

音声認識を導入した歌声自動伴奏

4T-2

井上 渉 橋本 周司 大照 完

早稲田大学 理工学部

1 はじめに

音楽の中でのヒューマンインターフェースの研究の一つとして、歌唱や演奏に合わせる自動伴奏システムの開発が試みられている[Vercoe 1984][Dannenberg 1984][Katayose et al. 1993]。筆者らも既に、歌唱に対する自動伴奏システムと歌声の音程の自動補正システムについて報告している[Inoue et al. 1993]。自動伴奏システムは、歌声の音程と主旋律の楽譜を比較しながら、歌唱テンポを検出し伴奏をつける。また音程補正システムでは、歌声と楽譜の音程のズレを検出し、歌声の音程を補正することにより、正しい音程の歌声を実時間で出力する。従来の自動伴奏システムでは、テンポは自由に歌うことができるが、歌声の音程が大きくはずれた場合、歌唱テンポに追従して伴奏することは困難である。また音程補正システムでは、音程をはずして歌うことができるが、歌唱テンポに追従した伴奏は考慮に入っていない。

そこで、音程をはずして歌っているときにも適応的な自動伴奏ができるシステムの開発を試みた。ここでは、簡単な音声認識を行うことにより演奏情報（歌唱位置、テンポ）を抽出し、これに合わせた伴奏を出力する。また同時に音程の自動補正も行うことができる。

唱位置、テンポを求め、歌唱テンポに合わせた伴奏をする。また音程補正を行うために、Pitch to MIDI Converter（音程-MIDI信号変換器）で、歌声の音程を求める。検出した音程と主旋律の楽譜から、正しい音程とのズレを求め、この情報をMIDI信号でエフェクターに送信する。エフェクターは受信した情報によりピッチチェンジを行い、音程を補正した歌声をスピーカーから出力する。

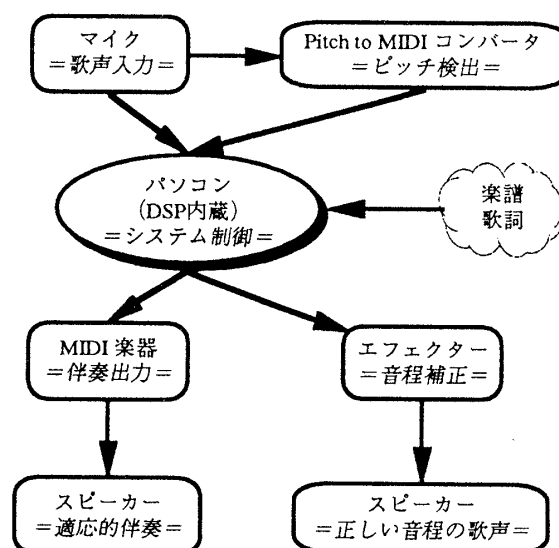


図1 システム構成

2 システム概要

本システムのハードウェア構成は、システム制御用のパーソナルコンピュータ、音声認識用のデジタル信号処理ボード（DSP）、音程検出用の Pitch to MIDI コンバータ、音程補正用のエフェクター、および伴奏出力用のMIDI音源からなっており、曲に関する知識としては主旋律（歌）と伴奏の楽譜、歌詞を持っている。

図1にシステムの基本構成を示す。まず自動伴奏を行うために、マイクからの入力信号をDSPで処理し、音声認識を行う。認識結果と歌詞情報から、歌

3 歌声の解析

3.1 音声認識

まず、各母音（あいうえお）の標準パターンの作成を行う。ここでは、スペクトル包絡線を各母音の標準パターンとしている。包絡線は図2に示すケプストラム法により求めているが、高速なデジタル信号処理が可能なDSPを用いて実時間処理を行う。

認識部では各母音の標準パターンとのマッチングをとり、どの母音であるかを判断する。認識結果としては、各母音（5種）、母音以外、音声入力なしの7種類があり、これをシステム制御部へ送る。

An Automated Accompaniment System for Human Singing Based on Speech Recognition

Wataru INOUE, Shuji HASHIMOTO, Sadamu OHTERU

Waseda University

3-4-1 Okubo, Shinjuku - ku, Tokyo 169, JAPAN

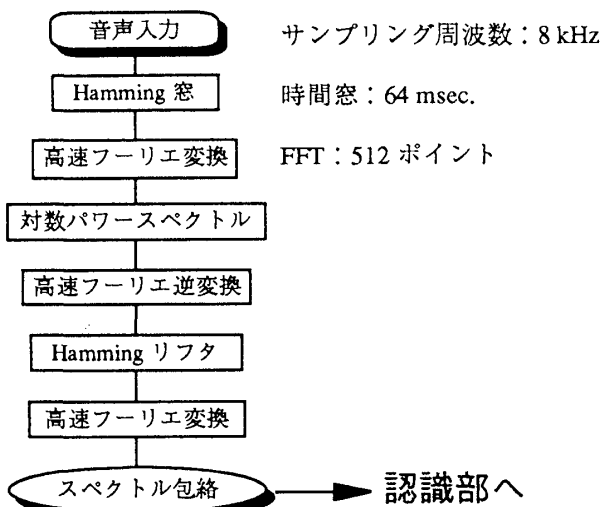


図2 ケプストラム法

3.2 ピッチ検出

マイクから入力された歌声の音程を Pitch to MIDI Converter (Roland: CP-40) により、実時間でMIDI信号に変換する。音程は半音単位で検出可能で、MIDI信号でシステム制御部へ送られる。検出精度を向上させるため、マイクからの入力信号をフィルタに通し高調波を除去した後、Pitch to MIDI Converterへ入力している。

4 システム制御部

4.1 自動伴奏

音声認識の結果とあらかじめ持っている歌詞情報とのマッチングをとり、演奏情報（歌唱位置、テンポ）を得る。演奏情報が得られるたびに、次のテンポを予測し、伴奏側のスケジューリングを行う。これにしたがって伴奏をMIDI音源により出力、人間の歌唱と協調した自動伴奏を実現する。伴奏側のテンポは、人間と機械の相互作用モデル [井川他1990] を用いて決定している。

4.2 音程補正

Pitch to MIDI Converterで検出した歌声の音程と主旋律の楽譜から、正しい音程とのズレを求め、MIDI信号でエフェクターに送信する。音程補正はエフェクター（ヤマハ: Digital Sound Processor SPX90II）のピッチチェンジプログラムを使用している。音程補正の量は、“BASE KEY”パラメータで指定した音名と、MIDI信号で入力された音名との音階差で決められる。音程は±1オクターブ（±12半音）の範囲で可変である。ピッチチェンジにより

声質（音色）の変化が生じるので、フィルターを用い、十分ではないが簡単な補正を行っている。

5 実験結果

図3に音声認識部の結果を示す。また図4にテンポ追従の様子を示す。図4の横軸の時間は曲の進行を表し、縦軸の最小音長時間は最小音長音符の持続時間を表す。最小音長音符とはシステムで扱う最も短い音符のことで、ここでは32分音符としている。この値が大きくなるとテンポが遅くなり、小さくなるとテンポが速くなることを意味する。図を見てわかるように、歌唱と伴奏のテンポに大きなズレはなく、歌唱テンポに伴奏が追従している。

歌詞情報: k i r a k i r a h i k a r u
 認識結果: --XXiiuieaaaXXiiXaaXXiiiXXXaaaaouuuu
 (×: 母音以外、-: 音声入力なし) → 時間

図3 実験結果-1

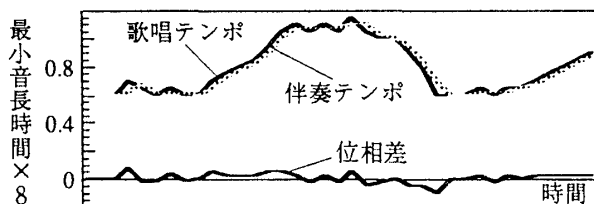


図4 実験結果-2

6 おわりに

音声認識を導入することにより、音程をはずして歌っている場合にも歌詞から歌唱位置を知ることができるようになり、歌唱テンポに合わせた自動伴奏が可能となった。また同時に、歌唱の音程補正も行うことによって、より適応的な自動伴奏システムとした。

参考文献

[Vercoe 1984] Vercoe, B., "The Synthetic Performer in the Context of Live Performance," Proceedings of International Computer Music Conference, 1984.
 [Dannenberg 1984] Dannenberg, R. B., "An On-Line Algorithm for Real-Time Accompaniment," Proc. of ICMC, 1984.
 [Katayose et al. 1993] Katayose, H. et al., "Virtual Performer," Proc. of ICMC, 1993.
 [Inoue et al. 1993] Inoue, W., Hashimoto, S., and Ohteru, S., "A Computer Music System for Human Singing," Proc. of ICMC, 1993.
 [井川他1990] 井川、直井、大照、橋本、"相互作用モデルによる実時間適応自動伴奏とその動作解析。" 電通春期全体、D-464、1990。