様式Ｃ－１９

# 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　25　年　6　月　5　日現在

機関番号：13102
研究種目：若手研究（B）
研究期間：2010 ～2012
課題番号：22700169
研究課題名（和文）　マルチチャンネル最小二乗平均を用いた複数話者の発話に頑健なハンズフリー音声認識
研究課題名（英文）　Distant-talking speech recognition based on spectral subtraction by multi-channel least mean square approach
研究代表者
王　龍標（WANG LONGBIAO）
長岡技術科学大学・産学融合トップランナー養成センター・産学融合特任准教授
研究者番号：30510458

研究成果の概要（和文）：
遠隔環境下で音の生成を定式化し、伝送路の伝達特性を自動的に推定し、様々な残響環境に対して頑健な残響除去および残響除去の信頼性を用いる後処理を行い、高精度な残響処理を実現した。また、パワースペクトル減算（SS）の代わりに，一般化 SS を用いたブラインド残響除去法を提案し，パワーSS に基づくブラインド残響除去法に対してエラー率が大幅に削減できた。さらに、実環境（会議室）の残響を含んだ音声を収録し評価に用いた。人工残響音声と同程度のエラー削減率を達成した。なお、非定常雑音である音楽を含む残響音声に対して，本提案のマルチチャンネル最小二乗平均を基づく一般化スペクトルサブトラクション（GSS）によるブラインド残響除去法と ICA（独立成分分析）に基づくブラインド音源分離を組み合わせる方法を提案しました。

研究成果の概要（英文）：
We proposed a blind dereverberation method based on spectral subtraction using a multi-channel least mean square algorithm (MCLMS). This method was evaluated in a simulated and real noisy reverberant environment with stationary noise. In this study, we also evaluate this method in a noisy reverberant environment with non-stationary noise like music. After suppressing the music, using a blind source separation based on Efficient FastICA (independent component analysis) algorithm, spectral subtraction based dereverberation method is employed to reduce late reverberation. The proposed method achieves an average relative word error reduction rate of 41.9% and 7.9% compared to baseline method and the state-of-art multi-step linear prediction (MSLP) based dreverberation in a real environment, respectively.

交付決定額

（金額単位：円）

| | 直接経費 | 間接経費 | 合　計 |
|---|---|---|---|
| 2010 年度 | 1,300,000 | 390,000 | 1,690,000 |
| 2011 年度 | 900,000 | 270,000 | 1,170,000 |
| 2012 年度 | 800,000 | 240,000 | 1,040,000 |
| 年度 | | | |
| 年度 | | | |
| 総　計 | 3,000,000 | 900,000 | 3,900,000 |

研究分野：音声情報処理
科研費の分科・細目：知覚情報処理・知能ロボティクス
キーワード:一般化スペクトルサブトラクション、ハンズフリー音声認、missing feature theory、
マルチチャンネル LMS、ブラインド残響除去

１．研究開始当初の背景

多くのアプリケーションにおいて自然で使いやすい音声インタフェースを構築するためには、マイクから離れた発声を可能とするハンズフリー音声認識を用いた対話インタフェースが必要となる。遠隔環境下では、さまざまな距離からの音声を認識する必要があり、直接音の減衰や反射音の重畳、背景雑音の影響により性能低下を招く。

このため、近年、実環境におけるハンズフリー音声処理／認識に関する研究が行われている。例えば、中村ら（奈良先端大（当時））の研究テーマ科研基盤（B）（平12～14)「ハンズフリー音声認識」では、遠隔発話の音声認識に関する様々な先行研究が行われた。ヨーロッパの十数大学と研究機関で共同参加するCHILプロジェクトでは、スマートルームにおける話者位置推定・追跡、音声／話者認識などの研究も行われている。

しかしながら、ハンズフリー音声認識の性能はまだ不十分であり、実用化は困難な状況である。これに対して本研究は、信号処理の観点から以上の問題を定式化して解決することによって、実世界環境下で雑音・残響特性の動的変化により高精度な音声認識を行うものである。

２．研究の目的

（１）実環境での高性能な残響処理：　実環境下での音の生成を定式化し、伝送路の伝達特性（残響特性）を自動的に推定し、異なる残響特性（異なる残響時間や部屋）に対して頑健な残響除去を行う。さらに、提案手法と雑音抑圧に有効なミッシングフィーチャ理論（雑音で歪んだ周波数帯域をマスクする理論）を効率的に融合し、補正した音声の信頼できる成分のみから音声を回復することによって、高精度なハンズフリー音声認識の研究を行う。

（２）定常雑音と残響の同時処理：　定常雑音と伝送路の伝達特性が影響する場合、加算性雑音と乗算性雑音（残響）の特性を考慮し、加算性雑音を抑圧手法と提案する残響をブラインドに除去する手法と統合し、残響や加算性妨害雑音（定常）を同時に除去することによる音声認識の研究を行う。

（３）定常雑音と非定常雑音と複数音源からの残響の同時処理：　各時刻で音源数の自動推定および残響音声の自動検出によって、定常／非定常雑音と複数音源からの残響が同時に存在しても、提案法を厳密に定式化するように拡張する。ある時刻では、ある音源位置だけから発話していることを仮定すると、この音源からマイクロフォンまでの伝達特性が推定できる。同様な方法を使って、複数の音源からマイクロフォンまでの伝達特性も求められる。定常雑音を補正した後、複数音源からの補正パラメータを利用し音声を補正すれば、非定常雑音だけを含む音声が正確に求められる。そして、実環境下での雑音・残響の動的特性を従来よりも厳密に定式化して、非定常雑音抑圧手法と本提案手法を統合し、定常／非定常雑音と複数音源からの残響のすべてを推定し、頑健なハンズフリー音声認識を行う。

３．研究の方法

（１）スペクトルサブトラクションを用いて異なる残響特性の違いに頑健な残響補正：既に、本研究発足のための先行研究として、インパルス応答の後部残響の影響を加算性雑音と見なし，スペクトルサブトラクションを使って，残響音声とインパルス応答のパワースペクトルを用いてクリーン音声のパワースペクトルを推定する方法を提案してきた。提案法は長い残響時間の音声に対しては認識率の改善が得られたが、短い残響時間の残響音声に対しては逆に音声が劣化した。この原因は、全ての残響音声に対して、スペクトル減算のため考慮する残響のフレーム数を同一にしたためである。そこで、自動推定した残響時間に応じて、異なる音声の長さ（フレーム数）を利用し、残響を補正する。また、従来の提案手法では、異なる残響による初期反射(early reverberation)の時間も考慮していなかった。後部残響（late reverberation）は音声認識への悪影響が大きい事実を利用し、初期反射時間を自動推定して後部残響だけを補正し、様々な残響環境に頑健な残響補正法を目指す。

（２）ミッシングフィーチャ理論を用いる残響補正：推定するインパルス応答の長さが実際のインパルス応答長より短いことやインパルス応答のパラメータの推定誤差などの原因で、ある区間のある周波数範囲でうまく補正できない場合もあり得る。ミッシングフィーチャ理論は、加算性雑音に対して効果がある。一方、乗算性雑音（残響）は、現在の信号は前の信号の反射などが加わり、全周波数に影響が与えられるため、ミッシングフィーチャ理論は直接に利用できない。本研究では、まずスペクトル減算によって残響を補正し、前時刻の信号の影響を軽減してから、各時刻の周波数毎に SIR（Signal-to-Interference Ratio）を自動的に算出し、SIR の値や補正されたスペクトルの振幅（負の振幅の信頼性を低減する）によってスペクトルをマスクする。

（３）残響と定常雑音の同時補正：　本研究では、加算性雑音も考慮する。加算性雑音が定常雑音の場合、まず、文先頭の無音区間を利用し加算性雑音のスペクトルを推定し、加算性雑音を除去する。次に、上記の方法を用いて残りの残響を補正することによって、二段階で雑音や残響を補正する。さらに、音声を正確に補正するために、上述の処理を繰り返して除去する方法も考えている。

（４）複数音源の残響と非定常雑音／定常雑音の同時補正：　定常／非定常雑音と複数音源からの残響が同時に存在しても、厳密に定式化するように提案法を拡張する。定常雑音を除去してから、音源分離により非定常雑音と残響音声を分離する。分離後の残響音声に対して残響除去を行う。

４．研究成果

（１）スペクトルサブトラクションを用いて異なる残響特性の違いに頑健な残響補正：既に、本研究発足のための先行研究として、インパルス応答の後部残響の影響を加算性雑音と見なし、スペクトルサブトラクションを使って、残響音声とインパルス応答のパワースペクトルを用いてクリーン音声のパワースペクトルを推定する方法を提案してきた。平成22年度で、大語彙連続音声認識による評価とこの手法に用いられるパラメータ変化による影響分析や改善手法の効果を比較評価した。提案法は様々な残響環境やタスクに対して頑健な結果が得られた。結果を図1に示す。

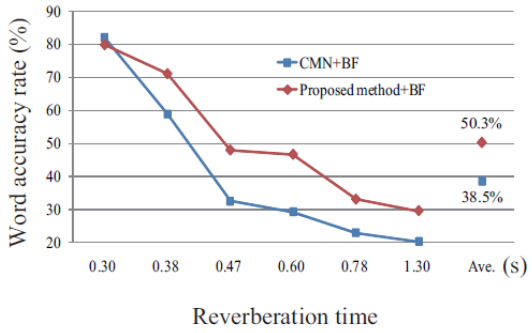


図1、提案手法を用いた様々な残響時間における単語認識率。(BF: Beamforming; CMN: Cepstral Mean Normalization)

（２）ミッシングフィーチャ理論を用いる残響補正：推定するインパルス応答の長さが実際のインパルス応答長より短いことやインパルス応答のパラメータの推定誤差などの原因で、ある区間のある周波数範囲でうまく補正できない場合もあり得る。本研究では、まずスペクトル減算によって残響を補正し、前時刻の信号の影響を軽減してから、各時刻の周波数毎にSRR（Signal-to-Reverberation Ratio）を自動的に算出し、SRRの値から計算したスペクトルの信頼度を補正されたスペクトルにかけることで重み付けを行う。従来法より良い認識性能が得られた。

（３）一般化ＳＳによる残響除去：　先行研究において、任意の指数パラメータを与える一般化 SS はパワーSS より効果的な雑音抑圧法であることが示されている。本研究では、一般化 SS を用いて後部残響を除去する方法を提案した。一般化ＳＳ（GSS）は残響環境下において、パワーＳＳと従来法より大幅な性能を改善した。

（４）残響と定常雑音の同時補正：　本研究では、加算性雑音も考慮する。加算性雑音が定常雑音の場合、まず、文先頭の無音区間を利用し加算性雑音のスペクトルを推定し、加算性雑音を除去する。次に、上記の方法を用いて残りの残響を補正することによって、二段階で雑音や残響を補正する。

（５）実環境での評価：実環境の残響を含んだマルチチャンネル残響音声を収録し、残響除去法の評価に用いた。実験の結果として，発話単位 CMN を利用しただけの場合と比べて，残響除去法を適用することで、使用したチャンネルの組み合わせ全てに対して大きな性能改善が見られた。

表１、実環境における雑音残響音声の認識結果。(DN: Denoising; DNR: Denoising and dereverberation)

| Speakers / Position | CMN only | Power SS | | GSS | |
|---|---|---|---|---|---|
| | | DN | DNR | DN | DNR |
| A | 60.2 | 67.7 | 78.9 | 64.7 | 79.5 |
| B | 75.6 | 72.2 | 78.5 | 72.5 | 83.2 |
| C | 67.4 | 63.2 | 69.4 | 66.7 | 77.5 |
| D | 59.1 | 53.9 | 74.9 | 60.8 | 78.7 |
| E | 42.9 | 51.0 | 62.8 | 50.0 | 61.7 |
| Average | 60.9 | 61.6 | 73.1 | 62.9 | 76.2 |

（６）非定常雑音である音楽を含む残響音声に対して，本提案のマルチチャンネル最小二乗平均を基づく一般化スペクトルサブトラクション（GSS）によるブラインド残響除去法と ICA（独立成分分析）に基づくブラインド音源分離を組み合わせる方法を提案しました。本研究では，ICA の代表的なアルゴリズムである FastICA を改善した Efficient FastICA (EFICA)を用いる。まず，EFICA に基づく音源分離によって音楽と音声を分離する。その分離音声から推定したインパルス応答を用いて GSS に基づく残響除去を適用し，後部残響を除去する。その後，特徴量抽出時の CMN によって初期残響の影響を除くように正規化する。この方法を評価するために，残響環境下において非定常的な雑音である音楽が背景雑音として重畳された音楽重畳音

声を用いる。SNR（信号雑音比）を変化させ人工的に作成した音楽重畳音声と実環境で収録した音楽重畳音声に対してこの手法を評価した。人工環境で，全てのSNRで音源分離と残響除去の適用によって大幅な改善が見られた。従来法に比べ，SNR 20 dB，10 dB，0 dB のときのエラー削減率はそれぞれ44.2%，48.9%，24.9%を達成した。実環境で，従来法に比べ 41.9%のエラー削減率を達成した。この結果はSNRが同程度である 0 dB と 10 dB の人工音楽重畳音声のときのエラー削減率に匹敵し，本手法は実環境で収録した音楽重畳音声に対しても有効であることが分かった。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　5件）


1. S. Nakagawa, L. Wang and S. Ohtsuka, "Speaker identification and verification by combining MFCC and phase information", IEEE Transactions on Audio, Speech and Language Processing, Vol. 20, No. 4, pp. 1085-1095, May 2012.
   DOI: 10.1109/TASL.2011.2172422
2. L. Wang, K. Odani and A. Kai, "Dereverberation and Denoising Based on Generalized Spectral Subtraction by Multi-channel LMS Algorithm Using a Small-scale Microphone Array", Eurasip Journal on Advanced in Signal Processing, 2012:12, Jan. 2012.
   DOI: 10.1186/1687-6180-2012-12
3. Y. Jiang, Z. Tang and L. Wang, "Identification of a distant speaker and its robustness", Chinese Journal of Electronics, Vol. 20, No. 2, pp. 278-282, Apr. 2011.
   http://www.ejournal.org.cn/Jweb_cje/EN/abstract/abstract1109.shtml
4. "Distant-talking speech recognition based on spectral subtraction by multi-channel LMS algorithm", L. Wang, N. Kitaoka, S. Nakagawa, IEICE Trans. on Information and Systems, Vol. E94-D, No.3, pp. 659-667, Mar. 2011.
   http://search.ieice.org/bin/summary.php?id=e94-d_3_659
5. "Speaker recognition by combining MFCC and phase information in noisy conditions", L. Wang, K. Minami, K. Yamamoto, S. Nakagawa, IEICE Trans. on Information and Systems, Vol. E93-D, No. 9, pp. 2397-2406, Sep. 2010.
   http://search.ieice.org/bin/summary.php?id=e93-d_9_2397


〔学会発表〕（計　27 件）


1. L. Wang, Z. Zhang, A. Kai and Y. Kishi, "Distant-talking speaker identification using a reverberation model with various artificial room impulse responses," Proc. of APSIPA ASC 2012, Dec. 2012 .
2. Z. Zhang, L. Wang and A. Kai, "Dereverberantion based on Generalized Spectral Subtraction for Distant-talking Speaker Recognition," Proc. of APSIPA ASC 2012, Dec. 2012.
3. Y. Hirano, L. Wang, A. Kai and S. Nakagawa, "On the Use of Phase Information-based Joint Factor Analysis for Speaker Verification under Channel Mismatch Condition,"

Proc. of APSIPA ASC 2012, Dec. 2012.
4. K. Odani, L. Wang and A. Kai, "Speech Recognition by Denoising and Dereverberation Based on Spectral Subtraction in a Real Noisy Reverberant Environment," Proc. of Interspeech 2012, Sep. 2012.
5. Kyohei Odani, Longbiao Wang and Atsuhiko Kai, "Blind Dereverberation Based on Generalized Spectral Subtraction by Multi-channel LMS Algorithm", Proc. of APSIPA ASC 2011, Oct. 2011.
6. Longbiao Wang , Kyohei Odani and Atsuhiko Kai, "Evaluation of Hands-free Large Vocabulary Continuous Speech Recognition by Blind Dereverberation Based on Spectral Subtraction by Multi-channel LMS Algorithm", Proc. of Text, Speech and Dialogue, pp. 131-138, Sep. 2011.
7. "Multimodal interface with N-best display including candidates of spoken word fragments," Y. Jang, A. Kai and L. Wang, Proc. of APSIPA ASC 2010, pp. 478-481, Dec. 2010.
8. "Compensation approaches for distant Speaker identification under reverberant environments", Y. Jiang, Z. Tang and L. Wang, Proc. of CCPR 2010, pp. 70-74, Oct. 2010.
9. Zhaofeng Zhang、Lee Kong Aik、Longbiao Wang、Atsuhiko Kai、Ma Bin、"Single-sided Approach to Discriminative PLDA Training for Text-Independent Speaker Verification"、Proc. of the 2013 Spring Meeting of the ASJ, 1-Q-46b, Mar. 2013.

〔図書〕(計 2 件)

1. Longbiao Wang, Kyohei Odani, Atsuhiko Kai, Norihide Kitaoka and Seiichi Nakagawa, "Dereverberation Based on Spectral Subtraction by Multi-channel LMS Algorithm for Hands-free Speech Recognition", Chapter in Modern Speech Recognition Approaches with Case Studies, S. Ramakrishnan (Eds.), IN-TECH, ISBN 978-953-51-0831-3, pp. 155-174 (2012).
2. Longbiao Wang, Kyohei Odani and Atsuhiko Kai, "Evaluation of hands-free large vocabulary continuous speech recognition by blind dereverberation based on spectral subtraction by multi-channel LMS algorithm", Ivan Habernal, Václav Matousek (Eds.), Lecture Notes in Artificial Intelligence, Springer LNAI6836, ISBN 978-3-642-23537-5, pp. 131-138, 2011.

〔その他〕
ホームページ等
http://sip.nagaokaut.ac.jp/wang-j.html

６．研究組織
(1)研究代表者
王　龍標 (WANG LONGBIAO)
長岡技術科学大学・産学融合トップランナー養成センター・産学融合特任准教授
研究者番号：30510458
(2)研究分担者 なし

(3)連携研究者 なし