

**1aSC9. Emotion to emotion speech conversion in phoneme level.**

Murtaza Bulut, Serdar Yildirim, Carlos Busso, Chul Min Lee, Ebrahim Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

Having an ability to synthesize emotional speech can make human-machine interaction more natural in spoken dialogue management. This study investigates the effectiveness of prosodic and spectral modification in phoneme level on emotion-to-emotion speech conversion. The prosody modification is performed with the TD-PSOLA algorithm (Moulines and Charpentier, 1990). We also transform the spectral envelopes of source phonemes to match those of target phonemes using LPC-based spectral transformation approach (Kain, 2001). Prosodic speech parameters ( $F_0$ , duration, and energy) for target phonemes are estimated from the statistics obtained from the analysis of an emotional speech database of happy, angry, sad, and neutral utterances collected from actors. Listening experiments conducted with native American English speakers indicate that the modification of prosody only or spectrum only is not sufficient to elicit targeted emotions. The simultaneous modification of both prosody and spectrum results in higher acceptance rates of target emotions, suggesting that not only modeling speech prosody but also modeling spectral patterns that reflect underlying speech articulations are equally important to synthesize emotional speech with good quality. We are investigating suprasegmental level modifications for further improvement in speech quality and expressiveness.