

De dag van de *Fonetiek* 2006

Over onderzoek naar
spraak en spraaktechnologie

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

Donderdag 21 december 2006 in de Sweelinckzaal, Drift 21 te Utrecht

Georganiseerd door de *Nederlandse Vereniging voor Fonetische Wetenschappen*

deelname gratis



Nederlandse
Vereniging
Voor
Fonetische
Wetenschappen

WORD LID VAN DE *VERENIGING VOOR FONETISCHE WETENSCHAPPEN*

Vul het formulier in en stuur het naar het onderstaande adres of email de gegevens naar Mirjam.Ernestus@mpi.nl.

achternaam:

voorletter(s) evt. titel:

afdeling/vakgroep:

postadres

werk- of priveadres:

postcode en plaats:

emailadres:

De contributie is 7 Euro / jaar

Aanmelding als lid bij:

Mirjam Ernestus
Max Planck Institute for Psycholinguistics
Postbus 310
6500 AH Nijmegen
tel: +31-24-3612970
email: Mirjam.Ernestus@mpi.nl

Voor meer informatie over de *Vereniging voor Fonetische Wetenschappen*:

Rob van Son
Leerstoelgroep Fonetische Wetenschappen
Universiteit van Amsterdam
Spuistraat 210-212
1012 VT Amsterdam
Tel.: 020-5252196
Fax: 020-5252197
Email: R.J.J.H.vanSon@uva.nl
URL: <http://www.fon.hum.uva.nl/FonetischeVereniging/>

Programma

9:00 Ontvangst met koffie

9:15 Welkom

9:20-10:40 Ochtendsessie I (voorzitter Bert Schouten)

9.20 Iets over rythme en timing in NL en GB

Luc van Buuren

9.40 Over het perceptieve belang van prosodie bij oude en jonge luisteraars

Hugo Quené

10.00 Prominent Words as Anchors for TRP Projection

Wieneke Wesseling & Rob van Son

10.20 Gevarieerde, natuurlijke intonatie voor difoonsynthese

Arthur Dirksen

10:40 Koffiepauze

11:10-12:10 Ochtendsessie II (voorzitter Mirjam Ernestus)

11.10 Stemhebbend-stemloosheid in Nederlandse tracheoesofageale sprekers

Petra Jongmans

11.30 How do alaryngeal speakers convey glottal stops?

Maya van Rossum

11.50 Spoken word processing and the effect of phonemic mismatch in aphasia and dyslexia

Esther Janse, Elise de Bree & Susanne Brouwer

12:10 Lunch

13:40-15:00 Middagsessie I (voorzitter Wilbert Heeringa)

13.40 Noord-Russische vraagintonatie: Een brede hoed voor brede focus?

Margje Post

14.00 E-learning resource 'Russian Dialectal Phonetics': Perspectives and first output

Sergej Kniazev (Engelstalige bijdrage)

14.20 Identifying Linguistic Structure in the Aggregate Analysis of Dialect Pronunciation

Jelena Prokic (Engelstalige bijdrage)

14.40 Vowel pronunciation in Swedish dialects

Therese Leinonen (Engelstalige bijdrage)

15:00 Thee

15:30-16:30 Middagsessie II (voorzitter Rob van Son)

15.30 JASMIN-CGN: Uitbreiding van het CGN met spraak van Jongeren, Anderstaligen en Senioren

Olga van Herwijnen & Catia Cucchiarini

15.50 "Landgenooten..." Verbeterde toegang tot historische toespraken

Willemijn Heeren & Laurens van der Werff

16.10 Automatische emotie-detectie in spraak in het project MultimediaN: Multi-Modal Interaction

Khiet Truong

16:30 Afsluiting

9.20 Iets over rythme en timing in NL en GB

Luc van Buuren

Linguavox

Mijn definitie van rythme luidt: het ‘rappen’, binnen een psychologisch/neurologisch praesens van $\pm 2-5$ seconden, van motorische eenheden (h.l. syllabes) in (hierarchieën van) MOON, TROchee, jamBOS, (+)DACtylus, (+)amPHIbrach(+) en anaPEST(+).

(De + eveneens zwakke syllabe aanduidend, komt dit neer op 11 mogelijke ‘voeten’. ‘RAPpen’ = groeperen in overeenstemming met een Rhythmisch Alternatie Principe inhoudende: niet meer dan 2 zwakkeren tussen Sterkeren en niet meer dan 1 aan begin en eind van locutie).

Deze theorie laat zich voor GB illustreren in (1), voor NL in (2): zelfde scandering (maar andere timing).

(1) the secretary for education was at the front door, not in the living-room.

(2) het vreselijkste gesodemier was met de m.e., niet met de kleuterjuf.

Vervolgens de vraag hoe de beschrijving voor GB/NL te voltooien.

Eerdere pogingen tot beantwoording van deze vraag o.m. in hfdst. 8 van mijn Indispensable Foundation (1990), afscheidscollege (2000), artikelen in Festschrift Barentsen (ed. Honselaar e.a.) (2003) en Lacus Forum (2004/5/6).

9.40 Over het perceptieve belang van prosodie bij oude en jonge luisteraars

Hugo Quené

Utrecht inst. of Linguistics OTS, Utrecht University

In een eerdere bijdrage (Dag van de Fonetiek 2003) heb ik laten zien dat een regelmatige 'timing' van woorden in een woordenlijst leidt tot snellere reactietijden in een foneem-detectie-taak. Temporele regelmaat maakt duidelijk dat luisteraars de gesproken woorden sneller kunnen herkennen. Dat eerdere onderzoek was uitgevoerd met jong-volwassen luisteraars. We voorspellen dat het perceptieve belang van (temporele) prosodie voor oudere luisteraars nog groter is dan voor jongeren, ter compensatie van auditieve en cognitieve handicaps bij eerstgenoemden. Een herhaling van een eerder experiment, nu met oudere luisteraars, toont wisselende resultaten. Temporele regelmaat heeft alleen effect bij jambische woorden, en niet bij trocheeën. Oudere luisteraars lijken dus meer gevoelig voor de metrische structuur van het doelwoord zelf, dan voor de al dan niet regelmatige timing van de woorden in een lijst.

10.00 Prominent Words as Anchors for TRP Projection

Wieneke Wesseling & Rob van Son
IFA/ACLC, University of Amsterdam

Het effect van het laatste geaccentueerde woord van een uiting op de projectie van (potentiele) beurtwisselingsmomenten (TRPs) is onderzocht met twee reactietijd experimenten. Proefpersonen werd gevraagd met minimale responsen te reageren op vooraf opgenomen dialogen en gemanipuleerde, verarmde versies van deze dialogen. De eerste manipulatie, de 'gehumde' stimulusvariant, bevatte enkel intonatie en timing informatie. In de tweede manipulatie, de 'gefluisterde' stimulusvariant, ontbrak juist alle periodische informatie. De distributie van de reactietijden van deze uitgelokte responsen was vergelijkbaar met die van natuurlijke beurtwisselingen. Het bleek dat de aanwezigheid van niet prominente woorden voor een uiting-einde de timing van zowel de natuurlijke als de uitgelokte responsen versnelt, zelfs in de 'gefluisterde' versie. Dit suggereert dat met de aanwezigheid van een prominent, informatief woord de projectie van een opkomend uitingseinde begint. De aanwezigheid van niet prominente, voorspelbare spraak helpt de luisteraars vervolgens om hun voorspelling van het uitingseinde exact te timen.

10.20 Gevarieerde, natuurlijke intonatie voor difoonsynthese

Arthur Dirksen
Fluency

Tekst-naar-spraaksoftware die gebaseerd is op unit-selectie heeft als belangrijk voordeel dat de prosodie voldoende gevarieerd is. Die is immers direct afkomstig uit spraak van een menselijke spreker. Regelgebaseerde systemen kunnen in vergelijking hiermee wel wat 'houterig' klinken: de regels produceren elke keer hetzelfde deuntje. Het is echter geen kwestie van alles of niets. Ook in klassieke difoonsynthese kan voor de beregeling van duur en melodie gespiekt worden bij een echte spreker. In deze lezing bespreek ik een recente - en nog enigszins experimentele - toevoeging aan de tekst-naar-spraaksoftware van Fluency, die, waar mogelijk, een zin of deelzin voorziet van een passende melodie uit een corpus van natuurlijke spraak.

11.10 Stemhebbend-stemloosheid in Nederlandse tracheoesofageale sprekers

Petra Jongmans

Universiteit van Amsterdam/Antoni van Leeuwenhoek ziekenhuis

Bij keelkanker is een totale laryngectomie vaak de laatste kans. Deze operatie verandert de anatomie en fysiologie van het spraakkanaal drastisch. Door de introductie van de stemprothese is de verstaanbaarheid van deze patienten verrassend goed, maar het is nog steeds veel slechter dan bij normale spraak. Een van de meest voorkomende problemen is de verwarring tussen stemhebbende en stemloze plosieven en fricatieven. Het is interessant te weten welke akoestische cues sprekers gebruiken die wel een correct onderscheid maken en of deze cues verschillen van normale sprekers. Daarom hebben wij metingen verricht in Praat op de goed waargenomen plosieven en fricatieven, waarbij we oa hebben gekeken naar lengte van de klinkers, sluitingsduur, burst en fonatie offset na een klinker. Eerste resultaten laten een verschil zien tussen labiale en alveolaire plosieven en het feit dat patienten meer cues gebruiken dan normale sprekers. Ook lange fonatie offset lijkt een probleem. Patienten lijken secundaire cues te gebruiken als compensatie voor de verminderde controle over hun 'neoglottis'.

11.30 How do alaryngeal speakers convey glottal stops?

Maya van Rossum

Antoni van Leeuwenhoek ziekenhuis

To convey glottal stops, laryngeal speakers control the degree of constriction at the larynx. Laryngectomy – usually necessitated by laryngeal cancer – involves surgical removal of the larynx. The new voice source (neo-glottis) is formed by the mucosa and muscles situated at the entrance to the esophagus (foodpipe). Control over the neo-glottis in these tracheo-esophageal (TE) speakers, might influence attempts to convey glottal stops accurately. This study investigated how TE speakers realize glottal stops in abutting words that end and begin with the same vowel. Acoustic analyses showed that TE speakers use different strategies, most notably the insertion of glides (/w/; /j/). Listeners rated the TE speakers' attempts to convey 'glottal' stops less favourably than the control speakers' glottal stops. The TE speakers' acoustic realizations that deviated most from what would normally be expected, received the least favourable ratings.

11.50 Spoken word processing and the effect of phonemic mismatch in aphasia and dyslexia

Esther Janse, Elise de Bree & Susanne Brouwer

Utrecht inst. of Linguistics OTS, Utrecht University

Initial lexical activation in typical populations is a direct reflection of the goodness of fit between the presented stimulus and the intended target. In this study, lexical activation was investigated upon presentation of polysyllabic pseudowords (such as *procodile*) for two atypical populations: aphasic and dyslexic adults. Results of an auditory lexical decision task supported the hypothesis that sensitivity to phonemic mismatch was lower in both atypical populations, compared to their respective control groups. However, whereas the dyslexic participants were outperformed by their controls only for minimal word-initial mismatches, lexical decision accuracy of the aphasic participants was lower across the board. Both atypical groups were affected by the acoustic/perceptual difference between the changed and canonical word form. The results are discussed in terms of lexical access and the role of verbal working memory.

13.40 Noord-Russische vraagintonatie: Een brede hoed voor brede focus?

Margje Post

Universitetet i Tromsø (N)/Universiteit Leiden

Hoewel de Russische dialecten in verhouding weinig van elkaar verschillen, is de onderlinge variatie op het gebied van de prosodie aanmerkelijk. Een opvallend intonatieel verschil tussen het Standaard-Russisch en het Noord-Russische dialect van Varzoega, een dorpje aan de Witte Zee is een "brede hoed"-patroon in ja/nee-vragen in het dialect. Waar in het Standaard-Russisch de toonhoogte onmiddellijk na de geaccentueerde lettergreep aan een sterke daling begint - LH*L volgens Odé's ToRI (Transcription of Russian Intonation) - kent dit dialect, naast ditzelfde patroon, de mogelijkheid om de daling pas te beginnen na de beklemtoonde syllabe van het laatste woord van de uiting. De eerste analyses doen vermoeden dat dit onderscheid in het dialect bepaald wordt door een verschil in focusdomein. Dit zou betekenen dat sprekers van dit dialect, in tegenstelling tot sprekers van het Standaard-Russisch, in vraagzinnen eenduidig het verschil aan kunnen geven tussen brede en nauwe focus.

14.00 E-learning resource 'Russian Dialectal Phonetics':

Perspectives and first output

Sergej Kniazev (Engelstalige bijdrage)

Russian Language Department, Philological Faculty, Moscow State University

The aim of the project supported by NWO is to create an electronic course of modern Russian dialectal phonetics suitable for learning, teaching and for scientific research. The course will comprise two components: a theoretical overview and a practical section (real texts and on-line tests, exercises, quizzes). The information provided to users is a multimedia one: original dialectal texts and their phonations, multimedia sensitive maps and photo- and video-fragments from the archive of dialectological expeditions with exercises in the practical section. All course units are constructed as SCORM-consistent reusable learning objects (RLOs). As an open-source product, the electronic course "Russian Dialectal Phonetics" can be enlarged and outspread to other levels of linguistic description of Russian dialects.

In my paper, after a short discussion of the multimedia course, I will present some results of our study on some Southern Russian dialects which are characterised by so called “dissimilative akan’je” (DA) - a specific type of vowel neutralization in pretonic syllables where the result of neutralization depends on the quality of the stressed vowel. We believe that originally DA showed compensatory dissimilation within the prosodic nucleus on the basis of vowel length: the longer (more open) the stressed vowel, the shorter (more reduced) the pretonic one. The qualitative and quantitative difference between long and short pretonic vowels [a] and [ə] in all so far known dialects with DA is striking, and there is an abrupt change from one vowel to another. While processing some texts with DA during our work on the project, previously supposed to show the so-called *Prokhor* or *Don* type of DA ([a] before high stressed vowels and [ə] before all non-high ones), we found a completely different situation: qualitative and quantitative differences of pretonic vowels depending on the stressed vowel are small, gradual and irrelevant. It can be suggested that this dialect reflects the most archaic type of DA with a vague dependence of the pretonic vowel on the length of the stressed vowel, which has not yet been fixed in the phonological system. This type of DA can be called *protodissimilative*.

14.20 Identifying Linguistic Structure in the Aggregate Analysis of Dialect Pronunciation

Jelena Prokic (Engelstalige bijdrage)

Humanities Computing, University of Groningen

Computational dialectometry is a multidisciplinary field that uses quantitative methods in order to measure linguistic differences between the dialects. The distances between the dialects are measured at different levels (phonetic, lexical, syntactic) by aggregating over the entire data set. These aggregate analyses do not expose the underlying linguistic structure, i.e. the specific linguistic elements that contributed to the differences between the dialects. The aim of the work presented here is to extract the linguistic structure from the aggregate analysis of Bulgarian dialects carried out on the basis of word pronunciations. The method is based on the identification of regular sound correspondences and is for the first time applied in dialectometry in order to extract linguistic structure from the aggregate analysis. All the analyses are based on the transcriptions of 117 words collected from 84 sites equally distributed over the Bulgarian dialect area. The results have shown that identification of regular sound correspondences can be successfully applied in the task of identifying linguistic structure in the aggregate analysis of dialects based on word pronunciation.

14.40 Vowel pronunciation in Swedish dialects

Therese Leinonen (Engelstalige bijdrage)

Humanities Computing, University of Groningen

At the University of Groningen computational methods are used to measure dialect distances and classify dialects. In my PhD-project the focus will be on pronunciation of vowels in Swedish dialects. The data comes from 105 Swedish sites with 12 speakers from each site. The speakers represent 4 groups: elderly women, elderly men, young women and young men. The data exist both as acoustic files and transcriptions, and both will be used for analyses. One aim is to analyse the transcriptions using different feature systems, since the feature systems used so far in Groningen have not led to an improve ment of the results compared to a simple phone representation. The relationship between acoustics and transcriptions will be analysed as well as the relationships between the 4 speaker groups. Cluster analysis and multidimensional scaling are used for classifying the dialects and results are shown on dialect maps.

15.30 JASMIN-CGN: Uitbreiding van het CGN met spraak van Jongeren, Anderstaligen en Senioren

*Olga van Herwijnen & Catia Cucchiarini
Radboud University, Nijmegen*

Large speech corpora constitute an indispensable resource for conducting research in speech processing and for developing real-life speech applications. In 2004 the Spoken Dutch Corpus (Corpus Gesproken Nederlands – CGN: a corpus of standard Dutch as spoken by adult natives in the Netherlands and Flanders) became available. Owing to budget constraints, CGN does not include speech of children, non-natives, elderly people and recordings of speech produced in human-machine interactions. Since such recordings would be extremely useful for conducting research and for developing HLT applications for these specific groups of speakers of Dutch, a project was started to extend CGN by collecting a corpus of contemporary Dutch as spoken by children of different age groups, non-natives with different mother tongues and elderly people in the Netherlands and Flanders (JASMIN-CGN). In addition, in this project speech material will be collected in a communication setting that was not envisaged in CGN: human-machine interaction. One third of the data will be collected in Flanders and two thirds in the Netherlands. In this talk I will discuss the rationale of the project, the corpus design, the speech material, the procedure and the use that can be made of the results of this project.

15.50 "Landgenooten..." Verbeterde toegang tot historische toespraken

*Willemijn Heeren & Laurens van der Werff
Human Media Interaction, Universiteit Twente*

Het CHoral project heeft tot doel om technologie te ontwikkelen ter verbetering van de toegang tot historische audiocollecties. Via spraakherkenning proberen we audio-archieven om te zetten in tekst, waardoor het mogelijk wordt om de audio te doorzoeken op woordniveau. Relevante fragmenten kunnen vervolgens direct worden beluisterd. Het automatisch verwerken van historische gesproken documenten is lastig, omdat de audiokwaliteit vaak slecht is, en de spreekstijl en het woordgebruik ouderwets zijn. Dit leidt tot een mismatch met bestaande akoestische modellen en taalmodellen, die zijn ontwikkeld op het nieuwsdomein. In deze presentatie willen we ons onderzoek introduceren aan de hand van experimenten op een specifieke, historische audiocollectie: de toespraken van Koningin Wilhelmina uitgezonden op Radio Oranje tijdens de Tweede Wereldoorlog. Deze collectie bestaat niet alleen uit audio; alle toespraken zijn tevens volledig uitgeschreven. Daarom was het mogelijk om door oplijning van de audio met de teksten tot een index op woordniveau te komen. In een demo zullen we laten zien dat - met behulp van deze index - online gezocht kan worden in de audio en de tekst als ondertiteling bij de audio kan worden gepresenteerd.

16.10 Automatische emotie-detectie in spraak in het project

MultimediaN: Multi-Modal Interaction

Khiet Truong

TNO Human Factors, Soesterberg

Binnen het project MultimediaN Multi-Modal Interaction doen we o.a. onderzoek naar het automatisch detecteren van emotie in spraak. De laatste jaren wordt er steeds meer gewerkt aan het automatisch detecteren van emotie in spraak, ofwel gezicht, ofwel gebaren en daarbij komt men vaak dezelfde knelpunten tegen. Dit is een van de redenen waarom wij zijn begonnen met het detecteren van gelach. In deze presentatie zal ik iets meer vertellen over de moeilijkheden die gepaard gaan met dit soort emotie onderzoek; hoe komen we aan spontane emotionele spraak data, hoe gaan we emotie annoteren etc. Verder is het van belang dat de classificatie/detectie experimenten goed opgezet worden (classificatie vs. detectie) en dat de ontwikkelde emotie detectoren ook goed geëvalueerd kunnen worden. Ook zal ik ingaan op een aantal emotie detectie experimenten die we hebben uitgevoerd op een database en wat we van plan zijn om te gaan doen binnen het project.

