Doktorandentag

1. Dezember 2011

Internationales Begegnungszentrum der Wissenschaft

Amalienstr. 38, 80799 München

09:00 – 09:15	**Begrüßung**
09:15 – 10:45	Christian Heinrich, LMU: *Untersuchungen zur Rhythmik von alkoholisierter und nüchterner Sprache* Barbara Baumeister, LMU: *Veränderungen der Grundfrequenz in alkoholisierter Sprache*
10:45 – 11:00	**Kaffeepause**
11:00 – 13:15	Theresa Schölderle, LMU: *Dysarthrie bei infantiler Cerebralparese (ICP)* Nele Salveste, LMU: *Focus and its prosodic means* Mareike Plüschke, LMU: *Peak Alignment in falling and rising accents in Estonian*
13:15 – 14:30	**Gemeinsames Mittagsessen**
14:30 – 15:15	**Postersession** Nicole Weidinger, LMU: *Pantomimische Darstellung des Objektgebrauchs bei Vor- und Grundschulkindern* Hanna Feiser, LMU: *Stimmähnlichkeit bei Familie*n Clara Tillmanns, LMU: *Phonetische/phonologische Sensibilität von spontaner phonetischer Imitation* Zixing Zhang, TUM: *Easing Data Sparseness in Acoustic Emotion Recognition* Veronika Neumeyer, LMU: *Akustische Analyse der Vokal-Artikulation von Cochlear Implantat-Trägern* Michele Nicoletti, TUM: *Model-based validation framework for coding strategies in cochlear implants*
15:15 – 15:30	**Kaffeepause**
15:30 – 17:00	Michele Nicoletti, TUM: *Model-based validation framework for coding strategies in cochlear implants* Felix Weninger, TUM: *Multi-Source Speech Recognition by Non-Negative Matrix Factorization*
17:00 – 17:15	**Schlusswort**

Abstracts

Barbara Baumeister, Institut für Phonetik und Sprachverarbeitung, LMU, bba@phonetik.uni-muenchen.de

Veränderungen der Grundfrequenz in alkoholisierter Sprache

Basierend auf dem Alcohol Language Corpus (ALC) werden Untersuchungen zur Veränderung der Grundfrequenz in alkoholisierter Sprache vorgestellt. Das Korpus beinhaltet Sprachaufnahmen von 162 Sprechern unterschiedlichen Alters in alkohlisiertem und nüchternem Zustand in den drei Sprachstilen read, spontaneous und command & control speech. Eine Langzeitanalyse des F0-Median und des Interquartilsabstands zeigt bei der Mehrzahl der Versuchspersonen einen Anstieg der Tonhöhe vom nüchternen zum alkoholisierten Zustand, sowie eine größere Variabilität des Stimmtons. Jedoch lässt sich bei einigen Sprechern auch ein gegenteiliger Trend beobachten. Der Grad der Alkoholisierung scheint hier allerdings keine Rolle zu spielen. Die Ergebnisse deuten darauf hin, dass zumindest anhand dieser Parameter keine allgemeine Aussage über das Verhalten der Sprecher unter Alkoholeinfluss zu treffen ist, sprecherspezifische Unterschiede zwischen alkoholisierter und nüchterner Sprache jedoch deutlich sind.

*****

Hanna Feiser, Institut für Phonetik und Sprachverarbeitung, LMU, feiser@phonetik.uni-muenchen.de

Stimmähnlichkeit bei Familien

Es wird der Frage nachgegangen, warum Stimmen verwandter Sprecher häufig verwechselt werden. Dies kommt gerade in der forensischen Phonetik häufig vor, wenn eine Aufzeichnung eines männlichen Sprechers vorliegt, die auch beispielsweise von seinem Bruder stammen könnte. Anhand der in diesem Bereich etablierten auditiv-akustischen Methode soll untersucht werden, ob die auditive Ähnlichkeit sich anhand von verschiedenen akustischen Parametern – wie beispielsweise die mittlere Grundfrequenz – erklären lässt. Anschließend wird mittels eines Perzeptionstests die auditive Ähnlichkeit getestet und versucht Muster zu finden, die zwei Stimmen als ähnlich oder nicht ähnlich klassifizieren.

*****

Christian Heinrich, Institut für Phonetik und Sprachverarbeitung, LMU, heinrich@phonetik.uni-muenchen.de

Untersuchungen zur Rhythmik von alkoholisierter und nüchterner Sprache

Der Alcohol Language Corpus (ALC) beinhaltet alkoholisierte und nüchterne Sprachaufnahmen von 162 Sprechern und deckt dabei die unterschiedlichen Sprachstile gelesen, spontan und command & control ab. Diese Datenbasis wurde als Grundlage zur Untersuchung der Rhythmik gesprochener Sprache herangezogen. Dabei werden auf phonetischer Segmentierung basierende Dauermuster, verschiedene Parameter aus dem Energieverlauf von Äußerungen und Konturen der Grundfrequenz und dem Energieverlauf analysiert. Grundsätzlich scheint alkoholisierte Sprache unregelmäßiger zu sein als nüchterne Sprache, jedoch gibt es große sprecherspezifische Unterschiede. Eine allgemeingültige Aussage ist daher bis jetzt nicht möglich.

*****

Veronika Neumeyer, Institut für Phonetik und Sprachverarbeitung, LMU, veronika.neumeyer@phonetik.uni-muenchen.de

Akustische Analyse der Vokal-Artikulation von Cochlear Implantat-Trägern

Inwiefern weicht die Artikulation von Vokalen bei gehörlosen Sprechern mit Cochlear Implantat (CI) von der Artikulation normal Hörender ab?

Diese Frage ist Grundlage einer Studie, in der 7 deutsche Langvokale /a:, e:, i:, o:, u:, 2:, y:/, gesprochen von CI-Trägern und normal Hörenden, untersucht wurden. Da CI-Träger in der Regel keine homogene Sprechergruppe darstellen, wurden sie je nach Zeitpunkt der Ertaubung und dem Zeitraum zwischen Ertaubung und CI-Versorgung in vier Gruppen eingeteilt. Dazu wurden vier in Alter und Geschlecht passende Kontrollgruppen aufgenommen. Zusätzlich wurden die vier CI-Gruppen untereinander verglichen. Untersuchungsgegenstand dieser Studie sind neben der Fläche des gesamten Vokalraums Distanzen zwischen einzelnen Vokalen, die je nach ihrer Lage im Vokalraum den Artikulationsparametern Zungenhöhe, Zungenlage und Lippenrundung zugeordnet werden können.

*****

Michele Nicoletti, Bioanaloge Informationsverarbeitung, TU München, michele.nicoletti@tum.de

Model-based validation framework for coding strategies in cochlear implants

Although modern cochlear implants (CI) are able to restore speech perception to a high degree, there is still a large potential for improvements e.g. in music perception and speech discrimination in noise.
To evaluate and optimize novel coding strategies, we have developed a toolbox which codes sound signals into spike-trains of the auditory nerve. We have previously developed a model of the intact inner ear, which we now have complemented with detailed models of a CI speech processor, the channel crosstalk and spiral ganglion neuron models. Here we use a model of spiral ganglion type I neurons with Hodgkin-Huxley type ion channels, which are also found in cochlear nucleus neurons. Their large time constants might be responsible to explain adaptation to electrical stimulation (Negem & Bruce 2008). We corrected conductances and time-constants to a body temperature of 37° and solved the differential equations in the time domain with a exponential Euler rule. Depending on the task, we model the neurons at different levels of detail. The electrode was modeled as an array of 12 current point sources at a distance of 0.5 mm from the spiral ganglion neurons. The coupling between electrode and excitation of the neuron was described by the activation function (second derivative of the extracellular potential with respect to the neuron's path). Cannel cross-talk was implemented by a convolution of the activation function with a spread function (symmetric, slope: 1dB/mm).
With our toolbox we present qualitative comparisons of neurograms elicited by different coding strategies with the situation in the healthy inner ear. Moreover, we conducted qualitative evaluations using two methods: With the framework of automatic speech recognition we evaluated speech discrimination using a noisy database. With the methods of information theory we are able to quantify the transmitted information coded in neuronal spike trains, which allows us to evaluate especially well how well temporal information is coded.
The major advantage of our approach is that we are able to evaluate both spectral- and temporal aspects of novel coding strategies before we conduct long-lasting clinical studies.

*****

Mareike Plüschke
Institut für Phonetik und Sprachverarbeitung, LMU, plueschke@phonetik.uni-muenchen.de

Peak Alignment in falling and rising accents in Estonian

Estonian is a quantity language with a three-way quantity contrast for both vowels and consonants. In the talk the influence of quantity on the peak alignment of falling H*+L and rising L*+H accents will be discussed comparing the peak alignment of vowel and consonant quantity with each other. Furthermore results about the influence of the number of post-focal unstressed syllables on the high target of H*+L and on the low target of L*+H accents will be presented.

*****

Nele Salveste, Institut für Phonetik und Sprachverarbeitung, LMU, nele.salveste@gmail.com

Focus and its prosodic means

Focus can be defined either through contrast (Rooth 1992) or through givenness (Schwarzchild 1990). Focusing of a constituent is motivated through its semantic and pragmatic properties in a discourse. On the phonetic level the focused constituent bears a prosodic marking (pitch accent). Though, prosodic focus marking of a constituent involves probably more than just being contrasted or discourse-new (Selkirk 2007, Baumann (to appear)). Despite the varying theoretic claims about the nature of the focus in the literature, it is generally accepted that English (Pierrehumbert and Hirschberg 1990, Breen et al. 2010), German (Kohler 1991, Baumann 2008) and Dutch (Swerts et al. 2002) use prosodic means for marking information status in a spoken discourse, whereas Italian has strong syntactic and morphological properties for encoding information status beside the prosody (Swerts et al. 2002). Other morphologically more complex languages like Estonian and Finnish are considered not to use prosody for the pragmatic and semantic purposes (Wiik 1991). However, the results delivered by the Vainio and Järvikivi (2006, 2007) suggest that there are some contexts, where also morphologically complex language makes use of prosodic focus marking for transmitting the pragmatics of a sentence.

Baumann, Stefan (2008). Degrees of Givenness and their Prosodic Marking. In: C. M. Riehl, A. Rothe (eds.): *Was ist linguistische Evidenz? Kolloquium des Zentrums „Sprachvielfalt und Mehrsprachigkeit“*, November 2006. Aachen: Shaker (= ZSM Studien 2), 35-55.

Baumann, Stefan and Arndt Riester (to appear). Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects. In: Elordieta, Gorka and Pilar Prieto (eds.): Prosody and Meaning. Berlin, New York: Mouton De Gruyter. Interface Explorations 25.

Breen et al. (2010) = Mara Breen, Evelina Fedorenko, Michael Wagner, Edward Gibson (2010). Acoustic correlates of information structure. Language and Cognitive Processes, Vol. 25 (7/8/9), pp. 1044–1098.

Kohler, Klaus (1991). Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics, In: J. Harrington, Chr. Mooshammer, *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* (Vol 25), 115-185.

Rooth, Mats (1992). A theory of focus interpretation. Natural Language Semantics Vol. 1 pp 75–116.

Schwarzschild, Roger (1999). Givenness, Avoid F, and other constraints on the placement of accent. Natural Language Semantics Vol. 7, pp. 141–177.

Selkirk, Elisabeth (2007). Contrastive Focus, Givenness and the Unmarked Status of “Discourse-new”. In Féry, C., Fanselow, G. and Krifka, M. (Eds.), *Working Papers of the SFB632*, Interdisciplinary Studies on Information Structure (ISIS) Vol *6*, pp. 125–146. Potsdam: Universitätsverlag Potsdam.

Swerts et al. (2002) = Marc Swerts, Emiel Krahmer, Cinzia Avesani (2002). Prosodic marking of information status in Dutch and Italian: a comparative analysis. Journal of Phonetics, Vol. 30, pp. 629–654.

Vainio, Martti and Juhani Järvikivi (2006). Tonal features, intensity, and word order in the perception of prominence. Journal of Phonetics, Vol. 34, pp. 319–342.

Vainio, Martti and Juhani Järvikivi (2007). Focus in production: Tonal shape, intensity, and word order. Journal of the Acoustical Society of America, 121(2), EL 55–61.

Wiik, K. (1991). Foneetika alused. Translated and adapted by J. Valge. Tartu: Tartu Ülikool.

*****

Theresa Schölderle, Entwicklungsgruppe Klinische Neuropsychologie, LMU,
Theresa.Schoelderle@extern.lrz-muenchen.de

Dysarthrie bei infantiler Cerebralparese

Hintergrund: Menschen mit infantiler Cerebralparese (ICP) zeigen häufig Dysarthrien, die jedoch bislang nur selten im Fokus empirischer Untersuchungen lagen (Pennington et al., 2005). Es ist nicht beschrieben,

ob die Einteilung der ICP-Subtypen nach Störungen der Körpermotorik sich in entsprechenden Dysarthriesyndromen widerspiegelt (Morgan & Liégeois, 2010).

Methoden: Die Stichprobe umfasst 22 Patienten. Alle Teilnehmer wurden in Kooperation mit dem *Integrationszentrum für Cerebralparesen (ICP) München* gewonnen. Es wurden funktions- und kommunikationsorientierte Parameter der Dysarthrie erhoben, die Analysen erfolgten auditiv und akustisch. Die Sprechstörung jedes Patienten wurde von drei erfahrenen Beurteilern entsprechend der Syndromeinteilung klassifiziert.
Ergebnisse und Diskussion: Patienten mit ICP zeigen schwere Dysarthrien. Vor allem artikulatorische und prosodische Merkmale üben deutliche Einflüsse auf kommunikationsrelevante Parameter aus. Es konnten klar differenzierbare Dysarthriesyndrome identifiziert werden. Jedoch spiegelt sich im Dysarthriesyndrom nur sehr bedingt der ICP-Typ wider, es ergaben sich deutliche Dissoziationen.
Literatur

- Morgan, A.T. & Liégeois, F. (2010): Re-Thinking Diagnostic Classification of the Dysarthrias: A Developmental Perspective. *Folia Phoniatrica et Logopaedica*, 62, 120-126.
- Pennington, L., Goldbart, J. & Marshall, J. (2005): Direct speech and language therapy for children with cerebral palsy: findings from a systematic review. *Developmental Medicine & Child Neurology*, 47, 57–63.

*****

Clara Tillmanns, Institut für Phonetik und Sprachverarbeitung, LMU, clara@phonetik.uni-muenchen.de

Phonetische/phonologische Sensibilität von spontaner phonetischer Imitation

Anhand manipulierter VOT-Dauer zeigt Nielsen (2011), dass spontane phonetische Imitation sensibel für phonologische Kategorien ist. Das dargestellte Experiment soll (1) die phonologische Kategoriensensibilität reproduzieren, (2) testen, ob phonotaktische Häufigkeiten diese Sensibilität weiter differenzieren und(3) ob die Imitation eines Parameters zu Kompensation durch einen anderen Parameter führt.

Es werden dafür mehrere Wörter-Gruppen mit medialen Vokal-Plosiv-Verbindungen sowie initialen Plosiven verwendet und systematisch manipuliert.

*****

Nicole Weidinger, Entwicklungsgruppe Klinische Neuropsychologie, LMU, Nicole.Weidinger@extern.lrz-muenchen.de

Pantomimische Darstellung des Objektgebrauchs bei Vor- und Grundschulkindern

In mehreren Studien wurde die Fähigkeit von Kindern zur Pantomime des Objektgebrauchs untersucht (Dick 2005, Njiokiktjien 2000, Boyatzis & Watson 1993). Bei dieser Aufgabe wird die pantomimische Darstellung des Gebrauchs einzelner Objekte geprüft (z.B. „Zeig mir, wie man sich mit einer Zahnbürste die Zähne putzt“). Dabei zeigen sich Altersunterschiede zwischen Vor- und Grundschulkindern. Kinder bis etwas 6 Jahre ersetzen den Gegenstand durch ein Körperteil und produzieren einen „Body-Part-as-Object“ (BPO). Zum Beispiel streichen sie mit dem Zeigefinger über die Zähne, anstatt den Griff der Zahnbürste zu zeigen. Der Anteil von Pantomimen mit richtiger Handstellung nimmt im Grundschulalter zu und erreicht mit 9 Jahren das Niveau von Erwachsenen (Njiokiktjien 2000). Die Schwierigkeit, sich von den „BPO“ Darstellungen zu lösen ist insofern bemerkenswert, da sie die gleichzeitige Darstellung des Objektes und seines Gebrauchs betrifft. Die Konfiguration des Handgriffs zeigt die Form des Objektes und die Bewegung der ganzen Hand seinen Gebrauch. In meiner Arbeit wird geprüft, ob der BPO bei Vorschulkindern durch die Schwierigkeit der Koordination des Objekts und seines Gebrauchs hervorgerufen wird.

*****

Felix Weninger, Mensch-Maschine-Kommunikation, TU München, weninger@tum.de

Multi-Source Speech Recognition by Non-Negative Matrix Factorization

Automatic speech recognition (ASR) over distant microphones enables natural human-machine communication, but is a challenging task in the presence of interfering sources as often encountered in domestic or car environments. Numerous approaches have been proposed to improve recognition rates, including front-end (attempting to reconstruct the clean speech signal) and back-end based robust ASR (learning to cope with noisy signals). This talk describes a unified algorithmic framework for both front-end and back-end based robust ASR using non-negative matrix factorization (NMF). NMF describes the spectrogram of a noisy speech signal as the product of speech and noise patterns with non-negative activations. This enables the reconstruction of a clean speech signal as well as joint decoding of the speech and noise components. Recent results show that NMF-based front-ends and back-ends can drastically improve the recognition rates in challenging acoustic environments compared to conventional Gaussian mixture modeling.

*****

Zixing Zhang, Mensch-Maschine-Kommunikation, TU München, zix@mmk.ei.tum.de

Easing Data Sparseness in Acoustic Emotion Recognition

Obtaining large amounts of realistic data for speech emotion recognition is currently considered one of the most important issues in the field. We present a large-scale study on unsupervised and supervised learning from multiple databases in cross-corpus acoustic emotion recognition. For evaluation, we use six common databases, including acted and natural emotional speech. Regarding supervised learning, data agglomeration and decision-level fusion are investigated in a cross-corpus scenario. Conversely, unsupervised learning is used to integrate large amounts of unlabeled data into labeled training sets through an iterative re-training process. Recent results show that employing data agglomeration or majority voting strategies can enhance recognition performance even in a challenging cross-corpus setting, and that adding unlabeled emotional speech to agglomerated multi-corpus training sets is a promising approach to remedy the issue of data scarcity.