

KEYNOTE 1:

SHIFTING THE PARADIGM FROM SOURCE-FILTER-MODELLING TO SEGMENTAL-SUPRASEGMENTAL MODELLING OF SPEECH

Hartmut R. Pfitzinger

*Inst. of Phonetics and Digital Speech Processing (IPDS),
Christian-Albrechts-University Kiel, Germany*

hpt@ipds.uni-kiel.de

Extended Abstract

The human voice conveys linguistic content but also information about the speaker's gender, age, physique, and health situation, about his attitudes and emotions, about his personality, about his social and educational background, about his dialect and places where he resided for a longer time, and also about the interlocutor to whom he adapts his speaking style (which at least becomes obvious in the case of child- or pet-directed speech). Many of these factors are discussed since a long time but it seems impossible to list or even consider or investigate all factors determining the human voice, although the number of studies concerned with interactions between some of these factors is continually increasing during the last decade. However, the absolute number of those studies is still small. We claim that this is mainly due to a lack of an appropriate speech model.

Well-established speech models provide a strictly sequential concept from the "sender" to the "receiver" via the speech production mechanism, the acoustic transmission, and the speech perception system. They are not capable of directly representing the multi-dimensional interactions between the above-mentioned factors.

Therefore, the keynote talk presents a new model of speech proposing a paradigm shift from the classical articulatory-acoustic view of speech, that is perfectly represented by the acoustical theory of speech production and the well-established source-filter-model, towards a functional speech model which separates each of both, the source as well as the filter, into two components: the segmental layer and the suprasegmental layer. The segmental layer is concerned with those excitation signal variations and vocal tract deformations induced by the articulatory rendering of words, whereas the suprasegmental layer contains (besides the linguistic information that prosody provides) also and mainly non-linguistic information, i.e. expressive, speaker-specific, para-, and extra-linguistic effects on the excitation signal as well as on the vocal tract.

The model is established by a data-driven parameter decomposition approach based on mathematical methods taken from linear algebra e.g. multiple linear regression (MLR), singular value decomposition (SVD), principal component analysis (PCA), and linear predictive coding (LPC). Details of this technique are presented at the conference. This functional speech model enables access to higher-level knowledge and parameter control in automatic speech analysis, modification, and synthesis.

On the presenter

- Hartmut R. Pfitzinger was born in September 1968 in Kiel, Germany
- 1989-1992: He studied Phonetics and Digital Speech Processing (Prof. Kohler), Computer Science, Mathematics, and Psychology at Christian-Albrechts-Universität in Kiel

- 1992: passing intermediate examination in Kiel, move to München
- 1992-1995: At Ludwig-Maximilians-Universität München he studied Phonetics and Speech Communication (Prof. Tillmann), Computer Science, and Computational Linguistics
- Feb. 1995: M.A. degree in Phonetics and Speech Communication, Computational Linguistics, and Computer Science
- Jul. 2001: He received the PhD degree in Phonetics and Speech Communication, Computational Linguistics, and Computer Science
- May 2007: He received the Habilitation degree for Phonetics from the Ludwig-Maximilians-Universität München
- Nov. 2007: He received the *venia legendi* for Phonetics from the Christian-Albrechts-Universität in Kiel
- Feb. 1995-Mar. 2007: Assistant lecturer and researcher at Institute of Phonetics and Speech Communication, München
- Feb.-May 2003 and Feb.-May 2004: Two stays as invited scientist at the JST/CREST Expressive Speech Processing (ESP) lab in ATR, Kyoto
- Dec. 2004-Jan. 2005: Stay as visiting scientist at MARCS Auditory Laboratories at the University of Western Sydney
- Sep.-Oct. 2009 and Feb.-Mar. 2010: Stay as invited scientist at Laboratory of Sensory Research, Instituto de Neurociencias Aplicadas, Hospital de Clinicas in Buenos Aires
- Apr. 2007-Mar. 2011: Temporary chair and visiting professor (*Lehrstuhlvertretung*) at Institute of Phonetics and Digital Speech Processing (IPDS) at Christian-Albrechts-Universität in Kiel