Vol.2013-MUS-100 No.21
2013/9/1

様々な歌手が同じ曲を歌った歌声の多様さを活用するシステム

都築 圭太[1,a)]　中野 倫靖[2,b)]　後藤 真孝[2,c)]　山田 武志[3]　牧野 昭二[3]

概要：本稿では，Web 上で公開されている「一つの曲を様々な歌手が歌った歌声」を活用する二つのシステムを提案する．一つは，それらの歌声を重ね合わせる合唱生成支援システム，もう一つは，それらの歌声同士や自分の歌声を比較できる歌唱力向上支援システムである．従来，複数の楽曲を用いた鑑賞や創作支援，自分が歌うだけの歌唱力向上支援は研究されてきたが，同一曲を複数人が歌った歌声を活用した合唱生成や歌唱力向上支援はなかった．合唱生成支援システムでは，歌声の出現時刻と左右チャネルの音量をマウスで直感的に調整できる．直感的な操作と，それぞれの歌が完成された作品であることを利用することで，創作と同時に鑑賞を楽しむ「創作鑑賞」も可能となる．また，歌唱力向上支援システムでは，声質（MFCC）と歌い回し（F_0 軌跡）が近い歌声同士を比較表示できる．Web 上で公開されていて再生数・マイリスト数があるため，それらの情報を活用しながら歌唱力向上に取り組める．これらのシステムを実現する信号処理技術についても説明し，特に，多数の歌声を活用した新たな F_0 推定法を提案する．

1. はじめに

近年，様々なエンドユーザによる歌声コンテンツが Web 上で大量に公開されるようになった（**図 1**）．ここで特徴的なのは，ある一つの曲（一次コンテンツ）を対象として，それを様々なユーザが自分なりに歌った二次コンテンツが大量に存在する点である [1]．さらに，それらの歌声のみで鑑賞して楽しまれるだけでなく，それらを重ね合わせた合唱[*1]（三次コンテンツ）が創作されて楽しまれている [2]．本稿では，以降これらを N 次歌声コンテンツと総称する．N 次歌声コンテンツを楽しむ視聴者は多数存在し，動画コミュニケーションサイト「ニコニコ動画」[3] には 2 次コンテンツである歌唱動画の投稿数は執筆時点で約 63 万件（208 件が再生数 100 万回以上），3 次コンテンツである合唱動画が約 1.7 万件（9 件が再生数 100 万回以上）である．動画共有サイト「YouTube」等他の様々な Web サイトにも投稿されていて，その数は増え続けている[*2]．

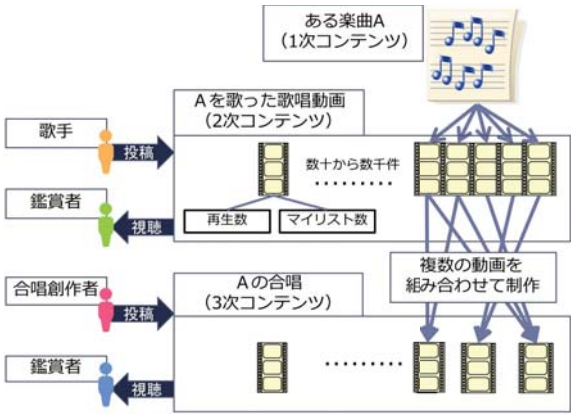

図 1　ある楽曲 A に対する N 次コンテンツ動画同士の関係と歌手・鑑賞者・合唱創作者の関係．それぞれの動画には再生数とマイリスト数（ブックマーク数）が付随している．

単独の歌声コンテンツのみでも，視聴者は同じ楽曲を様々な声質・歌い回し・アレンジで聴く楽しみが得られるが，合唱コンテンツによって，以下に示す新たな鑑賞の楽しみが可能となった．

歌の厚みや切り替わりを楽しむ　多人数が歌唱した歌声特有の音の厚みや，複数の歌声の切り替わり・掛け合い

[1] 筑波大学大学院 システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba
[2] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)
[3] 筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba
[a)] tsuzuki [at] mmlab.cs.tsukuba.ac.jp
[b)] t.nakano [at] aist.go.jp
[c)] m.goto [at] aist.go.jp

[*1] 同じ楽曲に対する歌声コンテンツはほとんどの歌手が同じメロディラインを歌っているため「斉唱」の方が音楽的には正しいが，Web 上において「合唱」と呼ばれていることが多いため，本稿でもそれに従って合唱と呼ぶ．

[*2] N 次歌声コンテンツのうち，VOCALOID 楽曲を歌ったコンテンツを含む，様々な派生コンテンツに関する増加傾向（2007〜2012 年）が，文献 [1] の Fig.3 に示されている．

等を楽しむことができる．100 人以上の規模の合唱も存在する．

合唱の違いを楽しむ　一つの楽曲に対して合唱が複数存在することも多く，異なる合唱間の違いを比較する楽しみ方がある．同じ楽曲の合唱であっても，使用する歌声や出現タイミングなど，制作者の個性が現れる．

個別の歌手の特性に気付く　一つの歌を聴いているだけでは分からなかった歌手の個性や特性に気付くことがある．その歌手の歌声をより深く理解することにつながって，歌を細部まで楽しむことができる．

新たな歌手を知る　自分の好きな歌手の歌声が使用されている合唱を聴くことで，自分が知らない歌手の歌声と出会って好きになれることがある．視聴者にとっては新たな歌手の発掘につながり，創作者にとっても自分の好きな歌手を人に推薦できる場として機能する．

ただし，公開されている N 次歌声コンテンツは，ほとんど全て伴奏が重畳されているため，このような合唱動画を制作するためには，個々の歌声に関して伴奏を抑制する処理が必要となる．さらにそれらは，全体的に時刻がずれていたり，キーが異なる場合もあるため，そのずれを補正する必要もある．また，合唱を生成するためには，個々の歌声を波形ベースで切り貼りする煩雑な手作業が必要となり，多くの鑑賞者にとって合唱創作をする敷居は高かった．

もし，直感的で簡便な合唱制作が行えれば，鑑賞者が創作と同時に自分好みの合唱の鑑賞を楽しむ「創作鑑賞」が実現できる．そのためのインタフェースは，能動的音楽鑑賞 [4] の一種として捉えることができる．従来から「楽曲」の多様さに着目した音楽鑑賞インタフェースは研究されてきたが [5]，「歌声」の多様さに着目したインタフェースは研究されていない．また，異なる楽曲を複数用いて創作を支援する研究は行われてきたが [6,7]，異なる歌声を複数用いた音楽創作については研究されていなかった．

さらに，前述した合唱動画の楽しみの一つである「**個別の歌手の特性に気付く**」を，歌手の立場から活用することで，歌唱力向上支援システムにつなげることができる．自分自身の歌がうまいかどうかを自分で判断したり，自分の歌唱のどこが悪いのかに自分で気づくことは，歌唱・音楽経験が浅いと難しい．しかし，合唱の特性を生かせば，「**自分自身の特性に気付く**」ことを支援できる．さらに，歌唱動画が Web 上で公開されている利点を生かし，再生数・マイリスト数等の客観的なメタデータも活用する歌唱力向上支援システムを提案する．従来，歌唱力向上支援として，自分の歌唱のみを対象とした研究 [8]，原曲の歌唱との自分の歌唱の比較を考慮した研究 [9] はあったが，それに対して提案システムは，様々な歌声を「お手本歌唱」として比較したり，そのメタデータを活用したりできる点が新しい．自分のお手本として，原曲の歌手よりも適した歌声を見つけることができる可能性もある．ここでは，歌声同士の比較や自分に適した歌声を見つける上で必要となる，新しい F_0（基本周波数）推定技術も併せて提案する．また，合唱における歌唱の学習を支援するシステムが香山らによって提案されている [10] が，それに対して提案システムでは自分一人での歌唱の上達を支援する点で異なる．

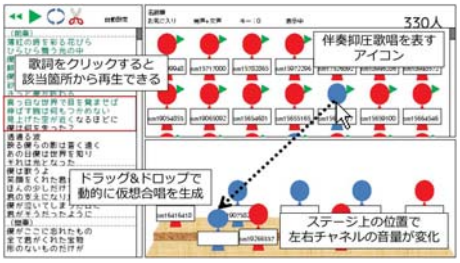

図 2　合唱制作支援システムの画面．あるタイミングで使用する伴奏抑圧歌唱を追加している．

2. 歌声の多様さを活用するシステム

本稿では，図 1 に示した N 次歌声コンテンツに関わるユーザである，「歌手」，「視聴者」，「合唱創作者」の全てを対象とするため，合唱制作支援システムと歌唱力向上支援システムの二つを提案する．この二つのシステムは，多様な歌手が同じ曲を多様に歌った歌声を活用するシステムであり，これによって，N 次歌声コンテンツの創作・鑑賞を促進し，歌声コンテンツとの関わりを豊かにする．

2.1 本研究で扱うデータ

本研究では伴奏音源が公開されている楽曲から，ニコニコ動画において歌唱動画の投稿数が最多の楽曲の歌唱動画（5247 件）とそれに付随した再生数やマイリスト数（図 1）を収集して利用した．以降では歌唱動画中の映像を除いた音響信号を伴奏付歌唱，3 章で説明する伴奏抑圧等の前処理を行った伴奏付歌唱のことを伴奏抑圧歌唱と呼ぶ．

2.2 合唱制作支援システム

合唱制作支援システム（**図 2**）は，複数の歌声を重ねた合唱コンテンツ（以下，単に「合唱」と呼ぶ）を動的に制作できるシステムである．本システムによって，ユーザは合唱を創作・鑑賞する楽しみを同時に味わうことができる．

2.2.1 インタフェースの設計

従来，合唱制作では伴奏抑圧歌唱を波形ベースで切り貼りしたり，左右チャンネルの音量を調整したり煩雑な作業を行う必要があった．本インタフェースではそれぞれの伴奏抑圧歌唱をアイコンで表し，出現タイミングをマウス操作だけで編集できる点が新しい．また，左右チャネルの音量の調節をシステム側が支援することで，音量調節の煩雑さを解消した．更に，歌詞を用いて楽曲をいくつかの時間区間（セクション）に分割し，セクション単位による伴奏抑圧歌唱の入れ替えや音量の調節を可能にした．

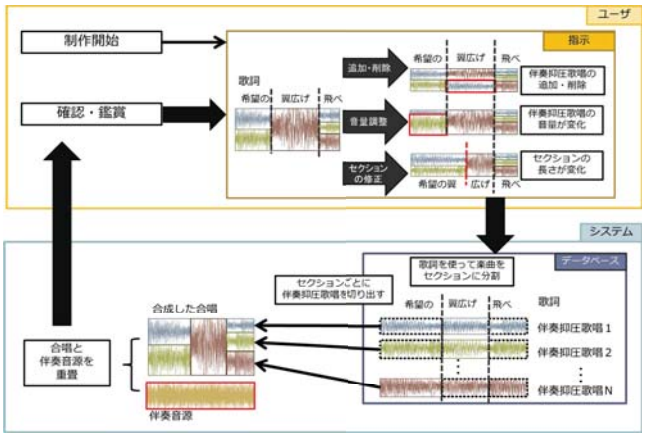

図 3 合唱創作支援システムとユーザのインタラクション.

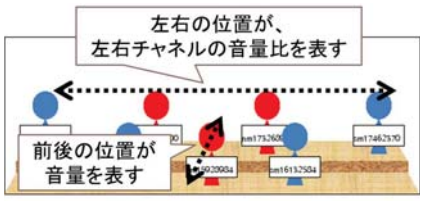

図 4 アイコンを用いた音量調節機能.

図 3 にユーザとシステムのインタラクションの流れを示す．ユーザは，伴奏抑圧歌唱のセクションへの追加・削除，セクションの範囲の変更，各セクションの伴奏抑圧歌唱の音量等をシステムに対して指示する．システムはそれに応じて動的に合唱を生成し，ユーザはそれを確認して，また指示を出す．この一連の流れを再生を止めずに行えるため，ユーザは創作と鑑賞を同時に楽しめる．

なお，本インタフェースを実現するためには歌詞と波形の対応情報と楽曲の区切り位置に関する情報を事前に用意しておく必要がある．これらについては 3 章で説明する．

2.2.2 インタフェースの機能

本インタフェースの特徴的な機能について説明する．

アイコンを用いた音量調節

DAW(Digital Audio Workstation) 等を用いた制作では，各伴奏抑圧歌唱の左右チャネルの音量を細かく設定できるものの，使用する伴奏抑圧歌唱が増えると設定が煩雑になる．そこで，本インタフェースでは同じセクションで使用している伴奏抑圧歌唱とその音量と左右チャネルの音量比が一目でわかるように，各伴奏抑圧歌唱をアイコン（以下，歌手アイコン）で表し，ステージを模した部分での歌手アイコンの配置で左右チャネルの音量を表現した．また，左右チャネルの音量比の設定を一部自動化した．ユーザは歌手アイコンを左右に動かすことで左右の音量比を，前後の段のどちらに配置するかで総合的な音量を操作する．

歌詞を用いた楽曲中の位置指定

効率的に合唱を制作するには，どのセクションが楽曲のどの範囲にあたるかを容易に把握できる必要がある．DAW

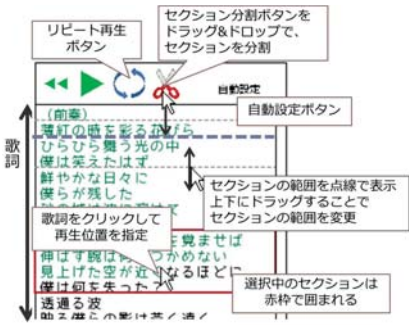

図 5 歌詞による楽曲中の位置指定と，セクションに関する操作.

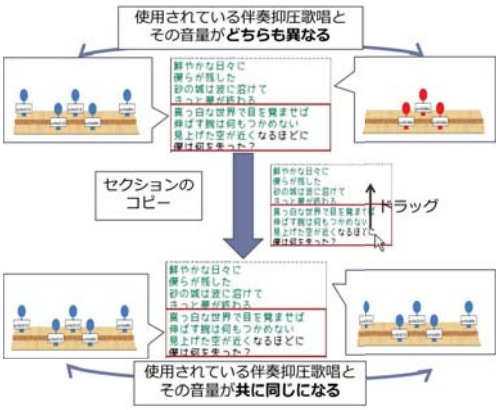

図 6 伴奏抑圧歌唱及びその音量の別セクションへのコピー.

等を用いた従来の編集では，波形を目で見ながら編集範囲を確定する必要があるため，選択した範囲が正しいかその都度聴取して確認する必要があり効率的でない．そこで歌詞を用いた楽曲中の位置の指定を行う（**図 5**）．歌詞を用いると波形に比べて楽曲中の位置の対応をつけやすいだけでなく，一覧性が高いため，効率の良い位置の指定が行える．

また，離れた時間で同じ伴奏抑圧歌唱を使用したいときに有用な操作として，本システムではコピー元のセクションからコピーしたい先のセクションへドラッグ&ドロップすることで，セクション内の各伴奏抑圧歌唱とそれぞれの音量をコピーできるようにした（**図 6**）．

フィルタリングによる検索支援

歌唱動画は数千件存在するため何らかのフィルタリングが必要である．そのため，本システムでは**図 7** で歌手アイコンが表示されているエリアの左上のボタンをクリックすることで，性別，お気に入り登録の有無（歌手アイコンをクリックでお気に入り登録できる），及びキーで伴奏抑圧歌唱のフィルタリングを行えるようにした．

2.3 インタフェースの操作方法

ユーザは前節で示した機能を用いて，合唱の鑑賞創作を行う．図 5 のように，歌詞をクリックすると再生位置が指定される．その状態で，図 2 のように歌手アイコンをドラッグ&ドロップすることで，再生中のセクションに，ドラッグした歌手アイコンの伴奏抑圧歌唱が追加される．そ

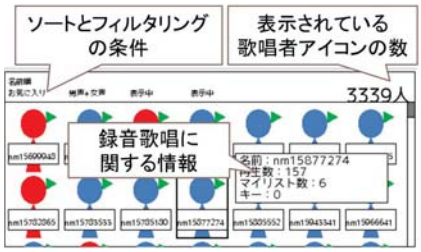

図 7 伴奏抑圧歌唱の検索に関するインタフェース．

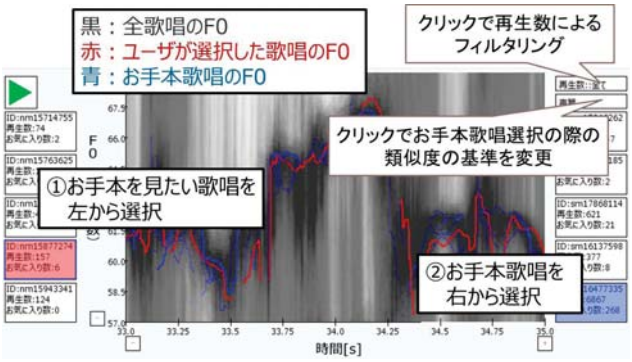

図 8 歌唱力向上支援システム．

の後，歌手アイコンを前後左右にドラッグ&ドロップすると，総合的な音量や左右の音量比が調節できる．

例えば，男声だけで合唱を制作したい，再生数やマイリスト数が高いものを使いたいなど，ある程度使用したい伴奏付歌唱が決まっている場合にはフィルタリング機能を用いることで，目的の伴奏付歌唱にたどり着きやすくなる．また，図 5 の自動設定ボタンをクリックすると，セクションと使用する伴奏付歌唱とその音量を全自動でシステムが設定するため，初期の検討に便利である．

2.4 歌唱力向上支援システム

歌唱力向上支援システム（**図 8**）は，同じ楽曲に対する様々な歌唱を活用した新しい歌唱力向上支援システムである．様々な歌手が同じ曲を歌った多様な歌唱があれば，そこから歌唱の多様性や自分の歌い方・声質の位置づけを知ることができて，歌唱力向上に有用である．従来はお手本となる歌唱は原曲の歌唱しか存在しなかったが，現在は同じ曲の歌唱が多数公開されているため，原曲の歌唱よりもユーザの参考になる歌唱が存在する可能性がある．本システムは多数の同じ曲を歌った歌唱の中から個々のユーザに適したお手本歌唱を推薦し，ユーザの歌唱とお手本歌唱の歌い回し（F_0 の軌跡）を可視化する．

本システムはユーザの歌い回し・声質がそれぞれ近いもの，もしくは両方とも近いものをユーザにお手本歌唱として提示した．例えば，歌い回しが似ている歌唱が提示されれば，同じ歌い回しに対してどんな声質の歌唱の人気が高いか参考にすることができる．声質が似ている歌唱に関しては，歌い回しに着目して歌唱同士を比較することで，自分の歌唱力向上につなげられる．また，お手本歌唱の再生数やマイリスト数も併せて表示することとした．例えば，提示された歌唱につけられた再生数・マイリスト数が高ければその曲における人気の傾向を把握する目的で活用できる可能性がある．

なお，本システムでは F_0 を MIDI ノートナンバー（以下，ノートナンバー）で表示する．周波数 f[Hz] からノートナンバー f_{MIDI} へは次のように変換した．

$$f_{MIDI} = 12 \log_2 \frac{f}{440} + 69 \qquad (1)$$

2.4.1 システムを用いた練習

自分の歌唱とお手本歌唱の F_0 軌跡を可視化し，お手本歌唱と自分の歌唱を聴き比べることで，自分の歌唱の間違いに気づくことができる．例えば，部分的に 0.5 半音程度低く歌唱しているというようなずれはある程度訓練を積まないと知覚することが難しいが，F_0 軌跡を可視化することで気づきやすくなる．逆に，ユーザが苦手であると意識している箇所については，その箇所に集中して様々な歌唱を聴くことで, 例えば抑揚のつけ方などの歌い方に関する指針が得られる．

2.4.2 システムの操作

提案システムの操作方法を説明する．

歌唱の入力

まず，画面（図 8）左側のボタンからどの歌唱に対するお手本歌唱を提示するかを選択する．表示されているボタンをクリックしたり，歌唱に関する情報が記録されたファイルをボタンへドラッグ&ドロップすることでお手本歌唱を探す際の基準となる歌唱を選択できる．ボタンには再生数とマイリスト数が表示されている．

お手本歌唱の選択

ユーザが選択した歌唱と同じキーのお手本歌唱が右側の 5 つのボタンに表示される．また，中央に選択した歌唱の F_0 軌跡が赤色で表示され，薄い青色でお手本候補の F_0 軌跡が表示される．右側のボタンをクリックするとそのボタンに対応する歌唱の F_0 軌跡が薄い青から濃い青色へ変化する．その状態で画面左上の再生ボタンを押すと，左側のスピーカから最初に選択した歌唱が，右側のスピーカからお手本の歌唱が再生される．これにより，ユーザは二つの歌唱を F_0 軌跡と実際の歌唱で比較することができる．

お手本歌唱の推薦基準の変更

右上の 2 つのボタンのうち下側のボタンを押すと，歌い回しの近さ・声質の近さ（MFCC）・両方の近さと類似度の基準を切り替えることができる．また，上部のボタンを押すと，お手本を提示する際に一定の再生数以上のものだけを提示するように設定できる．ユーザは大量の歌唱の中から自分の歌唱に近い歌唱のうち，自分が気に入ったもの，人気が高いものに絞って比較することができる．

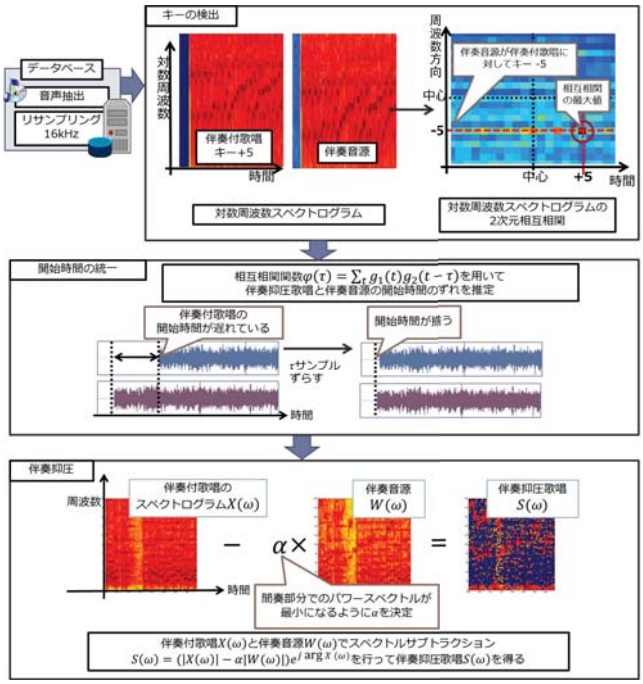

図 9 前処理の流れ.

表示する F_0 軌跡の移動

画面中央の F_0 が表示されているエリアはドラッグすると，注目する時間や周波数の範囲を動かすことができる．また，時間が書かれた軸上の＋・－ボタンをクリックすると時間方向にズームイン・ズームアウトが，周波数が書かれた軸上の同ボタンをクリックすると周波数方向にズームイン・ズームアウトが行える．

3. 提案システムを実現するための処理

本章では，両システムで必要な前処理，合唱制作支援システムで合唱を生成する際の音量の決め方やシステムの実現に必要なデータの説明，歌唱力向上支援システムにおける F_0 推定手法と類似度の計算法を説明する．特に，F_0 の推定手法では，同一楽曲を歌った多数の歌唱を活用して正しい F_0 の範囲を推定する新たな手法を提案する．

3.1 共通の前処理

各伴奏付歌唱はほぼ同じメロディを歌っているが，その中の一部は原曲に対してキーが異なり，各伴奏付歌唱の開始時間はそれぞれ異なる．伴奏抑圧のために，キーの推定と開始時間の統一を行う必要があるが，扱うデータ量が大きいため，まず粗いシフト幅でキーを推定し，その後時間方向に対して精密にずれの大きさを推定する．

なお，ユーザに提示される音も含めて，本システムで用いる全ての伴奏付歌唱と伴奏音源のサンプリング周波数は 16kHz，チャネル数は 1 とした．また，伴奏付歌唱に対応する伴奏音源は既知であるとする．**図 9** に前処理の流れを示す．

キーの推定 各伴奏付歌唱はキーと開始時間が異なっているため，伴奏付歌唱と伴奏音源の対数周波数スペクトログラムの 2 次元相互相関を計算することで，これらのずれを同時に推定した．対数周波数スペクトログラムを用いると，キーのずれを周波数方向の線形のずれとして扱える．対数周波数スペクトログラムは，窓幅 2048 サンプルのハニング窓を使用し，シフト幅は 16000 サンプル，128 個の周波数ビンがそれぞれ 1 半音を表すようにフーリエ変換をすることで求めた．
通常 1 オクターブ離れて歌唱する場合，伴奏のキーは原曲と同じものを使用するため，推定するキーの範囲はオリジナルの楽曲に対して ±6 半音以内とした．また，開始時間の統一と伴奏抑圧に用いる伴奏音源は，各キーごとに原曲の伴奏音源をピッチシフトしたものを事前に用意して利用した（今回は，Audacity[*3]のピッチシフト機能を用いた）．後述する開始時間の統一，伴奏音の抑圧はピッチシフトした音源を用いても問題なく機能した．

開始時間の統一 各伴奏付歌唱 $g_1(t)$ と伴奏音源 $g_2(t)$ の開始時間が何サンプルずれているかを，次の相互相関関数 $\phi(\tau)$ を用いて推定した．t, τ はサンプルを表す．$\phi(\tau)$ が最大となるような τ サンプルだけ，それぞれの伴奏付歌唱をずらすことで，全ての伴奏付歌唱の開始時間を伴奏音源に合わせられる．

$$\phi(\tau) = \sum_t g_1(t) g_2(t-\tau) \tag{2}$$

伴奏音の抑圧 スペクトルサブトラクション法 [11] を用いて伴奏付歌唱の伴奏音を式 (3) , (4) により抑圧する．$X(\omega)$ は伴奏付歌唱，$W(\omega)$ は伴奏音源のスペクトルであり，ω は周波数，α は伴奏音源を引くときの重み，j は虚数単位を表す．

$$S(\omega) = \begin{cases} 0 & (H(\omega) \leq 0) \\ H(\omega) e^{j \arg X(\omega)} & (otherwise) \end{cases} \tag{3}$$

$$H(\omega) = |X(\omega)| - \alpha |W(\omega)| \tag{4}$$

α の値が不適切だと音質が劣化するため，伴奏付歌唱ごとに適切な α を推定する必要がある．本研究では，$\alpha = 1$ として一度伴奏抑圧を行い，その結果から推定された非歌唱区間において $|S(\omega)|$ が一定未満になるように α を決定した．非歌唱区間においては伴奏付歌唱と伴奏音源はほぼ同じ信号なので，$|S(\omega)|$ は極めて小さくなり，非歌唱区間においては α の値によらず $|S(\omega)|$ の値が歌唱区間より小さくなる傾向がある．α の値を推定するために，まず非歌唱区間を推定する．$\alpha = 1$ で一度伴奏抑圧し，伴奏付歌唱の楽曲全体の平均パワーに対してパワーが半分未満である区間を非歌

[*3] http://audacity.sourceforge.net

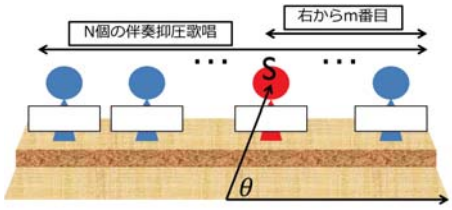

図 10 音像定位の角度決定に用いるパラメータ．

唱区間とした．パワーは窓幅 1024 サンプルのハニング窓を使用して，FFT 長 2048 サンプル，シフト幅 512 サンプルで計算した．次に α を決定するために，最長の非歌唱区間で α を 0 から 5 の範囲で 0.1 ずつ増加させながら伴奏抑圧を行った．伴奏抑圧後の非歌唱区間の平均パワーが，伴奏抑圧前の楽曲全体の平均パワーの 1%未満となる最小の α を推定結果とした．α の推定及び伴奏抑圧では，窓幅 2048 サンプルのハニング窓を使用し，FFT 長は 4096 サンプル，シフト幅は 1024 サンプルとした．

3.2 合唱制作支援システムの実現

本節では合唱制作支援システムにおいて，各伴奏抑圧歌唱の音量の設定方法と，インタフェースの実現のために必要なデータについて説明する．

3.2.1 合唱の合成

合唱を動的に生成できる鑑賞・創作システムにおいては，セクション毎に指定された伴奏抑圧歌唱を重畳することで最終的な合唱を得る．インタフェース上（**図 4**）の配置に基づいて音量と左右の音量比が決定される．

音量の決定 後段に配置した伴奏抑圧歌唱は，前段に配置した伴奏抑圧歌唱より音量が小さくなるように，振幅に 1/2 を乗じた．注目する伴奏抑圧歌唱 S の時間領域の波形が $s(t)$ であるとすると，音量が変化した波形 $s'(t)$ は次のように表せる．

$$s'(t) = \begin{cases} s(t) & (S \text{ が前段}) \\ \frac{1}{2}s(t) & (S \text{ が後段}) \end{cases} \tag{5}$$

左右チャネルの音量比の決定 まず各伴奏抑圧歌唱の音像定位の角度を決定する．同じ段に N 個の伴奏抑圧歌唱が配置され，注目する伴奏抑圧歌唱 S が右から m 番目にあり，ユーザから見た角度が右から左へ 0 から π であるとき（**図 10**），次の式（6）で S の音像定位の角度 θ が決定される．前段の伴奏抑圧歌唱は音像定位が中央（$\pi/2$）に寄りやすいようにした．

$$\theta = \begin{cases} \frac{m}{N+1}\pi & (N \neq 1 \text{ かつ } S \text{ が前段}) \\ \frac{m-1}{N-1}\pi & (N \neq 1 \text{ かつ } S \text{ が後段}) \\ \pi/2 & (N = 1) \end{cases} \tag{6}$$

このとき，伴奏抑圧歌唱 S の左チャネルのゲイン G_{SL} と右チャネルのゲイン G_{SR} は，

$$G_{SL} = \theta/\pi \tag{7}$$

$$G_{SR} = 1 - g_{SL} \tag{8}$$

とし，これを $s'(t)$ の振幅に乗じることで伴奏抑圧歌唱 s について左右チャネルの波形 $s'_L(t)$ と $s'_R(t)$ を得る．

$$s'_L(t) = G_{SL}s'(t) \tag{9}$$

$$s'_R(t) = G_{SR}s'(t) \tag{10}$$

3.2.2 インタフェースの実現に必要な情報

合唱制作支援システムのインタフェースを実現するためには歌詞と波形の対応情報，楽曲の区切り位置に関する情報の 2 つを事前に用意しておく必要がある．それぞれ，歌詞を用いた再生位置の指定，合唱の自動生成機能に使用する．現状ではどちらも手動で作成しているが，歌詞と歌声音響信号の対応情報については，これらを自動で対応付ける研究 [12] がなされているため，このような技術を組み入れることで自動化できる可能性がある．

3.3 歌唱力向上支援システムの実現

本節では歌唱力向上支援システムに利用した，F_0 推定手法と伴奏抑圧間の類似度の計算手法について説明する．

3.3.1 伴奏抑圧歌唱の F_0 推定

歌唱力向上のための鑑賞・練習システムを実現するために，各伴奏抑圧歌唱の F_0 を推定する必要がある．本研究では F_0 推定手法として，SWIPE [13] を用いて時間分解能を 10ms，周波数分解能を 0.1 半音（ノートナンバー）として F_0 推定を行った．SWIPE は高精度な手法であるものの，本研究で扱う伴奏抑圧歌唱に対しては**図 11** の赤線のように推定誤りが発生する．そこで本研究では，一つの歌声からの F_0 推定では誤りだと気付けなくても，複数の歌声からの推定結果を活用することで，F_0 の存在範囲に気付けるという仮定のもと，誤りを含んだ F_0 を多様な歌声で一旦推定し，それらの傾向から F_0 の存在範囲を推定することで，F_0 推定性能を向上させる．このような F_0 存在範囲の推定は，他の F_0 推定手法に適用することもでき，推定性能を向上させるための基本的で重要な処理である．

図 12 は不透明度 1/255 でそれぞれの伴奏抑圧歌唱の F_0 軌跡をプロットしたものを同じ曲を歌った伴奏抑圧歌唱全てについて重畳したものである．黒色の濃さはその時間にその F_0 が出現した頻度を表す．ここでプロットした F_0 軌跡はほとんどの場合，図 11 の赤線のように誤りを含むが，図 12 中に示されるように濃い黒の連続的で滑らかな軌跡とその少し下に同じ形をしたやや薄い軌跡が確認できる．このことは大量の推定結果を活用することで，誤りの少ない F_0 軌跡が得られることを示唆している．

図 13 はある楽曲の歌いだし 0.1 秒間の F_0 の出現頻度である．50 と 62 という 1 オクターブ離れた二つのノート

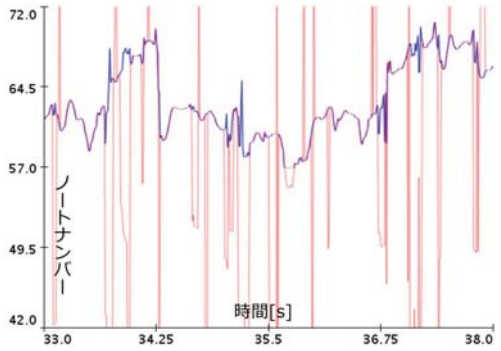

図 11 ある伴奏抑圧歌唱を SWIPE で推定して得られた軌跡（赤線）と，最頻 F_0 を用いて再推定して得られた軌跡（青線）．

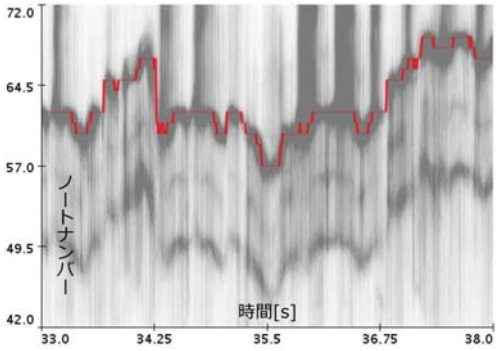

図 12 3203 人が同じ曲を歌った歌唱の F_0 軌跡の重畳表示．図中の赤い線（最頻 F_0）で示されるような濃い黒の軌跡とその下に同じ形をした軌跡があることが分かる．

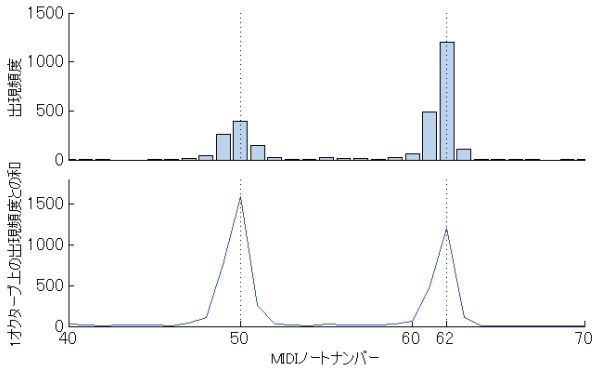

図 13 上の棒グラフはある楽曲の歌いだしにおける各伴奏抑圧歌唱の F_0 の出現頻度．12 半音離れて 2 つのピークが立っていることがわかる．下のグラフは各ノートナンバーとその 12 半音上のノートナンバーの出現頻度の和．

ナンバーでピークが立っている．この図からも誤りを含んだ大量の F_0 推定結果がいくつかの値に集中することが確認できる．以降このような多くの推定結果で共通している F_0 を最頻 F_0 と呼ぶ．最頻 F_0 は，あるノートナンバーと一オクターブ高いノートナンバーの出現頻度の和を全てのノートナンバーに対して求め，それが最大になるノートナンバーの組み合わせのうち，出現頻度が高い方のノートナンバーである．

各時間フレームにおける最頻 F_0 を用いて各歌唱の F_0 を再推定する．SWIPE を用いて時間分解能 10ms, 周波数分解能 0.1 半音，推定範囲を各時間フレームにおける最頻 F_0 から 3 半音以内として再推定した．ただし，女声曲を男性が 1 オクターブ低く歌うように，各歌手が最頻 F_0 に対して ±1 オクターブずれて歌っている可能性がある．そこで，本研究では各伴奏抑圧歌唱と最頻 F_0 及び最頻 F_0 の ±1 オクターブの軌跡との距離をそれぞれ調べ，最も小さいものを再推定に用いた．t をサンプルとするとき，伴奏抑圧歌唱の F_0 軌跡 $f(t)$ と最頻 F_0 の軌跡 $f_{mode}(t)$ の距離 D は次のように計算される．

$$D = \sum_{t=1}^{T} \sqrt{(f(t) - f_{\mathrm{mode}}(t))^2} \qquad (11)$$

以上の処理で得られた F_0 軌跡が図 11 の青線である．最初の推定に比べて F_0 軌跡の乱れが減少したことが分かる．

3.3.2 類似度計算

類似度の計算方法を説明する．いずれの類似度も小さいほど，二つの伴奏抑圧歌唱が似ていることを表す．

歌い回しの類似度 各歌唱の F_0 軌跡の類似度として計算した．ある 2 つの伴奏抑圧歌唱の歌唱の F_0 軌跡を $f_1(t), f_2(t)$，楽曲全体の時間フレーム数が T であるとき類似度 P_{12} は，次のようにして求めた．

$$P_{12} = \sum_{t=1}^{T} \sqrt{(f_1(t) - f_2(t))^2} \qquad (12)$$

声質の類似度 声質の類似度については各伴奏抑圧歌唱について MFCC をフレーム幅 25ms で 10ms ずつシフトしながら求め，得られたケプストラムの低次 13 次元について各次元ごとに平均が 0，分散が 1 になるように正規化を行った．n がケプストラムの次元のインデックスを表す時，正規化によって得られる伴奏抑圧歌唱に対応する行列は $F_1(t,n), F_2(t,n)$ と表せる．楽曲全体の時間フレーム数を T とすると類似度 Q_{12} は，次の式のようになる．

$$Q_{12} = \sum_{t=1}^{T} \sum_{n=1}^{13} \sqrt{(F_1(t,n) - F_2(t,n))^2} \qquad (13)$$

総合的な類似度 $F_1(t,n), F_2(t,n)$ に伴奏抑圧歌唱の F_0 軌跡を表す $f_1(t), f_2(t)$ を加え，この次元においても正規化を行って得られた 14 次元の行列 $F_1'(t,n), F_2'(t,n)$ に対して，声質の類似度と同様にして類似度を求めた．

4. 考察

本章では，本研究の意義や位置づけ，今後の可能性と課題について述べる．

4.1 本研究の意義・貢献

本研究では，音楽経験が浅いユーザでも活用できるようにインタフェースを設計した．創作支援や練習支援に関す

る研究は，クリエイターの裾野を広げたり，コンテンツの質を底上げし，UGC（User Generated Content）文化のさらなる発展に貢献する．また，歌唱動画は同じ楽曲を歌った歌声を含むデータベースとみなすことができる．本稿で提案した信号処理技術は，これらを用いた分析する際の基礎として重要である．

4.2 今後の展望

本研究をさらに発展させることで，ユーザが音楽を鑑賞する力を身につけられるようなシステムの実現を目指す．音楽を鑑賞する力とは，その音楽の作曲や演奏の意図を理解する力であり，意図を理解することで，その楽曲を聴取した際の感動が深まると考えられる．大串 [14] も演奏会で音楽を聴取した際に，受ける感動が大きくなる要因の一つとして，聴覚だけで鑑賞するよりも演奏している様子という視覚情報が加わることによって演奏者の意図がわかりやすくなることをあげている．

一方で，大串の指摘は音響情報だけで意図を理解することはさほど容易ではないことも示している．後藤は音響情報から楽曲の構造を可視化したりその一部を操作することで「音楽を聴く力」を豊かにする「音楽理解力拡張インタフェース」を提唱している [4]．本研究で提案した鑑賞と創作を組み合わせるインタフェースは，鑑賞しながら自分や他人の歌声の特徴に気づくことができるという意味で音楽理解力拡張インタフェースの一つと位置づけられる．

また，感動に関する研究として，大出らはある楽曲に対してなぜ感動したのか・どこに感動したのかは聴取者によって大きく異なることを指摘した [15]．通常，このような議論を行う際は，異なる楽曲を用いて実験を行うため，楽曲間の差が大きく比較が難しい．しかし，本研究で扱ったような歌唱動画においては，伴奏は同じで歌声だけが異なる楽曲が多数存在するため，聴取者が何に着目し，どこに感動しているのかを調べる上で非常に有益であると考えられる．

4.3 今後の課題

本研究では歌唱動画で用いられた伴奏音源が公開されていることを前提とした．これを伴奏がない歌唱動画についても実現できるように拡張できれば扱える歌唱を増やすことができる．また，本研究では鑑賞と創作を融合させたシステムと，練習支援のためのシステムをそれぞれ提案したが，これら 2 つのシステムを融合させることでユーザの音楽を聴く力をさらに向上させるようなインタフェースを生み出せないかと考えている．

5. おわりに

本稿では様々な歌手が同じ楽曲を歌った歌声の多様さを活用したシステムと，それらを実現するための信号処理技術を提案した．今後は鑑賞・創作・練習支援を融合した新たなインタフェースの検討を行っていく予定である．

謝辞 本研究の一部は，科学技術振興機構 OngaCREST プロジェクトによる支援を受けた．また，ニコニコ動画上の合唱動画を扱うために濱崎 雅弘氏，石田 啓介氏にご協力頂きました．深く感謝致します．

参考文献

[1] Hamasaki, M. and Goto, M.: Songrium: A Music Browsing Assistance Service Based on Visualization of Massive Open Collaboration Within Music Content Creation Community, *Proc. WikiSym + OpenSym 2013* (2013).
[2] 濱野智史：インターネット関連産業，デジタルコンテンツ白書 2009，pp. 118 – 124 (2009).
[3] ニワンゴ：ニコニコ動画，`www.nicovideo.jp`.
[4] 後藤真孝：音楽音響信号理解に基づく能動的音楽鑑賞インタフェース，情処研報，Vol. 2007-MUS-96, pp. 59–66 (2007).
[5] 堀内直明，薗田俊行，田中浩司，田中淳一，長沢秀哉，荻山真一：Song Surfing: 類似フレーズで音楽ライブラリを散策する音楽再生システム，*PIONEER R&D*, Vol. 17, No. 2 (2007).
[6] 宮島 靖：Music Mosaic Generator: 高精度時系列メタデータを利用した音楽リミックスシステム，WISS 2007 論文集，pp. 13–18 (2007).
[7] Tokui, N.: Massh!—A Web-based Collective Music Mashup System, *Proc. DIMEA 2008*, pp. 526–527 (2008).
[8] Hoppe, D., Sadakata, M. and Desain, P.: Development of Real-time Visual Feedback Assistance in Singing Training: a Review, *J. Computer Assisted Learning*, Vol. 22, pp. 308 – 316 (2006).
[9] Nakano, T., Goto, M. and Hiraga, Y.: MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data, *Proc. ISM2007*, pp. 75–76 (2007).
[10] 香山瑞恵，中西 将，岡部真実，浅沼和志，伊東一典，為末隆弘，橋本昌巳：指導者知識に基づく合唱学習支援システムの構築とその評価，情処学論，Vol. 51, No. 2, pp. 365–379 (2010).
[11] Boll, S. F.: Suppression of Acoustic Noise in Speech Using Spectral Subtraction, *IEEE Trans. Acoust. Speech Signal Process*, Vol. ASSP-27, pp. 113–120 (1979).
[12] Fujihara, H., Goto, M., Ogata, J. and Okuno, H. G.: LyricSynchronizer: Automatic Synchronization System Between Musical Audio Signals and Lyrics, *IEEE J. Selected Topics in Signal Processing*, Vol. 5, No. 6, pp. 1251–1261 (2011).
[13] Camacho, A.: SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech And Music, PhD Thesis, University of Florida (2007).
[14] 大串健吾：感心させる演奏と感動させる演奏，日本音響学会誌， Vol. 56, No. 5, pp. 349–353 (2000).
[15] 大出訓史，今井 篤，安藤彰男，谷口高士：音楽聴取における”感動”の評価要因—感動の種類と音楽の感情価の関係，情処学論，Vol. 50, No. 3, pp. 1111–1121 (2009).