

2N-2

音高・音長情報を利用したメロディ検索

藤山 哲也

日本電気(株) C&C 情報研究所

1 はじめに

音楽データベース検索の検索手がかりとして旋律を利用する方法がある[1]。筆者らはこの旋律手がかりの入力を誰でも簡単に行なえるものとするために、ハミング歌唱を手がかりとする音楽データベース検索手法「メロディ検索」を提案している[2]。

旋律には音高情報と音長情報が含まれているが、ハミングから得られる旋律情報は不完全なものである。この中から、いかに検索のために有効な手がかりを取り出すが、検索精度を決めるポイントとなる。これまでは音高情報だけを検索の手がかりとして楽曲データベースの照合に利用し、検索結果を決めていた。

本稿では検索精度の向上のために、まず実際のハミングデータの誤り方のパターンについて考察することにする。次に音高情報だけでなく音長情報も手がかりとして利用する照合方法について述べ、検索手がかりとしての音高情報と音長情報の有効性について考察する。

2 ハミング手がかりの不完全さ

メロディ検索が必要な状況とは、書誌的な手がかりが不足している場合であろう。したがって書誌的な情報の一つである楽譜があり、更にその楽譜に忠実な歌唱がなされるとは考えにくい。そのためハミングは利用者の記憶に基づくものであることを考慮する必要がある。音階やテンポは利用者の記憶と楽譜との間で異なる場合が多い。旋律の細部を間違えて記憶している場合も考えられる。これに対して、照合の対象となる楽曲データは、楽譜に書かれた旋律を照合のための音符列に変換したものである。ゆえに利用者の記憶している旋律とデータベース中の同じ楽曲の旋律は同一とは限らない。

また照合の前処理としてハミングから旋律の音符列を取り出す採譜処理[3]は、音程のふらついた歌唱や、音符の切り分けがあいまいな歌唱では正しく採譜できない。

3 手がかりの誤り

ビッチを揺らさず音符を明確に区切る歌唱では、かなり正確な採譜が行なえる。このような歌唱を行なえるかどうかは歌唱の訓練に依存する。誤りパターンの考察には、歌唱訓練した人1名(A)の28例、一般の人2名(B,C)の59例、合計3人の87例を用いて採譜結果と楽譜[4][5]とを比較した。誤りの形態別に、音符あたりの誤り頻度を表1に示す。

表1: 誤りの形態と頻度

誤りの種類	被験者 A	被験者 B	被験者 C
音高のずれ	0.179	0.255	0.249
音長のずれ	0.406	0.476	0.587
音符の挿入	0.022	0.046	0.043
音符の脱落	0.026	0.120	0.115

この音高のずれの60%は半音ずれた場合である。歌唱ビッチは不安定なので、採譜時の半音判別は特に困難である。これが半音ずれの原因だと考えられる。またビッチ変動のためか1音符だけ単独で音高がずれる例が音高のずれの32%あった。

ハミング入力時には採譜の基準となるテンポを指揮棒の画面で提示しており、得られる手がかりの音長値も十六分音符単位になっている。音長のずれた音符の中の42%は、付点八分音符と四分音符を間違えるというように2分の1より長く2倍未満の値となっているものである。

このように、音長値は取りえる値の種類も少なく、誤差も含まれている。手がかりとしての有効性は、音高情報よりも音長情報の方が少ないことが予想できる。

4 類似度定義とパラメータ設定

ハミングという不完全な手がかりと楽曲の照合する候補となるフレーズとの照合には、先に述べたような誤りを補正する必要がある。二つの次元パターンを伸縮を認めて照合を行なう計算法としてDP照合がある。ここでは、手がかりフレーズと候補フレーズを次元パターンとして表し、DP照合の枠組みでフレーズ類似度を定めることにする。以下の $a, b_1, b_2, b_3, b_4, c_1, c_2$ は類似度を定義するパラメータである。

1. 音符の挿入・脱落時に減点 a を加える。また照合の組み合わせの範囲を狭めるための整合幅を設ける。
2. 音高は前音との音程差を求めることでフレーズ間の音階のずれを補正する。音符単位の音高一致度 S_P を決めるため、フレーズのそれぞれ前音との音程差を求めて S_{P1} を次のように決める。

$$S_{P1} = \begin{cases} b_1 & (\text{音程差が一致したとき}) \\ b_2 & (\text{音程差が半音ずれたとき}) \\ b_3 & (\text{音程差が半音より大きくずれたとき}) \end{cases}$$

ある1音だけ音高値がずれると音程差系列では前後の二つの値がずれてしまう。この補正のためフレーズのそれぞれ2音前の音との音程差から S_{P2} を、

$$S_{P2} = \begin{cases} b_4 & (\text{音程差が一致したとき}) \\ 0 & (\text{音程差が半音以上ずれたとき}) \end{cases}$$

と決め、音高一致度 S_P を $S_P = S_{P1} + S_{P2}$ とする。

3. テンポの異なったフレーズを比較する必要があるため、音長は双方のフレーズ中の前音との音長比を比較する。音長一致度 S_T は次のように決める。

$$S_T = \begin{cases} c_1 & (1/2 < \text{前音との音長比} < 2) \\ c_2 & (\text{それ以外}) \end{cases}$$

4. 音高一致度 S_P と音長一致度 S_T の和を音符一致度 S とする。手がかりと楽曲フレーズ間の対応する音符の S の総和に音符の挿入・脱落の減点 a をそれぞれ加えた値は DP によって求まるので、この値を(手がかりの音符数 -1) で割ることによって正規化し、フレーズ類似度とする(音符一致度は完全一致時に 1 になるようなパラメータに定める)。

類似度定義で用いたパラメータ値は、検索手がかりとしての音高・音長の有効性を見るために、以下の3通りについて調べることにする。

1. 音長一致度だけで評価した場合
($b_1 = b_2 = b_3 = b_4 = 0$)
2. 音高一致度だけで評価した場合
($c_1 = c_2 = 0$)
3. 音高と音長の組み合わせ

他の類似度パラメータ値は、パラメータを変化させながら、「手がかりと意図した楽曲の類似度と他の類似度最大の曲の類似度との差の期待値」(Difference of Similarity, DS) が最大になるように決めた。この DS は意図した楽曲が他の楽曲集団と比べて、類似度でどの程度上回っているかを示す指標である。DS を求めるために用いる手がかりは3節で述べた3人のハミング87例(平均音符長 18.5)であり、楽曲データベースは童謡・歌謡曲・演歌のジャンルの合計 124 曲からなるものである。

なお、照合の組み合わせを制限する整合窓の幅は、手がかりフレーズについて音符の脱落個数の期待値 1.64、標準偏差 1.85 なので、手がかりの脱落方向は 4 とし、音符の挿入個数の期待値 0.69、標準偏差 0.85 なので、手がかりの挿入方向は 2 として固定した。なお、整合窓の幅をこれ以上増やしても、DS は殆ど変わらなかった。

5 類似度パラメータ値の評価

類似度の定義に基づいて決められたパラメータ値の妥当性の評価を行った。評価のための手がかりとしては、歌唱訓練した人 1 名を含む 10 名の 50 例(各人 5 例ずつ、平均音符長 12.4)を用いた。楽曲データベースは類似度パラメータ設定に用いたものを使用した。類似度パラメータ設定の際の検索結果 X と評価のための検索結果 Y をまとめて表 2 に示す。

この結果について次のことがいえる。

1. 音長系列については、X と Y どちらのハミング例についても DS が負となるなど単独の検索精度はあまり良くなかった。
2. 音高系列は音長系列に比べて検索精度も良いことから、音高の揺れにもかかわらず、多くの手がかりとなる情報を持っていることが分かる。

表 2: 類似度評価結果

条件		$P_1(\%)$	$P_{10}(\%)$	DS	検索パラメータ
音長のみ	X	29	44	-0.101	$a = -1,$ $b_1 = b_2 = 0,$ $b_3 = b_4 = 0,$ $c_1 = 1, c_2 = 0$
	Y	22	36	-0.156	
音高のみ	X	77	90	0.161	$a = -1,$ $b_1 = 0.8, b_2 = 0.4,$ $b_3 = -1, b_4 = 0.2,$ $c_1 = c_2 = 0$
	Y	62	72	-0.004	
組み合せ	X	76	91	0.199	$a = -1,$ $b_1 = 0.5, b_2 = 0.25,$ $b_3 = -1, b_4 = 0.25,$ $c_1 = 0.25, c_2 = 0$
	Y	62	72	0.039	

P_1 : 類似度 1 位を正解とするときの正解率

P_{10} : 類似度 10 位までを正解とするときの正解率

X : 3 人のハミング 87 例の場合

Y : 10 人のハミング 50 例の場合

3. 音高情報と音長情報との組み合わせは、音高情報だけのものと、正解率 P_1 、 P_{10} ではほぼ変わらなかったが、DS の値では最もよい結果を得た。
4. 音高情報と音長情報を組み合わせる場合のパラメータ値は、音符の完全一致時の音高一致度 S_P > 音長一致度 S_T であった。このことは今回の類似度定義の中で、音高は音長よりも多くの手がかりを含んでいることを示していると考えられる。
5. 評価に用いた 10 人の検索結果 Y は、パラメータ設定に用いた 3 人の X より悪くなっている。手がかりと意図した楽曲フレーズとの類似度自体は X と Y の間で差が見られなかったのに、Y のハミングの長さが短かったのが主な原因だと考えられる。

6 むすび

ハミングを手がかりとする音楽データベース検索手法について、手がかりとなるハミングデータの誤りパターンについて考察し、音高値と音長値を手がかりとした類似度計算方法について述べた。

また、手がかりの情報の中で、音高情報、音長情報がそれぞれ単独の場合、組み合わせの場合について検索の精度を比較し、本稿で述べた類似度定義では、音高系列情報がかなり有効な手がかりとなることがわかった。

参考文献

- [1] 山本: 音楽データベース, 情報処理, Vol.29, No.6, pp.599-607(1988).
- [2] 藤山ほか: メロディ検索 - ハミングで音楽 DB を検索する -, 第 43 回情報大全, 3M-5(1991).
- [3] 水野ほか: パーソナルコンピュータミュージックシステム - 歌声の自動採譜 -, 第 35 回情報大全, 5Ff-5(1987).
- [4] 中田喜直 編: 「こどものうた」, 野ばら社(1984).
- [5] 浅野純 編: 「歌謡曲のすべて」, 全音楽出版社(1987).