De dag van de *Fonetiek* 2008

Over onderzoek naar spraak en spraaktechnologie

(http://www.fon.hum.uva.nl/FonetischeVereniging/)

Donderdag 18 december 2008 in de Sweelinckzaal, Drift 21 te Utrecht

Georganiseerd door de *Nederlandse Vereniging voor Fonetische Wetenschappen*

deelname gratis

WORD LID VAN DE
VERENIGING VOOR
FONETISCHE WETENSCHAPPEN

Vul het formulier in en stuur het naar het onderstaande adres of email de gegevens naar W.F.L.Heeren@ewi.utwente.nl.

achternaam: .....

voorletter(s) evt. titel: .....

afdeling/vakgroep: .....

postadres

werk- of priveadres: .....

postcode en plaats: .....

emailadres: .....

De contributie is 7 Euro / jaar

Aanmelden als lid bij:
Willemijn Heeren, waarnemend secretaris NVFW
Human Media Interaction - CHoral project
Universiteit Twente & Gemeentearchief Rotterdam
w.f.l.heeren@ewi.utwente.nl
wwwhome.cs.utwente.nl/~heerenwfl/

Hier kunt u ook terecht voor meer informatie over de
Vereniging voor Fonetische Wetenschappen

16:20 Het modelleren van subtiele fonetische informatie in een computationeel model van menselijke woordherkenning

Odette Scharenborg, Radboud Universiteit Nijmegen

In de afgelopen jaren hebben verschillende psycholinguïstische experimenten laten zien dat luisteraars al voor het einde van 'ham' weten of de spreker het broodbeleg bedoelt of het huisdiertje. Maar hoe komt het dat mensen – in ieder geval in het laboratorium – in staat zijn om de grenzen tussen woorden zo snel en trefzeker te vinden dat ze al verschil kunnen maken tussen de woorden 'ham' en 'hamster' voor het einde van de 'ham"?

Het blijkt dat er subtiele fonetische informatie in het spraaksignaal zit die aangeeft of het einde van een woord in aantocht is of niet. Het is al langer bekend uit fonetisch onderzoek dat er in laboratoriumspraak (onder andere) subtiele verschillen in duur zijn die samenhangen met het aantal lettergrepen dat nog volgt tot aan het einde van het woord. Luisteraars blijken deze informatie dus te kunnen gebruiken tijdens het luisteren naar spraak. Maar hoe doen luisteraars dat nu eigenlijk?

Om een verklaring te vinden voor hoe mensen dat doen gebruiken wij in dit onderzoek een computermodel van de verwerking van spraaksignalen en de opslag van woorden in ons brein. De meest invloedrijke computationele modellen van auditieve woordherkenning kunnen echter deze subtiele fonetische informatie niet representeren en dus ook niet gebruiken tijdens woordherkenning. Wij presenteren een nieuw computationeel model, Fine-Tracker, dat dit wel kan. Fine-Tracker is een computermodel dat is ontwikkeld met gebruikmaking van technieken uit de automatische spraakherkenning en heeft net als automatische spraakherkensystemen echte spraak als input. Op deze manier slaat dit onderzoek een brug tussen de onderzoeksgebieden van de automatische spraakherkenning en de psycholinguïstiek.

16:00 Web-resource "Russian Dialectal Phonetics" as a model of effective authoring procedures for educational linguistic hypermedia e-learning content development

G. Kedrova (Engelstalige presentatie)

An original and efficient model of hypertextual authoring processes and guidelines for electronic multi-media educational and scientific resources' development will be presented for discussion. Basic authoring principles were tested in the course of development of the interactive Web-resource "Русская диалектная фонетика (Russian Dialectal Phonetics)". This new open-source electronic public educational and research resource in Russian linguistics is the first one based on the ideology of "Shareable Courseware Object Reference Model" (SCORM). The Russian dialectology electronic course's structure comprises two major components: a theoretical overview and practical sections (on-line self-tests, exercises). An interactive glossary of linguistic and other special terms is also appended through dense network of hyperlinks. As essential multi- and hypermedia product the course comprises texts, charts, sonagrams, intonograms and diagrams, images (dialect maps), authentic dialectal spoken language records from the archive of dialectological expeditions of the Philological Faculty of the MGU.

Programma

9:00 Ontvangst met koffie

9:15 Welkom

9:30-10:50 Ochtendsessie I

9:30 Meer stemmen voor Nederlandse spraaksynthese
Arthur Dirksen

9:50 RechtSpraakHerkenning: Nederlandse spraakherkenning in de rechtszaal
AJ van Hessen

10:10 Een menselijke benchmark voor automatische taalherkenning
Rosemary Orr en David van Leeuwen

10:30 What's in a name? Autonomata Too!
Henk van den Heuvel

10:50 Koffiepauze

11:20-12:20 Ochtendsessie II

11:20 Een audiovisuele spontane emotie-database van gamers
Khiet Truong, Mark Neerincx en David van Leeuwen

11:40 Perceptie-effecten van geografische variatie in (micro-)prosodische eigenschappen: In hoeverre kunnen luisteraars de leeftijd, de lengte en het gewicht van een spreker raden?
Marie Nilsenová

12:00 Modaliteit in spontane en geacteerde spraak
Deelnemers MA-onderzoekscollege Fonologie, Rijksuniversiteit Groningen

12:20 Lunch

13:50-15:10 Middagsessie I

13:50 Leren en doceren van klinkers
Luc van Buuren

14:10 Perceptie van onvolledig spraaksignaal
Bea Valkenier en Dicky Gilbers

14:30 Nederlandse baby's gebruiken statistische informatie om spraakklanken te leren onderscheiden
Desiree Capel, Elise de Bree, Annemarie Kerkhoff en Frank Wijnen

14:50 Voorspellers van 'audiovisual benefit' voor het perceptief scheiden van stemmen bij oudere luisteraars
Esther Janse en Alexandra Jesse

15:10 Thee

15:40-16:40 Middagsessie II

15:40 Transcriptie van Russische Intonatie ToRI, een interactieve module op het Internet
Cecilia Odé

16:00 Web-resource "Russian Dialectal Phonetics" as a model of effective authoring procedures for educational linguistic hypermedia e-learning content development
G. Kedrova (Engelstalige presentatie)

16:20 Het modelleren van subtiele fonetische informatie in een computationeel model van menselijke woordherkenning
Odette Scharenborg

16:40 Afsluiting en borrel

9:30 Meer stemmen voor Nederlandse spraaksynthese

Arthur Dirksen, Fluency, Amsterdam

In mijn bijdrage van vorig jaar heb ik een overzicht gegeven van de nieuwe spraaksynthesizer van Fluency, toen nog volop in ontwikkeling. Inmiddels is de software verder ontwikkeld, en voorzien van zeven levendige, levensechte stemmen: drie mannen, twee vrouwen, en twee tieners (een jongen van 13 en een meisje van 16).

Bijzonder aan de nieuwe synthesizer is dat het betrekkelijk eenvoudig is een nieuwe stem te maken. De spreker moet een corpus inspreken dat bestaat uit 387 woorden en 387 zinnen (totaal 774 items). Elk item wordt door de spraaksynthese voorgezegd, en de spreker moet dit vrij precies nazeggen, met name wat betreft pauzes. De opnames kunnen door Fluency grotendeels automatisch worden omgezet in een spraakdatabase voor de synthesizer.

De software om het corpus op te nemen is vrij beschikbaar, en het opnemen van een nieuwe stem vereist geen grote investeringen in hardware: met een usb-microfoon en een notebook kan al een goede kwaliteit bereikt worden.

In deze bijdrage wil ik nader ingaan op de mogelijkheden die dit biedt voor spraakgehandicapten, en hoe andere partijen hierop kunnen inspelen.

9:50 RechtSpraakHerkenning: Nederlandse spraakherkenning in de rechtszaal

AJ van Hessen, Universiteit Twente

In toenemende mate moeten verhoren door politie volledig worden opgenomen. Ingeval van twijfel, kan dan altijd de oorspronkelijke opname opnieuw beluisterd worden. Ook de Nederlandse rechtbanken experimenteren met geluidsopnamen. De griffier maakt altijd het verslag van de rechtszitting, maar omdat het soms lastig is alles direct tijdens de zitting correct te noteren, worden er al voor intern gebruik dikwijls geluidsopnamen gemaakt: alles wat er gezegd wordt op een cassettebandje!

Door iedere spreker echter op een eigen spoor op te nemen en de opnamen door de spraakherkenner te halen, kan veel meer bereikt worden. De opnamen worden namelijk doorzoekbaar op zowel spreker als spraak. Iedereen die straks toegang heeft tot de opnamen kan met een paar simpele klikken zoeken naar de woorden X,Y en Z, uitgesproken door verdachte A of Rechter B.

De griffier kan de spraakherkenningsresultaten gebruiken om sneller een verslag te maken en rechters kunnen naar een gesproken samenvatting luisteren; bedoeld om hun geheugen op te frissen als ze de zaak weer oppakken na een langdurige onderbreking.

De Taal- en Spraaktechnologie wordt in het RechtSpraakHerkenningsproject ingezet voor de ondersteuning van de rechtbank, niet als vervanging van medewerkers. Rechtspraak blijft vooralsnog toch echt mensenwerk.

14:50 Voorspellers van 'audiovisual benefit' voor het perceptief scheiden van stemmen bij oudere luisteraars

Esther Janse en Alexandra Jesse, MPI Nijmegen

Oudere luisteraars hebben over het algemeen meer moeite om de spraak van een bepaalde spreker gescheiden te houden van een of meerdere concurrerende sprekers op de achtergrond. In deze studie onderzoeken we hoeveel baat iemand heeft bij het zien van het gezicht van de doelspreker bovenop het alleen horen van een mix van twee concurrerende stemmen. Veertig oudere luisteraars (65-plussers), varierend in mate van gehoorverlies, deden mee aan deze foneemdetectiestudie. Daarnaast werden een aantal achtergrondtesten bij hen afgenomen: gehoorverlies, lipleesscore, informatieverwerkingssnelheid, executief functioneren (planning en organisatie), en selectieve aandacht. We onderzochten welke van deze achtergrondmaten correleerden met gemiddelde foneemdetectiescore en met de mate van 'audiovisual benefit'. De resultaten zullen besproken worden.

15:40 Transcriptie van Russische Intonatie ToRI, een interactieve module op het Internet

Cecilia Odé, IFA, Universiteit van Amsterdam

In een audiovisuele demonstratie zal ik een nieuw systeem presenteren voor het transcriberen van Russische intonatie: ToRI, gratis beschikbaar op het Internet. ToRI maakt gebruik van éénduidige symbolen voor de transcriptie van toonhoogteaccenten, verbindende toonhoogtebewegingen en grenzen van uitingen gemarkeerd door toonhoogte. De beschrijving van alle toonhoogteverschijnselen in ToRI is gebaseerd op de resultaten van perceptie experimenten met moedertaalsprekers van het Russisch. Het systeem geeft ook de fonetische correlaten voor de realisatie van toonhoogteaccenten. In ToRI worden de toonhoogteaccenten gepresenteerd met audiovisuele voorbeelden en oefeningen voor het leren herkennen van toonhoogteaccenten en grensmarkeringen. In de voorbeelden en oefeningen worden ook de communicatieve functies van de accenten gegeven. Een alfabetische woordenlijst verklaart de in het systeem gebruikte terminologie. Het systeem is zodanig opgezet dat het als leermodule voor linguisten en vergevorderde studenten kan worden gebruikt, individueel of in een klassituatie.

14:10 Perceptie van onvolledig spraaksignaal

Bea Valkenier en Dicky Gilbers, Rijksuniversiteit Groningen

Ito et al (2001) betogen dat de perceptie van /i,e,a,o,u/ niet alleen berust op de eerste twee formanten in het akoestisch signaal. Ook bij onderdrukking van een van deze formanten worden de vocalen goed geïdentificeerd.

In een vervolgexperiment hebben we 12 participanten 4 voorvocalen aangeboden, niet alleen de primaire kardinale vocalen /i,e/ maar ook de secundaire kardinale vocalen /y,2/. De participanten kregen alle vocalen zowel met als zonder tweede formant aangeboden. Wij concluderen evenals Ito et al (2001) dat primaire kardinale vocalen in beide condities goed worden waargenomen. Onvolledig gespecificeerde secundaire kardinale vocalen worden echter significant vaker als hun primaire kardinale tegenhanger waargenomen.

Wat zijn de perceptieve en/of cognitieve strategieën van de luisteraar die dit verschil in perceptie verklaren? Wij zullen betogen dat de luisteraar prototypes van klanken heeft opgeslagen en dat deze ideaalpatronen de perceptie beïnvloeden.

Ito, M., J. Tsuchida & M. Yano (2001) On the effectiveness of whole spectral shape for vowel perception. J.Acoust. Soc. Am. 110 (2).

14:30 Nederlandse baby's gebruiken statistische informatie om spraakklanken te leren onderscheiden

Desiree Capel, Elise de Bree, Annemarie Kerkhoff en Frank Wijnen
UiL-OTS Universiteit Utrecht

Baby's hebben aanvankelijk een 'universele' spraakperceptie. Zij zijn in staat om foneemcontrasten uit alle natuurlijke talen te onderscheiden. Dit vermogen verdwijnt echter gedurende het eerste levensjaar en wordt meer moedertaalspecifiek. Maye et al. (Cognition, 2002) suggereren dat (onder andere) statistisch leren verantwoordelijk is voor deze verandering. Maye et al. waren de eerste die aantoonden dat 6 en 8 maanden oude baby's bij het leren onderscheiden van spraakklanken gebruik maken van de statistische distributie van fonetische variatie. In een replicatie van dit experiment werden 10 tot 11 maanden oude Nederlandse baby's blootgesteld aan ofwel een bimodale ofwel een unimodale frequentiedistributie van een 8-staps spraakklankcontinuüm. Dit continuüm was gebaseerd op de Hindi stemhebbende en stemloze retroflexe plosieven (/ɖa/ en /ʈa/). De resultaten laten zien dat alleen baby's in de bimodale groep na de blootstelling reageren op het verschil tussen stemloos en stemhebbend. Dit wijst erop dat de spraakklanken voor deze groep in twee categorieën worden gerepresenteerd. Samenvattend kan gezegd worden dat de resultaten van het huidige experiment de hypothese ondersteunen dat baby's statistisch leren aanwenden om foneemcategorieën te vormen.

10:10 Een menselijke benchmark voor automatische taalherkenning

Rosemary Orr en David van Leeuwen
University College Utrecht/TNO Human Factors

Automatische taalherkenning heeft als doel het herkennen van de taal die gesproken wordt in een spraakfragment. Regelmatig worden wereldwijd systemen langs de lat gelegd in benchmark evaluaties, uitgevoerd door het Amerikaans NIST, en systemen worden steeds beter. Maar hoe goed kunnen mensen dat eigenlijk? En hoe meet je zoiets, en waar hangen de prestaties van af? We willen de resultaten presenteren van een onderzoek dat we bij het International Computer Science Institute in Berkeley en het University College Utrecht hebben uitgevoerd. En mensen doen het zo gek nog niet---als ze de taal in kwestie een beetje kennen.

10:30 What's in a name? Autonomata Too!

Henk van den Heuvel, CLST, Radboud Universiteit Nijmegen

Het Autonomata Too project is een project in het STEVIN-programma. In dit project proberen we de automatische spraakherkenning van Nederlandse, Vlaamse en buitenlandse namen te verbeteren door rekening te houden met uitspraakvarianten van namen ten gevolge van interculturele fenomenen, meer specifiek de oorsprong van een naam en de oorsprong van de spreker van een naam. Deze verschijnselen worden op een aantal niveaus onderzocht: variaties van uitspraken binnen de Nederlandse foneemset; het nut van het toevoegen van fonemen uit een buitenlandse foneemset; het aantal varianten dat leidt tot een optimale herkenning. Autonomata Too maakt hierbij gebruik van een namencorpus dat is opgenomen in een eerder project: Autonomata. Dit namencorpus wordt gebruikt om patronen op te sporen die ontstaan als Nederlanders namen van Nederlandse komaf uitspreken, c.q. namen van buitenlandse komaf, of als buitenlandse sprekers (met enige kennis van het Nederlands) namen van Nederlandse of buitenlandse oorsprong uitspreken. In de voordracht zal het Autonomata Too project worden voorgesteld en een aantal eerste onderzoeksresultaten worden gepresenteerd.

11:20 Een audiovisuele spontane emotie-database van gamers

Khiet Truong, Mark Neerincx en David van Leeuwen,
TNO Defensie en Veiligheid, Soesterberg

Een spontane audiovisuele emotie-database is opgenomen met het doel om automatische emotieherkenners te ontwikkelen. 28 proefpersonen hebben een videospel gespeeld (Unreal Tournament) waarin bepaalde spelelementen zijn gemanipuleerd om emoties uit te lokken. Er zijn spraak- en gezichtsopnamen gemaakt die na afloop door de gamers zelf zijn geannoteerd op emotie. Met deze opgenomen data is een aantal experimenten uitgevoerd. Ten eerste hebben we gekeken naar hoe het aanbieden van uni- of multimodale informatie (bijv. alleen audio, alleen video of beiden) de beoordeling van emotie beinvloedt. Ten tweede hebben we gekeken naar de betrouwbaarheid van de eigen emotiebeoordelingen van de gamers. Het uiteindelijke doel is om automatisch emotie in spraak te detecteren; we zullen voorlopige resultaten van een aantal emotieclassificatie-experimenten laten zien.

11:40 Perceptie-effecten van geografische variatie in (micro-)prosodische eigenschappen: In hoeverre kunnen luisteraars de leeftijd, de lengte en het gewicht van een spreker raden?

Marie Nilsenová, Universiteit Tilburg

Hoe goed zijn luisteraars in het inschatten van de leeftijd, lengte en gewicht van een spreker op basis van haar/zijn stem? De resultaten van een aantal experimenten van Lass et al (1979, 1980a, 1980b, 1980c, 1980d) voor het Engels waren positief, maar achteraf blijken de data statistisch op een niet betrouwbare manier geanalyseerd te zijn (Cohen et al. 1980). Ook andere experimentele resultaten (Gunter & Manning 1982, van Dommelen & Moxness 1995) doen twijfel rijzen omtrent de conclusies van Lass et al.. Toch bestaat er een duidelijke relatie tussen de leeftijd, lengte en gewicht van een spreker en spraakparameters zoals pitch en formanten, gebaseerd vooral op de lengte van de vocal tract en de grootte en dikte van de stembanden (Greisbach 1999). Een mogelijke reden voor de ruis in de data is de sociale en geografische vrije (niet-linguïstische) variatie tussen sprekers. In dat geval verwachten we dat sprekers van een geografische variant beter in staat zouden zijn om sprekers van dezelfde variant te beoordelen in vergelijking met sprekers van een andere variant. In onze experimenten hebben we Nederlandse (Vlaamse) luisteraars in Antwerpen met Nederlandse luisteraars in Tilburg vergeleken in hun inschatting van de leeftijd, de lengte en het gewicht van mannelijke en vrouwelijke sprekers van de twee varianten in twee tussen-proefpersoon condities; met spraak die achterstevoren wordt afgespeeld (`reversed speech' ≠ om het duidelijk hoorbaar verschil tussen Vlaams en Nederlands te maskeren) en met gewoon afgespeelde spraak. Onze verwachting dat luisteraars beter op hun eigen variant zouden presteren wordt duidelijk bevestigd voor alle drie de eigenschappen (leeftijd, lengte, gewicht) met significante interactie-effecten voor geslacht.

12:00 Modaliteit in spontane en geacteerde spraak

Deelnemers MA-onderzoekscollege Fonologie, Rijksuniversiteit Groningen

Frequentieanalyses van spraakfragmenten laten vaak één piek zien. In emotionele spraak daarentegen zijn soms meerdere pieken te vinden (modaliteit). Schreuder, van Eerten & Gilbers (2006) hebben een pilotstudie gedaan naar modaliteit in voorgelezen, emotionele spraak. Zij vinden in sombere passages mineurmodaliteit (3 semitonen afstand tussen de pieken) en in vrolijke majeurmodaliteit (4 semitonen afstand).

In ons vervolgonderzoek is behalve geacteerde emotionele spraak (Bert & Ernie, soapseries, motherese) ook spontane spraak (winnaars en verliezers in sportinterviews) onderzocht op modaliteit. De uitkomst is dat modaliteit frequent in geacteerde spraak voorkomt maar nauwelijks in spontane spraak. Mineur en majeur vinden we alleen duidelijk in overacting. In motherese vinden we wel veel modaliteit, maar niet altijd de verwachte majeurmodaliteit.

Volgens Boersma (2007) “beschrijven” cue constraints ideaalpatronen voor de productie en perceptie van spraak. De resultaten van het hier gepresenteerde onderzoek geven aan dat in spontane spraak cue constraints minder sterk opereren dan in geacteerde spraak.

Boersma, Paul (2007), Cue constraints and their interactions in phonological perception and production. ROA 944

Schreuder, Maartje, Laura van Eerten & Dicky Gilbers (2006), Mineur en Majeur in Emotionele Spraak, in: Tabu 35, 1/2, p. 1-14

13:50 Leren en doceren van klinkers

Luc van Buuren, Linguavox

L2 sprekers van het Nederlands hebben veelal grote problemen met de klinkers. Blijkbaar zijn wij niet (langer) in staat die te doceren. Hetzelfde geldt trouwens voor L2 (i.c. Engels) sprekende Nederlanders.

Uitgaande van Catford (1988) hoofdstuk 7-8 en Van Buuren (1993) hoofdstuk 7-9 (op Linguavox.nl/klinkers) wil ik weer eens Daniel Jones’ Cardinal Vowel benadering onder de aandacht brengen en enkele verbeteringen en verfijningen voorstellen.

J.C. Catford (1988, 2002). A Practical Introduction to Phonetics

L. van Buuren (1975, 1993). English Phonetics Course