Abstract

The prevalent approach for automatic segmentation in speech synthesis is Hidden Markov Model (HMM) - based. Even though an HMM-based approach is the most automatic and reliable, there are still several limitations, such as mismatches between hand-labeled transcriptions and HMM alignment labels which can lead to discontinuities in the synthetic speech, or the need for hand-labeled bootstrap data in HMM initialization. This final task introduces a new method which is combine an HMM-based approach and spectral boundary correction. This method generally consist of four phases; initialization with Speaker Independence (SI) HMM, reestimation of HMM parameters, segmentation with viterbi alignment, and segment boundary correction with Spectral Boundary Correction. In this proposed method, boundary of each segment from HMM-based approach will be corrected based on their spectral distribution to get the better result. The conclusion of this final task is an automatic speech segmentation combining an HMM-based approach and spectral boundary correction gives 73 % af accuracy, which is affected by the type of speaker and the speed of pronounciation.

Keywords: segmentation, HMM, bootstrap, spectral boundary correction, viterbi.