

歌声による曲検索システム

— 音程・音長情報の分布を利用した閾値の決定法 —

1J-06

園田 智也 後藤 真孝 村岡 洋一

早稲田大学 理工学部

1. はじめに

本稿では曲データベース (DB) 検索の手法として、曲のメロディを口ずさむとそのタイトルを出力するシステムについて述べる。

従来の研究 [1]~[4] の多くでは、音符の旋律情報 (音程・音長) のうち、主に音程情報が検索キーとして利用され、音長情報をキーとした検索は比較的精度が悪いことが指摘されていた [1][2]。しかし、音長情報は本来有効な情報であり、音長が適切に利用できれば、精度の良い検索ができるはずである。

そこで、本研究では DB 中の音符の情報分布を利用することで、音長情報による検索を音程による検索と同等の精度で行い、二つを組み合わせることでより正確な検索を実現するシステムを提案する。

2. システムの概要

本システムの概要を図 1 に示す。ユーザはメロディを口ずさんでマイク入力を行なう。システムではまず入力されたメロディの「音程・音長抽出」を行なう。次に、得られた音程・音長情報を DB 中の各曲の持つ情報と照合する。入力キー key と DB 中の曲 $song$ との音程・音長情報の距離 D_p, D_d を $\forall song \in DB$ についてそれぞれ計算し、その和 $D = D_p + D_d$ が最小となる $song$ のタイトルを検索結果として出力する。以下に各処理の概要を述べる。

2.1 音程抽出と音長抽出

入力旋律の A/D 変換後、まず入力音のパワーがある閾値以下の箇所を音符に区切る (メロディを ta や cha などで歌うと適切に区切りやすい)。次に各音符に対して、文献 [1] と同様の手法で音高を求める。また、各音符間の開始点の距離を音長とする。歌声は音高・音長が正確とは限らないため、音高・音長はさらに直前の音符からの相対音高差 (音程)、音長比に変換する。

2.2 検索

key と $song$ の照合には、時系列の照合の手法で代表的な DP マッチングを用いる。 key と $song$ の各音符間で削除・挿入・置換があるごとにペナルティを与え、その合計を両者の距離とする。

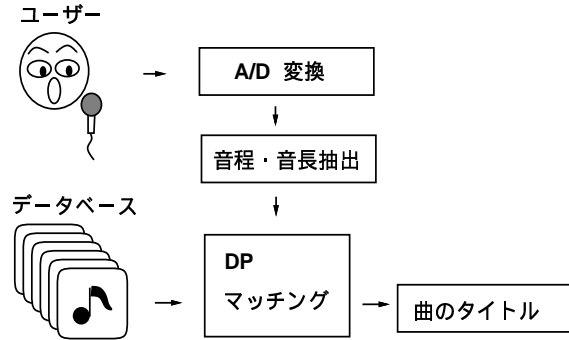


図 1: 曲検索システム

3. 検索キー抽出のための閾値決定問題

以上のシステムにおいて、全体の検索精度に大きな影響を与えるのがマッチング時の旋律情報の形式である。入力キーの歌声の音程・音長は DB と完全に一致することはない。そこで、各音符について検出された音程、音長比は、ある範囲の値を同一の記号とし、その記号列を検索キーとして用いる必要がある。マッチング時には DB 中の曲の旋律情報も同様の記号列に変換すれば良い。

例えば、音程情報において前音より「上がった、下がった、同じ」という情報を UP, DOWN, SAME の頭文字で表現するとした場合、「ドレミミレド」は「XUUSDD」という記号列に変換される (X は先頭のため相対値がないことを意味する)。音長も同様に「長くなった、短くなった、同じ」を表す記号列に変換する。

この際、従来の手法では、静的な閾値で「上がった、下がった、同じ」の判定を行なっていたため、以下で述べるように DB 性質によってはうまく曲の絞り込みが行なえない場合があった。本システムでは DB の性質によって動的に閾値を決定し、効率よく曲の絞り込みを行なうことを可能にする。

4. 静的閾値と動的閾値

図 2 は、ある DB 中の 100 曲中の第 2 音符目の音程情報のヒストグラムである (DB は歌声によって作成した)。横軸は第 1 音符目からの音程 (半音を 1 で表現する)、縦軸は度数である。

半音以下の音程を「同じ」と判定すると (静的閾値が半音の場合)、多くの音符が前の音符より「上がった」という右側のカテゴリに分類されてしまい、従来手法では効果的に曲の絞り込みができていないことが分かる。

本手法では、閾値によって分割される3つのカテゴリ内の合計度数が3等分されるように、その値を動的に決める。これにより入力のある音符がどのカテゴリに分類されても得られる情報量は等しくなり、入力キーの1音符ごとに曲は1/3に絞られていくことが期待される。

音長情報についても同様にヒストグラムを作成し、3等分するように動的に閾値を決定することで、効率よく曲の絞り込みが行なえる。

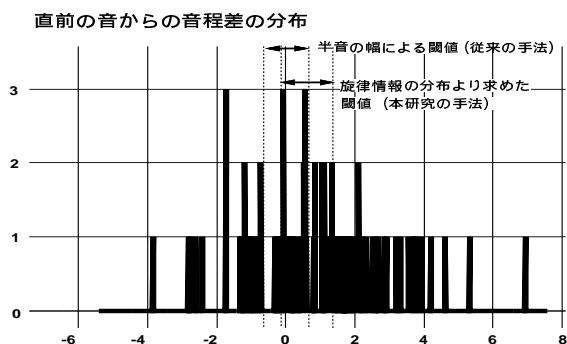


図 2: 音程情報の分布 (100 曲中の第 2 音符目)

5. 動的閾値の決定法

具体的な動的閾値の決定手順を以下に述べる。

1. DB 内の全曲中の i 番目の音程 (音長) 情報の値で図 2 のようなヒストグラムを作成する。
2. ヒストグラムの総度数を Sum とし、カテゴリ数を $Parts$ としたとき、1つのカテゴリ内の合計度数は $M = Sum/Parts$ となる。
3. 各カテゴリの合計度数が均等に M に近くなるようにヒストグラムを分割し、その各境界線を閾値 $Th(i, j)$ ($j = 0..Parts - 2$) とする。
4. この作業を $i = 1..N$ (N は全曲中の最大音符数) で繰り返し、 $Th(i, j)$ を決定していく。

こうして得られた閾値を、検索時の DP マッチングにおいて距離を計算する際に用いる。すなわち、入力キー中の音符を $song$ の s 番目の音符と比較する際には、対応する閾値である $Th(s, j)$ を用いて U, D, S などの記号列に変換していく。

6. 実験と考察

従来の静的閾値を利用した検索と、我々が提案する動的閾値を利用した検索の精度の比較を行なった。様々なジャンル (ポップス・演歌・童謡など) 100 曲の DB に対して、音程情報のみの検索、音長情報のみの検索、両者の組み合わせによる検索をそれぞれ 100 回行なった (被験者 12 人の入力)。閾値が静的か動的かという条件以外はまったく同じ条件で行ない、静的閾値は、

音程に関しては半音以内のずれを同音程とみなし、音長に関しては $1/2$ より大きく 2 未満の音長比を同音長とみなした ([2] を参考にした)。また、マッチングは曲の先頭からの一致で行ない、補正のために、2 つ前の音符からの音程差・音長比も使用している。

入力を 8 秒間で限定したところ、入力キーの平均音符数は 19.4 であった。以下に結果を示す。

	音程検索	音長検索	組合せ検索
正答率	73%	59%	93%
3 位以内	84%	65%	95%
10 位以内	96%	66%	98%

表 1: 静的閾値を利用した検索 (100 回)

	音程検索	音長検索	組合せ検索
正答率	74%	73%	99%
3 位以内	86%	84%	100%
10 位以内	96%	92%	100%

表 2: 動的閾値を利用した検索 (100 回)

現時点では DB 中の曲数が少ないため、各々の検索精度についての絶対的評価は難しい。しかし、従来の静的閾値を用いた手法では音長情報が音程情報に比べて 15% ~ 30% ほど検索精度が悪いのに対し、本システム的手法では音長情報も音程情報とほぼ同じ精度で検索結果を返している点が注目すべき点である。音程のみでの検索ではあまり差がなかったが、これは、音程情報の静的閾値での分類に偏りがなかったためだと考えられる。音程・音長を組み合わせた検索においても検索率は向上している。

7. まとめ

本研究では動的閾値を用いることで、音長による検索も音程とほぼ同じ精度で検索キーとして活用できるシステムを実現した。今後は大規模データベースに応用することが課題である。

参考文献

- [1] 貝塚 智慧, 後藤 真孝, 村岡 洋一: 歌声の旋律情報と歌詞情報をキーとした曲検索システム, 情報処理学会 第 54 回全国大会, 7J-06, 1997
- [2] 藤山 哲也: 音高・音長情報を利用したメロディ検索, 45 回情報学会全国大会, 1992
- [3] T.Kageyama, K.Mochizuki, Y.Takashima: *Melody Retrieval with Humming*, ICMC Proc., 1993
- [4] Asif Ghias, Jonathan Logan: *Query By Humming - Musical Information Retrieval in an Audio Database*, ACM Multimedia 95, Electronic Proc., 1995