

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24年 6月18日現在

機関番号：14701

研究種目：挑戦的萌芽研究

研究期間：2010～2011

課題番号：22650042

研究課題名（和文） 能の謡いの芸術表現の可視化に基づく聴覚潜在脳研究手段の開発

研究課題名（英文） Development of research tools based on visualization of artistic expression in vocal performance of Noh

研究代表者

河原 英紀（KAWAHARA HIDEKI）

和歌山大学・システム工学部・教授

研究者番号：40294300

研究成果の概要（和文）：我国の伝統的芸術である能の謡の表現に含まれる音声の分析合成技術の開発を通じて、これまでの分析合成技術では不可能であった情報の抽出と可視化ならびに制御と再合成を可能にする技術を開発した。特に、これまで不可能であった高い時間分解能での基本周波数およびスペクトルの分析を可能にした。これらに基づくモーフィングは、意識化が困難な脳における潜在的処理の計測の手段を提供することで、新しい研究の可能性を拓いた。

研究成果の概要（英文）：A new set of technologies for information extraction, visualization, control and resynthesis, which provide functions conventional methods failed to implement, are developed in this project aiming at analysis and synthesis of vocal artistic expressions found in Japanese traditional theatrical art, Noh. Especially, temporally finer F0 and spectral estimation procedures were invented to enable detailed analyses, which were impossible using conventional methods. Voice morphing based on these technologies opens new possibilities for investigating implicit brain functions which are not accessible from consciousness.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,600,000	0	1,600,000
2011年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,900,000	390,000	3,290,000

研究分野：総合領域

科研費の分科・細目：情報学、感性情報学・ソフトコンピューティング

キーワード：感性脳科学、芸術表現、音声情報の可視化、能の謡、伝統歌唱、非周期性

1. 研究開始当初の背景

当時、音声知覚、音声分析合成などの研究手段として、本課題の代表者の開発した STRAIGHT は、事実上の世界標準と言える位置を築いていた。この STRAIGHT は、本課題の提案の2年前に、新しいアイデアに基づき TANDEM-STRAIGHT という本質的に新しいシステムに進化していた。しかし、歌唱音声等の

芸術的表現の与える非言語的、下意識的感動を研究するための手段として用いるには、分析の深さも加工され再合成された音声の品質も全く不十分なものであった。幾つかの可能性は見えかけていたものの、明確な見通しは無く、挑戦的・探索的にそれらの検討を進めるといふリスクの高いステージに入る必要があった。

2. 研究の目的

本研究の目的は、能の謡の音声に含まれる芸術表現の可視化と変換音声のフィードバック技術などの導入により、聴覚に関する脳機能研究にブレークスルーをもたらす手段を開発することにある。人間の脳で行われている意識に上らない情報処理に関しては、脳活動の大きな部分を占めるにも関わらず研究が遅れている。特にその遅れは聴覚において顕著である。本研究では、この遅れを取り戻すための強力な研究手段と方法論を、唐突と取られかねない能の謡の音声の分析／可視化／加工／知覚の研究を通じて構築することを狙う。

3. 研究の方法

研究手段として用いる能の謡の音声の芸術的印象を再現することを可能とするために、前述の TANDEM-STRAIGHT をベースとして、高品質・高精度の音声の分析ならびに合成技術の開発を進め、その方法による分析結果の可視化技術のプロトタイプを開発することとした。実験に用いるための能の謡の音声については、研究協力者である藤村靖オハイオ州立大学名誉教授と金春流のシテ方である金春康之氏らの協力によって、以前に奈良新公会堂の能舞台において収録したものと、ATR 人間情報通信研究所の無響室において収録したものを中心に用い、補足資料として中山一郎大阪芸術大学教授が配布しているものを併用することとした。後者は、人間国宝 19 名を含む計 78 名の演者の歌唱や謡を、物理計測に耐える測定系で収録した貴重な資料である。

4. 研究成果

本課題により、以下に挙げる重要な多数の成果を得ることができた。これらの新しく発明／開発された方法は、これまで不可能であった深さの分析と可視化を可能とし、さらに、芸術表現の研究に耐える品質の合成をも可能にした。これらを音声モーフィングに応用することにより、これまで不可能であった前意識的な脳活動を計測し定量化するための手段とすることができる。

(1) XSX (excitation Structure extractor) の最適化。学会発表②

強い歌唱表現においては、通常の音声のように基本周波数という単一の指標を用いて記述することのできるものとは異なり、同時に複数の周期を有する駆動が生ずる場合がある。このような複雑な駆動情報を表現する方法として、著者らにより XSX が開発されてきた。この XSX は、現在の TANDEM-STRAIGHT において音源情報抽出法として用いられているが、細部の更新の繰返しにより、見通し

の悪いものとなっていた。

本課題では、まず、XSX を根本から再設計し、ダウンサンプリングと組み合わせることで、構造と内部の調整可能なパラメタの最適化を行なった。その結果、能の謡や狂言に含まれる二重音声等の分析と可視化において、従来法では見えなかった構造を明らかにすることができた。

図 1 に狂言に含まれる二重音声の分析結果の例を示す。図の上部には、参照のために音声波形を併せて示している。下は、XSX により求められた周期成分の複数の候補を 1ms 毎に示したものである。それぞれの候補の信頼性を、マークのサイズを用いて可視化することにより、特徴を明瞭に把握することができる。このような表示により、0.2 秒までの部分は、通常の音声と同様に単一の基本周波数で表すことのできる音源により駆動されているのに対し、0.3 秒から 0.7 秒までは、二重音声となっている。XSX は、この二重音声の区間において、駆動が短い周期と長い周期のペアから構成されていることを容易に把握することができる。このような表示と分析は、XSX により初めて可能になったものである。

(2) 周期信号における安定した瞬時周波数の計算法の発明。学会発表②⑨

音声の基本周波数は、時間とともに変化する。このような信号を適切に表現するには、数学的な周期信号に対して定義されている周波数ではなく、位相の時間微分として定義される瞬時周波数という、拡張された概念に基づく必要がある。しかし、音声のように複数の成分を含む信号の瞬時周波数には、周期的に不連続が生ずるという問題があり、応用する上での障害となっていた。

本課題では、瞬時周波数の有するこの問題を、分析対象が周期信号の場合に、実用上完全に解決する方法を明らかにした。具体的に

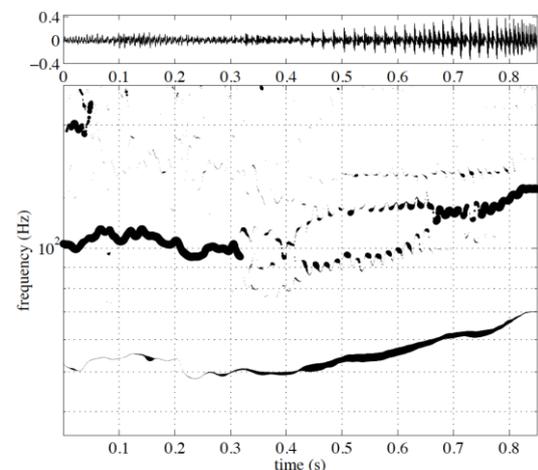


図 1 XSX による分析結果

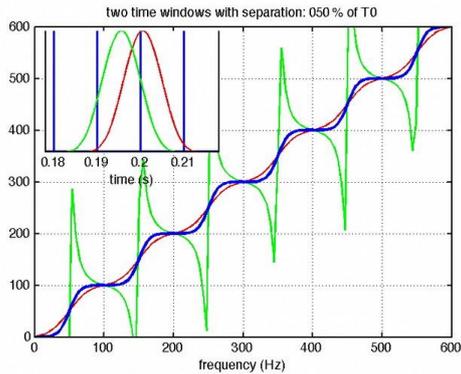


図 2 安定化した瞬時周波数の計算法

は、基本周期の 1/2 の時間間隔を隔てて求めた瞬時周波数からパワースペクトルを重みとして用いた加重平均を計算するという簡単な処理で、この問題を解決することができる。これは、歌唱音声の処理だけではなく、周期信号の分析一般に応用することのできる非常に基本的な発明である。

図 2 に例を示す。図の左上に周期信号（この場合はパルス列）と分析用の 2 つの窓を示す。これらの窓により求められた瞬時周波数を窓に対応した色で示す。発明した方法により加重平均として求められる瞬時周波数を青線で見せる。緑と赤の線で示されている従来の瞬時周波数は、窓と信号の相対位置に応じて、周期的に図に示すような鋭い不連続を生ずる。それに対し、青線で表される提案した方法は、パルスと窓の相対位置に関わらず、形状が固定している。

(3) 波形の対称性に基づく時間応答性が高く高速な基本周波数抽出法。学会発表③⑤

XSX の分析により、強い表現の歌唱には、非常に高速な基本周波数の変化が含まれていることが見いだされた。しかし、このような分析を大量の資料に対して行なう上で、XSX の処理の重さが大きな問題となっていた。

この問題を、波形の対称性からの逸脱に基づく指標を導入し、(2) で挙げた安定した瞬時周波数計算法と組み合わせることにより解決する方法を明らかにした。通常この種の分析では、成分の分離に帯域通過フィルタが用いられる。しかし、提案する方法では、帯域通過フィルタではなく、低域通過フィルタを用いることで、まず時間分解能を 2 倍としている。これを、間隔と対称性の計測という軽い処理を組み合わせることにより、高速性と処理の頑健性を両立させている。さらに、こうして高速に求められた基本周波数を初期推定値として用い、安定化された瞬時周波数を用いた二段階の更新処理を行うことにより、推定精度を大きく改善する方法を明らかにした。

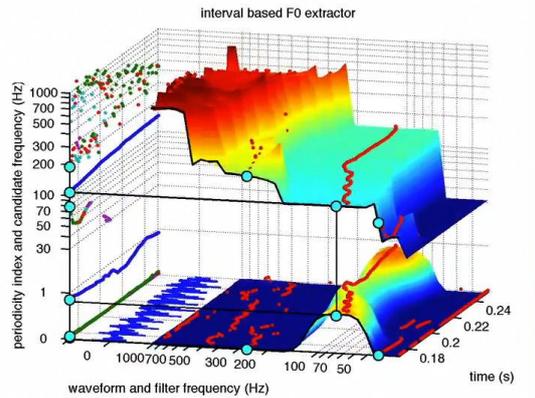


図 3 対称性に基づく基本周波数抽出法

図 3 に、この方法による初期推定値の抽出過程の可視化例を示す。下側の擬似カラーの曲面は、対称性からの逸脱に基づいて求められた周期性の確からしさを示す指標を表している。奥行き方向が時刻、左右方向が用いる低域通過フィルタの遮断周波数（対数周波数）、を表す。

推定では、まず、各時刻において、水色の円で示されている極値の中から、最大のものを候補として選択し、上の擬似カラーの曲面として表されている波形の平均間隔の中から、対応するフィルタの出力を読み出す。このようにして、左側の壁面に描かれている基本周波数の候補が求められる。図示していないが、この後で、最良の候補の値を初期値として、安定化した瞬時周波数に基づく推定値の改良が行われる。

この方法による処理時間は、通常のラップトップコンピュータを用いた場合、CD 品質の音声の基本周波数の初期推定に実時間の 1/5、安定化した瞬時周波数に基づく推定値の改良に実時間の 1/2 である。なお、これらの値は、分析結果を 1/1000 秒毎に求める場合について求められたものである。また、処理に必要な記憶領域も少なく、数分におよぶ全曲の処理を一括して行なうことが可能となっている。

(4) 高品質な音声処理を可能とするスペクトル推定法。

分析合成および加工音声の品質を向上させるためのスペクトル包絡の推定法の改良法を明らかにした(論文②)。この方法では、人間の聴覚の特性に基づいて、知覚される品質に大きく影響するスペクトルピーク周辺での推定精度を重視するとともに、聴覚末梢系における周波数分析を反映させた評価尺度を用いている。この評価尺度を用い、発声時の声道形状に関する MRI（磁気共鳴イメージング）データと、声帯振動モデルを用いたシミュレーションを遂行し、最適なパラメタ設定を明らかにした。実際の音声を用いた分析合成音声および変換音声を作成し、最終的

な評価となる被験者を用いた主観評価実験を遂行することにより、提案した方法が従来の方法を凌ぐ品質を実現していることを確認した。

(5)その他の成果

これらの要素技術の発明に加え、TANDEM-STRAIGHT が初めて可能にした高品質な音声モーフィングを、容易に利用できる技術とするためのシステム開発を進めた。その過程で導入された、声道長正規化(論文④)は効果の大きい方法であり、また、応用範囲も広い。本課題では、これらの成果を社会に還元させるための活動として、解説、招待講演、総合的展望の発表を精力的に行った。

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4件)

①河原英紀、音声分析合成技術の動向、日本音響学会誌、67巻、2011、40-45、査読無、招待

②赤桐隼人、河原英紀、他、スペクトルピークを強調した F0 適応型スペクトル包絡抽出法の最適化と評価、電子情報通信学会論文誌 A、J93-A、2011、557-567、査読有

③Hideki Kawahara and Masanori Morise, Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, SADHANA - Academy Proceedings in Engineering Sciences, vol. 36, 2011, 712-722、査読有

④Erika Okamoto and Hideki Kawahara et. al., Evaluation of voice morphing using vocal tract length normalization based on auditory filterbank, Journal of Signal Processing, vol. 15, 2011, 283-286、査読有

[学会発表] (計16件)

①Hideki Kawahara et. al., Analysis and synthesis of strong vocal expressions: Extension and application of audio texture features to singing voice, ICASSP2012, 2012. 3. 29, 京都

②和田芳佳、河原英紀、他、障害音声および歌唱音声における音声の周期構造分析について、日本音響学会春季研究発表会、3-11-1、2012. 3. 15、横浜

③河原英紀、他、異なった原理に基づく周期性検出器のアンサンブルによる音源情報の分析について、日本音響学会春季研究発表会、3-11-4、2012. 3. 15、横浜

④河原英紀、歌声のテクスチャに信号処理はどう迫るか、音楽情報科学研究会、2012. 2. 3、浜松

⑤河原英紀 他、基本波の FM と AM 成分に基づく高速な基本周波数推定法について、日本音響学会聴覚研究会、2011. 12. 10、熊本

⑥Erika Okamoto and Hideki Kawahara et. al., Auditory filterbank improves voice morphing, Interspeech2011, 2011. 8. 30, フイレンツェ・イタリア

⑦ Hideki Kawahara et. al., An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction, ICASSP2011, 2011. 5. 27, プラハ・チェコ

⑧ Hideki Kawahara, In search of perceptually relevant speech representations, STRAIGHT, TANDEM-STRAIGHT and beyond, NCSP' 11, 2011. 3. 2, 中国・天津

⑨Yoshika Wada and Hideki Kawahara et. al., Optimization of a Multiple Local Periodicity Detector for Vocal Excitation Structure Analysis, APSIPA2010, 2010. 12. 16, Singapore

⑩Hideki Kawahara, Exploration of the other aspect of Vocoder revisited, A-Z STRAIGHT, TANDEM-STRAIGHT and morphing, SSW7, Speech Synthesis Workshop, 2010. 9. 22, 京都

6. 研究組織

(1)研究代表者

河原 英紀 (KAWAHARA HIDEKI)
和歌山大学・システム工学部・教授
研究者番号：40294300

(2)研究分担者

()

研究者番号：

(3)連携研究者

()

研究者番号：