平成 20 年 1 月 10 日

電子・情報工学専攻		学籍番号	**035401**	指導教官氏名	**中川 聖一** 秋葉友良
申請者氏名	**王 龍標**				

論 文 要 旨 (博士)

論文題目	**A Study on Hands-Free Speech/Speaker Recognition (ハンズフリーによる音声認識と話者認識に関する研究)**

(要旨 1,200 字程度)

In this thesis, we first propose a robust speech/speaker recognition method by incorporating the estimated speaker position information into Cepstral Mean Normalization (CMN), which is called Position-Dependent CMN (PDCMN). The system measures the transmission characteristics (the compensation parameters for position-dependent CMN) from some grid points in the room *a priori*. Four microphones are arranged in a T-shape on a plane, and the sound source position is estimated by Time Delay of Arrival (TDOA) among the microphones using a proposed closed-form solution. The system then adopts the compensation parameter corresponding to the estimated position and applies a channel distortion compensation method to the speech (that is, position-dependent CMN) and performs speech/speaker recognition.

A speaker recognition method by combining speaker-specific GMMs with speaker-adapted syllable-based HMMs for close-talking speaker recognition is extended to distant-talking speaker recognition and integrate this method with the proposed position-dependent CMN. By integrating the position-dependent CMN into the combination use of speaker-specific GMMs and speaker-adapted syllable-based HMMs, a remarkable improvement was obtained.

In a distant-talking environment, the length of channel impulse response is longer than the short-term spectral analysis window. Therefore, conventional short-term spectrum based Cepstral Mean Normalization (CMN) is not effective under these conditions. In this thesis, we also propose a robust speech recognition method by combining a short-term spectrum based CMN with a long-term one. The proposed variable term spectrum based PDCMN achieved a relative error reduction rate of 60.9% over the conventional short-term spectrum based CMN and 30.6% over the short-term spectrum based PDCMN.

We also propose a blind dereverberation method based on spectral subtraction by Multi-Channel Least Mean Square (MCLMS) algorithm for distant-talking speech recognition. By treating the late reverberation as additive noise, a noise reduction technique based on spectral subtraction is proposed to estimate power spectrum of the clean speech using power spectra of the distorted speech and the unknown impulse responses. To estimate the power spectrum of the impulse response, a Variable Step-Size Unconstrained MCLMS (VSS-UMCLMS) algorithm for identifying the impulse response in a time domain is extended to the spectral domain. We conducted the experiments on distorted speech signals simulated by convolving multi-channel impulse response with clean speech. An average relative recognition error reduction of 17.8% over conventional CMN under various severe reverberant conditions was achieved.