Abstract

This paper proposes a method for tuning the weights of unit selection cost functions in syllable based text-to-speech (TTS) synthesis system, two-stage feedforward neural network (FFNN) based approach for modeling fundamental frequency (F0) values of a sequence of syllables. Unrestricted Text To Speech System (TTS) is capable of synthesize different domain speech with improved quality. A clustering technique is used in annotated speech corpus that provides
way to select the appropriate unit for concatenation, based on the lowest total join cost of the speech unit. Unit selection cost functions, namely target cost and concatenation cost, are designed appropriate to syllables. The method tunes the weights in such a way that perceptual preference patterns are appropriately considered while selecting the units. The method uses genetic algorithm to derive the optimal weights. From the evaluation, it is observed that prediction accuracy is better for two stage FFNN models, compared to the other different models.

References


Index Terms

Computer Science

Signal Processing
Keywords
Text To Speech Synthesis (tts)  Concatenative Synthesis Approach  Intonation Models  Feed Forward Neural Networks

Unit Selection

Target Cost

Tuning Of Weights.