Abstract

This paper describes the factors used in designing and recording large speech databases for applications requiring speech synthesis. Given the growing demand for customized and domain specific voices for use in corpus based synthesis systems, good practices should be established for the creation of these databases which are a key factor in the quality of the resulting speech synthesizer. This paper focuses on the factors affecting the designing of the recording prompts, on the speaker selection procedure, on the recording setup and on the quality control of the resulting database. One way to find the emotions in the speech is, once the speech has been recorded from the user it is converted into text, at the same time the
stressed word from the speech is recorded & then the frequency for that word is find out for recording the corresponding emotion.

References

- Gregor O. Hofer, “Emotional Speech Synthesis”, Master of Science School of Informatics University of Edinburgh 2004
- Inger S. Engberg, Anya V. Hansen, Ove Andersen and Paul Dalsgaard, “Design, Recording and verification of a Danish Emotional speech Database”
- Lu’s C. Oliveira, S’ergio Paulo, Lu’s Figueira, Carlos Mendes, Ana Nunes‡, Joaquim Godinho‡, “Methodologies for Designing and Recording Speech Databases for Corpus Based Synthesis”
- Voice”, Irvine, “Models of Speech Synthesis”, draft version of a paper presented at the "Colloquium on Human-Machine Communication California, February 8-9, 1993, organized by the National Academy of Sciences, USA.

Index Terms

Computer Science

Engineering and Technology
Keywords
Extraction of Emotion in Speech  Database  Recording