様式 C-19

科学研究費補助金研究成果報告書

平成 22 年 6 月 28 日現在

研究種目：若手研究（スタートアップ）
研究期間：2008～2009
課題番号：20800062
研究課題名（和文）音素変化の影響を受けにくい音声モーフィング技術の研究

研究課題名（英文）A study of voice morphing technology robust against phonemic change

研究代表者
森勢　将雅（MORISE MASANORI）
立命館大学・情報理工学部・助教
研究者番号：60510013

研究成果の概要（和文）：
本研究では，2 つの音声を入力とし，2 つの音声の中間的な印象の音声を出力できる音声モーフィングに基づく声質変換の高品質化を目的としている．一般的なモーフィング技術では，モーフィング対象とする 2 つの事象を対応付けることが要求される．従来の音声モーフィングにおける対応付けは，手動にて行われていた．本研究成果により，対応付けの自動化を実現すると共に，対応付けそのものが不要な声質変換技術が構築された．

研究成果の概要（英文）：
A new voice morphing technology was proposed to easily synthesize intermediate voices between two voices. Conventional voice morphing method requires aligning two input voices manually. However, the proposed method can align two input voices automatically. Furthermore, a new voice conversion method without aligning was proposed to overcome the problem. In this research, we confirmed that the proposed method can synthesize natural morphing voices without aligning.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	1,320,000	396,000	1,716,000
2009 年度	1,000,000	300,000	1,300,000
年度			
年度			
年度			
総　計	2,320,000	696,000	3,016,000

研究分野：音声合成
科研費の分科・細目：感性情報学・ソフトコンピューティング
キーワード：感性情報学，音声情報処理，音声モーフィング

１．研究開始当初の背景

音声分析合成に関する技術は，Vocoder が提案された 1930 年代以上より研究されてきた．従来研究されてきた音声分析合成に関する技術は，音声をより少ない情報量で記述し，効率よく転送することを目的として発展してきた．そのため，従来の合成音声の品質は，必ずしも高いものとはならなかった．近年計算機の処理能力向上に伴い，高品質な音声合成技術への需要が高まっている．特に，コン

テンツ業界において，音声合成システムを用いた作品が急速に増えていることからも，音声合成に対する需要の伸びは今後ますます加速すると予想される．

現在の音声合成技術は，(1) 大量のデータを予め蓄積し，素片を接続する方式と，(2) 音声を音響パラメタに分解し，音響パラメタを操作することにより，少量のデータから大量のデータを合成する Vocoder 方式とに大別できる．前者では，Vocaloid のような歌唱合成ソフトが市販されているなど，すでに実用化されつつある．人間の音声に基づいた素片を接続するため，比較的容易に高品質な音声を合成できるメリットがある．一方，膨大なデータを蓄積する必要があり，データベースに存在しない音声は合成できない．Vocoder 方式は，音響パラメタの操作により様々な表情を生み出せる可能性を秘めている一方，品質が大きく劣化する問題があった．

1997 年に河原らによって発明された音声分析変換合成方式 STRAIGHT は，Vocoder でありながら元音声に匹敵する品質の音声が合成できる技術として注目されている．STRAIGHT は，複数の音声から，その中間的な印象の音声を合成できる音声モーフィング技術を可能にした．音声モーフィングに関しては，これまでも提案されていたが，STRAIGHT は品質劣化を最小限に抑えつつ変換が可能な技術として，現在も幅広く利用されている．

STRAIGHT における音声モーフィングでは，時間周波数方向において，両音声を音素レベルで対応付ける必要がある．この対応付けの作業は，熟練者が目視によって行う必要があり，対応付けが不十分な音声間でモーフィングを行った場合，品質が著しく低下することが知られている．また，STRAIGHT を用いた音声の分析合成は多大な計算コストが要求される．そのため，実時間で加工を行うようなコンテンツ制作を支援するソフトウェアに音声モーフィングを用いることは現状では不可能といえる．

２．研究の目的

本研究は，従来要求されていた対応付けの制約を緩和させた音声モーフィング技術の提案を目的とする．また，歌唱を含むコンテンツ制作現場においても利用可能なシステムの確立を目指し，STRAIGHT の計算コスト削減や，Ｃ言語での実装と配布も行う．

３．研究の方法

本研究は，音声モーフィングにおける対応付けの自動化に関する研究と，音声モーフィングの基盤となる STRAIGHT の計算コスト削減に関する研究から構成される．以下に，それぞれの研究方法について示す．

(1) STRAIGHT の計算コスト削減

STRAIGHT は音声から 3 つの音響パラメタ「基本周波数」「スペクトル包絡」「非周期性指標」を取り出す分析法と，3 つの音響パラメタから音声を合成する方法から構成される．申請者らが 2008 年に提案した TANDEM-STRAIGHT は，スペクトル包絡推定の計算コストを大幅に削減した．本研究では，基本周波数推定における推定精度の改善と計算コストの削減を中心に検討する．

(2) 実時間分析合成を実現する TANDEM-STRAIGHT の API 策定，およびライブラリの実装と配布

従来の TANDEM-STRAIGHT は，分析開始から終了までを一括して行うように実装されていたため，例えば入力音声を実時間で変換して合成を行うようなアプリケーションに利用することができない．本研究では，分析・合成用 API を実時間アプリケーションに利用可能となるよう変更を加え，C 言語のライブラリとして実装する．実装されたライブラリは，Web を通じて配布を行い，成果を社会へ還元する．

(3) 歌唱モーフィングにおける自動対応付け

現在のデジタルコンテンツ制作において歌唱は重要な役割を担う．そこで，モーフィング対象を歌唱に限定し，対応付けを自動化可能なモーフィング技術を提案する．歌唱をモーフィングする場合，モーフィング対象となる 2 つの歌声は，同一のキーで歌うこと，歌詞のタイミングが概ね既知であるという制約を活用し，対応付けの自動化を試みる．

４．研究成果

本研究の主要な成果として，(1) STRAIGHT に必要な計算コストの大幅な削減，(2) 実時間アプリケーションの実装を可能にする STRAIGHT Library の実装と配布，(3) 対応付けを必要としない歌唱モーフィング技術の構築が挙げられる．それぞれの詳細を以下で述べる．

(1) STRAIGHT の計算コスト削減

従来の基本周波数推定は，自己相関，相互相関を用いた時間軸の周期を検出する方法と，ケプストラムのようにパワースペクトルが有する調波構造に着目した方法とに大別される．STRAIGHT における基本周波数推定は，時間軸・周波数軸の特徴だけではなく，瞬時周波数など新たな特徴量も併用した，極めて計算コストの大きいものであった．提案法は，低域通過フィルタとゼロ交差検出を組み合わせた簡易な方法である．従来提案された高精度な方法の計算量が $O(n \log(n))$ であるのに対し，提案法は，従来法よりも高い精度を

達成しつつ計算量が O(n)となる特長を有する．さらに，従来法と比較してほぼ等価な性能を達成できる．

① 推定精度の評価結果

提案法を従来の STRAIGHT，および近年提案された高精度な基本周波数抽出法 YIN，SWIPE＇と比較した結果，提案法は，YIN より高い推定精度を達成し，STRAIGHT や SWIPE＇とほぼ等価な精度で基本周波数を抽出可能であることが示された．

② 計算時間の評価結果

提案法の計算時間は，STRAIGHT よりも約 80 倍高速であり，YIN と比較しても 29 倍，SWIPE＇と比較しても 42 倍高速に基本周波数を抽出可能であった．以上の成果より，STRAIGHT における基本周波数推定の大幅の計算コストの削減に成功した．

(2) STRAIGHT Library の実装と配布

STRAIGHT は，各音響パラメタを推定する分析と，音響パラメタから音声を合成するための API から構成される．従来の STRAIGHT は，波形全体に対する処理を行う API のみ実装されていたが，本研究では，カラオケやヴォイスチェンジャーのような用途にも利用できる実時間処理 API を策定し，実装した．実装されたライブラリは，Web を通じて，研究用とであればフリーで利用できるように配布した．

(3) 歌唱モーフィングにおける自動対応付け

音声のモーフィングには，対象となる 2 つの音声を時間周波数平面上で対応付ける必要がある．モーフィング対象を歌唱に限定した場合，同一楽曲を歌うことから，音素の発声タイミングが比較的近く，声質の類似した歌手同士にてモーフィングを行う場合が多い．

本研究では，歌唱のモーフィングにおける対応付けの問題に対し，音声認識ソフトウェアにおける自動ラベリングによる自動化と，周波数方向では簡易モーフィング（同性かつ声質が類似した場合，周波数軸上の対応付けを行わなくとも品質が損なわれないという特徴に着目したモーフィング）を組み合わせた自動化法を提案した．

提案法の有効性を確認するために，主観評価を実施した．図１，図２は，女性ボーカル２名が同一楽曲を発声した歌唱をモーフィングし，モーフィング率により品質がどのように変化するか示している．横軸はモーフィング率を示し，縦軸は被験者 10 名が判定した品質（1 が最低で 5 が最高）の平均値を示している．

図１より，提案法（Automatic alignment）

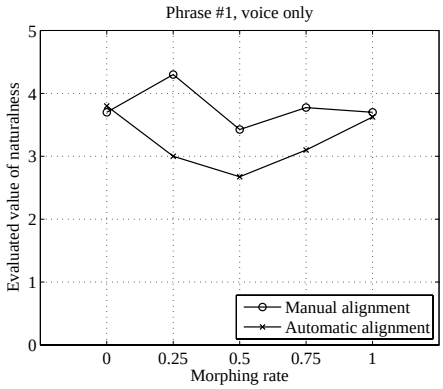

図１．手動対応付けと自動対応付けによる品質の違い（サンプル１）

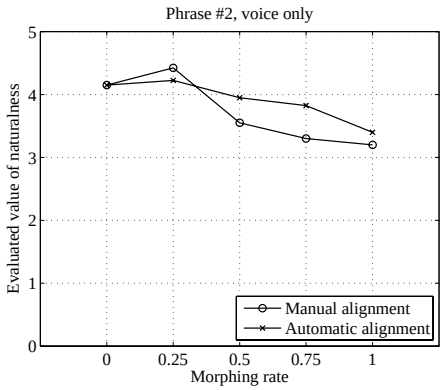

図２．手動対応付けと自動対応付けによる品質の違い（サンプル２）

は，従来法と比較するとモーフィング率 0.5 において特に品質が劣化する．一方で，歌詞が異なるサンプル２の実験結果では，提案法のほうが従来法よりも高い品質であることが確認できる．提案法は既存の音声認識ソフトウェアにより音素境界を算出しているが，算出精度が歌詞に依存して大きく変化するため，モーフィング歌唱の品質に差が生じたと考えられる．しかし，歌詞によっては，手動で対応付けを実施した場合よりも高い品質が達成されることも分かった．

本実験結果より，提案法は，対応付けの作業を自動化することに成功したといえるだろう．

(4) その他

ある話者が発した音声の母音を他者の母音へと置き換えることによる声質変換技術を提案した．提案法は，モーフィングとは異なり，時間周波数平面上での対応付けを必要とせず，他者が発した日本語五母音のみを用いて変換が可能である．

本研究では，提案と予備的な検討を実施したにとどまったが，今後は正式な主観評価実

験を行い，有効性を示す予定である．

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２件）

① 著者名：森勢将雅, 西浦敬信, 河原英紀, 論文標題：基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法，雑誌名：電子情報通信学会 論文誌 D，査読：有，巻：J93-D，発行年：2010，ページ：109-117

② 著者名：森勢将雅，高橋徹，河原英紀，入野俊夫，論文標題：分析時刻に依存しない周期信号のパワースペクトル推定法を用いた音声分析，雑誌名：電子情報通信学会 論文誌 A, 査読：有, 巻：J92-A, 発行年：2009，ページ：163-171

〔学会発表〕（計１１件）

① 発表者名：中野皓太, 発表標題：STRAIGHT スペクトルと格子型フィルタに基づく音声の極推定と評価，学会名等：日本音響学会 2010 年春季研究発表会，発表年月日：2010 年 3 月 9 日, 発表場所：東京（電気通信大学）

② 発表者名：松原貴司，発表標題：振幅相関雑音を用いた高品質合成音声の主観評価，学会名等：日本音響学会 2010 年春季研究発表会，発表年月日：2010 年 3 月 8 日，発表場所：東京（電気通信大学）

③ 発表者名：森勢将雅，発表標題：能の発生における非周期的な声帯振動について，学会名等：日本音響学会 2010 年春季研究発表会，発表年月日：2010 年 3 月 8 日，発表場所：東京（電気通信大学）

④ 発表者名：松原貴司, 発表標題：Proposal of an advanced modulated noise reference unit to evaluate high-quality synthesized voices，学会名等：NCSP ’10，発表年月日：2010 年 3 月 4 日，発表場所：ホノルル（アメリカ）

⑤ 発表者名：松原貴司，発表標題：室内残響時間が TANDEM-STRAIGHT 分析合成音声の品質に与える影響，学会名：平成 21 年電気関係学会関西支部連合大会，発表年月日：2009 年 11 月 8 日，発表場所：大阪（大阪大学）

⑥ 発表者名：森勢将雅，発表標題：Rapid F0 estimation for high-SNR speech，学会名等：WESPAC X 2009，発表年月日：2009 年 9 月 22 日，発表場所：北京（中国）

⑦ 発表者名：森勢将雅，発表標題：歌唱モーフィングにおける対応付けの自動化に関する検討，学会名：日本音響学会 2009 年秋季研究発表会，発表年月日：2009 年 9 月 15 日，発表場所：福島（日本大学）

⑧ 発表者名：中野皓太，発表標題：音声合成を目的とした励起信号抽出に関する初期的検討，学会名：日本音響学会 2009 年秋季研究発表会，発表年月日：2009 年 9 月 17 日，発表場所：福島（日本大学）

⑨ 発表者名：森勢将雅，発表標題：v.morish ’09: A morphing-based singing design interface for vocal melodies，学会名：ICEC2009，発表年月日：2009 年 9 月 3 日, 発表場所：パリ（フランス）

⑩ 発表者名：森勢将雅，発表標題：高 SNR の音声を対象とした高速な F0 推定法の最適化および性能評価，学会名：日本音響学会 2009 年春季研究発表会, 発表年月日：2009 年 3 月 17 日，発表場所：東京（東京工業大学）

⑪ 発表者名：森勢将雅，発表標題：Fast and reliable F0 estimation method based on the period extraction of vocal fold vibration of singing voice and speech，学会名：AES 35th International conference，発表年月日：2009 年 2 月 12 日，発表場所：ロンドン（イギリス）

〔図書〕（計０件）

〔産業財産権〕
○出願状況（計０件）

○取得状況（計０件）

〔その他〕
ホームページ等
http://www.crestmuse.jp/cmstraight/

６．研究組織
(1)研究代表者
森勢　将雅（MORISE MASANORI）
立命館大学・情報理工学部・助教
研究者番号：60510013