ジェスチャ併用型Voice-to-MIDIシステムの提案

A Voice-to-MIDI pitch input method with concurrently using tap gestures

伊藤 直樹
Naoki Itou
北陸先端科学技術大学院大学　知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology
n-itou@jaist.ac.jp, http://www.jaist.ac.jp/~n-itou/

西本 一志
Kazushi Nishimoto
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, Japan Advanced Institute of Science and Technology
knishi@jaist.ac.jp, http://www.jaist.ac.jp/~knishi/

keywords: Voice-to-MIDI, Gesture, Tapping, Rhythm segmentation, Pitch correction

Summary

Voice-to-MIDI, one of the input methods for MIDI sequence data, has a merit that users can input melodies intuitively. However, sometimes the quolity of pitch translation is not satisfactory. To solve this issue, we propose a method to correct such translation mistakes by concurrently using rhythm taps and gestures with the Voice-to-MIDI. Our method allows the users to input 3 level （high / low / hold） pitch transition information by tap gesture when they start to sing and tap. After singing and tapping, the note peers have the paradox between pitch transition by pitch translation algorithm and pitch transition by tap gesture are corrected by the pitch correction rules. We developped the prototype system and had the experiments to evaluate the pitch correction accuracy and its usability with 2 subjects. According to an example of the results, For total 4 paradoxical note peers, 1 peer was corrected propery, but others are corrected with mistakes. For our system, the subjects said it is heavy work that they should sing, tap and imagine the pitch transition concurrently. Our method shows some usable case, but we found the issues of our method and correction rules.

1. は じ め に

コンピュータを用いた音楽制作では，音楽ソフトウェアにメロディや伴奏のフレーズ，そしてそれらを自動演奏させたときによりよく聞かせるための演奏制御情報（音量の上下，ビブラートをかける等）を記録してゆく．このような制御を行うために MIDI(Musical Instruments Digital Interface）*1 という通信規格が一般的に利用され，特に音楽自動演奏のためにこの制御情報を時間ごとに並べた符号列は MIDI シーケンスデータと呼ばれている．

MIDI シーケンスデータを記録するには，鍵盤楽器などを実際に演奏して入力したり，音楽ソフトウェア上の楽譜などにマウスや PC キーボードで入力したりするなどの方法が存在している．しかしこれらの入力法では，たとえば自作曲のような楽譜などの音高やリズムが記述された情報がない場合などに，記憶されたメロディやフレーズから 1 音ずつ自ら音高やリズムを探る必要があり，特に音楽的スキルが十分でないユーザにとってはこの作業が負担となりうる．ユーザのこのような負担を解消するためには，歌唱するだけで音符の入力が可能な鼻歌入力法（Voice-to-MIDI）[YAMAHA 03, INTERNET 06] が有用だと考えられる．Voice-to-MIDI は，音高やリズムなどの音楽情報の特定をソフトウェアが行うため，特に絶対音感や相対音感を持たないユーザにとって，もっとも理想的な入力方法であり，入力時間の短縮によるスループットの向上も期待できる．

しかし，既存の Voice-to-MIDI は，入力された歌唱のみに対して音楽情報処理を行うことで変換を行ってきたが，これら従来の手法は 1 音ごとの区切りが上手くいかず，音数が誤認識されることによる変換精度の低下がしばしばみられた．そこで，我々は，これまでにユーザが歌唱と同時に鍵盤やマウスなどをタップにすることによって音符区切り情報を入力し，音数を正しく認識させることで安定的な変換精度を得る，「タップ併用型 Voice-to-MIDI」手法 [伊藤 06] を提案した．実験により，歌唱と同時にタップを行う身体的な負荷はあるものの，従来手法と比べ精度のばらつきが少なく，また従来手法では区切りが難しかった歌詞歌唱のような様々な母音・子音が含まれる歌唱に対しては，変換精度が向上した．

ところが，このタップ併用型 Voice-to-MIDI により解決したのは，Voice-to-MIDI がより高い変換精度を実現

*1 http://www.amei.or.jp/

するための主要素である「正確な区切り数（音数）の認識」と「各区切り（発音区間）における正確な音高の認識[*2] 」のうち，前者についてのみであり，後者についてはまだ解決できておらず課題であった．

そこで本稿では，音高補正に適用可能な新たな情報をタップ時に入力することによって音高の誤認識の補正を行う手法を提案する．そして，この手法を実現するために，単純なタップ動作のみに対応していたこれまでのシステムを拡張し，より高度な動作，つまりジェスチャに対応させた「ジェスチャ併用型 Voice-to-MIDI」システムを構築し，ジェスチャからの音高補正情報の入力と補正への適用についての評価を行ったので，結果を報告する．

2. 提案システムの概要

2·1 Voice-to-MIDI における音高の誤変換

Voice-to-MIDI は，ユーザが音高を意識せずに使用できる反面，意図した音高に変換されないことがあり，変換精度改善のための課題となっている．

発生要因は主に 2 つある．1 つは，音高を一意に決めるために微小時間の音高（瞬時ピッチと呼ぶ．また「ピッチ」は半音よりもさらに細かく表現した音高と定義する）を歌唱から抽出するが，声質やマイクの状態，音声情報処理アルゴリズムなどによって正しくピッチを取得できなくなるためである．本来のピッチから外れて認識されるため，意図しない音高に変換される．

2 つ目は，歌唱はピッチの変動が発生しやすいため，ド〜シという固定された音高からずれないように歌唱するのは至難であり，意図した音高に変換されないためである．そのため，例えば同一の音高を連続して歌唱したつもりでも実際にはピッチが変動していたために異なる音高に変換されてしまうことや全体的にド〜シから外れた曖昧なピッチで歌唱を行ったためにうまく変換されないことが起こりうる．

市販ソフト [YAMAHA 03] に用いられているような，変換した音高列の調性から外れる音を補正する手法では，両方を一度に解決可能であるが，特に音数が少ないときには調性を一意に決めるのは難しい．また全体的なピッチのズレを考慮して，音階側のチューニングを歌唱のピッチに合わせる手法 [来海 07] が有用と考えられるが，前者のシステム側の誤認識には対応できない．

また楽曲検索分野において，Voice-to-MIDI を応用した QBH(Query-by-Humming) と呼ばれる，口ずさんだメロディを入力としてデータベースからそのメロディを持つ楽曲を検索する手法 [Lutz 01，Alexandra 99，Sonoda 98] が存在する．そこで用いられている前の音との関係が音楽的に妥当か否かによる補正 [小杉 02] では，該当箇所は必ず補正が行われる．Voice-to-MIDI においては，例えば 1 音目の音高決定に失敗すると以降の全ての音に影響が出る可能性がある．

2·2 ジェスチャ併用型 Voice-to-MIDI

上記の課題解決のために，我々が先に提案したタップ併用型 Voice-to-MIDI 手法を拡張し，タップにより多くの意味を持たせたジェスチャ併用型 Voice-to-MIDI 手法を用いて音高補正を行う方法を提案する．

§1 音高補正手法の概要

タップ併用型 Voice-to-MIDI システムでは，1 音毎の区切りが明確になるように歌う替わりに新たに区切りを示す情報を歌唱と同時に入力する．具体的には，歌唱するメロディのリズムに合わせて鍵盤楽器や PC キーボード，ボタンなどをタップすることにより 1 音毎のリズム区切りを作り出し，これを併せて入力する．

ジェスチャ併用型 Voice-to-MIDI システムでは，タップによるリズム区切りに加えて「ジェスチャ」にあたる情報をタップ位置やタップ中の動きから取得する仕組みを追加した．そのためマウスなどの座標情報が入力できるようなディバイスでタップすることが望ましい．しかし，通常のマウスではマウスポインタの瞬間的かつ直接的な移動が容易とは言えないため，液晶ペンタブレットの使用を前提としている．ここで液晶ペンタブレットにおける音高補正情報の入力方法について述べる．

仕組みとしては，画面上のタップ位置の Y 座標の遷移を利用しており，ユーザは歌唱と同時にタップ位置を変えることによって音高の上下あるいは音高維持の情報を入力すればよい．具体的には，同じ音高が続く場合は，あらかじめ設定されている Y 軸方向のマージン内でタップし，上昇した場合はマージンより上で，逆に下降した場合はマージンより下でタップすることにより，3 パターンある音高の変化を入力してゆく．

次に音高補正方法について述べる．歌唱とタップ情報の入力が終了した後，歌唱の音響信号から求めた各音の音高候補から，後述する音高補正ルールに基づいて，タップ位置の遷移との矛盾がもっとも少ない音高列を採用し，最終的な出力とする．

なお本研究で用いるタッピングは，手で拍を打つように打拍後すぐに手を離すようなタッピングではなく，例えばピアノのダンパーペダルのように動作に必要な時間だけ押下し続けるようなタッピングを想定している．そのため，基本的には押下開始〜終了まで (MIDI シーケンスデータではそれぞれ note on，note off メッセージに対応) が 1 音となる．

§2 音高補正ルール

音高補正ルールは，タップ遷移情報は正しいと仮定し，タップ遷移にもっともあてはまりのよい音高列を求めるものとし，歌唱とタップの動きが一致しない箇所を「補正候補」として補正が可能かを判定する．基本的には以

*2 Voice-to-MIDI における「正確な音高」とは，おおまかにはシステム視点：忠実に変換できたか，とユーザ視点：意図通りに変換できたか，の 2 点があると考えられ，本稿では前者をシステム的正解，後者をユーザ的正解と使い分ける．

下のルールとなる．

- タップが同じ位置　→　前音の音高に現在音を移動したときと現在音の音高に前音を移動したときのあてはまりのよい方を採用
- 歌唱：下降，タップ：上昇　→　前音の音高を低くしたときと現在音を高くしたときのあてはまりのよい方を採用
- 歌唱：上昇，タップ：下降　→　前音の音高を低くしたときと現在音を高くしたときのあてはまりのよい方を採用
- 音高が移動できないとき　→　補正せず

あてはまり具合を決定する方法についてであるが，我々のシステムではある区間の音高を決めるために，区間内で微小時間の音高（瞬時ピッチ）を位置を移動させながら求め，そのヒストグラムの最頻値を出力する方法をとっている．そこで最初に各音のヒストグラム中でタップ遷移関係を保ったまま移動可能かをみる．もしどちらかの音のみ可能であればその音の音高を移動して解決すればよい．しかしどちらの音も移動可能な場合は，各音のヒストグラム中の移動先音高の占有度（移動先音高の個数／ヒストグラム全体の個数）が高い方を「よりあてはまりがよい」音であると考え，この音を移動して解決する．

音高が移動できない場合は，処理順に次の 2 つである．

(1) 音高移動によって補正を試みる音の前の音とのタップ遷移関係に矛盾が生じる場合
(2) 音高移動先候補がない，つまり移動可能な音高のヒストグラムが 0 の場合

次に全体の処理の流れを示す．

(1) 音高の高低の関係が発生する 2 音目から音高遷移とタップ遷移が不一致の箇所を探し，補正候補箇所とする．
(2) 補正候補箇所の 2 音それぞれについて，移動可能な音高があるかを評価し，なければその箇所は補正しない．
(3) もし一方の音が可能であればその音を補正する．
(4) 両音とも可能であれば移動後のあてはまりのよい方を補正する．

現行のルールでは，一度処理の終わった箇所については再処理を行わないようにすることで，補正可能な組み合わせが増加しないようにしている．なお [小杉 02] 同様，本補正ルールでも周辺の音との関係を用いて補正しているものの，本ルールでは補正した影響がなるべく局所的にとどまるようにし，また補正該当箇所でも状況に応じて適用されないようなルールとしたため，比較的“ ゆるい ”ものである．

2·3 実　　装

上述した要領に基づいて作成したシステム（図 1）について述べる．

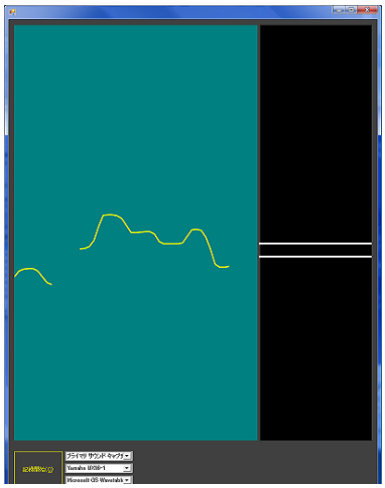

図 1　ジェスチャ併用型 Voice-to-MIDI 入力システム

§1 仕　様

システムは Microsoft Visual C # 2005, 瞬時ピッチ算出部は Visual C++2005 の DLL で作成した．また音声録音には DirectSound を用いている．入出力される情報については，入力は音声波形とタップから得られる情報 (ノートの区切りおよびタップ位置)，出力は E2-G5（A4 = 440Hz）の半音単位の音高列，つまり MIDI シーケンスデータとなる．入力音声は 22.05kHz, 16bit, モノラルでサンプリングされ，タップ情報は液晶ペンタブレットを用いて入力する．

タップの Y 軸方向のマージンは，タップ位置の Y 座標から上下にそれぞれ 20pixel，合計で 40pixel（実験に用いた機材では約 1cm に相当）とした．今回は予備実験によって 40pixel に設定したが，使用機材やユーザによって変える必要があると思われる．なお画面上のタップを行う領域には，このマージンを境界線によって表示しており，ユーザはタップを行う際に視覚的にこのマージンを把握可能である．

また評価実験などのために，note on/off や瞬時ピッチ列などの情報をテキスト形式で記録し，入力波形の Wave ファイル形式による録音も行えるように拡張した．今回の試作システムでは，音高補正処理をこのテキスト形式ファイルに対して，EXCEL のマクロ機能を用いて行った．

§2 処理の流れ

次に入力～出力までの処理の流れについて述べる．

録音が開始されると，入力されている音声波形に対して短時間フーリエ変換 (STFT) による瞬時ピッチ算出処理が録音終了までオンラインで繰り返される．そしてこの間に，タップ情報が入力されたらそれらを保持しておく．録音が終了したら各音の発音区間をタップの押下開始～終了タイミングから求め，次に瞬時ピッチの時系列から各音について半音単位の音高ヒストグラムを求める．またこのときに，タップ位置が一つ前のタップ位置のマージンを超えたか否かを判定し，各音の音高の変化の有無

表 1 実験で行った Condition の一覧

	タップ方法	歌唱	テンポ
Cdn 1	上下あり	歌詞	遅い
Cdn 2	同位置	歌詞	遅い
Cdn 3	上下あり	タタタ	遅い
Cdn 4	上下あり	歌詞	速い
Cdn 5	----	タタタ	遅い

を取得する．

最後に，歌唱の音響信号から求めた各音の音高ヒストグラムから，音高補正ルールに基づいて，タップ位置の遷移にもっともあてはまりのよい音高列を採用し，出力する．比較のための補正を行わない音高列については，各音のヒストグラムの最頻値を出力する．

ここで瞬時ピッチ算出方法について述べる．

瞬時ピッチは，入力波形に対する FFT(フレームサイズ= 2048samples : 約 100ms) から求めたパワースペクトルの E2-G5 間に存在するピークと，そのパワースペクトルに対する FFT によって求めた自己相関のピークを組み合わせて総合的に判断する．しかし，そこで決定したピークは周波数解像度の低さゆえに特に低音域では音高を一意に決定できない．そこでスペクトルの内挿 [原 83] を用いて cent 単位の音高推定を行い，この値を瞬時ピッチとして出力する．STFT フレーム移動間隔は 256samples（約 12ms）である．

3. 評 価 実 験

タップ位置の上下動作の情報が音高補正に有用であるか，またその作業負荷や使用感などについて評価するため，試作システムを用いた評価実験を行った．

3·1 概 要

実験は表 1 に示すような 5 つの設定で行った．タップ位置の上下動作の情報が音高補正に有用であるかどうかについては，Condition 1 の設定で歌唱させたときの結果を用いて，出力された音高列に対する音高補正の有無による比較を行い，被験者自身がどちらを気に入るかによって提案法の効果について評価する．

また，作業負荷や使用感などについては，主にアンケートを用いた主観評価により，提案法を用いた Condition 1, 3, 4 と他の Condition とを比較し評価する．既存 Voice-to-MIDI 手法を用いる Condition 5 では，YAMAHA XGWorks ST[YAMAHA 03] を使用して入力を体験させた．アンケートの項目を以下に示す．

- 精神的負荷について（7 段階評価）
- リズム通りにタップできたと思うかについて (6 段階評価)
- メロディに対する理解の深まりについて（6 段階評価）
  もし「深まった」と回答した場合で可能であれば，その具体的な内容の記入もお願いした．

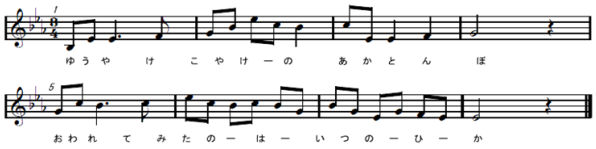

図 2 課題曲「赤とんぼ」 作曲：山田耕作

表 2 予備調査の結果と各被験者の音楽歴

	絶対音感		音程感	楽器経験
	正解	半音違い		
被験者A	0	1	6	ピアノ10年，ギター7年など
被験者B	3	3	6	ピアノ11年

- 意見や感想の自由記述

歌唱に用いる曲については，作曲させるのは難易度が高く時間もかかり，また，被験者自身が，どのようなメロディを考えたかというユーザ的正解を記憶していなければ音高補正の評価が難しくなるため，多くの人が楽譜のような視覚的情報としては知らないが，歌唱できる程度に知っていると思われる童謡「赤とんぼ」（全 31 音符）を用いた.「赤とんぼ」は図 2 の楽譜に示す通り，広い音域で起伏に富む中にも同一音高が連続する箇所があり，タップ位置の違いを用いる提案手法の効果を測るためには最適だと思われる．

被験者は，筆者らが所属する大学の男子学生 2 名である．実験に先立ち，予備調査により被験者の音高知覚能力を調べた．その内容を以下に示す．

- 弾かれた単音の音名を回答:「絶対音感」
- 連続して弾かれた 2 音の単音の高低を回答:「音程感」

いずれの項目とも全 6 問ある．表 2 に被験者ごとに結果と各被験者の音楽歴を示す．

以上より両被験者とも音高の高低は知覚できると言え，また被験者 B は絶対音高を知覚できると考えられる．

次に実験環境について述べる．実験は筆者が所属する大学内の防音室で行った．マイクには，音声チャット用の安価なヘッドセットマイク（型番不明）を用い，タップディバイスには，hp 2710p ノートブック PC の液晶ペンタブレット機能を用いた．

なお誤認識の原因にもなりうるためマイクの選択には議論の余地があるが，我々は最終的にはタブレット機能を備えたモバイルツールで手軽に使用できるシステムを目指しており，マイクはツール内蔵のものか携行しやすいものの使用を前提としているため，敢えて音楽用のマイクは用いなかった．また，実験を防音室で行ったのは，これまで歌唱を伴う実験を行ってきた経験の中で歌唱を他人に聞かれることを嫌う被験者が多かったため，他人に歌唱を聞かれにくい場所を提供し作業に集中させることが目的である．*3

実験手順について述べる．最初に「赤とんぼ」のメロ

*3 今後，より環境音が多い場所における実験も考えている．

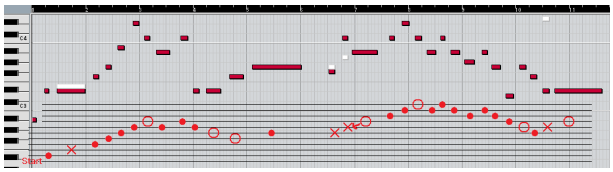

図 3 被験者 A の音高補正結果．横軸は時間，縦軸は MIDI データでは音高，タップ遷移情報ではタップの上下動の累積値を示す．上半分の黒のバーが補正なしの MIDI データ，白のバーが補正箇所である．下半分はタップ遷移情報であり，●が音高遷移とタップ遷移が一致した箇所，○が不一致な箇所（補正候補），×が実際に補正された箇所を示す．

ディのみを 1 番の歌詞を見せながら 3 回聴取させた．次に，提示された Condition に従ってメロディを 3 回歌唱後アンケートに回答させるという作業を全 5 Condition 分行った．なお各 Condition はランダムな順番で提示した．最後に，Condition 1 の 3 つの歌唱から音高遷移とタップ遷移が不一致な箇所が多いものを 1 つを選び，音高補正ルールに従って第 1 筆者が補正を行い，後日，被験者に実験時の歌唱 Wave ファイルを聞かせた後，補正前後の MIDI シーケンスデータをブラインドテストで試聴させ，どちらが好みかを述べさせた．

提案システムおよび XG WorksST については，ともに実験前に操作法の簡単な説明のみを行った．被験者に聴取させた「赤とんぼ」のメロディは，野ばら社刊「童謡」に収められた変ホ長調の楽譜に従い MIDI シーケンスデータで作成した．演奏テンポは BPM=60 とした．

必ずしも楽譜通りのメロディを覚えているとは限らないが，覚えなおすのは負担となるので被験者の覚えているメロディで歌唱してもよいこととし，歌唱テンポは「遅い」の場合 BPM=60 程度,「速い」の場合 BPM=110 程度を目安に歌唱するように指示した．おおよそのテンポを把握させるため，各 Condition 開始前にメトロノームでこれらのテンポを体感させている．ただし，メトロノームのクリック音は許容しがたいレベルで入力音声に影響を与える可能性があるため，歌唱中はメトロノームを使わせていないため，実際の歌唱テンポは変動を含んでいる．

3·2 音高補正結果

音高補正結果について述べる．ただし，被験者 B はタップ抜け (タップ併用型 Voice-to-MIDI の問題点の一つ) が発生したため，現行のルールでは対応できないため評価から除外し，被験者 A の結果についてのみ述べる．

図 3 に音高補正結果として補正なしの MIDI データ・補正された箇所・タップ遷移情報を示す．音高遷移とタップ遷移が不一致な箇所は全 30ヶ所中 10ヵ所であり，そのうち実際に補正の対象になった箇所は 4 か所であった．2 つ目の補正箇所である 16・17 音目は，直接の補正候補 17 音目ではなく一つ前の 16 音目が補正されたが，他 3 か所は補正候補音が補正されていた．なお A は終盤の「のーひーか」の付近でオリジナルメロディと違うメロディで歌唱している．

被験者 A に対するブラインドテストの結果,「補正前の方が好みである」との回答を得た．この要因として，今回のサンプルでは本来は変換しなくてよいと思われる補正箇所が変換されてしまったことが挙げられる．

1 つ目の補正箇所である 2・3 音目は D ＃ 3 → D ＃ 3 から D ＃ 3 → E3 に補正されているが，図 2 の楽譜によるとこの箇所は楽譜上は同一音高が正しい．音程が楽譜と同じに保たれていればキーは変わるものの違和感はなく聞こえると考えられるため，A が覚え間違いなどをしていないと仮定すれば，ここでは必要のない補正がされたと言える．これは 4 つ目の補正箇所である 30・31 音目でも同様のことが言える．

3 つ目の補正箇所である 16・17 音目では，タップは上昇したが音高は下がったため，タップ遷移情報へのあてはまりのよい補正音高候補を音高ヒストグラムから選ばれている．しかし，補正前の 15 と 16 音目の完全 4 度上昇（15 音目は補正されているため，補正後の音高 G3 が基準になる），および 16 と 17 音目の長 2 度下降という音程は，いずれも楽譜上の音程と一致している．つまりこの箇所でも必要のない補正がなされたと言える．

A が覚え間違いなどをしていないと仮定すると，これらはいずれもミスタップをルールが無視できなかったために発生したと結論づけられる．今後タップの優先度を減らしたルールの採用などによる改良が必要であり，また 4 つ目のような音楽的にあまりみられない跳躍は除外することも必要である．

しかしながら，2 つ目の補正箇所である 14・15 音目は楽譜上の音程関係的には正しく補正され，ルールがうまく機能した例と考えられる．ここでは 15 音目の音高ヒストグラム中で G3 が占める割合が，14 音目の音高ヒストグラム中で F ＃ 3 が占める割合より大きかったため，F ＃ 3 → G3 と補正された．これにより 2 音が楽譜通りに同一音高になっただけでなく，もともと楽譜通りであった 13 音目と 14 音目の音程（長 2 度上昇）が保たれた上での補正であることがわかる．

このほかの部分についてみてみると，6 音目は補正候補ではないが，本来半音低い A ＃ 3 と認識されるべきだと思われ，補正が必要と思われるがタップ遷移情報的には矛盾がない場合には対応できない，という問題点の存在がわかった．

以上より提案法および現行ルールについてまとめると，

- ルール通りに音高補正が機能した
- 必要がないと思われる箇所の誤補正あり
- 必要と思われるがタップの遷移情報的に矛盾がない箇所が未補正
- 不適切なタップ遷移情報が入力される可能性あり

ということが言える．

表 3 被験者 A のアンケート結果

	タップ方法	歌唱	テンポ	負荷	タップできたか	理解の深まり
Cdn 1	上下あり	歌詞	遅い	非常に高い	全く思わない	深まった
Cdn 2	同位置	歌詞	遅い	低い	非常に思う	どちらかというと深まった
Cdn 3	上下あり	タタタ	遅い	非常に高い	全く思わない	深まった
Cdn 4	上下あり	歌詞	速い	非常に高い	全く思わない	深まった
Cdn 5	----	タタタ	遅い	非常に低い	----	全く深まらず

表 4 被験者 B のアンケート結果

	タップ方法	歌唱	テンポ	負荷	タップできたか	理解の深まり
Cdn 1	上下あり	歌詞	遅い	どちらかというと低い	どちらかというと思う	深まった
Cdn 2	同位置	歌詞	遅い	普通	どちらかというと思わない	どちらかというと深まらず
Cdn 3	上下あり	タタタ	遅い	どちらかというと高い	どちらかというと思う	どちらかというと深まらず
Cdn 4	上下あり	歌詞	速い	どちらかというと高い	思わない	どちらかというと深まらず
Cdn 5	----	タタタ	遅い	普通	----	どちらかというと深まらず

3·3 アンケート結果

作業負荷などに関するアンケート結果について述べる．

被験者 A の結果を各 Condition の差異も併せて表 3 に示す．表 3 より提案法である Condition 1 と単純なタップ併用型 Voice-to-MIDI 法である Condition 2 の差異はタップ遷移情報の入力の有無だけであるが，A にとって音高の上下を意識しながらのタップは負荷が高い作業であることがわかる．これは Condition 3 や Condition 4 についても同様の回答であった．また Condition 1 と Condition 4 よりテンポに関係なく負荷が高いという傾向が読み取れる．そのほか，タップをテンポに乗ってうまくできたとは感じておらず，これも負荷が高いことを示している．

Condition 5 は既存 Voice-to-MIDI 法であるが，負荷は非常に低いことがわかる．しかし，歌唱したメロディへの理解は深まっていないと回答した．一方，提案法については，理解は深まったという回答を得，そのコメントとして「前後の音との関係 (上か下か) がわかる」というような回答があった．

次に被験者 B の結果を各 Condition の差異も併せて表 4 に示す．表 4 の Condition 3 や Condition 4 と Condition 2 を比較すると，A ほどではないものの，B にとっても音高の上下を意識しながらのタップは負荷が高い作業であることがわかる．なお Condition 1 の「どちらかというと低い」という回答は，B がテンポの速い Condition 4 の次に行ったため「少しやりやすくなった」とのコメントから順序効果によるものと推測されるため考慮しない．タップをテンポに乗ってできたかについては，アンケート結果では Condition 1，3，4 で回答が違うが，実験時に録った歌唱の Wave ファイルでテンポが不安定になっている箇所が見られたことから，提案法は負荷が高いと考えられる．

Condition 5 には普通と回答している．しかし，歌唱したメロディへの理解はどちらかというと深まらなかったと回答し，A 同様，提案法については，理解は深まったという回答を得た（順序効果の可能性はある）．

以上から，提案法に対する作業負荷や使用感についてまとめると，提案法については，歌唱・タップ・音高把握を同時に行うことは負荷が高いと言える．しかし，継続な使用を行った場合に変化がみられるかについて調査の必要がある．また，提案法は自分の歌唱に注意が向く効果が期待でき，音楽制作以外の歌唱練習などへの応用が考えられることがわかった．

4. 結　　論

我々は，各発音区間における音高の誤認識という Voice-to-MIDI の精度向上における問題点に対して，音高補正情報をタップジェスチャ入力することによって解決を行う手法を「ジェスチャ併用型 Voice-to-MIDI」として提案した．また実際のシステムを構築し，ジェスチャによる音高補正情報の入力と補正への適用の評価を行った．

その結果，ルール通りに音高補正が機能した箇所がある一方で，必要がないと思われる箇所の誤補正や本来必要と思われるがタップの遷移情報的に矛盾がない箇所の未補正があった．また，不適切なタップ遷移情報が入力される可能性あることがわかった．今後，タップの優先度を減らしたルールなどによる改良を行う予定である．

また，提案法に対する作業負荷や使用感については，提案法については，歌唱・タップ・音高把握を同時に行うことは負荷が高いことがわかった．しかし，継続な使用を行った場合に変化がみられるかについて今後調査の必要がある．また，提案法は自分の歌唱に注意が向く効果が期待でき，音楽制作以外の歌唱練習などへの応用が考えられることがわかった．

その他，モバイルツール移植のための実験やジェスチャ機能の拡充を行う予定である．

◇ 参　考　文　献 ◇

[YAMAHA 03] ヤマハ株式会社：XGworks ST，http://www.yamaha.co.jp/product/syndtm/p/cmp/xgwstw/index.html.

[INTERNET 06] 株式会社インターネット：SingerSongWriter Lite5，http://www.ssw.co.jp/products/ssw/win/sswlt50w/index.html.

[伊藤 06] 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 2006-EC-5, Vol.2006, pp.43-48，2006.

[来海 07] 来海 大輔，江村 伯夫，三浦 雅展，柳田 益造：音高・音価テンプレートを用いた単音節歌唱の採譜精度の向上，日本音響学会，音響研資 MA2007-73，Vol.26, No.6, pp.99-104，2007.

[Lutz 01] Lutz P.，Rainer T.：An Interface for melody input, ACM Trans. on Computer-Human Interaction（TOCHI），Vol.8，No.2，pp133-149，2001.

[Alexandra 99] Alexandra U.，Justin Z.：Melodic matching techniques for large music databases：Proc. of the seventh ACM int. conf. on Multimedia, MULTIMEDIA '99，pp57-66，1999.

[園田 98] Tomonari Sonoda, Masataka Goto, Yoichi Muraoka：A WWW-based Melody Retrieval System：ICMC 98 Proc.，pp349-352，1998.

[小杉 02] 小杉 尚子，小島 明，片岡 良治，串間 和彦：大規模音楽データベースのハミング検索システム，情処論，Vol.43, No.2, pp.287-298，2002.

[原 83] 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，pp718-723，1983.