社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－HCI－128 (19)
2008－MUS－ 75 (19)
2008／5／29

Voice-to-MIDIシステムのためのジェスチャを用いた音高補正手法の検討

伊藤 直樹† 西本 一志†

音楽制作におけるMIDIシーケンスデータ入力の一手法に，鼻歌入力(Voice-to-MIDI)がある．この入力法の主たるメリットは，覚えたメロディを手動で音名に変換する必要がない点であるが，良好な入力結果を得るのは難しかった．そこで筆者らはこれまでに，各音の区切り情報としてメロディのリズムと一致したタップ情報を歌唱と同時に入力することにより，入力された歌唱に対して安定した精度の獲得や自由なスタイルの歌唱に対する音高データ変換性能の向上を狙った，タップ併用型Voice-to-MIDIシステムを開発した．本研究では，入力歌唱通りではなくユーザの意図通りに変換可能なシステムを目指して音高変換精度をさらに向上させるために，タップをジェスチャに拡張し，ジェスチャ情報を利用した音高補正アルゴリズムについて検討する．

A Pitch Correction Method Using Pen Gestures for A Voice-to-MIDI Pitch Input Method

NAOKI ITOU†, KAZUSHI NISHIMOTO†

Voice-to-MIDI, one of the input methods for MIDI sequence data, has a merit that users can input melodies intuitively, so that they are released from tasks to translate their memorized melodies into chromatic pitches manually. However, the quality of translation by the ordinary VtoM was not satisfactory. To improve an objective performance, we proposed a VtoM concurrently using rhythm taps to input boundary information of notes. In this paper, to improve a "subjective" performance, we propose a new VtoM concurrently using gestures to input not only the boundary information but also information on relative transitions of pitches that a user intends to sing. We investigate a pitch recognition algorithm using the gesture information to achieve subjectively correct results.

1.はじめに

音楽制作におけるMIDIシーケンスデータ入力法のひとつに，鼻歌入力（Voice-to-MIDI：以下VtoM）法[1)2)]がある．VtoMでは，音高やリズムをコンピュータが特定する．ユーザは，マイクに向かって，頭に浮かんだメロディや記憶しているフレーズを歌うだけで音符を入力できるので，特に絶対音感や相対音感を持たないユーザにとって有用な入力方法である．また，音楽制作用途だけではなく，曲の一部を口ずさんで楽曲検索を行うQBH（Query-by-Humming）手法[3)4)5)]にも応用されている．

従来のVtoMシステムは，入力された歌唱の音響信号のみを用いて処理を行ってきた．しかし，各音を正確に区切ることが難しいため，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によって変換精度を向上する「タップ併用型Voice-to-MIDI」手法[6)]を提案した．実験により，歌唱と同時にタップを行うことや，短い音長を入力する連続タップ動作がやや難しいものの，従来手法と比べ変換精度が向上することがわかった．また従来手法では全く扱えなかった，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も，良好な精度で変換できることを示した．

VtoMにおける音高変換精度の基準として，次の2つが考えられる．第1は客観的基準であり，入力された歌唱の音響データを忠実に変換できたかどうかで精度を評価する．従来のVtoMの評価は，我々自身の事例も含め，この客観的基準に基づいて行われてきた．第2は，主観的基準である．VtoMのユーザは，ほとんどの場合音楽的な訓練を受けていないため，歌唱技術は一般に低く，安定した音高で歌唱することができない場合が多い．このため，自分で歌っているつもりの音（主観的歌唱音）と，実際に発声されて歌われている音（客観的歌唱音）とが食い違っていることがしばしばある．客観的基準が客観的歌唱音との一致度を問う基準であったのに対し，主観的基準は，主観的歌唱音との一致度を問う基準である．

従来のVtoMは，客観的歌唱音のみを用いて変換を行うため，得られる変換結果が主観的歌唱音と一致せず，ユーザにとって満足できない結果となることがある．つまり，客観的基準で

†北陸先端科学技術大学院大学
†Japan Advanced Institute and Science of Technology

の変換精度は高いが，主観的基準での変換精度は必ずしも高く無い場合があった．我々が開発したタップ併用型 VtoM 手法も，客観的精度を向上することを目指したものであったが，主観的精度の向上のための機能は組み込まれていなかった． ユーザの満足度を向上させるためには，主観的精度の向上手段を実現することが不可欠である．

そこで我々は，主観的精度向上のための第 1 歩として，タップ併用型 VtoM 手法を拡張した，ジェスチャ併用型 VtoM 法を提案する．この手法では，タップを入力する際，同時に音高の相対的推移に応じてタップ位置を上下してもらう．これによって，現在入力しようとしている音の音高は，1 つ前の音の音高よりも主観的には高いのか，低いのか，あるいは同じかに関する情報を取得する．この情報を用いて，客観的歌唱音を変換した結果を，主観的歌唱音にできるだけ一致するように補正する．本稿では，被験者実験によって，考案した手法の有用性と問題点を検証する．

2. 提案システムの概要

2.1 従来の VtoM における誤変換の補正方法

VtoM の変換精度低下の要因は主に 2 つある．

第 1 は，システム側の要因である．ほとんどのシステムでは，微小時間の音高（以下「瞬時ピッチ」と呼ぶ．）を歌唱から抽出し，その集合から音高を決定すると推察される．その際，声質やマイク，瞬時ピッチ抽出アルゴリズムの質などの要因で，実際のピッチから外れて認識される場合がある．この誤りは，客観的精度にも主観的精度にも影響する．

第 2 は，入力者側の要因である．訓練されていない者にとって安定した正確な音高で歌唱するのは至難の技である．このため，例えば高い音が正しい音高まで上がりきらなかったり，同一の音高を連続して歌唱したつもりでも，不安定な発声のためにピッチが変動し，異なる音高に変換されてしまったりすることなどが起こる．この誤りは，おもに主観的精度に影響する．

従来，第 1 の要因に対する対策による精度向上が図られてきた．文献 7)では，歌唱データから個々の音高の境界となる周波数閾値を推定することにより，システムを各ユーザに適応する手法が提案されている．このような手法によって客観的精度は向上されるが，主観的精度の向上には第 2 の要因に対する対策が必要である．

しかしながら，第 2 の要因については，入力データにそもそも誤りが含まれているため，入力データだけによって誤りを修正することはできない．何らかの外部知識ないし補正のための別のデータが必要となる．これまでは，外部知識として音楽理論を用いる手法が主として試みられてきた．市販ソフト [1)]では，歌唱データから楽曲の調性を推定し，その調性に対応するスケールに属する音高に修正する．また文献 8)では，前の音との関係が音楽的に妥当か否かによって補正を行う．これらの方法は，局所的な音のずれには対応できるが，歌唱全体が不安定な場合には対応が難しいと思われる．また，スケール外音を多く含む楽曲には適用できない．

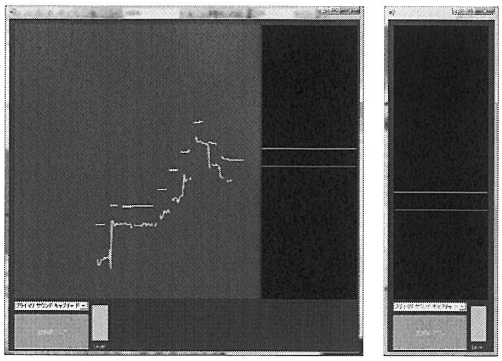

図 1 システム外観（左：通常用，右：実験用）
通常用では右に縦長を十分にとったタップスペース，左に歌唱ピッチとタップ位置のアニメーション画面を備える．3 章の実験に用いたものはアニメーションを表示しない．

2.2 ジェスチャ併用型 Voice-to-MIDI

我々は，第 2 の要因に対する対策として，楽理などの外部知識を用いるのではなく，歌唱と同時に歌唱音の主観的な音高推移を示すデータをユーザに入力してもらうことによる音高補正を試みる．そのために，先に提案したタップ併用型 VtoM 手法を拡張したジェスチャ併用型 VtoM 手法を提案する．

2.2.1 主観的音高推移情報の入力手法

タップ併用型 VtoM システムでは，1 音毎の区切りを示す情報を，鍵盤楽器などをタップすることで歌唱と同時に入力していた．ジェスチャ併用型 VtoM システムでは，タップデバイスとして液晶ペンタブレットを用いる．図 1 にシステムの外観を示す．ユーザは，タップ併用型 VtoM システムと同様，歌唱時に同時にタッピングを行い，音の区切り情報を入力する．その際，さらに主観的な音高の上下推移に応じて，タップスペース上で Y 軸方向にタップ位置を上下させる．1 つ前の音より現在の音を高く／低く歌っているつもりであれば，予め設定されているマージン幅よりタップ位置を上げる／下げる．あるいは 1 つ前の音と同じ音高の音を歌っているつもりであれば，マージン幅内で前と同じ位置をタップする．なお，マージン幅は，タップ毎に自動的に液晶画面上に表示される．こうして，タップ毎に 3 パターンの主観的な音高の相対的推移情報を入力してゆく．

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，後述する音高補正処理に基づいて，タップの推移情報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．

2.2.2 システムの仕様

システムの作成には Microsoft Visual C＃2005 を用い，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．入力は音声波形とタップによるノートの区切りおよび Y 軸位置，出力は E2-G5（A4 = 440Hz）の半音単位の音高列，つまり MIDI データとなる．入力音声は 22.05kHz, 16bit, モノラルで記録され，タップ情報は hp 製液晶ペンタブレット PC 2710p 上で入力する．

タップの Y 軸方向のマージンは，予備実験より，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．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．

この他，評価用に入力波形の Wave ファイル，タップ時刻や瞬時ピッチ列，各音のヒストグラム，補正を行わない音高列を記録している．

次に入力～出力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，フレームサイズ t_{win} = 2048samples：約 100ms，フレーム移動間隔Δt=128samples：約 6ms)による瞬時ピッチ算出処理が録音終了まで繰り返される．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，音高補正処理に基づいて，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．比較用音高列は各音のヒストグラムの最頻値を用いて出力する．なお note off～次の note on までの間隔が 200ms 未満の場合は，次のタップの準備のためのやむをえないタップ終了とみなし，note on～次の note on 直前までを 1 音とする．

瞬時ピッチ算出は，基本的に入力波形に対する FFT から求めたパワースペクトルの E2-G5 相当の周波数間に存在するピークのうち，こ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最大ピーク近傍の周波数のものを用いて求める．次にスペクトルの内挿 [9]を用いて cent 単位の音高推定を行い，瞬時ピッチとして出力する．これは周波数解像度不足を補うためである．

2.2.3 音高補正処理

本研究における，個々の音（あるタップの開始位置から次のタップの開始位置まで）の候補音高の求め方は以下の通りである．1 つの音区間の時間幅 t_{note} より十分に小さな時間幅 t_{win} を持つ「窓」を，その音区間の開始位置から終了位置まで，微小時間Δt（$\Delta t \ll t_{win}$）ずつシフトさせつつ，各窓について瞬時ピッチを求める．こうして得られた n= $((t_{note} - t_{win}) / \Delta t) + 1$ 個の瞬時ピッチのヒストグラムを作成し，最も高い頻度をとるピッチを，その音のピッチ候補とする．

こうして得られたピッチ候補の推移と，タップ位置の推移とを比較し，両者の推移が一致しない箇所を探し，補正候補箇所とする．本研究では，タップ位置推移情報は主観的な音高推移を正しく反映しているものと仮定している．ゆえに，不一致箇所については，候補音高を頻度が次点以下の候補に差し替えて，両者の推移が一致するものを見出すことにより，主観的な音高の上下推移に一致するように補正する．

不一致のパターンと，各パターンにおける補正方法を以下に示す．

- 歌唱：上昇または下降，タップ：同位置
  前音の音高に現在音を移動した場合と，現在音の音高に前音を移動した場合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場合と，現在音を高くした場合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場合と，現在音を高くした場合のあてはまりのよい方を採用

なお，あてはまり具合は，以下のようにして判断する．2 音それぞれについてヒストグラム中で 1 以上の頻度の音高（ただしタップが同位置のときは頻度 0 も対象とする）に着目し，前後音とのタップ推移関係を満たすような移動が可能かをみる．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する．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前音高の占有度（移動前音高の個数 / その音のヒストグラム全体の個数）が低い方を「移動すべき」音であるとして移動する．

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．

補正先となる音高候補を音高推移情報によって絞り込める点が本手法の特徴である．

3. 評価実験

3.1 実験概要

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，システムを用いた評価実験を行った．

実験は，被験者にメロディの歌詞歌唱と同時に上下つきタップを入力させて行い，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る．

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)
- 意見や感想の自由記述

歌唱に用いる曲については，作曲は負荷が高く確実に主観的正解が得られるとは限らず，また新たに曲を覚えるのは負荷があるため，被験者にとって既知と思われる童謡「赤とんぼ」（全 31 音符）とした．よって，この実験では楽譜またはそれを移調したものを主観的正解とみなして代用することとする．今回は，野ばら社刊「童謡」に収められた変ホ長調の楽譜（図2）に従い，被験者に聴取させるメロディのみの MIDI データで作成した．

「赤とんぼ」は，広い音域で起伏に富む中にも同一音高が連続する個所があり，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女子学生1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．

1. 指された鍵名を回答:「音名」
2. 弾かれた単音の音名を回答:「音高聴取」
3. 連続して弾かれた2音の単音の高低を回答:「音程感」

表1　項目 1-3 の 6 問中の正解数と音楽経験

	音名	音高聴取		音程感	楽器経験
		正解	半音違い		
被験者A	5	1	4	6	電子オルガン(3～4歳)，ピアノ(4～11歳)

いずれの項目とも全 6 問ある．表 1 にその結果と被験者の音楽歴を示す．

被験者 A は過去に楽器学習経験はあるが，現在はない．A の音高聴取の結果は比較的よいが，回答までに各問とも 10 秒程度の時間がかかっていたため，聴取したメロディを即座に楽譜情報に変換できるような VtoM が不要なレベルではないと判断し被験者として採用した．

実験手順は，最初に 3 分間歌唱の練習時間を設ける．次に，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行った．実験では，比較のために上下をつけないタップによる入力（音高補正をしない）も行わせた．実験前にはシステムの操作法の説明を行った後，3 分間の練習を行わせている．歌唱はなるべくビブラートなどのせずに一定のピッチで行うように依頼し，移調を認めた．一連の過程で楽譜は一切呈示しなかった．

3.2 実験結果

3.2.1 各正解の導出法

客観的正解および主観的正解を求めるために以下に示すような作業を行った．

客観的正解は，各歌唱の音響波形から，第一筆者が各音の音高の特定を行った．その音高とシステムの音高変換結果を比較して正解個数を割り出した．

音高の特定は，録音波形をループ再生しながら，PCM 音源の持続音を同時に鳴らし，適宜ピッチベンドホイールも用いて，うなりを聞くことによって行った．各音の区切りは波形の目視によって行い，各音について子音部を除いた数か所に分けて音高を判定した．2 音にまたがる曖昧な音や音高の変化がある場合はとりうる音高の候補として全て記録した．うなりの聴取だけでは決め難い場合は，1 波長の時間から周

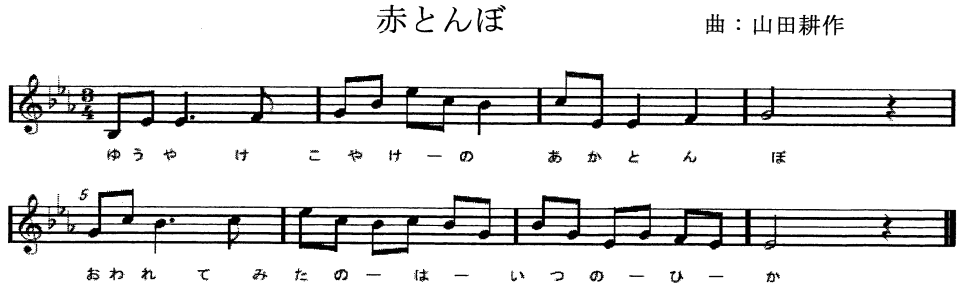

図 2 課題曲「赤とんぼ」

波数を求めるなどの手段も援用した．

客観的正解の判定は，システムが求めた音高が割り出した音高候補のいずれかに該当すれば正解とした．また，タップタイミングがずれていてもタップ区間の一部が該当する発音区間に重なっていればOKとした．その上で，

a. 一致した音
b. 一致しなかった音
c. 1音を複数音に認識し，かつそのいずれの音も正解と一致しなかった音
d. 本来の31音に対して欠落した音
e. 余分な音

に分類して個数を集計した．a〜dを合計すると31音となる．

主観的正解は，移調して歌唱されている可能性を考慮して，音程関係が維持される音がもっとも多いキーに移調したときを正解として採用した．

3.2.2 結果と考察

結果の一例として，被験者Aの1回目の歌唱の結果を表2に示す．また歌唱ピッチの軌跡およびタップ開始・終了点を図3に示す．

客観的精度は90.3%（28音 / 31音）であった．この例では，タップの抜けや余分なタップはなく，タップタイミングの多少のずれはあったものの，全く関係のない箇所におけるタップはなかった．誤変換の理由について考察する．

1音目については，録音波形を調べた結果，1音目が歌唱者の発声音域の下限に近く安定したピッチ・声質で歌唱できなかったことに起因してシステムがピッチを約1オクターブ誤認識したと思われる．15, 26音目については，タップ開始タイミングのズレや十分な時間タップし続けられていないことが原因と考えられる．

次に音高補正結果について述べる．

「タップ推移」と「タップ推移の正解」欄より誤ったタップ推移情報が入力された箇所はないことがわかる．音高とタップの推移が不一致な箇所は4か所あり，このうちの3か所の音高

表2 被験者Aの1回目の歌唱に対する音高補正処理結果

No.	歌詞	得られたデータ 出力音高と音高推移		タップ推移	音高とタップの不一致	音高の補正結果	正解 客観的正解	主観的正解	タップ推移の正解
1	ゆ	A3	--	--	--	A3	G#2>F#2>A2,A#2	A#2	--
2	う	D#3	Down	Up	○	D#3	D#3	D#3	Up
3	や	D#3	Keep	Keep		D#3	D#3	D#3	Keep
4	け	F3	Up	Up		F3	E3,F3	F3	Up
5	こ	G#3	Up	Up		G#3	G3,G#3	G3	Up
6	や	B3	Up	Up		B3	A#3,B3	A#3	Up
7	け	D#4	Up	Up		D#4	D4,D#4>D#4	D#4	Up
8	ー	C4	Down	Down		C4	C4,C#4	C4	Down
9	の	A#3	Down	Down		A#3	A#3	A#3	Down
10	あ	C4	Up	Up		C4	C4,C#4	C4	Up
11	か	D3	Down	Down		D#3 ○	C#3>D3	D#3	Down
12	と	D#3	Up	Keep	○	D#3	D#3	D#3	Keep
13	ん	F3	Up	Up		F3	F3	F3	Up
14	ぼ	G3	Up	Up		G3	G3,G#3	G3	Up
15	お	F2	Down	Keep	○	G3 ○	G3	G3	Keep
16	わ	C4	Up	Up		C4	C4	C4	Up
17	れ	A#3	Down	Down		A#3	A#3	A#3	Down
18	て	C4	Up	Up		C4	C4,C#4	C4	Up
19	み	D#4	Up	Up		D#4	D#4	D#4	Up
20	た	C4	Down	Down		C4	C4	C4	Down
21	の	A#3	Down	Down		A#3	A#3	A#3	Down
22	ー	C4	Up	Up		C4	C4,C#4	C4	Up
23	は	A#3	Down	Down		A#3	A#3	A#3	Down
24	ー	G3	Down	Down		G3	G3	G3	Down
25	い	A#3	Up	Up		A#3	A#3,B3	A#3	Up
26	つ	F2	Down	Down		G3 ○	G3	G3	Down
27	の	D3	Up	Down	○	D3	D3>D#3	D#3	Down
28	ー	G3	Up	Up		G3	G3	G3	Up
29	ひ	F3	Down	Down		F3	F3	F3	Down
30	ー	D#3	Down	Down		D#3	D#3	D#3	Down
31	か	D#3	Keep	Keep		D#3	D3>D#3	D#3	Keep

注：「音高の補正結果」欄の丸がついた音は補正が行われた音，「正解」欄で太枠に囲まれた音は誤変換された音となる．「客観的正解」は「得られたデータ」欄の「出力音高」と比較し，「主観的正解」は「音高の補正結果」欄と比較する．

「客観的正解」において ">" 途中で音高が変化したことを示し，"," は音高が2音にまたがることを示す．この他，短時間でピッチが大きく変動し多くの音高を含むなど場合に "..." を用いるなどしている．

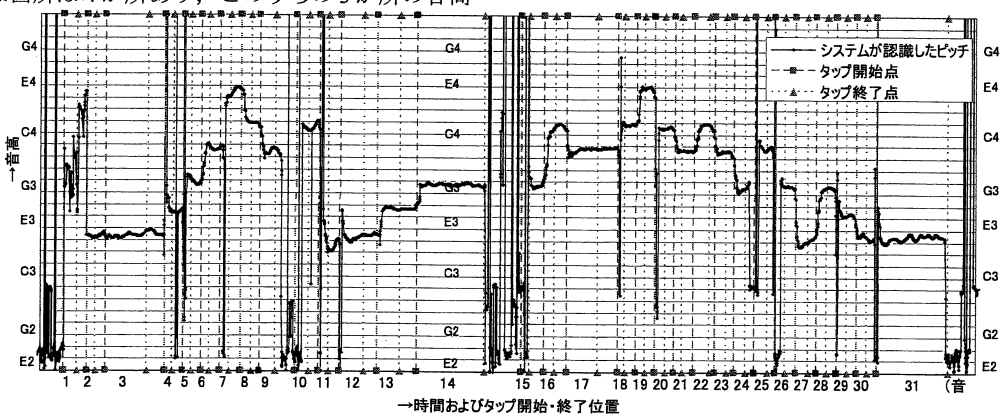

図3 被験者Aの1回目の歌唱ピッチ軌跡とタップ開始・終了点

が補正された．

個々の補正箇所について考察する．

1-2 音目については，1, 2 音目とも移動先の音高候補が存在しなかったため，補正が行われなかった．基本的に移動先の音高候補はヒストグラムの頻度 1 以上の音高が対象となるため，この箇所のように正解の音高が頻度 0 の場合は補正する手段がなくなってしまっている．

11-12 音目については，11 音目は 10 音目から大きく音高が下がるため過剰なアンダーシュートが発生し，目標の音高までピッチが上がりきらなかったことに加え，十分な時間タップし続けられていないが，12 音目の移動前音高のヒストグラム占有度が高いため，11 音目が 12 音目と同じ音高まで移動された．

14-15 音目については，14 音目の移動前音高のヒストグラム占有度が 15 音目より高いため，15 音目が 14 音目と同じ音高に移動された．

26-27 音目については，26 音目に移動先候補が存在したためである．

主観的精度は 87.1%（27 音 / 31 音）となり，全く補正が行われない場合の 77.4%（24 音 / 31 音）と比べて 10%の改善となった．

しかし，5, 6, 27 音目についてみると，補正が必要だがタップ推移情報的には矛盾がないため，提案手法では対応できない問題点がわかった．

11 音目や 26 音目は，タップ時間が短すぎるために誤変換されており，タップを用いない既存の VtoM 手法では誤変換されないとも考えられる．しかし，11 音目のようにピッチの変動が比較的大きい箇所では，まず波形処理のみで正しく 1 音として区切れるのかという疑問もある．ピッチのゆれや変動は，意識して歌唱しても起こりうる．このことから，タップを用いることにより，ピッチ変動がある場所でも 1 音の区切りを明確にし，誤変換された場合でも補正によって音高を主観的正解に導く，という利用方法が考えられる．一方で，より正確なヒストグラム作成のために，常時音量などをチェックしてタップとの同期性を高めたり，タップ時間が不十分な場合の音長（ヒストグラム作成対象となる区間）補正機構を追加したりするなどの方策もやはり必要である．

この他，この例には示されていないが，提案手法の課題として，不適切なタップ推移情報が入力された場合に必要がない箇所が誤って補正される可能性があることがわかっている．

アンケート結果について述べる．音高の上下を意識しながらの歌唱やタップの負荷は精神的負荷がどちらかというと高く，タップはどちらかというとリズム通りにはできていないと感じていた．上下をつけないタップによる入力では，それぞれ低い，どちらかというと思わないという回答であったので，被験者 A にとって提案法は負荷が増していることがわかる．

4. 結論

我々は，VtoM の主観的正解（意図通りどおりに変換されたかどうか）の精度向上に対して，音高の誤変換をタップ位置の上下移動による音高補正情報入力で解決する手法およびそれを入力するための「ジェスチャ併用型 Voice-to-MIDI」システムを提案し，評価した．

その結果，主観的精度が向上することが分かった．一方で本来必要と思われるがタップ推移情報的に矛盾がない箇所の未補正があった．また，すでに不適切なタップ遷移情報が入力されることによって誤補正される可能性あることがわかっている．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．その他，既存の補正手法との組み合わせによって相補的に精度を上げられる可能性もある．

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．

その他，現在タップしづらい速いテンポや短い音長の入力への対応などが課題である．

参考文献

1) ヤマハ株式会社：XGworks ST, http://www.yamaha.co.jp/product/syndtm/p/cmp/xgwstw/index.html.
2) 株式会社メディアナビ: 鼻歌ミュージシャン 2, http://medianavi.co.jp/product/hana2/hana2.html.
3) Lutz P., Rainer T.: An Interface for melody input, ACM Trans. on Computer-Human Interaction（TOCHI）, Vol.8, No.2, pp133-149, 2001.
4) 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.
5) Tomonari Sonoda , Masataka Goto , Yoichi Muraoka: A WWW-based Melody Retrieval System, ICMC 98 Proc., pp349-352, 1998.
6) 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.
7) 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.
8) 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.
9) 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).