

# 脳性麻痺障がい者のマルチモーダルコミュニケーション支援技術の研究開発（112307005）

Multimodal assistive technologies for people with articulation disorders

## 研究代表者

滝口哲也 神戸大学

Tetsuya Takiguchi Kobe University

## 研究分担者

有木康雄

Yasuo Ariki

神戸大学

Kobe University

研究期間 平成 23 年度～平成 24 年度

## 概要

本提案課題では、平成 24 年度までに騒音を含む環境において、発話の聞き取りが困難な脳性麻痺障がい者の音声コミュニケーションが実現可能となるように、脳性麻痺障がい者の機械学習法を用いた音声認識における音声特徴量抽出法、音響モデルの構築、更に聞き取りが困難な音声を健常者の声を活用した聞き取りやすい声への声質変換、リップリーディング情報の統合による騒音下音声認識手法の研究に取り組む。

## 1. まえがき

近年、情報技術が向上し、福祉分野への情報技術の適用が行われている。例えば、画像認識技術を用いた手話認識や、文書内の文字の音声化などが行われている。また、音声合成を用いて、発話障害者支援のための音声合成器の作成なども行われている。

音声認識技術は近年、飛躍的に進歩し、様々な環境や場面での利用が期待されている。例えばカーナビゲーションの操作や会議音声の議事録化など様々な分野に応用されている。しかしながら、言語障がいなどの障がい者を対象としているものは非常に少ない。

本研究は、ユビキタス社会における脳性麻痺障がい者の自立生活を情報通信技術に支援する新しいユニバーサルデザインの研究開発を目的としている。脳性麻痺障がい者の発話は、その発話内容を聞き取る事が困難な場合があり、地域社会から取り残される事がある。本システムが具体化されて、地域生活において活用されることにより、全ての人々が共存したコミュニケーション社会の構築が可能となる。本稿では、ランダムプロジェクションを用いた構音障がい者の音声認識、唇特徴と音声特徴の統合による発話認識、更に構音障がい者の声質変換の結果について述べる。

## 2. 研究開発内容及び成果

### 2.1 ランダムプロジェクションを用いた構音障がい者の音声認識

ここでは、アテトーゼ型の脳性麻痺による構音障がい者を対象とした音声認識の実現を目指している。彼らは意図的な動作時や緊張状態にある場合に筋肉の制御が難しくなり、アテトーゼと呼ばれる不随意運動を伴う。アテトーゼ型の構音障がい者の発話スタイルは健常者と大きく異なり、認識精度が著しく低下する。

ランダムプロジェクションとは、空間写像の一手法で、その変換写像行列の各要素がある確率分布に従うランダムな値として定義される点に特徴を持つ。

提案手法では、複数のランダム写像行列を用いて音声特徴量を変換する。各々の特徴量を用いて音声認識を行い、各認識結果を投票により統合することで最適な認識結果を得る（図 1）。

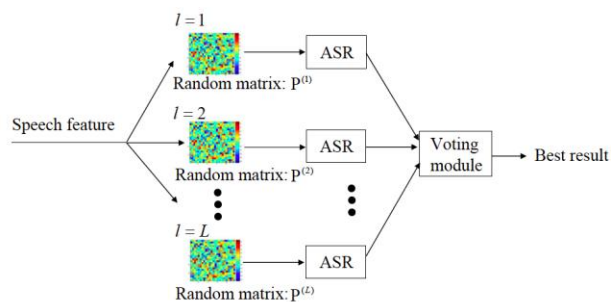


図 1 ランダム写像と投票統合

構音障がい者 1 名を対象とした孤立単語認識実験（215 単語認識タスク）を行った。音声の標本化周波数は 16kHz、音響モデルは monophone HMM (Hidden Markov Model) で、各 HMM の状態数は 5、状態あたりの混合分布数は 8 である。

ランダム写像行列の数を 20、40、60、80、100 と変化させて実験を行った。語認識実験の結果を図 2 に示す。図 2 より、提案手法を用いて認識結果を統合することで、安定して高い認識率が得られることが示された。また、統合特徴量数を増やすことで認識率が上がる傾向があるが、（今回の実験では）元特徴量と比べて、20～40 個程度の統合で十分な認識率が得られている。

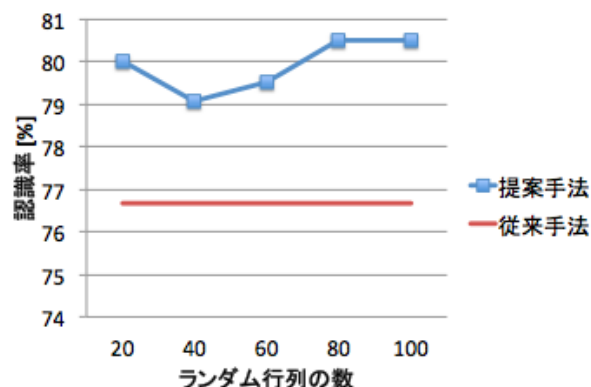


図 2 ランダム写像を用いた音声認識実験

## 2.2 唇特徴と音声特徴を用いた構音障がい者の音声認識

アテトーゼ型の構音障がい者の場合、筋肉の緊張のため発話が不安定になりやすく、発話時に頭が動いてしまう場合がある。本研究では発話時の頭部の動きに対して、AAM (Active Appearance Model)を用いることで画像から顔方位にロバストな唇領域特徴を抽出し、音声特徴と共に用いることで、雑音の影響を受けず発話変動を考慮したマルチモーダル音声認識を検討した。

音声特徴量、画像特徴量をそれぞれ用いて、音声 HMM、画像 HMM を構築する。音声 HMM を画像 HMM と統合することで、音響的な雑音にロバストな認識が可能であるだけでなく、雑音がない環境下においても、構音障害者の音声認識率が低下するという問題に対する精度の改善が期待できる。認識時において、両 HMM の尤度の重み付き線形和を用いて統合を行う。

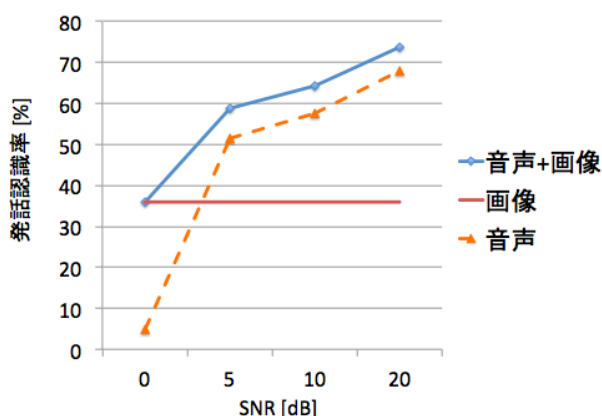


図3 音声と画像の統合による発話認識結果

図3に、音声特徴量と画像（唇）特徴量の統合による発話認識実験の結果を示す。画像情報との統合によって、全ての雑音条件（音声と背景雑音のSNR (Signal to Noise Ratio)を0、5、10、20 dBとした）に対して音声のみを用いた時よりも認識率が改善された。画像情報を用いることで、雑音の影響を受けず、発話変動を考慮することができ、認識率の改善が得られたと考えられる。

## 2.3 構音障がい者の声質変換

本研究では、脳性麻痺の一種であるアテトーゼ型構音障がい者を対象とした話者性を維持した声質変換を提案する。アテトーゼ現象は意図的な動作に緊張状態を発生させるために、障がい者の発話、特に子音が不安定になる。ここでは、非負値行列因子分解 (Non-negative Matrix Factorization: NMF)を用いた Exemplar-based な声質変換を構音障がい者の発話に適用し、不安定な発話音声をより聞き取りやすく変換することを目指す。従来の統計モデルを用いた声質変換技術は、主として話者変換を目的としていたため、入力話者の声質は完全に別の話者の声質に変換されてしまう。

本研究では、障害者の母音と健常者の子音を組み合わせた Combined Dictionary を用いることで、入力障がい者音声の話者性を維持しつつ、より聞き取りやすく変換することを可能にした。

図4に「話者性」と「自然性」の評価結果を示す。提案手法である NMF に基づく声質変換は、従来手法と比較して話者性を維持できていることがわかる。これは、従来手法は全ての入力スペクトル包絡を健常者のものに変換しているのに対し、提案手法では Combined dictionary を用いて子音のみの変換を行っているためである。自然性に

ついても、提案手法は異なる話者の子音と母音を組み合わせているのにも関わらず、従来手法よりも高いスコアを得ている。これは、従来手法の変換ノイズの問題に加え、提案手法は母音部分をほぼ無変換で合成できるためと考えられる。

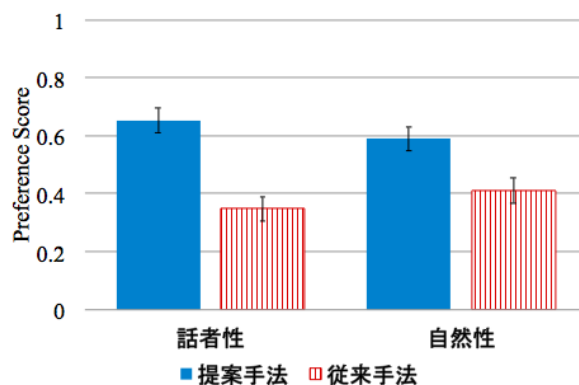


図4 声質変換の主観評価結果

## 3. 今後の研究開発成果の展開及び波及効果創出への取り組み

近年、家庭生活、社会生活などにおいて様々な機器の情報化が進み、情報機器が身の回りの生活環境にて浸透しつつある。しかし、そのような機器は操作が複雑であり、高齢者や障がい者が使いこなすには困難である場合が多い。

本コミュニケーション支援技術が具体化されれば、障がい者のコミュニケーション機能を増大させることはもとより、障がいの無い人々と同等の社会活動の参加に貢献するものであると期待出来る。

## 4. むすび

本稿では、ランダムプロジェクションを用いた構音障がい者の音声認識手法を提案し、音声認識実験により有効性を示した。また音声情報と画像情報の統合によるマルチモーダル発話認識手法を検討し、雑音環境下において、音声認識率の改善が得られた。更に、構音障がい者の話者性を維持した（聞き取りやすい）声質変換技術を提案し、聴取実験により有効性を示した。今後は話者数を増やして提案手法の有効性を確認し、システムの実用化を目指して行く。

### 【誌上発表リスト】

- [1]Tetsuya Takiguchi, Mariko Yoshii, Yasuo Ariki, and Jeff Bilmes, "Acoustic Model Transformations Based on Random Projections," IEEE International Conf. on Acoustics, Speech, and Signal Processing, pp. 1933-1936 (2012年3月29日)
- [2]Ryo Aihara, Ryoichi Takashima, Tetsuya Takiguchi, and Yasuo Ariki, "Consonant Enhancement for Articulation Disorders Based on Non-negative Matrix Factorization," APSIPA, 4 pages (2012年12月4日)
- [3]Ryo Aihara, Ryoichi Takashima, Tetsuya Takiguchi, and Yasuo Ariki, "Individuality-Preserving Voice Conversion for Articulation Disorders Based on Non-Negative Matrix Factorization," IEEE International Conf. on Acoustics, Speech, and Signal Processing, pp. 8037-8040 (2013年5月26日)