

Towards a Functional Model that Integrates Speech Production, Acoustics, and Perception

Hartmut R. Pfitzinger

20th June 2004

This abstract summarizes our recent research efforts in functional modelling of the relationship between speech production, acoustics, and perception and outlines our first ideas regarding a holistic functional model of the speech chain.

In 1995 and 2003 [1, 2, 3] we developed and improved a functional model that uses F0, F1, and F2 to predict the perceptual vowel quality in terms of the Cardinal Vowel Diagram of Daniel Jones. It is also able to predict the formant frequencies given a perceptually specified vowel quality and the target fundamental frequency. In 1996 [4] a syllable detection mechanism was developed which uses the lowpass filtered amplitude envelope of the bandpass filtered speech signal. It is able to segment syllable centers of phonetic syllables with fair accuracy. In 1999 [5] we described the relationship between acoustic features and the perceptual local speech rate by means of a simple linear model (linear models are well suited to account for *trading relations*). And in 2003 [6] we investigated the correlation between articulatory kinematics and perceptual local speech rate. Currently, we are using an articulatory-acoustic speech database (recorded with the *Carstens AG-500* 3D-EMA system) to further investigate the functional relationships between kinematic and acoustic properties of speech.

Functional modelling is central to all these investigations. It involves not only understanding the function of a component and its impact on other components of the speech chain. It also provides a formal description (usually in the form of a computer program) which allows the quality of the model to be evaluated since all functional models are in some sense simplified copies of natural real-world processes and thus always show a more or less imperfect behaviour.

The main questions to be discussed are *i)* how to discover research questions relevant in the context of functional modelling of the speech chain, *ii)* how to design and conduct experiments to drive model-building, and *iii)* how to functionally model the entire speech chain?

The answers will be given by outlining a first approach to a holistic functional model which is based on the assumption that human speech communication doesn't only follow statistical rules (which are mainly acquired empirically by humans and modelled best by Hidden Markov Models and Bayesian classification). We claim that there are fundamental relationships between the articulation, acoustics, and perception of speech sounds which for example enable humans simply by listening to sounds to draw conclusions about articulatory strategies and to imitate and reproduce them. And we also claim that these fundamental relationships not only exist on the segmental level but also on the supra-segmental level (at least in the case of intonation). We will also try to provide evolutionary evidence by referring to investigations of basic genetic programs of monkeys' communication.

References

- [1] H. R. Pfitzinger, "Dynamic vowel quality: A new determination formalism based on perceptual experiments," in *Proc. of EUROSPEECH '95*, vol. 1, Madrid, 1995, pp. 417–420.
- [2] ———, "Acoustic correlates of the IPA vowel diagram," in *Proc. of the XVth Int. Congress of Phonetic Sciences*, vol. 2, Barcelona, 2003, pp. 1441–1444.
- [3] ———, "The /i/-/a/-/u/-ness of spoken vowels," in *Proc. of EUROSPEECH '03*, vol. 1, Geneva, 2003, pp. 809–812.
- [4] H. R. Pfitzinger, S. Burger, and S. Heid, "Syllable detection in read and spontaneous speech," in *Proc. of ICSLP '96*, vol. 2, Philadelphia, 1996, pp. 1261–1264.
- [5] H. R. Pfitzinger, "Local speech rate perception in German speech," in *Proc. of the XIVth Int. Congress of Phonetic Sciences*, vol. 2, San Francisco, 1999, pp. 893–896.
- [6] H. G. Tillmann and H. R. Pfitzinger, "Local speech rate: Relationships between articulation and speech acoustics," in *Proc. of the XVth Int. Congress of Phonetic Sciences*, vol. 3, Barcelona, 2003, pp. 3177–3180.