歌唱トレーニングシステムVSGの改良

片寄 晴弘* 平井 重行* 村尾忠廣** 金森 務* 井口 征士*
*イメージ情報科学研究所 **愛知教育大学
katayose@image-lab.or.jp

我々は，ビジュアルフィードバックと各種診断機能により，ゲーム間感覚で音痴（調子外れ）を改善するためのシステムVSG の開発を行ってきた．本稿では，最近改良を行った，1）邦楽等，ピッチが微妙に変化する対象の登録となぞり，2）各種，診断のためのパターン認識処理，3）追従型カラオケを中心にシステムを紹介する．

Some Improvement of Singing Training System VSG

We have been developing Singing Training System VSG based on visual feedback and various diagnose. This paper introduces 1) trace mode for traditional vocal melodies as shown in "hougaku", 2) pattern recognition for singing diagnosis, and 3) adaptive KARAOKE function, which have been recently engaged in.

1．はじめに

日本語で俗に言う音痴（調子外れ）は音楽教育における重要な問題である．最近の研究で，歌の苦手な人でも言語のアクセント制御を行っていることから，トレーニングによって音痴がなおるということが明らかになってきた．我々は，自分の発声状態を視覚的にフィードバックし，さらにゲーム感覚で，歌唱のトレーニングを行うことの出来るVoice Shooting Game（以下，VSG）の開発を行っている[1]．同様のシステムとしては，Howard らが1987 年にSingard[2] というシステムを発表を行っているが，1）歌唱ピッチのグラフィック上での表示，2）目標ピッチの表示，3）区間内での平均ピッチの計算とそれに基づいた評価という機能に限定されている．また，Atari でしか使用できないこととピッチ抽出の不安定さが問題となっていた．その他の関連システムとしてはカラオケにアミューズメント機能を盛り込んだ「セガカラ」が存在するが，歌唱のトレーニングを目指したものではない．

本稿では，システムの概要を紹介した上で，最近開発を行った診断機能，追従型カラオケ機能について述べる．

2．システム概要

図1にシステムの概略を示す．この図に示すようにシステムはピッチ抽出センサとMacintosh コンピュータとマルチメディアビジュアルプログラミング環境MAX およびMAX プログラムから構成される．プログラムについては，処理負荷の関係上，シューティングゲームとトレーニングモード

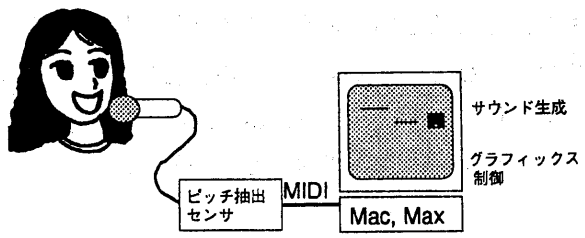

図1．システムの概略

のプログラムを別々に分けて構成している．

初期バージョンで持たせた機能（Singard との差違）としては次のようなことがらがあげられる．

1）目標ピッチの音としての出力と音色の変更
2）シューティングゲームとしての要素の追加
3）ユーザによる簡易な教師データ記述
4）なぞりモード

その後，追加・強化した機能としては

5）邦楽等，ピッチが微妙に変化する対象の登録

となぞり

6）各種，診断のためのパターン認識処理

7）追従型カラオケ

がある．このうち，追従型カラオケについては1993年に発表した別個のシステム[3]を上記フレームワークで実施できるように改良を行ったものである．

2．1ピッチ抽出センサ

リアルタイムのピッチ抽出装置はさまざまなものが開発されており，簡単に入手することができる．しかしながら，安価なものは別途ローパスフィルタを使用する必要があったり，音声を対象とした場合，高周波を含む発音（「い」や「え」の発音）の誤認識が問題になることが多かった．そこで，音声信号を2つに分け，2種類の遮断周波数を持った周波数可変型ローパスフィルタ（男性の標準設定で場合で100Hzと400Hz）に通し，それぞれのゼロクロスインターバルを計算することによってもとめられた周波数のうち，低い周波数を優先して採用するという方式に基づいたハードウエアピッチ抽出センサを開発した．本センサは上記問題に強く，また，比較的安価に実現できるという特徴を有している．なお，本センサからの出力はピッチ（セント単位）と振幅であり，MIDIのポリフォニックプレッシャーの形式で送信している．

2．2 Vocie Shooting Game

Vocie Shooting Gameの画面を図2に示す．Voice Shooting Gameはその名の通り，声でシューティングを行うゲームである．モードとして図2に示したタイプの他に，戦闘機でターゲットを打ち落とすというものもある．ゲームの動作について簡単に説明を行う．ゲームをスタートするとガイド音とともにターゲット（ここでは肉）および雲が画面の右から左へスクロールするようになっている．プレーヤはブタの位置を音程でコントロール

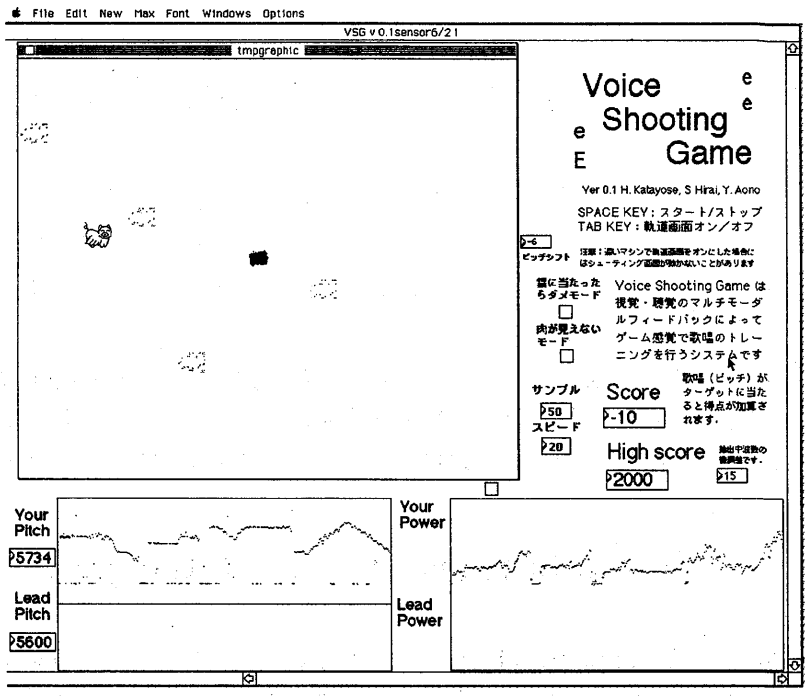

図2．Voice Shooting Gameの画面

して，ターゲットに当たったとき（肉を食べたときに）得点が加算されるようになっている．ターゲットが当たらずに通り過ぎたときには減点される．ここでは，肉を表示するかどうかのモードを選択することができ，聴覚情報にのみに基づいたゲームの使用も可能である．また，プレーヤはMacintoshの内部音源を用いて，各種楽器音を選択することができる．ガイド音の音列情報については，ピッチと音長を一つのセットとするデータ形式をテキストライタ等で容易に作成することが可能である．

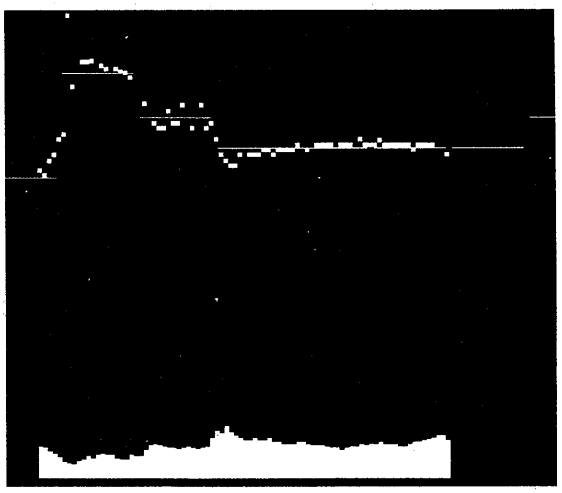

図3．”なぞり”の一例

2．3カウンセリング

音痴（調子外れ）を治して行くための特徴的なVSGの使用法を示す．

0）クライアントにピッチ変化のある言葉を話させて，ピッチの制御が可能なことを納得してもらう．

1）音の表示はまずは，本人の声を使用し，徐々に他の人，人間の声から離れた音を使用する．

2）音の提示時間を変える．

3）ビジュアルフィードバックを徐々にはずす

4）シューティングゲームで遊ぶ

5）ガイド音を徐々になくして，最終的には追従型カラオケ（ピッチがあわないと伴奏データが出ない）を楽しむ．

3．改良

3．1トレースモード

民謡等日本の邦楽においては，各種技法，唱法のほとんどが，伝承という形で引き継がれている．楽譜に明に記載されない部分が表現の根幹となっている部分も多く，効率的な学習法が模索されていた．トレースモードはこのような微妙な表現の学習を支援するためのモードである．

まず，アプリケーションからマッキントッシュ内のサウンドドライバーを用いて手本となるデータの入力を行っておく．そのデータを再生し，ピッチ抽出センサを用いることにより，手本となるデータの音高及び音量データを取得しておく．トレーニング時には手本データを再生すると同時に，音高及び音量を視覚的に提示し，「なぞり」を行っていく．ビブラートやポルタメントの他，演奏上の微妙な表現を視覚を通じて定量的に把握することが可能であり，後述のパターン認識に基づいた診断のほか，独習にも効果が期待されている．

3．2診断機能

開発当初のVSGでは，ピッチとガイド音のズレが上下50cent以内に収まるかどうかという基準で正しい音高か否かの判別を行っていた．今回，より高次での診断を行うために以下のようなパターン判別の機能をインプリメントした．

1）ノートオンイベント

音の開始点でのピッチをノート音イベントとして扱えるようになれば，コンピュータ音楽等幅広い応用が考えられる．歌声を対象とした場合，ポルタメントやビブラートによるピッチの変化を新たな音のイベントと見るかどうかを自由に設定できることが望ましい．ここでは，ある程度の音量が確保されたところで，ノートオンイベントを出力し，音量が十分に下がるまで，あるいはノートオンイベントを出力してからの経過時間がユーザーが設定した時間を越えるまで出力ゲートを閉じるにするという方法を用い，上記機能を実現している．

2）演奏技法

a)ビブラート

発声音のピッチ抽出はハードウエアで行われている．ビブラートについてはそのセンサから出力されたピッチのデータの揺れを計算することにより求められる．方法としては，ピッチ時系列データのPeak To Peak の周期を求めるのとして 微分値のゼロクロス周期を計算している．ビブラートの振幅は，一定間隔内の Peak の振幅という算出している．ビブラートとして認識するにあたり，次の２つの条件を設定している．

ピッチの揺れの周波数１以上の場合

ピッチの揺れの振幅が閾値以上の場合

（閾値はユーザが設定できる．Defaultは20cent である）これらの条件にマッチしない場合，ビブラートとしては認識しない．MAX における具体的な実現例を図４に示す．

b）ポルタメント

認識方法としてはビブラートと同じく簡易なものであり，ピッチ時系列データの微分値を取り，300ms 内の平均値が閾値以内の値であればポルタメントとみなすことにする．閾値はユーザが定義できるようになっている．閾値を，30 - 100 の範囲とすることで比較的安定にポルタメントが検出できることを確認している．MAX における具体的な実現例を図５に示す．

3）テンプレートマッチング

上記のような一括りの演奏技法以外に固有の歌

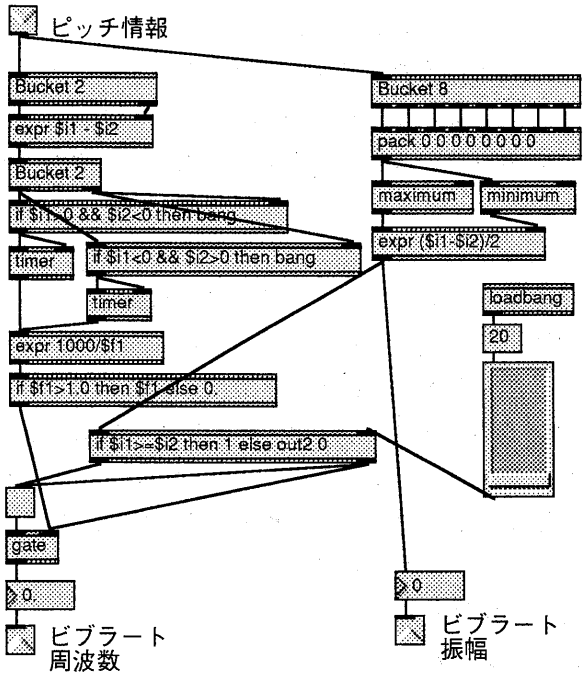

図４．ビブラートの検出

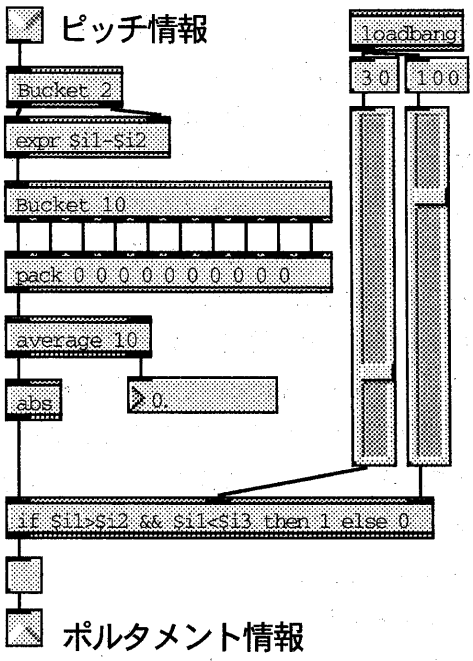

図５．ポルタメントの検出

唱技法の実現度具合（類似性）を計測するために相互相関に着目したテンプレートマッチング機構を用意した．ユーザはMAXプログラム上で，machingというパッチャーを選択し，実際の歌唱例を与えるだけで，テンプレートを作成することができる．ピッチ・音量のデータシーケンスに対し，新たなイベントが入るごとに，それぞれピッチ・音量時系列とテンプレートとの類似性を計算する．マッチングを行う個々のテンプレートは可変長であり，複数のパターンの同時検出が可能である．現在使用している計算機環境(MAX Centris 650)では，30msごとに新しいイベントが入力されており，テンプレート長の総計としては，約６秒（210イベント）を持つことができる．大量データの処理への対策としてパターンマッチングに関しては別途ハードウエア化を検討している．

３）診断

上記の技法の認識の組み合わせおよび教師データと入力データとのズレに基づいた歌唱の診断機能を実現した．診断は事後に提示するという形だけでなく，ゲームでのリアルタイムフィードバックを行うことにより，楽しみながら歌唱制御の訓練を行うということが可能である．

４．追従型カラオケ

４．１　ピッチ訓練としてのシステム

３章で述べたノートオンイベント検出機能を利用することにより，正確に発声した場合に伴奏をつける，というカラオケシステムを容易に実現できる．メロディとそれに伴うコードをデータとして与えておき，登録されていたメロディと入力される歌唱ピッチを照合し，その照合位置における伴奏コードが演奏するというのが基本コンセプトである．但し，どのメロディの音にもコードが伴奏されるわけではない．曲の始めはシステムがカウントを取り，曲のデータによっては前奏のコードが伴奏される．その後，ユーザが歌を入力していくが，あらかじめ設定されたキーとなる音に対する伴奏コードしか演奏されない．キーの音が正しく認識されて始めて伴奏がつくことになる．要するに，カラオケのルールとして調子よく歌うためには，どれだけテンポに合った位置に正しい音が出せるか，ということにかかっており，これによりVSGとしてのゲーム性が実現されている．

４．１　自然な追従型カラオケ

自動伴奏は音楽情報処理の分野で積極的に取り組まれてきた対象の一つである．自動伴奏システムの多くは，入力された楽器音のピッチが確定してから

演奏位置を認識を行うと処理を行っている．この原理による追従型の歌唱伴奏システムも提案されている[4] が，実際の歌唱においては，ポルタメント等にみられるピッチ変動のため，音の立ち上がり時にピッチを確定するのは困難である．結果的に，歌唱を対象をする場合には唱歌風に歌わないとうまくシステムが動作しないという制約があった．我々は，音量の立ち上がりに基づいてスケジューリング*を行い，アタックから100m経過したところから平均によって安定したピッチを求め，それに基づいて歌唱位置を確認するという方式に基づいた追従型カラオケシステムを開発した[3]．スケジューリングに用いるデータはメロディから母音でつながるシーケンスを外したものを用いているため，安定してスケジューリングが可能となっている．このシステムでは，歌唱者のテンポの変化，間の挿入，フレーズの抜かしなどに対応し，他のシステム比べて，自然な形で追従することができる．

このうち，テンポの変化，間の挿入に追従した追従型カラオケをVSGの一つの機能として実現した．この追従型カラオケは歌唱者がテンポや間に変化をつけながら伴奏をひっぱていくことが可能である．そして，どれだけのキーを外したかが，歌唱者に提示される．

5．おわりに

本稿では，ビジュアルフィードバックにより、歌唱のトレーニングをおこなうVSG の現状について紹介した．本システムは，MAXを用いており，使用法に応じて容易にプログラムを拡張していくことが出来る．今後は，歌唱のうまさを評価するための診断機能をより強化するにするとともに「楽しみながらうまくなる」という意味でゲーム性の向上をはかって行きたいと考えている．さらに，著者の一人は音楽教育学を専門としており，既にカウンセリングを通じての音痴治療の事例を蓄積しつつある[5]．今後，これらの知見に基づいたカウンセリングの方法論を確立し，教育現場での実用化を進めて行く予定である．

診断機能の各節で述べた特徴量は正確に歌うときの基本情報となることには間違いないが，音楽のうまさと直接的に結びつくものではない．今後は，さまざまに実験を重ねながら歌のうまさの客観的な測定法を確立して行きたい．

* イベントが検出されたと同時に和音をならすということに加え，現状テンポに基づいて，アルペジオ等の時系列音イベント発生させること

参考文献

[1] 片寄他：MAX を利用した Voice Shooting Game, 情報処理学会研究報告, 音楽情報科学, MUS-11, 7 (1995)

[2] 第１４回国際音楽教育リサーチセミナー及び歌唱障害児（音痴）に関する国際シンポジウム委員会編資料集 (1992)

[3] H. Katayose et al. : Virtual Performer, Proc. Intl. Computer Music Conf., pp.138-145 (1993)

[4] W.Inoue, S.Hashimoto and S.Ohteru : A Computer Music System
for Human Singing", Proc. Intl. Computer Music Conf., pp.150-153 (1993)

[5]村尾忠廣：調子外れを治す, 音楽之友社 (1996)

謝辞

本研究の一部は中山隼雄科学技術文化財団の助成によるものです．