G-014

鼻歌入力による音楽検索のための特徴量の比較
Comparison of features for Query-by-Humming MIR

伊藤彰則† 許盛弼‡ 鈴木基之† 牧野正三†
Akinori Ito Sung-Phil Heo Motoyuki Suzuki Shozo Makino

1. はじめに

鼻歌（ハミング）を利用して音楽を検索する手法が各種提案されている [1, 2]．我々は，3次元DPマッチングを用いたハミング音楽検索システムを作成している [3, 4]．検索のための特徴量として，Ghias ら [1] は量子化された音高値，園田ら [2] は階層的に量子化されたピッチ比 (deltaPitch) と音長比 (IOIratio) を用いている．また，我々のシステムでは，deltaPitch と IOIratio の差の絶対値を距離として利用している．このようにさまざまな特徴量が音楽検索のために用いられているが，それらの特徴量の優劣については十分検討されていない．そこで，本稿では従来用いられていた「距離による方法」「量子化による方法」に加え，新しい方法として「ファジイ量子化による方法」を提案し，これらの手法を比較する．本稿では deltaPitch のみを用い，マッチング手法としては連続DPマッチング [5] を用いる．

2. 特徴量の定義

データベースの m 番目の曲の i 番目の値と，入力の j 番目の値との類似度を $S_m(i,j)$ とする．この $S_m(i,j)$ の定義を変えることで，異なる特徴量の比較を行う．

本稿では，deltaPitch を基本とした方法を用いており，音長の情報は用いていない．音長情報の利用は今後の課題である．ある連続した2つの音符に対応する周波数を f_1, f_2 とすると，deltaPitch は

$$\Delta P = 1200 \log_2 \frac{f_2}{f_1} \quad \text{(cent)} \tag{1}$$

として計算される．

2.1 距離に基づく方法

距離に基づく方法は，データベースと入力音符の deltaPitch の値の差をそのまま特徴量として用いる方法である．データベースの m 番目の曲の i 番目の deltaPitch を $d_m(i)$，入力の j 番目の値を $h(j)$ とするとき，類似度は

$$S_m(i,j) = -|d_m(i) - h(j)| \tag{2}$$

で与えられる．

2.2 量子化に基づく方法

量子化コードを $0, 1, \ldots, K-1$ とし，それぞれの量子化レベルの中央値を $\mu_0, \ldots, \mu_{K-1}$ とする．このとき，量子化関数を

$$Q(x) = \underset{k}{\operatorname{argmin}} |x - \mu_k| \tag{3}$$

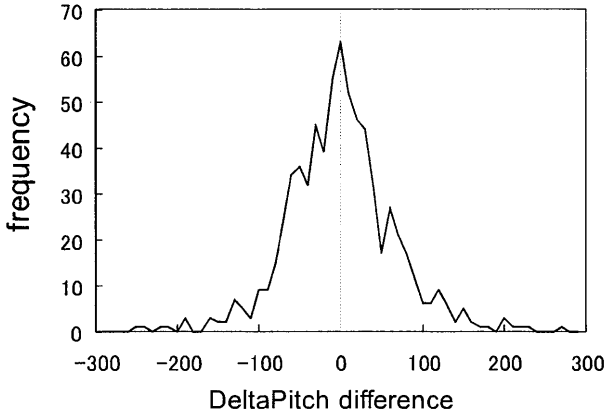

図 1: Histogram of deltaPitch error.

とし，類似度を次のように定義する．

$$S_m(i,j) = \begin{cases} 1 & \text{if } Q(h(j)) = Q(d_m(i)) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

2.3 ファジイ量子化による方法

データベースの deltaPitch と，対応する入力の deltaPitch の差のヒストグラムを観測すると，Laplace 分布に似た中央の尖った分布を示す．これを図1に示す．そこで，データベースと入力との差を一般化された Laplace 分布で近似し，そこからメンバーシップ関数を設計することによってファジイ量子化を行う．まず，deltaPitch の差の分布を

$$\phi(x) \equiv \frac{\gamma}{2\Gamma\left(\frac{1}{\gamma}\right) v^{\frac{1}{\gamma}}} \exp\left\{-\frac{|x|^\gamma}{v}\right\} \tag{5}$$

と仮定する．γ と v は定数であり，v は分布の分散を，γ は尖度を制御するパラメータである．この分布は，$\gamma = 1$ の場合は Laplace 分布に等しく，$\gamma = 2$ の場合は正規分布に等しい．次に，入力 x の量子化レベル k についてのメンバーシップ関数を

$$R(x,k) \equiv \frac{\phi(x - \mu_k)}{\sum_i \phi(x - \mu_i)} \tag{6}$$

とする．これを用いて，類似度は次のように定義される．

$$S_m(i,j) = R(h(j), Q(d_m(i))) \tag{7}$$

†東北大学，Tohoku Univ.
‡Korea Telecom

表 1: Large-scale database.

音楽 DB	曲数	童謡:	155
		自動生成:	10,000
		計:	10,155
	平均音符数		57.8
鼻歌データ	歌唱者		男性 5 名
	データ数		320
	平均音符数		11.7

3. 評価実験

実験に用いたデータベースの概要を表 1 に示す．音楽データベースのうち，155 曲は童謡を MIDI 化したデータであり，10000 曲は実データ 155 曲で学習 k した trigram モデルにより自動生成されたデータである．ハミングデータは男性 5 名の歌唱した 320 曲である．量子化およびファジイ量子化における中央値 $\mu_i\,(i=0,\ldots,K-1)$ は，

$$\mu_i = 100\left(i - \frac{K-1}{2}\right) \tag{8}$$

とした．

検索実験の結果を図 2 に示す．量子化に基づく方法における量子化レベル，およびファジイ量子化に基づく方法における量子化レベル・v・γ は，男性 1 名を用いた予備実験によってあらかじめ最適化してある．量子化およびファジイ量子化の量子化レベルは 9，$v=50, \gamma=0.8$ である．

5 名についての検索実験の結果を図 2 に示す．評価基準として，検索スコアの良かった上位 10 曲の中に正解が含まれる割合 (Top-10 recall) を用いた．この結果から，距離に基づく方法が最も性能が良く，ファジイ量子化による方法がそれに次ぐことがわかる．この傾向は話者によってほとんど変動しないこともわかる．単純な量子化による方法は性能が良くなかった．

次に，距離に基づく方法とファジイ量子化に基づく方法を組み合わせて性能を向上させる方法について検討した．m 番目の曲に対する距離に基づく方法の検索スコアを V_m^{dist}，ファジイ量子化によるスコアを V_m^{FQ} とするとき，全体の検索スコアを

$$V_m = \lambda V_m^{dist} + (1-\lambda) V_m^{FQ} \tag{9}$$

として計算する．5 名に対する平均の結果を図 3 に示す．この結果から，$\lambda = 0.3$ の場合に，距離による方法よりも 1 ポイント性能が向上することがわかる．

4. まとめ

鼻歌入力に基づく音楽検索システムの特徴量について比較検討した．その結果，距離に基づく方法が最も高い検索性能を示すことが明らかになった．また，ファジイ量子化を用いた方法を併用することにより，検索性能を向上させることができた．今後は，複数ピッチを用いた特徴量などについて検討していきたいと考えている．

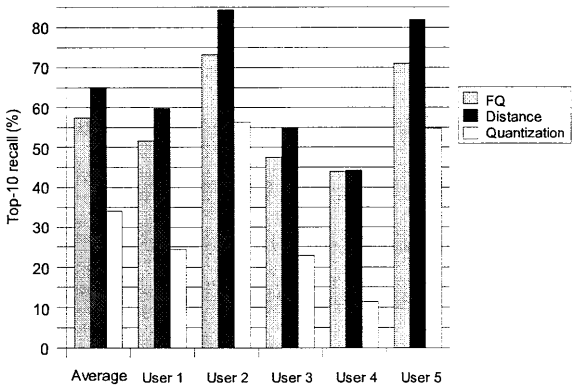

図 2: Retrieval result for five users.

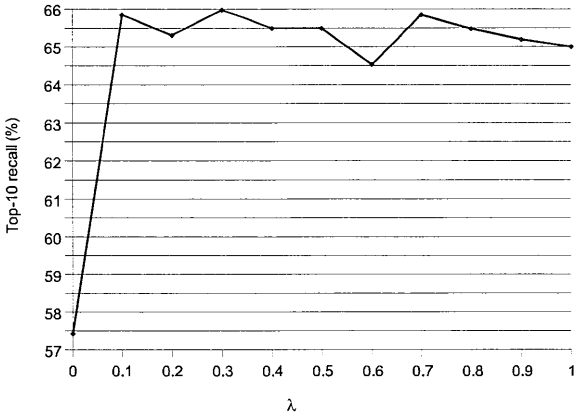

図 3: Retrieval result using the combined score.

参考文献

[1] A. Ghias, J. Logan, D. Chamberlin and B. C. Smith, *Query by Humming: Musical Information Retrieval in an Audio Database,* Proc. ACM Multimedia, 1995.

[2] T. Sonoda and Y. Muraoka, *A WWW-based Melody Retrieval System–An Indexing Method for A Large Database–,* Proc. ICMC, 2000.

[3] S.-P. Heo, M. Suzuki, A. Ito and S. Makino, *Three Dimensional Continuous DP Algorithm for Multiple Pitch Candidates in Music Information Retrieval System,* Proc. ISMIR2003, pp. 235-236, 2003.

[4] S.-P. Heo, M. Suzuki, A. Ito, S. Makino and H. Chung, *Multiple pitch candidate based music information retrieval method for query-by-humming*, Proc. Int. Workshop on Adaptive Multimedia Retrieval, 189–200, 2003.

[5] 岡隆一：「連続 DP を用いた連続単語音声認識」，音響学会音声研資料 S81-65, 1978.