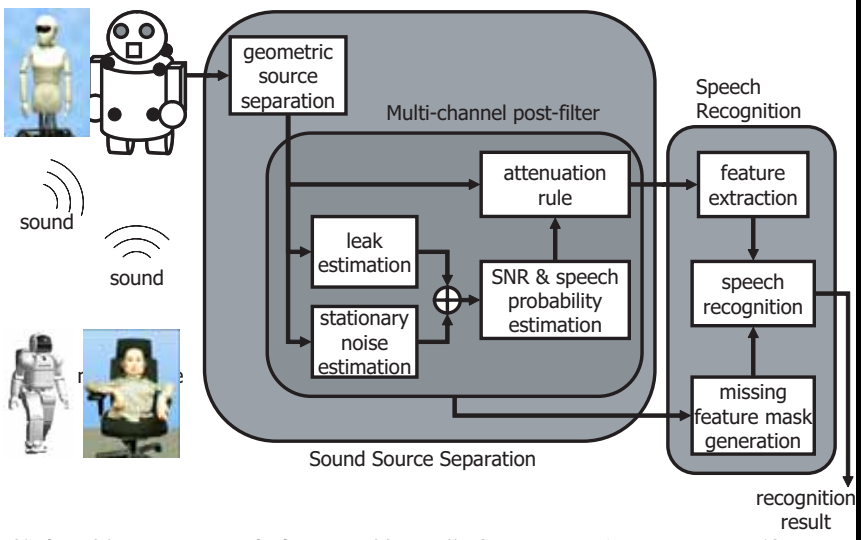
混合音分離と音声認識を統合したミッシングフィーチャ理論による音声認識手法の開発（人を観る）

研究担当者：奥乃　博（京都大学大学院・情報学研究科）
研究代表者：奥乃　博（京都大学大学院・情報学研究科）[公募研究]
研究期間：平成１５年度～平成１７年度

研 究 成 果 概 要

【目的】

ロボットが動的な環境で複数の人との間で高度なコミュニケーションを行う上で、聖徳太子のように複数同時発話を認識できる機能が不可欠である。ロボット聴覚に必要な機能は、音源定位、音源分離、および、分離音認識である。とくに、分離音認識では、他人の話し声が雑音になるという従来研究での想定外の動的変化の大きい雑音が課題となる。さまざまな音響的環境でロボットは使用されるので、対象環境を想定した multiconditioning 学習による音響モデルではなく、単一音響モデルを使用した音声認識が不可欠である。また、提案手法が形状の異なるロボットに対してもうまく機能するという汎用性も、ロボットヒューマンインタラクションの観点からは重要である。



【技術内容】

上記の課題に対して、以下の3つの技術に取り組んだ：

①ミッシングフィーチャ理論に基づいた音源分離と音声認識の統合手法

②同統合法の汎用性の検証

③マスク自動生成

ミッシングフィーチャ理論による音声認識では、マスクされる特徴量を使用せずに尤度計算を行い、音声認識を行う。まず、2本のマイクロフォンを使用した既開発の ADPF (Active Direction-Pass Filter) の分離音に対してクリーン音声との比較から作成されるアプリオリマスクを使用したシステムを構築した。さらに、形状の異なる3体のロボット（ASIMO, SIG2, Replie）に実装し、評価を行った。3話者同時発話に対して、クリーン音声で学習したトライフォンの音響モデルを使用した語彙サイズ200語の孤立単語認識実験により、3話者同時発話認識の単語正解率の平均は ASIMO, SIG2, Replie の場合、それぞれ, 79.7%, 78.7%, 82.7% であった。

次に、音源分離過程での情報を利用したマスク自動生成に取り組んだ。ただし、2本のマイクロフォンから得られる情報では不十分であるので、8本のマイクロフォンアレイを使用した。音源分離には, Geometrical Source Separation と multi-channel post-filter を使用し, 後者から得られるチャネル間リーク情報と背景雑音情報を基にマスクを自動作成する. 自動生成されたマスクを使用し, マルチバンド版 Julius を用いて認識を行った。 ここで、特徴量をスペクトル歪みに強い MSLS とした。同じベンチマークにより、アプリオリマスクの場合と比較し、約62%の性能を達成した。

【結論】

ミッシングフィーチャ理論による音声認識はマスク自動生成が困難であり、文献ではほとんど成功していなかった。これに対して、音源分離情報を活用することにより、マスク自動生成が可能となり、聖徳太子ロボットの基盤技術が確立できた。今後、マスク自動生成の高性能化、距離や方向を勘案した音源分離技術の向上を通じて、より高度なロボットヒューマンインタラクションを具体的に実現していく予定である。

論 文 発 表 等

1. 山本 俊一, 中臺 一博, 辻野 広司, 奥乃 博: ミッシングフィーチャ理論を利用した音源分離と音声認識のインターフェースと複数ロボットへの適用, *日本ロボット学会誌*, Vol.23, No.6 (Aug. 2005) 743-751.
2. Shun'ichi Yamamoto, Jean-Marc Valin Kazuhiro Nakadai, Hiroshi Tsujino, Jean Rouat, Francois Michaud, Tetsuya Ogata, Kazunori Komatani, Hiroshi G. Okuno: Enhanced Robot Speech Recognition Based on Microphone Array Source Separation and Missing Feature Theory. *Proceedings of IEEE-RAS International Conference on Robots and Automation (ICRA-2005)*, 1489-1494, IEEE, Barcelona, Apr. 2005.
3. Shun'ichi Yamamoto, Kazuhiro Nakadai, Jean-Marc Valin, Jean Rouat, Francois Michaud, Tetsuya Ogata, Kazunori Komatani, Hiroshi G. Okuno: Making A Robot Recognize Three Simultaneous Sentences in Real-Time, *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2005)*, pp.897-892, IEEE, RSJ, Edmonton, Aug. 2005.