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 5 月 24 日現在

機関番号：14501
研究種目：若手研究(A)
研究期間：2009～2012
課題番号：21680054
研究課題名（和文）　脳性麻痺構音障がい者の発話スタイルの解析及びハンズフリーコミュニケーションの研究
研究課題名（英文）　Analysis of articulation disordered voice and research of hands-free communication
研究代表者　滝口　哲也（TAKIGUCHI TETSUYA）
神戸大学・都市安全研究センター・准教授
研究者番号：40397815

研究成果の概要（和文）：
脳性麻痺構音障がい者の音声コミュニケーションの実現を目指し，機械学習法を用いた構音障がい者の音声特徴量抽出法を提案した．また，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．これに対して，顔方位に頑健な発話認識法を提案した．更に声質変換法を用いて構音障がい者の音声を健常者の音声に変換し，子音を強調する手法を提案した．

研究成果の概要（英文）：
For speech communication of persons with articulation disorders resulting from cerebral palsy, we proposed a robust feature extraction method using a machine learning algorithm. Also, Also, in the case of a person with an articulation disorder, there may be a problem due to the tendency of his/her erratic head movement. We investigated a pose-robust audio-visual speech recognition method to solve this problem. Also, we presented consonant enhancement on a voice for persons with articulation disorders.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	6,100,000	1,830,000	7,930,000
2010 年度	5,900,000	1,770,000	7,670,000
2011 年度	5,200,000	1,560,000	6,760,000
2012 年度	3,500,000	1,050,000	4,550,000
年度			
総　計	20,700,000	6,210,000	26,910,000

研究分野：総合領域
科研費の分科・細目：科学教育・教育工学、教育工学
キーワード：ヒューマン、インターフェイス

１．研究開始当初の背景

近年，音声認識技術の発展に伴い，様々な環境下や場面での利用が期待されている．また，これまでは成人を対象とした音声認識が多くみられたが，最近では成人だけでなく高齢者や子供など成人と発話スタイルが異なる人も対象としており，様々な人が利用する機会が増えている．しかし，健常者とは発話スタイルが異なる構音に障がいを持つ人を対象とした音声認識は非常に少なく，手足の不自由を患っている場合など音声に頼るほかない場合，構音障がい者を対象とした音声認識の実現が期待される．

本研究では，構音障がいを持つ脳性麻痺の被験者を対象としている．多くの場合，発話が困難であることと共に手足が不自由である．発話は自分の気持ちをよく表す手段であり，手足が不自由な場合，発話は重要なコミュニケーション手段の一つであると考えられ，障がい者における音声認識の実現が期待される．音声認識が実現し，発話内容が分かれば，会話時にお互いの理解がより深まり，また障がい者の就業機会の増加や講演時の補助等への活用などが期待される．

更に，情報機器と人（ユーザ）との音声コミュニケーションを実現する上で，ユーザがマイクロホンから離れて発話するため，ハンズフリーなインターフェースに関する研究開発が必要となる．しかしこれまでの研究の多くは，マイクロフォンスイッチが必要であり，つまりユーザがマイクロホンの位置を常に意識し，マイクロフォンスイッチを押してから，発話を開始する必要があった．本研究では，ハンズフリーインターフェースを実現するための要素技術の一つである，音源位置推定手法についても提案する．

２．研究の目的

(1) 構音障がい者の音声認識における動的特徴量の考察

構音障がい者の発話スタイルは，筋肉の緊張のため，健常者と大きく異なり不安定であるため，特定話者モデルでの音声認識には限界がある．特に構音障がい者の動的特徴量(デルタケプストラム）の認識精度は健常者に比べて大きく低下する．これは構音障がい者において，時間変化を十分に表現できていないと言える．本研究では，この問題点に着目し，構音障がい者の音声認識精度の改善を行うため，セグメント特徴量を用いた音声特徴量抽出法を提案する．動的特徴量の代わりに，デルタケプストラム係数のセグメント特徴量を用いることで，より時間特徴を表現する．

(2) スパース表現に基づく構音障がい者の発話スタイル変動にロバストな特徴量抽出

構音障がい者は発話スタイルが健常者と大きく異なるため，従来用いられている健常者の不特定話者モデルでは認識が困難である．また，構音障がいの原因であるアテトーゼ型脳性マヒによる不随意運動のため，最初の発話スタイルが最も不安定になるという課題がある．そこで，構音障がい者の特定話者モデルを作成することで認識率の改善を行い，最初の発話スタイルが不安定になるという課題に対して，非負値行列因子分解(Non-negative matrix factorization: NMF)をベースとした，発話スタイル変動にロバストな特徴量抽出法を提案する．

(3) 非負値行列因子分解による構音障がい者の声質変換

アテトーゼ型脳性麻痺による構音障がい者の発話の特徴として，音声の子音が不明確になることを挙げることができる．アテトーゼ現象により，子音を発音する際の筋肉の動きが制限されるためにおこる．本研究では，声質変換技術を構音障がい者に適用し，音声の子音強調を行う．アテトーゼ型脳性麻痺による構音障がい者の多くは，身体が不自由であるため，手話や文章読み上げ装置を使うことは困難である．そのため，構音障がい者のための声質変換には十分なニーズがあり，研究の必要性があるといえる．

(4) 顔方位にロバストな唇領域特徴抽出と音声特徴による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．これに対して，発話時の頭部の動きに対しては，Active Appearance Model (AAM) 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する．

(5) 音響伝達特性を用いた単一チャネル音源位置推定

これまでに提案されてきた音源方向や位置の推定方法は，マイクロホンアレーにおける各観測信号の位相差などを用いた手法が多く，複数のマイクロホンが必要であった．単一マイクロホンで音源位置を推定することができれば，コスト削減やシステムの縮小化など様々な利点が期待できる．本研究では音響伝達特性を用いた音源位置推定法について検討する．

３．研究の方法

(1) 構音障がい者の音声認識における動的特徴量の考察

音声認識システムにおいて従来は，音声特徴量として対数スペクトルに対し離散コサイン変換を適用したメルケプストラムや，その線形回帰係数であるデルタケプストラムが広く用いられている．しかし，構音障がい者の発話スタイルは健常者と大きく異なり不安定であるため，メルケプストラムを用いた特定話者モデルでの音声認識には限界がある．特に動的特徴量であるデルタケプストラムを用いた音声認識において，構音障がい者の認識精度は健常者に比べて大きく低下

する．そこで，デルタケプストラムの代わりに，デルタケプストラム係数のセグメント特徴量を音響特徴量として用いる．当該フレームとその前後数フレームの計 n フレームを連結させ，主成分分析（Principal Component Analysis）により N 次元に圧縮を行ったものを音響特徴量とする．実際にはケプストラムと組み合わせたものを音響特徴量として用いる．

(2) スパース表現に基づく構音障がい者の発話スタイル変動にロバストな特徴量抽出

本研究では，最初の発話スタイルが不安定になるという課題に対して，非負値行列因子分解(Non-negative matrix factorization: NMF)をベースとした発話スタイル変動にロバストな特徴量抽出法を提案する．提案手法では，不安定な発話のスペクトルを，安定した発話のスペクトルパターンのスパースな線形結合(結合係数のほとんどがほぼ 0)で近似表現し，その結合係数を特徴量として音声認識を行う．

(3) 非負値行列因子分解による構音障がい者の声質変換

本研究では非負値行列因子分解(Non-negative Matrix Factorization: NMF)に基づく声質変換法を用いて構音障がい者の音声を健常者の音声に変換し，子音強調を行う．

学習に使う音声データからスペクトル包絡を抽出し，入力話者（構音障がい者）の辞書と出力話者（健常者）の辞書から構成されるパラレルな辞書行列を作成する．入力音声のスペクトルは，入力話者の辞書のスパース表現に変換できる．このとき，入力話者の辞書行列から選ばれた基底を，出力話者の辞書行列の同一アライメントの基底と交換することで，入力話者のスペクトルは出力話者のスペクトルと置き換えられる．NMF に基づく手法には，統計モデルが導入されていないため，過学習が起こりにくいと考えられる．

(4) 顔方位にロバストな唇領域特徴抽出と音声特徴による構音障がい者の音声認識

AAM は shape（特徴点の座標値）と texture（輝度値）をそれぞれ主成分分析によって次元削減することにより，少ないパラメータで顔の形状の変化とテクスチャの変化を表現できるようにしたモデルである．変形を伴う物体を高速かつ安定して追跡することが可能であり，顔特徴点抽出や発話認識において広く用いられている．また，AAM は顔画像の平均形状から学習により得られるパラメータにより，学習サンプルに十分近い画像を生成することもでき，入力された顔が横顔であっても正面の画像に補正することが可能である．

音声特徴量，画像特徴量をそれぞれ用いて，音声隠れマルコフモデル（Hidden Markov Model: HMM），画像隠れマルコフモデルを構築する．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，構音障がい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 の尤度に対して，重み付き線形和を用いて統合を行う．

(5) 音響伝達特性を用いた単一チャネル音源位置推定

本研究では音響伝達特性を用いて音源の位置を推定している．音響伝達特性は音源の位置によって異なる値を持つため，あらかじめこれを位置毎に学習しておけば，評価音声に対してもその音響伝達特性を識別することで音源位置を推定することができる．

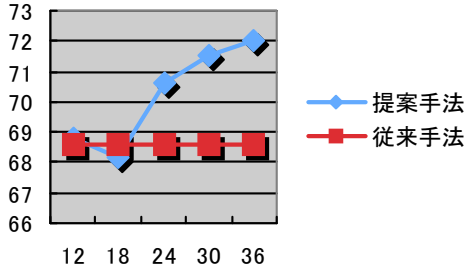
本手法は大きく二つのステップに分けられる．まず，ある位置から発話された音声から最尤推定法に基づき音響伝達特性を推定する．そして，推定された音響伝達特性を用いて音源位置を学習，識別を行う．

４．研究成果

(1) 構音障がい者の音声認識における動的特徴量の考察

動的特徴量であるデルタケプストラムを用いた音声認識において，構音障がい者の認識精度は健常者に比べて大きく低下することに着目し，デルタケプストラムの代わりにセグメント特徴量を用いることで音声認識識精度の改善を試みた．

ATR 音素バランス単語 216 単語と ATR 音声データベース 2,620 単語を用い，評価データは 1080 発話（216 単語×5 回発話），学習データは 5,240 発話（2,620 単語×2 回発話）を使用した．



特徴量次元数

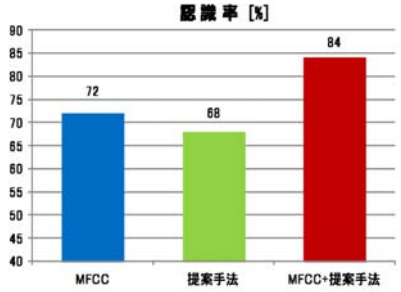
上図に音声認識率を示す（縦軸は認識率[%]を示す）．横軸は，主成分分析の次元数を

示す．従来手法のデルタケプストラムに比べ，セグメント特徴量が認識率の改善に貢献していることが分かった．

(2) スパース表現に基づく構音障がい者の発話スタイル変動にロバストな特徴量抽出

構音障がい者の最初の発話スタイルが不安定になるという課題に対して，スパース表現に基づく発話スタイル変動にロバストな特徴量抽出法を提案し，評価実験によりその有効性を確認した．

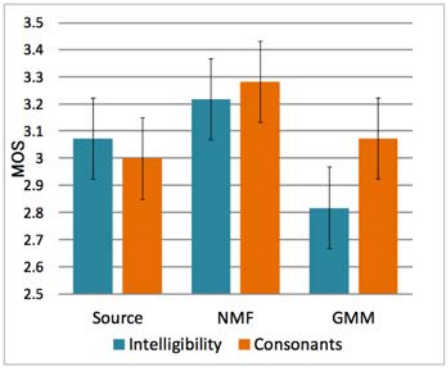
音声認識実験結果より，提案手法で得られた特徴量と従来のケプストラム音響特徴量を統合することで，不安定な第 1 発話において 12 ポイント(72%から 84%)の認識率の改善が得られた．



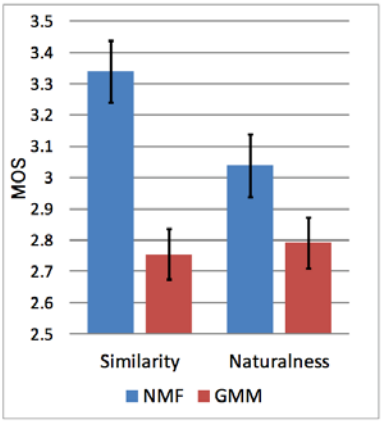
(3) 非負値行列因子分解による構音障がい者の声質変換

声質変換法を用いて構音障がい者の音声を健常者の音声に変換し，子音強調を行った．実験データとして，男性のアテトーゼ型構音障がい者 1 名のデータを収録した．発話内容は，ATR 音素バランス単語 216 語から選択した 50 語を用いた．対となる健常者の音声データは，ATR 音声データベースに収録されている男性話者のものを使用した．

成人男女 5 名による聴取実験を行った．音質と子音の明瞭性の 2 つの項目について，MOS 評価基準に基づく 5 段階評価(5:とてもよい，4:よい，3:ふつう，2:わるい，1:とてもわるい) の主観評価実験を行った．



上図に「聞き取りやすさ (intelligibility)」と「子音の明瞭性(consonants)」の評価結果を示す．提案手法に基づく声質変換は，無変換の障がい者音声と比較して聞き取りやすさと子音の明瞭性を向上させていることがわかる．一方，従来手法である GMM (Gaussian Mixture Model)に基づく声質変換では子音の明瞭性は向上しているものの，聞き取りやすさは無変換音声と比較して劣化している．これは，変換ノイズによるものと考えられる．提案手法の NMF に基づく変換手法も変換ノイズを発生させるものの，GMM (Gaussian Mixture Model) に基づくものよりは少ない変換ノイズとなっている．

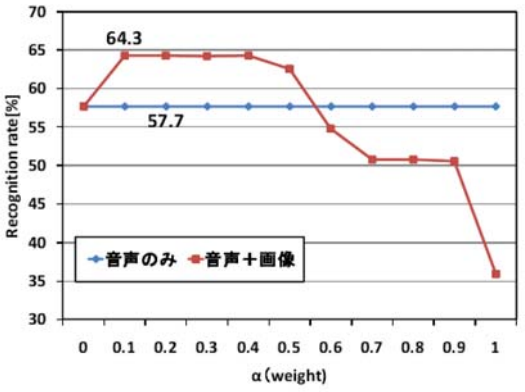


上図に「話者性(similarity)」と「自然性(naturalness)」の評価結果を示す．提案手法である NMF に基づく声質変換は，従来手法と比較して話者性を維持できていることがわかる．これは，従来手法は全ての入力スペクトル包絡を健常者のものに変換しているのに対し，提案手法では Combined dictionary を用いて子音のみの変換を行っているためである．自然性についても，提案手法は異なる話者の子音と母音を組み合わせているのにも関わらず，従来手法よりも高いスコアを得ている．これは，従来手法の変換ノイズの問題に加え，提案手法は母音部分をほぼ無変換で合成できるためと考えられる．

(4) 顔方位にロバストな唇領域特徴抽出と音声特徴による構音障がい者の音声認識

実験用データとして構音障がい者 1 名のデータを収録した．発話内容として ATR 音素バランス単語 1,065 発話 (216 単語×5 回) と ATR 音声データベース 5,240 発話 (2,620 単語×2 回) を使用した．画像データの解像度は 720×480，フレームレートは 30 fps である．

Signal to noise ratio が 10 dB の場合における重みを変化させたときの結果を下図に示す．音声情報に画像情報を統合することで発話認識率が改善されているのが分かる．しかし統合重みの値を変えることにより，音声情報のみの認識率を下回る場合もあり，最適な重みを選択して統合することが必要であることが分かる．



(5) 音響伝達特性を用いた単一チャネル音源位置推定

提案手法を評価するために特定話者によるシミュレーション実験を行った．音響伝達特性の学習データと評価データは，RWCP 実環境音声・音響データベースより音源とマイクロホンの距離が 2 m，残響時間が 300 msec のインパルス応答をクリーン音声に畳み込むことで作成した．音源位置は 30 度，90 度，130 度の 3 種類である．識別実験の結果，従来手法ではシングルマイクロホンによる音源方向推定は不可能とされていたが，本提案手法により平均 89.1%の識別率を達成することが出来た．

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 22 件）

① Toshiya Yoshioka, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Ariki, “Robust Feature Extraction to Utterance Fluctuations Due to Articulation Disorders Based on Sparse Expression,” APSIPA, 査読有, 2012, 4 pages

②Ryo AIHARA, Ryoichi TAKASHIMA, Tetsuya TAKIGUCHI, Yasuo ARIKI, “Consonant Enhancement for Articulation Disorders Based on Non-negative Matrix Factorization,”
APSIPA, 査読有, 2012, 4 pages

③Yuto Komai, Nan Yang, Tetsuya Takiguchi, Yasuo Ariki, “Robust AAM-Based Audio-Visual Speech Recognition against Face Direction Changes,” ACM Multimedia, 査読有, 2012, pp. 1161-1164

④T. Takiguchi, M. Yoshii, Y. Ariki, J. Bilmes, “Acoustic Model Transformations Based on Random Projections,” IEEE ICASSP, 査読有, 2012, pp. 1933-1936

⑤ Y. Komai, Y. Ariki, T. Takiguchi, “Audio-Visual Speech Recognition Based on AAM Parameter and Phoneme Analysis of Visual Feature,” The Fifth Pacific-Rim Symposium on Image and Video Technology, 査読有, 2011, pp. 97-108

⑥Ryoichi Takashima, Tetsuya Takiguchi, and Yasuo Ariki, “Feature Selection Based on Multiple Kernel Learning for Single-channel Sound Source Localization Using the Acoustic Transfer Function,” IEEE ICASSP, 査読有, 2011, pp. 2696-2699

⑦Chikoto Miyamoto, Yuto Komai, Tetsuya Takiguchi, Yasuo Ariki, Ichao Li, “Multimodal Speech Recognition of a Person with Articulation Disorders Using AAM and MAF,” IEEE International Workshop on Multimedia Signal Processing, 査読有, 2010, pp. 517-520

〔学会発表〕（計 34 件）

①相原龍，“Sparse Coding を用いた唇情報からの音声変換”，電子情報通信学会技術研究報告，2012 年 12 月 21 日，東京

②吉岡利也，“スパース表現に基づく構音障害者の発話スタイル変動にロバストな特徴量抽出”，日本音響学会 2012 年春季研究発表会，2012 年 3 月 15 日，神奈川

６．研究組織

(1)研究代表者

滝口　哲也（TAKIGUCHI TETSUYA）
神戸大学・都市安全研究センター・准教授
研究者番号：40397815