Form 2
Dissertation Abstract

Report no.	(Course-based)	No. 1097	Name	MD. ARIFOUR RAHMAN
Dissertation title	Accurate Spectrum Estimation Based on Linear Prediction for Speech Signals (音声信号のための線形予測に基づく高精度なスペクトル推定)			

Abstract

Speech signal is produced by the convolution of excitation source and time varying vocal tract system components. These excitation and vocal tract components are to be separated from the available speech signal to study these components independently. The discrete-time samples of a speech signal do not exactly follow the model defined for speech generation. This makes the separation and representation of speech signal component inconvenient for the application such as recognition, coding, transmission, storage, and so on. Thus, efficient representation of speech signal in terms of a small number of slowly varying parameters is a problem of considerable importance in speech research.

Linear prediction (LP) analysis has solved the problem greatly and has been applied to speech system over the last few decades. LP technique is well-suited for speech analysis due to its ability to model speech production process approximately. Hence LP analysis has been widely used for speech enhancement, low-bit-rate speech coding in cellular telephony, speech recognition, characteristic parameter extraction (vocal tract resonances frequencies, fundamental frequency called pitch) and so on. There are basically two different formulations of linear prediction, namely, autocorrelation method and covariance method. However, the performance of the conventional LP method is degraded in an ill-conditioned environment. Ill-condition of the correlation matrix occurs in the case of high spectral dynamic range of the input speech signal. The problem of ill-conditioning is severely presented in the bone conducted (BC) speech, which is generally recorded by bone conductive microphone, mounted on the talker's head. BC speech has amplitude attenuation in high frequency regions being more than 1 kHz usually. If the LP method is directly applied to the BC speech, then the performance is degraded. In order to improve the performance of LP analysis, it is necessary to make the correlation matrix well-conditioned, which is a most challenging task for LP analysis. In this dissertation, we focus on resolving the factors that degrade the performance of LP analysis and new approaches have been proposed and implemented. We developed some approaches to improve the performance of the LP analysis based on prediction error filtering. Unlike some conventional techniques, which uses the pre-emphasis filter to emphasis the high frequency component of the input speech signal, we utilize a simple prediction error filter as a pre-processor in LP analysis. The approach can improve the performance of LP analysis under noisy

environment.

In the first proposed method, the LP based power spectrum estimation is compensated by the spectrum characteristics of the designed prediction error (PEF) filter. The accuracy of formant frequency estimation is verified on synthetic speech. Averaged condition number was estimated for different synthetic vowels to observe the ill-conditioning of the correlation matrix. Spectral bias value was also estimated to investigate closeness of the estimated spectrum to the true spectrum. The validity of the proposed method was also illustrated by inspecting real air conducted and bone conducted speeches. The experimental results, based on synthetic and real speeches, demonstrate the effectiveness of the new approaches for improving the performance of the LP analysis.

The second method proposes a new approach of LP by averaging the LP coefficients. A modified spectral subtraction method was also used here for the further improvement. The proposed method was applied to reduce the spectral distortion of the noisy bone conducted speech and enhance the quality also, without using any spectral characteristics of the corresponding air conducted speech. Log-spectrum distortion (LSD) measurement is used for the objective evaluation of the proposed method. Also, we carried out mean opinion score (MOS) tests to measure the sound quality of the reconstructed speech.

The third method proposes an approach of pitch synchronous linear prediction (LP) for bone conducted (BC) speech. A combination of the spectral compensation (SC) method with pitch extension LP is used to obtain more accurate power spectrum of the BC speech signal. The pitch extension approach requires the information of pitch (inverse of fundamental frequency) of speech signals. While we cannot avoid pitch detection errors in implementing pitch detectors on normal speech, pitch could be detected accurately from BC speech. This property of BC speeches very beneficial for utilizing pitch-synchronous analysis. We generated the BC synthetic speech by low-pass filtering the AC synthetic speech. Log-spectrum distortion (LSD) measurement was used for the objective evaluation of the proposed method. After the investigation on the synthetic BC speech, the proposed method was also applied to real BC speech. Experiments through synthetic and real BC speeches have demonstrated that the proposed method provides more accurate power spectrum estimation.