

7J-06 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

貝塚 智憲 後藤 真孝 村岡 洋一

早稲田大学 理工学部

1. はじめに

本稿では音楽データベースを検索する手法として、ある曲のメロディ（フレーズ）を歌うとそのタイトルを出力する曲検索システムについて述べる。従来の研究 [1]~[3] では、歌声の旋律情報（ハミング）のみをキーとして検索を行っていたが、利用者の音高、音長についての記憶とデータベースとの食い違いが正答率を下げる原因となっていた [3]。

これに対し、本研究では歌声の旋律情報に加え、歌詞情報もキーとして用いることで曲の絞り込みをし、より正確な検索をすることができるシステムを提案する。今回、旋律情報として音高を、歌詞情報として母音をキーとしたシステムの実装を行った。

2. 歌詞（母音）情報の利用

歌詞情報を利用するために、子音は扱わず、母音 (a,i,u,e,o) のみをキーとして用いる。歌唱中は会話中よりも母音の発音時間が長いという利点があり、母音の抽出は、厳密な認識でなければ比較的簡単に実装ができる。（我々の実験での母音の認識率は約 80%程度ある。）

また、文献 [4] より入力した 100 曲のデータベースにおける歌詞中の母音の出現確率は（あ:32% い:20% う:14% え:10% お:22%）で、極端な偏りがいないため母音だけでも検索の有効な手がかりとなる。なお、音長+音高の利用も試みたが、効果がないため本研究では音長による検索は採用しなかった。

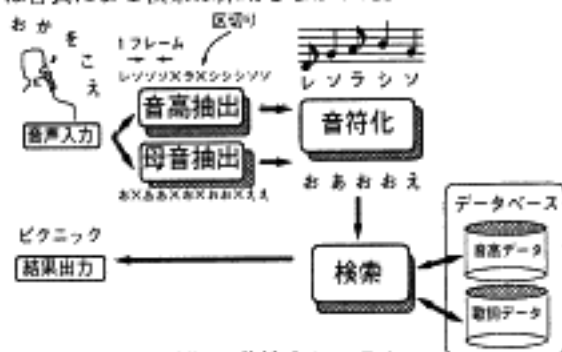


図1: 曲検索システム

3. システムの概要

システムの概要を図1に示す。本システムは、歌詞のついたメロディの一部（フレーズ）を歌って入力するとその曲のタイトルを出力する。まず、「音高抽出」、「母音抽出」が、入力された歌声から処理の単位時間（1フレーム）あたりの音高、母音を決定し、「音符化」が音符ごとの区切りを決める。

次に、こうして得られた音高列、母音列をキーとして、「検索」がデータベース (DB) とのマッチングをする。キーと DB 中の曲 k との音高列、母音列の距離 $D_p(k), D_v(k)$ を $k \in DB$ についてそれぞれ計算し、それを組み合わせた距離 $D(k)$ が最小となる k を検索結果として出力する。このシステムでのデータベース、音高列、母音列の表現方法は次の通りである。

データベース データベース内の曲は音符ごとの音高列、母音列で表す。また、曲の先頭からだけでなくフレーズ単位で検索できるように、1曲をフレーズごとに手作業で区切っておく。

音高列 音高列は半音を1とした前音との相対的な音高の差で表す。例えば図1のレゾラシソは5,2,2,-4となる。

母音列 各音符に対し (a,i,u,e,o) のいずれか1母音を対応させる。

4. 問題点と解決法

実現上問題となる点と本研究での解決法を述べる。

1. 歌声の音符の区切りの決定

音符の区切りを決める処理は、全体の検索の精度に大きく影響するため重要である。歌声では、ハミングに比べて入力音声のパワー変化が少なく、これだけを音符の区切りとして判断すると、2音符を1音符とみなしてしまう場合がある。

そこで、パワー変化に加え「母音抽出」の判定結果も考慮する。子音と判断された場合にその位置を音符の区切りとすることで、同じ音高を連続して発声した場合などにも有効な手がかりとなる。

2. キーとデータベースの曲との距離の測定

歌声から抽出した音高列、母音列は、必ずしもデータベースの検索対象とは一致しない。そこで、DP マッチングを用いて挿入、削除（欠落）、置換をペナルティとして加算した合計をフレーズ間の距離として定義する。

3. 旋律情報と歌詞情報を組合せた検索

音高に基づく距離と母音に基づく距離をどう組み合わせるかが問題となる。

本システムではまず、求めた $D_p(k), D_v(k)$ ($k \in DB$) の平均 \bar{D}_p, \bar{D}_v と標準偏差 $\sigma_{D_p}, \sigma_{D_v}$ をもとめ、式(1)により平均0、標準偏差 c_p, c_v の分布へと変量の変換を行う。 c_p, c_v を変えることで、音高、母音の分布に重みをつけて検索をすることができる。実験では $c_p = 10, c_v = 7$ のときに良い検索結果を得た。

$$D_i(k) = c_i \frac{D_i(k) - \bar{D}_i}{\sigma_{D_i}} \quad (i=p, v) \quad (1)$$

$$D(k) = D_p(k) + D_v(k) \quad (2)$$

変換した2つの値 $D_p(k)$, $D_v(k)$ から、音高、母音を組合せた距離 $D(k)$ を式(2)により求め、組合せ検索の距離とする。

5. 各処理の手順

システムの各処理の手順について述べる。

1. 音高抽出

文献[5]の手法を参考に高速フーリエ変換を用いて基本周波数を同定し、1フレームごとの音高を決定する。このとき基本周波数のパワーが閾値以下の場合、あるいは周波数軸方向のパワー分布において、基本周波数から5倍音までの隣接する極大極小の差が1つでも閾値以下の場合には音符の区切りとする。

2. 歌詞抽出

LPC(線形予測)分析を用いて入力波形のスペクトルの包絡を求め、フレームごとに、あらかじめ被験者ごとに登録した、標準母音とのLPCケプストラム距離が最小の母音を求める。距離が閾値以上の場合には子音と判断し、音符の区切りとする。

3. 音符化

「音高抽出」、「母音抽出」から受けとった区切りの情報をあわせて音符の区切りを決定する。決定した区間に対して音高、母音のヒストグラムをつくり、最大のものをその区間の音高、母音とする。

4. 検索

キーの先頭をデータベースの各曲のそれぞれのフレーズの先頭に固定してDPマッチングを行なう。DPマッチングにおける一致、置換、挿入、削除のペナルティの値は、実験により経験的に表1のように定めた。

	一致	置換	挿入/削除
音高	0	音高差	2
母音	0	a-o,i-e 間 2 その他 3	3

表1: DP マッチングにおけるペナルティ

6. 実験結果

実験はSGI indigo2 Impactで行なった。8kHzでサンプリングされた音声をもとに1フレーム64点(8msec)ずつシフトしながら、FFT,LPC(12次)ともに512点で分析した。文献[4]より入力した100曲(577フレーズ)のデータベースに対し、以下の2つの実験を行なった。

実験1 (音高検索、母音検索、組合せ検索の比較)

4人の被験者(男2人、女2人)が合計132回検索を行なった。歌声の録音時間を8秒間以下に制限し、何も手がかりを与えず、被験者の自由な音高とテンポで入力したところ、平均音符数は11.84音符であった。

表2に「音高検索」、「母音検索」、本研究で提案する「組合せ検索」の3つについて、正解の曲との距離が1位となった確率(正答率)、3位以内となった確率、10位以内となった確率を示す(括弧内はその曲数)。ただし、複数の曲が同距離となった場合はその曲を含む累積度数を順位とする。例えば最小距離の曲が4曲ある場合はそれらすべてを4位とみなす。

	音高検索	母音検索	組合せ検索
正答率	78.7% (104)	67.4% (89)	91.6% (121)
3位以内	89.3% (118)	81.8% (108)	94.7% (125)
10位以内	95.4% (126)	91.6% (121)	97.7% (129)

表2: 音高検索・母音検索・組合せ検索の比較

実験の結果から、組み合わせることで音高検索よりも正答率を12.9%向上できた。誤検索の原因は、主に音符が必要以上に細かく区切られ、距離が大きくなったからであった。

実験2 (キーの長さに対する正答率)

DPマッチングをするキーの音符数(母音数)の長さを変えて正答率を測定した。キーの上限を2~12に変化させ、実験1と同じ条件で実験した。また、キーを正解の音高、母音に置き換えてDPマッチングをして、アルゴリズム上の上限値も同時に求めた。

図2にこれら結果を示す。グラフの横軸はキーの音符(母音)数の平均値である。グラフから検索結果、上限値ともに短いキーの方が組み合わせ検索の効果が大きいことがわかる。

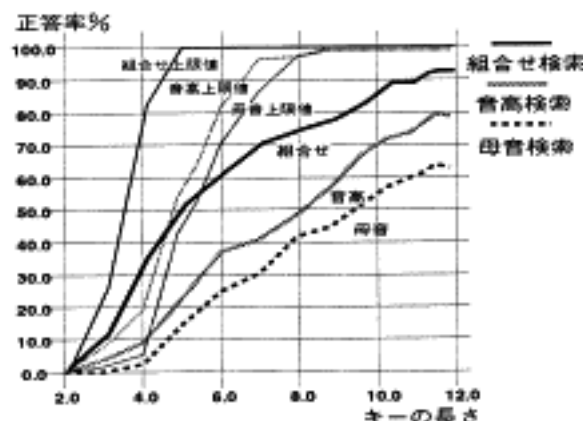


図2: キーの長さと正答率のグラフ

7. まとめ

歌詞情報を利用することにより曲検索の正答率を上げるシステムを提案し、実験によりそれが有効であることを確認した。特に、短いキーの場合に効果が大きいことがわかった。今後は大規模なデータベースに対しての有効性も検証したい。

参考文献

- [1] T.Kageyama, K.Mochizuki, Y.Takashima: *Melody Retrieval with Humming*, ICMC Proc., 1993
- [2] Asif Ghias, Jonathan Logan: *Query By Humming - Musical Information Retrieval in an Audio Database*, ACM Multimedia 95, Electronic Proc., 1995
- [3] 葦山哲也: 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
- [4] 秀学社編集部: 中学愛唱歌集, 秀学社
- [5] 新原高水, 今井正和, 井口征士: 歌唱の自動採譜, 計測自動制御学会論文集, 20-10, 1984