Integrating speech production and perception: an embodied optimization approach

Juraj Simko, Uni Bielefeld

Speech is a complex motor action harnessing a vast number of degrees of freedom in order to satisfy broad set of constraints imposed by communication requirements. However, these constraints alone come short of explaining a number of well known sequencing and coordination phenomena occurring in speech.

An embodied dynamical model of speech production enables us to principally integrate physiological and perceptual aspects of speech in terms of tradeoffs between complementary efficiency requirements. These requirements give rise to realistic articulatory trajectories and temporal patterns that are optimal with respect to the interplay between anatomic and neurophysiological properties of the embodied system and communication requirements in realistic environmental context. Rich prosodic patterns ubiquitous in speech can emerge as a result of high-level parameterization of these tradeoffs. Fine-tuning of lower-level optimization parameters of the unified system can be interpreted in terms of development and acquisition of the phonetics and phonology of a given language.

I will present an outline of our embodied task dynamical model combining speech production and perception that is based on the outlined optimization approach. I will then discuss our recent progress, plans and challenges encountered on the way to develop it into a fully fledged speech synthesis platform.