "Three-mode principal components analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation", *Speech Communication* 2(4): 275-293.
- [40] Pols, L.C.W. (1986), "Variation and interaction in speech", In: J.S. Perkell & D.H. Klatt (Eds.): 140-154.
- [41] Pols, L.C.W. (1991), "Quality assessment of text-to-speech synthesis-by-rule", In: S. Furui & M.M. Sondhi (Eds.), *Advances in Speech Signal Processing*, Marcel Dekker Inc., Chapter 13: 387-416.

- [42] Pols, L.C.W. (1994), "Speech technology systems: Performance and evaluation", In: R.E. Asher (Ed.), *Encyclopedia of Language & Linguistics*, Pergamon Press, Oxford: 4289-4296.
- [43] Pols, L.C.W. (2000), "Even better synthesis", In: W.F. Sendlmeier (Ed.), *Speech and signals - Aspects of speech synthesis and automatic speech recognition*, Hector, Frankfurt am Main: 15-25.
- [44] Pols, L.C.W. (2004a), "Current developments in phonetics", In: F. Slifka, S. Manuel and M. Matthies (Eds.), *Proc. 'From Sound to Sense: 50+ Years of Discoveries in Speech Communication'*, MIT Cambridge: B52-B61.
- [45] Pols, L.C.W. (2004b), "Expanding Phonetics", In: H. Quené and V.J. van Heuven (Eds.), *On speech and language: Studies for Sieb G. Nootboom*, LOT Occasional Series 2: 141-148.
- [46] Pols, L.C.W., Kamp, L.J.Th. van der & Plomp, R. (1969), "Perceptual and physical space of vowel sounds", *J. Acoust. Soc. Amer.* 46: 458-467.
- [47] Pols, L.C.W. & Olive, J.P. (1983), "Intelligibility of consonants in CVC utterances produced by dyadic rule synthesis", *Speech Communication* 2(1): 3-13.
- [48] Pols, L.C.W. & SAM-partners (1992), "Multi-lingual synthesis evaluation methods", *Proc. ICSLP'92*, Banff, Canada, Vol. 1: 181-184.
- [49] Pols, L.C.W. & Schouten, M.E.H. (1987), "Perception of tone, band, and formant sweeps", In: M.E.H. Schouten (Ed.), *The psychophysics of speech perception*, Martinus Nijhoff Publishers, Dordrecht: 231-240.
- [50] Pols, L.C.W. & Son, R.J.J.H. van (1993), "Acoustics and perception of dynamic vowel segments", *Speech Communication* 13: 135-147.
- [51] Pols, L.C.W., Tromp, H.R.C. en Plomp, R. (1973), "Frequency analysis of Dutch vowels from 50 male speakers", *J. Acoust. Soc. Amer.* 53(4): 1093-1101.
- [52] Pols, L.C.W., Wang, X. & Bosch, L.F.M. ten (1996), "Modelling of phone duration (using the TIMIT database) and its potential benefit for ASR", *Speech Communication* 19: 161-176.
- [53] Rechziegel, A. (2001), "Consonants in contact: On assimilation and cross-language contrast", *Proc. Inst. of Phonetic Sciences, Univ. of Amsterdam* 24: 103-115.
- [54] Reetz, H. (1996), *Pitch perception in speech: A time domain approach. Implementation and evaluation*, Ph.D. thesis Univ. of Amsterdam: 236 pp.
- [55] Santen, J.P.H. van, Pols, L.C.W., Abe, M., Kahn, D., Keller, E. & Vonwiller, J. (1998), "Report on the Third ESCA TTS Workshop Evaluation Procedure", *Proc. Third ESCA TTS Workshop*, Jenolan Caves, Australia: 329-332.
- [56] Schouten, M.E.H. & Pols, L.C.W. (1983), "Perception of plosive consonants. The relative contributions of bursts and vocalic transitions", In: M.P.R. van den Broecke, V.J. van Heuven & W. Zonneveld (Eds.), *Sound Structures: Studies for Antonie Cohen*, Foris Publications, Dordrecht: 227-243.
- [57] Schwippert, C. (2000), "Perception of a b-d contrast by adult dyslexic and control subjects", IFA internal report.
- [58] Schwippert, C. E., Koopmans-van Beinum, F.J. & Leeuwen, T.H. van (1999), "Phoneme boundary perception in relationship to developmental dyslexia", *Proc. ICPHS'99*, San Fransisco, CA, Vol. 2: 877-880.
- [59] Son, R. J. J. H. van (1993), *Spectro-temporal features of vowel segments*, Ph.D. thesis Univ. of Amsterdam: 195 pp.
- [60] Son, R. J. J. H. van, Binnenpoorte, D., van den Heuvel, H. & Pols, L.C.W. (2001), "The IFA corpus: a phonemically segmented Dutch 'open source' speech database". *Proc. Eurospeech'01*, Aalborg, Denmark, Vol. 3: 2051-2054.
- [61] Son, R.J.J.H. van, Bolotova, O., Lennes, M. & Pols, L.C.W. (2004), "Frequency effects on vowel reduction in three typologically different languages (Dutch, Finnish, Russian)", *Proc. Interspeech 2004 - ICSLP*, Jeju Island, Korea, Vol. II: 1277-1280.
- [62] Son, R. J. J. H. van & Pols, L. C. W. (1992), "Formant movements of Dutch vowels in a text, read at normal and fast rate", *J. Acoust. Soc. Amer.* 92(1): 121-127
- [63] Son, R. J. J. H. van & Pols, L. C. W. (1997), "The correlation between consonant identification and the amount of acoustic consonant reduction", *Proc. Eurospeech'97*, Rhodes, Vol. 4: 2135-2138.
- [64] Son, R. J. J. H. van & Pols, L. C. W. (1999a), "An acoustic description of consonant reduction", *Speech Communication* 28(2): 125-140.
- [65] Son, R. J. J. H. van & Pols, L. C. W. (1999b), "Perisegmental speech improves consonant and vowel identification", *Speech Communication*, 29(1): 1-22.

- [66] Son, R.J.J.H. van & Pols, L.C.W. (2001), "Structure and access of the open source IFA-corpus", *Proc. IRCS Workshop on Linguistic Databases*, S. Bird, P. Buneman & M. Liberman (Eds.): 245-253.
- [67] Son, R. J. J. H. van & Pols, L. C. W. (2003), "Information structure and efficiency in speech production", *Proc. Eurospeech '03*, Geneva, Vol. 1: 769-772.
- [68] Steeneken, H. J. M. (1992), *On measuring and predicting speech intelligibility*, Ph.D. thesis Univ. of Amsterdam: 165 pp.
- [69] Stelt, J. M. van der (1993), *Finally a word: A sensori-motor approach of the mother-infant system in its development towards speech*, Ph.D. thesis Univ. of Amsterdam: 226 pp.
- [70] Stelt, J.M. van der, Wempe, T.G. & Pols, L.C.W. (2003), "Progression in infants' vowel space: An analysis of deaf and hearing infants' sounds", *Proc. 15th ICPHS*, Barcelona, Vol. 3: 2225-2228.
- [71] Streefkerk, B.M. (2002). *Prominence. Acoustic and lexical/syntactic correlates*, Ph.D. thesis Univ. of Amsterdam: 168 pp.
- [72] Streefkerk, B. M., Pols, L. C. W. & ten Bosch, L.F.M. (2001). "Up to what level can acoustical and textual features predict prominence", *Proc. Eurospeech '01*, Aalborg, Denmark, Vol. 2: 811-814.
- [73] Tielen, M. T. J. (1992), *Male and female speech. An experimental study of sex-related voice and pronunciation characteristics*, Ph.D. thesis Univ. of Amsterdam: 180 pp.
- [74] Verdonck-de Leeuw, I. M. (1998), *Voice characteristics following radiotherapy: the development of a protocol*, Ph.D. thesis Univ. of Amsterdam: 137 pp.
- [75] Wang, X. (1997), *Incorporating knowledge on segmental duration in HMM-based continuous speech recognition*, Ph.D. thesis Univ. of Amsterdam: 190 pp.
- [76] Wempe, A.G. & Boersma, P. (2003): "The interactive design of an F₀-related spectral analyser", *Proc. 15th ICPHS*, Barcelona, Vol. 1: 343-346.
- [77] Wesseling, W. & Son, R.J.J.H. van (presented for publication), "Timing of experimentally elicited minimal responses as quantitative evidence for the use of intonation in projecting TRP's", *Proc. Interspeech 2005*, Lisbon.
- [78] Wieringen, A. van (1995), *Perceiving dynamic speechlike sounds. Psycho-acoustics and speech perception*, Ph.D. thesis Univ. of Amsterdam: 256 pp.
- [79] Wieringen, A. van & Pols, L.C.W. (1995), "Discrimination of single and complex consonant-vowel- and vowel-consonant-like formant transitions", *J. Acoust. Soc. Amer.* 93(3): 1304-1312.