De dag van de *Fonetiek* 2007

Over onderzoek naar spraak en spraaktechnologie

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

Donderdag 20 december 2007 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 R.J.J.H.vanSon@uva.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:

Rob van Son, secretaris NVFW
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/

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

16:10 De perceptie van modale partikels in het Nederlands als tweede taal

Johanneke Caspers & Ton van der Wouden

Het Nederlands kent tal van modale partikels, zoals *zeker* in *Jij bent zeker Jan?* en *maar* in *Kom maar hier*. Deze ongeaccentueerde woorden geven de hele zin een bepaalde kleuring, die lastig te omschrijven is. De voorspelling dat verwerving van deze woorden moeilijk moet zijn voor tweedetaalsprekers (Foolen 1986, Wenzel 2002) werd getoetst in een perceptieproef met de modale partikels *toch*, *wel* en *zeker* en hun geaccentueerde niet-modale tegenhangers (i.e., 'gewone' bijwoorden). NT2-sprekers en NT1-sprekers kregen telkens een contextzin aangeboden (bv. *Emma wil haar fiets verkopen*) en moesten daarna aangeven wat ze de best passende vervolgzin vonden: met een accent op het doelwoord (*Ze fietst TOCH nooit*) of met een accent elders in de zin (*Ze FIETST toch nooit*). De resultaten laten duidelijk zien dat NT2-sprekers inderdaad grote moeite hebben met modale partikels: bij de modale contexten scoren NT2-sprekers slechts 51% correct, d.w.z. conform de voorspelling, tegenover 94% voor NT1-sprekers.

16:30 SpeakGoodChinese - De tonen van het Mandarijnchinees leren spreken

David Weenink & Rob van Son
Amsterdam Centre for Language and Communication, University of Amsterdam

Nederlandse studenten, en vele anderen, hebben grote moeite om de tonen van het Mandarijnchinees te leren verstaan en spreken. In de praktijk kunnen beginnende studenten niet zelfstandig oefenen wat de vooruitgang ernstig vertraagt. In een samenwerkingsproject tussen de Hogescholen van Rotterdam en Amsterdam, Fontys en de universiteiten van Amsterdam en Twente is op basis van *PRAAT* een applicatie ontwikkeld, SpeakGoodChinese (<http://speakgoodchinese.org/>), die de uitspraak van de tonen herkent waarmee studenten individueel kunnen oefenen. De student spreekt een in het *pinyin* gespeld woord uit. Daarna geeft de applicatie feed-back over de uitspraak van de tonen. SpeakGoodChinese kan ook synthetische referentietonen genereren vanuit de *pinyin* notatie.

Op de spraak van referentiesprekers maakte de SpeakGoodChinese herkenner 6% fouten op voorgelezen woorden en minder dan 15% op spraak verzameld met een functionele testvariant met vrije woordkeuze (vals negatief). Op nagesproken, geschaduwde, goede en foute tonen van 8 sprekers, zowel goede als slechte, maakte de herkenner minder dan 15% fouten in beide richtingen, vals positief en vals negatief. De volledige applicatie is beschikbaar voor MS Windows XP en Linux onder de GNU GPL-licentie.

Programma

9:20 FUSS - Unit-selectie met een (relatief) klein corpus

Arthur Dirksen
Fluency

FUSS is een nieuwe unit-selectie synthesizer, die deel uitmaakt van een poging om met eenvoudige middelen, en zo veel mogelijk geautomatiseerd, opnames te maken van een spreker, en die om te zetten in een stem-voor-spraaksynthese.

In mijn bijdrage zal ik aandacht besteden aan de volgende onderwerpen: constructie van het corpus, opnameprocedure, automatische labeling, de architectuur van de synthesizer en de integratie in de tekst-naar-spraaksoftware van Fluency.

De eerste stem die voor FUSS gemaakt is, wordt gedemonstreerd in twee versies: zowel vóór als ná handmatige correctie van de automatische labeling.

9:40 Emotionele spraaksynthese

Melanie Kroes
TNO DenV , Soesterberg

In dit onderzoek heb ik drie methodes om synthetische spraak emotioneel te laten klinken met elkaar vergeleken. Twee methodes passen de grondfrequentie en de klankduren aan. Voor deze twee methodes gebruiken we hetzelfde difoonsynthesesysteem om de initiële intonatie en klankduren te bepalen. De derde methode kopieert de intonatie en klankduren uit natuurlijke emotionele spraak. De spraak is geselecteerd uit een Engelstalige emotionele database en de transcripties van de fragmenten zijn ook gebruikt voor synthese met de andere methodes. Om de spraak te genereren is voor alle methodes dezelfde Engelse MBrola-stem gebruikt. In een luisterexperiment is aan 20 proefpersonen gevraagd om van ieder fragment aan te geven wat de 'arousal' (activatie) en 'valence' (waardering) is en om het emotielabel (afraid, angry, happy, neutral, relaxed en sad) te kiezen dat het beste past bij het fragment. Ik zal de resultaten van dit experiment presenteren

10:00 “Hoe kan ik u van dienst zijn?” – Spraakgestuurde routeringsapplicaties

Diana Binnenpoorte, Christophe Van Bael, Johan de Veth
LogicaCMG Nederland, Nieuwegein

Het doel van routeringsapplicaties in callcenters is om bellende klanten door te verbinden met de meest geschikte dienst of medewerker. Routeringsapplicaties kunnen op verschillende manieren geïmplementeerd worden, zowel toetsgestuurd als spraakgestuurd. Bij toetsgestuurde applicaties moeten klanten zelf bepalen met welke keuzes in een gelaagd en dus vaak ook ingewikkeld keuzemenu hun vraag het beste beantwoord kan worden. Spraakgestuurde routeringsapplicaties vormen wat dat betreft een veel klantvriendelijker alternatief. Bij spraakgestuurde applicaties hoeven klanten niet langer een keuzemenu te doorlopen; ze moeten enkel nog antwoord geven op de open beginvraag:”Hoe kan ik u van dienst zijn?”. Spraakgestuurde routeringsapplicaties herkennen vervolgens de ingesproken zin en beslissen op basis van deze gegevens automatisch bij welke dienst of medewerker de klant het best kan worden verder geholpen.

15:30 Articulatory settings in spraakproductie

Sybrine Bultena & Wander Lowie
Rijksuniversiteit Groningen

Er wordt verondersteld dat de combinatie van de stand van articulatoren die gebruikt worden voor het spreken (met name de tong, kaken en lippen), per taal verschillend is; dit fenomeen is bekend als ‘articulatory settings’. Eerdere studies over dit onderwerp hebben gebruik gemaakt van technieken variërend van analytisch luisteren tot moderne scantechnieken; geen van de tot nu toe uitgevoerde studies heeft echter eenduidig kunnen aantonen dat taalspecifieke settings tijdens spraakproductie meetbaar zijn. Met deze studie proberen we verschillen tussen de Nederlandse en Engelse setting akoestisch te meten onder optimale omstandigheden: op basis van metingen van vergelijkbare klinkerparen binnen sprekers. Hiervoor zijn de formantfrequenties gebruikt van acht verschillende Nederlands-Engelse klinkerparen voorkomend in interlinguale homofonen, uitgesproken door vijf gevorderde Nederlandse leerders van het Engels als tweede taal. Statistische analyses van deze akoestische data laten zien dat er significant verschillende globale patronen voorkomen in de Engelse en Nederlandse data, die verklaard kunnen worden door de taalspecifieke settings van deze twee talen. Deze uitkomsten laten bovenal zien hoe dynamisch het articulatieproces is, wat gezien kan worden als een verklaring voor de moeilijkheden die voorgaande studies hebben ondervonden.

15:50 Automatische meting van spreeksnelheid in gesproken Nederlands

Nivja de Jong & Ton Wempe
Amsterdam Centre for Language and Communication, University of Amsterdam

In het kader van een grootschalig project aan de UvA (What is Speaking Proficiency, <http://www.hum.uva.nl/wisp>) hebben wij een methode ontwikkeld om objectief te meten hoe snel mensen spreken, hoeveel pauzes mensen laten vallen en hoe lang de pauzes duren. In deze voordracht gaan wij in op de methode om spreeksnelheid objectief te meten: een programma, geschreven in *PRAAT*, detecteert syllabes met behulp van informatie over intensiteit (dB) en stemhebbendheid van het spraaksignaal. Met dit programma is het mogelijk om spreeksnelheid te schatten zonder voorbewerking van het spraaksignaal, en zonder te hoeven transcriberen. Het programma is gevalideerd op twee verschillende corpora van gesproken Nederlands. Voor zover het mogelijk is het succes van het programma te vergelijken met bestaande methodes om spreeksnelheid te schatten, functioneert het programma goed en is het makkelijk te gebruiken omdat geen enkele voorbewerking van de spraakbestanden nodig is.

14:20 It's all in your head - How to get abstract representations in there

Diana Apoussidou
Amsterdam Centre for Language and Communication, University of Amsterdam

How do we get the phonological representation of a word into our heads from the speech signal? If we assume that the mental lexicon contains something like underlying representations of words, we have to account for how children acquire them in the course of language learning. For instance, how can children learn that the word for 'rat' in Dutch, pronounced [rat], ends in a final voiceless stop, while the word for 'wheel', also pronounced [rat], actually ends in a final voiced stop, as the plural form [radə] shows? In this talk, it will be demonstrated how computer-simulated learners can acquire both the phonological grammar causing the final devoicing effect plus the correct underlying forms from the phonetic forms and their meanings (e.g. from pairs such as [rat] – 'rat' or [radə] – 'wheels') by combining a parallel learning procedure of the different levels of representation with subsequent serial production of the words.

14:40 Competitieprocessen tijdens het begrijpen van spontane spraak

Susanne Brouwer[1], Holger Mitterer[1], & Mirjam Ernestus[21]
[1]Max Planck Institute for Psycholinguistics, Nijmegen
[2]Radboud University, Nijmegen

In spontane spraak worden woorden vaak niet volledig uitgesproken. Zo kan '*oktober*' bijvoorbeeld uitgesproken worden als '*tower*'. Wij onderzochten hoe luisteraars gereduceerde woorden herkennen door het meten van hun oogbewegingen terwijl ze keken naar 4 gedrukte woorden - het doelwoord (bv. *oktober*), een concurrent die fonologisch op het ongereduceerde woord lijkt (bv. *octopus*), een die op het gereduceerde woord lijkt (bv. *toveren*), en een ongerelateerde distractor - terwijl ze naar spontane zinnen luisteren met gereduceerde en ongereduceerde woorden. Proefpersonen keken meer naar de concurrenten dan naar de distractor, zonder verschil tussen de twee concurrenten. We voerden twee experimenten uit om te onderzoeken wanneer de concurrenten *wel* van elkaar verschillen. Bij presentatie van alleen ongereduceerde vormen in zorgvuldige spraak keken proefpersonen meer naar de ongereduceerde dan naar de gereduceerde concurrenten. Met de ongereduceerde woorden uit het eerste experiment, maar nu niet gemixt met gereduceerde woorden, bleek er naar beide concurrenten niet gekeken te worden. De resultaten laten zien dat lexicale competitie afhankelijk is van de mate van reductie.

Vandaag presenteren we verschillende aspecten die te maken hebben met de bouw van spraakgestuurde routeringsapplicaties. We zullen spreken over dialoogdesign, de open spraakherkenning (OSR) en de training en tuning van de semantische component die herkende zinnen classificeert om zo de meest geschikte dienst of medewerker te vinden.

10:20 Een computationeel model voor taalverwerving - Woorddetectie op basis van multimodale input

Louis ten Bosch[1], Lou Boves[1] & Hugo Van hamme[2]
[1]Radboud University, Nijmegen
[2]ESAT, KU Leuven

Taalverwerving bij baby's en jonge kinderen is een interessant proces, want baby's beginnen zonder woordenschat en in gesproken taal zijn woordgrenzen als zodanig niet hoorbaar. Toch zijn baby's en jonge kinderen heel goed in staat auditieve (multimodale) stimuli te gebruiken om woorden en betekenissen te leren die hen in staat stellen met de omgeving te kunnen communiceren. In deze presentatie laten we een rekenmodel zien dat dit woordleerproces simuleert. Het leerdermodel is in staat zonder voorafgaand gedefinieerd lexicon woorden (en woordachtige eenheden) te leren uit 'ruwe' multimodale stimuli die in een dialoog worden aangeboden door de 'verzorger'.

Het leerder-model bestaat uit 4 ingrediënten: een waarnemingsmodule ('zintuig'), een geheugen, een drijfveer om te leren, en een module die de communicatie met de 'verzorger' regelt.

In de presentatie bespreken we de resultaten van de leerder als functie van een aantal parameters (zoals hoeveelheid trainingstokens, sprekerafhankelijkheid, leren versus vergeten), voor drie talen (Nederlands, Fins en Zweeds). Resultaten worden gerelateerd aan wat bekend is uit de taalverwervingsliteratuur.

11:10 De oplijning van het begin van de eindstijging in Nederlandse "falling-rising" intonatiecontouren

Marco van de Ven & Carlos Gussenhoven
Radboud Universiteit Nijmegen

In hoeverre wordt de oplijning van het begin van de intonationele eindstijging beïnvloed door beklemtoonde lettergrepen in postnucleaire woorden of door Second Occurrence Focus (SOF)? Om deze vragen te beantwoorden is een corpus met Nederlandse "falling-rising" intonatiecontouren met een vroeg nucleair accent opgenomen met negen sprekers. De resultaten laten zien dat noch een postnucleaire beklemtoonde lettergreep, noch SOF de positie van het begin van de eindstijging beïnvloedt; dit punt heeft een vaste afstand tot het zinseinde. Dit suggereert dat de positie van postnucleaire tonen kan worden bepaald door (1) fonologische associatie met een postnucleaire beklemtoonde lettergreep, zoals elders aangetoond voor het Atheens-Grieks en het Roermonds, of (2) oplijning met het zinseinde of andere tonen, en dat een graduele aantrekkingskracht door klemtoon niet voorkomt.

11:30 Prominentie van onbeklemtoonde lettergrepen in Noord-Russische dialecten

Margje Post
Universitetet i Tromsø

In Noord-Russische dialecten wordt de eerste lettergreep van prosodische woorden vaak zo prominent uitgesproken dat het moeilijk is om te horen waar de klemtoon ligt. De prominentie lijkt vooral veroorzaakt te worden door de toonhoogtebeweging, maar mogelijk spelen relatieve lengte en luidheid toch een grotere rol

Dit verschijnsel is mogelijk een gevolg van taalcontact met de omringende Fins-Oegrische talen met vaste klemtoon op de eerste lettergreep, zoals Samisch en Karelisch. In tegenstelling tot deze talen hebben de Noord-Russische dialecten geen obligatorisch beginaccent, maar wordt het accent alleen gesignaleerd bij een specifieke prosodische structuur van de uiting, wat laat zien dat de klemtoon in de betreffende woorden niet verschoven is naar de eerste lettergreep.

Frequente nadruk van de eerste lettergreep komt ook voor in het Russisch van de Samen en Komi uit hetzelfde gebied, maar deze lijkt een lichtelijk andere vorm te hebben.

11:50 Waargenomen spreektempo en articulatorische inspanning

Hugo Quené
Utrechts instituut voor Linguïstiek OTS, Universiteit Utrecht

Oordelen van luisteraars over spreektempo zijn niet alleen gebaseerd op de gehoorde spraakklanken, maar ook op het aantal *bedoelde* (en wellicht onhoorbare) spraakklanken. Dat roept de vraag op, of het subjectieve spreektempo mede beïnvloed wordt door de (door luisteraars geschatte) articulatorische inspanning die nodig is om de spraak te produceren. De klankreeks "tisitisiti" gesproken met 5 syll/s zou dan beoordeeld worden als *trager* dan de klankreeks "sokesokeso" gesproken met hetzelfde tempo, vanwege de geringere articulatorische inspanning voor de eerste klankreeks. Deze hypothese is onderzocht in een "magnitude estimation" experiment, met klankreeksen die relatief weinig of veel articulatorische inspanning vereisen, in tempi tussen 3.0 en 4.5 syll/s.

De voorlopige resultaten bevestigen bovengenoemde hypothese, althans voor het snelste tempo. Luisteraars compenseren blijkbaar voor de articulatorische inspanning (vergelijkbaar met compensatie voor coarticulatie), indien het gehoorde spreektempo mogelijk begrensd is door articulatorische beperkingen. Deze resultaten bevestigen het algemene idee dat spraakperceptie mede wordt beïnvloed door impliciete kennis die luisteraars hebben over spraakproductie.

13:40 Van *arm* tot *zwerven* - Sjwa-insertie in het Standaardnederlands van Vlamingen en Nederlanders

Hanne Kloots
Universiteit Antwerpen – Centrum voor Nederlandse Taal en Spraak

In deze bijdrage bestuderen we sjwa-insertie in het spontaan gesproken Standaardnederlands van 80 Vlaamse en 80 Nederlandse leraren Nederlands. Bij de samenstelling van de steekproef werd rekening gehouden met de variabelen *leeftijd*, *sekse* en *regio*. Er wordt gefocust op woorden van het type *arm* en *zwerven*, d.w.z. woorden waarin de (enige) volle klinker gevolgd wordt door een consonantencluster met /r/ als eerste element en een niet-homorganische medeklinker als tweede element. Voor elk woord is nagegaan of in het consonantencluster een sjwa werd ingelast, bv. *arm* > *arrem*, *zwerven* > *zwerreven*. Sjwa-insertie bleek vaker voor te komen in Vlaanderen dan in Nederland. Bij de generatie geboren voor 1955 is het verschijnsel iets prominenter aanwezig dan bij de generatie geboren na 1960, zeker in Nederland. Ten slotte bleken er ook opvallende verschillen te bestaan tussen de respectieve regio's. De tendens om sjwa's in te voegen is het sterkst in de regio Antwerpen/Vlaams-Brabant.

14:00 Lexical-stress information rapidly modulates spoken-word recognition

Eva Reinisch, Alexandra Jesse, James M. McQueen
Max Planck Institute for Psycholinguistics

The time-course of the effect of suprasegmental stress information on word recognition was investigated by tracking Dutch listeners' looks to arrays of four printed words on a computer screen as they listened to spoken sentences. Target trials included word pairs that did not differ segmentally in their first two syllables but differed in their stress placement (e.g., 'BArometer' and 'baroNES'; capitals marking stressed syllables). The listeners' eye-movements showed that they used stress information to disambiguate rapidly between word candidates. For example, when hearing 'baroNES', participants looked more at 'baroNES' than at its competitor 'BArometer' even before segmental information could disambiguate the words. Furthermore, there was an asymmetry in the amount of competition. Words with stress on the first syllable provided stronger competition than words with non-initial stress. Lexical stress information thus affects the degree to which words compete, and it is used immediately to modulate the recognition process.