氏名・(本籍)	Md. Ekramul Hamid（バングラデシュ）
学位の種類	博　士　（工　学）
学位記番号	工博甲第　288　号
学位授与の日付	平成 19 年 3 月 23 日
学位授与の要件	学位規程第 5 条第 1 項該当
研究科・専攻の名称	電子科学研究科　電子応用工学
学位論文題目	**A single-channel noise reduction method of speech by noise subtraction and blind source separation.** (雑音減算とブラインド音源分離による音声のシングルチャネル雑音低減法)
論文審査委員	(委員長) 教 授　中 井 孝 芳　　教 授　北 澤 茂 良 教 授　竹 林 洋 一　　教 授　深 林 太計志

論 文 内 容 の 要 旨

Varieties of environmental sources of noise and distortion can degrade the quality of the speech signal in a communication system. This thesis explores the effects of these interfering sounds on speech applications and introduces some techniques for reducing their influence and enhancing the acceptability and intelligibility of the speech signal. In this thesis, a noise reduction system using single-microphone method in time domain to improve SNR of noise-contaminated speech is proposed. A novel two-stage noise reduction technique is applied, namely Noise Subtraction (NS) stage followed by Blind Source Separation (BSS) stage.

In the first noise reduction stage, time domain noise subtraction (NS) method (equivalent to spectral subtraction) is applied. The amount of noise in this stage (first noise) is estimated from the valleys of the spectrum based on the harmonic properties of noisy speech, called Minimum Value Sequences (MVS). However, the valleys of the spectrum are not pronounced enough to warrant reliable noise estimates. We therefore propose the estimated Degree of Noise (DON) to adjust the amplitudes of the MVS. Also we perform some suitable steps to minimize the residual noise problem. It shows that the estimated degree of noise and the minimum value sequences can be used to estimate the noise spectrum almost accurately and is an efficient tool to reduce noise. To estimate the degree of noise, a function was previously prepared using the least-squares (LS) method, from the given (true) DON and the estimated parameter of DON. This parameter is obtained from the Auto-Correlation Function (ACF) of the noisy speech on a frame basis. On the basis of this concept, a new technique is developed that continuously estimates noise on a frame-by-frame basis without the need for Voice Activity Detector (VAD). This stage performs well to remove noise from the voiced segments of high SNR noisy speech signals.

In the next stage, we study the time domain blind source separation (BSS) of convolutive mixtures modeled by adaptive FIR filters to further reduce the effect of noise and the noise residuals from the first noise-reduced speech from previous stage. The coefficients of the FIR filters are estimated from the decorrelation of two mixtures. The process is blind because we have no idea about the mixtures of the input signals. We discuss the technique for achieving BSS when only a single-channel noisy speech is input to the system. Because the method is a single-channel, the other input to the BSS is a noise dominant signal, called second noise here. The second noise is generated by taking the first noise characteristics into consideration. The generated signal plays a very important role in the performance of the BSS system. In terms of computational complexity, NS in the time domain (TD) is much more efficient than SS when combining with BSS in TD. We find that the BSS (alone) shows poor performance but by controlling noise using the NS, the BSS improves the performance of this stage. We obtain noise reduction (words) of about 8.4dB for white, 4.9dB for crowd and 5.2dB for white, 3.2dB for crowd noise at 0dB and 10dB input SNR respectively. It shows that combination of the two methods results in a high noise reduction score than a single processing method.

In the next phase, we further study to improve the performance of the NS+BSS algorithms. Then we propose Weighted Noise Subtraction (WNS), improved blind source separation and combination of the two (WNS+BSS) methods. It is experimentally proved that when the weight is the square of the complement of the DON, the performance of the BSS is improved. Like before, initially, input noise is reduced to obtain the first noise- reduced speech by the weighted noise subtraction.

A method of and a system for noise suppression, consisting of two adaptive filters for the first noise-reduced speech and the estimated noise-dominant (second noise) signal, are developed. The least-mean-squares (LMS) algorithm that is based on the steepest-descent method is implemented to update the filter coefficients in an iterative manner. The method addresses the situations in which the input signal-to-noise ratio (SNR) varies substantially and performing the specified number of iterations of the LMS algorithm for each SNR is time-consuming. Therefore, we propose a function that can be used to estimate the number of iterations required for a given value of the DON of noisy speech. By adopting the proposed number of iterations, the system becomes computationally inexpensive and the signal regeneration problem is minimized in the silent parts. Moreover, good efficiency of the algorithm is achieved by appropriate block length processing. The method is tested with five types of noise. We obtain noise reduction (words) of approximately 11.3dB, 8.1dB and 6.6dB for white, crowd and factory noise respectively at 0dB input SNR and at 10dB input SNR, the noise reduction is 5.7dB, 4.3dB and 3.5dB by the WNS+BSS. The experimental results with both subjective and objective evaluation confirm the acceptability of the proposed WNS+BSS method.

論 文 審 査 結 果 の 要 旨

種々の環境で音声機器を使用する機会が多くなり、雑音低減が重要な問題になっている。本論文は雑音低減機能組み込み機器のコンパクト性を考慮し、シングルチャネル入力の音声の雑音低減法を提案したもので、全8章からなる。

第1章～第3章では本研究の背景と目的、雑音の低減に関する過去の研究紹介や雑音の音声に対する影響等を述べ、本研究の内容を理解する上で必要な基礎を与えている。第4章ではシングルチャネル入力の雑音混入音声（観測信号）から、音声の特徴を利用しての雑音推定問題を扱っている。本研究で重要な役割をする音声の雑音の程度(DON)の定義とそれを推定する実験式を導出している。この実験式を用いることで、SNRが0～15dBの音声に対してほぼDONを推定できることを示している。DONの推定はシングルチャネル入力で、しかも処理フレーム内で行えるところに特徴がある。第5章ではDONを用いて推定した雑音スペクトルを基に、スペクトル減算法と等価な時間領域での雑音減算法(NS)、それによる実験結果とその問題点を述べている。第6章ではNSで用いた推定雑音と観測信号からDONと音声のピッチ情報を基に雑音入力を生成して、2入力音源分離法（BSS）を適応した雑音低減法を提案し、それ単独では雑音低減効果が小さいが、NSとBSSの組み合わせ(NS+BSS)の2段階処理にすると雑音低減効果が大きくなることを実験結果で示している。第7章ではDON情報を基にNSを重み付き(WNS)とした、すなわち、雑音の減算量を抑圧したNSとBSSの組み合わせ(WNS+BSS)を提案し、BSSによる同一フレーム内での処理の繰り返しを制御することにより、さらに雑音低減効果を高めている。繰り返し制御のために、DONと繰り返し回数の関係式を導出して利用している。白色雑音、人混み雑音、工場雑音など5つのタイプのそれぞれの雑音混入音声を用いた実験により、提案法の雑音低減性能を求めている。その性能はSNR 0dBと10dBの白色雑音混入音声については、それぞれ11.3dB程度と5.7dB程度、SNR 0dBと10dBの人込み雑音混入音声については、それぞれ8.1dB程度と4.3dB程度になることなどを述べている。第8章は結論である。

以上のように、本論文はシングルチャネル入力で、2段階処理による、音声の雑音低減法を確立しており、今後の音声の雑音低減技術の改良に役立てることができるものである。よって、本論文は博士（工学）の学位を授与するに適当な内容を具備していると認定する。