

平成26年度 メディア科学専攻修士論文要旨

大西 研究室	氏 名	出木浦 悠人
修士論文題目	同時発話の高性能な音声認識 -スペクトル減算による分離と分離音を用いた音響モデル学習-	

音声認識技術は人の声をコンピュータに認識させる技術である。本研究では、同時発話の高性能な音声認識を目的とする。現在の音声認識システムは1人の発話を対象としたものが多く、その認識性能は非常に高水準となっているが、同時発話認識においては他人の発話の重複によって認識対象とする音声不明瞭となり、認識率は大幅に低下する。

同時発話認識のためには音源分離が必要であるが、その過程で生じる音響的な歪みや残留雑音により、認識精度の改善幅はある水準に留まる。また、従来用いられてきた独立成分分析(ICA)による分離は、処理時間が長いことも問題である。そこで本研究では、スペクトル減算法を用いた音源分離手法、歪みを含む分離音を用いて音響モデル学習を行うことを提案する。

前者は、一般的に定常雑音の抑圧に用いられ、計算量は少ないが比較的良好な結果が得られるという特徴がある。今回は、混合音声から他者発話スペクトルを推定し、観測信号から減算することで分離する。

後者では、あらかじめ話者-マイク間の伝達関数を収録しておくことで、分離音(学習データ)を多数生成できることを利用した手法である。

計算機シミュレーションによる実験と実際の発話による実験(2~4話者)で、提案法による分離音の音声認識精度は従来法に比べて約5%~20%改善し、処理時間が約10%~50%に短縮された。また、分離音を学習データに用いることにより、認識精度がさらに約10%~20%改善されることを示した。図1と図2に、2音源の場合についてそれぞれの実験結果を示す。

今後の課題として、伝達関数推定に用いていた単独発話区間の自動推定や、自由発話でシステムの性能評価を行うことが挙げられる。

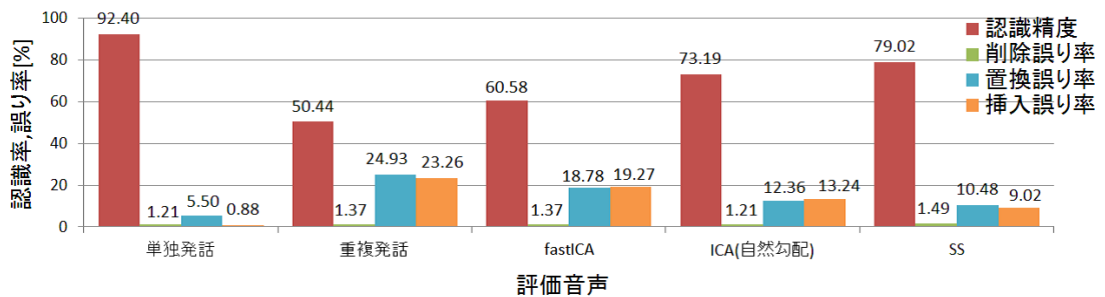


図1 音声認識精度 (既存の音響モデル)

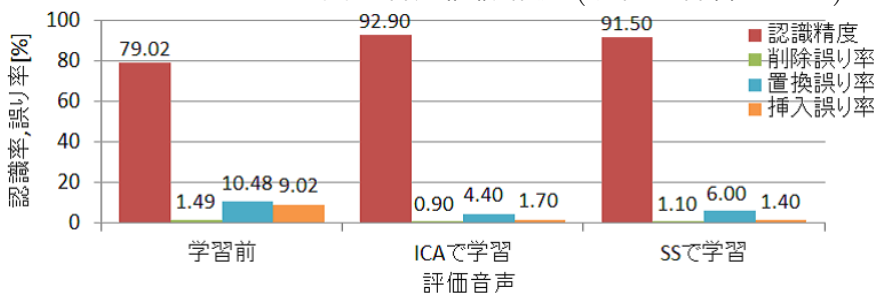


図2 音声認識精度 (分離音で学習したモデル)