

SimPhon.Net

workshop 2

**Exemplar-theoretic models and alternative approaches in
usage-based models of language and speech
production/perception**

December 4–7, 2016

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 (Ludwig-Maximilians-Universität München)
- Elena Safronova (Berlin / Universität 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)

This workshop is aimed at computational approaches to the study on human speech segmentation and acquisition. In addition, contributions related to other topics of general interest to the research network are presented. The workshop is organized by the members of SimPhon.Net, funded by *Deutsche Forschungsgemeinschaft (DFG)*.

Organizers:

Ingmar Steiner, Daniel Duran and the members of SimPhon.Net.

Venue:

The workshop is hosted by SimPhon.Net at Schloss Dagstuhl:

Schloss Dagstuhl - Leibniz-Zentrum für Informatik
Oktavie Allee
66687 Wadern
Germany
<http://www.dagstuhl.de/16495>



Abstracts

(in alphabetical order)

“Exemplar theory and sound change”

James Kirby

University of Edinburgh

In this presentation I present a simple multivariate exemplar model that can be used to model sound change at the sub-phonemic level. After illustrating the model with a case study from Seoul Korean, I will discuss some extensions to the framework which appear necessary to handle cases of phonological change. I conclude that, at least in some cases, explicit sublexical representations may not be necessary to explain sound changes such as tonogenesis, despite the fact that such changes appear to involve the emergence of a new phonological contrast.

“Acoustic and 3D tongue motion statistical speech synthesis”

Sébastien Le Maguer

Saarland University

Nowadays, text-to-speech (TTS) synthesis is a widely used technology. As a step forward, some researches are also focusing on coupling the speech signal with another modality. The presented work focuses on the generation of a synchronized speech signal and 3D tongue motion. Like a classical speech synthesis system, this generation is achieved only from the text.

The presentation will include two parts. First, I will present and justify the architecture put in place to achieve this goal. Then we will focus on the evaluation and the analysis protocol. The objective of this part is not only to show that it is possible to achieve our goal but also to point the different problems about it.

“Information Density and the predictability of phonetic structure: preliminary results”

Bernd Möbius, Bistra Andreeva, Erika Brandt, Zofia Malisz, Yoonmi Oh⁺, Frank Zimmerer
Saarland University, ⁺ Université de Lyon

Variation in information density affects phonological and phonetic structure (Aylett and Turk 2004, 2006; Levy and Jaeger 2006; Jaeger 2010). Reduction and expansion of acoustic-phonetic features is also a characteristic of prosodic variability. In this study, we assess the impact of information density and prosodic structure on phonetic encoding, both independently of each other and in interaction. We model segmental duration and vocalic quality as a function of information density measures (e.g. surprisal), as well as syllable prominence, phrase boundary and speech rate manipulated experimentally. Correlates of phonetic encoding density are extracted from a subset of the BonnTempo corpus (Dellwo et al. 2004) for six European languages (Czech, English, Finnish, French, German, Polish). Surprisal is estimated from n-gram language models trained on text corpora.

“Hybrid exemplar models”

Janet B Pierrehumbert

Oxford e-Research Centre, University of Oxford

Exemplar theory was a response to findings that the mental representations of words and word phrases include a lot of phonetic and contextual detail. The initial computational models made the assumptions that 1) Words are associated directly with parametric phonetic representations and 2) Every experience with a word caused the memory representation to be updated. In this talk, I will present findings that require more complex assumptions.

First, I will present a repetition memory study that shows that words can be recognized but fail to create a memory trace if they are pronounced in a way that is atypical in relation to the contextually relevant standard. Then, I will continue to describe results showing the need for an intermediate level of representation — the phonological level. These are a dialect imitation study and an analysis of word frequency effects in the New Zealand vowel push chain. I conclude that a hybrid exemplar model is needed, combining much of the detail present in the first-generation models with the phonological parsing present in classical models of sound structure.

“The CoPaSul toolkit for intonation stylization”

Uwe Reichel

Hungarian Academy of Sciences

The purposes of the CoPaSul toolkit are automatic prosodic annotation and prosodic feature extraction from syllable to utterance level.

At the current state automatic annotation comprises:

- segmentation into interpausal chunks
- syllable nucleus extraction
- unsupervised localization of prosodic phrase boundaries and prominent syllables

F0 and partly also energy feature sets can be derived for:

- standard measurements (as median and IQR)

- register in terms of F0 level and range
- prosodic boundaries
- local contour shapes
- bottom-up derived contour classes
- Gestalt of accent groups in terms of their deviation from higher level prosodic units
- rhythmic aspects quantifying the relation between F0 and energy contours and prosodic event rates

All feature sets will be presented along with an application scenario. The toolkit is implemented in Python3 and will be made freely available on GitHub by beginning of 2017.

“The PaIntE model of intonation”

Antje Schweitzer

Universität Stuttgart

In this talk I will present the PaIntE model, which was originally developed for speech synthesis. I will argue in this talk that PaIntE is also well suited for intonation research beyond speech synthesis. To this end, I will show that PaIntE shares assumptions with autosegmental models of intonation. I will also discuss the parameters that PaIntE uses to describe intonation events, arguing that they represent several dimensions that contribute to the meaning of tonal events. Furthermore, PaIntE intonation events can be related to categories posited by autosegmental approaches to intonation, while they are also compatible with an exemplar-theoretic perspective.

“Exploring the similarity of intonation events”

Katrin Schweitzer

Universität Stuttgart

I will outline two lines of research that are concerned with the similarity of intonation events. One is a number of studies which explored how the assumptions of exemplar theoretic models and the assumptions about intonation in Germanic languages relate. The other one is the development of an instance-based research tool at the interface of written and spoken language. Both of these rely heavily on an definition of similarity by means of acoustic parameters. I will describe a planned perception experiment that aims at finding a precise way of modelling human perception of tonal similarity on the basis of these parameters.

“DevOps for Speech Research”

Ingmar Steiner

Saarland University

The term *DevOps* can be linked to a movement in the software industry that places emphasis on *tools* and *processes* and integrates them into a rapid release cycle. Many of these tools and practices can be adapted and integrated into research, specifically language and speech science and technology. In my talk I will give an overview of some of the ways in which a DevOps approach offers practical solutions for the challenges faced by researchers working with large collections of speech data and its annotations,

especially open data, and how these solutions can streamline research activities. In particular, concepts such as *dependency management*, *build automation*, and *continuous delivery* can be adapted for research projects and experiments with speech data.

“Modelling multimodal integration – The case of the McGurk-Effect”

*Fabian Tomaschek*¹ & *Daniel Duran*²

¹ Universität Tübingen, ² Universität Stuttgart

Our first aim in this study is to model the McGurk-Effect using two algorithms: NDL and an exemplar-theoretic model. Our hypothesis (w.r.t. NDL) is: [d] shares visual and acoustic cues with [b] and [g]. When presented separately, higher frequency of [d] attracts stronger association strength for [d] than for [b] or [g]. We test if the same holds for an Exemplar model.

Examining the influence of pitch accents on word learning in German

Michael Walsh

Universität Stuttgart

This paper investigates the relationship between pitch and the lexicon in the context of a pitch-accented-word learning experiment in German. Participants were presented with novel abstract objects with nonsense words for names, and were required to remember these object-name pairs. The nonsense names were presented during training with either rising or falling pitch accents. In the testing phase participants were asked if auditory stimuli matched subsequently presented visual stimuli. In order to examine the effect of pitch accents on word learning, the auditory stimuli either matched or varied from their training equivalents with respect to pitch accent. The results show that this variation subtly influences reaction times despite the fact that German is not a tone language.

“Examining exemplar effects in L2-learning”

Frank Zimmerer

Saarland University

A very basic form of exemplar models assume that every time a word (or phrase...) is heard, a new detailed entry is stored in the memory of a listener. Frequency of use is thus stored directly in the mental representation and can be considered a central concept in exemplar models. The assumed representations are also crucial for subsequent perception. However, it is impossible to control the perception history listeners. Additionally, the very basic form of exemplar models has been shown to be implausible. Pierrehumbert (2016) argued for a hybrid model that also assumes abstraction as an additional integral part of the architecture of lexical representations. Furthermore, recent research also suggests that there are individual differences between speakers and their abstraction strategies. I will present an experimental set-up that tries to investigate how the hybrid representations are built irrespective of listeners' perception histories by using an artificial language. A control over what listeners hear for the first time can help to investigate episodic and abstract effects, because all participants can be assumed to have the same perception of this language (i.e., = 0). The poster does not present results, it is rather aimed at discussing pros and cons of such an approach.