

Lab name (Supervisor)	Augmented Human Communication (Prof. Satoshi Nakamura)		
Name (surname) (given name)	Aprilyanti, Fine Dwinita	Date	2015/12/03
Title	Microphone Array Processing based on Blind Source Separation for Robust Distant Speech Recognition System		
<p>Abstract:</p> <p>Distant speech recognition system, in which a single or an array of microphone is utilized to capture the user's utterance as opposed to the user-attached microphone, is greatly affected by the presence of background noise and reverberation. Many researches have been conducted on developing an array processing method to improve the quality of the captured speech signals. Most of these methods are optimized to obtain the clean target speech, as measured by signal-to noise ratio or human perception. However, speech recognition system works as a statistical pattern classifier of features extracted from the speech waveform. Therefore, array front-end processing can only be expected to increase the recognition accuracy if it maximizes the likelihood of the correct hypothesis.</p> <p>In this study, I propose array processing techniques based on blind source separation to suppress the background noise and late reverberation, optimized to maximize the likelihood to the acoustic model of speech recognizer. The first method utilizes frequency-domain blind signal extraction (BSE), which is an alternative to the conventional blind source separation specifically designed for the case of speech in the presence of diffuse noise, combined with two stages of multichannel Wiener filter. I extend this method by integrating information from the image sensor to achieve optimum performance regardless the interference level. In the second method, I combine BSE with multichannel generalized minimum mean-square error estimator of short time spectral amplitude (MMSE-STSA), which can provide less distortion to the output signal owing to the use of speech spectral amplitude statistical model assumption and decision-directed signal-to-noise ratio estimation approach.</p>			

氏 名	Fine Dwinita Aprilyanti
-----	-------------------------

(論文審査結果の要旨)

近年、遠隔位置で発話された遠隔発話の音声認識の必要性が高まっている。遠隔発話の受音においては環境雑音、残響の影響を受け性能劣化が避けられない。このため、複数のマイク素子からなるマイクロフォンアレーや単一のマイクでも高度な抑圧処理が用いられる。従来は雑音抑圧を中心にした研究が多かったが、本研究では音声認識の性能を向上する研究はあまり行われてこなかった。本研究では、3つの方法でブラインド音源分離を音声認識性能が改善するように雑音、残響に最適化した。まず、第1にブラインド信号抽出(BSE)を用いた方法を提案する。この方法では、2ステップのウィーナフィルタをBSEに統合して用いる。さらに画像情報による音源方向情報を利用して性能を向上した。次に、第2の方法として、BSEを一般化最小2乗誤差によるスペクトル推定法に統合し歪みを改善する方法を提案した。これらの方法は、雑音が拡散性の時に有効であるため、雑音に指向性がある際にも改善を行うため、第3の方法として、音源に適応的なブラインド音源分離を行うことで多様な雑音や残響に頑健な方法を実現し、実験により有効性を示した。

これらの成果は、従来技術では本質的に解決困難であった問題に対して、解決策を示し、非常に高く評価できる。本研究成果は、1編の学術論文、2編の査読付き国際会議、3編の国内会議論文として発表されており、研究成果の有効性を見ることができる。

以上より、平成28年2月19日に開催した公聴会の結果も参考にして、本博士論文の審査を行い、本論文は、博士論文(工学)として十分な価値があるものと判断した。