

Separate paper suggestions doctoral course

Fall semester 2003

Finite State Transducers

Jurafsky, D. & Martin, J. (2000), *Speech and Language Processing*, Prentice Hall, pp 71-90

Riley, M. Pereira, F., Mohri, M. (1997) "Transducer Composition for Context-Dependent Network Expansion", *Proceedings from EUROSPEECH '97*, pp 1427-1430.

Pronunciation variability

Jurafsky, D. & Martin, J. (2000) *Speech and Language Processing*, Prentice Hall, pp 142-189

Strik, H., Cucchiaroni, C. (1999) "Modeling pronunciation variation for ASR: A survey of the literature", *Speech Communication Vol. 29*, pp 225-246.

McAllaster, D., Gillick, L., Scattoni, F., Newman, M. (1998) "Fabricating Conversational Speech Data with Acoustic Models: A Program to Examine Model-Data Mismatch", *Proc. ICSLP '98*, Sydney.

Jurafsky, D., Ward, W., Jianping, Z., Herold, K., Xiuyang, Y., & Sen, Z. (2001) "What Kind of Pronunciation Variation is Hard for Triphones to Model?", *Proc. ICASSP 2001*.

Alternatives to standard HMM

Livescu, K., Glass, J., Bilmes, J. (2002) "Hidden Feature Models for Speech Recognition Using Dynamic Bayesian Networks", *Proc. EUROSPEECH 2002*

Weber, K., Bengio, S., Bourlard, H. (2000) "HMM2- A Novel Approach to HMM Emission Probability Estimation", *Proc. ICSLP 2000*.

Dupont, S., Bourlard, H. (1997) "Using Multiple Time Scales In A Multi-Stream Speech Recognition System", *Proc. EUROSPEECH '97*

Blomberg, M. (1991) "Modelling articulatory timing variation in a speech recognition system based on synthetic references", *Proc. EUROSPEECH '91*, pp789-792.

Deng, L., Ramsay, G., Sun, D. (1997) "Production models as a structural basis for automatic speech recognition", *Speech Communication, Vol 22*, pp 93-111.

Gao, Y., Bakis, R., Huang, J., Xiang, B. (2000) "Multistage Coarticulation Model Combining Articulatory, Formant and Cepstral Features", *Proc. ICSLP 2000*.

Prosody

Crystal, T, House, A. (1986) "Characterization and Modeling of Speech-Segment Durations", *Proc. ICASSP 86*, pp 2791-2794.

Chen, K., Borys, S., Hasegawa-Johnson, M., and Cole J. (2003) "Prosody Dependent Speech Recognition with Explicit Duration Modelling at Intonational Phrase Boundaries", *Proc. EUROSPEECH 2003*.