歌詞歌唱による入力が可能な Voice-to-MIDI 手法の提案

伊藤　直樹[†]　　西本　一志[††]

A Method of Voice-to-MIDI for Singing with Lyrics

Naoki Itou[†] and Kazushi Nishimoto[††]

1. はじめに

コンピュータを用いた音楽制作のための MIDI シーケンスデータ入力の一手法に，鼻歌入力 (Voice-to-MIDI：V-to-M) がある．この入力法では，ユーザ自身がフレーズから音名や音価を探す必要はなく，頭に浮かんだり耳コピのために記憶したりしたフレーズを直感的に入力可能であるといえ，特に楽器演奏技術や音高・リズムの知識が乏しいユーザにとって有用と考えられる．しかし，良好な入力結果を得るために発音（例えば「タタタ・・・」）や歌い方が制限[1)2)]されており，例えばビブラートをつけて歌ったり，歌詞をつけて歌うなどの多様な歌唱法には対応しきれていなかった．そこで本稿では，歌詞歌唱でも入力精度を大きく向上可能なタッピング併用 V-to-M 手法を提案する．提案手法の有効性に関する基礎評価のために，歌詞歌唱を用いた実験を行い，入力された音数と各音の音高の精度について調査した．

2. 提案手法と試作システム

2.1 提案手法

実際に市販鼻歌入力システムに歌詞歌唱入力を行ってみた結果，精度が下がる主原因は，歌唱を適切に 1 音ずつ区切れないことにあり，これが連鎖的に音高や音長の誤認識を引き起こしているのではないかと考えた．そこで提案手法では，歌唱と並行してメロディの各音を区切る情報（リズム区切り情報）を入力して解決をはかった．具体的には，歌唱と同時にメロディのリズムに合わせてボタンや鍵盤などをタッピングすることによって，区切り情報を入力する(図 1)．なお同様の仕組を用いた研究[3)]では音声認識への応用を試みているが,V-to-M 手法としての適用･有用性の検証はされていない．また歌唱中の歌詞認識は,その難しさ[4)]もあり現時点では対象としていない．

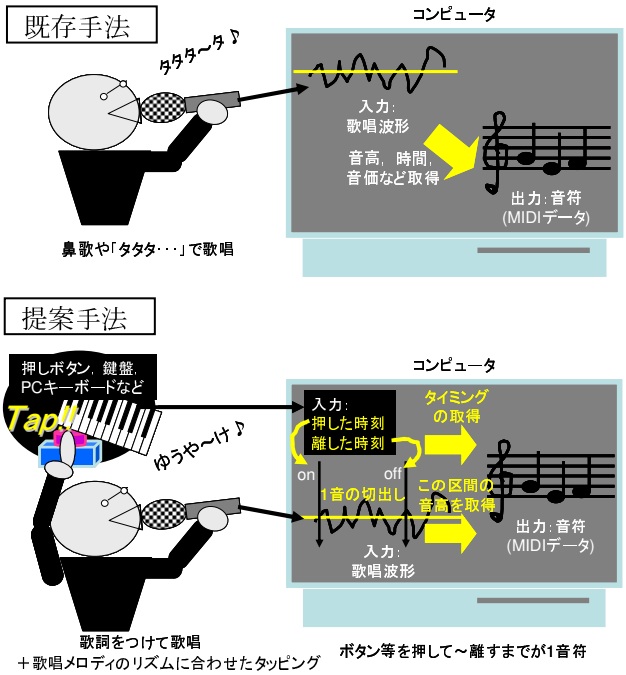

図 1 既存手法と提案手法の動作構造比較

2.2 試作システム[5)]

2.2.1 入出力される要素

入力は音声波形とリズム区切り情報，出力はE2-G4の半音単位の音高(A4 = 440Hz)である. 入力音声は44.1kHz, 16bit, モノラルでサンプリングされ，リズム区切り情報にはMIDI鍵盤楽器の打鍵と離鍵情報(いずれも入力タイミングのみ利用)を用いた.各種処理はオンライン（リアルタイム）で行う．

2.2.2 動作概要

システムに打鍵情報が入力されたらこれをトリガーとして，離鍵まで歌唱からの瞬時ピッチ算出処理を繰り返し,瞬時ピッチの時系列を記録する．離鍵後，瞬時ピッチ時系列から半音単位でとったヒストグラムの最頻値を求め，これをこの区間（1 音分）の音高として出力する．

瞬時ピッチ算出は短時間フーリエ変換 (フレームサイズ= 4096samples, STFT 間隔 = 512samples)の結果から，E2-G4 の範囲で最も優勢なスペクトルピークを探し，スペクトルの内挿法[6)]を用いて cent 単位で求めた．

3. 評価実験

3.1 実験概要

リズム区切り情報追加による効果と問題点を見るため，入力された音数とそれらの音高認識精度の評価を行った（入力位置や音長については今回評価しない）．

課題曲を歌詞をつけて歌唱したときの音高認識結果を市販の鼻歌入力ソフト 2 種 (XGworksST[1)] (以下 XGW), SingerSongWriter Lite5[2)] (以下 SSW))と比較した．被験者は，筆者らが所属する大学の男子学生 6 名であり，楽器演奏経験者 4 名，未経験者 2 名であった．

実験は大学内の防音室を用いて行った．記録用に各システムにつき 1 台のコンピュータを用意した．被験者が鍵をタッピングしながらマイクに歌唱するとミキサーを通じて

†北陸先端科学技術大学院大学　知識科学研究科
School of Knowledge Science, JAIST
††北陸先端科学技術大学院大学　知識科学教育研究センター
Center for Knowledge Science, JAIST

歌唱音声データが各 PC に分配され，既存システムは歌唱のみから音高抽出を行い，提案システムではタッピングによるリズム区切り情報も併用して音高抽出を行う．課題曲は，童謡「赤とんぼ」(全 31 音符)とし，最初にメロディを 3 回聴取させた後，楽譜を見ずにメロディに 1 番の歌詞をつけて 3 回メロディリズムのタッピング併用型歌唱（移調可）をさせた．最後に提案手法の使用感についてのアンケートを行った．

3.2 実験結果

音高認識精度の評価 [5)7)]は，被験者が音を外すことを考慮し，誤認識の原因が歌唱の誤りなのか，システムにあるのかを弁別して行う必要がある．そこで，実験時の歌唱の音響波形から第一筆者が 1 音毎に音高の特定を行い（音高を 1 つに特定できないときは可能性のある音高全てを候補とした），その音高と各システムの音高認識結果を時間位置を考慮して比較し，正解個数を割出して評価を行った．なお歌唱した音の数は，いずれの被験者もリズムの覚え間違えや音を飛ばすなどによる過不足はなく，楽譜通り 31 音，3 回で 93 音であった．

被験者ごとに集計を行った結果，いずれの被験者とも提案手法は音高の正解率(図 2)が大幅に高く（全被験者の平均正解率は提案手法が 65.9%, XGW が 24.0%, SSW が 10.2%），また入力されなかった音（音高欠落）も少なかった．正解数，音高欠落数はともに，両側 t 検定で提案手法が XGW,SSW それぞれに対して 0.1%以下で有意差があった．また XGW,SSW では余分な音符が入力されることがあったが，提案手法ではミスタッチしない限り原理的に発生しないため，余分な音は入力されなかった．なお既存システムでは歌唱時以外に喋り声なども入力されていた（音高認識精度の集計には加えていない）．これらより提案手法は，歌詞歌唱入力においても高精度に音高を入力できることが分かった．

一方で，一部被験者については，提案手法による音高欠落が多かった．これは例えば「いつのーひーか」は本来 7 音であるが，音を伸ばす箇所を省いて「いつのひか」と 5 音分しかタッピングされていないためである．

また提案手法では，オクターブの誤認識が多く見られた．これは試作システムが単純にピークを取得して瞬時ピッチ算出しているためと考えられる．これは今後改善可能であると考えているが，逆に考えると提案手法は，非常に簡易な処理法でも高い音高認識精度を実現していると言える．

3.3 アンケート結果

アンケートで提案法の使用感を聞いた結果，楽器経験者である被験者 4 名は，

「鍵盤を押すタイミングが混乱した」

「歌のリズムと鍵盤のリズムを合わせるのが難しい」

「フィードバックがないから正しく歌えているか不安」

「テンポを保つために発音以外でも押してしまう」

と述べた．これより楽器経験者がテンポの拍打ちにタッピングを用いることに慣れており，メロディリズムに合わせたタッピングが負荷を感じた可能性が伺える．

一方で，楽器演奏経験のない被験者 2 名は，

「リズムを取りながら歌えるので，それほど歌のみと感覚的には変わらなかった」

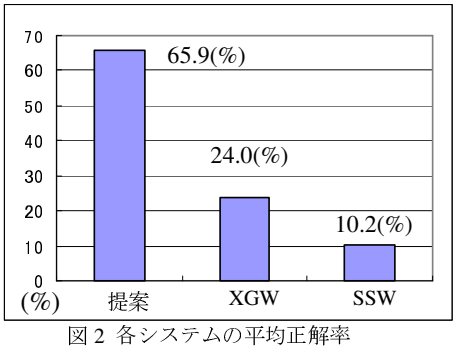

図 2 各システムの平均正解率

「普段意識的に行う行為ではないが，特に違和感はない．感覚的なものではあるがリズムをきざみながら歌ったので若干歌いやすかったようにも思う」

と回答した．ともにメロディのリズムタッピングを行うことに違和感を感じずに受け入れていることが分かる．この 2 名だけでは判断しかねるが，提案手法は楽器演奏未経験者や苦手な者には有用である可能性が高い．

4. 結論と今後の予定

提案手法はリズム区切り情報の付加により，歌詞歌唱でも高い音高入力精度を得られることが分かった．また楽器演奏経験者にはメロディリズムのタッピングが負荷である可能性がある一方，楽器演奏未経験者からは良好な評価を得た．しかし,これについては今後更なる実験が必要である．

今後は，より多くの被験者による実験，安定した瞬時ピッチ算出アルゴリズム，音高推定アルゴリズムの検討・導入を行う予定である．また実験において，しゃくりあげなどの音程的な「表情」がついた歌唱を行った被験者がいたが，そのような「表情」がついた歌唱から適切に音高を推定する方法の検討も行う予定である．

謝辞

本研究は，科学技術研究費補助金基盤研究(Ｃ)(２)課題番号 16500580 の支援を受けて実施したものである．

参 考 文 献

1) ヤマハ株式会社：XGworks ST, http://www.yamaha.co.jp/product/syndtm/p/index.html
2) 株式会社インターネット：SingerSongWriter Lite5, http://www.ssw.co.jp/
3) 番弘光，伊藤克亘，武田一哉，板倉文忠: タッピングを利用した音声認識の検討: 情報処理学会研報，SLP-47, pp71-76, 2003.
4) 尾関弘尚，鎌田貴幸，後藤真孝，速水悟: 歌声の歌詞認識における音高の影響について: 日本音響学会秋季講論集, pp637-638, 2003.
5) 伊藤直樹，西本一志: MIDI シーケンスデータの 2step 打ち込み法への鼻歌による音高入力の適用: 情報処理学会研報, EC-5, pp.43-48, 2006.
6) 原　裕一郎，井口　征士: 複素スペクトルを用いた周波数同定: 計測自動制御学会, pp.718-723, 1983.
7) 清水 純, 丸山 剛志, 三浦 雅展 柳田 益造: ハミングによる単旋律の自動採譜，日本音響学会音楽音響研究会研資，Vol.23, No.5, pp.95-100, 2004.