



## INVITED PAPER:

# Speech models and speech synthesis

Mary E. Beckman

*Ohio State University, Dept. Linguistics, Columbus OH 43210-1298*

Basic research in speech science over the last half century has enjoyed enormous benefits from our endeavours to synthesize speech by machine. For example, developing programs for simulating the time course of fundamental frequency variation over sentences and longer utterances has been an indispensable research tool in our basic understanding of intonation. Synthesis systems, in turn, have directly benefited from being able to incorporate the models of linguistic control of F0 originally built to test one or another theory of intonation. Moreover, incorporating these models in turn has made synthesis into an important research tool in another area of linguistics – namely, the examination of phonetic correlates of discourse structure and pragmatic intent, research that should lead us that much closer to our ultimate goal of going beyond mere synthesis to speech generation. Models of temporal control are another area that can see important cross-fertilization of results and ideas between basic research and synthesis. Since the 1970s, many synthesis systems have modeled timing control as the computation of context-sensitive durations for acoustic intervals corresponding to phoneme segments or their subunits. Such models have allowed us to take full advantage of statistical tools and the large speech data-bases now available, while yet incorporating the insights of several decades of

smaller controlled laboratory experiments on segment durations. A fruitful next step might be to explore how synthesis systems can incorporate a new consensus about timing control that has emerged from the recent explosion of basic research on articulation. Studies of articulatory kinematics suggest that a closer modeling of the spectral effects of articulator movement will be an important element in improving how well our synthesis systems capture the salient phonetic correlates of stress and phrasing. They also suggest that duration can no longer be treated quite so independently of fundamental frequency, amplitude variation, and other aspects of the spectrum now often modeled by the choice of concatenative unit.





