論文 / 著書情報
Article / Book Information

題目(和文)	スペクトル特徴を用いた効率的な音声区間検出と音声強調アルゴリズム
Title(English)	Efficient Voice Activity Detection and Speech Enhancement Algorithms based on Spectral Features
著者(和文)	MaYanna
Author(English)	Yanna Ma
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第9632号, 授与年月日:2014年9月25日, 学位の種別:課程博士, 審査員:西原　明法,國枝　博昭,山田　功,府川　和彦,篠田　浩一,入野　俊夫
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第9632号, Conferred date:2014/9/25, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

A novel and robust Voice Activity Detection (VAD) algorithm utilizing long-term spectral flatness measure (LSFM) and an efficient speech enhancement (SE) algorithm based on modified Wiener filtering method have been proposed in this thesis.

In chapter 1, the research significance and problem statement about voice activity detection and speech enhancement were introduced. The motivations and objectives of doing research on voice activity detection and speech enhancement were also given, respectively.

In chapter 2, a brief review of the three standard VAD schemes, namely, G.729B, AMR1 and AMR2 were provided for the comparison with our proposed VAD scheme. Also, the classical statistical model-based VAD and the newly proposed LTSV based VAD were presented in this chapter.

Four kinds of SE algorithms including spectral-subtractive, subspace, statistical-model based and Wiener-type methods were also introduced briefly.

Finally, the TIMIT corpus and NOISEX-92 noise database for VAD and the NOIZEUS corpus and AURORA database for SE algorithms are also described.

Chapter 3 introduced an efficient long-term spectral flatness measure-based VAD algorithm. The motivation of exploring flatness measure along time frames using a long window was clarified by the LSFM feature distributions as a function of the long-term window length R. The discriminative power of the LSFM feature was verified in terms of the separateness of its distribution for noisy speech and non-speech signals. The decision threshold was adapted according to the previous 100 LSFM measures of speech and non-speech. Experiments were done on core TIMIT test set for 12 kinds of noises (11 from NOISEX-92 database and speech-shaped noise) across five different SNRs ranging from -10 to 10 dB. No a priori knowledge of noise characteristics was needed for training purposes. The performance of our proposed method was compared with the three standards (namely

G.729B, AMR1, and AMR2) and with a classical statistical model based VAD and an emerging LTSV-based VAD algorithm. The results were analyzed by accuracy rate and error rate. Through extensive experiments, we showed that our proposed LSFM-based VAD achieved the best CORRECT, HR0, and NDS and among all tested schemes that averaged all 12 kinds of noises. Furthermore, we investigated the individual performance on each noise type. Our proposed LSFM-based VAD outperformed LTSV-based VAD for 6 out of 12 noise types tested especially for non-stationary impulsive machine gun noise and speechlike babble noise.

In chapter 4, a new SE algorithm based on imposing constraint on Wiener gain function was introduced.

Experiments were done on NOIZEUS database for eight kinds of noise (AURORA database) across seven different SNRs ranging from -8dB to 15dB. The Wiener_Clean algorithm was taken as the ground truth. The performance of our proposed algorithm was compared with WT and KLT methods. The results were analyzed mainly by three composite measures and the SNR_LOSS measure to predict the performance on subjective quality and speech intelligibility, respectively.

The SNR_LOSS measure values obtained from each algorithm were subjected to statistical analysis in order to assess their significance differences. A highly significant effect ($p<0.005$) was found in all SNR levels and all types of noise by analysis of variance (ANOVA). Following the ANOVA, multiple comparison statistical tests according to Tukey's HSD test were done to assess significance between algorithms. The difference was deemed significant if the p value was smaller than 0.05.

Through extensive experiments, we showed that when averaged over all eight kinds of noises, our proposed SE algorithm achieved the best results in terms of predicting signal distortion C_sig and overall quality C_ovl when SNR is no more than 10dB. Furthermore, we investigated the individual

performance on each noise type. Our proposed SE algorithm outperformed the KLT algorithm for all noise types tested in terms of both C_sig and C_ovl.

On the other hand, the SNR_LOSS measure comparisons with both the UP and other SE algorithms predicted that our proposed algorithm was able to improve speech intelligibility for low SNR levels and outperform WT and KLT algorithms for most conditions examined.

It is important to point out that the three composite measures and the SNR_LOSS measure used in this chapter are adopted for predicting the subjective quality and intelligibility of noisy speech enhanced by noise suppression algorithms because of their high correlation with real subjective tests. Further subjective tests on both normal-hearing listeners and hearing impaired listeners are needed to verify the effectiveness of the proposed algorithm on improving both subjective quality and speech intelligibility.

It is also worth mentioning that depending on the nature of the application, some practical SE systems may require very high quality speech, but can tolerate a certain amount of noise while other systems may want speech as clean as possible even with some degree of speech distortion. Therefore, it should be noted that according to different applications, different SE algorithms should be chosen to meet the variant requirement.

In chapter 5, our main contributions for VAD algorithm and SE algorithm are summarized. The parameter setting and trade-off between HR1 and HR0 for the proposed LSFM-based VAD are also discussed. Furthermore, the requirement for real subjective tests to evaluate our proposed SE algorithm is also stated in this chapter.