

ハミング楽曲検索のための音響信号からのデータベース生成の検討

小杉 優† 伊藤 仁† 伊藤 彰則† 牧野 正三†

† 東北大学大学院工学研究科 電気・通信工学専攻

1 はじめに

近年、楽曲検索の必要性が高まっており、メロディーのハミング入力による楽曲検索に関する研究が進められている [2]。ハミング楽曲検索は、入力音声から特徴量を抽出し、楽曲データベースとマッチングを行うことにより実現される。従来の研究では、その楽曲のメロディーラインを記述した MIDI 等による楽曲データベースを仮定しているが、実際には CD 音源のような音響情報しか存在していない場合も多い。そこで音響信号から直接データベースを作成する必要が生じるが、多数音源が含まれる音響信号からのメロディーラインの基本周波数 (F0) を正しく推定することは困難を極める。本研究では、楽曲の F0 を一意に決めずに、複数個の F0 候補をもったデータベース生成を検討する。また、上述のデータベースを用いた検索手法についても検討する。

2 データベース作成の流れ

本研究におけるデータベース作成および楽曲検索の一連の流れは次のようになる。

1. 音響信号の周波数解析
2. 基本周波数 (F0) の存在確率の推定
3. 基本周波数 (F0) 候補の選択
4. データベースの作成

F0 推定には、「PreFEst (Predominant F0 Estimation)」 [1] の PreFEst-core 部を用いる。この PreFEst では、モノラルの音楽音響信号に対し、その信号中のメロディーラインの推定を実現している。このシステムにより、ある時刻 t における周波数 F に基本周波数 (F0) が存在する確率密度関数 $p_{F0}^{(t)}(F)$ を推定することが可能となる。この操作をすべての時刻 t に対し行うことにより、楽曲の確率密度関数が求まる。

PreFEst により求めた確率密度関数から、 $p_{F0}^{(t)}(F)$ の値の高いピークをいくつか選択し、それを F0 の候補とする。このようにして、楽曲の時刻 t での数個の F0 候補をひとつの楽曲のデータとし、データベースの作成が実現できる。

3 楽曲検索手法

3.1 線形マッチング

複数個 F0 候補があるデータベースとのマッチングについては、入力音声系列を線形シフトすることにより、

表 1: 線形マッチング使用楽曲

	RWC ポピュラー音楽データベース (RWC-MDB-P-2001)
入力クエリ	01,02,03,05,06,07,08,09
データベース	01,02,03,04,05,06,07,08,09, 10,11,12,13,14,15

最もスコアの高い位置を見つけ検索することにより実現される。

入力する検索クエリとしては、正確なメロディーラインの F0 (ただし、高さのみ相対化) であると仮定している。

入力音声の F0 系列を $in(t, f)$ 、データベース中の楽曲 i の F0 存在確率を $db_i(t, f)$ とすると、楽曲 i のスコアは以下のように定式化される。

$$\operatorname{argmax}_{T=0 \sim t_{db}, F=0 \sim f_{range}} \left\{ \sum_{t=0}^{t_{in}} in(t, f) * db_i(t+T, f+F) \right\} \quad (1)$$

ただし、 t_{in}, t_{db} はそれぞれ入力および楽曲 i の長さ、 f_{range} は考慮する周波数帯域の幅を示す。 f は cent による周波数表現である。なお、今回のシステムにおける $in(t, f) * db_i(t+T, f+F)$ はマッチングの判定結果を表す。

3.2 テンポ非依存な時間軸の導入

ハミング検索を行う場合、歌唱者によりその歌唱するテンポには個人差が現れる。そこで、構築されているデータベース楽曲の時間軸については、実時間軸ではなく、64 分音符単位を 1 フレームとするテンポ非依存な時間軸を導入した。これにより、クエリのテンポが異なる場合も検索が可能となる。

4 線形マッチングシステムの実装と検証

4.1 実装条件

生成されたデータベースに対し線形マッチングを行った。データベース楽曲には、RWC 研究用音楽データベース [3] のポピュラー音楽のなかから 15 曲を使用した。また入力する検索クエリとしては、同データベースの 8 曲を選択し、楽曲のサビの一部分のメロディーラインの F0 を採譜し手動でラベル付けしたものを入力した (表 1)。

マッチングにおける判定については、以下の 4 パターンについて比較した。

1. F0 候補を 1 個使用し、入力 F0 との差が 50 cent 以内という正解判定によりスコアをつける場合。
2. F0 候補が上位 3 個使用し、(1) と同様に正解判定を順に行う場合。この際、1 位から 3 位 F0 候補の間で傾斜配点 (1 位:2 位:3 位 = 3:2:1 の比率) を行う。

Database generation from acoustic signal for music information retrieval system with Query-by-Humming

† Yu Kosugi, Masashi Ito, Akinori Ito, Shozo Makino

† Graduate School of Engineering, Tohoku University

表 2: 各手法の検索順位

入力楽曲	手法 (1)	手法 (2)	手法 (3)	手法 (4)
楽曲 01	1 位	1 位	1 位	4 位
楽曲 02	2 位	2 位	1 位	6 位
楽曲 03	1 位	1 位	1 位	1 位
楽曲 05	1 位	1 位	1 位	1 位
楽曲 06	11 位	9 位	8 位	13 位
楽曲 07	1 位	1 位	1 位	1 位
楽曲 08	1 位	1 位	1 位	1 位
楽曲 09	11 位	8 位	9 位	9 位
平均順位	3.6 位	3.0 位	2.9 位	4.5 位

3. F0 候補が上位 3 個使用し, (1) と同様に順に行う場合. 傾斜配点なし
4. PreFEst-core で出力された確率密度 $p_{F0}^{(i)}(F)$ を対数変換し, スコアを足し合わせた結果を用いる場合.

4.2 結果

結果を表 2 に示す. 手法 (1) でも, 8 曲中 5 曲で 1 位を取れているが, 楽曲 06,09 では 15 曲中 11 位と低い結果となってしまっている. なお, 手法 (3) では楽曲 06,09 を除き 1 位を実現した. 一方, 手法 (4) では, 他の手法では上位に来ていた楽曲まで順位が下がるといった傾向が見られた.

4.3 考察

手法 (1) について, 楽曲 06,09 が 15 曲中 11 位と低い結果となってしまっている原因は, F0 推定の不正確さだと考えられる. 本来のサビのメロディーラインの F0 が図 1 のような軌跡をとるにも関わらず, F0 推定して出てきた結果は図 2 のようになり, 一致していない. 逆に, 一例として楽曲 09 を入力した際に正解楽曲よりも高いスコアを出した楽曲 01 の該当箇所 (図 3) を見ると比較的マッチしているといえる. このため正解精度が低くなったと考えられる.

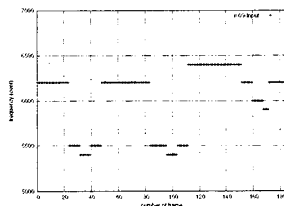


図 1: 楽曲 09 の正解 F0 (手動ラベル付け)

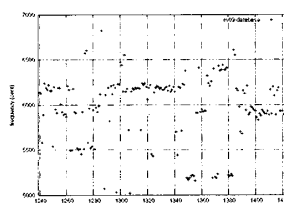


図 2: 楽曲 09 の F0 推定結果 (最大スコア区間)

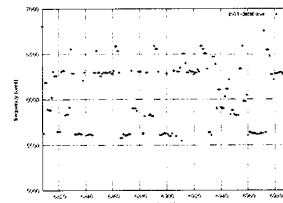


図 3: 楽曲 01 の F0 推定結果 (最大スコア区間)

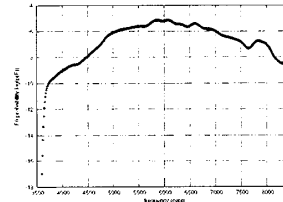


図 4: 楽曲 09 の F0 確率密度分布 (1 フレーム分)

しかし, 手法 (3) を用いると, 楽曲 06 は 8 位まで改善されていることがわかる. このように, 複数個 F0 候補を用いることにより, データベース生成の段階での F0 抽出の精度を埋め合わせることが可能となることが考えられる.

手法 (4) については, データベース作成において得られた対数確率密度 $\log p_{F0}^{(i)}(F)$ の値の差があまり小さくなく (図 4), F0 の正解値とそれ以外との差があまりつかなかったことが原因と考えられる.

5 まとめ

本研究では, ハミング検索のために, 複数個の F0 候補をもったデータベース生成を行い, そのデータベースを用いた検索手法を検討した. F0 の複数候補を用いることで, 候補 1 つの場合の平均順位 3.6 位から, 平均 2.9 位まで向上した. 今後は, マッチングアルゴリズムについての検討が必要である.

参考文献

- [1] Masataka Goto, "A Real-time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals", *Speech Communication* Vol.43, No.4, pp.311-329, September 2004.
- [2] 後藤真孝, 平田圭二「音楽情報処理の最近の研究」, *日本音響学会誌* 60 卷 11 号 (2004), pp.675-681
- [3] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一 "RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース", *日本音響学会 2002 年春季研究発表会 講演論文集*, 2-6-7, pp.705-706, March 2002.