

Integrated acoustic and language modeling of speech disfluencies. Elizabeth E. Shriberg, Rebecca A. Bates, and Andreas Stolcke (Speech Technology and Research Laboratory, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025)

This work investigates the use of prosodic features in modeling disfluencies (filled pauses, repeated words, and self-repairs) in spontaneous speech. The main goal is to automatically detect and correct disfluencies, so that a “fluent” version of a disfluent utterance can be used as input for speech understanding and other applications. A second goal is to develop explicit acoustic and language models for disfluencies to improve speech recognition performance for spontaneous speech. The prosodic features examined include duration, fundamental frequency, amplitude, and features correlated with voice quality. Decision trees serve as the acoustic models that relate these prosodic features to disfluency events. To integrate the disfluency model into speech recognition, decision tree probabilities are combined with standard acoustic model scores and probabilities from a “Clean-up” language model to rescore N-best hypotheses. The Clean-up language model represents disfluencies as hidden events, and predicts words following a disfluency from the corresponding fluent word sequence. A linguistically hand-annotated version of the Switchboard corpus is used for model training and evaluation. [Work supported by the National Science Foundation under Grants IRI-9314967 and IRI-8905249.]

Number of words in abstract: 178

Suggested Special Session: Invited, Prosody modeling and its application to recognition and synthesis

Technical Area: Speech Communication

(PACS) Subject Classification number(s): 43.70.Fq, 43.72.Ne

Telephone number: 415-859-3798 email address: [ees@speech.sri.com](mailto:ees@speech.sri.com)

Send notice to: E.E. Shriberg. EJ 196 SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025

Special Facility:

Method of Presentation: