社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－MUS－74 (2)
2008－SLP－70 (2)
2008／2／8

複数の音程特徴量によるハミング入力楽曲検索システムの高精度化

市川 拓人† 鈴木 基之† 伊藤 彰則† 牧野 正三†

† 東北大学大学院工学研究科
E-mail: †{tackt.213,moto,aito,makino}@makino.ecei.tohoku.ac.jp

あらまし 本稿では，基本周波数 (F0) の抽出を行わないハミング入力楽曲検索システムについて検討する．F0 の抽出は，どれほど高精度なものでも抽出誤りを避けることが完全にはできず，検索精度を低下させる原因となっている．また，F0 の抽出は適切に抽出されても，歌唱者の音高自体が誤っていることで，検索性能が低下するという問題も存在する．これらの問題に対し我々は以前，2 つの対数周波数領域パワースペクトルの相互相関関数を音程特徴量として提案し，F0 の代わりに音程特徴量，さらには音程特徴量の確率モデルを用いたハミング検索システムを構築した．検索実験の結果，提案手法を用いることで検索システムが高精度化することが確かめられた．本稿では前述の相互相関関数のピーク音程を音程特徴量として抽出し，検索システムの性能をさらに向上させることを検討する．また，以前に提案した音程特徴量，今回提案する音程特徴量それぞれを用いた時の検索結果を統合することで，それぞれの検索誤りを補正することを検討する．そして実際に検索実験により提案手法を導入した検索システムは，F0 を用いた時の検索精度を 13.2 %上回る結果となった．

Improvement of a Query-by-Humming Music Information Retrieval System using Multiple Musical Interval Features

Takuto ICHIKAWA†, Motoyuki SUZUKI†, Akinori ITO†, and Shozo MAKINO†

† Graduate School of Engineering, Tohoku University
E-mail: †{tackt.213,moto,aito,makino}@makino.ecei.tohoku.ac.jp

Abstract This paper describes a query-by-humming (QbH) music information retrieval (MIR) system without F0 extraction. In F0 extraction based system, F0 extraction errors inevitably occur that degrades performance of the system. Furthermore, errors in pitch of sung data degrade performance of the system, too. To improve these problems, we have propose an MIR system that used a musical interval feature and probabilistic models. The performance of the proposed system exceeded the system based F0 extraction. In this paper, we use peak interval of the cross-correlation function as a tonal feature to improve performance of the system. In addition, we integrated multiple retrieval result to obtain better recognition result. From an experimented result, the top retrieval accuracy given by the proposed method have exceeded the system based F0 extraction by 13.2 %.

1. はじめに

ハミングを入力とした楽曲システムはこれまでいくつか提案されている [1]〜[3]．これらのシステムは，入力されたハミングを音符単位に分割し，それぞれの音符区間で音高・音長を抽出する．さらに連続する音符区間同士で音高・音長を相対化することで音程・音長比を求め，これらをデータベースの楽曲と照合するという流れが一般的である．

しかし，この検索の流れには問題がある．それは音高の誤りである．音高の誤りはシステムと歌唱者の双方で引き起こされる．音高には対数基本周波数がよく用いられるが，システムはしばしば基本周波数 (F0) の抽出誤りを引き起こし，

また歌唱者は記憶の不正確さや歌唱の不安定さにより歌唱音程の誤りを引き起こす．これは楽曲検索の性能低下に直結する重大な問題である．

この問題に対し，Heo らが F0 の複数候補を用いた検索手法を [4]，Shih らが F0 から計算される音程の確率モデルの構築 [5] を提案している．

Heo らのシステムでは，FFT ケプストラム分析法により F0 の抽出がなされている．FFT ケプストラム分析法では，入力のケプストラムに対して，F0 の存在範囲に対応する探索区間でピークとなるケフレンシーを求め，ここからさらに F0 へと変換する．従来であれば，ピークを 1 つだけ求め，その結果から F0 を求めるが，このシステムではケプストラムのピークを大きい順に複数個求め，それぞれに対応する周波数を F0 候補として抽出している．そして，これを全ての音符区間で行い，ピッチ候補の全ての組み合わせについて入力のメロディを構成しデータベースと照合する．Heo らの実験によって，ピッチを 3 候補抽出した場合，その候補中に正解の F0 が含まれている精度は 99.7 %であり，またこれを用いて楽曲検索を行ったところ，検索精度は 86.5 %となったことが報告されている．この精度は従来システムに比べてよいものであるが，全ての組み合わせについて検索を行うことから，検索時間は従来法に比べて増加するという問題がある．3 候補によって検索を行う場合，その計算量は従来法の約 9 倍になってしまう．また，99.7 %という F0 の抽出精度はあくまで実際に歌唱された音声の高さに対する抽出精度であり，歌唱すべき音高に対する抽出精度ではない．そして，抽出される F0 候補は歌唱された音声の音高とその倍音・半音がほとんどである．つまり歌唱の誤りに頑健とはいえない．

Shih らは抽出された F0 から計算される音程の確率モデルを構築している．Shih らは大量の歌唱データを収集し，それぞれの歌唱データの全区間に対し F0 の抽出を行っている．そして隣り合う区間で音程を計算し，これを学習サンプルに用いて楽曲中に出現が想定される全ての音程について，統計的なモデル化を行っている．このモデル化により，歌唱誤り時に表れる音程のわずかな揺らぎに頑健になっている．しかし，F0 の倍音・半音抽出誤りまではこのモデルでは考慮していない．つまりこの手法はシステムによる F0 の抽出誤りに対しては頑健とはいえない．

以上のように，F0 の抽出を行う手法は常に何らかの問題を抱えている．しかし，ハミング検索に用いられるのは音程であって，F0 の値そのものではない．連続する 2 区間の音程を直接的に表す特徴量 (以下，これを音程特徴量と呼ぶ) を抽出することができれば，F0 を抽出する必要はない．つまり，音程特徴量を用いることでシステムによる抽出誤りの問題を避けることができると考えられる．抽出誤りを避けることができれば残りの課題はユーザの歌唱誤りのみとなる．そしてこの課題は Shih らの手法のような確率モデルを用いることで対処することができる．すなわち，音程特徴量とその確率モデルを用いることで，従来システムの持つ音高誤りの問題に全て対処できると考えられる．また，確率モデルを用いて検索を行うことにはさらに利点がある．各データベースから得られる音程系列どおりに確率モデルを並べ，それに対する尤度を計算する手法を用いることで，音程を決定的に扱うのではなく，確率的に全ての可能性を評価することができるため，より頑健な検索が実現できるということである．

この考えのもと，我々は音程特徴量として「連続する 2 区間それぞれの対数周波数領域パワースペクトルから計算される相互相関関数」を用い，さらにこの音程特徴量の確率モデルを用いたハミング検索システムを構築した [6]．システムの性能は従来の F0 抽出による検索性能を上回ることが実験により確かめられている．しかし，複数 F0 候補を用いた検索性能には及んでおらず，提案システムには何らかの改良が必要であると考えられる．

そこで本稿では，これまでに提案した音程特徴量をさらに改良し，検索に用いることで検索性能の向上を目指している．具体的には，相互相関関数のピーク音程を音程特徴量として抽出している．

また本稿ではさらに，これまでに提案した音程特徴量，及び今回提案する音程特徴量それぞれを用いた時の検索結果を統合することでさらなる性能向上を目指している．

2. 音程特徴量と音程モデルを用いた楽曲検索

本章では，以前提案した音程特徴量と音程モデルを用いたハミング検索システムについて述べる．

2.1 システム概略

本システムは

（1） F0 に代わる，音程特徴量の抽出

（2） 各音程毎に音程特徴量を確率分布でモデル化

（3） 各データベースの曲毎に，確率モデルを並べ，音程系列らしさを計算

という 3 つの特色をもつ．

検索に先立ち，まず音程の確率モデル (以下，音程モデルと呼ぶ) を作成する．楽曲中に出現するであろう全ての音程について，ハミングデータを収録し，このデータから音程特徴量を抽出する．得られた大量の音程特徴量を音程毎に分類し，それぞれの音程について統計的なモデルを作成する．

楽曲検索は次のようにして行われる．入力ハミングに対して，音符に相当する区間を検出する．隣り合う区間同士から音程特徴量を抽出し，これらの特徴量群を入力系列として並べる．一方，データベースには楽曲のメロディが MIDI の系列として格納されている．ここからデータベース中の楽曲の音程系列を得ることができる．得られた音程系列は事前に準備された音程モデルの系列として並べられる．入力ハミング

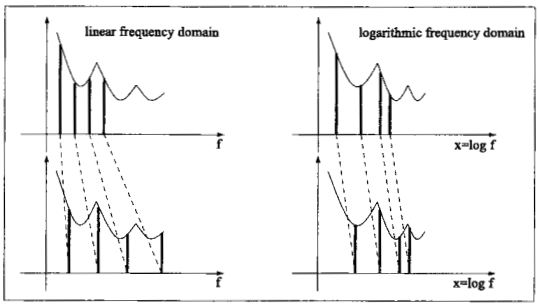

図 1　線形周波数領域と対数周波数領域でのパワースペクトルの相対的位置関係

から抽出された音程特徴量系列を楽曲の音程モデル系列で評価することにより，データベース中のある楽曲に対する入力ハミングの尤度を計算することができる．音程特徴量と音程モデルの照合に連続 DP マッチングを用いることで，入力の挿入・脱落に対処でき，さらに曲のどこから歌い始めても検索が可能になる．

2.2　音程特徴量の抽出

音程特徴量の抽出は，対数周波数領域パワースペクトルに表れる「調和拘束」[7] の性質を利用して行う．

線形周波数領域では，基本周波数が Δw 変化すれば第 n 高調波成分は $n\Delta w$ 変化する．一方対数周波数領域では，第 n 高調波成分は基本周波数から常に $\log n$ 離れたところに存在する．そのため基本周波数が変化しても基本波と高調波の相対的位置関係は一定に保たれ，基本周波数の変化は対数周波数スペクトルの平行移動として表現される (図 1)．

ある入力信号 $x(t)$ が観測され，この信号のパワースペクトルを $X(\omega)$ とすると，$\xi = \log_2(\omega)$ として対数周波数領域パワースペクトルは $X(\xi)$ となる．この時，$x(t)$ よりも α オクターブ高い基本周波数を持つ信号 $y(t)$ が観測されたとすると，$y(t)$ のパワースペクトルは $Y(\omega) \approx X(2^{-\alpha}\omega)$ であり，対数周波数領域では $Y(\xi) \approx X(\xi - \alpha)$ となる．つまり，この 2 つの信号 $x(t)$ と $y(t)$ から対数周波数パワースペクトルの相互相関関数を求めると

$$C_{XY}(l) = \sum_{\xi} X(\xi)Y(\xi+l) \approx \sum_{\xi} X(\xi)X(\xi-\alpha+l) \quad (1)$$

となり，$l \approx \alpha$ にピークを持つ関数が得られる．音程特徴量はこの相互相関関数から計算される．

以前の報告では，相互相関関数 $\boldsymbol{C}_{XY}$ について

$$\boldsymbol{C}_{XY} = (C_{XY}(0), \ldots, C_{XY}(D)) \qquad (2)$$

とし ($D = 69$，$\pm$ 1400cent 以内のシフトまで考慮)，これを主成分分析することにより，K 次元ベクトル $\boldsymbol{z} = (z_1, ..., z_K)$ へと変換した結果を音程特徴量とした (以下，これを音程特徴ベクトルと呼ぶ)．

2.3　音程モデル

音程モデルは，各音程を 1 つのクラスとして，それぞれの音程毎に音程特徴量を収集して作成される．例えば，+200cent の音程クラスは「F3 → G3」「G3 → A3」など，-400cent の音程クラスは「E3 → C3」「A3 → F3」などの音程特徴量から学習する．この学習により，各音程クラスの確率密度関数 $p(\boldsymbol{z}|T_i)$ が得られる．ここで，確率密度関数 $p(\boldsymbol{z}|T_i)$ は音程クラス T_i に属する特徴量 $\boldsymbol{z}$ の生起確率を表している．確率密度関数には，単一正規分布・2 混合正規分布・ラプラス分布の 3 種類を用いる．なお，これらはそれぞれ次式で表される．

まず単一正規分布は，

$$p(\boldsymbol{z}|T_i) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} \exp\left\{-\frac{(z_k - \mu_{ik})^2}{2\sigma_{ik}^2}\right\} \qquad (3)$$

である．ここで，μ_{ik}，σ_{ik}^2 は音程クラス T_i の学習サンプルの音程特徴量の k 次元目の平均と分散を表している．

混合正規分布は，

$$p(\boldsymbol{z}|M_i) = \sum_{m}^{M} \alpha_{im} \prod_{k}^{K} \frac{1}{\sqrt{2\pi\sigma_{ikm}^2}} \exp\left\{-\frac{(z_k - \mu_{ikm})^2}{2\sigma_{ikm}^2}\right\} \qquad (4)$$

である．α_{im} は音程クラス T_i における第 m 混合要素の混合重みで，μ_{ikm}，σ_{ikm} は音程クラス T_i における，第 m 混合要素の k 次元目の平均と分散である．

最後にラプラス分布は，

$$p(\boldsymbol{z}|T_i) = \prod_{k=1}^{K} \frac{1}{2b_{ik}} \exp\left\{-\frac{|z_k - \hat{\mu}_{ik}|}{2b_{ik}}\right\} \qquad (5)$$

$$b_{ik} = \frac{1}{N} \sum_{n_i=1}^{N} |x_{n_i k} - \hat{\mu}_{n_i k}| \qquad (6)$$

である．N は学習サンプル数，x_{n_i} が学習サンプルの音程特徴量である．また，$\hat{\mu}_{ik}$，b_{ik} は音程クラス T_i に属する音程特徴ベクトルの k 次元目の学習サンプルの中央値，尺度パラメータである．

なお，単一正規分布・ラプラス分布のパラメータは学習サンプルから得られる，理論的な最尤推定量を用い，混合正規分布のパラメータは EM アルゴリズムで推定される．

2.4　検索手法

楽曲検索時，入力は音符区間に区切られ，それぞれの区間から得られた特徴量が楽曲の特徴量と比較される．この区間検出は帯域通過フィルタと差分フィルタを用いて行われる [4]．入力ハミングは/ta/で歌唱されるため，この/a/のフォルマントが大きい 600~1,500Hz の帯域通過フィルタをかけることで，/a/の部分を強調し，それ以外の部分を誤って検出することを防ぐ．さらに差分フィルタを用いて，パワーの時間的変動のエッジを抽出する．単純にパワーにより

検出を行っていないのは，パワーでは歌唱者による違いが大きく，閾値による検出が困難なためである．

入力から得られた特徴量系列とデータベース中の楽曲から得られる特徴量系列は連続 DP マッチングを用いて比較される．本稿に示すシステムでは，入力の第 i 区間と参照側の第 j 区間の格子点で，確率密度関数 $p(\boldsymbol{z}(i)|T(j))$ から得られる対数事後確率 $\log P(T(j)|\boldsymbol{z}(i))$ と音長比距離 $t(i,j)$ を統合し，これをスコアとしてマッチングが行われる．事後確率 $P(T(j)|\boldsymbol{z}(i))$ は，音程 $T(j)$ の生起確率 $P(T(j))$ を一様分布とみなし計算する．すなわち，

$$p(\boldsymbol{z}(i)) = \sum_{c=1}^{C} p(\boldsymbol{z}(i)|T_c) \tag{7}$$

$$P(T(j)|\boldsymbol{z}(i)) = \frac{p(\boldsymbol{z}(i)|T(j))}{p(\boldsymbol{z}(i))} \tag{8}$$

ここで C は，音程モデルを構成する音程クラスの数である．さらに，入力の i 番目の音符区間の音長 $L_h(i)$，楽曲の j 番目の音符区間の音長 $L_m(j)$ に対し，

$$\Delta L_h(i) = \log \frac{L_h(i+1)}{L_h(i)} \tag{9}$$

$$\Delta L_m(j) = \log \frac{L_m(j+1)}{L_m(j)} \tag{10}$$

$$t(i,j) = |\Delta L_h(i) - \Delta L_m(j)| \tag{11}$$

と定義し，スコアリング重み $weight$ を用いて

$$s(i,j) = weight \times \log P(T(j)|\boldsymbol{z}(i)) - (1-weight) \times t(i,j) \tag{12}$$

と計算される $s(i,j)$ を用いて連続 DP マッチングを行い検索する．ただし厳密には，マッチング時に挿入・脱落の判定がなされた時には特徴量を再計算する手法を用いている [4], [6] 従来の DP マッチングでは，入力側・参照側ともに各区間の特徴量は DP パスに依らず一定である．しかし，本手法のように，相対的な値を特徴量に用いている場合は，このような計算方法はふさわしくなく，DP パスによっては特徴量を再計算する必要がある．例えば，ハミングのある音符に着目してスコアを計算している時，前の音符が余計に挿入されたという判定がなされたとする．この時，音程を計算する相手は，…一つ前の音符と計算しては挿入の判定に反することになる．そこでこのケースでは音程は二つ前の音符と計算する．このように挿入・脱落時には特徴量を再計算して検索を行う．また挿入・脱落時にはスコアにペナルティを付与する [8].

3. 相互相関関数のピーク音程の抽出及び検索結果の統合

これまでは音程特徴量として 2.2 節で述べた音程特徴ベクトルを用いてきた．今回はこれに加えて，2.2 節で示した相互相関関数がピークとなる音程を検出し，これを音程特徴量に用いることを検討する．なお，以降この相互相関関数のピーク音程を，単に「ピーク音程」と表す．そしてこれまでと同様に，「ピーク音程」「音程特徴ベクトル」それぞれの音程特徴量を用いて，「単一正規分布」「混合正規分布」「ラプラス分布」で音程モデルを構築し検索に用いる．

ここで，音程特徴量と音程モデルそれぞれの組み合わせを 1 つの手法と見立てると，ある手法では正解するが，別の手法では検索誤りを起こすというケースは十分に考えられる．これは，例えばピーク音程と音程特徴ベクトルの持つ情報の違いによるもの，音程モデルの歌唱者への適応の度合など様々な原因で発生すると考えられる．そこで，このような誤りを補正するため，それぞれの手法での検索結果を統合することを検討する．統合の手法には，最も単純な

- 各検索手法による 1 位検索結果を多数決により統合する方法
- 各検索手法による検索順位の和をスコアとしてリスコアリングする方法

の 2 手法を検討する．

表 1 実験条件

音程モデル作成	学習データ	男性 10 名の歌唱データ 音程特徴量 225,000 パターン
	音程候補	25 候補 -1200 ～ +1200cent 100cent 刻み
音響分析	サンプリングレート	16kHz
	窓幅	64ms (ハニング窓)
	分析周期	8ms
	区間検出法	BPF: 600~1500 Hz DF: 一次差分
特徴量	音程	ピーク音程 音程特徴ベクトル
	音長	IOI 比
評価データ	音程推定実験	男性 5 名の歌唱データ 音程特徴量 18,000 パターン
	楽曲検索実験	男性 5 名の歌唱データ 326 曲
楽曲データベース		童謡 156 曲

BPF:帯域通過フィルタ, DF:差分フィルタ

4. 評価実験

今回，音程推定実験と楽曲検索実験，検索結果の統合実験の 3 つの実験を行い，本手法の精度を検証した．音程推定実験は，特徴量と音程モデルの評価のために行っている．

4.1 実験条件

表 1 に今回行った 3 つの実験の条件を示す．

学習データおよび音程推定実験のテストセットはヘッドセットマイクを通じて DAT に収録した．歌唱者には，ボイストレーニングで用いられる 5 音の発声法 (例 : C → E →

表 2　音程推定実験結果

特徴量	確率分布	1 位	2 位以内	3 位以内
ピーク音程	単一正規分布	63.3 %	91.6 %	98.6 %
ピーク音程	ラプラス分布	64.2 %	91.7 %	98.8 %
音程特徴ベクトル (8 次元)	単一正規分布	63.0 %	91.4 %	98.9 %
音程特徴ベクトル (12 次元)	2 混合正規分布	64.1 %	91.5 %	98.9 %
音程特徴ベクトル (8 次元)	ラプラス分布	63.1 %	90.8 %	98.4 %
F0	(距離)	68.38 %	-	-

表 3　楽曲検索精度

特徴量	確率分布	1 位検索率	10 位以内	検索時間
ピーク音程	ラプラス分布	86.8 %	93.6 %	16.7 sec
音程特徴ベクトル (8 次元)	単一正規分布	81.9 %	93.6 %	17.5 sec
音程特徴ベクトル (12 次元)	2 混合正規分布	83.7 %	92.0 %	20.8 sec
音程特徴ベクトル (8 次元)	ラプラス分布	83.7 %	93.6 %	17.7 sec
単一 F0	(距離)	74.2 %	89.0 %	13.0 sec
複数 F0	(距離)	86.5 %	94.1 %	116.7 sec

表 4　統合実験結果

統合手法	1 位検索率	10 位以内
多数決	87.4 %	-
ランキング	84.0 %	94.5 %
ピーク	86.8 %	93.6 %
単一 F0	74.2 %	89.0 %
複数 F0	86.5 %	94.1 %

G → E → C) を検索と同様に/ta/で歌唱してもらい，第一音を半音毎あげて 1 オクターブ中の音が全て出現するまでの計 6 種類の発声を繰り返した．発声前には，ピアノ音を基準音として提示し，この基準音を聞きながら歌唱する声の高さを覚えるまで数回練習してもらった．収録時には基準音を提示せず，歌唱者の記憶だけを頼りに歌唱してもらった．これにより，疑似的に知っている曲を歌う環境を作りだした．なお，一種類につき 5 回ずつ発声してもらっている．

続いて，収集された歌唱データから，手作業で音声区間を検出し，/a/にあたる部分を切り出した．そして切り出されたデータ全ての組み合わせで音声を繋げ，学習データ・テストセットの特徴量を計算した．この特徴量を用いて実験を行っている．

また，本手法の性能比較のため，F0 抽出による同様の実験も行った．F0 抽出は FFT ケプストラム分析法により行った．音響分析は 64ms のハミング窓により行い，各フレームの F0 の中央値を F0 として採用した．

4.2　音程推定実験

特徴量と音程モデルの評価のため，入力の音程推定実験を行った．この実験は，1 つの音符に対するハミングを 2 つ入力し，その 2 音の音程を推定するものである．評価としては，「最尤の音程が正解しているか」に加え，「音程尤度上位 2 位以内に正解が存在するか」「音程尤度上位 3 位以内に正解が存在するか」の点から行った．これは，最尤の音程が正解でなくても上位に正解が含まれていれば，正解の音程に対し高い尤度を得ることから，検索に効果的であることが示されるためである．実際に結果は表 2 のようになった．音程推定実験の結果としては，

- いずれの手法でも最尤音程の正解率は F0 による音程計算の正解率に及ばなかった
- 3 位以内にほぼ正解が含まれており，検索への有効性が望める
- ピーク音程も音程特徴ベクトルと同等の性能を示している
- 確率分布の違いによる影響はほとんどない

ということがいえる．

4.3　楽曲検索実験

続いて，提案手法による楽曲検索実験を行った．実験結果を表 3 に示す．結果としては，

- ピーク音程を特徴量に，ラプラス分布でモデル化したときに最も良い検索率が得られる
- 検索時間は単一ピッチによる検索からわずかに増加したのみである

ということが挙げられる．

4.4　統 合 実 験

最後に統合実験結果について述べる．統合実験結果を表 4 に示す．

多数決の結果は「ピーク音程をラプラス分布でモデル化した手法」「音程特徴ベクトルをラプラス分布でモデル化した手法」「音程特徴ベクトルを 2 混合正規分布でモデル化した手法」での検索結果を統合したときのものであり，順位の和による結果は「ピーク音程をラプラス分布でモデル化した手法」「音程特徴ベクトルをラプラス分布でモデル化した手法」での検索結果を統合した時のものである．結果としては，

- 多数決による統合が，順位の和による統合の性能を上回った
- 多数決による統合で，統合前よりも 0.6 %精度が向上した

ということが挙げられる．ただし，ピーク音程による 1 位検索率と多数決による検索率について，有意水準 5%で母比率の検定を行ったところ，その差には有意差は見られなかった．

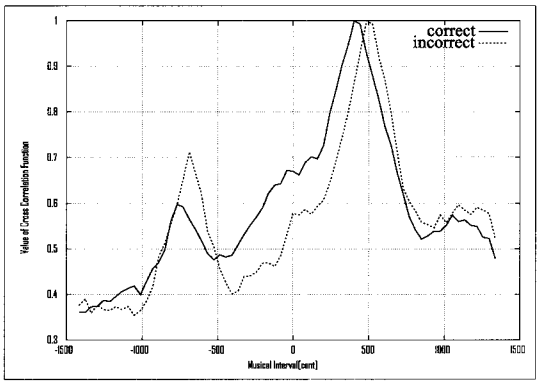

図 2　相互相関関数の違い

4.5 考　　察

4.5.1 提案手法による検索率向上

音程推定実験結果は，F0 から計算される音程の方が正解率が高いという結果であったが，検索実験の結果は提案手法の方がよい結果となった．この原因について考察する．

まず，F0 の抽出について着目する．音程推定実験に用いたテストセットに対し，F0 から計算される音程と正解音程の差のヒストグラムを調べたところ，倍音・半音違いを生じていないことが判明した．それに対し，楽曲検索実験に用いた歌唱データに対しては F0 抽出時に倍音・半音誤りが発生している．ピーク音程の抽出ではどちらも倍音・半音誤りを引き起こしておらず，この違いから検索では提案手法の方がよい性能となったと考えられる．また，F0 の抽出が適切になされていても，歌唱が不正確であったために検索誤りを引き起こしている場合がある．このようなケースに対し，音程モデルを導入することで歌唱誤りに頑健になり，かつ特徴量を確率的に評価したことで検索誤りを修正できている．

4.5.2 音程特徴量の違いによる性能の差

まず，ピーク音程と音程特徴ベクトルを用いた時の性能の違いについて考察する．図 2 に音程推定実験で正しく音程が推定されるデータの相互相関関数と，推定を誤るデータの相互相関関数を示す．なお，2 つの相互相関関数は，共に 400cent を歌唱した時の相互相関関数である．図 2 を見る限り，これらの違いはピークの音程にあり，それ以外の部分にはほとんど違いが見られない．この他にも音程推定誤りを起こしたときの相互相関関数を定性的に評価してみたが，ほとんどがピークの音程がずれているために推定誤りを起こしていた．このことから，相互相関関数のピーク音程以外の部分は余計な情報がほとんどである可能性が高いといえる．ピーク音程を用いることで，この雑音成分がほぼ除去され，その結果実験的には良い性能を示したと考えられる．

4.5.3 統合手法による性能の違い

今回の検証では，多数決により検索結果を統合することで検索性能が向上した一方，順位の和による統合では検索性能が低下した．この原因は，統合する前の検索順位によるものだと考えられる．多数決では 2 位以下の場合は何位でも影響はないが，順位の和では影響がある．上位の検索結果に対し，下位の検索結果を統合してしまうことが悪影響となり，順位の和では統合は逆効果となってしまった可能性が高い．

5. 結論と今後の課題

本稿では，複数の音程特徴量を提案し，さらにこの確率モデルを構築した．これらをハミング検索システムに導入することで検索性能は従来の単一の F0 を用いた検索精度を 12.6%上回った．また，複数の F0 候補を用いた検索精度を 0.3%上回った．一方で検索時間の増加を抑えることができており，本手法が有効であることが立証された．

さらに，複数の検索手法による検索結果を多数決により統合したところ，検索率は 0.6%向上した．

今後の課題としては，統合手法の検討，異なるデータベースでの提案手法の有効性の検証が挙げられる．

文　　献

[1] 園田 智也，後藤 真孝，村岡 洋一，“WWW 上での歌声による曲検索システム”，電子情報通信学会論文誌，Vol.J82-D-II, No.4, 1999

[2] Steffen Pauws, “CubyHum: A Fully Operational Query by Humming System”, *Proc.3rd International Conference on Music Information Retrieval(ISMIR2002)*, pp.187-196, 2002.

[3] Jyh-Shing Roger Jang, Hong-Ru Lee, Jian-Chun Chen, “Super MBox:an efficient/effective content-based music retrieval system”, *Ninth ACM Multimedia Conf. (Demo Paper)*, pp.636-637, 2001

[4] Sung-Phil Heo, Motoyuki Suzuki, Akinori Ito, Shozo Makino, “An Effective Music Information Retrieval Method Using Three-Dimensional Continuous DP”, *IEEE Trans. on Multimedia*, vol.8, No.3, pp.633-639, 2006.

[5] Hsuan-Huei Shih, Shrikanth S. Narayanan, C.-C.Jay Kuo, “A Statistical Multidimensional Humming Transcription using Phone Level Hidden Markov Models for Query by Humming Systems”, *Proc. the International Conference on Multimedia and Expo*, vol.2, pp61-64, 2003

[6] 市川拓人，鈴木基之，伊藤彰則，牧野正三，“音程特徴量の確率分布を考慮したハミング入力楽曲検索システム”，情報処理学会研究報告，2007-MUS-71, Vol.2007, No.81, pp.33-38 (2007)

[7] Shigeki Sagayama Keigo Takahashi, Hirokazu Kameoka, Takuya Nishimoto, “Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signals by Deconvolution of Log-Frequency Spectrum”, *Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, 2004.

[8] Akinori Ito, Sung-Phil Heo, Motoyuki Suzuki, Shozo Makino, “Comparison of Features For DP-Matching Based Query-by-Humming System” *Proc.5th International Conference on Music Information Retrieval (ISMIR 2004)*, pp297-302, 2004