

吟たあ：ギター型インタフェイスによる 弾弦併用型 Voice-to-MIDI システム

伊藤 直樹[†] 西本 一志[†]

ギターの MIDI フレーズ作成を行う際に発音タイミングやベロシティなどのばらつきを意識してエディットするのは手間がかかる。MIDI ギターシステムが市販されているが、ギターをある程度弾けたり、弦を押さえられる必要がある。また楽譜情報がないフレーズの場合、音高が分からず押さえるポジションの特定が難しいこともある。そこで我々は、鼻歌入力機能と市販のギター型電子楽器インタフェイスを組み合わせ、歌唱から音高を、弾弦からタイミングやベロシティを取得して、パワーコードによるギターフレーズや単音フレーズの入力(および演奏)が可能なシステム「吟たあ」を開発した。ギター上級者 1 名にシステムの試用を依頼し、評価を行った。

VOUITAR: A Voice-to-MIDI System with Guitar like User Interface

NAOKI ITOU[†], KAZUSHI NISHIMOTO[†]

VOUITAR is a Voice-to-MIDI system that employs a guitar like user interface. VOUITAR merges two kinds of input; singing voice for obtaining pitch data and picking/cutting strings for obtaining rhythm data. VOUITAR can input not only monophonic melody data but also chord backing accompaniment data: it provides a function to add the 5th note to the input note to make it a power-chord by natural string cutting performances. Using VOUITAR, the user can input his/her cool phrases as well as backings by the hot guitar-like performances only if he can sing the phrases, even if he/she does not know fret positions of the guitar. We evaluate VOUITAR with a guitarist.

1. はじめに

近年の DTM (Desk Top Music) システムの普及により、楽器を弾けない者でも様々な楽器を用いた音楽作品の制作を楽しむことができるようになった。しかし、個々の楽器にはその楽器特有の発音の特徴があり、これをシーケンサ上での編集によつて的確に再現することは極めて困難である。本研究ではギターを取り上げる。ギターには複数の弦(通常は 6 本の弦)があり、これを弾弦して発音する。演奏の際、個々の弦の弾弦タイミングや弾弦の強さは一様ではなく、通常バラツキがあり、これがギターらしさを生み出す重要な要因となっている。

このバラツキを的確かつ簡易に入力するためには、ギターそのものを演奏し、そのデータを入力することが望ましい。このため、ギターの弾弦を MIDI 情報に変換し、出力できる MIDI ギターシステムが市販されている¹⁾²⁾³⁾。しかし、これを使うためにはギターを演奏できることが前提となり、楽器を弾けない者でも音楽制作を楽しめるという DTM の最大の魅力が損なわれてしまう。

そこで本稿では、いわゆる鼻歌入力(Voice-to-

MIDI)機能と市販のギター型電子楽器インタフェイスを組み合わせ、歌唱から音高を、弾弦からタイミングやベロシティを取得することによって、ギターを演奏できない者でも簡易にギターらしいフレーズを入力できるシステム「吟たあ(英語名:VOUITAR)」を提案する。さらに、システムデザインの基礎的問題点を洗い出すために、今回はギターについてよく理解していると考えられるギター上級者 1 名に試用してもらい評価した。

2. 関連研究

鼻歌入力に関する研究は、楽曲検索分野まで含めると様々⁴⁾⁵⁾⁶⁾⁷⁾存在する。またギターインタフェイスに着目した研究もいくつか存在する。

Mountain Guitar⁸⁾は、音楽経験がなくても簡単に演奏することのできる楽器インタフェイスであり、構える高さで音の高さを指定できる、弾く真似をすることによりコード演奏が可能であるなどの特徴がある。本研究とは機構や楽器からフレーズ入力インタフェイスかの違いはあるがフレーズ入力用途への応用可能性はある。

近年 Air Guitar が流行っているが、Virtual Air Guitar⁹⁾は、画像処理を用いることにより Air Guitar のようにギターがなくても演奏している

[†]北陸先端科学技術大学院大学

[†]Japan Advanced Institute and Science of Technology

ふりをするだけで、指の位置に応じて実際に発音可能な楽器である。ギターは不要となるが、指の位置と音のマッピングを理解している必要がある。

3. 提案システムの概要

3.1 動作概要および操作方法

吟たあは、ギター型電子楽器インタフェースから出力された弾弦情報から NoteOn タイミングと On ベロシティを取得し、Voice-to-MIDI 機能により弾弦中の歌唱の音高を取得し、これらをマージして MIDI シーケンスを作成する、という動作を行う。音高を歌唱から取得するためユーザは出した音とその楽器上のマッピングを覚えずとも入力が可能となる。よって、ユーザは歌唱しながら各音の発声開始に合わせて弾弦を行えばよい。

現在下記のような入力モードがある。

[1] 単音入力モード(単弦ピッキング)

[2] 単音入力モード(カッティング)

[3] パワーコード入力モード

となる。以下、各モードごとに詳細を説明する。

[1] 単音入力モード(単弦ピッキング)

入力前にあらかじめ選択した 1 弦の弾弦情報を取得し、それに歌唱の音高を組み合わせ、単音フレーズを入力する。確実に特定の弦を弾く必要がある。

[2] 単音入力モード(カッティング)

[1]の単弦ピッキングの代わりにカッティング(複数弦を鳴らす)による弾弦を用いる。弦がカッティングされたときの最初に弾弦された弦の情報を取得し、それに歌唱の音高を組み合わせ、単音フレーズを入力する。特定の弦を弾弦しなくてよい。

どの弾弦までがひとかたまりの弾弦であるかについての判定法については、全ての弾弦がカッティングでなされることを前提として、カッティングのダウン時には弦番号が単調減少し、アップ時には弦番号は単調増加することを利用し、弦番号がひとつ前に弾弦された弦の番号と同じか不連続、あるいは単調減少→単調増加、単調増加→単調減少への移行があったときに新規の弾弦のかたまりと判断するルールとした。

[3] パワーコード入力モード

常にパワーコード(基本的にはルート音と完全 5 度上の音から成り立つギター特有の和音)フレーズのルートを取っていると仮定し、ルート音と完全 5 度上の音を入力する。ルート音の弾弦タイミングとベロシティは 6 弦のものを、5 度上の音の弾弦タイミングとベロシティは 5 弦

のものを適用する。その他の弦の弾弦情報は出力しない。

ルート音(歌唱の音高)は常に 6 弦に割り当てられるため、仮に 6 弦を弾弦せずに 5 弦だけ弾弦しても、5 弦に割り当てられる音高は歌唱されているルート音の 5 度上の音高となる。

その他以下のような機能を備えている。

[A] ピッチベンド付き MIDI データ保存機能

本システムでは 1 音毎の音高はその区間の最頻値で決定しているが、本機能は最頻値の音高を基準としてその区間の短時間ピッチの変動をピッチベンド情報に変換し、保存する機能であり、単音モード、パワーコードモードとも対応する。なおピッチが基準音高より大きく離れた場合は無視する。歌唱によるスライドなどの表情入力が可能となる。

[B] 演奏機能

弾弦後最初に E2-G5 内に存在した短時間ピッチを基準音高として MIDI NoteOn メッセージを生成し、MIDI 音源をリアルタイムで発音させることにより、本システムを楽器として用いることができる。これは単音モード、パワーコードモードとも対応している。

弾弦と歌唱タイミングが合うとは限らず、意図した歌唱音高を取得できないことがありうるので、歌唱から得たピッチ情報をピッチベンド情報にし、発音後でもある程度音程調整できるようにした。

現状では、録音機構に用いた DirectSound のレイテンシおよび FFT フレームサイズに起因する遅延の問題により 100ms 程度の発音遅延が発生する。

3.2 システムの仕様

システムの開発には Microsoft Visual C#2005 を用い、短時間ピッチ算出部は 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。入力は歌唱波形と弾弦によるノートの区切り情報およびベロシティ、出力は E2-G5 (A4 = 440Hz とする)の半音単位の音高列、つまり MIDI データとなる。ギターの音域と若干異なるのは、人間の音域を考慮したためである。入力音声は 22.05kHz, 16bit, モノラルで記録される。この他評価用に入力波形の Wave ファイル、弾弦時刻や短時間ピッチ列、各音のヒストグラムを記録している。

3.2.1 動作概要

次に単音モード/パワーコードモードに共通する入力～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 $t_{win} = 2048\text{samples}$: 約 100ms, フレーム移動間隔 $\Delta t = 128\text{samples}$: 約 6ms)による短時間ピッチ算出処理が録音終了まで繰り返される。この間に弾弦情報が入力されたら、短時間ピッチを E2-G5 の範囲で半音単位で音高ヒストグラム化してゆく。これは次に同じ弦が弾弦されるかギター型デバイスの指板上の弦スイッチや消音センサに触れる,あるいは後述する自動発音終了機構によって Note Off が発行されるまで続く(図 1)。その後、弾弦された弦に応じた MIDI チャンネル, 弾弦の強さをベロシティ, ヒストグラムの最頻音高を用いて音符情報を生成し, 出力する。これを録音停止まで続ける。

最後の音については, 何らかの理由で録音停止時に終了していなければそこで終了させる。

3.2.2 歌唱入力用マイク

歌唱の入力には, ヘッドセットマイクか環境に応じてスロートマイクを用いる。通常は PC 用に市販されているヘッドセットマイクのようなものでもよいが, 激しく弾弦すると弦を弾く音が混入したり, 演奏機能使用時にスピーカからの出力音を拾うことがあるためこのような場合にはスロートマイクを用いる方がよいと思われる。後述する実験では, マイクにはバッファローコクヨサプライ製 USB ヘッドセットマイク BMHUN01SVA を用いた。

3.2.3 弾弦情報入力装置

弾弦情報の入力には YAMAHA EG-AG¹⁰⁾を用いた。本デバイスは,

- ・ 弾弦時に弦自体から音が出ない
- ・ 音源のローカルオフ設定が可能
- ・ 弾弦情報の取得が比較的容易

という理由により採用した。

以下に吟たぁ制作に当たって我々が調べた EZ-AG の出力可能な情報や挙動について記す。

- [A] 弦ごとに独立した出力チャンネルを持つ (1 弦=1ch, ..., 6 弦=6ch)
- [B] 弦ごとに独立したベロシティを出力
- [C] ブリッジ付近にある消音センサで鳴っている全弦の Note Off を出力 (Note Off は Velocity = 0 の Note On メッセージを使用)

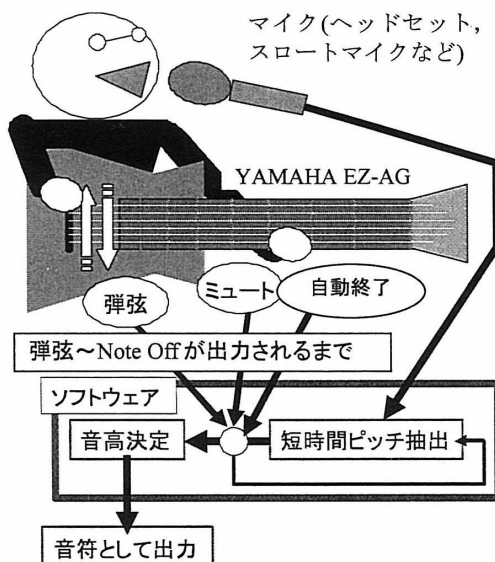


図 1. 処理の流れ

弾弦情報が入ると, 短時間ピッチのヒストグラム化開始。自動終了 or 弦スイッチ or 消音センサ or 次の弾弦により, ヒストグラムの最頻値から音高を決定する。パワーコードを出力する場合は, 5 度上の音高を付加して出力する。

- [D] 次の弾弦か弦スイッチや消音センサに触れないと Note Off が発行されない
- [E] 指板上の弦スイッチを押すと Velocity = 16 の Note On を発行される。そのときに弾弦中であれば先にその音を消音するための Note Off が発行される (今回は, 歌唱中であれば指板上の弦スイッチを押したときの Note Off による消音は行わすが, Note On は全く用いないので除去する)。

なお YAMAHA から過去市販されていた EZ-EG とはトレモロアームの有無, 弦の共振の再現の有無などの点で違いがある。EZ-EG では共振による弾弦情報が出力されるが, 本システムではそれらを除去する処理は行っていないため EZ-EG には対応していない。

3.2.4 短時間ピッチ抽出処理

短時間ピッチ算出は, 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 IFFT によって求めた循環自己相関の正の最大値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¹⁾を用いて cent 単位で音高推定を行い短時間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。

なお歌唱のような優勢な音が存在する場合、半周期の位置で自己相関が負の最大値になるため、正の最大値の出現位置は、相関が一度ゼロクロスし負側に振れた後に出現するはずである。よって、正の最大値は一度負側に振れて以降の位置から求め、それ以前に最大値があっても無視する。

波形に直流成分があると相関の値が全体的に上昇してしまいゼロクロス位置の位置がずれたり消滅するなどの影響が出るので、FFT 結果より直流成分を除いてから IFFT を行った。

循環自己相関の系列は、パワースペクトルの値を振幅を持つ各周波数の波の合成波と表せるので、直流成分を除去しても全体的に等量が減ぜられるだけで、自己相関係列の形状自体には影響しない。

3.2.5 自動発音終了機構

3.2.3 に列挙した挙動にあるように次の弾弦か弾弦中に弦スイッチや消音センサに触れないと Note Off メッセージが発行されない。よって弦スイッチや消音センサを駆使しなければ、短い音価で入力したいときでもレガートになってしまったり、最後の音が消音されないことが起こりうる。そこで、一音ごとの歌唱の終了を検知し、Note Off メッセージを発行する機構を実装している。

具体的には、本システムではピッチ抽出に循環自己相関を用いているため、弾弦後に E2-G5 の範囲に最大の正相関値がなくなれば終了と判断する。なお終了しないまま次の弾弦になった場合は、レガートと判断する。

この手法により対象とする音高範囲内に目立つ音がなければ、音量閾値などの手法を用いなくても有音/無音を判別可能となり、周期性がはっきりとした音が混じっていなければ環境音の音量変化に動的に対応したり、小音量でも判別可能となるメリットがある。

一方でこの手法では、非歌唱中でも範囲内に最大の正相関値が出現していれば消音されない可能性があるが、本システムの使用を想定している室内において、PC 内蔵マイクやヘッドセットマイク、コンデンサマイクなど数種類のマイクで調査したところ、概ね良好に作動した。しかしながら、本手法に限らず音量閾値などによる方法でもコンデンサマイクのような感度の高いマイクは避けた方がよいと思われる。

なお、弾弦と歌唱開始が同期するとは限らないため、弾弦開始～200ms までは自動終了しない。また、歌唱中にもかかわらず音高範囲内にピッチが無いと判定されることを想定し、音量

(パワースペクトルの合計値)が直前のフレームの音量の 90%以上であれば終了しない。

4. 評価実験

本システムについての評価のために実験を行った。

4.1 実験概要

3.1 に挙げた機能のうち、単音入力(単弦ピッキング)およびパワーコード入力機能について、その作業負荷や使用感などを評価する実験を行った。

- [1] 課題メロディの単音入力 (テンポ自由)
- [2] 課題メロディの単音入力 (オルタネイトピッキング&テンポ自由)
- [3] 課題メロディの単音入力 (BPM=120)
- [4] 被験者が用意したフレーズのパワーコード入力 (テンポ自由)

単音入力時の課題メロディは、作曲は負荷が高く、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」とした。

パワーコード入力に用いたフレーズは、被験者が用意した LUNA SEA「ROSIER」の冒頭などで聴くことのできる 8 小節程度のものである。被験者は、主にビジュアル系ロックや UK ロックを嗜好するギター上級者 1 名である。システムが本来想定している歌唱やリズムをとることはできるが、十分なギター演奏はできない人である。しかし、ギターは自由度の高い楽器であり、弦のマッピングなどシステム設計に無理がないかを調べ、またギター上級者ならではの工夫やコツなどの点からアドバイスをもらうため、今回はギター上級者に評価を依頼した。

実験前に EZ-AG 自体の試奏時間を設けた。当初弾弦について弾きづらいとの意見を得たが、しばらく後、慣れてきたとのことであったので EZ-AG のインタフェース自体が実験に影響を及ぼすことはなかったと思われる。

評価はアンケートによって使用感などを訊いて行った。各条件に共通するアンケートの項目を以下に示す。

[A] 精神的負荷について

(7 段階:非常に低い-低い-どちらかというと低い-普通-どちらかという和高い-高い-非常に高い)

[B] リズム通りに弾弦できたと思うかについて

(6 段階:非常に思う-思う-どちらかというと思う-どちらかというと思わない-思わない-全く思わない)

[C] 歌唱について

(7段階:非常に歌いやすい-歌いやすい-どちらかという-歌いやすい-普通-どちらかという-と歌いづらい-歌いづらい-非常に歌いづらい)

[D] 意見や感想の自由記述

またパワーコード入力については、吟たあの出力であるルート音に6弦の、5度上の音に5弦の弾弦情報を用いて生成したフレーズ(最初のフレーズ)とルート音、5度上の音とも6弦の弾弦情報を用いて生成したフレーズ(後のフレーズ)をブラインド試聴によって比較させ、以下の評価を行った。試聴に用いた音色はDistortion Guitarである。

[E] 気に入った方を選択 &

その理由を自由記述

(2択:最初のフレーズ, 後のフレーズ)

そして実験の最後に以下の2項目について自由に記述させた。

[F] 使用してみて満足な点

[G] 使用してみて不満な点

4.2 実験結果および考察

結果について実験条件ごとに述べてゆく。

[1] 課題メロディの単音入力 (テンポ自由)

[A] 負荷は[非常に低い]

[B] リズム通りに弾弦できたと[思う]

[C] [非常に歌いやすい]

[D] 「鍵盤を使ったときよりもスイッチ的感覚がなかった。」

・著者注: 被験者は過去に我々が作成したタップ併用型 Voice-to-MIDI システム¹²⁾を使用して「赤とんぼ」を歌唱したことがあり、弾弦という動作をタップの代わりに音符を区切る手段として捉えている。

被験者自身に合ったテンポであり、また普段から弾弦には慣れている部分もあるのか、歌唱と弾弦を同時に行っても余裕があると思われる。

[2] 課題メロディの単音入力 (オルタネイトピッキング&テンポ自由)

[A] 負荷は[どちらかという]と低い

[B] リズム通りに弾弦できたと[非常に思う]

[C] [歌いやすい]

[D] 「どこをアップにするか考えてしまうときがある→精神的負荷となる。」

オルタネイトピッキング (ダウンピッキングだけでなくアップピッキングも用いる奏法) に

よる弾弦では、設問[A]の結果より、条件[1]と比べて負荷が上がっている。この理由については設問[D]で述べられている通りであるが、更に被験者は「ダウンピッキングは表拍、アップピッキングは裏拍のように決めている」と回答している。よってメロディのような音価がばらつきやすいフレーズの場合、ギタリストにとっては多少混乱を起こすことがあると考えられる。

[3] 課題メロディの単音入力 (BPM=120)

[A] 負荷は[非常に低い]

[B] リズム通りに弾弦できたと[非常に思う]

[C] [歌いやすい]

[D] 特になし

メトロノームに合わせて入力したが、テンポが速くなくても条件[1]と比べて大きく悪くなることはなく、設問[B]ではよりよい評価を得た。ただし更にテンポが速くなるか、短い音価のフレーズでは負荷が増すことが考えられるため、アンケート結果は変わると思われる。

[4] 被験者が用意したフレーズのパワーコード入力 (テンポ自由)

[A] 負荷は[非常に低い]

[B] リズム通りに弾弦できたと[非常に思う]

[C] [非常に歌いやすい]

[D] 「入力できたのちちょっと不安である」

[E] 気に入ったのは[最初のフレーズ]

理由: 「雰囲気が入力しようとしていたものに近かったから。」

テンポは原曲より遅めで入力され、フレーズに細かなノートが少なかったこともあり、条件[1]と同じように良好な評価を得た。しかし入力できたのか不安を感じている。その理由を尋ねたところ、「口ずさんだフレーズの調が正しかったか、不安だったから」とのことであった。これは直接的に「吟たあ」の仕様などが影響したものではないと考えられる。しかしこのような不安を解消するために伴奏、CDなどに合わせて弾くことはありうる。この点については、スロートマイクを使用するか伴奏をヘッドホンで聴くことで解決しようと思われる。

気に入ったのは吟たあの出力したフレーズであると回答した。タイミングとペロシティが独立していることがフレーズに表情を与え、また聴感上でも知覚できたと思われる。

最後に全体的なシステムへの意見や感想について示す(原文のまま記載)。

[F] 使用してみて満足な点

「速い旋律の入りにアップピッキングが使え
るのは良い。」

実験条件[2]において、アップピッキングも用いることに若干の負荷を感じているが、同時に利点でもあると考えていると察せられる。ギターにそれほどなじみがない人は、拍の表裏を気にしないと思われるためアップピッキングの使用についての評価は異なる可能性がある。

[G] 使用してみて不満な点

「左手を動かしたくなるときがある。パワーコード入力はベースラインを口ずさんでもらうのを限定した方がよいかも。」

実は、被験者はルート音ではなく 5 度上にある音を歌唱していた。我々は必ずルート音の歌唱であると仮定していたが、選択できるような仕組みが必要であろう。

全体として、おおむね良好な評価を得たと言える。また、ギター上級者ならではの意見やアドバイスの得ることもできた。

5. 結論

本稿では、歌唱から音高を、弾弦からタイミングやベロシティを取得することによって、パワーコードによるギターフレーズや単音フレーズの入力が可能なシステム「吟たあ(英語名:VOUITAR)」を提案した。

ギター上級者 1 名に使用してもらい評価を行い、おおむね肯定的な意見を得た。また今後の開発に向け貴重なアドバイスを得ることができた。今後、本来対象と想定している歌唱やリズムをとることはできるが十分なギター演奏はできない人を対象とした評価を行う予定である。またシーケンサでエディットして作成したギターフレーズ、タップ併用型 Voice-to-MIDI システムでギターを意識して入力したフレーズを比較する予定である。

今後の予定として

- ・パワーコード入力モードにおいて、ミュート弦の弾弦も取得
- ・より弾きやすくするため、弾弦する弦を 6, 5 弦固定からそれぞれ最初に弾弦された弦および 2 番目に弾弦された弦に変更
- ・パワーコードと単音入力のシームレスな切り替えが可能となるモードの追加
- ・単音入力のために、どの弦を弾弦しても入力可能なモードの追加
- ・弾弦する弦によって上下 1 オクターブトランスポーズして入力可能なモードの追加 (女性による低音入力, 男性による高音入力, ベースパート入力など)
- ・指板上の弦スイッチの活用: スライド奏法等
- ・歌唱からのエクスペリメンテーション情報の抽出

といった機能について検討し、実装を行う。

また、現在吟たあソフトウェアの配布 (<http://www.jaist.ac.jp/~n-itou>) を予定している。

参考文献

- 1) ローランド株式会社: GK シリーズ, <http://www.roland.co.jp/GK/>.
- 2) ヤマハ株式会社: G50, <http://www.yamaha.co.jp/product/syndtm/p/cont/g1b1g50/>.
- 3) Godin: <http://www.jes1988.com/godin/index.html>.
- 4) Ghias A., Logan D., Chamberlin D., Smith S.C.: Query by humming – musical information retrieval in an audio database, Proc. of ACM Multimedia'95, San Francisco, Ca., Nov. 1995.
- 5) Lutz P., Rainer T.: An Interface for melody input, ACM Trans. on Computer-Human Interaction (TOCHI), Vol.8, No.2, pp133-149, 2001.
- 6) 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.
- 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) Junichi K., James G., Dr. Laurent M.: Mountain Guitar: a Musical Instrument for Everyone, Proc. of the 7th int. conf. on NIME2007, pp.396-397, 2007.
- 9) Karjalainen M., Mäki-Patola T., Kanerva A., Huovilainen A., Jänis P.: Virtual Air Guitar, Proc. of the 117th Audio Engineering Soc. Conv. San Francisco, CA, USA, October 28-31, 2004.
- 10) ヤマハ株式会社: EZ-AG, <http://www.yamaha.co.jp/product/epiano-keyboard/ez-ag/index.html>
- 11) 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.
- 12) 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.