New paradigms for speech analysis and processing: the source-filter model revisited and gesture-controlled analysis-by-synthesis

Christophe d'Alessandro, CNRS-LIMSI, Orsay, France

Knowledge discovery in speech analysis and processing is based on both static and dynamic features of the speech signals. Static features are corresponding to parameters of a model or "settings". Dynamic features are corresponding to parameter trajectories, or "gestures".

In a first part, the source filter-model of speech production is revisited. Although the voiced source component is usually described by non-linear time-domain glottal flow models, spectral modelling suggests that it can be considered as a mixed-phase filter, with an anticausal component corresponding to glottal open phase and a causal component corresponding to glottal closure. Identification of this causal-anticausal model can be achieved exploiting the phase properties of the glottal flow. Two signal representations taking advantage of this description have recently been investigated at LIMSI (Orsay) and FPMs (Mons): Zero of the Z Transform representation and lines of maximum phase of the wavelet transform. The performances of these representations for source-filter separation and estimation of various parameters (glottal closure instants, open quotient, glottal flow asymmetry and spectral richness) demonstrate their viability as alternatives to inverse filtering and ElectroGlottoGraphic analysis.

Speech synthesis has been for a long time one of the most fruitful tools for knowledge discovering in speech analysis and processing. In a second part, it is argued that this paradigm can be extended to dynamic features analysis, using real-time gesture-controlled instruments for analysis-by-synthesis. Experiments are reported, showing that such instruments allow for real-time manual control of glottal flow parameters, voice source aperiodicities and vocal tract formants. This could bring new insights into the dynamics of voice and speech in tasks such as expression of attitudes and prosody mimicking.