母音認識とピッチ検出による歌声のテンポ抽出 2

4M－7　東　英司　橋本 周司
早稲田大学理工学部

1、はじめに

人が伴奏付きで歌を歌う際に「意識的にここのフレーズはゆっくり歌いたい」などの意志を反映するよう人の歌のテンポに合わせて伴奏を出力する自動伴奏システムについて述べる。以前から様々な自動伴奏システムの研究について報告がなされている[1][2]が、メロディが歌声などのいわゆる非電子音の場合、一般に歌唱位置やテンポを細かく抽出するのは困難であり、種々の手法が提案されている[3][4][5]。そこで、本システムにおいてはテンポ抽出の手法としてケプストラム法を利用し、歌い手の発声するピッチと母音を実時間で同時に獲得することで楽譜情報から歌唱位置、テンポを割り出すことを検討している[6][7]。その際、ピッチ検出には倍音構造を用いた比較的小さな窓長でも正確に検出できる方法を用いている。更に母音認識率向上の手段として複数の標準スペクトル包絡からのマッチングを行った。又、歌声の入力をパソコンに標準的に内蔵されている音声入力デバイスで行っている為、ソフトウェアだけでの構成が可能になった。ここではシステムの概要と音声認識の改良について報告する。

2、システム構成

システムの概要を図1に示す。本システムは歌唱者のメロディからテンポを割り出し、歌唱者のテンポに合わせた追従伴奏を実現していく。まず、音声の入力を計算機（Macintosh 本体）に標準的に内蔵されている"Sound-Input-Device"で行い、AD 変換する。そのデジタルデータに対しケプストラム法を適用することにより、リアルタイムでの母音認識とピッチ検出を行う。この歌唱位置情報（母音、ピッチ）と拍情報からテンポ抽出する。そのテンポを用いてMIDI音源により伴奏を出力することで人のテンポに合わせた伴奏が可能になる。

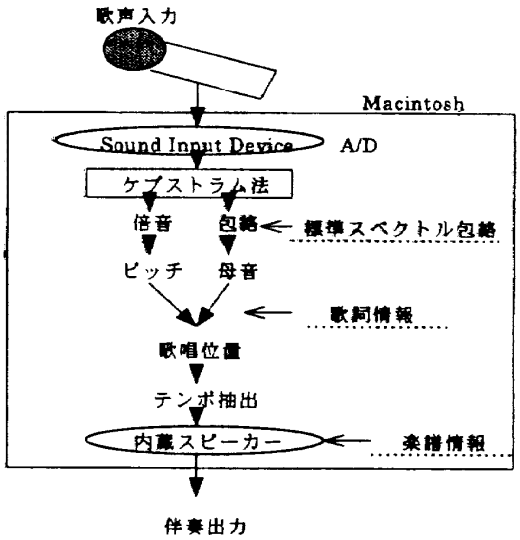

図1　システム概要

3、アルゴリズム

3.1　ピッチ抽出

一般的なピッチ抽出方法では長いフレームやLow-Pass-Filterなどのハードウェアを必要とする。しかし精密な実時間ピッチ抽出を考えた場合、処理に適さない。そこでフレームを短くした時でも高調波成分を着目する手法を扱うことで、高い周波数分解能でのピッチ抽出を可能にした。つまり基音の第N倍音周波数をNで割ることで基本周波数を推定した。実際には精度向上のため特定の倍音のみで推定せず、周波数軸上の雑音の影響が少ないと思われる領域の倍音に対し、重みをつけて基本周波数の推定を行うことにした（式1）。尚、パワーが小さいフレームについては無音もしくは子音の部分と判断し、

The Extraction of Singing tempo by Vowel recognition and Pitch detection 2,
Eiji Azuma, Shuji Hashimoto,
School of Science and Engineering, Waseda University

母音認識、ピッチ検出の対象としない。本システムでは歌唱者が男性の場合、スペクトルの周波数刻みは約43.1Hz（フレーム長約23.2ms）とし、N＝12 とした。したがって、C(48) からF(65) まで半音の区別ができる。女性歌唱者の場合は周波数刻みは約86.1Hz、フレーム長約11.6ms となり、C(60)～F(77) の判別が可能である。

$$f_1 = \sum_{k=1}^{N} f_k \Big/ \sum_{k=1}^{N} k \qquad \cdots \text{（式1）}$$

f_1：基本周波数

f_k：第k倍音目の周波数

3.2 母音認識

汎用性を考えれば不特定話者認識が望まれる所だが、本システムでは現段階、認識率を重視し特定話者認識を採用した。つまりいかなる音程についてもスペクトル包絡が保存されているとし、スペクトル包絡の平均と分散を標準スペクトル包絡パターンとしてマッチングさせればよい。しかし、実際にスペクトル包絡は入力時のピッチに少なからず依存してしまう。そこで、まず歌う曲の音域内の異なる音程で1母音につき複数の標準スペクトル包絡をあらかじめ用意しておく。そして実際の自動伴奏中に、予想されるピッチの近傍の標準パターンを優先しマッチングをすることで母音認識を行う。

3.3 歌詞位置推定とテンポ抽出

歌詞情報（つまりここでは楽譜上のピッチと母音情報）と歌われたピッチ、母音とで随時マッチングを行う。この場合、いづれかが発声された時にその歌詞（音符）が歌われたと判断する。又このシステムでは音程が一定となるフレーズにおいては母音認識のみを用いることでトラッキングが可能であり、逆に同じ母音の続くフレーズではピッチで推定できる。これらのトラッキングから歌唱のテンポを計算する。伴奏部はMIDIで構成され、楽譜情報としてMIDI ファイル、歌詞情報として楽譜のピッチ、母音、拍の長さを持つ。

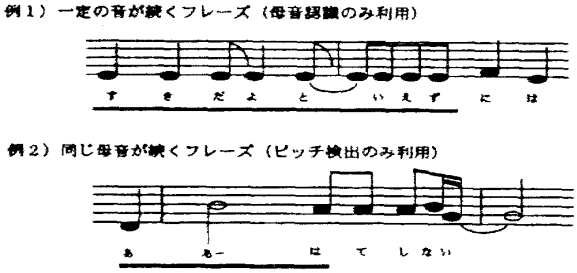

図2　歌詞位置推定の例

4、 実験

11.025kHz サンプリング、量子化ビット数16bit、1フレームあたり256点でのFFTにおいて、標準スペクトル包絡パターンの数を変えて男声の自動伴奏を行った。以下に図3に結果を示す。

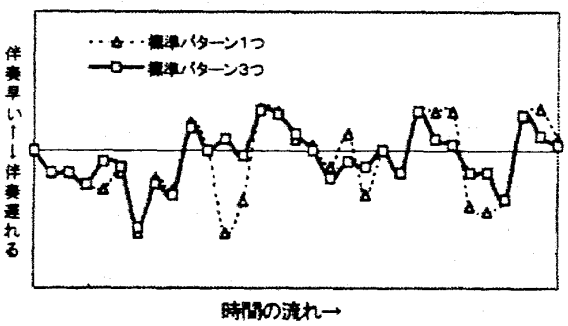

図3 曲と伴奏のずれ

5、 おわりに

限られた実時間処理の中で、テンポ抽出がソフトウェア上で可能になり、自動伴奏に適用できた。今後は不特定話者への対応、音声補正機能の導入を検討する予定である。

参考文献

[1] Dannenberg,RB. and Mont-Reynand,B. "Following an Improvisation in Real-Time", Proc.of ICMC, pp241-pp.248 (1987)

[2] 直井、大照、橋本、"実時間拍検出機能を用いた伴奏システム"、日本音響学会講演論文集、pp.465-pp.466 (March 1989)

[3] Katayose, H., Kanamori, T., Kame, K., Nagashima, Y., Sato, K., Inokuhci, S. and Simura,S. "Virtual Performer ", Proc.of ICMC, pp138-pp.145 (1993)

[4] 井上、橋本、大照、"適応型歌声自動伴奏システム"、情報処理学会論文誌、vol.37 pp.31-pp.38 (1996)

[5] Grubb,L., Dannenberg,R. "A Stochastic Method of Tracking a Vocal Performer", Proc.of ICMC, pp301-pp.308 (1997)

[6] 東、尾上、橋本、"母音認識とピッチ検出による歌声のテンポ抽出" 情報処理学会第54回全国大会講演論文集(2)、pp.283-pp.284 (1997)

[7] 東、橋本、"音声認識とピッチ検出を併用した歌声の自動伴奏" 情報処理処理学会 音楽情報科学97-MUS-22 pp.1-pp.5 (1997)