

推薦論文

リアルタイム日本語歌唱鍵盤楽器のための 文字入力インタフェースの開発

山本 和彦^{1,a)} 加々見 翔太^{1,b)} 濱野 桂三^{1,c)} 柏瀬 一輝^{1,d)}

受付日 2012年6月26日, 採録日 2012年10月10日

概要: 本論文では, リアルタイムに日本語の歌詞とその音高を同時に入力して, 歌唱合成を利用した演奏をすることを可能とするリアルタイム日本語歌唱鍵盤楽器のための文字入力インタフェースを提案する. このボタン配列を用いた文字入力インタフェースでは, ローマ字入力方式とかな入力方式を組み合わせることによって非常に少ない数のボタンのみを利用して任意の日本語の文字を高速に入力することが可能である. また, 片手で操作可能なようにボタン配置が最適化され, 一般的な 12 音階の鍵盤を音高入力装置として同時に組み合わせることで, 歌唱合成に必要なパラメータの入力をリアルタイムに行うことができる. 本論では, この文字入力インタフェース単体の評価をまず行った後, 実際に歌唱演奏を行うための楽器のプロトタイプとして, 音高入力装置として広く普及している 12 音階の鍵盤と組み合わせることでリアルタイム日本語歌唱鍵盤楽器を開発しその検証を行った.

キーワード: 文字入力インタフェース, 歌唱合成, リアルタイム演奏, 電子楽器

The Development of a Text Input Interface for Realtime Japanese Vocal Keyboard

KAZUHIKO YAMAMOTO^{1,a)} SHOTA KAGAMI^{1,b)} KEIZO HAMANO^{1,c)}
KAZUKI KASHIWASE^{1,d)}

Received: June 26, 2012, Accepted: October 10, 2012

Abstract: In this paper, we present a lyric input interface for an instrument, called “Realtime Japanese Vocal Keyboard”, that enables us to perform music using singing synthesis in real-time by inputting japanese lyrics and the pitch simultaneously. The interface of button matrix for inputting lyrics, which combines “romaji” method with “kana” method, enables us to input arbitrary japanese characters using a few combination of buttons very quickly. This interface is optimized for single hand use, and by combination with common twelve-tone musical keyboard, we can input the parameters need for singing synthesis in real-time. In this paper, after evaluating this lyric input interface individually, actually we developed “Realtime Japanese Vocal Keyboard” as a prototype by combining with the common twelve-tone musical keyboard and verified it.

Keywords: text input interface, singing synthesis, real-time performance, digital musical instrument

1. はじめに

現在の音楽文化を取り巻く状況において, UGC (User

Generated Contents), CGM (Consumer Generated Media) と呼ばれる分野は非常に重要な位置を占めるようになってきている. その中でも一般のユーザやアーティストが VOCALOID [1], [2] をはじめとした歌唱合成ソフトウェアを用いて楽曲制作を行い, Web 上で発表することが

¹ ヤマハ株式会社
YAMAHA Corporation, Hamamatsu, Shizuoka 430-8650, Japan

a) yamotulp@gmail.com

b) mirrorboy@gmail.com

c) keizo_hamano@yamaha.gmx.com

d) kazuki_kashiwase@gmx.yamaha.com

本論文の内容は 2012 年 3 月のインタラクシオン 2012 にて報告され, 同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である.

さかんに行われており、歌唱合成ソフトウェアというものが我々にとって、より身近な存在として認知されるようになってきている。こうした歌唱合成ソフトウェアは歌を歌うことが得意でないユーザでも自分の楽曲に歌唱を導入することができ、コンピュータで音楽制作を行う多くのユーザにとって非常に魅力的なツールとなっている。

従来の歌唱合成ソフトウェアで音楽を制作するための手順には現状、歌詞と音高をそれぞれ入力する作曲プロセス、それをレンダリングして再生する再生プロセスの2段階のプロセスを必要とする。これは特に歌詞と音高を同時に素早く入力することが困難であるということに起因し、このため、リアルタイムには意図した歌唱演奏を行うことはできないという問題がある。こうしたリアルタイムに歌唱合成を利用した演奏が容易に可能となれば、今までオフラインでしか歌唱合成を利用した音楽活動を行うことができなかったユーザがオンライン上で直接音楽を演奏できるようになったり、NETDUETTO [3] といったネットワークを利用したリアルタイムセッションにおいて歌を歌えないユーザが歌唱合成演奏を導入したりできるようになり、音楽表現の方法に新たな自由度を与えることができ有用である。

そこで本論文では、リアルタイムに日本語の歌詞とその音高を同時に入力して歌唱合成を利用した演奏することを可能とするためのインタフェースを構築する (図 1)。特に本論文では歌詞を入力するための文字入力インタフェースにまず着目し、その開発と評価を行う。さらにその後、考案した文字入力インタフェースを用いて実際に歌唱演奏を行うための楽器のプロトタイプとして、音高入力装置として広く普及している鍵盤と組み合わせることでリアルタイム日本語歌唱鍵盤楽器を開発し、その検証を行う。

ここで、対象を日本語に限定したのは、他の多くの言語がすべての文字を平等に扱うのに対し、日本語では母音と子音というように文字の役割が完全に分離しているために文字の扱いが平等ではないため、この性質を有効に利用した入力手法の構築が期待できると考えたためである。また、日本語はモーラ言語であることから、ユーザの入力した歌詞と音高の対応付けをするうえで曖昧さが少ないとい

う特徴もあり、リアルタイムの歌唱合成インタフェースと相性が良いと判断した。

2. 関連研究

歌唱合成を実現するためのデータ入力方法としては、歌詞と音高をあらかじめ入力しておくという方法以外にも過去にいくつかの研究が提案されている。VocaListener [4] では、人間が実際に歌唱した歌データから自動的に抽出した特徴ベクトルを声色のベクトル空間にマッピングを行い、歌を歌えば自動的に VOCALOID の歌唱が出力される。しかしながら、歌唱した結果をリアルタイムに分析することはできず、生演奏には適用できないため、本研究の目的のためには利用できない。また予測変換や機械学習等による文字入力の補佐システムを利用するというアプローチがあり、マイクを用いて歌唱によって歌詞認識を行う研究 [5] や、手に装着するグローブ型デバイスを利用した指の形による日本語入力 [6]、カメラを使った指モーション認識 [7] 等があるが、これらは認識精度の点で解決しなければならない問題が多い。竹川らによる小型鍵盤を用いた文字入力インタフェース [8], [9] では、鍵盤のみを用いて片手での日本語のリアルタイム入力を実現している。これは鍵盤で分散和音を入力して、その基音となる音からの他の音の音程とその順番、さらにアーティキュレーションを利用して鍵盤入力のみで語彙を入力する方法であり、鍵盤演奏の技術を有する 12 音階の鍵盤に慣れたユーザにとって文字列を素早く入力することが可能である。しかし、リアルタイムに歌唱合成を利用した演奏を行うためには任意のリズムに合わせた入力が必要であり、こうした分散和音型の入力は、複数の鍵を決まった順番に押さなくてはならないため必然的に入力するまでの時間が長くなってしまい、ユーザにとってタイミングの制御が難しいという問題があり、本研究の目的のためには適さない。鍵盤のみを使って歌詞の入力と音高の入力を同時に行うものとしてはフォルマント兄弟の手法 [10] がある。この手法では五線譜の楽譜としてあらかじめ歌詞の文字列も含めて演奏する音符を書き起こしておくことを前提としており、演奏者は歌詞を入力しているということをまったく意識せずに楽譜どおり鍵盤を弾くことのみで歌唱演奏が可能である。この手法では五線譜を読むことのできるユーザにとっては文字を入力するための特別な鍵盤配置等をいっさい習得する必要がないという利点もあるが、必ず楽譜を用意しておく必要があり、その場でリアルタイムに歌詞やメロディをあらかじめ想定しておいたものから変化させて演奏することが困難であるという問題がある。

ここで、加藤ら [11], [12] による日本語入力をテンキーで行う試みは、ボタンのみを使う構成であり、実現が容易なほか、母音キーと子音キーが分離し、その組合せで入力するという点で、コンピュータのキーボード上でのローマ

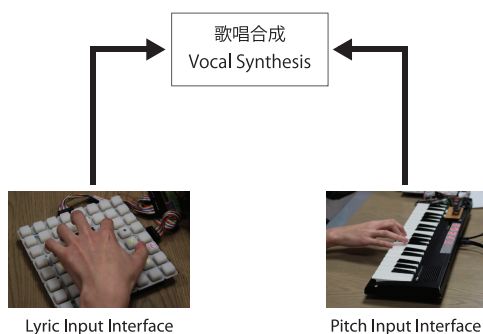


図 1 歌唱合成のリアルタイム演奏

Fig. 1 The real-time performance using singing synthesis.

字入力に慣れたユーザなら比較的理解されやすいインタフェースであるという利点がある。また片手で入力できるので、もう一方の手で音高の指定操作を行える等、歌詞を入力しながら何らかの楽器演奏をするということと親和性が高いと考えられる。加えて所望の文字を入力するためにボタンを押す順番を意識する必要がないため、発音のタイミング制御も容易である。そこで本研究では、加藤ら [11] の手法を発展させてボタン配置を手の形状にフィットさせることにより、より高速に片手で歌詞入力を行うことを可能とするインタフェースを構築する。

3. 提案手法

本論文では、日本語と音高を同時に入力してその場でリアルタイムに楽器に歌わせる・演奏することを目的としたリアルタイム日本語歌唱鍵盤楽器のためのインタフェースを開発することを目的としている。リアルタイムに日本語を歌唱させることができる鍵盤楽器を実現するためには、

- 高速な日本語入力インタフェース
- 音高入力インタフェース
- 歌唱の抑揚や声質を制御するインタフェース
- それらを処理し、合成歌唱を出力するシステム

がそれぞれ必要である。本論文では、これらの要求を満たすため、まず高速に日本語の入力が可能な歌詞入力のインタフェースを構築する。リアルタイムに歌唱演奏を行うためには音高と歌詞を同時に入力する必要があるため、歌詞は片手で入力できることが望ましい。また、非常に高速に入力が可能なインタフェースが求められる。

3.1 日本語入力インタフェース

本システムでは、目的とする歌詞をできるだけ素早く入力できるように、加藤ら [11] による日本語入力手法を改良する。日本語入力インタフェースとして入力方法の複雑さを最小限に抑えるため、可能な限り少ないボタン数で日本語の発音記号を満たすように子母音を 16 キーに絞り込む。母音ゾーンと子音ゾーンは分離して配置している (図 2)。ホームポジションとしては左手を置いた際に親指近くに母音、その他の指に子音が近くなるような配置とした。親指で 5 母音、それ以外の指で各指 2~3 子音+濁点・半濁点を担当する。これは、コンピュータのキーボードでのローマ字入力方式に慣れたユーザには子音と母音の組合せという入力方法は理解しやすく、また子音と母音キーをそれぞれ近い位置にまとめて親指を母音に、その他を子音にといった具合に指とボタンの役割を紐付けることによって、手の移動を最小限にすることができ高速な入力が可能になると考えたためである。基本的にはローマ字入力方式を踏襲しており、子音+母音を同時に押すことで目的とする文字を入力する。通常のローマ字入力方式と異なる点としては、濁音と半濁音のために別途濁点ボタンと半濁点ボタンを用



図 2 文字入力インタフェース

Fig. 2 The text input interface.

意し、濁音と半濁音入力時には、これを同時に押すこととした (例 ベ = h + ゝ + e)。そのためローマ字入力における B, D, V, G, Z, P といったボタンの削減を可能にしている。これはかな入力の濁音・半濁音を入力において、濁点・半濁点を入力する方法を踏襲しており、ボタンの数を削減することで入力スピードを向上させる狙いがある。また、本入力システムの特徴としては、母音が「あ」である発音は、「A」ボタンの母音入力をしなくても、子音のみ押した状態で発音が可能である。例外としては、日本語で頻出する「ん」の入力は「n」のみで行えるようになっている。入力が複雑になりやすい促音に関しては、複数の入力方法やローマ字入力が使われない組合せのショートカット入力を用意している (例：みゅ 通常 m + y + u / ショートカット m + h, ぴゃ 通常 h + ゝ + y + a / ショートカット h + ゝ)。このような入力方法は方式に慣れたユーザがさらに素早く入力する際に利用できる。押下ボタンと発音文字列の対応に関しては (図 13) に詳細を載せた。このようにローマ字入力とかな入力を組み合わせ、最小限の押下数で目的とする発音を行うことを可能にすることで、入力の速度・リアルタイム性の向上を狙った日本語入力インタフェースを考案した。

4. 検証

一般的に広く普及しており、PC 等に採用されているため多くの人にとって馴染みのあるフルキーボードによる入力方式と多くの携帯電話で採用されているフリック入力方式を比較対象として、考案した文字入力インタフェースの性能の評価を行った。

4.1 実験方法

マルチタッチによる入力が可能なタブレット端末上で動作する実験用のアプリケーション (図 3) を作成した。このアプリケーションではフルキーボード配列 (図 4)、フリック入力方式 (図 5)、提案手法 (図 3) の 3 種類の入

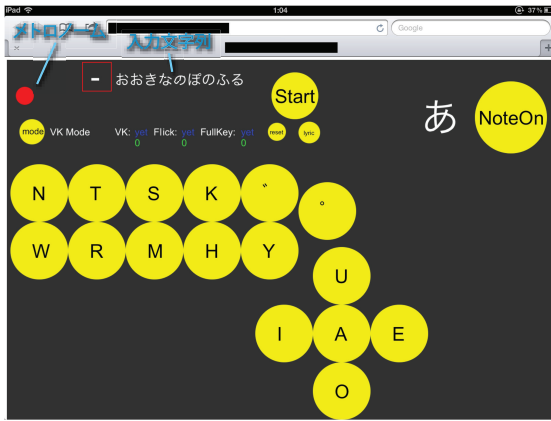


図 3 実験用テストアプリ (提案手法)

Fig. 3 The test application for the proposed method.



図 4 実験用テストアプリ (フルキーボード入力方式)

Fig. 4 The test application for full keyboard text input method.



図 5 実験用テストアプリ (フリック入力方式)

Fig. 5 The test application for flick text input method.

力方式を切り替えることができ、左手で文字をタッチ入力しながら右手で NoteOn ボタンを押すとそのときに左手で押されていた文字が入力される。たとえばフルキーボード方式では「S」と「A」を同時に押さえたまま NoteOn ボタンを押すと「さ」、提案手法では「K」と「A」のボタンを両方押さえたまま NoteOn ボタンを押すと「か」が入力され

る。フリック入力方式のみはローマ字入力ではなく所望のひらがなをフリックで1つだけ押したまま NoteOn ボタンを右手で押して文字の入力を行う。また、濁音、半濁音に関しては「わ」行ボタンの左右に独立したキー(図5)を設け、他のひらがなキーと同時に押すことで入力を行う。実験では、被験者にこの実験用アプリケーションの画面左上に表示されるメトロノーム(図3左上)の表拍のタイミングとちょうど同時になるように1拍ごとにタイミング良くテスト用文字列を入力させ、そのときの入力誤り率を測定した。メトロノームは表拍と裏拍が提示され、表拍を表すグラフィック表示は裏拍よりも大きめに提示した。入力させる文字列は数文字先まで画面上に表示されており、メトロノームに合わせて次に入力すべき文字が一番端に移動するように全体をスライドさせた。メトロノームのテンポは設定されたテスト用文字列がひとつおき提示され終わると次第に速くなっていく。テンポが速くなる直前の裏拍は次のテンポで提示されるため、その裏拍のタイミングから被験者は次のテンポを把握することができる。テンポが変化した際にはテスト用文字列は同じものが再び繰り返される。実験開始時には4拍分最初のテンポを提示してカウントダウンした後に入力を開始させた。テンポはBPM30から開始し、BPM100までBPM10刻みで増加させた。また、入力するテスト用文字列は保富康午作詞の「大きな古時計」[13]の冒頭一部分の21文字を採用した。

4.2 実験結果

20代~30代の全部で24人の被験者を対象に実験を行った。被験者には同じテスト用文字列を使って10分程度それぞれの入力方法で練習してもらった後、実験を行った。それぞれの入力方式において、入力テンポを変化させたときの入力誤り率の平均値を近似した指数曲線を(図6)に示す。近似曲線の近似式および近似式と実際のデータとのあてはまりの良さを示す重決定係数を表1上段に示す。なお、近似式の x はテンポ、 y は入力誤り率に対応し、重決定係数 R^2 は90%以上となり近似式の信憑性は高い。ここでの入力誤り率とは、メトロノームに合わせてある文字が提示される半拍前から半拍後までの間に正しい文字を入力することができなかったことと定義する。この入力誤り率に対して二元配置分散分析法により検定を行ったところ、入力方法に対する有意差($F = 20.56, p < 0.01$)を確認できた。結果ではテンポが変化してもフリック方式と提案手法は同程度の入力誤り率となり、フルキーボード方式がこの2方式と比較して明らかに低い誤り率に抑えられている。ここで、フルキーボード方式とフリック入力方式に関しては被験者のほぼ全員がふだんから慣れ親しんでいた。フリック入力とフルキーボード方式で差がついたのは、フリック入力の性質上、文字を選択するために1度ボタンを押してからスライドさせるという2段階の動作を必要とす

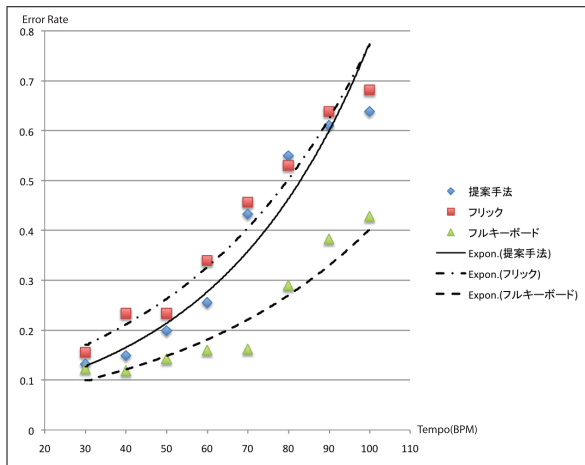


図 6 初心者の平均入力誤り率 (横軸：テンポ (BPM), 縦軸：エラー率)

Fig. 6 The mean error rate for beginners (the horizontal axis: tempo, the vertical axis: error rate).

表 1 文字入力誤り率の近似式

Table 1 The approximate expression for the rate of error of text input.

項目	提案手法	フリック	フルキーボード
初見	$y = 0.059e^{0.03x}$ $R^2 = 0.96$	$y = 0.090e^{0.02x}$ $R^2 = 0.97$	$y = 0.055e^{0.02x}$ $R^2 = 0.90$
練習後	$y = 0.015e^{0.06x}$ $R^2 = 0.94$	$y = 0.024e^{0.05x}$ $R^2 = 0.98$	$y = 0.018e^{0.06x}$ $R^2 = 0.93$

るため必然的に入力速度が遅くなってしまうことに起因しているとみられる。この実験は被験者がどの入力方式に最も慣れているかに依存する部分が大きく、それゆえに、すべての被験者が初めて操作する提案手法にとっては非常に不利な条件となっているが、初めて触れてから 10 分程度というほぼ初見の状態でもすでに習熟しているフリック入力方式には匹敵する速度で入力が可能となったことが分かる。しかし提案手法の有効性を検証するためには被験者がより長い時間の練習を重ねた後に他の入力方式と比較する必要がある。

そこで、さらに被験者の中から 6 人に 3 日間、提案手法で練習してもらった後に再び実験を行った。このときのそれぞれの入力方式において、入力テンポを変化させたときの入力誤り率の平均値を近似した指数曲線を図 7 に示す。また、近似曲線の近似式および近似式と実際のデータとのあてはまりの良さを示す重決定係数を表 1 下段に示す。この結果でも入力誤り率の平均値に対して二元配置分散分析法により検定を行ったところ、入力方法に対する有意差 ($F = 7.79, p < 0.01$) を確認できた。初見のときと比べて明らかに提案手法の入力誤り率が低下しているのが分かる。フルキーボードによる入力方式と提案手法では遅いテンポのときはほぼ等しい性能を示すが、提案手法の方が速いテンポに対して低いエラー率で入力できている。こ

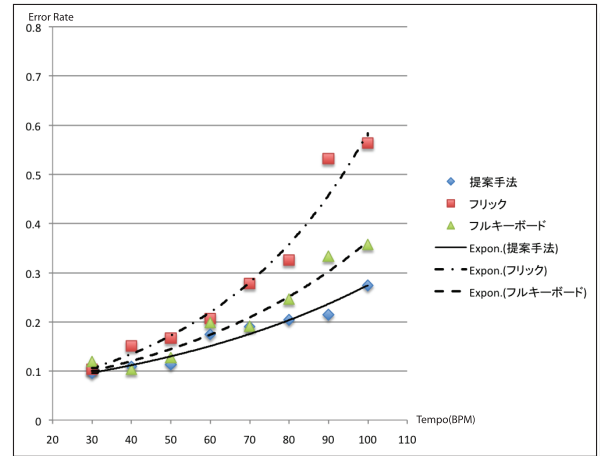


図 7 習熟者の平均入力誤り率 (横軸：テンポ (BPM), 縦軸：エラー率)

Fig. 7 The mean error rate for well-practiced (the horizontal axis: tempo, the vertical axis: error rate).

れは、フルキーボードの配置が両手でタイピングすることを前提として設計されているため、片手では文字を選択するときの手の移動量が多くなってしまいうためであると考えられる。この実験によって、ユーザにとっては 1 度方式に馴れてしまえば、提案手法に優位性があることが検証できた。

5. プロトタイプ実装

実際に考案した文字入力インタフェースと鍵盤を組み合わせ、リアルタイムに歌唱演奏を行うためにリアルタイム歌唱鍵盤楽器のプロトタイプ (図 9) の開発を行った。このプロトタイプは、本論文で提案した歌詞入力インタフェース、広く普及し利用者も多い 12 音階の鍵盤、歌唱の抑揚や声質を制御するインタフェースとしてのツマミ、それらから歌唱を実際に合成し出力する音源部から成り立つ。本論文では、歌唱合成のエンジンとして VOCALOID-board (図 8) を利用する。これは MIDI のシステムエクスクルーシブにて音高と文字、歌唱パラメータを入力することにより、リアルタイムに歌唱合成を行うことのできる組み込みボードである。システムの I/O 処理には ArduinoMega2560 [14], 日本語入力インタフェース部に tkrworks パッドキット [15], 鍵盤部分に CBX-K1 [16], 文字列表示部分に 8 × 8 マトリクス LED × 7 個を利用した (図 10)。動作の流れとしては、文字入力インタフェースから 1 発音分の文字を入力し、鍵盤で音高を指定、それらの信号を Arduino 上で VOCALOID-board へ入力できるように MIDI の Sysex に変換し送信、VOCALOID-board 上で音声合成し出力する。LED では現在発音している文字とすでに過去に発音した文字列を表示し、ユーザが入力している文字をフィードバックする。これは、VOCALOID-board の歌詞の発音は明瞭でないこともしばしばあり、ユーザへのフィードバック



図 9 リアルタイム歌唱鍵盤楽器のプロトタイプ
 Fig. 9 The prototype for Real-time Vocal Keyboard.



図 8 VOCALOID-board
 Fig. 8 VOCALOID-board.

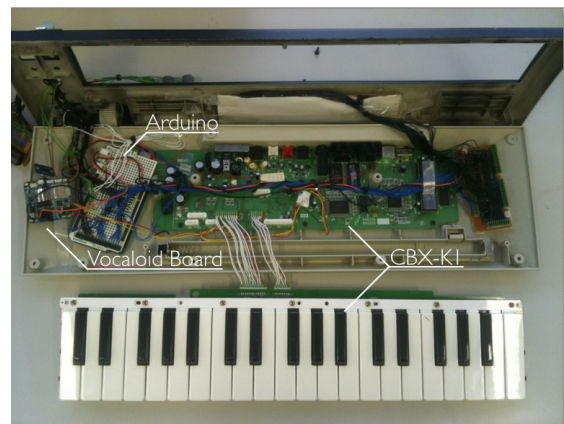


図 10 プロトタイプ内部
 Fig. 10 The inside of our prototype.

クとして音だけでは分かりにくい場合があるため、それを補助する役割を持っている。また、ユーザが歌詞入力インタフェースにおいてブラインド入力を練習する際に、自分の入力した文字の正誤が判断できなくてはならない。そのため自分の入力を確認するために必要となる表示装置でもある。一方、歌唱合成をするためのパラメータは歌詞と音高以外にも数種類存在するが、その中でもビブラートの深さ、音声の明るさ、フォルマント等をつまみによって変更できるようにし、自分の好みの声質を作り込むことができるようにした。LED への発音文字の表示やつまみからの入力も Arduino からの制御で行っている。実際のユーザの演奏方法としては、任意の文字を歌詞入力インタフェースを左手で押さえて入力したうえで、右手の音高入力インタフェースである鍵盤で音高を指定し発音を行う (図 11)。

音高の入力装置としてすでに一般的に普及している 12 音階の鍵盤を利用したのは、本論文で提案する歌詞入力インタフェース単体の有効性の検証を行うためには音高入力装置においてすでに多くの被験者が演奏技術を持ち合わせているインタフェースを採用し、練習においては歌詞入力インタフェース自体の習得に集中してもらうことが適していると判断したためである。実際の発音タイミングは鍵盤が押されたタイミングとなり、この時点で歌詞入力インタ



図 11 演奏方法
 Fig. 11 The play style.

フェースにおいて左手で押さえられていたボタンに対応した歌詞を発音する。たとえば、左手で「あ」を押したまま鍵盤を連打すると「あ」が何度も連続して発音され、逆に鍵盤を押したまま左手の文字を「か」変えても再発音はされず、次に鍵盤を押なおしたタイミングから発音文字が「か」に変更される。これは日本語がモーラ言語であるからこそ可能なアプローチである。

6. 評価と考察

開発したリアルタイム歌唱鍵盤楽器のプロトタイプを使って、実際に歌唱の演奏が可能かどうかを検証した。

まず、幾人かのアマチュアの鍵盤演奏経験を有する被験者に試奏してもらいその様子を観察した。歌詞入力しながら鍵盤による音高入力を同時に行うため、鍵盤楽器の演奏技術をまったく有さない被験者であると、鍵盤と歌詞入力インターフェースのそれぞれの有用性を分離して検証できない。そのため今回は鍵盤楽器の演奏経験を有する被験者を対象とした。

6.1 鍵盤演奏中級者の結果

鍵盤楽器の演奏歴が10年の被験者に本システムを使って練習・実演奏を行ってもらった結果では、3時間程度の練習で日本語入力が無理なくできるようになり、最終的には、簡単な童謡をスムーズに演奏できるようになった。また、鍵盤楽器の演奏歴が3年の被験者にも体験してもらった結果、最初に触れてから2時間ほどで、最終的に1曲オリジナル曲を作曲し演奏することが可能になった。歌詞入力のリアルタイム性についても、あまりテンポの速い曲を演奏するのでもなければ十分に対応できていた。また、まったく楽器の演奏経験がない人に比べ、鍵盤演奏経験者は、指それぞれに対して役割を分けて楽器を演奏することに慣れており、提案した親指に母音、その他の指に子音という役割を割り当てる歌詞入力インターフェースと非常に親和性が高いことが実際の演奏している姿から見て取れた。このことから、ある程度の鍵盤演奏経験があれば本システムは特に習熟が困難でなく、特に高い演奏技能を持ち合わせていなくても遅いテンポの曲であれば歌詞を入力しながら歌唱合成を使った演奏を行うことができることが分かった。

6.2 鍵盤演奏上級者での検証

次に、プロのキーボーティストにも2週間ほど、楽器を貸し出して練習を重ねてもらい、実際にライブでの演奏に使用してもらった。この演奏では非常に速いテンポの曲においても高速に歌詞を打ち込みながら音楽的な歌唱演奏ができており、演奏者の熟練によって高度な歌唱の演奏も可能であることが分かり、楽器としての可能性の高さを確認することができた。演奏者からは、子音の発音が鍵盤を押してから開始されるため、特に「さ」行では前のめりに演奏しないと音が遅れて聞こえてしまうという問題が指摘された。しかし、これは実際の人間が歌唱する場合にも無意識に演奏技術として行われていることで、演奏者の熟練によって解決していく問題とも考えられる。また、現状のプロトタイプでは、歌詞と音高以外の歌唱のパラメータについてリアルタイムに操作することがまったく考えられていないが、その重要性を指摘された。



図 12 展示、演奏風景

Fig. 12 The exhibitions and performances scene.

6.3 ユーザの意見・要望

2011年から2012年にかけて国内で複数回の展示・演奏を行い体験者に意見をいただいた。男女ともに幅広い体験者に興味を持っていただくことができた。鍵盤を使ってリアルタイムに任意の歌詞で歌唱合成を使った演奏ができる楽器は過去に存在しないため、非常に驚くユーザも多かった。そのような中で実際にユーザから得た意見、要望を以下に示す。

(1) 左手での日本語入力インターフェースのボタン配置の改良。

多くのユーザからより手にフィットした押しやすいボタン配置にしてほしいとの要望があった。現状の日本語入力のボタンは、実装の都合上から子音ボタン群が直線的に配置されているため、手を大きく広げないと入力の難しい文字もある。これは手を丸めたときによりそれぞれの指の位置に近い位置にボタンが配置されるように丸みを帯びた配置にすることが解決策として考えられる。ユーザの手の大きさによっても、ボタンの組合せの押しやすさは影響されるため、この配置に関してはマルチタッチのタブレット等で再び実験を行い最も最適なものを検討していく必要がある。ただし、このような意見の背景には、既存の日本語入力インターフェースとまったく違った入力方式・形状をしているため、展示会等で少し触れただけでは、慣れるというレベルまで到達することが難しいといった事情も考慮して対応する必要がある。

(2) 文字入力をタッチパネルを使ったフリック入力方式にしてはどうか。

日本語入力に対してタッチパネルを使ったフリック入力方式にしてはどうかといった意見が多かった。しかし、現在のタッチパネルは実際に手が触れてからソフトウェアまで信号が到達するまでのレイテンシが非常に大きく、リズムカルな演奏表現には不向きである。またフリック入力、

わ	ら	や	ま	は	な	た	さ	か	あ
w / wa	r / ra	y / ya	m / ma	h / ha	na	t / ta	s / sa	k / ka	a
うい	り	ゐ	み	ひ	に	ち	し	き	い
wi	ri	yi	mi	hi	ni	ti	si	ki	i
う	る	ゆ	む	ふ	ぬ	つ	す	く	う
wu	ru	yu	mu	fu	nu	tu	su	ku	u
うえ	れ	いえ	め	へ	ね	て	せ	け	え
we	re	ye	me	he	ne	te	se	ke	e
を	ろ	よ	も	ほ	の	と	そ	こ	お
wo	ro	yo	mo	ho	no	to	so	ko	o

ん
n

ば	ば
h ^o	h ^o
び	び
h ^o i	h ^o i
ぶ	ぶ
h ^o u	h ^o u
べ	べ
h ^o e	h ^o e
ぼ	ぼ
h ^o o	h ^o o

だ	ざ	が
t ^o /t ^o a	s ^o /s ^o a	k ^o /k ^o a
でい	じ	ぎ
t ^o i	s ^o i	k ^o i
づ	ず	ぐ
t ^o u	s ^o u	k ^o u
で	ぜ	げ
t ^o e	s ^o e	k ^o e
ど	ぞ	ご
t ^o o	s ^o o	k ^o o

ヴ
u ^o

りゃ
ry / ria
りゅ
ryu / rh / rwu / riu
りえ
rie / rm
りょ
ryo / rho / rw / rio

みゃ	ひゃ	にゃ	ちゃ	しゃ	きゃ
my / mia	hy / hia	ny / nia	ty / tia	sy / sia	ky / kya/ kha / kma / kwa / kia
みゅ	ひゅ	にゅ	ちゅ	しゅ	きゅ
myu / mh / mwu / miu	hyu / hwu / hiu	nyu / nh / nwu / niu	tyu / th / twu / tiu	syu / sh / swu / siu	kyu / kh / khu kmu / kwu / kiu
みえ	ひえ	にえ	ちえ	しえ	きえ
mie	hm / hie	nm / nie	tm	sm / sie	kye / khe / km / kme / kwe / kie
みょ	ひょ	にょ	ちょ	しょ	きょ
myo / mho / mw / mio	hyo / hw / hio	nyo / nho / nw / nio	tyo / tho / tw / tio	syo / sho / sw / sio	kyo / kho / kmo / kw / kwo / kio

てい
tei

びゃ	びゃ	ふぁ	ぢゃ	じゃ	ぎゃ	ヴぁ
h ^o y / ^o / ia	h ^o y / ^o / ia	hua	t ^o y	s ^o y / s ^o ia	k ^o y / k ^o ha / k ^o ma / k ^o wa / k ^o ia	u ^o a
びゅ	びゅ	ふい	どゅ	じゅ	ぎゅ	ヴい
h ^o yu / ^o u / iu	h ^o yu / ^o u / iu / h ^o iu	hui	t ^o yu / t ^o h / t ^o wu	s ^o yu / s ^o h / s ^o wu / s ^o iu	k ^o yu / k ^o h / k ^o wu / k ^o iu	u ^o i
びえ	びえ	ふえ	どえ	じえ	ぎえ	ヴえ
^o m / ^o ie	^o m / ^o ie / h ^o ie	hue	t ^o m	s ^o m / s ^o ie	k ^o m / k ^o ie	u ^o e
びょ	びょ	ふぉ	どょ	じょ	ぎょ	ヴぉ
h ^o w / ^o o / io	h ^o w / ^o o / io / h ^o io	huo	t ^o yo / t ^o ho / t ^o w	s ^o yo / s ^o ho / s ^o w / s ^o io	k ^o yo / k ^o ho / k ^o w / k ^o io	u ^o o

図 13 ボタン-かな対応表

Fig. 13 Button - Kana table.

本論文での検証からも明らかのように、押してからスライドさせる、という2段階の操作を要求するため、高速な文字入力を行うためには限界がある。

(3) ボタンを離れたタイミングによる文字のミスが多い。提案した文字入力インタフェースでは、右手で鍵盤を押していた際に左手で押さえておいた文字を発音する、というアプローチを採用しているため、ちょっとしたタイミングのずれで母音と子音のボタンのどちらかを次の文字の入力のために鍵盤を押して発音するより先に離してしまうと誤った文字が発音されてしまうという問題が頻発した。これに関しては意図して押さえていた組合せの文字がどちら

かのボタンを離れた後もしばらく残留するような仕組みを取り入れていくことも検討していく必要がある。

(4) 文字の事前入力をしたい。

事前に文字列を入力しておき、鍵盤を押すことで保存されていた歌詞データに1音1音音高をつけて演奏する方式の方が良いのではないかという意見が多かった。これに関しては、過去に「あいうえお VOCALOID ボード」[17]という試作があり、より気軽に VOCALOID を発音させる体験が得られることが確かめられている。しかし、この手法ではある1文字を保ったままのレガート奏法（たとえば1度「か」をある音高で発音させておき、さらに「か」の子音部

分を再発音させないまま異なる音高に連続的に移る) や、文字列をその場で思いついたメロディに合わせて追加、変更させたいときに対応できないという問題がある。さらに、一般的な歌唱楽曲においては演奏時にリアルタイムに作詞をするという状況は稀であり単純に既存の楽曲を楽譜どおり演奏するだけであるならば事前入力方式は非常に有効ではあるが、実際の演奏時にユーザがミスをした場合にまったく対応できないことや、2段階の操作をユーザに要求することにより音楽的な即興性を損ねてしまうという問題がある。事前入力方式と提案手法には双方それぞれ利点があるが、事前に入力した歌詞とその場で鍵盤で弾いた音高との対応付けをユーザがリアルタイムに制御できるようにすると同時に新たな文字列も事前入力した文字列の間にリアルタイムに追加していくことができる、といったハイブリッド方式を構築していくことも今後検討していく必要がある。

7. まとめ・今後の課題

本論文では、リアルタイムに歌詞と音高を同時に入力して歌唱合成を使った演奏を行うことを可能にするインタフェースとしてのリアルタイム日本語歌唱鍵盤楽器のプロトタイプ開発を行った。特に歌詞を高速に片手で入力するためのインタフェースを考案し、その評価を行った。また、考案した歌詞入力インタフェースと12音階の鍵盤を組み合わせて試作を行い、数人の被験者に対して実際に演奏してもらいその有効性を確認した。今後の課題としては、より高速でリズムカルな歌詞入力を可能にするための改良を行っていく。また、本論文では歌詞と音高以外の歌唱のパラメータをリアルタイムに操作することに関してはいっさい検討されていない。特に今回の手法を用いると、歌詞と音高入力で両手が塞がれている状態になるため、歌唱のパラメータはツマミのようなインタフェースでは操作することができないという問題がある。こうしたパラメータも含め、強弱、抑揚等の歌唱の微妙なニュアンスを歌詞と音高を入力すると同時にリアルタイムに演奏するインタフェースを今後構築していく。

参考文献

- [1] ヤマハ株式会社：VOCALOID™，入手先 (<http://www.VOCALOID.com/>).
- [2] 剣持秀紀：歌声合成とその応用，情報処理，Vol.50, No.8 (Aug. 2009).
- [3] ヤマハ株式会社：NETDUETTO™，入手先 (<http://www.y2lab.com/project/netduetto/>).
- [4] 中野倫靖，後藤真考：VocaListener：ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案，情報処理学会音楽情報科学研究会研究報告，Vol.2008, No.50, pp.49-56 (May 2008).
- [5] 伊藤直樹，西本一志：歌詞歌唱による入力可能な Voice-to-MIDI 手法の提案，情報処理学会シンポジウム論文集，Vol.2007, No.4, pp.71-72 (2007).

- [6] 福島大志，宮崎文夫，西川 敦：指文字入力インタフェース「Fingual」の開発，情報処理学会インタラクティブ2011 (2011).
- [7] 西田好宏，鈴木雅実：VecType：動きベクトルを利用した携帯電話文字入力システムの提案，情報処理学会インタラクティブ2005 (2005).
- [8] 竹川佳成，寺田 努，塚本昌彦，西尾章治郎：歌唱機能をもつ黒鍵追加型小型鍵盤楽器モバイルクラヴィア IV の設計と実装，情報処理学会研究報告（音楽情報科学研究会），2004-MUS-57, Vol.2004, No.111 (Oct. 2004).
- [9] 竹川佳成，寺田 努，西尾章治郎：鍵盤奏者のための小型鍵盤を用いた文字入力インタフェースの構築，情報処理学会論文誌，Vol.49, No.4, pp.1234-1244 (2008).
- [10] 三輪真弘，佐近田展康：フォルマント兄弟，入手先 (<http://formantbros.jp/j/top/top.html>).
- [11] 加藤善也，北上義一：複合入力処理によるパーム型キーボードの実験，ヒューマンインタフェース 60-3 (1995).
- [12] 加藤善也，北上義一，酒井靖夫，小野充一：テンキー型片手打鍵文字入力方式の実験と評価，ヒューマンインタフェース 82-1 (1999).
- [13] 保富康午：大きな古時計，入手先 (<http://ja.wikipedia.org/wiki/大きな古時計>).
- [14] Arduino: Arduino Mega 2560, available from (<http://arduino.cc/>).
- [15] (有) 山本製作所：PICnome，入手先 (<http://atelier.tkrworks.net/doc/about-picnome>).
- [16] ヤマハ株式会社：CBX-K1，入手先 (<http://jp.yamaha.com/products/music-production/midi-controllers/cbx-k1/?mode=model>).
- [17] ヤマハ株式会社：歌唱合成制御装置および歌唱合成装置，特開 2012-83569.

推薦文

インタラクティブ2012では、87名から構成されるプログラム委員会によって投稿数43件の中から優秀な論文18件を一般講演発表として採択し、インタラクティブ発表は149件の投稿から19件をファイナリストとして選出いたしました。本論文は、これらの37件からさらにプログラム委員会による投票によって、論文誌に推薦すべき論文であるとの評価を得たものであり、論文誌編集委員長としてもぜひ推薦したいと考えました。

(インタラクティブ2012プログラム委員長 宮下芳明)



山本 和彦 (正会員)

2008年九州大学芸術工学部音響設計学科卒業。2010年九州大学大学院芸術工学府コミュニケーションデザイン科学コース修了。2010年ヤマハ株式会社にて音楽情報処理の研究開発業務に従事。現在グラフィックス、ユーザインタフェース、数値シミュレーション、音響信号処理に関する研究を行う傍ら、メディアアーティスト Yamo として活動。日本音響学会、IEEE、ACM 各会員。



加々見 翔太

2007年千葉大学工学部情報画像工学科卒業。2010年慶應義塾大学大学院政策・メディア研究科修了。クロスシンセシスを用いたDJ向けエフェクター、マルチエージェントエフェクターの研究開発を行う。2008年任天堂「Re:コエティスト」ディレクター。2010年ヤマハ株式会社入社。2012年同社を退職後@mirrorboy/コバルト爆弾αΩとしてDJ, 楽器・アプリ開発, イベント企画, 自宅警備に従事。



濱野 桂三

2004年静岡大学工学部電気電子工学科卒業。2006年静岡大学大学院理工学研究科電気電子工学コース修了。2006年ヤマハ株式会社入社。電子楽器のアナログ・デジタル回路等のハードウェア設計業務に加え, ソフトウェアプラットフォーム開発業務に従事。



柏瀬 一輝

2010年多摩美術大学美術学部生産デザイン学科プロダクトデザイン専攻卒業。2010年ヤマハ株式会社デザイン研究所。楽器, 音響機器のプロダクトデザイン業務に従事。