音声認識とピッチ検出による歌声のトラッキング

3G－1

東　英司　　橋本　周司
早稲田大学理工学部

1　はじめに

我々は人の歌声と機械の伴奏との自動伴奏についての研究を進めてきた。この自動伴奏を実現するにはまず人が歌っている場所をシステム側が実時間で認識する必要がある。従来の自動伴奏の研究では、ピッチを手がかりとしてトラッキングをする場合が多い[1][2][3]。この場合、認識率や音の立ち上がりの悪さなどにより、リアルタイムでのトラッキングは容易でない。そこでピッチ検出の他に歌声の母音認識[4][5]と音量測定を行うことで、より効果的なトラッキングを実現した。その後、これらをMax上で構成した[6]。ここではこれを改良したトラッキング法による歌唱タイミング解析の Max パッチと実験の概要について紹介する。

2　トラッキングシステム概要

トラッキングの過程を図1に示す。

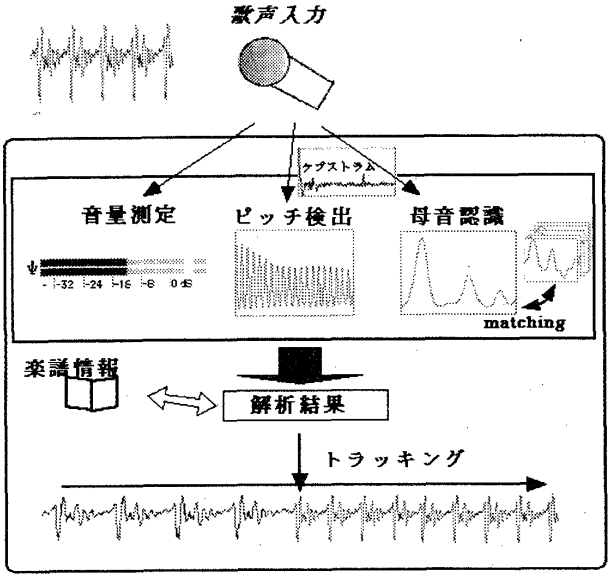

図1　トラッキングまでの流れ

Singing Tracking Using Vowel Recognition and Pitch Detection
Eiji Azuma, Shuji Hashimoto
School of Science and Engineering,
Waseda University

まず歌声を AD 変換した後、音量・ピッチ・母音を抽出し楽譜情報とのマッチングによりタイミング解析を行う。ピッチ、母音の検出には主にケプストラム法を用いている。

3　歌声のトラッキング

3-1　歌声の解析

(VowelPitchVolume オブジェクト)

3-1-1　音量測定

音量を測定(式1)し、休符後の音の立ち上がりによる歌声トラッキングや音声区間の抽出などを行う。

$$Volume(\mathrm{dB}) = 10\log_{10}\left(\frac{1}{N}\sum_{n=0}^{N-1}x(t+n\Delta t)\Big/\frac{A^2}{2}\right) \quad (式1)$$

N：　1フレーム中のサンプル数

A：　最大振幅

3-1-2　ピッチ検出

高域ケプストラムをフリーエ変換して得られる倍音構造から基本周波数を同定する。この場合、周波数分解能が大きいため第1ピークで基本周波数を特定するのは困難である。そこで第 k ピーク(k>1)周波数を k で割った値を基本周波数とし、標本化による誤差を減らした。ここでは$1 \le k \le 12 (= p)$間のすべてのピークに対し重みをつけて基本周波数を同定した(式2)。

$$f_1 \approx \sum_{k=1}^{p} f_k \Big/ \sum_{k=1}^{p} k \quad (式2)$$

3-1-3　母音認識

低域ケプストラムをフーリエ変換したスペクトル包絡をマッチングすることで母音を認識する。そのために、トラッキング前にあらかじめ歌い手に1秒程度発声してもらい5つの母音(i =0～4)のスペクトル包絡平均と分散を計算する。そしてフレーム毎に得られるスペクトル包絡とその標準パターンとを式3を使ってマッチングする。式3の左辺の値が大きいものほど類似度が高いと言える。

$$B_i = \sum_{j=0}^{N-1} \left[-\frac{\log(2\pi\sigma_{ij})}{2} - \frac{(y_j - \alpha_{ij})^2}{2\sigma_{ij}^2} \right] \quad (式3)$$

以上の解析はすべて VowelPitchVolume オブジェクト(サンプリング周波数 11.025(kHz)、量子化ビット数 16(bits)、フーリエ時間窓約 23.2(ms))で行われる（図2)。

3-2 歌唱位置の決定(SingTrack パッチ)

フレーム毎に得られる歌声の解析結果(音量・ピッチ・母音)を楽譜情報と照合する。楽譜情報とは曲の音程・母音(歌詞)・音長(休符も含む)を指す。歌唱イベントが予想される付近において、休符の後の急激な音量の増加、ピッチと母音のいずれかの検出などがあった場合、「歌われた」と判断する。尚、この楽譜情報はファイルとして作成、保存が容易にできる。歌声トラッキングの Max パッチを図2、トラッキングの例を図3に示す。

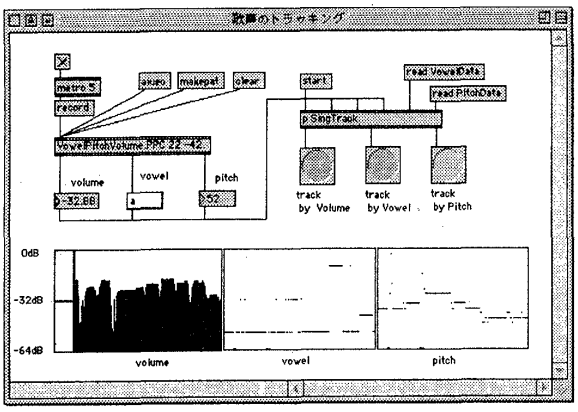

図2　歌声トラッキングパッチ

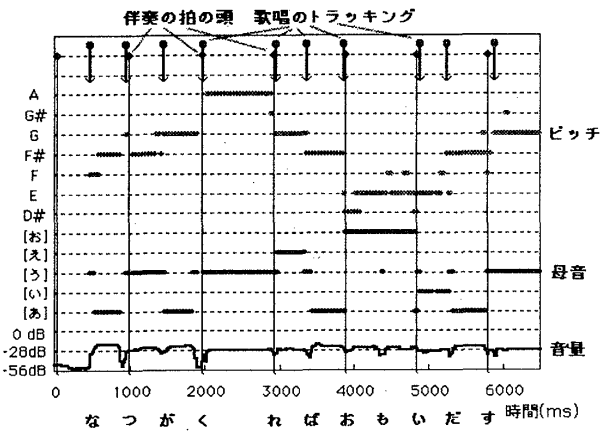

図3　トラッキングの例

4　歌声トラッキングの実験

伴奏テンポを変化させた時の歌声の伴奏との同期の様子を解析する。「ずれ」とはこの場合、歌声の拍の頭でトラッキングされた箇所と伴奏の拍の頭のずれを指す。テンポ変化の影響によるずれが生じてもその後修復している様子がわかる(図4)。

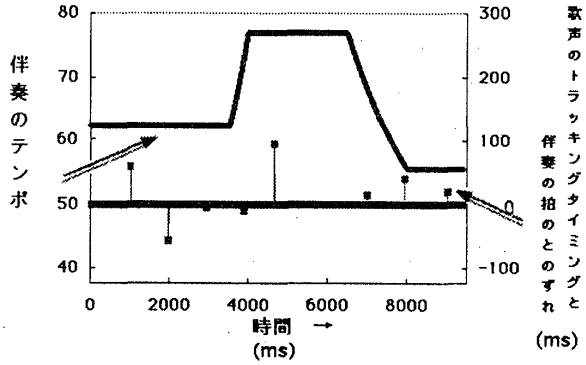

図4　伴奏テンポの変化による歌声の同期

5　自動伴奏への応用

この歌声のトラッキング手法を用いて歌声の自動伴奏を試みている。Max 上での MIDI 出力(seq オブジェクト)を伴奏に用い、伴奏と歌声の同期制御と歌声のテンポ予測の2つを行うことでテンポ変化に対応していく。これにより本システムではマルチモーダルな自動伴奏が可能になるばかりでなく、伴奏との相互作用の様子を定量的に観察することができる。また歌声の自動伴奏だけでなく、マウスクリック、手拍子、指揮棒などによる自動伴奏なども実現可能になっている。

6　おわりに

母音認識などの複数の手がかりを用いた歌声のトラッキングについて紹介した。本手法はソフトウェアだけで構成されており、Max 上での動作により MIDI との融合も容易である。

参考文献

[1] Katayose,H., Kanamori,T., Kame,K., Nagashima,Y., Sato,K., Inokuchi,S. and Shimura,S. "Virtual Performer", Proc.of ICMC, pp138-145(1993)

[2] Grubb,L. and Dannenberg,R. "A Stochastic Method of Tracking a Vocal Perfomer", Proc.of ICMC, pp.301-pp.308 (1997)

[3] Horiuchi,Y. and Tanaka,H. "A Computer Accompaiment System With Independence", Proc.of ICMC, pp.418-420(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] 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" 情報処理学会 97-MUS-22 pp.1-pp.5(1997)

[6] 東、橋本、"母音認識とピッチ検出を用いた歌声のテンポ抽出第3報" 情報処理学会 98-MUS-26 pp.17-pp.22 (1998)