歌声の旋律と動的変動を特徴付けるための確率的な表現手法に関する検討

大石 康智 †　　後藤 真孝 ††　　伊藤 克亘 †††　　武田 一哉 †

† 名古屋大学大学院情報科学研究科

†† 産業技術総合研究所,　††† 法政大学情報科学部

†ohishi[at]sp.m.is.nagoya-u.ac.jp, kazuya.takeda[at]nagoya-u.jp

††m.goto[at]aist.go.jp, †††itou[at]k.hosei.ac.jp

あらまし　本研究では，歌声の旋律とビブラートのような動的変動を同時に特徴付けるための新しい表現手法を提案する．歌声の旋律を表す基本周波数 (F0) が動的システムによって生成されると想定し，F0 とその時間微分 ΔF0 からなる相平面を利用して F0 軌跡を表現する．この相平面では，歌声の音高遷移が動的システムのアトラクタによって特徴付けられる．つまり，旋律を構成する音符の音高はアトラクタの位置に対応し，動的変動はアトラクタの渦軌跡の形状に現れる．したがって，動的変動のような揺れを含む歌声に対しても，アトラクタの位置から旋律情報を正確に抽出できると考えた．そこで，相平面におけるアトラクタの位置を確率的に表現し，その確率分布をハミング検索のための旋律の類似尺度に適用した．実験結果より，提案法による類似尺度が，従来の DTW に基づく尺度よりも効果的に検索結果を絞り込むことが可能であった．

A stochastic representation of sung melody and the dynamic characteristics

Yasunori OHISHI†　　Masataka GOTO††　　Katunobu ITOU†††　　Kazuya TAKEDA†

†Graduate School of Information Science, Nagoya University

††National Institute of Advanced Industrial Science and Technology (AIST)

†††Faculty of Computer and Information Sciences, Hosei University

Abstract　In this paper, we propose a stochastic representation of a sung melodic contour, which can characterize both musical-note information and the dynamics of singing behaviors included in the melodic contour. We assume that the F0 trajectories are generated by a dynamic system and represented in a F0-ΔF0 phase plane. In this plane, a fluctuation in a sung melody can be modeled by a damped oscillation of the dynamic system and appears as a curling trajectory around a certain target point, i.e., an attractor of the system. The advantage of this modeling is that the location of each attractor corresponds to the F0 of its target musical note and typical singing behaviors can be characterized by the shape of curling trajectories. By using this representation, we also define a melodic similarity measure for query-by-humming (QBH) applications. Our experimental results show that the proposed similarity measure is superior to a conventional dynamic-programming-based method.

1　はじめに

本研究では，歌声の旋律情報 (音の高低や長短) だけでなく，歌唱者ごとの演奏表現や癖が現れる音高の連続的な遷移，ビブラートのような動的変動を特徴付けた信号モデルの構築を目指す．例えば，音高を探りながらゆっくりと遷移させる歌唱者もいれば，素早く的確に次の音高を発声できる歌唱者もいる．音高が安定する箇所では，表現豊かなビブラートをかける歌唱者もいる．歌声は，多くのジャンルの音楽を特徴付ける

重要な要素の一つであり，現在様々な研究 [1, 2, 3, 4, 5] がされているが，歌唱者ごとに異なる，歌声の動的変動についてはまだ十分に検討されていない．

また，ハミング検索などで利用される歌声の旋律情報の表現方法といえば，信号のパワーを利用して，基本周波数 (F0) の軌跡を音高と音長を表すシンボル列に変換し，ngram モデルのような離散的な確率表現を利用することが一般的であった [5, 6, 7, 8]．しかし，タタタやチャチャチャのような閉鎖音によるハミングに比べて，歌詞付きの歌声，さらに動的変動を含む歌声は，旋律情報を正しくシンボル列で表現することが難しい．これに対して，F0 軌跡の DTW に基づく照合方法が提案され，高い検索性能が報告されているが [9, 10, 11]，どの程度の歌声の動的変動にまで耐えうる照合方法であるのか十分に検討されていない．

その他，HMM や制動 2 次系のインパルス応答を利用した F0 制御モデルにより，自然性かつ明瞭性のある歌声合成が実現されている [4, 12]．ただ，同じ旋律でも人それぞれ歌い方が異なるように，合成音声の多様性については，さらに検討する必要がある．

そこで我々は，歌声の旋律情報と動的変動を同時に特徴付けるための新しい表現手法を提案する．歌声の F0 が動的システムによって生成されると想定し，F0 とその時間微分 ΔF0 からなる相平面で F0 軌跡を表現すると，旋律の音符の音高はアトラクタの位置，動的変動はアトラクタの渦軌跡の形状によって特徴付けられる．さらに，相平面におけるアトラクタの位置を確率的に表現し，その確率分布をハミング検索のための旋律の類似尺度に適用した．実験結果より，提案法による類似尺度が，従来の DTW に基づく尺度よりも効果的に検索結果を絞り込むことが可能であった．

以下，第 2 章では相平面における F0 軌跡の特性を述べ，第 3 章では相平面を利用して歌声の旋律情報を確率的に表現する手法を提案する．第 4 章では提案法をハミング検索のための旋律の類似尺度に適用し，その有効性を確認するための評価実験を行う．第 5 章では実験結果を考察し，第 6 章でまとめと今後の課題について述べる．

2　相平面における歌声の F0 軌跡

歌声の旋律を表す F0 が動的システムによって生成されると想定する．図 1(b) は，図 1(a) の歌声の F0 軌跡を 2 次元の相平面 $\vec{f}(x, \dot{x})$ 上に図示した例である．

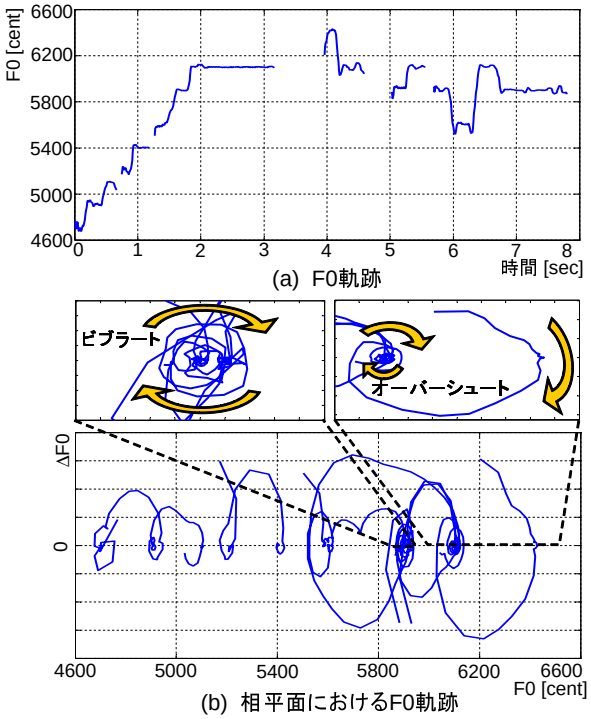

図 1　F0-ΔF0 の相平面に表現される歌声の F0 軌跡：音高の遷移が，複数のアトラクタとそれらを遷移する動きによって表現される．歌声に特有な動的変動であるビブラート，オーバーシュートが，楕円または螺旋を描く軌跡によって表現される．

相平面 $\vec{f}(x, \dot{x})$ は，F0 軌跡の局所的な方向 (ベクトル) を表現することができる．ここで x は F0，$\dot{x}$ は F0 の時間微分を表す．F0 の時間微分は，微小区間の回帰係数 ΔF0 で近似した．この平面には，F0 軌跡が渦を描きながら，ある点に引き寄せられる動き，すなわち動的システムのアトラクタが複数個観測される．また，アトラクタから別のアトラクタに遷移する動きが観測される．これらのアトラクタの位置は，歌声の旋律を構成する音符 (楽譜に記される音符) の音高に対応する．一方，アトラクタの渦軌跡は，歌唱者の意図する演奏表現や癖を特徴付ける．例えば，音高安定時に準周期的な振動を繰り返すビブラートは，音符の音高を中心に，その周りで楕円を描く軌跡として観測される．これは動的システムの非減衰調和振動と同じ動きである．また，音高が遷移する時に目的音高より大きく振れてしまうオーバーシュートは，螺旋を描きながら目的音高に引き寄せられる軌跡として観測される．これは動的システムの減衰振動と対応づけられる．

さらに同じ旋律を異なる歌唱者が歌ったときの F0 軌跡を図 2 に示す．歌唱者 A は持続して振幅の大きい

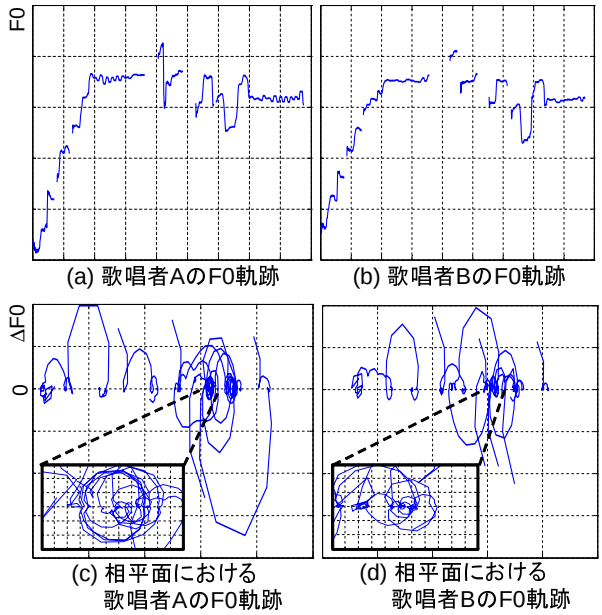

図 2　2 人の歌唱者が同じ旋律を歌唱したときの F0 軌跡：相平面で F0 軌跡を表現することにより，音高の遷移の仕方，ビブラートのかけ方のような歌い方の違いが特徴付けられる．

ビブラートをかけるため，相平面では一定の楕円軌跡が描かれる．一方で，歌唱者 B はビブラートをかけないため，相平面のある 1 点に軌跡が集中する傾向にある．ビブラートだけでなく，ある音高から別の音高への遷移の仕方にも違いがみられる．以上のように F0 軌跡を相平面に表現することで，旋律情報と歌唱者の演奏表現や癖に基づく動的変動を同時に特徴付けることができる．

3　相平面を利用した歌声の旋律情報の確率的表現手法

相平面を利用して，歌声の旋律情報を確率的に表現する手法を提案する．前章で述べたように相平面に現れるアトラクタの位置は，歌声の旋律を構成する音符の音高に対応する．一方で，歌唱者の演奏表現や癖による “揺れ” は，その周囲の軌跡の形状に現れる．したがって，揺れを含む歌声であっても，F0 軌跡が密となるアトラクタの中心を特定することにより，安定して旋律情報を抽出できると考えた．

本手法では，事前に学習データを利用して相平面を有限個の領域に分割し，各領域が占める特徴ベクトル (F0，ΔF0) の割合から，アトラクタの中心を特定することを試みる．以下にその操作手順を説明する．

3.1　特徴抽出

歌声の F0 は，de Cheveigne らの提案した YIN[13] を利用して 10ms ごとに推定された．なお，本論文では以下，対数スケールの周波数を cent の単位 (本来は音高差 (音程) を表す尺度) で表し，Hz で表された周波数 f_{Hz} を，次のように cent で表された周波数 f_{cent} に変換する．

$$f_{cent} = 1200 \log_2 \frac{f_{Hz}}{440 \times 2^{\frac{3}{12}-5}} \tag{1}$$

さらに時刻 t における F0 の時間微分は，以下の式から計算される微小区間の F0 軌跡の傾き ΔF0 で近似する．ΔF0 の計算区間は 50ms とした．

$$\Delta f[t] = \frac{\sum_{k=-2}^{k=2} k \cdot f[t+k]}{\sum_{k=-2}^{k=2} k^2} \tag{2}$$

ここで，$f[t]$ は時刻 t における F0(単位: cent) であるとする．以上によって求めた F0 と ΔF0 の系列を本論文では以下，特徴ベクトル列と呼ぶ．ただし，無声音または休止のため，F0 が推定されない，もしくは ΔF0 が計算できない区間は取り除く．

3.2　相平面の領域分割と旋律の確率表現

まず，学習データの特徴ベクトルをクラスタリングすることによって，相平面を有限個の領域に分割するための符号帳 $\boldsymbol{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_M\}$ を作成する．ここで M はクラスタの数 (相平面を分割する領域の数) であり，$\boldsymbol{v}$ を以下，重心ベクトルと呼ぶ．クラスタリングの方法として，LBG アルゴリズムを利用した．

長さ L の特徴ベクトル列 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L\}$ によって描かれる相平面の F0 軌跡から，以下に定義される符号帳 $\boldsymbol{V}$ の出現確率 $\boldsymbol{p} = (p_1, p_2, \ldots, p_M)$ を計算することによって，歌声の旋律情報が確率的に表現される．

$$p_m = \frac{K_m}{L} \qquad (m = 1, 2, \ldots, M) \tag{3}$$

$$K_m = \sum_{l=1}^{L} w_{lm} \tag{4}$$

$$w_{lm} = 1/\Big(\sum_{n=1}^{M} \big(\frac{||\boldsymbol{x}_l - \boldsymbol{v}_m||_2}{||\boldsymbol{x}_l - \boldsymbol{v}_n||_2}\big)\Big) \tag{5}$$

ここで，$||\cdot||_2$ はベクトル間のユークリッドノルムを表す．式 (5) の w_{lm} は特徴ベクトル $\boldsymbol{x}_l$ と周囲の重心ベクトルとの距離に基づいて計算される，重心ベクト

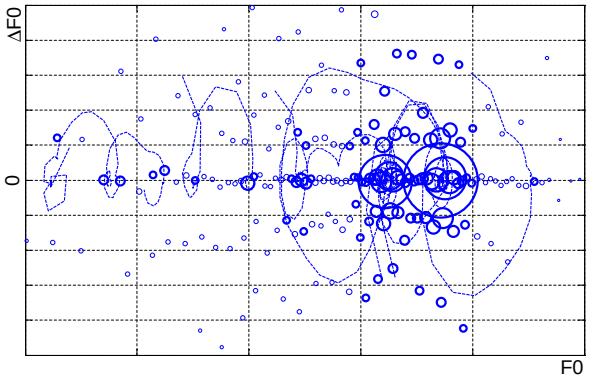

図 3　相平面を利用した歌声の旋律情報の確率的表現：○印の位置は，相平面を 256 個の領域に分割したときの重心に対応し (符号帳の大きさ $M = 256$)，○印の大きさは，重心の位置が旋律を構成する音符の音高 (アトラクタの中心) となる確率を表す．図 1(b) の相平面に適用した結果である．

ル $\boldsymbol{v}_m$ の重みである．式 (4) では，特徴ベクトルごとに計算されるこの重みを総和し，式 (3) において長さ L で正規化したものが $\boldsymbol{v}_m$ の出現確率 p_m となる [14]．アトラクタの中心では F0 軌跡が密になるため，その付近に配置された重心ベクトルの出現確率は大きくなる．したがって，アトラクタの位置を確率的に表現することができる．

図 3 は，図 1(b) の F0 軌跡 (特徴ベクトル列) から計算される符号帳の各重心ベクトルの出現確率を○印の大きさによって示したものである．特に出現確率の大きいものは，印を太線で示した．符号帳の大きさ M は 256 である．符号帳の学習に利用したデータについては 4.2 節で述べる．ΔF0 が 0 付近で，出現確率が大きい重心ベクトルの位置と，F0 軌跡のアトラクタの位置がおおよそ一致していることがわかる．以上のような相平面を利用した歌声の旋律情報の確率的な表現手法を *Stochastic Phase Representation* (以下，SPR) と呼ぶことにする．

4　評価実験

前章で提案した SPR が，揺れを含む歌声に対しても適切に旋律情報を表現できることを検証するために，SPR をハミング検索のための旋律の類似尺度に適用する．つまり，ユーザーによって入力される歌声 (以下，入力信号と呼ぶ) と，データベースにおける楽曲の旋律 (以下，参照信号と呼ぶ) との類似尺度に，それぞれの信号から作成される SPR を利用し，その検索性能について評価する．

4.1　実験条件

「RWC 研究用音楽データベース：ポピュラー音楽」(RWCMDB-P-2001)[15] の計 100 曲から，歌唱の出だしの部分と一番代表的な盛り上がる主題の部分の 2 箇所を切り出し，全 200 種類の参照信号からなる楽曲データベースを構築した．これらの信号の切り出し区間は，その部分の歌詞の始まりから区切りの良いところまでとし，平均 11.7 秒であった．また，本来ならばこれらの信号から F0 を推定することが望ましいが，今回は提案法の性能の上限を調べるために，楽曲データベースに関しては F0 を手作業でラベル付けした結果 [16] を用いた．

歌声研究用音楽データベース「AIST ハミングデータベース」[17] の一部である，日本人歌唱者 75 名 (男性 37 名，女性 38 名) が，上記の楽曲データベースの 200 種類の参照信号のうち 50 種類を歌詞付きで歌唱した計 3,750 サンプルを，入力信号として利用する．歌唱者は伴奏なしで，自由なテンポで歌唱した．歌唱時間は，平均 12.0 秒であった．歌唱者は，初めて聴くポピュラー音楽をうろ覚えの状態で歌唱したため，収録された歌声は原曲の旋律に比べて多少の揺れを含んでいる．これらの揺れは，ビブラートのような演奏表現によるものばかりでなく，うろ覚えのために生じた音符の挿入や置換，削除によるものでもある．

4.2　照合方法

前処理として，各信号ごとに F0 の平均値を計算し，F0 軌跡からこの平均値を減算する．これは，歌唱者が原曲の旋律とは異なる音の高さで歌う，移調に対応するためである．また，参照信号の特徴ベクトル列すべてを相平面に図示し，LBG アルゴリズムを利用して M 個の重心ベクトルからなる符号帳を作成する．

入力信号と参照信号からの SPR の作成と照合方法の概略を図 4 に示す．まず，各信号の特徴ベクトル列を長さ L で分割し，符号帳を利用して各分割区間内で SPR を作成する．長さ L に満たない，信号の最後の分割区間に対しても同様に SPR を作成する．特徴ベクトル列を分割する理由は，音符の順序を考慮した照合を行うためである．現状の SPR では，複数のアトラクタの位置を特定できたとしても，それらのアトラクタが出現する順序，つまり音符の順序 (ド－ミ－ソと

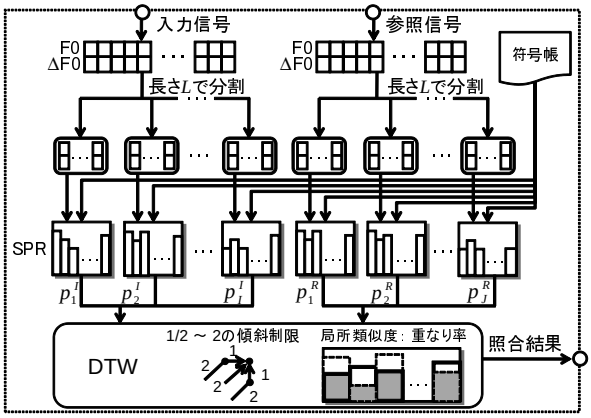

図 4　SPR を利用した入力信号と参照信号の照合方法

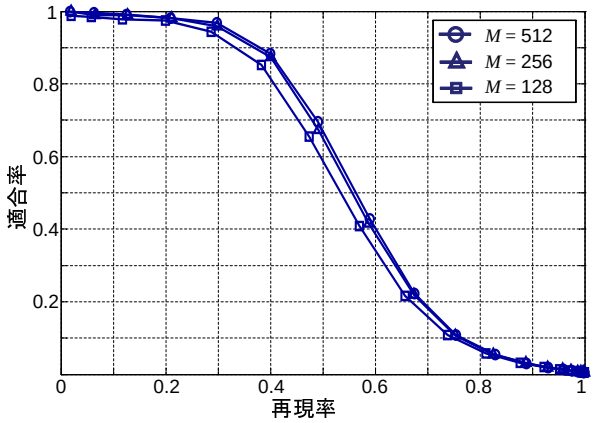

図 5　符号帳の大きさ M の変化に対する検索性能の推移：M を大きくするにつれて検索性能が向上した．$M = 1024$ の結果は $M = 512$ の結果と大きな差はみられなかった．

ミーソード) を表現できないためである．以上により，入力信号から I 個の SPR，参照信号から J 個の SPR が作成されたとする．

最終的に DTW を利用して，時間伸縮を考慮した SPR 間の照合を行う．図 4 の下部に示すように，DTW では局所的な傾斜を 1/2 と 2 の間に制限し，以下の再帰式を利用して類似度を計算する．

$$S(i,j) = \max \begin{cases} S(i-2,j-1) + 2s(\boldsymbol{p}_{i-1}^I, \boldsymbol{p}_j^R) \\ \qquad\qquad + s(\boldsymbol{p}_i^I, \boldsymbol{p}_j^R) \\ S(i-1,j-1) + 2s(\boldsymbol{p}_i^I, \boldsymbol{p}_j^R) \\ S(i-1,j-2) + 2s(\boldsymbol{p}_i^I, \boldsymbol{p}_{j-1}^R) \\ \qquad\qquad + s(\boldsymbol{p}_i^I, \boldsymbol{p}_j^R) \end{cases} \tag{6}$$

SPR 間の局所類似度 $s(\boldsymbol{p}_i^I, \boldsymbol{p}_j^R)$ は，ヒストグラムの重なり率 [18] を利用する．

$$s(\boldsymbol{p}_i^I, \boldsymbol{p}_j^R) = \sum_{m=1}^{M} \min(p_{im}^I, p_{jm}^R) \tag{7}$$

ここで $\boldsymbol{p}_i^I$ と $\boldsymbol{p}_j^R$ は，それぞれ入力信号と参照信号の i 番目，j 番目の時間分割における SPR であり，p_{im}^I，p_{jm}^R はそれぞれの m 番目の重心ベクトルの出現確率に対応する．$S(1,1) = 2s(\boldsymbol{p}_1^I, \boldsymbol{p}_1^R)$ として計算を繰り返し，最後に $S(I,J)/(I+J)$ として，入力信号と参照信号との時間正規化後の類似度が求まる．

4.3　符号帳の大きさに対する性能評価

3.2 節で作成する符号帳の大きさ M を変化させたときの検索性能を評価する．評価尺度として，再現率と適合率の関係を利用する．3,750 サンプルの入力信号と 200 種類の参照信号との照合から得られる 750,000(3,750×200) 個の類似度に対して，閾値 σ を設定しながら以下のように再現率と適合率を計算する．

$$\text{再現率} = \frac{R}{C} \qquad \text{適合率} = \frac{R}{N} \tag{8}$$

R：入力信号とその正解にあたる参照信号との類似度 3,750 個のうち，閾値 σ を上回った数
C：入力信号の数 (3,750 サンプル)
N：すべての入力信号と参照信号との類似度 750,000 個のうち，閾値 σ を上回った数

再現率と適合率の関係を表す曲線が右斜め上方向に移動するほど，入力信号から，正解にあたる参照信号を絞り込む検索性能が改善されることを意味する．

図 5 は M を 128, 256, 512 と変化させたときの結果である．照合における時間分割の長さ L は 10 フレーム (100ms) とした．M を大きくするにつれて，わずかであるが検索性能が改善された．$M = 1024$ の結果は，$M = 512$ の結果と大きな差はみられなかった．M の大きさについては，情報量基準などを利用して最適に決定することがさらなる課題である．

4.4　時間分割の長さに対する性能評価

図 4 の照合方法における時間分割の長さ L を変化させたときの検索性能を評価する．図 6 は L を 10 フレーム (100ms) から 200 フレーム (2s) の間で変化させたときの結果である．L を短くするにつれて検索性能が改善された．これは，F0 軌跡の分割区間ごとに SPR を作成することにより，長時間の信号から作成される SPR では表現できない，音符の順序と時間伸縮を考慮できたためであると考えられる．L が 5 フレームの結果は，L が 10 フレームの結果と大きな差はみら

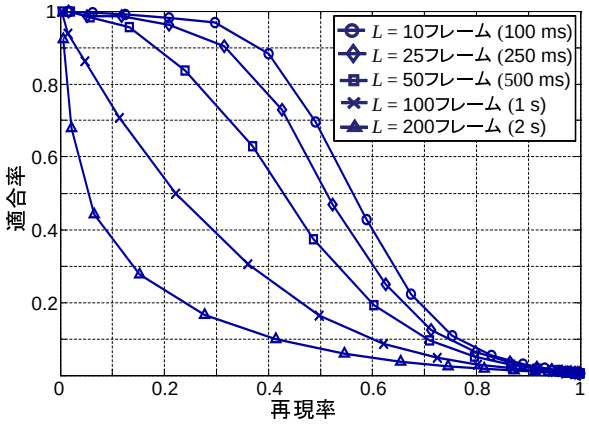

図 6　時間分割の長さ L の変化に対する検索性能の推移：L を小さくするにつれて検索性能が向上した．L が 5 フレームの結果は L が 10 フレームの結果と大きな差はみられなかった．

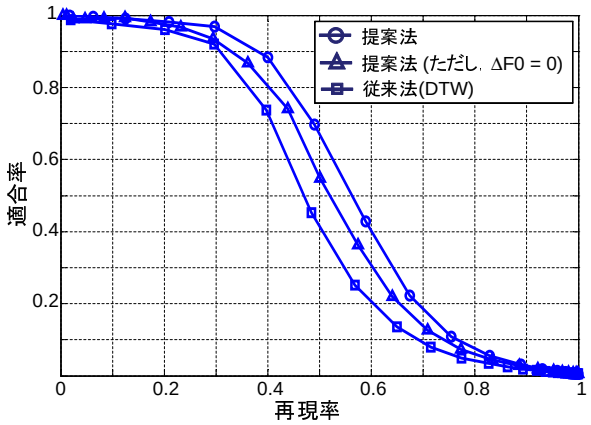

図 7　再現率と適合率の関係からみた従来法と提案法の性能比較：従来の DTW による照合に比べて，提案法の有効性が確認できる．また ΔF0 を 0 としたときの性能と比較しても，提案法において相平面を利用することの有効性が確認できる．

れなかった．

4.5　従来の DTW による照合方法との比較

従来の，F0 軌跡を DTW によって時間的に対応づける照合方法 [10] と提案法の性能を比較する．提案法のパラメータは $M = 512$，L は 10 フレームとした．また，相平面を利用することの有効性を確認するために，ΔF0 を 0 として提案法を実行した結果も示す．この場合のパラメータは $M = 256$，L は 10 フレーム (パラメータを変化させて最も性能が高かった設定) である．図 7 より，相平面を利用した提案法の検索性能が最も高いことがわかる．ただし，表 1 に示す 1 位検索率に基づいて性能を比較すると，提案法と従来法とでは大きな性能の差はみられなかった．以上のことから，提案法によって 1 位検索率は改善されないものの，入力信号とその正解にあたる参照信号との類似度が改善されたことで，提案法が検索結果を絞り込むことに有効な手法であると考えられる．

表 1　1 位検索率からみた従来法と提案法の性能比較：従来法と提案法では 1 位検索率に大きな性能の差はみられなかった．

	提案法	提案法 ($\Delta F0 = 0$)	従来法 (DTW)
1 位検索率 [%]	62.2	56.9	62.6

5　考察

5.1　従来法と提案法の類似尺度の違い

図 8，9 の上図は，同じ旋律からなる入力信号と参照信号を従来の DTW によって時間正規化させた結果である．一方，下図は，入力信号と参照信号を提案法によって時間正規化させた結果である．正規化させた各時刻ごとの入力信号と参照信号の SPR の重なり具合を示すために，出現確率が一定値以上大きい重心ベクトルの F0 値の範囲を図中に示した．2 つの範囲の重なる部分が入力信号と参照信号の SPR の重なる部分として見ることができる．また各時刻の SPR において，出現確率の最も大きい重心ベクトルの F0 値を線で結ぶことにより，入力信号と参照信号のおおよその本来の F0 軌跡が表される．

従来法を利用した場合，図 8，9 の旋律の検索結果は，それぞれ 39 位，28 位であったが，どちらも提案法によって類似度が改善され，1 位に検索された．各図の上図より，従来法では DTW によって大方の時間的な対応はとれているものの，参照信号の動きに対して，入力信号が追随できていない様子がわかる．つまり，楽曲の旋律の細かい動きを，歌唱者の歌声では正しく表現できておらず，部分的に音を外す箇所もみられる．このような従来の DTW の 1/2〜2 の傾斜制限では考慮できない F0 の挿入，置換，削除によって，入力信号と参照信号との類似度が結果的に低下してしまったと考えられる．

一方，提案法では，照合に DTW を利用しているものの，相平面を利用して歌声の旋律を確率的に表現する SPR という粗い表現方法が，ある種のフィルタを

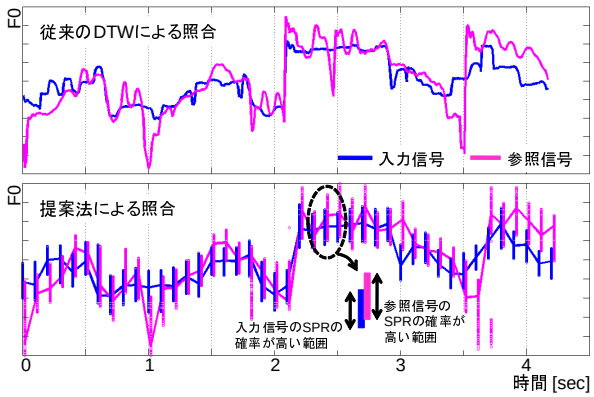

図 8　従来法と提案法による旋律の照合結果例 1：従来法では，DTW によって時間正規化しても，入力信号が参照信号の細かい動きに追随できていないことがわかる．一方，提案法は，旋律に範囲をもたせて照合を行うため，多少の揺れを吸収できる照合方法であると考えられる．

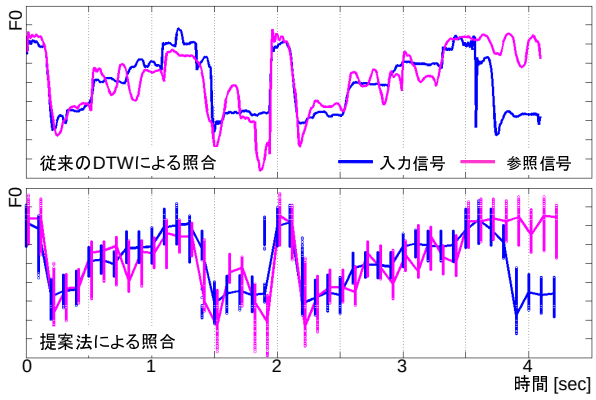

図 9　従来法と提案法による旋律の照合結果例 2

F0 軌跡にかけることと同等になり，揺れを含む歌声に対しても，正解の楽曲の旋律と類似度が高いものとして計算されたと考えられる．また，各時刻においてSPR の重なり率を局所類似度に用いたことによって，より柔軟な照合を行うことができたと考えらえれる．

5.2　旋律のテンポと検索性能の関係

図 8，9 の上図の F0 軌跡から，入力信号が参照信号の細かい動きに追随できていないことがわかった．これは，歌唱者が旋律の詳細を理解せずにうろ覚えの状態で歌唱したため，局所的に音符を挿入したり，削除していることを意味する．このことを踏まえ，旋律のテンポと検索性能の関係について考察する．テンポの遅い旋律であれば，歌唱者も歌いやすいため検索性能が高い，一方，旋律のテンポが速いと，歌唱者にとっては旋律の詳細にまで理解が及ばないため，検索性能が低いのではないかという考えに基づく．

4.1 節で説明した歌声データベースの歌唱された 50 種類の旋律のうち，テンポが遅いものから順に 10 種類の旋律とテンポが速いものから順に 10 種類の旋律を選んだ．テンポの遅い旋律の平均 bpm(beat per minute) は 72.7，テンポの速い旋律の平均 bpm は 169.3 であった．これらの旋律を 75 名の被験者が歌唱しているため，テンポの遅い旋律を歌唱した 750 サンプルの評価セットとテンポの速い旋律を歌唱した 750 サンプルの評価セットが作成される．これらの評価セットを利用した検索性能を図 10 に示す．予想し

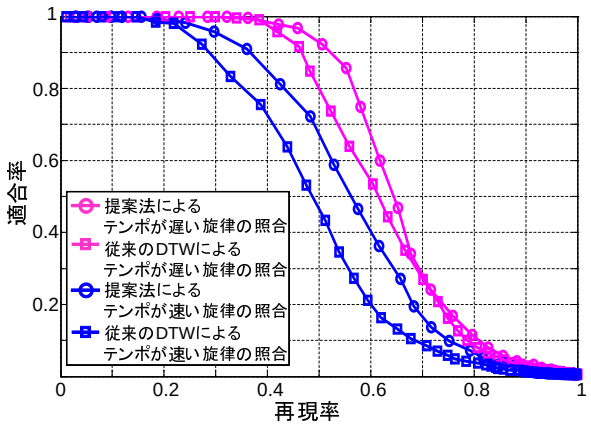

図 10　旋律のテンポの違いによる検索性能の比較：テンポの速い旋律の検索性能に比べて，テンポの遅い旋律の検索性能は高い．また，提案法によって，特にテンポの速い旋律の検索性能が改善された．

たとおり，テンポが遅い旋律の方が検索性能が高いことがわかる．また，提案法によって，テンポが遅い旋律よりも速い旋律の方が検索性能の改善が大きいことがわかる．例えば，再現率 0.6 のときに，適合率は，テンポが遅い旋律に対しては，従来法に比べて 0.12 ポイント改善されたが，テンポが速い旋律に対しては，0.2 ポイント改善された．これは前節で考察した，提案法による F0 軌跡のスムージングの効果によるものであると考えられる．すなわち，歌唱者はうろ覚えの状態であるためテンポの速い旋律の詳細を正しく歌唱できない．しかし，提案法による粗い照合方法によって検索性能が改善されたと考えられる．

6　まとめと今後の課題

歌声の旋律情報と動的変動を同時に特徴付けるために，F0 と ΔF0 で構成される相平面を利用した新しい

F0 軌跡の表現手法を提案した．この相平面に現れるアトラクタの位置は，旋律を構成する音符の音高に対応し，アトラクタの渦軌跡の形状は，歌唱者の意図する演奏表現や癖などによる動的変動を特徴付ける．

また，この相平面から歌声の旋律情報だけを確率的に表現する手法を提案した．これは事前に相平面を有限個の領域に分割し，入力された F0 軌跡に対して，各領域が占める特徴ベクトル (F0，ΔF0) の割合からアトラクタの位置を確率的に特定する手法である．この手法をハミング検索のための旋律の類似尺度に適用した結果，動的変動や部分的な音符の挿入や削除による揺れを含む歌声に対しても，効果的に検索結果を絞り込むことが可能であった．特に，歌唱者にとって難しい，テンポの速い旋律の歌声に対して有効な検索手法であることを確認した．今後の課題は，さらに楽曲データベースを拡大させたときの提案法の検索性能を検証すること，あらかじめ検索する旋律を切り出すことなく，楽曲全体の旋律のどの部分を歌唱したかを特定できる検索手法に発展させることである．

また，相平面における歌声の F0 軌跡の表現を，ハミング検索だけでなく，その他様々な分野に適用することを考えている．例えば，歌唱者の演奏表現や癖を特徴付ける F0 軌跡の “動き” をモデル化することによって，これまでの先行研究とは異なる視点に基づいた歌唱力評価や歌唱支援，歌唱者の識別などへの応用が考えられる．また楽譜が与えられたときに，歌唱者 A の歌い方，歌唱者 B の歌い方というように多様性のある歌声合成への応用も考えられる．したがって，相平面を利用して，歌唱者の演奏表現や癖の違いが大きく表れるアトラクタの渦軌跡の形状をさらに分析し，その動きをモデル化するための技術を検討することが今後の大きな課題である．

参考文献

[1] 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).

[2] 藤原弘将，北原鉄朗，後藤真孝，駒谷和範，尾形哲也，奥乃　博：伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法，情報処理学会論文誌，Vol. 47, No. 6, pp. 1831–1843 (2006).

[3] 中野倫靖，後藤真孝，平賀　譲：楽譜情報を用いない歌唱力自動評価手法，情報処理学会論文誌，Vol. 48, No. 1, pp. 227–236 (2007).

[4] Saitou, T., Unoki, M. and Akagi, M.: Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis, *Speech Communication*, Vol. 46, pp. 405–417 (2005).

[5] Dannenberg, R. B., Birmingham, W. P. et al.: A Comparative Evaluation of Search Techniques for Query-by-Humming Using the MUSART Testbed, *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 5, pp. 687–701 (2007).

[6] Song, J., Bae, S. Y. and Yoon, K.: Mid-Level Music Melody Representation of Polyphonic Audio for Query-by-Humming System, *Proc. ISMIR 2002* (2002).

[7] Pauws, S.: CubyHum: A fully operational query by humming system, *Proc. ISMIR 2002* (2002).

[8] Pardo, B., Shifrin, J. and Birmingham, W. P.: Name that tune: a pilot study in finding a melody from a sung query, *Journal of the American Society for Information Science and Technology*, Vol. 55, No. 4, pp. 283–300 (2004).

[9] Hu, N. and Dannenberg, R. B.: A Comparison of Melodic Database Retrieval Techniques Using Sung Queries, *Joint Conference on Digital Libraries*, pp. 301–307 (2002).

[10] Adams, N. H. et al.: Time Series Alignment for Music Information Retrieval, *Proc. ISMIR 2004* (2004).

[11] 橋口博樹，西村拓一，張　建新，滝田順子，岡　隆一：モデル依存傾斜制限型の連続 DP を用いた鼻歌入力による楽曲信号のスポッティング検索，電子情報通信学会論文誌 D-II，Vol. J84-D-II, No. 12, pp. 2479–2488 (2001).

[12] 酒向慎司，宮島千代美，徳田恵一，北村　正：隠れマルコフモデルに基づいた歌声合成システム，情報処理学会論文誌，Vol. 45, No. 3, pp. 719–727 (2004).

[13] de Cheveigne, A. and Kawahara, H.: YIN, a fundamental frequency estimator for speech and music, *JASA*, Vol. 111, No. 4, pp. 1917–1930 (2002).

[14] 杉山雅英：セグメントの高速探索法，情処研報音楽情報科学，　Vol. 1998, No. 29, pp. 87–93 (1998).

[15] 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).

[16] Goto, M.: AIST Annotation for the RWC Music Database, *ISMIR 2006* (2006).

[17] 後藤真孝，西村拓一：AIST ハミングデータベース: 歌声研究用音楽データベース, 情報処理学会 音楽情報科学研究会研究報告，Vol. 2005, No. 82, pp. 7–12 (2005).

[18] Kashino, K., Kurozumi, T. and Murase, T.: A Quick Search Method for Audio and Video Signals Based on Histogram Pruning, *IEEE Trans. Multimedia*, Vol. 5, No. 3, pp. 348–357 (2003).