November 16, 2009

Identifying informative and salient words to improve the expressiveness of English Text-to-Speech Synthesis

Leonardo Badino

Centre for Speech Technology Research
Informatics Forum
Edinburgh

ABSTRACT OF THE TALK

In this talk I will address the problem of improving and enriching the expressiveness of English Text-to-Speech (TTS) synthesis by identifying and generating natural patterns of intonational prominence.

Although in spoken English the degree of intonational prominence to which a word is uttered is closely related to its degree of informativeness and salience, which in turn is largely determined by the context surrounding that word, the identification of prominent words in most of the state-of-the art TTS systems relies on gross and context-independent linguistic features.

After briefly introducing the two most successful speech synthesis techniques I will present my work on the prediction of pitch accent placements (intonational events that tend to signal salient and informative words) and its impact on TTS synthesis, and cast doubt on the classic accent vs. no-accent dichotomy.

The last part of the talk is devoted to my work on a novel and challenging task, the automatic identification of emphatic words from text and their generation in Hidden-Markov-Models based TTS synthesis. I will show results from perceptual experiments and argue that this is the kind of task that can significantly make TTS synthesis sound more expressive and natural.