SimPhon.Net

workshop 3

Prosody

July 2–5, 2017

Abstracts

SimPhon.Net is a network of close interdisciplinary collaboration between linguists and computer scientists. It addresses the challenge to model and simulate phonetic variability. Through experiments with computer simulations we can pose a variety of questions to unobservable or inseparable aspects of phonetic processes and phonological systems.

Team:

- Jun.-Prof. Dr.-Ing. Peter Birkholz (TU Dresden)
- Dr. phil. Daniel Duran (Universität Stuttgart)
- Dr. James Kirby (University of Edinburgh)
- Leonardo Lancia (Max Planck Institute for Evolutionary Anthropology)
- Dr. phil. Natalie Lewandowski (Universität Stuttgart)
- Prof. Dr. Bernd Möbius (Universität des Saarlandes)
- Dr. phil. Uwe Reichel (Hungarian Academy of Sciences)
- Elena Safronova (Berlin / Universitat de Barcelona)
- Dr. Ingmar Steiner (Universität des Saarlandes / DFKI)
- Dr. Fabian Tomaschek (Universität Tübingen)
- Prof. Dr. Petra Wagner (Universität Bielefeld)
- Andrew Wedel, PhD (University of Arizona)
- Dr. Laurence White (Plymouth University)
- Dr. phil. Frank Zimmerer (Universität des Saarlandes)

The focus of this workshop is on prosody and prosodic models. The workshop is organized by the members of SimPhon.Net, funded by *Deutsche Forschungsgemeinschaft (**DFG**)*.

Organizers:
Uwe Reichel, Katalin Mády, Daniel Duran and the members of SimPhon.Net.

Venue:
The workshop is hosted by SimPhon.Net at Mátraházai Akadémiai Üdülő
http://www.udulo.mta.hu/uduloink/matrahazai-akademiai-udulo/

This abstract booklet was set in LaTeX by Daniel Duran. © 2017 by the individual authors. Cover photo: “Kékes seen from the roadbend beneath Galya-tető” — By Susulyka (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC BY-SA 4.0-3.0-2.5-2.0-1.0 (http://creativecommons.org/licenses/by-sa/4.0-3.0-2.5-2.0-1.0)], via Wikimedia Commons

http://www.simphon.net/workshops.html

Abstracts

(in alphabetical order)

"KaMoso: Agent-based models of speaker–hearer interaction and sound change"

Daniel Duran

Universität Stuttgart

Some researchers propose to regard language as an emergent, self-organising system. Developing formal models under this hypothesis needs to take into account the dynamics of speaker–hearer interactions. Inspired by previous work from Wedel (e.g. 2006) we developed "KaMoso" – a multi-agent simulation framework to model processes of speaker–hearer interactions and sound change. KaMoso incorporates exemplar-theoretic principles of speech production and perception: Perceived speech items are stored in a phonetically rich memory (where they form episodic memory traces or exemplars). Linguistic categories emerge from collections of exemplars which form clusters of high density within the phonetic/linguistic feature space. Production of new speech items is based on the collection of stored exemplars. Thus, production and perception are linked within a closed feedback loop. KaMoso also incorporates principles from social impact theory, inspired by work from Nettle (1999): Some individuals within a population have a higher impact on the behaviour of others.

KaMoso was developed in order to examine the interplay between exemplar-theoretic speech production and perception and social factors like different types of interactions according to social status or different network topologies. I discuss the (potential) applications of the KaMoso framework.

Nettle, D. (1999). Using Social Impact Theory to simulate language change. Lingua, 108(2–3), 95–117.

Wedel, A. B. (2006). Exemplar models, evolution and language change. The Linguistic Review, 23, 247–274. http://doi.org/10.1515/TLR.2006.010

"Acoustic and Social Correlates of Perceived Voice Attractiveness"

Natalie Lewandowski & Daniel Duran

Universität Stuttgart

We investigate correlates of voice attractiveness in spontaneous mixed-gender dialogs, as perceived by raters in an independent perception study. Multiple voice samples of 20 speakers (ten of which were female) engaged in 29 female-male dialogs were extracted from the extended German Conversations (*GECO2*) database of spontaneous speech and rated on attractiveness by 20 test subjects (twelve female).

Several linear mixed models were fitted with the perception ratings of attractiveness as the dependent variable (again, only for mixed-gender speaker-rater constellations), and both acoustic and social data as predictors. In order to investigate potential acoustic correlates of voice attractiveness, a set of predictor variables, including various measures for pitch, intensity, shimmer, jitter, and HNR was extracted from the recordings. The social predictors used were the social attractiveness scores (e.g., for *likeability*, *competence*, or *success*) every speaker received from their dialog partners in the *GECO2* database.

The model including the social scores shows a significant effect of *success*, mediated by an interaction with *speaker gender*, and an effect for *cheerfulness*. The acoustic model demonstrates a significant influence of *pitch slope variation*, including and interaction with *speaker gender*, and an effect of *mean intensity* on perceived voice attractiveness.

"Testing attention in the GECO2 database"

Natalie Lewandowski
Universität Stuttgart

[abstract not available]

"The impact of syntax and pragmatics on the prosody of dialogue acts"

Katalin Mády & Uwe D. Reichel
Hungarian Academy of Sciences

Task-oriented spoken dialogues have several advantages: (1) Since speakers are involved in a non-linguistic task, they tend to concentrate less on the fact that they are being recorded, (2) various settings allow for the elicitation of repetitions of certain elements (words, names, etc.). (3) Since these tasks create a specific setting, intentions of speakers are more easy to control in terms of information structure, e.g. whether a certain element is given, new, contrastive etc.

In this talk, the dialogue structure coding scheme by Carletta et al. (1997) is used in order to test whether dialogue acts can be classified based on their prosodic features according to the CoPaSul tool (Reichel 2016). Dialogue acts (DA) are investigated based on their (1) sentence type such as interrogatives, declaratives etc., and (2) on their informational weight within the same sentence type, e.g. explaining new information vs. assuring that previous information is understood correctly in declaratives. The goal is to find out whether syntactic and pragmatic categories can be distinguished by different prosodic features. Utterances were taken from the Hungarian version of the Columbia Games Corpus.

Dialogue acts referring to different sentence types were mostly distinguished by local features such as pitch accent shape. DAs with higher amount of new information were characterised by global features such as higher overall energy, syllable rate and longer durations compared to DAs containing all-given or not game-relevant information.

This attempt shows that the dialogue acts suggested by Carletta et al. (1997) can be characterised by stylised prosodic parameters. While DAs that express grammatical categories such as various question types seem to be connected to different prosodic categories based on local features, DAs that belong to the same sentence type but carry different pragmatic meaning tend to be distinguished along global prosodic parameters. A mid-term goal is to predict DAs on the basis of automatic prosodic feature extraction.

"Entrainment profiles: Comparison by gender, role, and feature set"

Uwe Reichel & Štefan Beňuš
Research Institute for Linguistics, Hungarian Academy of Sciences & Constantine the Philosopher University, Nitra & II SAS, Bratislava

We examine speech accommodation in cooperative games for established and new prosodic features representing register, pitch accent shape, as well as rhythmic aspects of utterances. Entrainment profiles for multiple feature sets, gender-role combinations, and distance measures show (1) that a speaker's way of entrainment highly depends on her/his gender and role in the game: female describers entrain most while male describers entrain least. This might reflect different strategies in cooperative solution-oriented interactions, such as: to create a common ground vs. to strengthen the hierarchy.

(2) Shapes rather converge and show less local entrainment, whereas overall mean and maximum values rather show synergy. These findings can be accounted for by different degrees of vulnerability to linguistic variation and different amounts of non-linearities in the form-meaning mapping.

“A new flexible development paradigm in MaryTTS”

Ingmar Steiner & Sébastien Le Maguer

Universität des Saarlandes / DFKI

The latest development in text-to-speech synthesis introduced numerous methodologies. Unit selection, HMM-based and now DNN-based synthesis are the current state of the art. Furthermore, even the signal processing community proposed numerous techniques: STRAIGHT, WORLD for the vocoding part or Wavelet for the F0 modeling.

To do so we are currently redesigning MaryTTS to focus on the modularity. Details of this are described in Le Maguer & Steiner (ESSV 2017). Based on this new version, we have started to introduce a new voice building paradigm. We have applied this workflow for our participation in the 2017 Blizzard Challenge for speech synthesis, and will provide a behind-the-scenes look at this process, and discuss some of the issues encountered.

“Digital language typology”

Juraj Šimko

University of Helsinki

[abstract not available]

“Analyzing prosody with wavelets”

Invited Talk: *Martti Vainio*

University of Helsinki

[abstract not available]

“Multimodal prosody – some data and some thoughts”

Petra Wagner

Universität Bielefeld

[abstract not available]

“A Semi-Automatic Method for Discovering Prosodic Constructions”

Invited Talk: *Nigel Ward*

University of Texas at El Paso

Prosodic constructions are meaning-bearing temporal configurations of prosodic features. I describe a method for semi-automatically discovering prosodic constructions from large unlabeled corpora, in which constructions may occur superimposed, and may be present weakly or strongly. The key innovation is the application of Principal Component Analysis to a couple hundred contextual prosodic features, sampled over a half a million timepoints taken from dialog data. The resulting components are generally interpretable as prosodic patterns with meanings, as determined by examining factor loadings and listening to timepoints with high values on the component. I illustrate with four constructions of

English: the backchanneling construction, the minor-third construction for cuing action, the bookmarked narrow pitch construction for introducing a new consideration, and the particle-assisted floor change construction.

"Prosody and regional accent in dynamic judgements of trustworthiness"

Laurence White

Plymouth University

Spoken communication is predicated on relative trust between interlocutors, but our assessment of a speaker's trustworthiness is affected by both their behaviour towards us and their intrinsic and extrinsic vocal characteristics. Prosodic features such as pitch and articulation rate are known to affect immediate trust judgements, but the unfolding dynamics of such voice-based attributions have been little explored. We used an investment game to explore how prosody interacts, in the formation of trust judgements, with actual speaker behaviour and with the more intrinsic indexical cue of regional accent. Regression analyses show that speaker accent, mean pitch and articulation rate all influence participants' investment decisions, our implicit measure of trust, but interpretations of prosodic features interact with how the virtual player's behaviour unfolds, whilst their accent appears to exercise a fairly consistent influence on trustworthiness over time.