

母音認識とピッチ検出による歌声のテンポ抽出

7J-9

東 英司 尾上 直之 橋本 周司

早稲田大学理工学部

1 はじめに

我々はこれまで歌声のテンポに合わせて演奏する自動伴奏システムを試作してきた。そこでは歌声のテンポを抽出するために、ピッチの検出結果とメロディデータとをマッチングする手法 [1] と母音の認識結果と歌詞データとをマッチングする手法 [2] という2通りの手法を扱っていた。しかし、同じ母音やピッチが連続して現れる箇所では、テンポ追従が困難であった。

そこで現在は、母音とピッチの情報を複合して用いることにより、細かい歌唱位置の判定を目指している。母音とピッチの認識にはDSPを使わずMacintosh内蔵のSound-Input-Deviceを使用し、主にケプストラム法を用いて行っている。特にピッチ検出では歌声の基本周波数の倍音構造に注目した手法を用いることでローパスフィルタも不要になった。これらにより、マイクとMIDI音源さえあれば手軽に自動伴奏が行えるという見通しを得たので報告する。

2 システムの構成

システムの概要を図1に示す。前奏に続いて歌声をマイクから入力しSound-Input-Deviceで計算機へ入力する。その後ケプストラム法により解析し、歌声の母音とピッチを得る。これを楽譜情報とマッチングさせることで歌唱位置を判定、歌唱イベントの時間間隔からテンポを予測し、MIDI音源に送信する。

3 アルゴリズム

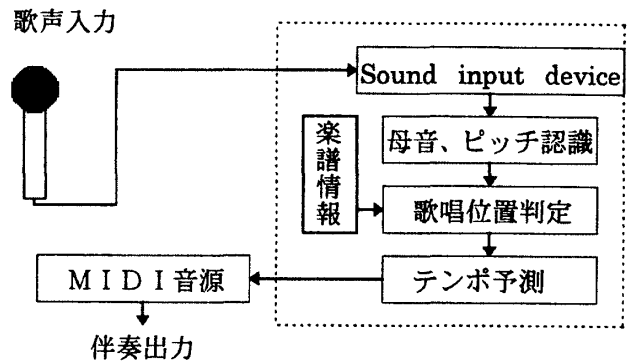


図1 システムの構成

3. 1 母音認識とピッチ検出

3. 1. 1 母音認識

まず歌い手の5つの母音を入力し、ケプストラム法によりスペクトル包絡を取り、標準パターンとする。歌唱時にこれらと随時マッチングを行い、母音を認識する。

3. 1. 2 ピッチ検出

従来はpeak-to-peak法やMIDI-Converterなどでピッチ検出を行ってきたが、ここではケプストラム法を用いる [3]。しかし、高次ケプストラムからフーリエ変換して得られる倍音構造の中の基本周波数のピークでは、周波数分解能が悪く、半音を区別するだけのピッチ検出が不可能である。そこで式1を十分満たすようなn、つまり第n倍音の周波数を測定しnで割ることによりピッチ検出を行うことにした (式2)。

$$n > \Delta f / \Delta s \quad \dots\dots \text{(式1)}$$

$$f_1 = f_n / n \quad (n=2, 3, \dots) \quad \dots\dots \text{(式2)}$$

Δf : 周波数分解能 Δs : 半音の周波数差

Δf_1 : 基本周波数 Δf_n : 第n倍音の周波数

3. 2 歌詞とのマッチング

自動伴奏を実現するためには歌い手がどこを歌っているかを知る必要がある。そこで楽譜情報と約

25ms 毎に得られる母音、ピッチとをマッチングする。この時、ピッチと母音のいずれかが楽譜情報と一致したときに歌われたと判断する。但し楽譜上の母音に変化がない部分はピッチで、ピッチに変化がない部分や半音階進行の部分などは母音でマッチングする(図2)。

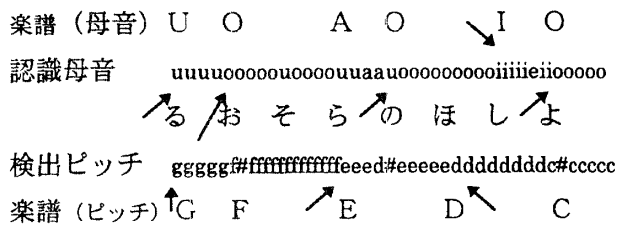


図2 歌詞とのマッチングの例

3.3 テンポ予測

伴奏を歌声に追従させる為にテンポを予測する。歌唱イベントの時間間隔により歌手のテンポを計算し、引き続きそのテンポで歌われると仮定し、次の歌唱タイミングに伴奏が合うよう、テンポを求めている。第*i*+1番目の伴奏テンポ*a_{i+1}*は、下式で表される。

$$a_{i+1} = \frac{60 N_{i+1} - a_i T_i}{A_{i+1}} \quad (i \geq 1)$$

$$A_{i+1} = \frac{N_{i+1}}{N_i} \times B_i$$

$$A_1 = 60 N_1 / a_1$$

$$T_i = B_i - A_i$$

$$b_i = 60 N_i / B_i$$

- A_i : 第*i*番目の伴奏予定時間 (s)
- B_i : 第*i*番目の音符の発声時間 (s)
- a_i : 第*i*番目の伴奏のテンポ
- b_i : 第*i*~*i*+1番目の音符の間のテンポ
- T_i : 第*i*+1番目の音符の伴奏タイミングと歌唱認識のタイミングの差 (s)
- N_i : 第*i*番目の音符の拍数

4 実験

サンプリング周波数 11,025 (kHz)、256点サンプリングで歌声を処理した。母音とピッチは約25msに1回認識される。テンポ100の4分

音符が600msであるから歌声のテンポを抽出する上で十分な速さと言えよう。実験では、被験者に意図的にテンポを揺らして歌ってもらった。テンポの推定結果を図3.1に、その時の歌声と伴奏のずれを図3.2に示す。

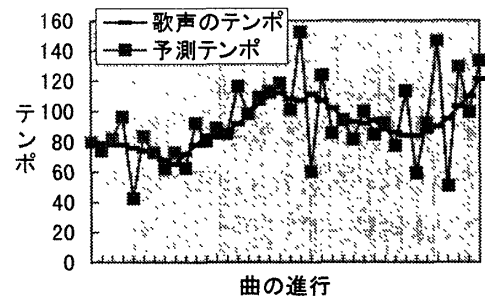


図3.1 テンポ予測

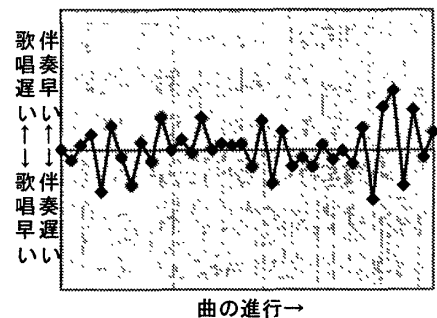


図3.2 位相のずれ

図3.1、図3.2より特に位相のずれの大きい部分は母音とピッチの誤認識やテンポの急な加速、減速によるものと考えられるが、現在検討中である。

5 おわりに

母音とピッチの両方を用いた歌声の自動伴奏システムについて述べた。今後は認識率の向上、テンポ予測方式の改善、不特定話者への対応、子音の認識などを検討して行きたい。

参考文献

- [1] W.Inoe,S.Hashimoto,S.Ohteru,"A Computer Music For Human Singing",ICMC Proc,1993
- [2] 井上,橋本,大照,"適応型歌声自動伴奏システム",情処論,vol.37(1993)
- [3] 鈴木,佐野,島村,八嶋,"調波構造を利用した精密基本周波推定",信学技報,sp93-5,pp.25-32(1993)