6T-07

時間および音高方向に伸長した音声入力を用いたMIDIパラメータ編集方法についての検討*

八木　陽介†　　菅野　由弘†

（早稲田大学 基幹理工学研究科）

1. はじめに

MIDIを用いたDTM(Desktop Music)における音楽制作では，入力された発音のためのMIDIイベントに対し，ピッチベンドやコントロールチェンジを用いて演奏の「表情付け」が行われる．このような編集には一般的にマウスやMIDIコントローラなどの外部機器が用いられるが，それぞれのデバイスの操作方法や，MIDIにおけるパラメータ設定と実際の楽器演奏における各種奏法の対応を学ぶ必要があるなど，初学者がスムーズに意図した編集を行うのは困難である．このような問題を解決するため，音声入力によって声のニュアンスをMIDIパラメータに反映させる提案がされている[1]．しかし，メロディーに沿った歌声から抽出される基本周波数変動やエネルギー変動などの特徴をMIDIパラメータに反映させる場合，ユーザの歌唱能力や経験などによって意図した編集が行えない可能性がある．

そこで本稿では，ユーザ間の歌唱能力による差が出にくい音声入力として，時間方向および音高方向に意図的に伸長させた音声，すなわち「ゆっくりとした，大げさな音程幅のある」音声を入力として利用し，その入力波形から得られる特徴に対し適当な平滑化処理や線形変換を行った後，MIDIパラメータにマッピングすることによって前述の編集を行う手法を提案する．この手法により音声入力の利点を活かしつつ，ユーザの歌唱能力の差に影響を及ぼされにくい編集が可能となり，初学者に対して音楽的参画の間口を広げることができると考えられる．

2. システム構成

まず，編集対象となるMIDIトラックを適当な倍率で時間的に引き伸ばして再生し，それに合わせてユーザは音声入力を行う．続いて，入力音声から得られる基本周波数変動および二乗平均エネルギー変動を抽出し，音声に含まれる微細変動と考えられる成分を除去するため，カットオフ周波数を10Hzとしてローパスフィルタによる平滑化処理を行う．その後，平滑化済みの波形に対し適当な倍率で線形変換による縮小を行い，MIDIパラメータへのマッピングを行う．使用するパラメータは，擦弦楽器におけるビブラートなどの演奏表現を再現する際に用いられることの多いピッチベンド，モジュレーション(CC#1)，および演奏中の音量変化に用いられるエクスプレッション(CC#11)とし，それぞれのマッピング範囲は表1のようにヒューリスティックに定めた．但しVは発音中のMIDIイベントにおけるベロシティとする．

表1 使用パラメータとマッピング範囲

パラメータ名	最小値	最大値
ピッチベンド	-4096	4096
エクスプレッション	-64 + V	64 + V
モジュレーション	0	127

*A discussion about the method of editing MIDI parameters using voice extended in time and pitch
† Yosuke Yagi，Yoshihiro Kanno (Waseda University)

最後に，以上の操作によって得られる数値列を元に作成したMIDIイベントを，100サンプルごとに元のMIDIトラックに付与することによって，編集済みのMIDIファイルを得る．

3. アプリケーションの制作

提案手法をユーザが利用できるアプリケーションの形で実装した．図1に実行画面を示す．インタフェース部分はProcessingを，音声からの特徴抽出や波形処理にはpythonを用いた．また，基本周波数変動の抽出にはSPTK(Speech Signal Processing Toolkit)により提供されるSWIPE[2]を，二乗平均エネルギー変動の抽出にはlibrosa[3]を用いた．

スタンダードMIDIファイルを画面上にドラッグアンドドロップすることで画面上部に全トラックに含まれるMIDIイベントのノートオン，ノートオフ情報が可視化される．ユーザは画面上部のスライダを操作して時間伸長率(再生速度)を任意に設定した後，RECボタンを押すことで入力音声を録音できる．録音後は画面下部に入力音声の基本周波数変動およびエネルギー変動が可視化され，画面左上部のスライダを操作して各波形の振幅縮小率を変更することができる．このとき，ユーザはPLAYボタンを押すことで暫定的な編集結果をプレビュー再生することができ，満足のいく編集ができたと感じた時点でCONVERTボタンを押すことにより，編集結果が反映されたスタンダードMIDIファイルが保存される．

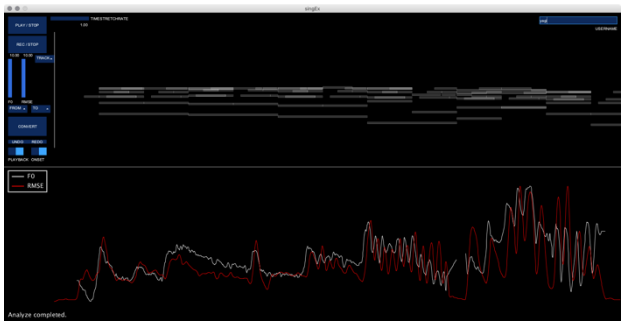

図1 制作アプリケーションのユーザインタフェース

4. 評価実験

4.1 実験概要

制作したアプリケーションを用い，次のような2種類の主観評価実験を，20歳代の男性10名を参加者として行った．

実験1. 提案手法と既存のマウスを用いたMIDI編集手法の「直感性」および編集に対する満足度の比較実験

実験2. 提案手法における時間伸長率を等倍，1.5倍，2.0倍と変化させた場合の編集に対する満足度の比較実験

マウスを用いたMIDI編集にはApple Logic Pro Xを，MIDI音源の再生にはRoland Sound Canvas VAを用いた．また，実験における編集対象として，リスト「愛の夢 第3番」の冒頭部分を打ち込みによって再現したスタンダー

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.

ドMIDIファイルを用意した．なお，主旋律はバイオリン音源，伴奏はピアノ音源を用いて再生されるように設定した．

4.2 実験1手順

まず，実験参加者に実験概要について説明した後，リファレンスとして人間の手による課題曲の演奏および，演奏表現をもたらす編集を加えていないMIDIファイルを再生した音源を聴取させた．続いて，参加者グループをA，Bの2つに分け，Aグループには先にマウスを用いた編集を行わせた後，音声入力による編集を行わせ，次のような主観評価尺度を用いたアンケートに回答させた．

(a) マウス編集と比較して，音声編集はどれくらい「直感的」な編集だと感じたか(1点～5点の5段階)

(b) マウス編集および音声編集について，編集に対する満足度はそれぞれどれくらいだと感じたか(1点～10点の10段階)

なお，本実験で編集対象とするMIDIパラメータはエクスプレッションのみとし，制限時間を3分として自由に再録音および波形の縮小率の変更ができるものとした．また，Bグループについては Aグループとマウス編集，音声編集の順番を入れ替えて同様の実験を行った．

4.3 実験2手順

実験1に続いて，提案手法における時間伸長率を等倍，1.5倍，2.0倍と変化させた場合のユーザの編集に対する満足度を比較する実験を行った．課題曲については実験1と同一とし，編集対象とするMIDIパラメータはピッチベンドのみとした．編集操作後のアンケートには次のような主観評価尺度を用いた．

(c) 3種類の時間伸長率について，編集に対する満足度はそれぞれどれくらいだと感じたか(1点～10点の10段階)

また，実験後にインタビュー形式で各編集操作に対する感想などを回答させた．

4.4 実験結果

実験1において，アンケート項目(a)の回答の平均は3.3であった． また，アンケート項目(b)の回答より得られた結果を図2に示す．Studentのt検定の結果，それぞれの編集手法の満足度について5%水準で有意差があった．

実験2において，アンケート項目(c)の回答より得られた結果を図3に示す． Studentのt検定の結果，等倍の場合と1.5倍の場合，1.5倍の場合と2.0倍の場合の間に5%水準で有意差があった．

5. 考察

実験1の結果より，音声編集はマウス編集に比べて「直感的」ではあるが，編集の満足度に関しては劣ることが分かった．これは編集をやり直す際に再度音声入力が必要であることの煩雑さなどに起因すると考えられる．

実験2の結果より，時間伸長率を適当に変化させることで音声編集における満足度が有意に向上する可能性が示された．しかし，参加者間で満足度にばらつきがあるため，編集する楽曲やユーザの需要に応じて適切な時間伸長率をユーザ自身が選択できることが望ましいと考えられる．また，実験後のインタビューでは「普段あまり使わない発声方法が必要であり，慣れるまで時間が必要と感じた」という意見があったことから，本実験において音高方向に意図的に伸長させた音声入力を困難に感じる参加者が多く，全体的な満足度の低下に繋がってしまったと考えられる．

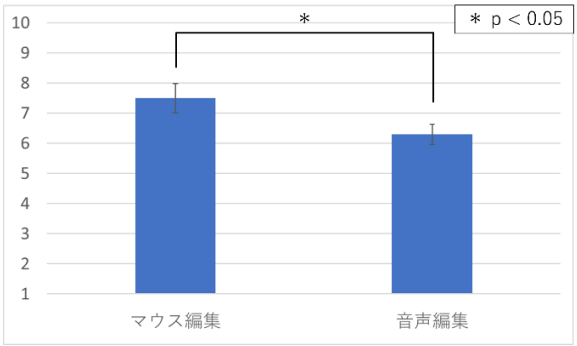

図2 各編集手法における満足度の平均

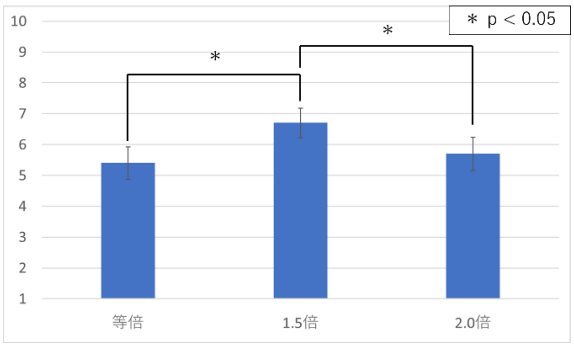

図3 各時間伸長率における満足度の平均

6. 結論および今後の展望

提案手法に基づくアプリケーションの制作およびそれを用いた実験の結果，適切な倍率で時間方向に意図的に伸長させた音声入力を用いることで，MIDI編集に対する満足度が向上することが確認された．しかし， 既存のマウスを用いた編集手法と比較して，音声を用いた編集に対する満足度はユーザの音楽経験などに依拠し，また，音高方向に意図的に伸長させた音声入力を用いた編集に関してはユーザの「大げさな音程幅」の解釈や歌唱能力に左右され，依然として個人差の大きいものであることが分かった．

それを踏まえ，今後の課題として，各ユーザに対し，歌唱能力による差をさらに軽減し，編集に対する満足度を向上させるような各伸長率などのパラメータの最適化を効率的に行う手法についての検討が必要と考えられる．

また，本システムの発展的な応用として，動的時間伸縮法などを用いた局所テンポの編集や，音声に含まれる音韻情報を利用することで，音声入力中に編集対象となるトラックを切り替えたり，MIDIプログラムチェンジに対応させることで音色そのものを変更したりするような編集を可能にすることなどが考えられる．

参考文献

[1] 螻川内努，伊藤克亘，"歌声を用いたDTM向け演奏表現パラメータの入力"，第76回情報処理学会全国大会講演論文集，2014-1，pp.299-300，2014.

[2] CAMACHO A，"SWIPE：A Sawtooth Waveform Inspired Pitch Estimator for Speech And Music"，Ph.D. Thesis，University of Florida，2007.

[3] McFee，*et al*.，"librosa: Audio and music signal analysis in python．"In Proceedings of the 14th python in science conference，pp．18-25．2015.

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.