2007－MUS－71 (6)
2007／8／1

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

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 〒980-9579 宮城県仙台市青葉区荒巻字青葉 6-6-05
E-mail: †{tackt.213,moto,aito,makino}@makino.ecei.tohoku.ac.jp

あらまし 本稿では，ピッチ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．ピッチ抽出は，どれほど高精度なものでもピッチ抽出誤りを避けることができず，検索精度を低下させる原因となっている．本システムでは，従来用いられているデルタピッチの代わりに，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用い，さらに楽曲中に存在しているであろう全音程の確率モデルを用意しておく．連続する2つの音符が観測された時，この特徴量と確率モデルを用いて，全音程についての尤度を計算する．このシステムの利点は，統計的なモデル化を行うことにより，ピッチ抽出誤りのような致命的な誤りを起こしにくいということである．そして実際に検索実験を行ったところ，ピッチによる検索精度を最大 4.9 %上回る結果となった．

キーワード ハミング入力楽曲検索システム，音程モデル，多次元確率分布

Query-by-Humming Music Information Retrieval System using Probabilistic Distribution for Tone Interval Features

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

† Graduate School of Engineering, Tohoku University Aza-Aoba 6-6-05, Aramaki, Aoba-ku, Sendai-shi, Miyagi, 980–9579 Japan
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 pitch extraction. In pitch extraction based system, pitch extraction errors inevitably occur that degrades performance of the system. In this system, a cross-correlation function between two logarithmic frequency spectra is extracted as a tonal feature instead of deltaPitch, and probabilistic models are prepared for all tone intervals assumed to exist in the music. When two signals corresponding to two contiguous notes are given, likelihoods are calculated for all possibility of tone intervals. The advantage of this system is that it is hard to occur a fatal error such as a pitch extraction error because extracted features are modeled stochastically. From a experimented result, the top retrieval accuracy given by the proposed method have exceeded the system based pitch extraction by 4.9 %.

Key words Query-by-Humming, Tone Interval models, multi-dimensional probabilistic distribution

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]．これらのシステムでは，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のピッチや音価を抽出する．さらに連続する音符区間同士でピッチや音価の比を計算し，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，一般にピッチ抽出の精度はあまり高くない．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．この問題を解決する方法の一つが，Heo らによるピッチの複数候補を用いた検索法である [4]．このシステムでは，FFT ケプス

トラム分析法により，ピッチ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，基本周波数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，ここからさらに基本周波数へと変換する．従来であれば，ピークを 1 つだけ求め，その結果から基本周波数を求めるが，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ピッチ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 らの実験では，ピッチを 3 候補抽出した場合，その候補中に正解のピッチ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莫大なものとなり，3 候補によって検索を行う場合，従来法の約 9 倍になってしまうという問題がある．

そこで，本稿ではピッチ抽出を行わない楽曲検索システムを提案する．このシステムではピッチに代わる音程特徴量として，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相互相関関数を求める．また，楽曲中に出現が想定される全ての音程について，大量のサンプルを用いてその特徴量を統計的にモデル化しておく．未知の入力に対して，各データベースから得られる音程系列どおりにモデルを並べ，それに対する尤度を計算することで検索を行う．このようにすることで，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可能になり，より頑健な検索が実現される．

入力の音程を確率的に扱う研究として，Shih らはピッチ抽出の結果に確率を付与する手法を提案している [5]．これに対し，本システムではピッチ抽出そのものをシステムから除去することで，ピッチ抽出に伴う誤りを回避し，検索精度改善を目指す．

2. 音程の確率モデルを用いた楽曲検索

2.1 システム概略

本システムは

（1） デルタピッチに代わる，音程特徴量の抽出
（2） 各音程毎に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．

音程を確率的に扱う手法を用いた検索システムの概要を図 1 に示す．検索に先立ち，まず音程の確率モデルを作成する．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，このデータから特徴ベクトルを抽出する．得られた大量の特徴ベクトルを音程毎に分類し，それぞれの音程について統計的なモデルを作成する．

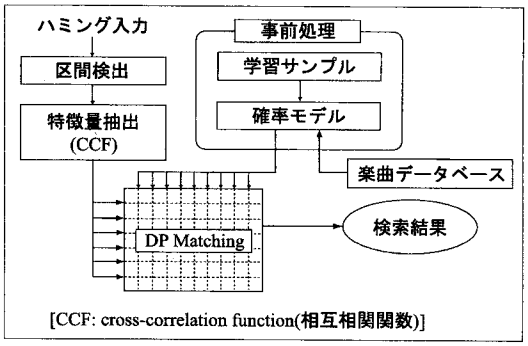

図 1 システム概略図

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を表す特徴量を抽出し，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確率モデルの系列として並べられる．入力ハミングから抽出された特徴ベクトル系列を楽曲の確率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．特徴ベクトルと確率モデルの照合に，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

本節では，特徴ベクトルの生成方法について述べる．今回，我々は対数周波数領域の周波数スペクトルに着目した．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 2) [6]．

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

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

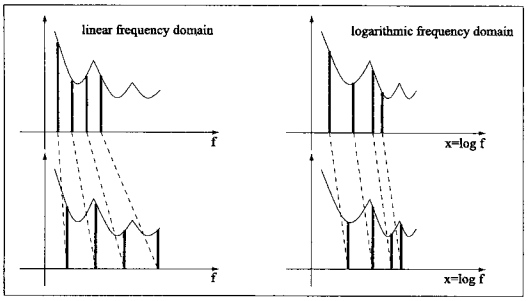

図 2　線形周波数領域と対数周波数領域でのパワースペクトルの相対的位置関係

となり，$l \approx \alpha$ にピークを持つ関数が得られる．なお，音程は対数周波数の差に比例するので，このピークから音程を一意に定めることもできるが，ピークが $l = \alpha$ に対応しないことも考えられるので，本手法ではピークの代わりに $C_{XY}(l)$ 全体を一つの特徴ベクトルとする．すなわち，

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

である．今回，$N = 69$ であり，± 1400cent 以内のシフト量についての相互相関関数を計算した．

さらに，この $\boldsymbol{C}_{XY}$ 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，これを音程特徴量として用いる．これは，$C_{XY}(l)$ をそのまま用いたのでは，「検索に時間が掛かってしまう」，「特徴ベクトルに冗長な情報が含まれており，検索精度が低下する」という 2 つの問題を解決するためである．

2.3　音程の確率モデル

音程の確率モデル (以下，音程モデルと呼ぶ) は，各音程毎に特徴ベクトルを収集して作成される．例えば，+200cent の音程モデルは「F3 → G3」「G3 → A3」などから，-400cent の音程モデルは「E3 → C3」「A3 → F3」などから形成される．

これらの特徴ベクトルから，確率密度関数を学習する．今回は，(3)(4) 式に示す 2 種類の確率密度関数について検討した．(3) 式は混合正規分布，(4) 式はラプラス分布に基づく．

$$p(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{P}) = \prod_k \sum_m \frac{P_{km}}{\sqrt{\pi B_{km}}} \exp\left(\frac{-(z_k - \mu_{km})^2}{B_{km}}\right) \qquad (3)$$

$$p(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{P}) = \prod_k \frac{1}{2B_k} \exp\left(\frac{-|z_k - \mu_k|}{B_k}\right) \qquad (4)$$

ここで，(3) 式においては $\boldsymbol{P} = (\boldsymbol{P}_1, ..., \boldsymbol{P}_K)$ $(\boldsymbol{P}_k = (P_{k1}, ..., P_{kM}))$ は混合重み，$\boldsymbol{\mu} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K)$ $(\boldsymbol{\mu}_k = (\mu_{k1}, ..., \mu_{kM}))$ は平均ベクトル，$\boldsymbol{B} = (\boldsymbol{B}_1, ..., \boldsymbol{B}_K)$ $(\boldsymbol{B}_k = (B_{k1}, ..., B_{kM}))$ はモデルの分布形状を表すパラメータで，各次元・各要素の標準偏差 $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, ..., \boldsymbol{\sigma}_K)$ $(\boldsymbol{\sigma}_k = (\sigma_{k1}, ..., \sigma_{kM}))$ に対し，$\boldsymbol{B}_k = (2\sigma_{k1}^2, ..., 2\sigma_{kM}^2)$ となる．これらのパラメータは EM アルゴリズムにより最尤推定される．一方 (4) 式では，$\boldsymbol{\mu} = (\mu_1, ..., \mu_K)$ が平均ベクトル，$\boldsymbol{B} = (B_1, ..., B_K)$ はモデルの分布形状を表すパラメータで，$\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_K)$ に対し，$\boldsymbol{B} = (\sigma_1/\sqrt{2}, ..., \sigma_K/\sqrt{2})$ となる．(4) のパラメータは学習データの平均や標準偏差から直接的に求められる．

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とデータベース中の楽曲から得られる特徴量とは連続 DP マッチングを用いて比較される．楽曲から得られた音程系列に対応する音程モデルが DP 平面の参照側に並べられる．この音程モデルを用いて，入力の音程系列らしさが (3) 式，(4) 式により計算される．そして音程系列らしさが最も大きい楽曲が，検索結果として返される．

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 パスに依っては特徴量を再計算する必要がある (図 3，4 参照)．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．音程は二つ前の音符と計算すべきである [4]．

この問題を定式化すると，次のようになる．まず，入力特徴ベクトルは次の $\boldsymbol{z}_{i1}$，$\boldsymbol{z}_{i2}$ の 2 通りを考える．

$$\boldsymbol{C}_{i1}(l) = \frac{1}{N-l} \sum_{\xi=1}^{N} S_i(\xi) S_{i+1}(\xi + l) \qquad (5)$$

$$\boldsymbol{C}_{i2}(l) = \frac{1}{N-l} \sum_{\xi=1}^{N} S_{i-1}(\xi) S_{i+1}(\xi + l) \qquad (6)$$

ここで，$\boldsymbol{C}_{i1}(l)$，$\boldsymbol{C}_{i2}(l)$ は入力の i 番目の区間の相互相関関数の l 次元目であることを表している．S_i は入力の i 番目の区間の対数周波数領域パワースペクトルである．さらに，

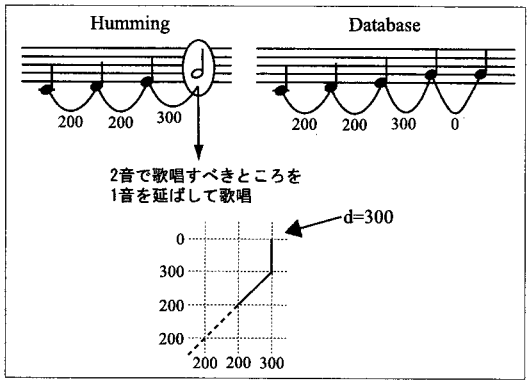

図 3　従来の特徴量とマッチング方法

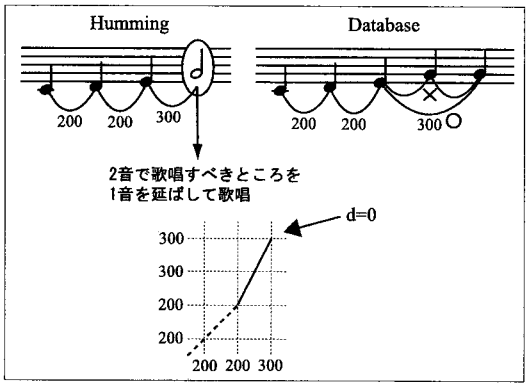

図 4　特徴量の再計算とマッチング

$$\boldsymbol{C}_{i1} \overset{PCA}{\rightarrow} \boldsymbol{z}_{i1} \tag{7}$$

$$\boldsymbol{C}_{i2} \overset{PCA}{\rightarrow} \boldsymbol{z}_{i2} \tag{8}$$

と変換する．

データベース中の楽曲の音程に関しても同様に次の T_{j2}, T_{j2} の 2 通り考えることができる．

$$T_{j1} = m_p(j+1) - m_p(j) \tag{9}$$

$$T_{j2} = m_p(j+1) - m_p(j-1) \tag{10}$$

ここで，T_{j1}, T_{j2} は楽曲の j 番目の区間の音程であり，$m_p(j)$ は楽曲の j 番目の区間の対数基本周波数である．

この時，挿入・脱落発生時には以下のようにして音程に対する確率密度を計算する．

- 挿入：$p_1(i,j) = p(\boldsymbol{z}_{i2}|\boldsymbol{\mu}(T_{j1}), \boldsymbol{B}(T_{j1}), \boldsymbol{P}(T_{j1}))$
- 通常：$p_2(i,j) = p(\boldsymbol{z}_{i1}|\boldsymbol{\mu}(T_{j1}), \boldsymbol{B}(T_{j1}), \boldsymbol{P}(T_{j1}))$
- 脱落：$p_3(i,j) = p(\boldsymbol{z}_{i1}|\boldsymbol{\mu}(T_{j2}), \boldsymbol{B}(T_{j2}), \boldsymbol{P}(T_{j2}))$

ここで，$\boldsymbol{\mu}(T)$, $\boldsymbol{B}(T)$, $\boldsymbol{P}(T)$ は音程 T の確率モデルのパラメータである．

また，検索には音長情報も用いる．音長情報には IOI (Inter-Onset-Interval) 比を用いる．これらについては以下のように考える．入力の i 番目の区間の音長 $h_t(i)$ とデータベースの j 番目の区間の音長 $m_t(j)$ から，入力・参照側それぞれ 2 つのケースの音長の比を定義する．

$$\Delta_1 h_t(i) = \log \frac{h_t(i+1)}{h_t(i)} \tag{11}$$

$$\Delta_2 h_t(i) = \log \frac{h_t(i+1) + h_t(i)}{h_t(i-1)} \tag{12}$$

$$\Delta_1 m_t(j) = \log \frac{m_t(j+1)}{m_t(j)} \tag{13}$$

$$\Delta_2 m_t(j) = \log \frac{m_t(j+1) + m_t(j)}{m_t(j-1)} \tag{14}$$

検索時には次の式に基づいて入力とデータベースの音長の比較をする．

- 挿入：$t_1(i,j) = |\Delta_2 h_t(i) - \Delta_1 m_t(j)|$
- 通常：$t_2(i,j) = |\Delta_1 h_t(i) - \Delta_1 m_t(j)|$
- 脱落：$t_3(i,j) = |\Delta_1 h_t(i) - \Delta_2 m_t(j)|$

これにより，特徴量の再計算を実装している．

さらに，音程尤度は高い方がよいスコアで，音長は距離でスコア付けするために小さい方がよいスコアとなる．そこで，各格子点でのスコアは，音程尤度から音長距離を重みをつけて引くという操作を行った．つまり検索を行うときは，次の漸化式から検索を行う．

$$g(0,0) = l_2(0,0) \tag{15}$$

$$g(i,0) = -\infty (i > 1) \tag{16}$$

$$g(0,j) = l_2(0,j) (j > 1) \tag{17}$$

$$g(i,j) = \max \begin{cases} g(i-2, j-1) + l_1(i,j) + penalty_1, \\ g(i-1, j-1) + l_2(i,j), \\ g(i-1, j-2) + 2l_3(i,j) + penalty_3 \end{cases} \tag{18}$$

$$l_e(i,j) = weight \times \log p_e(i,j) - (1 - weight) \times t_e(i,j) \quad (e = 1,2,3) \tag{19}$$

ここで，$weight$ はスコアリング重み，$penalty_1$, $penalty_3$ は，それぞれ挿入・脱落のペナルティであり，このペナルティを導入することにより検索精度が向上することが報告されている [7]．

3. 実験結果

今回，音程推定実験と楽曲検索実験の 2 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．

3.1 実験条件

表 1 に今回行った 2 つの実験の共通の環境を示す．

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 特徴ベクトル 225,000 個
	音程候補	25 候補 -1200 ~ +1200cent 100cent 刻み
音声分析	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	69 次元 CCF の 10 主成分
	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 特徴ベクトル 18,000 個
	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
CCF:相互相関関数

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 → 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, ピッチ抽出による同様の実験も行った. ピッチ抽出は 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, 各フレームのピッチの中央値をピッチとして採用した.

3.2 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 3 位以内に正解が存在するか」「音程尤度上位 5 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. 音程推定の結果としては,

- 最尤音程の正解率はピッチによる音程計算の正解率を大きく上回った
- 3 位以内にほぼ正解が含まれており, 検索への有効である

ということがいえる. また, ラプラス分布と単一正規分布, 混合正規分布による確率密度関数の計算による推定精度はほとんど同等であり, どちらが有効であるかについては言及できない.

表 2 音程推定実験結果

	1 位	3 位以内	5 位以内
ラプラス分布	63.2 %	98.7 %	99.9 %
単一正規分布	63.2 %	99.1 %	100.0 %
混合正規分布 (2 混合)	62.8 %	98.5 %	99.9 %
ピッチ	17.0 %	-	-

表 3 楽曲検索精度

	1 位検索率	10 位以内検索率	検索時間
ラプラス分布	79.1 %	92.0 %	17.7 sec
単一正規分布	77.3 %	90.2 %	17.5 sec
混合正規分布 (2 混合)	75.5 %	90.5 %	20.8 sec
単一ピッチ	74.2 %	89.0 %	13.0 sec
複数ピッチ	86.5 %	94.1 %	116.7 sec

3.3 楽曲検索実験

続いて, 本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,

- 確率分布の違いにより, 検索率には若干の優劣が見られた
- 提案手法における検索精度は, 単一ピッチによるものからわずかに改善が見られるものの, 複数ピッチ候補を用いた検索には及ばない
- 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.

3.4 性能に関する考察

今回, 性能が複数ピッチに及ばなかった原因として, 音程モデルの近似が不十分であったということが考えられる.

図 5, 6 に学習データの分布の一例を示している. これらは特徴ベクトルのある 1 次元の分布を図示したものである. これらはあくまで一例であるが, 他の学習データの分布でも同様の形状で分布していることが確かめられた. 学習データのヒストグラムから,

- 学習データの分布は単一正規分布ではない.

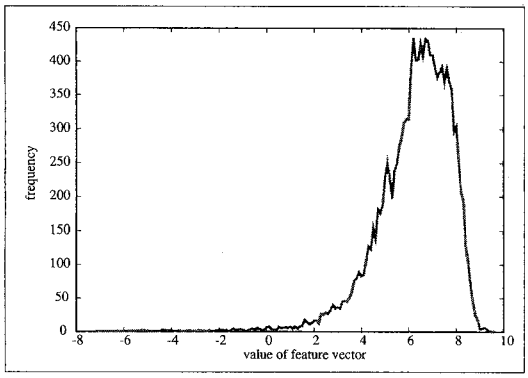

図 5　学習データの特徴量分布 (音程：+300cent 第 1 次元)

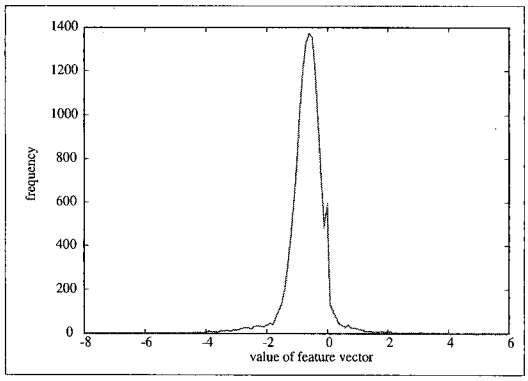

図 6　学習データの特徴量分布 (音程：+300cent 第 2 次元)

- 多くがラプラス分布に近い形状であるが，明らかにラプラス分布とは違う分布をするものがある．

ということが確かめられた．これらから，ラプラス分布・単一正規分布によるモデル作成が十分に効力を発揮しなかったと考えられる．

また，混合正規分布によるモデル化では，今回 2 混合でモデルを作成したためにフィッティングが完全になされなかったということが判明した．特に急峻なピークをもつ分布ほどその傾向が強く現れており，実際のヒストグラムよりも分散の大きい確率分布が推定されている．これにより混合正規分布の性能が悪かったと考えられる．この現象が発生した原因は，学習を頑健にするために分散閾値を設定しており，その閾値よりも実際の分散が小さくなってしまったということが考えられる．

今回，混合正規分布は低い性能となってしまったが，このフィッティングが完全になされれば，性能が向上すると考えられる．

4. 結論と今後の予定

ハミング入力の楽曲検索システムの高精度化を目指し，確率モデルを用いた楽曲検索システムを構築した．このシステムの実現のため，デルタピッチに代わる，音程を表す新たな特徴量を提案し，この特徴量を用いて検索を行った．

この手法により，楽曲検索時間をほとんど増加させることなく，検索精度を 4.9 %向上させることができた．しかし，複数ピッチ候補の検索率には及ばず，さらなる性能向上をしなくては効果的な手法とはいえない．

今後は，さらなる性能向上のため，混合数を増加しての音程モデル作成を行う予定である．

文　献

[1] N.Kosugi *et al.*, "A practical query-by-humming system for a large music database", *ACM Multimedia 2000*, pp.333-342, 2000

[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 *et al.*, "Super MBox:an efficient/effective content-based music retrieval system", *Ninth ACM Multimedia Conf. (Demo Paper)*, pp.636-637, 2001

[4] Sung-Phil Heo *et al.*, "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 *et al.*, "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] Shigeki Sagayama *et al.*, "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.

[7] Akinori Ito *et al.*, "Comparison of Features For DP-Matching Based Query-by-Humming System" *Proc.5th International Conference on Music Information Retrieval(ISMIR2004)*, pp297-302, 2004