

ハミングからの階名と音価の推定

Estimation of syllable name and tone values from Humming voice

清水 純† Jun Shimizu 丸山 剛志† Tsuyoshi Maruyama
三浦雅展‡ Masanobu Miura 柳田益造† Masuzo Yanagida

1. まえがき

携帯電話端末からメロディーを音符で入力すると、それに和音を付与するシステムを開発した¹⁾が、携帯電話端末からの音符入力のインターフェースが貧弱であるため、このシステムは「面白いけれども非常に使いにくいもの」となっている。本研究はこの状況を解決するべく、携帯電話端末から簡単にメロディーを入力できるようにすることを目的としている。

これまでに、自動採譜に関しては知識処理的アプローチによるピアノ・ギター・三味線の自動採譜²⁾、クラスタリング手法を用いたバイオリン二重奏の採譜³⁾、弦楽四重奏の採譜⁴⁾などがあるが、歌唱の採譜については、その困難さの故、性能のあまりよくない市販品⁵⁾がある程度で、本格的な研究は見られない。

本研究では、ハミングで歌った音声に対し、音価の階層性と音階の構造に基づくテンプレートマッチングによって、音価・階名の推定を行い、音符情報として出力することを目標としている。

2. 音価と階名の推定方法

2.1 音価の推定

2.1.1 音価の推定方法

入力されたハミングによる歌の、各音の立上りを検出し、音の立上りから次の音の立上りまでを求め、各音の時間長とする。そして、音の時間長の頻度分布より、歌の中で最も短い音符の時間長を推定し、それを音価の基準の時間長として、各音の音価を推定する。

2.1.2 音の立上りの検出

時間軸上で、基本周波数の値が連続的に近い値をとるとき、それは一つの音であると判定する。また、音と音の間に無音の状態が生じることがあるが、これは duty ratio が小さいことによるもので、特に長い無音でない限り、譜面上では前の音が続いていると見なす。

また、基本周波数抽出には、ケプストラム法を用いる。

2.1.3 基準となる音符の長さの推定

全ての音の時間長を求め、頻度分布をとると、歌の中で最も短い音符の時間長のところで最初のピークが存在し、また、その時間長のほぼ整数倍、あるいは2のべき乗倍のところでピークが存在することになる。つまり、基準となる音符の時間長を推定するには、最初に現れるピークを与える時間長を求めればよい。

そこで、図1のような、音の時間長の頻度分布を評価するための関数 $T_t(t, \tau)$ を定義する。 $T_t(t, \tau)$ は t と τ の関数であるが、 $t = 2^n \tau$ で 1、 $t = k\tau$ (k は 2^n 以外の整数) で 0.5、 $t \neq k\tau$ (k : 整数) で 0 をとる関数である。

† 同志社大学 工学部

‡ 龍谷大学 理工学部

$$T_t(t, \tau) = \begin{cases} 1 & \text{for } t = 2^n \tau \quad (n = 0, 1, 2, \dots) \\ 0.5 & \text{for } t = k\tau \quad (k \text{ は } 2^n \text{ 以外の整数}) \\ 0 & \text{for } t \neq k\tau \quad (k: \text{整数}) \end{cases} \quad (1)$$

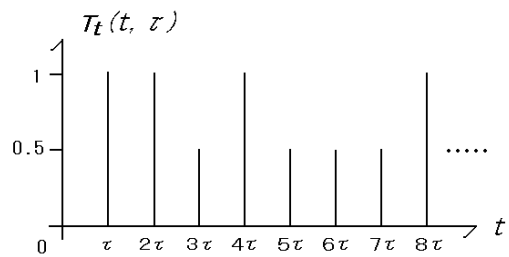


図1 音の時間長の頻度分布を評価するための関数 $T_t(t, \tau)$

この関数 $T_t(t, \tau)$ と音の長さの頻度分布 $F(t)$ との時間軸上の適合度 $C_t(\tau)$ を以下の計算式により求め、 $C_t(\tau)$ に最大値を与える $\tau = \tau_0$ を、基準となる音符の時間長であるとみなす。

$$C_t(\tau) = \int T_t(t, \tau) \cdot F(t) dt \quad (2)$$

$$C_t(\tau_0) = \max C_t(\tau) \quad (3)$$

2.1.4 各音の音価の推定

各音の時間長が τ_0 の何倍かを求める。この値は2のべき乗値、あるいは整数値に近い値であることが期待される。本システムでは、基準となる音価を8分音符とするか、16分音符、あるいは32分音符とするかは別に決める。

2.2 階名の推定

2.2.1 推定方法

一つの音と判定される時間領域内での周波数の平均値を求め、それがどの階名の基準周波数に最も近いかを調べることで、階名の推定を行う。ただし、音階全体は対数的に浮動するが、各階名間の周波数比は平均律に保たれると仮定する。また、本システムでは、階名の推定を半音単位で行う。

人が歌う場合、絶対音感を持たない人は、譜面上の調で歌うことは困難であるため、相対的に正確に歌えていれば、絶対的な音の高さは正規の高さでなくても正確に歌えていると考えるべきであり、絶対的な音の高さは特に考慮する必要はないし、それをユーザーに要請することは不可能である。そこで、相対的に正確に歌えている歌について、階名の推定を行うことを考える。このとき、音階の基準となる周波数を絶対的な位置に決めてしまうのではなく、基本周波数の頻度分布をもとに音階の基準周波数を定め、各音の階名を推定する方法をとる。

2.2.2 音階の基準周波数の決定

相対的に正確に歌えている歌について、基本周波数を抽出し、頻度分布をみると、全音、または、半音間隔で強い分布を示すことが期待される。そこで、この頻度分布に対して、以下のような、基本周波数の頻度分布を評価するための関数 $T_f(f, f_c)$ を定義する。

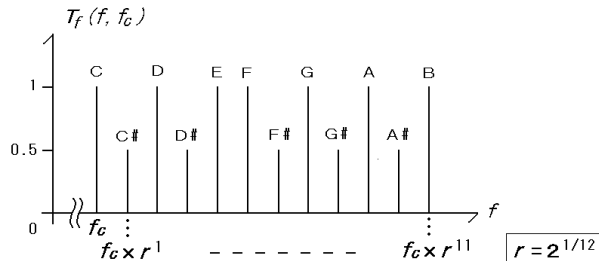


図2 基本周波数の頻度分布を評価するための関数 $T_f(f, f_c)$

図2に示すようにピアノの白鍵に対応する相対音高で1、黒鍵に対応する相対音高で0.5、その他で0となる関数 $T_f(f, f_c)$ と基本周波数の頻度分布 $F_f(f)$ との周波数軸上の適合度 $C_f(f_c)$ を以下の計算式により求め、 $C_f(f_c)$ に最大値を与える $f_c = f_{c0}$ をその歌唱でのドの基準周波数と定める。

$$C_f(f_c) = \int T_f(f, f_c) \cdot F_f(f) df \quad (4)$$

$$C_f(f_{c0}) = \max C_f(f_c) \quad (5)$$

同様に、 f_{c0} に2の1/12乗をかける毎に平均律で半音ずつ上がるので、ド#、レ、レ#、ミ、ファ、...と、全ての階名の基準周波数を定めることができる。

2.2.3 各音の階名の推定

各音のローカルな平均周波数が、どの階名の基準周波数に対数的に最も近いかを調べ、階名を推定する。

3. 評価実験

3.1 動作確認1

アルゴリズムの動作確認のために、MIDI 打ち込みの出力音（音色はピアノを使用）を分析したところ、音価・階名とも100%の推定結果を得た。（分析結果に基づく再生は、元の音と全く同一であった。）

3.2 動作確認2

$A_4 = 430 \text{ Hz}$ として調弦したギターで、童謡の旋律を弾き、それをマイクで拾い、分析したところ、音価の認識率は89.2%、階名の認識率は98.3%であった。

3.3 動作確認3

3.3.1 動作確認3の概要

ハミングに対して、各音の音価・階名がどれだけ正確に推定できたかを比較実験により調べる。

5人の被験者に対して、簡単な7曲の童謡(最初の4小節)を歌ってもらい、それらに対するシステムの音価の認識率、階名の認識率を調べる。

3.3.2 留意点

比較実験を行うにあたって、留意しなければならないことが4点ある。それは、(1)被験者が音符を間違えて記憶している場合があること。例えば、8分音符が2つ続けて存

在するところを、4分音符1つと思い込んでいる場合がある。(2)譜面上の調と、被験者が歌う歌の調が違う場合が多いこと。(3)テンポが途中で変わってしまう可能性があること。(4)被験者は必ずしも相対的に正確に歌うことができないこと。

従って、譜面と本システムの出力を比較することは、正しい比較にはならない。しかし、被験者が思い描いている曲と、本システムの出力を比較することは不可能であるため、以下の点を許容し、譜面との比較を行う。

(1)同じ階名の音符が続けて存在する場合、本システムがそれらの音符を結合した音価を出力しても、正しく推定できていると評価する。逆も同様である。(2)階名の評価においては、絶対的な音の高さが正規の高さでなくても、相対的に正確であれば、正しく推定できていると評価する。(3)メトロノームを用い、一定のテンポで歌ってもらう。

3.3.3 動作確認3の結果

音価の認識率は92.9%、階名の認識率は76.4%であった。しかし、階名の推定において、誤認識とみなされた音の中には、3.3.2節留意点の(4)で述べたように、システムの誤認識ではなく、被験者が正確に歌えなかった音も含まれている。

4. まとめと今後の課題

前節の評価実験の結果より、相対的に正確な演奏については、精度の良い推定が行われていることが分かる。今後の課題としては、相対的に正確ではない、人のハミングによる歌に対して、どのように補正を加えれば、歌った本人が思い描いている曲に近い出力を行えるようになるかを考察しなければならない。

以上の内容を踏まえて、今後の具体的な課題を述べる。

(1)音階の基準周波数の決定において、入力された歌全体に対して音階の基準周波数を決定したが、歌の途中で補正しながら、音階の基準周波数を変更していく方法を検討している。(2)メトロノームを用いて評価実験を行ったが、システム自体にテンポを取りながら歌を入力できるような機能を導入することを検討している。(3)基本周波数の抽出にケプストラム法を採用したが、基本周波数の抽出の精度を上げるために、他の基本周波数抽出法(SIFTアルゴリズムなど)の採用を検討している。

本研究は、同志社大学学術フロンティア事業の援助を受けた。

参考文献

- [1] 三浦雅展, 柳田益造他: “ポップス系の旋律に対する和声付与システム AMOR”, 日本音響学会 音楽音響研究会 MA2003-7(2003).
- [2] 片寄晴弘, 井口征士: “知的採譜システム”, 人工知能学会誌, Vol.5, No.1, pp.59-66(1990).
- [3] 小林政義, 柳田益造: “時間連動性を考慮した多重調波構造モデルによる多重奏音の自動採譜に関する検討”, 日本音響学会 音楽音響研究会 MA2003-41(2004).
- [4] Shirado, T. and Yanagida, M.: “Extraction of Fundamental frequencies from the sound of a violin duet”, Proc. of the Joint meeting of ASA/ASJ, pp.291-396(1996)
- [5] 鼻歌ミュージシャン2, (株)メディア・ナビゲーション