

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 27 日現在

機関番号：23703

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23520175

研究課題名(和文)MIDIアコーディオンによる合成音声の発話及び歌唱の研究

研究課題名(英文)Research of speech and singing voice synthesis by MIDI accordion

研究代表者

三輪 眞弘(MIWA, MASAHIRO)

情報科学芸術大学院大学・メディア表現研究科・教授

研究者番号：20336647

交付決定額(研究期間全体)：(直接経費) 3,600,000円、(間接経費) 1,080,000円

研究成果の概要(和文)：本研究は、MIDIアコーディオンをインターフェースとして用いた実時間「演奏」による人工音声発話システムの開発と、それを前提とする新しい「歌謡」の作曲、そこで生成される「誰のものでもない声」の哲学的考察を統合的に行うものである。その成果として、多言語に対応した音素制御のための規格「兄弟式国際ボタン音素変換標準規格(BBPTSI)」を発表し、その規格に基づいた楽曲「夢のワルツ」が作曲された。さらに、任意の旋律線(連続的音高変化)を、微分音高制御のために開発された「和音平均化アルゴリズム」を用い、制約プログラミングの手法によってピアノ鍵盤用楽譜として記譜し近似する研究も行われた。

研究成果の概要(英文)：This research is for the development of an artificial voice speech system via a real-time "performance" that utilizes a MIDI accordion as an interface, the composition of a new "song" with that system as a prerequisite, and synthetically conducts a philosophical examination of the "ownerless voice" generated in that "song". As a result of that research, we presented the "Brother's Button-to-Phoneme Transfer Standard for International languages (BBPTSI)", a standard for phoneme control that corresponds to multiple languages, and was able to compose "Yume no Waltz", a musical composition that is based on that standard. In addition, using the "Chord Avarave Algorithm", which was developed for microtonal pitch control, we developed a method to transfer from an optional melody line (continuous pitch change) to a musical score for a piano keyboard via a constraint programming technique.

研究分野：人文学

科研費の分科・細目：芸術学・芸術史・芸術一般

キーワード：フォルマント合成 電子音楽 音素変換標準規格 人工音声制御 ヒューマンインターフェース 楽器論 メディアアート 演奏身体論

1. 研究開始当初の背景

1950年代以降、コンピュータによる音声合成技術については、膨大な数の研究が蓄積されて来た。機械の声で「喋らせる」技術については、近年では極めて高品位で自然なイントネーションによる音声合成が実現し、さまざまな情報伝達の場面で広範囲に実用化されている。一方、機械の声で「歌わせる」技術はそのスタートこそ遅かったものの、2007年以降の初音ミク等 VOCALOID ブームにより、今日ではひとつの文化現象と呼べる規模の流行を見せている。

しかし、これら現行の「喋らせる／歌わせる」技術すべてに共通しているのは「テキスト読み上げ (Text-to-Speech = TTS) の概念である。つまり、あらかじめ用意されたテキストを音声合成システムに入力することによって、事後的に (非リアルタイムに) 音声を生産するという方式だ。音声合成技術の出発点として歴史的に評価される VODER (1939年) は、人間が操作することでリアルタイムに発声するアナログ方式スピーチシンセサイザーであった。しかし、これ以降リアルタイム音声合成に目立った技術的進展はなく、TTS の設計思想に基づく非リアルタイム音声合成のみが進歩して来た。

しかし、人間の音声コミュニケーションの最も重要な特徴は、話し／歌う主体が、発声行為の途中であっても、他者の反応を含んだ状況変化に応じて、次の発声を刻々と決定する「リアルタイム性」にある。これが、音声コミュニケーションに微妙な感情／情動表現やニュアンスを与えている。これまで音声合成研究において棚上げされて来たこのリアルタイム性こそ、人間／機械の関係性を新しいフェーズに移行させるためのキーになる概念だと我々は考える。とりわけ音楽等の芸術表現においては、リアルタイム音声合成が新たな表現の可能性を直接的に指し示していると考えられる。

2. 研究の目的

本研究は、合成音声の発話および歌唱をリアルタイムに制御する新しいヒューマン・インターフェイス・システムの開発を目的としている。

これを考えるに当たって問題になるのが、音声のリアルタイム生成に必要なパラメータ操作の数の多さである。最低でも音声のピッチ (高さ) ラウドネス (大きさ) 音素の3つの要素を同時に制御しなければならない。とりわけ音素の制御については、特定言語 (例えば日本語) に含まれるすべての音節から任意のものを瞬時に選択できるようにしなければならない。

そこで注目したのが、既存の楽器インターフェイスである。なぜなら、既存の楽器は、長年の歴史のなかでヒューマン・インターフェ

イスとしての改良が重ねられ、記譜法、教授法、練習法などが確立している。そして、何より訓練を受けてその楽器を演奏できる人々がすでに多数存在している。まったく新しい音声合成用のインターフェイスをデザインすることは可能であろうが、その習得が極めて時間のかかる困難な作業であることは容易に想像できる。既存の楽器インターフェイスを活用することは、こうした歴史的な文化資源を応用しつつ、合成音声のリアルタイム制御の実現に現実的な解決を与える発想である。

すでに三輪と佐近田は、2003年よりピアノ式鍵盤を用いた合成音声のリアルタイム制御を研究／開発／実現して来た。このノウハウを応用し、今回研究においては「アコーディオン」という、より直観的・身体感覚的・構造的に発声行為に近い楽器をインターフェイスとして用い、合成音声の発話および歌唱をリアルタイムに制御するシステムの開発を目指した。

3. 研究の方法

研究の具体的な方法として、次の3つの側面を進めた。

(1) インターフェイス (ハードウェア)
インターフェイスとなるアコーディオンは、ROLANDが開発した MIDI 対応電子アコーディオン「V-Accordion」シリーズの FR-1 (右手ピアノ式鍵盤タイプ) および FR-1b (右手ボタン・タイプ) モデルを購入し用いた。この電子楽器は、ボタン・鍵盤・蛇腹などの演奏情報を MIDI メッセージとしてリアルタイムに送信する機能を持つ。



図1 ROLAND V-Accordion FR-1 (左) と FR-1b (右)

(2) 音声合成ソフトウェア
合成音声をリアルタイムに生成するソフトウェアは、すでに研究分担者の佐近田が制作していたものを改良する形で使用した (開発プログラミング環境: Max/MSP)。また後述する通り、アコーディオンの各部操作を合成音声のパラメータにマッピングする規格も、同プログラミング環境により実装し、音声合成ソフトウェアに追加した。

(3) 演奏家によるモニタリングおよび公開実演
研究協力者として岡野勇仁氏 (ピアニスト) 長坂憲道氏 (アコーディオン奏者) の協力を

仰ぎ、実技テストによるインターフェイス・システムのモニタリングを継続的に行い、演奏しやすさ、習得しやすさ、運指法、発声の表情・表現力など多面的な評価を得ながら、システムの改良を重ねた。また、両氏による公開での実演演奏を計7回行い、一般観客ならびに美学者、工学研究者、音楽学者、作曲家など有識者からの多数の評価を得た。

4. 研究成果

(1) MIDI アコーディオンの解析

まず初年度に MIDI アコーディオン (ROLAND V-Accordion FR-1/FR-1b) を購入し、鍵盤、ボタン、蛇腹等の各部操作により出力される MIDI メッセージを解析した。製品の仕様で、左手ボタンにアサインされた MIDI ノートメッセージに一部重複があり、受信側からは全ボタンを一意に識別できない問題に直面し、対応に工夫を要した。そのうえで、次の3つの基本方針のもとインターフェイスの規格を定義することとした。

- 右手鍵盤の操作：合成音声のピッチ制御
- 左手ボタンの操作：音素ラベルの指定
- 蛇腹の操作：合成音声のラウドネス制御

(2) 「左手ボタン / 音素ラベル対応マッピング規格」の考案 / 実装

本研究において、MIDI アコーディオンの左手ボタンに言語音素ラベルを合理的にマッピングする規格を2種類考案し、音声合成システムに実装した。

「兄弟式日本語ボタン音素変換標準規格 (BBPTSJ)」

最初に考案・実装したこの BBPTSJ 規格の特徴は、「ひとつのボタンを押せばひとつの音節を発声する」という指針のもと、日本語のカナ文字五十音表の並びを基本として、MIDI アコーディオンのすべての左手ボタンに、日本語音節 (モーラ) 単位の音素ラベルをアサインする考え方に基づいている。濁音、拗音、/n/等の音節は、起点となる任意のボタンと隣接あるいは周辺ボタンのコンビネーション操作により合理的に指定できるようにした。

この規格を整理した PDF ドキュメントを web サイトにて公開している。

http://formantbros.jp/BBPTSJ_v2.pdf



図2 兄弟式日本語ボタン音素変換標準規格 (BBPTSJ)のボタン配列 (シングルボタン)

この BBPTSJ 規格では、研究当初の目標である「すべての日本語音素をアコーディオンのボタン操作で指定すること」を実現しており、ボタン配列にも合理的な規則を適用して、覚え易さに一定の配慮をしている。しかし、実際の演奏 (操作) では、ボタンを押す指の大きな跳躍が避けられず、インターフェイスの身体性の面で改善の余地を残したものとなった。

「兄弟式国際ボタン音素変換標準規格 (BBPTSJ)」

前 BBPTSJ 規格の反省をふまえ、まったく新しい発想のもとに考案したのが「兄弟式国際ボタン音素変換標準規格 (BBPTSJ)」である。最大の特徴は、これまでのカナ文字ベースの音節 (モーラ) 単位から、それを構成する音素単位へと発想を転換した点にある。具体的には、左手ボタンに子音音素だけをマッピングし、それを押した時に動的に発生する母音エリアから母音音素を選択する。つまり音節 (モーラ) を、一貫して子音ボタン + 母音ボタンのコンビネーションによって指定する方式である。

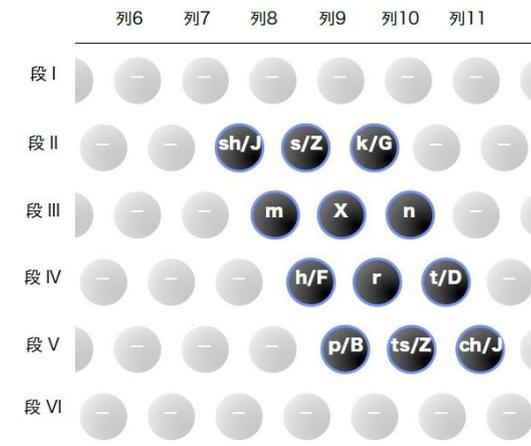


図3 兄弟式国際ボタン音素変換標準規格 (BBPTSJ)の子音ボタン配列

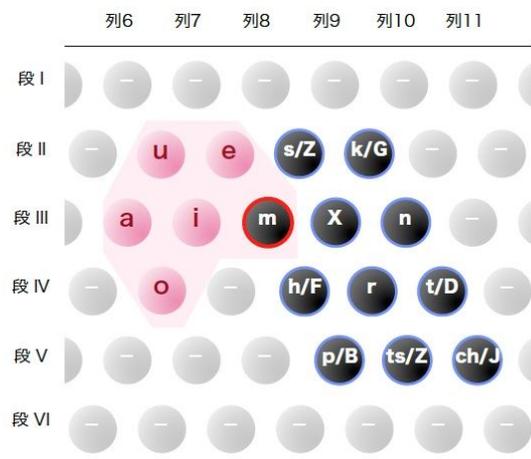


図4 同規格の動的母音エリア

また使用するボタンも、全領域ではなく、わずかに 12 個に制限したことで、演奏（操作）時の指の大きな跳躍を回避できた。また動的母音エリア内の母音ボタンの配列はつねに一定であり、人差し指、中指、薬指の一貫したフォームで母音を身体感的に覚えることが可能になり、インターフェイスの身体性の面で格段の進歩を遂げた。

さらに、こうした発想の転換を行った結果、子音単独発声（/s/や/t/など）および中間母音の発声（/a/と/e/の中間など）というこれまで想定していなかった制御に道が開け、実際にこれらを実装することができた。音声合成ソフトウェアに秘められたポテンシャルを開花させたとも言える。この意義は大きく、音素データを交換あるいは追加すれば、日本語以外の外国語の発声にも原理的に対応できる可能性が開けたのである。この可能性を明示するため、規格の名称に「国際」という文字を入れ、現行のバージョンを日本語ローカライズ版として再定義した。

この BBPTSI 規格を整理した PDF ドキュメントを web サイトにて公開している。

http://fornambros.lolipop.jp/BBPTSI_v1.0.pdf

(3) 蛇腹によるラウドネス制御

アコーディオンという楽器の最大の特徴は、その蛇腹（ペローズ）にある。これにより空気を送りリードを振動させるのが基本的な発音構造である。この構造が人間の呼吸や発声のメカニズムを直観的に連想させることは言うまでもない。本研究においても、蛇腹の操作によって合成音声のラウドネス（大きさ）を制御するようにした。

ただし、聴感上の声の大きさは、単に音量（アンプリチュード）だけで決まるのではなく、我々が「声帯緊張度」と呼んでいる声帯波形の変化によっても大きく左右される。囁き声、ノーマルな声、叫び声では、声帯部で発生する声源波形そのものが違っているのだ。我々の音声合成ソフトウェアでは、これをシミュレートするために声源波形の折り返しを実装している。声帯緊張度パラメータとしてスレッシュホールド値を設定し、それを超える波形部分は内側に折り返される。これにより豊かな高次倍音が発生し、結果として聴感上の声の力が増すことになる。MIDI アコーディオンの蛇腹を操作するとそれにかかる空気圧が MIDI コントロールデータとして送信されるが、本研究では、この MIDI データで音量と声帯緊張度パラメータを連動させ制御した。結果として、蛇腹に力を入れる奏者の身体的な感覚が、合成音声のラウドネスにダイレクトに反映するようになり、合成音声の表情・情動的な表現力が劇的に高まった。

(4) 人工音声歌唱作品の作曲・実演発表

上述した「兄弟式日本語ボタン音素変換標準規格 (BBPTSJ)」および「兄弟式国際ボタン音素変換標準規格 (BBPTSI)」に基づいて、MIDI

アコーディオンにより合成音声で歌わせる楽曲作品を 2 曲作曲し、実演発表を行った。

『ぼっくりきのくつ』（2012）作詞：村山 籌子、作曲：フォルマント兄弟（三輪真弘 + 佐近田展康）MIDI アコーディオン演奏；岡野 勇仁。（初演：「新・劇場の三科 1925 2012」神奈川県立近代美術館葉山館、2012 年 3 月 3 日）

『夢のワルツ』（2012）作詞・作曲：フォルマント兄弟（三輪真弘 + 佐近田展康）MIDI アコーディオン演奏；岡野 勇仁。（初演：サントリーサマーフェスティバル「ジョン・ケージのミュージサーカス」サントリーホール、2012 年 8 月 26 日。ほか再演多数）

(5) 制約プログラミングを用いた和音平均化鍵盤運指法の自動生成

すでに三輪と佐近田は、2003 年よりピアノ式鍵盤を用いた合成音声のリアルタイム制御を研究／開発／実現して来た。その際、ピッチの制御法として「和音平均化アルゴリズム」を考案・実装している。これは、同時に押さえた複数の鍵盤キーのピッチを単純平均して、合成音声のピッチを決定するものである。結果として、半音単位のピアノ式鍵盤を用いながらも、1/4 音、1/6 音、1/8 音といった微分音程のピッチ制御を実現している。本研究においても右手鍵盤による合成音声のピッチ制御はこの「和音平均化アルゴリズム」を継承している。

しかし、微分音程を含むピッチ系列を、具体的にどのような和音で指定するかについては、人間の手の身体的構造、押し替えにかかる時間、直前の指の状態からのスムーズな移行など極めて複雑な運指法的条件を勘案しなければならない。これまでは三輪が経験的な試行錯誤により手作業で運指法を考え、歌唱ピッチの楽譜化を行って来た。

本研究に研究協力者として参加した山崎雅史氏は、制約プログラミングのエキスパートであり、その手法を用いて、和音平均化アルゴリズムを前提に、微分音程を含むピッチの系列を実現するキーの組み合わせと運指の系列を自動的に求めるプログラムを作成した。出力されたキーの組み合わせを楽譜化し、実際に演奏してみると、訓練された鍵盤奏者であれば十分に演奏可能であり、得られるピッチの結果は極めて良好なものであることが分かった。

この研究は、合成音声のピッチ制御という具体的な課題を超えて、音楽事象の記譜と再現に関して大きな展望を秘めており、今後の展開が大いに期待される。

(6) 研究成果のビデオ公開

本研究の目的と成果を「兄弟式国際ボタン音素変換標準規格 (BBPTSI)」の紹介を中心に、

美学的な含意にも言及しながら、分かりやすくまとめたビデオ解説「フォルマント兄弟の長くまっすくな道」を作成し、youtube で公開した。

<http://youtu.be/zl0voJfgzvo>

「兄弟式国際ボタン音素変換標準規格(BBPTSI)」に基づくボタン運指を楽曲作品『夢のワルツ』を例に分かりやすく紹介したビデオを作成し、youtube で公開した。

<http://youtu.be/yvxWEExE0Az0>

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

佐近田展康、“兄弟式リアルタイム音声合成演奏システム”の概要と背景、名古屋学芸大学メディア造形学部紀要、査読無、vol.6、2013、pp.21-33

〔学会発表〕(計2件)

三輪眞弘、佐近田展康、研究発表「兄弟式国際ボタン音素変換標準規格に至る長くまっすくな道」、先端芸術音楽創作学会第17回研究会、2013年10月5日、名古屋市立大学。

三輪眞弘、佐近田展康、研究発表「兄弟式国際ボタン音素変換標準規格」およびシンポジウム、日本記号学会第33回大会、2013年5月18日、京都精華大学。

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ「MIDI アコーディオンによる合成音声の発話及び歌唱の研究」

<http://www.iamas.ac.jp/kaken/formant/>

6. 研究組織

(1)研究代表者

三輪 眞弘 (MIWA Masahiro)

情報科学芸術大学院大学・メディア表現研究科・教授

研究者番号：20336647

(2)研究分担者

佐近田 展康 (SAKONDA Nobuyasu)

名古屋学芸大学・メディア造形学部・教授

研究者番号：20410897

(3)研究協力者

岡野 勇仁 (OKANO Yujin)

ピアニスト、キーボード奏者

(4)研究協力者

長坂 憲道 (NAGASAKA Norimichi)

アコーディオン奏者

(5)研究協力者

山崎 雅史 (YAMAZAKI Masashi)

(株)NTTデータセキスイシステムズ