音声の声道・声帯波・韻律特性を分離し連続変化させる柔らかな声質変換方式の開発

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 31 日現在

機関番号：１２１０２
研究種目：基盤研究（C）
研究期間：2010～2012
課題番号：２２５００１４５
研究課題名（和文）　音声の声道・声帯波・韻律特性を分離し連続変化させる柔らかな声質変換方式の開発
研究課題名（英文）Development of Continuous Voice Morphing Using Separated Vocal Tract Area Functions, Glottal Source Waves, and Prosodic Features
研究代表者
田中　和世（TANAKA　KAZUYO）
筑波大学・図書館情報メディア系・教授
研究者番号：70344207

研究成果の概要（和文）：本研究の目的は、従来の声質変換、すなわちある話者（特徴空間上の特徴点）から目標話者（目標特徴点）への変換ではなく、音韻性を保持した状態で、特徴空間上の点ではなく面（領域）への声質変換を可能とする技法を開発することである。ここでは、音声発話時の声道・声帯波・韻律特性を分離し独立に変換すること、韻律特徴パターンの離散コサイン変換による圧縮表現などの導入により、これを可能にする手法を開発した。開発した手法は、客観評価においても、また聴取実験による主観評価においても、元話者と目標話者の中間点において音韻性の保持が従来手法に比べ高いことが確認された。

研究成果の概要（英文）：In this project, we have developed a flexible voice morphing method, which is based on a conversion using a linear combination of the vocal tract area functions estimated from speech signals and targeted on realization of the continuity of the phonological identity of the overall interpolated area. The main features of the method are 1) to separate characteristics of the vocal tract resonances from those of glottal source waves, 2) independent morphing of the vocal tract resonances and glottal source wave characteristics, and 3) conversion method of prosodic features based on DCT(digital cosine transform) domain. We have established that a morphing system constructed from the proposed method improves the continuity of the phonological identity and the speech quality in the intermediate morphing rate.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,200,000	360,000	1,560,000
2011 年度	1,100,000	330,000	1,430,000
2012 年度	900,000	270,000	1,170,000
総　計	3,200,000	960,000	4,160,000

研究分野：情報工学
科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス
キーワード：音声モーフィング、声質変換合成

１．研究開始当初の背景

近年、テキスト音声合成技術は一定の進歩は達成したものの、高品質音声が要求される現実のアプリケーションでは限界があり、むしろ、大容量ディジタル技術の進展により録音再生方式にとって替わられる傾向にあった。これは、音声の了解性を超えた個性的で肉声感のある音声の規則合成が依然として極めて困難であること、また、実際のシステムはそれぞれに固有で多様であり、単一の調子で任意の大量のテキストを音声に変換するというニーズは少ないこと、などによるも

のである。より効果的な手法・方式の開発が望まれて来た。その一つの有力な方式が、高水準な音声モーフィング（音韻性を保持したまま声質を連続的に変換する手法）である。この手法を開発すれば、声質を好みに応じてパラメータにより制御し、個性的な声をデザインできるシステムへの道が拓かれる。

２．研究の目的

本研究では、原音声の肉声感を活かした音声の声質変換により個性的で肉声感のある音声変換合成を可能とする基本技術を開発する。具体的には、新しい音声分析手法を駆使し、音声の特徴変量を声道特性、声帯波特性、韻律特性に分離し、それぞれを数理的モデルにより記述し、これらの変量を連続的に変化させることによる声質変換「柔らかな声質変換」を実現する。また、このような特性や機能を必要とする音声応答・案内などの対話文に適用し、その有効性を検証する。

３．研究の方法

（１）概要：本研究では、音声の声道特性、声帯波特性、韻律を個別に制御することに基づく声質変換を、音韻性を保持したまま特徴空間において連続的・面的に変換できる音声変換合成技術を確立する。研究課題として以下の２項目に分け、3 年計画で研究開発を行うものとする。

(A) 声道特性と声帯波特性の分離により、自由度が高く、かつ連続的な声質の変換を可能とする音声変換手法の開発。

(B) ピッチとパワー、発話時間要素の数理モデルとパラメータ制御による連続的な変換を可能とする手法の開発。

平成 22、23 年度において上記の項目について手法開発を行い、最終年度においては、上記を統合したシステムを構築する。

（２）課題項目（A）に対する方法：(A) についての問題は、特徴空間上の点（原音声）から点（参照話者音声）への変換ではなく、中間点や外挿も含む連続的・面的変換（より正確には空間領域への変換）を可能にすることである。図１に示すように複数の参照話者に広がる領域全体について、音質と音韻性の劣化を抑えて変換できる必要がある。これは特徴変量に係る問題であり、現行の主要な変換手法は音声のケプストラムを特徴変量としているため、中間点での変換音声の音韻性劣化は避けられない。変量を対数声道断面積とすると中間点でも連続的な声質変換が期待でき、そのための効果的な手法を確立する。

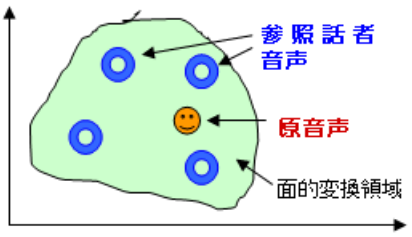

図１原音声からの面的領域への変換

基礎となる音声分析手法は AR-HMM（自己回帰 - 隠れマルコフモデル）分析であり、これにより推定された声道断面積関数空間において、声質変換手法を開発する。AR-HMM は声帯波を HMM、声道特性を AR モデルで表現して、音声信号を分析する手法であり、従来の音声の代表的分析手法である線形予測分析を発展させたものである。この結果、声帯波特性と声道特性を自然な形で分離推定できる。

推定された声道断面積関数に対して従来手法を単純に適用するだけでは（声道長の個人差などがあり）高品質な変換音声は得られない。このため、これを克服する手法の開発が必要である。

（３）課題項目（B）に対する方法：(B) についての従来の手法は、各音声サンプルの平均的レベルの平行移動が主であり、それ以上の詳細な変換手法は見当たらない。そこで、本研究では、まず、音声の韻律特徴パターンの分析とルール化について研究を行う。この結果に基づき、ルールベースの変換手法の開発を行う。

また、自動処理可能な手法開発のため、韻律パターンの特徴パラメータ化（ないし情報圧縮）について研究する。その上で、統計的手法に基づく韻律特徴パターンの変換手法を開発する。

（４）全体システムの構成：システム全体の処理の構成は図２に示すようになる。

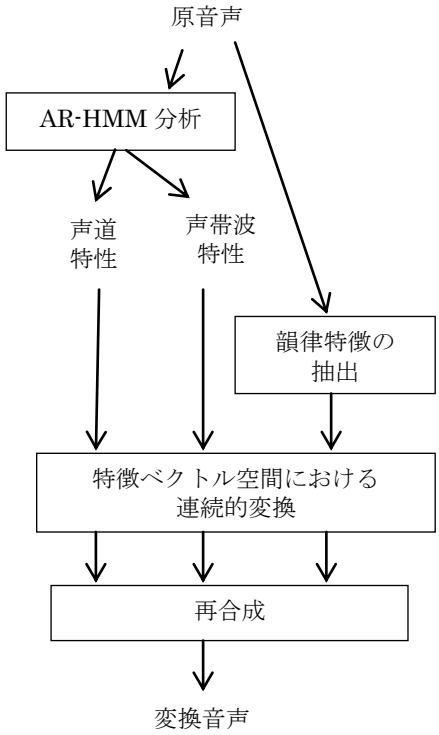

図２　システムの処理の流れ

ここで、主要素となる課題は特徴空間における音韻性を保持した状態での連続的変換手法の開発である。ただし、各ブロックの処理は互いに密接に関連しており、各ブロックのすべてについて手法開発と検証を進める必要がある。

（５）手法の評価： 課題項目（A)、(B）のいずれの場合においても、手法の評価は数理的指標による客観評価と共に、聴取実験に基づく評定者による主観評価を実施する。

４．研究成果

(1) 課題項目（A）に挙げた声道特性と声帯波特性を分離した連続的な音声モーフィングの手法開発：

この技術は、音韻性を保持したまま連続的な声質変換を可能とする技術である。開発した具体的手法として、まず AR-HMM 音声分析により音声から声帯波特性を分離し、声道断面積関数（反射係数）の精度の良い推定手法を開発した。次に、声帯波に相当する音源波形は、声道特性を求めて原音声からその特性を逆フィルタすることにより求めた。

予備実験として、声道断面積関数の対数値領域での統計的変換関数に基づくパワースペクトル変換が元話者と目標話者の中間点においても滑らかなホルマントになることを示した。とくにこの場合、話者により声道長が異なるため、これを補正する手法を開発し導入たことで、中間段階におけるモーフィング率に関してより滑らかな遷移特性を得ることが可能となった。(図３参照)。

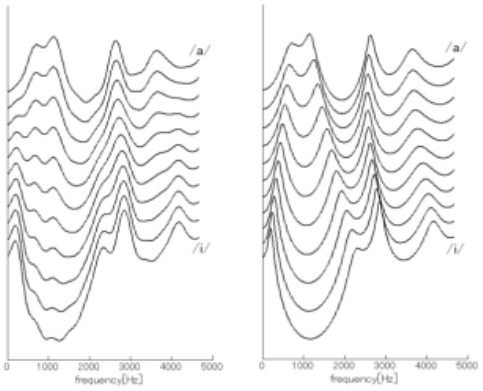

図３　左側図はケプストラム空間での補間によるパワースペクトルの遷移（従来法)、右側図は対数声道断面積空間での補間による場合（提案法）

以上の変換手法を導入して、最終的に声道特性と声帯波特性を分離した声質変換システムを以下の手順により構築した。

①学習フェーズ：

・AR-HMM 分析により声道断面積関数（反射係数）と声帯波を得る。声道断面積関数は対数値を取り、声帯波はケプストラムで表現。

・元話者と目標話者のサンプルデータの対応関係から変換関数を得る。

②変換フェーズ：

・上記の変換関数を利用して、元話者の音声から得られた特徴ベクトルを目標話者音声の対応する値に変換する。ただし、中間段階の変換音声を得るために、モーフィング率を設定する。(モーフィング率 100％が目標話者に相当する点である。)

・次に、目標音声の音源波形を高精度音声分析・合成ソフトウエア STRAIGHT によりケプストラムを用いて合成する。

・変換された声道断面積関数（反射係数）からなる線形予測フィルタを構成して、上記の音源波形を入力としたフィルタ出力を得て、目標音声波形を合成する。

（２）課題項目（A）に対する変換手法の客観評価と主観評価：

①変換手法の客観評価として、パワースペクトル歪（単位＝dB）を尺度として、目標話者音声への変換精度を測定した。元話者と目標話者の組合せとしては、男声→男声、男声→女声、女声→女声の 3 通りについて評価を行った。その結果、100％モーフィング率において、いずれの場合も従来手法であるケプストラムを特徴ベクトルとした場合に比べ変換精度がよく、３通りの平均で 4.40[dB]から 4.29[dB]と改善された。また、モーフィング率が 100％に至る中間段階におけるパワースペクトルの観察においても、先の図３に示したように、従来手法ではホルマントが消失したりするのに対し、提案手法ではホルマントが滑らかに遷移するパターンが明確に観測された。

②次に、変換した音声を評価者に提示して音声品質を評価する主観評価実験を実施した。実験は①と同じ 3 通りの組合わせについて、MOS（Mean Opinion Score）を用いて、従来手法と提案手法の評価を行った。MOS 値は 5 段階（1：悪い、2：やや悪い、3：普通、4：やや良い、5：良い）である。

この結果、モーフィング率 100％の場合は、3 通りの MOS 値の平均で、従来手法による変換音声 MOS 値＝2.7、提案手法による変換音声 MOS 値＝2.6 であり、(有意水準５％の統計的検定では）有意差は認められなかった。しかし、モーフィング率 70％の変換音声に関しては、従来手法による変換音声 MOS 平均値＝2.9 に対し、提案手法による変換音声 MOS 平均値＝3.4 となり、有意水準５％で有意差が認められた。この結果は、モーフィング率の中間段階においても音韻性が保持される変換という提案手法の有効性を実証するものである。

（３）韻律パターンの変換による中立音声から感情音声への変換：
前述したように、従来手法は基本周波数の平均レベルを平行移動する方式が主であり、韻律要素（意図や感情などを表す要素）を変換するものではない。これに対して、提案手法は基本周波数やパワーの時間変化パターン自体の変換を行うもので、韻律要素の変換を行う。このための手法開発の前段として、韻律の音響的特徴と意図や感情との関係分析を実施し、知見を整理した。これをルール化したモデルのパラメータ制御により、ある程度の制御が可能であることを確認した。
変換の過程を自動処理化する手法として以下の手順を開発した。
①基本周波数およびパワーの時間パターンを離散コサイン変換（DCT）によりベクトル表現する。次元数は 50 次元とする。
②各次元の相関を無視（対角要素のみ処理）して、元パターンと目標パターンの間の統計的写像を学習し、変換関数を作成する。
③この変換関数に基づいて、基本周波数、パワーパターンの変換を行う。
以上の変換手法の有効性を確認するため、評価実験を実施した。実験では、元音声として中立的発話音声を用い、目標音声として同じ文を感情を入れて発話した音声とした。実験の結果、従来手法では基本周波数などの韻律特徴の時間パターンについては平均レベルの移動のみであったのに対し、提案手法により、韻律特徴の局所的時間変化パターンの変換が可能となったことが観測により確認できた。また、変換音声を評定者に提示した主観評価実験によりその効果を確認できた。これらの実験により、特徴量としてパワーパターンの変化が感情の表現には主要因であることを示した。

（４）以上まとめると、本研究プロジェクトでは、当初の開発目標である「音韻性を保持した連続的な声質変換」の手法について、対数声道断面積関数を特徴ベクトル空間とする変換手法を新規に開発し、音韻性を保持した連続的声質変換を実現した。また、感情表現などの韻律要素を変換する韻律特徴パターンの有効な変換手法も提案した。これらの研究成果は学会論文誌、国際会議等において発表した。全体を統合したシステムの開発評価までには至らなかったが、基本要素となる新規手法を開発し提示したので、概ね当初の目標を達成した。

５．主な発表論文等
〔雑誌論文〕（計 4 件）

① Kazuyo Tanaka, Yoshiki Nambu, “Continuous Voice Morphing Using Separated Vocal Tract Area Functions and Glottal Source Waves,” International Journal of Multimedia Technology, ISSN:2226-7875(online), 査読有, 7 pages (accepted to be published in June, 2013).
② Tomoko Nariai, Kazuyo Tanaka, ″A Study on Pitch Patterns of Japanese Speakers of English in Comparison with Native Speakers of English,″ Acoustical Science and Technology, Vol. 33, No. 4, pp. 247-254, 査読有, Aug., 2012.
③ Tomoko Nariai, Kazuyo Tanaka, ″A Study on Pitch Patterns in Japanese Speakers of English with Verification by Speech Re-synthesis,″ IEICE Transactions on Information and Systems, Vol.E94-D, No.12, pp. 2495-2502, 査読有, Dec., 2011.
④ Tomoko Nariai, Kazuyo Tanaka, Yoshiaki Itoh, “A Comparative Study of Focal Lengthening in the Speech of Native Speakers and Japanese Speakers of English,” Acoustical Science and Technology, edited by Acoustical Society of Japan, Vol.32, No.2, pp.54-61, 査読有, March, 2011.

〔学会発表〕（計 12 件）

① 石 睿, 田中 和世, 三河 正彦, 羅 志偉, “韻律特徴パターンの DCT 次元圧縮による韻律の異なりを考慮した声質変換手法の検討”, 日本音響学会 2013 年春季研究発表会論文集 3-P-44d, 2 pages, 査読無, 東京工科大, 2013-3-15.
② Shi-wook Lee, Hiroaki Kojima, Kazuyo Tanaka, Yoshiaki Itoh, “Experimental Evaluation of Probabilistic Similarity for Spoken Term Detection,” Proc. of International Conference on Pattern Recognition Applications and Methods (ICPRAM) 2013, 6 pages, 査読有, Sants Hotel, Barcelona,(Spain), Feb. 17, 2013.
③ Tomoko Nariai, Kazuyo Tanaka, Tatsuya Kawahara, “Comparative Analysis of Intensity between English Speakers and Japanese Speakers of English,” Proc. of Interspeech 2012, Paper Tue.P4a.04, 4 pages, 査読有, Hilton Hotel, Portland, (USA), Sep. 11, 2012.
④ 李 時旭, 児島 宏明, 田中 和世, 伊藤 慶明、“混合正規分布間の誤差推定値近似に関する実験的考察”、日本音響学会 2012 年春季研究発表会論文集 3-P-19, 2pages, 査読無, 神奈川大学, 2012-3-15.
⑤ Yasuharu Hashimoto, Masahiko Mikawa, Kazuyo Tanaka, “Development of Prototype sound Direction Control System Using a Two-dimensional Loudspeaker Array,” Proc.

of 19th European Signal Processing Conference (EUSIPCO) 2011, pp.264-268, 査読有, Palau de Congressos, Barcelona, (Spain), Aug. 31, 2011.
⑥ Itoh, Yoshiaki; Iwata, Kohei; Ishigame, Masaaki; Tanaka, Kazuyo; Lee, Shi‐wook, "Spoken Term Detection Results Using Plural Subword Models by Estimating Detection Performance for Each Query," Proc. of Interspeech 2011, pp.2117-2120, 査読有, Palazzo dei Congressi, Florence, (Italy), Aug. 29, 2011.
⑦ Tomoko Nariai, Kazuyo Tanaka, "An Experimental Analysis of Pitch Patterns in Japanese Speakers of English with Verification by Speech Re-synthesis," Proc. of Interspeech 2011, pp.1169-1172, 査読有, Palazzo dei Congressi, Florence, (Italy), Aug. 29, 2011.
⑧ 南部良季, 三河正彦, 田中和世, "声道特性から分離した音源特性の異なりを考慮した声質変換手法の検討", 日本音響学会 2011 年春季研究発表会論文集 1-Q-56, pp.417-418, 査読無, 早稲田大学, 2011-3-9.
⑨ Tomoko Nariai, Kazuyo Tanaka, "A study of pitch patterns of sentence utterances by Japanese speakers of English in comparison with native speakers of English," Proc. of Interspeech Satellite Workshop: Second Language Studies, Paper No.P2-1(4 pages), 査読有, Waseda Univ., Tokyo, Sep. 23, 2010.
⑩ 成合智子, 田中和世, "日本人の英語文発声におけるパワーパターンの解析", 日本音響学会 2010 年秋季研究発表会論文集 3-P-20, pp.375-376, 査読無, 同志社大学, 2010-9-16.
⑪ 南部良季, 三河正彦, 田中和世, "複数話者間における声道長の差異に着目した音声モーフィング手法の検討", 日本音響学会 2010 年秋季研究発表会論文集 3-Q-14, pp.327-328, 査読無, 同志社大学, 2010-9-16.
⑫ Yoshiki Nambu, Masahiko Mikawa, Kazuyo Tanaka, "Flexible Voice Morphing based on Linear Combination of Multi-speakers' Vocal Tract Area Functions," Proc. of 18th European Signal Processing Conference (EUSIPCO) 2010, pp.790-794, 査読有, Congress Center, Aalborg, (Denmark), Aug. 25, 2010.

６．研究組織

(1)研究代表者

田中　和世（TANAKA　KAZUYO）
筑波大学・図書館情報メディア系・教授
研究者番号：70344207

(2)研究分担者

三河　正彦（MIKAWA MASAHIKO）
筑波大学・図書館情報メディア系・准教授
研究者番号：40361357

伊藤　慶明（ITOH　YOSHIAKI）
岩手県立大学・ソフトウエア情報学部・准教授
研究者番号：90325928