

実世界環境下における頑健なハンズフリー音声認識

Hands-Free Speech Recognition in Real World

H24海自08

派遣先 シンガポール国立情報通信研究所 (シンガポール)

期 間 平成24年8月26日～平成24年9月2日 (8日間)

申請者 長岡技術科学大学 特任准教授 王 龍 標

海外における研究活動状況

研究目的

実環境下では反射による残響や定常／非定常雑音などの動的な変動音が原音声に混入するため、ハンズフリー音声認識が非常に難しくなる。これに対して本共同研究は、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、実世界環境下における頑健な音声認識と話者認識の研究を行う。

海外における研究活動報告

2012年8月26日にシンガポールに到着し、27日から31日までに訪問先であるInstitute for Infocomm ResearchのHaizhou Li先生、Dr. Bin Maなどと実環境における遠隔発話の音声認識に関する共同研究の打合せを行いました。その後、9月1日の便で日本に戻った。日本に戻ってから、音声認識の共同研究を続けている。

最近、スマートフォンの普及により、音声認識技術が注目を集めている。スマートフォンは接話型マイクロフォンを用いるため、近接発話音声認識となる。近接発話環境下では周囲の雑音や残響の影響を受けにくいいため、現在の音声認識技術によって高精度の認識が可能である。しかし、ハンズフリー音声認識のようなマイクロフォンと音源の距離が遠い遠隔発話環境下では、周囲の雑音や残響の影響を多大

に受けてしまうため、精度の高い認識を行うことができない。もし、ハンズフリー音声認識システムが実現すれば、マイクロフォンの位置や手の拘束を受けずに音声認識を行うことが可能になる。これによって、カーナビゲーションシステムや自動議事録作成の使い勝手がよくなることが期待できる。遠隔発話環境で収録された歪んだ音声は、学習時と認識時の収録環境にミスマッチを生じ、認識精度を著しく低下させるという問題がある。したがって、このミスマッチを低減させることで環境変動に頑健な音声認識技術が重要な課題となっている。

そこで、本研究はマルチチャンネル最小平均二乗 (Multi-channel Least Mean Square: MCLMS) アルゴリズムに基づくSS (Spectral Subtraction) 法を用いた残響除去法という新たな残響除去技術を提案した。これは、スペクトル領域における後部残響が付加雑音とみなし、MCLMSアルゴリズムに基づいて推定したインパルス応答から得られる後部残響を雑音除去技術であるSS法を用いて除去する方法である。MCLMSアルゴリズムは、時間領域のインパルス応答を自動的に識別するために提案された方法であるが、処理時間が非常に長く、推定精度もあまり良くない。そのため、本研究では、MCLMSアルゴリズムを周波数領域へ拡張し、インパルス応答のスペクトルを推定している。また、MCLMSに基づく残響除去法の場合、

雑音の影響によってインパルス応答の推定精度が低下してしまう問題も考えられるため、残響を処理する前に雑音を除去する必要がある。雑音には、エアコンやコンピュータのように時間的にパワーが変化しない定常雑音と音楽や人の声のように時間的にパワーが変化する非定常雑音の2種類あり、それらに対する雑音除去技術が提案されている。一般的に、定常雑音除去技術としてSS法に基づく雑音除去法が、非定常雑音除去技術としてICA (Independent Component Analysis) に基づくブラインド音源分離が用いられており、これらの雑音除去技術と併用することで、雑音と残響に対応することができる。

本研究では、一般化SS法を残響除去法に適用した改善手法や従来の雑音除去技術との併用手法を提案し、実環境を含む様々な環境下における音声を用いて評価を行う。図1は提案手法の流れである。マルチチャンネル残響音声声を離散フーリエ変換で得られたスペクトルを用いて周波数領域のマルチチャンネル最小二乗平均 (LMS) でインパルス応答のスペクトルを推定する。このように推定したインパルス応答と推定したクリーン音声により求める後部残響を一般化スペクトル減算 (SS) 法で残響音声

から除去し、逆離散フーリエ変換で音声信号に戻す。後処理として、遅延和ビームフォーミングを適用することで指向性音声を強調し、ケプストラム平均正規化 (CMN) で初期残響を正規化したメル周波数ケプストラム係数 (MFCC) を出力する。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) や独立成分分析を効率的に融合し、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った。

提案手法の評価を行う前に、先行研究の残響除去法を大語彙連続音声認識タスクで評価し有効性を確認するとともに、MCLMSで推定するインパルス応答の長さやインパルス応答推定に使用するマイクロフォンの数の最適な条件を見つける。また、GSS法を用いた雑音除去法の音声認識による評価を行い、音声認識性能に関しても、パワーSSより効果的であることを確認する。

実環境で収録した雑音残響音声に対して、提案法は音声認識率を従来法の50%から実用化できる世界トップレベルの82%までに改善した。

この派遣の研究成果等を発表した 著書、論文、報告書の書名・講演題目

[国際会議]

L. Wang, K. Odani, A. Kai and W. Li, "Speech Recognition Using Blind Source Separation and Dereverberation Method for Mixed Sound of Speech and Music," Proc. of APSIPA ASC 2013, 2013 (Accepted).

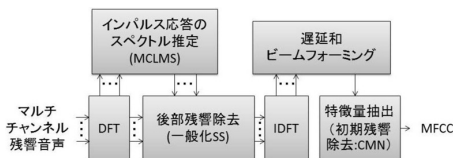


図1 一般化スペクトル減算による残響除去