歌唱音声モーフィングに基づく声質と歌い回し転写の知覚的検討

河原英紀[†1] 生駒太一[†2] 森勢将雅[†2]
高橋徹[†1] 豊田健一[†4] 片寄晴弘[†3]

歌唱における演奏デザインの転写技術の確立を目指して，高品質音声分析変換合成システムSTRAIGHT に基づくモーフィングの研究を進めている．STRAIGHT を用いることにより，基本周波数，時間周波数表現，非周期性指標の 3 種類の物理パラメタのみから高度に自然な音声を合成することができる．ここでは，それらのパラメタと時間軸および周波数軸を併せた 5 種類のパラメタを独立にモーフィングできるように拡張することで，歌手の声質と歌い回しを転写することを試みた．男子学生と男性のプロ歌手，女性のプロ歌手 2 名の歌唱を用いた主観評価実験は，目的とした変換が高い品質で実現されていることを示した．

Perceptual study on design reuse of voice identity and singing style based on singing voice morphing

HIDEKI KAWAHARA,[†1] TAICHI IKOMA,[†1]
MASANORI MORISE,[†1] TORU TAKAHASHI,[†1]
KEN'ICHI TOYODA[†4] and HARUHIRO KATAYOSE[†3]

Investigations on singing voice morphing has been conducted to establish a design reuse framework based on a high-quality speech analysis, modification and resynthesis system STRAIGHT. STRAIGHT enables high-quality speech reconstruction using only three parameters; fundamental frequency, time-frequency representation, and time-frequency aperiodicity map. In this paper, an extension of STRAIGHT-based morphing, which enables individual control of morphing rate using five parameters (in addition to three parameters mentioned above, time and frequency axes were introduced), was implemented to test design reuse of singers' voice identity and their singing style. Subjective evaluations of two sets of manipulated samples were conducted. First set was generated from a male student and a professional male singer and the second set was generated from two professional female singers. Test results illustrated that intended reuse was perceptually verified and generated samples were generally in high-quality.

1. はじめに

モーフィング技術には，物理パラメタと心理的属性との対応関係についての明示的な知識に依存せずに，２つの試料の中間状態を容易に制作できることや，現実に存在しない効果を生み出すことができるという利点がある．これらの特長により，モーフィングはコンテンツデザインの有用なツールとして映像メディアの領域において広く実用に供されている．しかし，聴覚メディアの領域におけるモーフィングは，適切なツールが不在であったなどの要因により，広く利用されるには至っていない．

この状況は，河原らによって発明されたSTRAIGHT[7]と，それを用いたモーフィング技術の導入により変わりつつある[8),16),20)]．STRAIGHT は，聴覚の情景分析[3)] に代表される聴覚機能の生態学的理解に立脚して開発されたシステムである．STRAIGHT は，電気的音声処理の原点である channel VOCODER[4)] に遡り，開発された当時は二義的とみなされていた『音声を聴覚における情報表現と機能的に等価なパラメタ群に分解する』という側面を，現代の技術で追求することか

†1 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University
†2 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University
†3 関西学院大学理工学部
School of Science and Technology, Kwansei Gakuin Univ.
†4 （元）関西学院大学大学院理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin Univ.

ら生み出された[5]．

STRAIGHT は，音声を基本周波数，スペクトル包絡，非周期性指標に分解する．注意深く相互の干渉を取り除かれて抽出された非負の実数で表されたそれらのパラメタは，操作による品質劣化の少ない音声変換を可能にする．STRAIGHT を，パラメタの値を変換しない単なる分析合成システムとして用いることもできる．その場合には，処理により波形が保存されないにも関わらず，再合成された音声は，元の音声に匹敵する自然な聴覚的印象を与える[10]．STRAIGHT に基づくモーフィングは，この特徴を利用して，パラメタおよびそれらの時間周波数座標を区分的一次関数を用いて補間することにより実現されている[8]．

3 種類のパラメタだけから高い自然性を有する刺激連続体を作成できるという STRAIGHT の特長[10),16),20)]は，例えば，宇多田ヒカルの曲を美空ひばりの歌声で聴きたいという願い[17)]や，自分の声がプロのように歌うのを聴きたいという願いを実現するための基盤を提供する．本論文では，そのような新しい音楽の愉しみ方[17),19)]を可能にするための第一歩として，異なった歌手による歌唱音声の声質と歌い回しのそれぞれを独立に操作する方法[15)]を提案し，知覚実験により有効性を検討した結果について報告する，

2. モーフィングによるデザインの転写

モーフィングは，2つの事例となる試料が提供されたときにそれらの中間となる性質を有する試料を合成する操作として定義される．しかし，前節で紹介したような願望を実現するためには，特定の知覚的属性のみの操作を実現する必要がある．ここでは，まず STRAIGHT に用いられているパラメタの性質と，モーフィングへの応用について簡単に紹介した後，目的とする操作を実現するために必要となる拡張について説明する．

2.1 STRAIGHT の情報表現[5),7)]

STRAIGHT は，音声を基本周波数，スペクトル包絡，非周期性指標に分解する．再合成音声は，必要に応じて変形されたこれらのパラメタから高い周波数における群遅延にランダムな変動を加えられ非周期成分を加えられた音源信号と，最小位相応答システムとして実現されたフィルタを用いて生成される．

2.1.1 スペクトル包絡

このパラメタの分析では，まず，音声の基本周波数に適応して変化する相補的時間窓を用いて，分析位置による変動のないパワースペクトルを求める．次いで，spline 関数の性質を利用した周波数方向の平滑化により，調波位置でのスペクトルのレベルを保存したスペクトル包絡を抽出する．これらの処理によって基本周波数の情報がほぼ完全に除去されたスペクトル包絡は，周波数軸の伸縮や基本周波数の変換に起因する品質劣化が少ない音声の合成を可能とする．

2.1.2 基本周波数

STRAIGHT のスペクトル分析では，基本周波数に適応して窓関数が設計される．そのため，基本周波数の抽出誤りは品質劣化を引き起こす．音声の収録時に用いられる高域通過フィルタや低域に主要な成分を有する空調騒音および商用電源からの誘導雑音の影響は，周期性が弱くまた乱れることのある母音の開始および終了時の抽出誤りにつながる．これらの問題を解決するために，基本波および低次調波の瞬時周波数の情報と帯域毎の正規化された自己相関とを併用し，さらに後処理を行う高精度の方法を開発した[6]．既定値では 1 ms 毎に求められる．

2.1.3 非周期性指標

有声音であっても声門閉止が完全に行われない場合には，声門での乱流等に起因する非周期成分が含まれる．この現象は，女性の音声に典型的に認められる．男性の場合でも，有声摩擦音や弱い気息性の発声の場合には，顕著な非周期成分が含まれる[14)]．これらを表現するために，STRAIGHT では非周期性を表す指標を用いている．指標は，帯域フィルタを通過するエネルギー全体と，その中に含まれている非周期成分のエネルギーの比 (dB) として定義されている．帯域幅は，聴覚末梢系における周波数分解能を近似する ERB_N(Effective Rectangular Bandwidth)[12)] に比例するように設定されている．実装では，基本周波数の瞬時位相を用いた時間軸の変換で仮想的に基本周波数を一定としたパワースペクトル[1),2)] の上側包絡と下側包絡の差から，非周期性指標を求めている[9)]．

2.1.4 実　　装

STRAIGHT の実装には，科学技術計算用の環境である Matlab を用いた．STRAIGHT を構成するサブシステムは，基本周波数抽出用関数，非周期性指標抽出用関数，スペクトル包絡抽出用関数，合成音声作成用関数として実装されている．以下のモーフィングの実装も，Matlab 上でこれらの関数を用いて行った．

2.2 STRAIGHT に基づくモーフィングの拡張

STRAIGHT を用いることにより音声を独立性の高い非負の実数の組により表現できるため，モーフィングを事例間の線形補間により容易に実装することができる．また，スペクトル包絡と音源情報という情報表現は，聴覚における属性との類似性が高く直感的に

理解し易い．ここでは，まず，同一の楽譜に基づいて演奏した歌唱音声のモーフィングの手順を簡単に説明する．

2.2.1 スカラー値によるモーフィング

モーフィングは，2つの演奏の時間軸と周波数軸の対応する点が重なるように時間-周波数座標を変形させることから始まる．次いで，変形された時間-周波数座標の各点において，指定されたスカラー値（モーフィング率）に応じて2つの演奏のパラメタの値を補間／補外し，モーフィング率に応じて変形された時間-周波数座標にそれらの値を設定する．こうして作成されたパラメタを STRAIGHT の音声合成部に与えることで，目的とするモーフィング音声が合成される．

2つの演奏の時間軸と周波数軸の対応づけは，次のように行われる．まず，文献 8) と同様に，フォルマント軌跡の変曲点や音素境界などを手がかりとして，特徴となる点が利用者の手作業によりそれぞれの演奏の時間-周波数座標の上に設定される．次に，それぞれの演奏に付与された特徴となる点が重なり合うように，双一次変換により実装された関数を用いて時間-周波数座標が変形される．

2.2.2 モーフィング率の拡張とオブジェクト化

前節で説明したモーフィングでは，スペクトル包絡，基本周波数，非周期性指標と時間軸および周波数軸の計 5 個のパラメタが，同一のモーフィング率で変換されていた．しかし，パラメタ空間の 2 点間として表現される事例間を結ぶ経路は，この方法により実現される直線のみに限られない，モーフィング率 r を，媒介変数 λ を用いてパラメタ空間の各次元に対応する 5 個の係数 $(r_1(\lambda), r_2(\lambda), r_3(\lambda), r_4(\lambda), r_5(\lambda))$ から構成されるベクトルに拡張することにより，2 点を結ぶ任意の経路を表現することができる．なお，実用上は，拘束条件（$r_k(0) = 0, r_k(1) = 1$，ただし $k \in \{1, 2, 3, 4, 5\}$）と単調性（$\lambda_1 \leq \lambda_2$ ならば $r_k(\lambda_1) \leq r_k(\lambda_2)$）を要請しておくと都合が良い．

これらの要請に柔軟に対応するために，モーフィングオブジェクトという概念を導入し，モーフィングを，2つのモーフィングオブジェクトから一つのモーフィングオブジェクトを生成する演算として定義した．演算をこのように定義することで再帰的なモーフィングの適用が可能となり，2つ以上の事例間のモーフィングへの拡張が容易となる．

実装では，モーフィングオブジェクトを Matlab の構造体として定義した．前述の 5 個のパラメタに加え，付与された特徴点の座標と分析条件やオブジェクトの履歴情報等を構造体のそれぞれのフィールドとした．

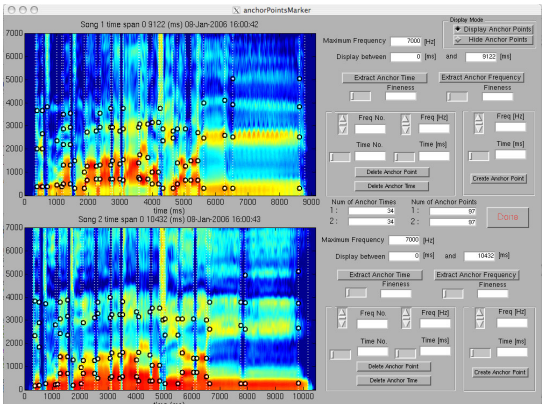

図 1 対応点付与作業支援ツール
Fig. 1 GUI tool for editing time-frequency markers

2.2.3 支援環境

モーフィングのための特徴となる点は，当初，研究者が実験目的に応じて注意深く設定することを想定していた[8]．しかし，音声生成および知覚についての基本的知識を欠く一般の利用者では適切な点の設定ができず，また，研究者にとっても，注意深い手作業には多くの時間を必要とするという問題がある．十分な信頼性を有する自動設定ツールがまだ存在していないため，ここでは，設定作業の支援に文献 15) の方法を利用した．

文献 15) の方法は，ガウス関数による多重解像度分析であるスケールスペースフィルタ[11] に基づいている．この方法では，まず特徴的な点の時刻の候補を，音量が大きく変化する点に相当するスケールスペースフィルタ出力の変曲点から求める．次いで，選択された時刻における周波数方向の特徴点の候補を，スペクトル包絡にスケールスペースフィルタを適用した出力の極大点から求める．求められる候補点は，用いるスケールスペースフィルタに含まれるガウス関数の標準偏差 σ に依存する．今回は，時間方向では 50 ms から 120 ms，周波数方向では 65 Hz から 215 Hz の範囲で幾つかの σ を用いた候補点の抽出を行い，**図 1** に示すような対話的ツールを用いて修正を行った．

図 1 の左側には，モーフィング対象の二人の歌手による歌声のスペクトログラム（STRAIGHT によるスペクトル包絡の系列の時間-周波数表現）が表示されている．白抜きの点は，設定された特徴点である．この例では，「結末はまだ誰も知らない」という 9〜10 秒の一節に，34 個の特徴となる時刻が設定され，合計 97 個の特徴点が設定されている．右側のパネルには，選択した特徴点や時刻を削除したり，周波数や時刻を編集するための GUI 要素が配置されている．それぞれ

の歌唱音声に同じ個数の特徴となる時刻が設定され，それぞれの時刻において同じ個数の特徴点が設定されると，モーフィングの準備は完了する．

3. 声質と歌い回しのモーフィング

STRAIGHT を用いた実験により，聴覚は，音を生成している物体の形状やサイズの情報を無意識のうちに自動的に抽出しているらしいことが分かって来た[13]．歌唱音声の場合には，声道の形状は，それぞれの歌手の解剖学的な構造による拘束を受けているため，スペクトル包絡の張る空間での特定の歌手の歌唱音声の軌跡は，埋め込まれた小さな部分空間内にとどまると考えられる．STRAIGHT で求められるスペクトル包絡には，単一の声帯体積流波形のスペクトル形状も含まれるため，この部分空間は，声帯音源の個人性までを反映した，声質を総合的に表す特徴を表してもいる．したがって，ある歌手から別の歌手への声質の転写は，この部分空間同士の写像を指定することに他ならない．これは，周波数軸とスペクトル包絡を選択的にモーフィングすることにより歌手の声質の転写が実現できることを意味する．

一方，基本周波数の軌跡や音声を発するタイミングおよび音量は，調音器官の動特性による拘束を受けるものの，訓練により意図的に制御することが可能である．また，拘束の表れる調音器官の動特性の時定数の変動[14] と比較すると，基本周波数に深く関わっているピッチ知覚の時定数[12] は遥かに大きい．そのため，楽譜により基本周波数の平均値が拘束されている歌唱音声☆ の聴取時に，基本周波数およびその軌跡から推定可能な動特性の時定数が歌手の個人性を表すものとして利用されているとは考えにくい．これらの予備的な考察から，歌い回しのスタイルを転写するには，音素境界等に設定されている特徴点の相対時刻と基本周波数の軌跡とを歌手間で選択的にモーフィングすれば良いことが示唆される．

なお，声質と歌い回しに関連するパラメタは，それぞれのグループ内で同一のものを用いることとするため，それぞれモーフィング率 r_{timbre}，r_{way} と表記することとする．

4. 声質と歌い回しの転写実験

前節で提案した仮説を検証するため，実際の歌唱音声を用いて主観評価実験を行った．以下に，用いた素材および手続きについて紹介する．

☆ 正確には，拘束されるのは知覚される属性であるピッチの系列である．

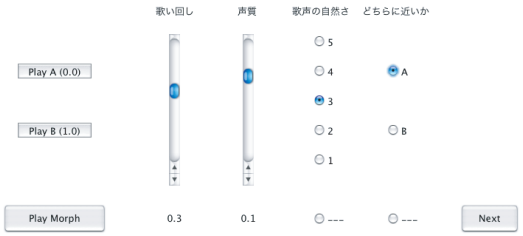

図 2 実験用アプリケーション
Fig. 2 GUI for subjective evaluation of morphed singing

4.1 予 備 実 験

実験に用いた楽曲は，RWC 研究用音楽データベース[18] のポピュラー音楽に収録された楽曲 “So Long”（RWC-MDB-P-2001 No.64）のうち，サビ部分にあたる約 10 秒のフレーズである．音声試料は，上記の楽曲に収録されたスタジオミュージシャンによる歌声（歌声 A とする）と，同一曲を歌ったプロではない 20 代男性の歌声（歌声 B とする）である．歌声 B の歌声の収録において，歌唱者は楽曲の伴奏をヘッドホン（audio-technica ATH-A 1000）で聴きながら歌った．この歌声をマイク（SHURE SM57-LCE）により収録し，44.1kHz，16bit で記録した．なお，歌声 A はくせの強い歌い回しであったため，モーフィングの相手となる歌声 B の収録においては，なるべくピッチのゆらぎを少なくしてフラットに歌うよう指示した．

実験における刺激は，2 つの歌声を元にモーフィング率 r_{timbre}，r_{way} をそれぞれ-0.2 から 1.2 まで 0.2 ステップで変化させて生成したモーフィング音声 64 個と，歌声 A と歌声 B の計 66 個の歌声である．なお，歌声 A と歌声 B の特等点は 7000 Hz 以下に付与することとし，各時刻における特徴点の個数の上限を 5 とした．

4.1.1 評価用アプリケーション

被験者にはランダマイズされた刺激をヘッドホン（audio-technica ATH-A 1000）により提示し，各音声について以下のような 4 つの項目に対する評価を求めた．実験用に作成した評価用アプリケーションの外観を図 2 に示す．被験者は左側の 3 個のボタンを用いることで，歌声 A，歌声 B，および刺激であるモーフィングされた歌声を何度でも聴くことができる．右下のボダンをクリックすることで評価が記録され，次の条件の刺激が提示される．

- 歌い回しのモーフィング率の評価
  刺激を聴取して感じた歌い回しのモーフィング率を左側のスライダーを用いて評価する．スライ

ダーの下にフィードバック情報として表示される数値は-0.4 から 1.4 の間で 0.1 刻みとした．

- 声質のモーフィング率の評価
  刺激を聴取して感じた声質のモーフィング率を右側のスライダーを用いて評価する．スライダーの下にフィードバック情報として表示される数値は-0.4 から 1.4 の間で 0.1 刻みとした．
- 歌声の自然性の評価
  刺激の自然性を「非常に自然」から「非常に不自然」までの 5 段階で評価する．
- 歌唱者の判別
  刺激が歌声 A と歌声 B のいずれにより近いと感じられるかを 2 択で選択する．

歌唱音声のモーフィングは，例えば「もう少し○○さん風の歌い回しにしたい」「もう少し○○さんの声に似せたい」「○○さんの感じを誇張して加えたい」という要求を実現するための手段である．このような要求を，任意の値を連続的に設定できるスライダーの操作にマッピングする際の知見を得ることを狙い，評価用アプリケーションの GUI を決定した．

4.1.2 モーフィング率の評価結果

3 名の被験者によるモーフィング音声の声質のモーフィング率の評価結果を箱ひげ図を用いて**図 3** に示す．横軸は，操作に用いた声質のモーフィング率 r_{timbre} を表し，縦軸は，被験者によって評価されたモーフィング率を示す．箱の外に飛び出した横棒（ひげ）は，評価の最大値最小値を示し，箱の下辺と上辺とは，それぞれ累積分布の 25%点と 75%点を示す．太線で示した箱の中の横棒は，評価値の平均値を表す．モーフィングが補間である領域では，操作したモーフィング率と主観的に評価されたモーフィング率は，ほぼ単調に対応している．しかし，モーフィングが外挿となっている部分では，この単調性は崩れている．また，操作した範囲が-0.2 から 1.2 であったにもかかわらず，全ての評価値が 0 から 1 の範囲に入っている．

図 4 に，歌い回しの評価結果を示す．横軸は，操作に用いた歌い回しのモーフィング率 r_{way} を表し，縦軸は，被験者によって評価されたモーフィング率を示す．箱ひげ図の表現方法は図 3 と同じである．歌い回しの評価は，声質と比較すると評価値が広い範囲に分布している．評価値の個人差も大きく，声質と比較すると歌い回しは評価が困難であるものと考えられる．

これらの結果は，モーフィング操作についての十分な説明が与えられていない場合には，被験者が 0 から 1 の範囲を超える値の意味のイメージを持てないことを示唆する．同様の傾向は文献 16) にも認められる．

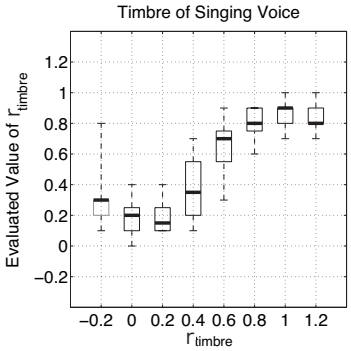

図 3 声質のモーフィング率の評価
Fig. 3 Timber rating results for r_{timbre} manipulation

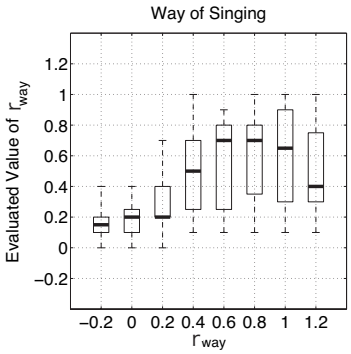

図 4 歌い回しのモーフィング率の評価
Fig. 4 Style rating results for r_{way} manipulation

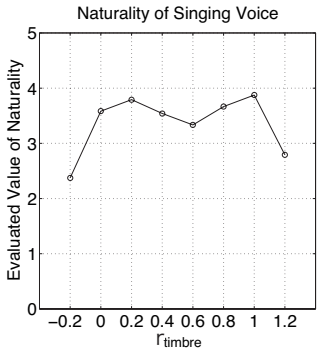

図 5 声質のモーフィング率と歌声の自然性
Fig. 5 Naturalness rating for r_{timbre} manipulation

4.1.3 自然性の評価結果

図 5 に声質のモーフィング率に対する自然性の評価値の平均値，**図 6** に歌い回しのモーフィング率に対する自然性の評価値の平均値を示す，

モーフィングの元試料となる 2 つの歌声の自然性の評価値の平均はいずれも 4.5 であった．単なる分析合成音であるモーフィング率 0 の歌声の自然性の評価値

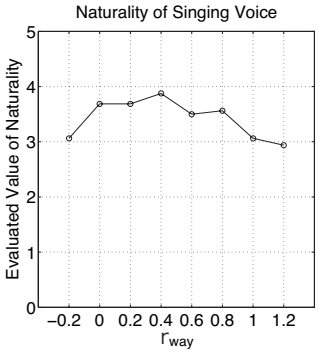

図 6 歌い回しのモーフィング率と歌声の自然性
Fig. 6 Naturalness rating for r_{way} manipulation

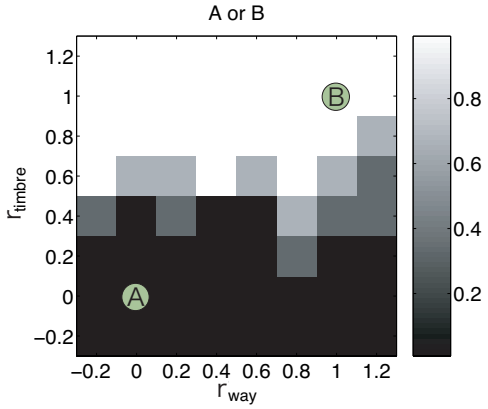

図 7 歌声の判別（A と判断された場合を 0，B と判断された場合を 1 とし，被験者の評価値の平均値を表している）
Fig. 7 Singer identification map for r_{timber} and r_{way} manipulations. Dark color represents singer A and white color represents singer B

は 3.8 であり若干の低下が認められるが，内挿を行った場合のモーフィング音声の自然性は，単なる分析合成と同程度である．しかし，外挿を行った場合のモーフィング音声では自然性が大きく下がる．この傾向は，声質のモーフィングの場合に顕著である．この評価は，外挿を行ったモーフィング歌声について被験者が「声質が荒れているように感じた」とコメントしたことと一致している．

なお，被験者の中には 5 段階の評価値において 4 あるいは 5 の評価値のみを与えている者もいた．この点について実験後にコメントを求めたところ「どのモーフィング歌声も，本物の人間の歌声のように聴こえた」との回答を得た．

4.1.4 歌声の判定

被験者によるモーフィング音声の歌唱者の判別結果を図 7 に示す．横軸は歌い回しのモーフィング率 r_{way}，縦軸は声質のモーフィング率 r_{timbre} に対応している．評価値は，A ならば 0，B ならば 1 とし，グラフの色は被験者 3 人による評価値の平均値を表している．グラフの黒い部分ほど歌声 A，白い部分ほど歌声 B と判断されたことを表している．

この結果から，歌唱者の判別においては，歌い回しよりも声質のほうが支配的であることがわかる．ただし，実験のサンプルに用いた 2 つの歌声の歌い回しが類似していたという可能性も否めない．そこで，歌唱者と被験者を変え次の実験を企画した．

4.2 実　　験

ここでは，2 名のプロの女性歌手による 8 秒間の歌唱音声を用いて学生を中心とする 14 名の被験者による実験を行った．曲は，著作権処理の問題がクリアされた新曲 (Love affair) を用いた．被験者による実験の手続きは予備実験と同様である．ただし，被験者は他の研究において音声モーフィングを用いており，モーフィング率の意味と外挿のイメージを既に有している．

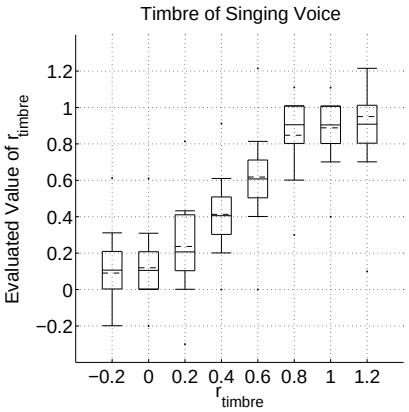

図 8 声質のモーフィング率の評価
Fig. 8 Timber rating results for r_{timbre} manipulation

4.2.1 モーフィング率の評価結果

図 8 と**図 9** にそれぞれ声質と歌い回しのモーフィング率の評価結果を示す．横軸と縦軸は，予備実験と同じである．今回の実験では被験者数が多く，分布を詳細に見ることができるため，黒点で表した評価の最大値と最小値に加え，ヒゲを用いて 10%および 90%の点を表し，箱内の実線で分布の中央値を，破線で平均値を表すこととした．

図 8 に示した声質の評価では，外挿に相当する評価を示す被験者もおり，評価値は，より広い範囲に分布している．しかし，予備実験と同様に，物理的な操作量が外挿にあたる領域では，評価値が飽和する傾向が同様に認められた．

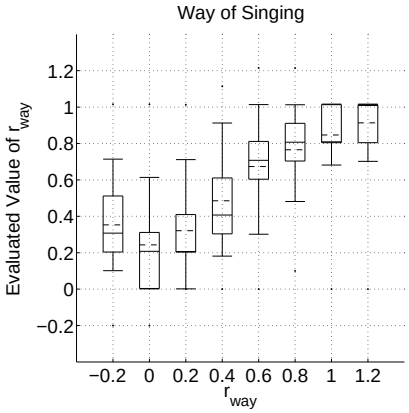

図 9 歌い回しのモーフィング率の評価
Fig. 9 Style rating results for r_{way} manipulation

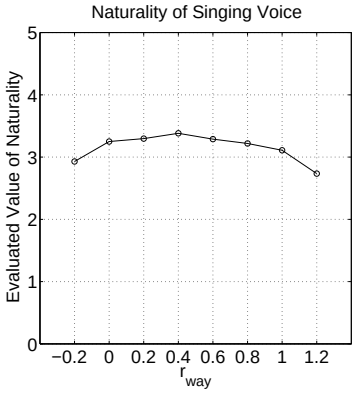

図 11 歌い回しのモーフィング率と歌声の自然性
Fig. 11 Naturalness rating for r_{way} manipulation

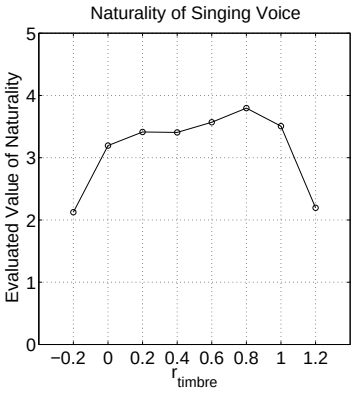

図 10 声質のモーフィング率と歌声の自然性
Fig. 10 Naturalness rating for r_{timbre} manipulation

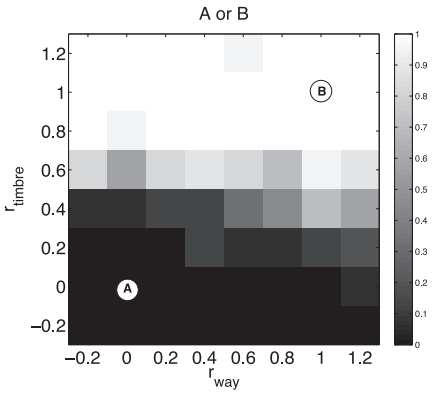

図 12 歌声の判別（A と判断された場合を 0，B と判断された場合を 1 とし，被験者の評価値の平均値を表している）
Fig. 12 Singer identification map for r_{timber} and r_{way} manipulations. Dark color represents singer A and white color represents singer B

図 9 に示す歌い回しの評価では，予備実験と比較すると，より分布が狭いことが分かる．今回の 2 名のプロ歌手は，声質も歌い回しも，予備実験の男性と比較すると違いがはっきりとしているため，判定がより容易であった結果であると思われる．ただし，実験終了後のインタビューでは，被験者の多くは，歌い回しは分かりにくく回答に迷ったとコメントしていた．得られた評価値も，声質の判定と比較すると外挿領域での飽和傾向が強く，また，相対的に分布が広がっていることが分かる．

4.2.2 自然性の評価結果

図 10 と**図 11** に，それぞれ声質のモーフィングと歌い回しのモーフィングによる自然性の評価結果を示す．

図 10 の声質のモーフィング率により整理した結果には，予備実験の場合と同様に，外挿における自然性の顕著な低下が認められ，はっきりとはしないが，内挿の中央付近での僅かな凹みが認められる．

一方，図 11 に示す歌い回しのモーフィング率により整理した結果には，そのような凹みは認められない．これらの結果は，用いているスペクトル包絡の外挿方法がスペクトルの自然な構造を壊すものであることを意味する．

4.2.3 歌声の判定

図 12 にモーフィングされた歌声の歌唱者の判定結果を示す．予備実験と同様に，被験者は主に声質によって歌唱者を判定していることが分かる．予備実験とは異なり，今回の 2 名のプロ歌手による演奏は，歌い回しの判定が比較的容易であったにもかかわらず，予備実験と同様に，被験者は声質に基づいて歌唱者を判定した．今後，より多くの実験によって検証することが必要ではあるが，歌唱者の判定には声質が支配的であると結論づけて良いであろう．

これらの結果は，収録された歌唱のポストプロダク

ションにおいて，歌手の声質の転写や歌い回しの転写を可能にするインタフェースを導入する際の有用な知見を与える．また，歌手の判定がほぼ声質のみに依存していることは，インタラクティブなリアルタイムの歌手変換が実現可能であることを意味する．

5. おわりに

歌唱音声の歌手の声質や歌い方の自由な操作を目的に，STRAIGHT に基づくモーフィングを拡張し，パラメタ操作と知覚印象との関係を調べた．実験結果は，基本周波数と時間軸の操作により歌い回しが，スペクトル包絡および非周期性指標と周波数軸の操作により声質が選択的に変換されることを示し，歌手の個人性の判断が主に声質により行われることを示した．この最後の結果は，はじめにで紹介した願いを実現する上で，大きな応用価値を持つ．

参考文献

1) Abe, T., Kobayashi, T. and Imai, S.: Harmonics estimation based on instantaneous frequency and its application to pitch determination, *IEICE Trans. Information and Systems*, Vol.E78-D, No.9, pp.1188–1194 (1995).
2) Abe, T., Kobayashi, T. and Imai, S.: The IF Spectrogram: A New Spectral Representation, *Proc. ASVA-97*, Tokyo, pp.423–430 (1997).
3) Bregman, A.S.: *Auditory Scene Analysis*, MIT Press, Cambridge, MA (1990).
4) Dudley, H.: Remaking speech, *J. Acoust. Soc. Am.*, Vol.11, No.2, pp.169–177 (1939).
5) Kawahara, H.: STRAIGHT, Exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds, *Acoust. Sci. & Tech.*, Vol.27, No.6, pp.349–353 (2006).
6) Kawahara, H., de Cheveigné, A., Banno, H., Takahashi, T. and Irino, T.: Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT, *Interspeech'05*, Lisboa, pp.537–540 (2005).
7) Kawahara, H., Masuda-Katsuse, I. and de Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication*, Vol.27, No.3-4, pp.187–207 (1999).
8) Kawahara, H. and Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation, *Proc. ICASSP 2003*, Vol. I, Hong Kong, pp.256–259 (2003).
9) Kawahara, H., Katayose, H., de Cheveigné, A. and Patterson, R.D.: Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity, *Proc. Eurospeech'99*, Vol.6, pp.2781–2784 (1999).
10) Matsui, H. and Kawahara, H.: Investigation of Emotionally Morphed Speech Perception and its Structure using a High Quality Speech Manipulation System, *Proc. Eurospeech'03*, Geneva, pp.2113–2116 (2003).
11) Mayer, H. and Steger, C.: A New Approach For Line Extraction and its Integration in a Multi-Scale, Multi-Abstraction-Level Road Extraction System, *Mapping Buildings, Roads and other Man-Made Structures from Images* (Leberl, F., Kalliany, R. and Gruber, M., eds.), Wien, R. Oldenbourg Verlag, pp.331–348 (1996).
12) Moore, B. C. J.: *An introduction to the psychology of hearing: fifth edition*, Academic Press (2003).
13) Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H. and Irino, T.: The processing and perception of size information in speech sounds, *J. Acoust. Soc. Am.*, Vol. 117, No. 1, pp.305–318 (2005).
14) Titze, I. R.: *Principles of voice production*, Prentice Hall (1994).
15) 豊田健一，片寄晴弘，河原英紀：STRAIGHT による歌声モーフィングの初期的検討，情報処理学会研究報告，Vol.2006-MUS-64 (2006).
16) Yonezawa, T., Suzuki, N., Mase, K. and Kogure, K.: HandySinger: Expressive Singing Voice Morphing using Personified Hand-puppet Interface, *Proc. NIME2005*, Hamamatsu, pp.121–126 (2005).
17) 河原英紀，片寄晴弘：高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案，情報処理学会論文誌， Vol.43, No.2, pp.208–218 (2002).
18) 後藤真孝，橋口博樹，西村拓一，岡 隆一：RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌， Vol.45, No.3, pp.728–738 (2004).
19) 片寄晴弘，後藤真孝：音楽のデザイン転写技術の開発にむけて - CrestMuse プロジェクトの「価値」創出視点からの紹介 -，人工知能学会近未来チャレンジ，pp.1D1–4 (2006).
20) 曽我部優子，筧 一彦，河原英紀：感性情報に曖昧さがある場合の音声の心理的評価とその物理特性，聴覚研究会資料，H-2003-14, pp.77–82 (2005).