Abstract

The paper explores to what extent Voice Conversion techniques can help incorporate dialect specific features into synthesized speech. A popular Voice Conversion technique using Gaussian Mixture Models, has been used to develop mapping functions, between speech synthesized by a Text-to-Speech System for the standard form of the language to parallel speech recorded from a speaker of the target dialect. Mel Cepstral Coefficients are used to represent the spectral envelope and pitch, intensity and duration values have been selected to represent the prosody of speech.

References

2. Elisabeth Zetterholm. Same speaker–different voices. a study of one impersonator and some of his different imitations. In Proceedings of the 11th Australian International Conference
Incorporating Dialectal Features in Synthesized Speech using Voice Conversion Techniques


Index Terms

Computer Science

Signal Processing
Keywords

Voice Conversion, Gaussian mixture models, Mel Cepstral Coefficients, Formants, F0, Assamese, Nalbaria, Dialect, Pitch, Intensity, Duration, Text-to-Speech System