

## 03

## 「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

“Brother’s realtime vocal synthesis performing system” : the overview and background

映像メディア学科・教授  
Department of Visual Media・Professor  
佐近田 展康 Nobuyasu SAKONDA

本稿は、コンピュータ音楽およびメディアアートの領域で活動続ける「フォルマント兄弟 Formant Brothers」の多岐にわたる制作の技術的な基盤である「兄弟式リアルタイム音声合成演奏システム」の概要と背景について解説する。

「フォルマント兄弟」(以下「F兄弟」と表記する)は、三輪眞弘(作曲家、情報科学芸術大学院大学教授)と佐近田展康(音楽家・メディアアーティスト、名古屋学芸大学教授)によって、2000年に結成されたアーティストユニットである。これまでに、作曲、録楽作品(1)、音楽パフォーマンス、レクチャー形式のパフォーマンス、メディアインスタレーションなど、さまざまな形態で10作以上の作品を発表し、さらに執筆、ソフトウェア開発、学会会議シンポジウム発表、ワークショップやアーティスト育成企画の開催など、広範囲の活動を行っている。

F兄弟の制作・活動に一貫しているのは、独自に開発した「声」の音声合成システムを人間の手で操作(演奏)し、リアルタイムに発話・歌唱させる創作のスタイルである。同時に、このような人工的な「声」を創ること、演奏すること、およびそれを聞く聴取体験に含まれる意味を考察し、現代のメディアテクノロジー・身体・芸術の関係をめぐる哲学的／美学的問題の提起を続けている。

F兄弟がつねに「声」をテーマにし、しかも人工的な合成音声により作品を作り続けている主要な理由を思いつくまに列記してみると、次のようになるだろう。①あらゆる音響のなかでも声は極めて複雑な音響的特性を持つ現象であること、②声の聴取において人間は他の音響と比較にならない細部まで鋭敏に聞き分けており、些細な異常も聞き逃さない。したがって声の人工的合成は容易ではないこと、③声は人間の言語活動の根源的なメディアであること、④声はそれを発する主体にとって最も無媒介的で直接的なメディアだとされ、人間の思考や意識といった精神活動を根源的に支えていること、⑤他者とのコミュニケーションにおいて、声はつねにそれを発する主体に結び付けられ認知されていること、⑥声の特徴は個人のアイデンティティを構成する主要な要素のひとつであること、⑦声は言語と同時に歌唱の根源的なメディアであること……などである。したがって、声を人工的に合成してそれを操作する行為は、人間にとって最も身近で直接的な現象をあえて技術的プロセスによって迂回させ、操作可能な形で外在化し、人間が人間であることの根源的な部分にテクノロジーを使って触れる行為であると、F兄弟は考えている。この行為を芸術としか呼びようのない文脈のなかで体験可能な形で作品化するのがF兄弟のコンセプトであり、そのために考案・開発して来た技術的基盤が「兄弟式リアルタイム音声合成演奏システム」である。

本稿は、このシステムの概要を「喋る／歌う機械」をめぐる技術の歴史のなかに位置づけながら紹介し、特にインターフェイスと

身体性の問題に焦点を当てつつ、その意義と展望を示したいと考える。

## 1 フォルマントと発声のメカニズム

### 1.1 フォルマントの概念

F兄弟のユニット名の由来となり、またその作品でじっさいに使用される音声合成法に関連する「フォルマント formant」の概念について、まず簡単に説明しておく。フォルマントとは、音響物理学上の用語であり、音のスペクトルに見られるエネルギーのピークを指している。スペクトルは音を周波数領域で表したものであり、特定の瞬間において音を構成する成分（部分音）の大きさを周波数に従ってグラフ化する。そこに見られる山型のピークをフォルマントと呼び、その分布は一般的に「音色」と呼ばれる音の属性を記述する。しかし、声という特殊な音響の場合、フォルマントは「音色」ではなく「音素 phoneme」として言語的に認知されるため、特に重視される。つまり、私たちが「アイウエオ」を識別するのは、それぞれのフォルマント分布の特徴を弁別しているのだ。

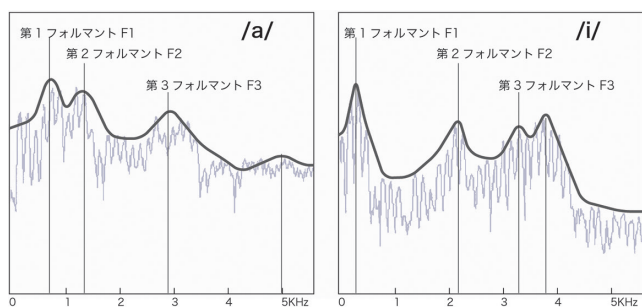


図1: 日本語母音/a/ /i/のスペクトルとフォルマント

この図は、ある日本人話者の母音/a/、/i/のスペクトルに基づき、各スペクトルのピークを接線で結んだ包絡を示している。図中の「山型」の部分フォルマントと呼び、周波数の低い方から順に第1、第2、第3…フォルマントと区別する。各フォルマントはフォルマント周波数（山型の中心周波数）、エネルギー強度（山型の高さ）、帯域幅（山型の裾野の広がり）によって定義される。

母音の認知のためには、とりわけ第1フォルマント周波数(F1)と第2フォルマント周波数(F2)の関係が重要とされ、多くの文献では各母音をF1-F2平面にプロットする方法が見られる。発声の強弱・性別・年齢・個人偏差等の変移があり、各母音は図2のような範囲を持ったおおよその領域として表現できる。また、第3フォルマント以降は音素の識別より声質の個人的特徴(声紋)を決定するうえで重要な役割を果たしているとされる。後述するように、このフォルマントをさまざまな仕組みで人工的に生成する音声合成法を「フォルマント合成」と総称している。

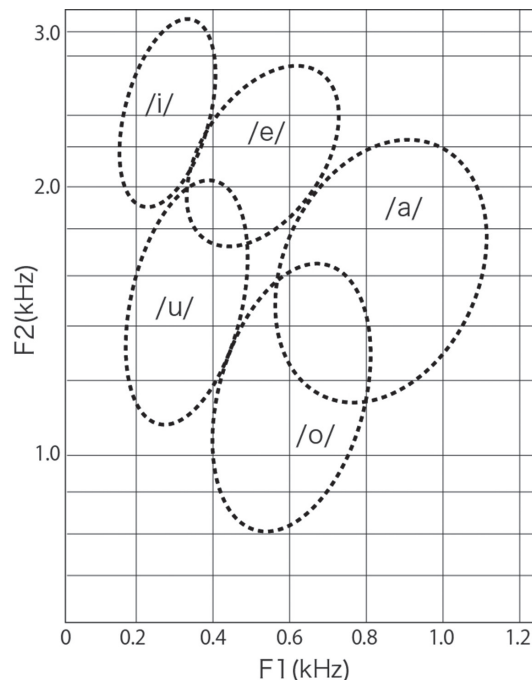


図2: 日本語母音のF1-F2平面

### 1.2 発声のメカニズム

このようなフォルマントの変化をもたらすのは、人体の発声メカニズムにおいて共鳴体として機能する声道(vocal tract)の高速かつ柔軟な形状変化である。有声音についての人間の発声は、肺呼吸の空気圧と声帯の緊張により生じる声帯振動が音源となり、声道を通過して放出される。この声道の大きさや形状が刻々とめまぐるしく変化することで複雑な共鳴が生じるのが発声の特徴だ。例えばギターであれば共鳴体の大きさ・形状はつねに固定している。しかし、人間の声道においては顎の開閉だけでなく、口唇、舌、口蓋、軟口蓋などの運動により共鳴体そのものの大きさや形状が目まぐるしく変化することになる。また軟口蓋の開閉により声道内の共鳴は鼻腔にも拡大される。これらの運動が人声のフォルマントにおけるバリエーションとダイナミックな変化を可能にしているのだ。

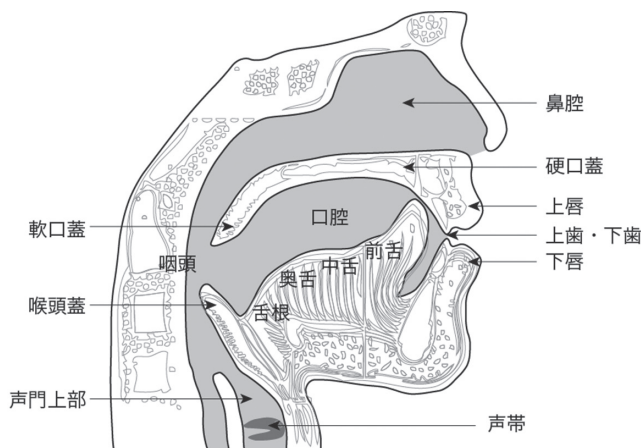


図3: 発声器官と声道

## 2 「喋る機械」の小史

### 2.1 機械式スピーチマシン

以上のような人体の発声メカニズムを器具や機械で模倣して「喋る機械(スピーチマシン)」を制作する試みは古くからある。F 兄弟の活動にも深く関連するため、ここで簡単に歴史的展開を紹介しておく。

まず次の図は、1779年にドイツの科学者C. G. Kratzenstein が発表した5つの母音の声道形状のモデルである。このモデルに従って彼は実際に5本の共鳴管を制作し、リードに接続して空気を送り込み人工的な声を発声させた。母音ごとに共鳴管は取り替えなければならない、連続的な母音変化を聞くことはできない(Dudley & Tarnoczy 1950)。

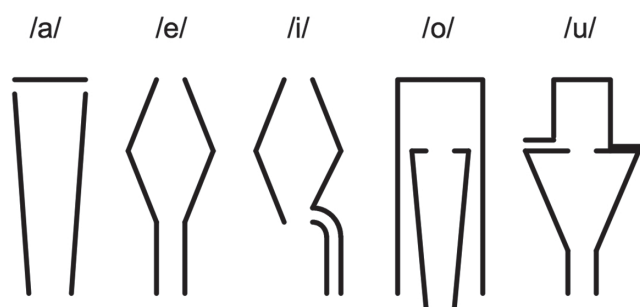


図4: Kratzensteinの声道共鳴管

自動機械「チェスをするトルコ人」の製作者として知られるハンガリーの物理学者・発明家 W. von Kempelenは、Kratzensteinよりも以前から人間の発声メカニズムの実証的・科学的研究を進めていた。彼は、声道の解剖学的構造をモデル化したうえで調音のメカニズムを機械化し、世界で最初の機械式スピーチマシンを開発し、1791年に研究の詳細を記した大部の著作を刊行している(Kempelen 1791)。このスピーチマシンは各種の母音、半母音、鼻音を生み出し、19種類の子音も発声することができたと言われる(Dudley & Tarnoczy 1950)。

図6はvon Kempelenの記述を元にしてイギリスの物理学者C. Wheatstoneが制作したスピーチマシンである。Wheatstoneは1835年に英国科学発展協会の会合でこれを発表している。von Kempelenの記述の完全な再現ではなく、声道に見立てた革製の管を手で握って形状を変えることにより母音をコントロールするなどオリジナルな発展が見られる。とはいえ、蛇腹で空気圧を生み出しリードを振動させ、その振動や空気の流路をレバーで切り替え、子音用に摩擦音を出す笛を装備するなど、基本的な構造は同じである。革製の管の微妙な握り方で目的の音素を確実に発声させる操作技術は極めて難しいものだったと容易に想像できる(Dudley & Tarnoczy 1950)。

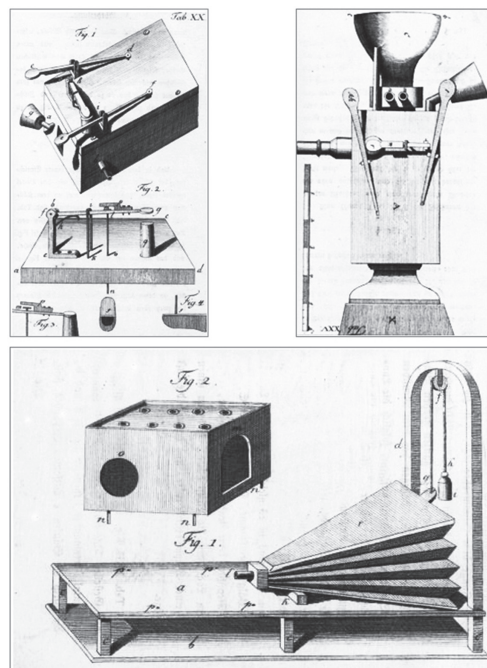


図5: von Kempelenのスピーチマシン (Kempelen 1791, p.418, 428, 439)

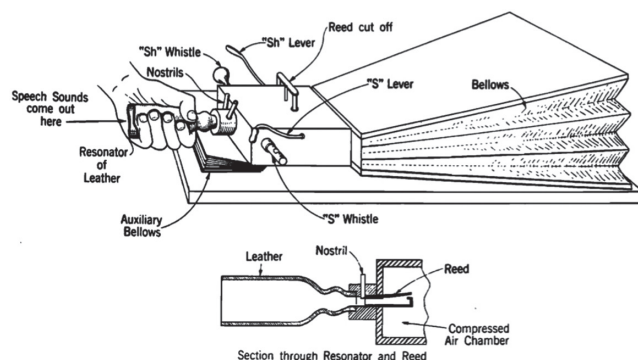


図6: Wheatstoneのスピーチマシン

後に電話の発明で知られるA. G. Bellは、16歳の時(1863年)にエジンバラでWheatstoneのスピーチマシンを実際に見ている。彼はこれに大いに魅了され、自分でも同様の機械を作ろうと独語で書かれたvon Kempelenの前掲書と格闘する(Groundwater 2005, p.25)。製作に当たってBellは、生理音声学者であった父(A. M. Bell)のアドバイスを受けて、頭部、唇、舌、口蓋、軟口蓋、歯、肺の人体構造をできる限り解剖学的に忠実に模倣しようとした。唇はワイヤーフレームにラバーを貼付けて作り、とりわけ複雑な動きをする舌についてはいくつかの木製ブロックに分けて細かく動かす工夫をした。図7は、1937年にベル研究所のR. R. Rieszが制作したスピーチマシンだが、少年時代のBellのアイデアを受け継ぎ、唇と舌に当たる部分を複数のブロックに分けてピストン式のキーを指で押すことでコントロールできるようにしている(Flanagan, et al. 2008, p.271)。ちなみにRieszは、ベル研究所の同僚であるH. Dudleyを中心に進められたVODER(後述)開発プロジェクトにも携わっており、この機械はVODER設計の一環として実験用に製作されたと考えられる。



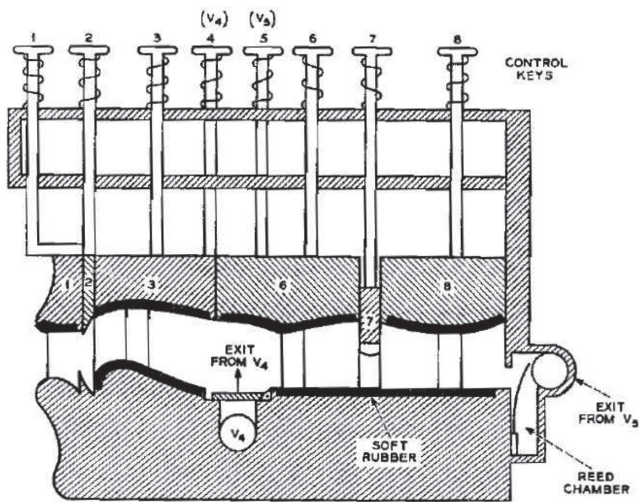


図7: Rieszのスピーチマシン(口腔の向きが上下逆)

時代は前後するが、同様に von Kempelenの研究に触発されたウィーンのJ. Faberは、EUPHONIAと命名したオルガン形状のスピーチマシンを製作し、1846年にロンドンのピカデリーでデモンストレーション興行を行っている。この機械は通常の声質だけでなく囁き声でも喋り、アリアを歌い、デモンストレーションの最後を英国国歌の歌唱で締めくくったと言われている。構造の詳細は明らかではないが、自由なピッチコントロールによる初めての「機械歌唱」パフォーマンスを実現した点で特筆すべき記録である (Dudley & Tarnoczy 1950)。

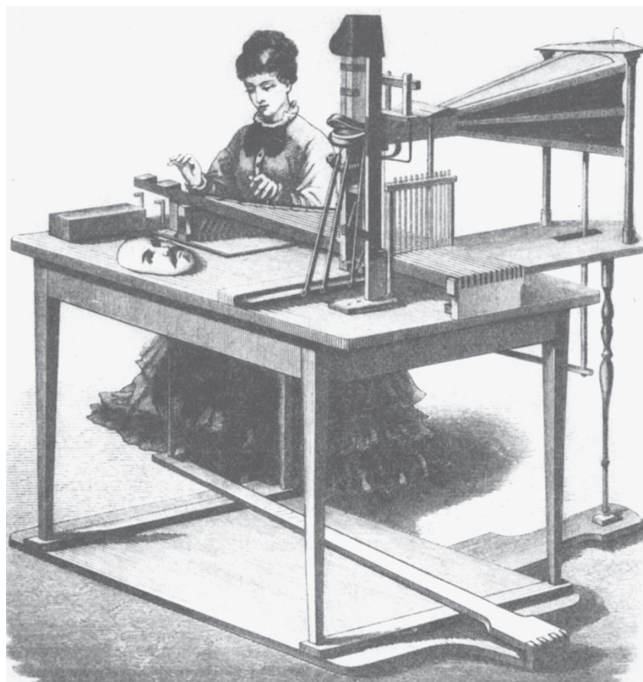


図8: FaberのスピーチオルガンEUPHONIA

## 2.2 スピーチ・シンセサイザー／VODER

VODERは、ベル研究所で開発され、世界で初めて完全に電子化された真空管式スピーチ・シンセサイザーである。開発の中心人物は、音声分析／再合成機ボコーダー (VOCORDER) の開発で知られるH. Dudleyであり、1937年のニューヨーク万国博覧会で発表された。構成は、発振回路で生成されたブザー音のような信号とランダム・ノイズを並列フィルタバンクに通しスピーカーから出力するものであり、アナログ式のスペクトル合成あるいは減算フォルマント合成だと言える。つまりVODERは、これまで紹介した機械式スピーチマシンのように人体において声が生み出される物理的過程をシミュレートする発想ではなく、音響としての声のスペクトルを直接的に生成する音響合成の考え方に立っている。音源となるブザー音は波形で言えば鋸歯状波形や三角波形に近く、そのスペクトルは豊富な倍音列を含む。これをバンドパス・フィルタ(レゾナンス・フィルタ)を通し濾過することでひとつの人工フォルマントが作れる。これらを周波数帯域別に複数個ミックスして母音など有声音のフォルマントを合成している。またランダム・ノイズは/s/や/sh/など摩擦音の音源となり、同様に帯域別にバンドパス・フィルタを通すことで無声音を合成する。フィルタは計10個使用されている。

ひとりのオペレーターが操作するために専用に設計されたコントロール部は、10個の白鍵により母音と一部の子音を、3個の黒鍵により/t/、/d/、/k/、/p/などの破裂音を制御し、左手を乗せるリストバーで有声／無声音源を切り替え、右足のペダルでピッチ(抑揚)をコントロールする。これらを同時に操って自然なスピードでVODERを「喋らせる」ためにはアクロバットのような操作技術が必要で、20人に長期間にわたる訓練を行っても満足できる操作ができたオペレーターはHelen Harperという女性ひとりだったという (2)。(Gold, et al. 2011, p.13)

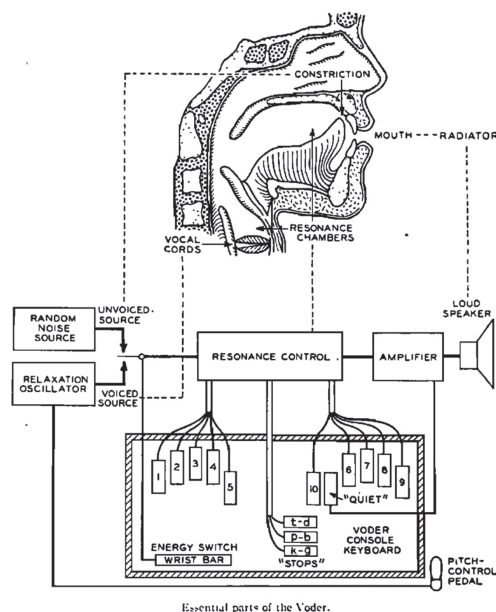


図9: VODERの構成 (Dudley & Tarnoczy 1950)



図10: ニューヨーク万国博覧会でVODERをコントロールするH.Harper

### 3 デジタル音声合成における歌声合成

人工音声研究は、1960年代以降、コンピュータや専用ハードウェアによるデジタル音声合成の時代を迎える。とりわけ1980年代以降になると、入力された文字テキストを抑揚を付けて自動的に音声化する「テキスト読み上げ」(Text to Speech : TTS)技術が一般化し、実用化が一挙に加速した。音声学研究の進展、さまざまな音声合成アルゴリズムの開発、デジタル信号処理技術の進展、ハードウェアレベルの演算処理の高速化などにより、今日では極めて高品位かつ自然な音声を生成させ、社会生活のさまざまな場面の情報伝達や障害者の支援などにおいて大いに実用化されている。ここでその概要を論じるにはあまりにも話題が多岐にわたり情報量が膨大になるため、本稿の目的に関連した2点についてのみ記したい。

#### 3.1 規則合成 VS 連結合成

コンピュータによる最初のデジタル音声合成は、1961年に発表されたベル研究所のJ. KellyとL. Gerstmanの研究である(Klatt 1987)。Kellyらの合成法は、本稿で紹介した機械式スピーチマシンと同様の発想で、物理現象としての声帯振動と声道伝達をシミュレートする規則を関数化し、コンピュータで演算処理することで音声波形を合成する「物理モデリング合成」の考え方であっ

た。他方、VODERの発想のように音響としての音声スペクトルをシミュレートする規則を関数化しコンピュータで演算処理することで音声波形を生成するのが「スペクトル合成」である。声の産出過程に着目するか、産出された声の音響的特性に着目するかの違いはあるが、両者とも一定の生成規則を設定し演算処理することで無から音声を生成させる「規則合成 synthesis-by-rule」である点は共通している。

概して規則合成は、生成規則における多数のパラメータを自在にコントロールできるため、発話のピッチやスピードまた声質の微妙な表情変化などにおいて極めて柔軟な合成法である。ただし、モデルの単純化のため生成される音声品位はいわゆる「ロボットの」なものになる傾向がある。これを改善するために、実際の人間の声を線形予測符号化(LPC)によって分析しフィルタ関数を得たり、隠れマルコフモデル(HMM)によって学習させその結果から直接に音声波形を生成する試みなどが進展している(赤川 2007)。

一方、実際の人間の声をレコーディングし、そのサンプル断片(音声素片)を時間領域で連結することで音声进行合成するのが「連結合成 concatenative synthesis」の考え方である。音声素片として、音素単位、音節単位、2音素間の変移単位(diphone)などさまざまな単位を用いた合成法が存在している。連結合成では、実際の人間の声を素材として使用しているために極めて生々しく高品位な合成音声が可能である。その一方で、最大の問題は連結部のギャップ、ノイズ、不自然さで、素片同士をいかにスムーズに連結させるかが課題となる。さらに、自由な抑揚や声の表情のバリエーションを得ようとすればするほど素片の数が膨大になり、素片の収集、データベース管理や検索アルゴリズムの最適化などの課題もある。

#### 3.2 歌声の音声合成

「規則合成 VS 連結合成」の問題を、本稿の目的に沿って「歌声」の音声合成分野でもう少し説明しておこう。言うまでもなく音声合成研究のメインストリームは実用的なTTS研究であり、歌声の音声合成は周縁的なテーマとして一部のコンピュータ音楽研究において議論されて来た。そもそも「歌声」は、TTSが対象とする「喋り声」と比較して、有声/無声の比率、音量の変化幅、ピッチの変化幅、歌唱特有のビブラートやフォルマントの変移など音声学的に重要な差異があり、TTS研究で得られた研究成果がそのまま歌唱合成に適用できるわけではない(Font 2001, p.12)。

「歌唱」のコンピュータによる最初のデジタル音声合成は、喋り声と同様にKellyらのベル研究所チームが達成した。1961年にJ. KellyとC. Lochbaumはベル研究所のIBM 7094を用い最初の歌声合成に成功し、同研究所のM. Mathewsが合成した楽器音色を伴奏にしてフォークソング「デージー・ベル」の録音を残している。これはコンピュータが歌った最初の歌として広く知られること



になる(3)。Kellyらの歌声合成法は、先述した物理モデリング合成であった(Kelly & Lochbaum 1962)。同種の考え方の発展はP.R.CookのSPASMなどに受け継がれる。

スペクトル合成の中では「フォルマント合成」が早くからポピュラーな合成法として知られている。これは、声のスペクトルにおけるフォルマントの分布を人工的に合成する方法の総称である。なかでも倍音成分を豊富に含んだパルス波や鋸歯状波オシレータとノイズジェネレータを音源にし、複数のバンドパス(レゾナンス)フィルタで不要なスペクトル成分を濾過して目的のフォルマントを得る「減算方式のフォルマント合成」は、シンプルかつ柔軟な方法である。フィルタの接続法には直列と並列があり、1980年にMITのD. Klattが提唱した直列+並列フィルタ型のフォルマント・シンセサイザーは、TTSの分野で大きな成果をあげ、DECTalkという名で広く商用化された(Klatt 1980)。これは筋萎縮性側索硬化症で発声機能を失った物理学者ホーキング博士の声として知られている。後述するF兄弟の音声合成エンジンも、この並列型を大いに参考している。

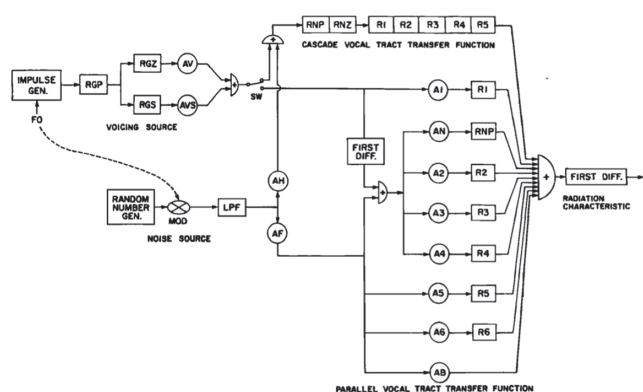


図11: Klattの直列+並列フォルマント・シンセサイザーのブロックダイアグラム

さらにこれとは異なり、基本周波数・フォルマント周波数・帯域幅のパラメータを与えるとひとつのフォルマントを生成するジェネレータを考案し、加算方式でフォルマントを得るフォルマント波形関数(FOFs)が、1979年にフランス国立音響音楽研究所IRCAMで開発され、CHANTという名で歌声合成に広く活用された。他にはサイン波の集合からフォルマントを得る方式などもある。

一方、実際の人間の声をレコーディングし、そのサンプル断片を時間領域で連結する連結合成は、大きなピッチ変化、長い母音持続、微妙な表情変化などを重視する歌唱の合成には不向きとされて来たが、YAMAHAの剣持秀紀らが2003年に開発したVOCALOIDは、クリプトン・フューチャー・メディアから2007年に発売されたライセンス商品「初音ミク」の爆発的ヒットにより社会現象と呼べるまでに大きなブームを巻き起こしている。VOCALOIDの音声合成アルゴリズムは、実際の人間の声より採取した音声素片のデータベース(歌声ライブラリ)から、スコアエディタ部に入力

された歌詞と音符データによって素片を選択して読み出し、連結して合成音声を生成する。この際、発声と音楽上の拍節とのタイミング調整など歌唱特有の工夫が盛り込まれている。また、ライブラリを構成する音声素片は、diphoneと伸ばし音で構成され、ある素片の最終フレームと次の素片の冒頭フレームのスペクトル包絡を滑らかに補間することで連結合成特有の不自然なギャップやノイズを実用レベルで回避している(剣持&大下 2008)。さらに、音声素片は単に日本語音素のすべての組み合わせを網羅するだけでなく、さまざまな声の高さで歌わせても自然に聞こえるように数種類のピッチで採取されている。

歌唱に特化しているとはいえ、VOCALOIDの基本的な設計思想は、あらかじめ用意されたテキストを読み上げるTTSに近い。例えば、音楽上の拍節に合わせて歌うためには、いくつかの音素では拍節に先立って発音を開始しなければならない、音素やピッチ情報に加えて持続時間やディレイ時間の情報をまとめて発声開始の前に合成エンジンに送る工夫がなされている。VOCALOID 2からは歌詞のみをエディタに入力しておきピッチはリアルタイムにキーボードで演奏できるようになり、また2012年には歌詞の指定を含めてリアルタイムに演奏が可能な音源搭載型専用インターフェイス「VOCALOIDキーボード」の試作機が発表されている(加々見, 他 2012)が、基本的な使用法は事前に時間をかけてエディタに入力した歌を歌わせる方式である。

## 4 フォルマント兄弟の人工音声システム

### 4.1 音声合成エンジンの基本構成

はじめに書いた通り、F兄弟の制作・活動に一貫しているのは、独自に開発した「声」の音声合成ソフトウェアを人間の手で操作(演奏)して発話・歌唱させる創作のスタイルである。これを実現するために音声合成エンジンの「リアルタイム・コントロール」が不可欠になる。第2章で紹介した機械式スピーチマシンからVODERまでの例では、オートマチックに機械に喋らせることは技術的に不可能であり、必然的にリアルタイムに人間の手でコントロールする必要があった。しかし、デジタル音声合成の時代を迎え、「テキスト読み上げ(TTS)」技術が一般化するに及んで音声合成研究の主流はTTSの発想のもとに発展し、リアルタイム・コントロールはいつしか忘れられた。同時に機械と対峙する人間の身体性や両者を媒介するインターフェイスの問題も忘却されて行く。その忘却されたものをF兄弟は創作のなかで取り戻そうとしている。

ユニット結成の以前から、すでに三輪は作曲作品「言葉の影、またはアレレヤ——Aのテキストによる」(1998)において、4人の女性キーボード奏者が、基本メロディとそれに対応する3つのフォルマント周波数をそれぞれ独立して演奏し、実空間のなかで「ひ

とつ「声」をフォルマント合成する作品を発表している。佐近田もサウンド・インスタレーション作品「watermachine」(1998)において、水面の揺らぎから声が合成される作品を発表し、フォルマント合成の手法を取り入れていた。佐近田はその後さらに自身の音声合成プログラムを発展させ「トワノコエ——ピアノ、ソプラノ、機械歌唱のための」(2003)では、子音を含めすべての日本語音素を網羅したリアルタイム音声合成エンジンのプログラミングに成功し、フォルマントの分布パターンを調整して声質のキャラクター設定も行っている。具体的には、乳児、幼女、ソプラノ歌手、ロック歌手Freddie Mercuryの声質をフォルマント合成したうえで、声質相互の連続的モーフィングも実現している。その後F兄弟の作品で用いられている「兄弟式リアルタイム音声合成演奏システム」の音声合成エンジン部分は、この時のプログラムを発展的に改良したものだ。

この音声合成エンジンは、すでに紹介したDudleyのVODERとKlattの直列+並列フィルタ型フォルマント合成から大きなヒントを得ながら、佐近田独自の解釈でプログラミングされた減算フォルマント合成型のエンジンである。開発環境はCYCLING'74のMax/MSPである。次の図はその基本構成になる。

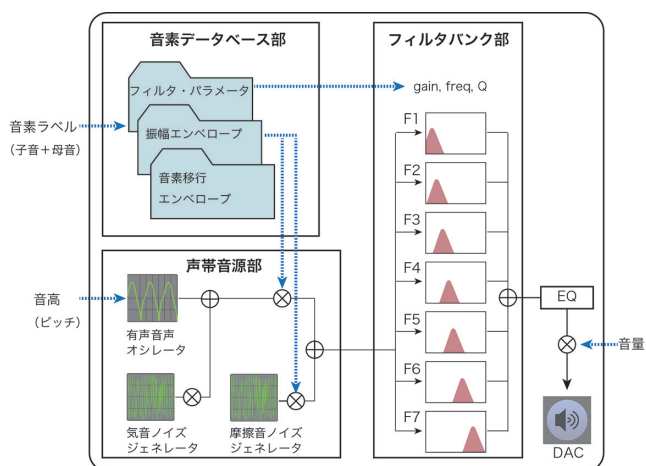


図12: 合成音声エンジンの基本構成 (佐近田 2003-)

この音声合成エンジンは、大きく分けて声帯音源部、フィルタバンク部、音素データベース部から構成されている。

「声帯音源部」において有声音声の音源となるオシレータ波形は、G.Fantほかの声帯波形研究を参考にしながら512サンプルの波形テーブルを作成して使用している(Fant, et al. 1985)。無声音声の音源は、気音ノイズと摩擦音ノイズに分け、それぞれホワイトノイズ・ジェネレータを用いている。

「フィルタバンク部」は、7つのバンドパス(レゾナンス)フィルタを並列に配置し、第7フォルマントまで表現可能である。本プログラムでは子音と母音でフィルタを共用している。この構成はVODERと類似しており、Klattのように並列と直列フィルタを使い分ける考え方は導入していない。各フィルタはゲイン・中心(フォルマント)

周波数・Q値(フィルタの傾斜値)の3パラメータで制御する。Max/MSPのreson~オブジェクトの仕様により、ゲインは負の値を取ることができ、原則として奇数フォルマントのゲインは正の値、偶数フォルマントは負の値を与えている。これにより合成音のスペクトルにおいて隣り合うフォルマントの相互干渉が低減される。

各音素は、[A], [E], [I], [O], [U], [k], [g], [s], [sh], [z], [j]....等のラベルで管理し、それぞれフィルタ・パラメータ、振幅エンベロープ、音素移行エンベロープのデータを持つ。これが「音素データベース部」を構成する。音素移行エンベロープとは子音音素から母音音素への移行タイミングと時間的傾斜を表しており、[s]や[h]などではゆるやかに、[k], [t], [p], [m], [r]などでは直角的に移行する。こうして例えば音声合成エンジンが [s\_A] というメッセージを受け取れば、ラベル[s]のフィルタ・パラメータと振幅エンベロープがただちに呼び出されて子音の発声を始め、[s]の音素移行エンベロープに従って自動的にラベル[A]のフィルタ・パラメータが呼び出され母音へと移行する基本仕様である。

合成音声の性別・年齢・声質を決定するのは母音音素のフィルタ・パラメータである。とりわけ第3〜第6フォルマント(おおよそ3〜7KHz帯域)のパラメータが声質を大きく左右する。その値は、目標とする音声のスペクトルと合成音声のスペクトルを並べて表示させ、各フィルタを調整しながら目と耳で近づけて行く試行錯誤で決定している。これら母音音素のフィルタ・パラメータを2セット読み込み、両者のあいだでリニアにデータを補間することで、2種類の声質モーフィングも可能である。他方、子音音素についてはすべての声質に共通したものを使用しており、Klattの論文(Klatt 1980)にあるデータを参考にしつつ試行錯誤で決定した。

## 4.2 音声合成エンジンの拡張

F兄弟の活動のなかで、音声合成エンジンの各所に改善および拡張を続けているが、そのいくつかを紹介しよう。

### ・声帯緊張度パラメータの追加

F兄弟の「フレディの墓／インターナショナル」(2009年)は、ユニットにとって最初の録音作品であり、音声合成エンジンの多数のパラメータをフル・コンピュータ制御でコントロールすることにより、どこまで人工音声歌唱の表現力を高められるかに挑戦した作品である。声質としてはすでに佐近田が作成していたロック歌手Freddie Mercuryの母音フィルタ・パラメータを土台にさらに本人の声紋に近づける工夫を重ね、「いまは亡きフレディが日本語で革命歌インターナショナルを歌っていた」という想定で歌唱を制作した。この時、新たに音声合成エンジンに導入したのが「声帯緊張度パラメータ」である。

F. Mercuryのロック的な歌唱を表現するうえで、声質の表情変化は極めて重要である。同一人物の声であっても、甘い声、ストレートな声、張り上げたシャウト、意図的なダミ声などの声の張り

具合に関する表情変化がある。これらは「喉を締める／緩める」という身体感覚の通り声帯音源部で実装するのが合理的だ。本プログラムでは、有声音オシレータの波形に限界値(上限・下限値)を設定し、それを越えた波形部分を内側に折り返すことで声帯緊張度を表現している。実際に波形の振幅と限界値を独立して操作することにより、高次倍音の比率が大きく変化し、ロック歌唱特有のダイナミックな表情変化が得られた。

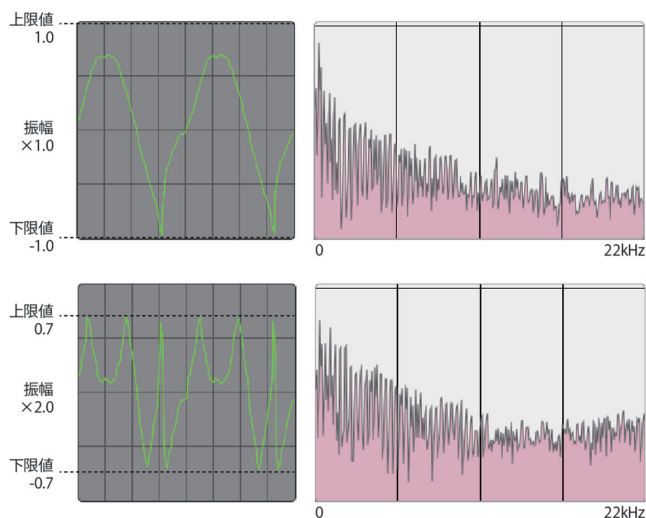


図13: 声帯緊張度による波形とスペクトルの変化

#### ・HARUMIモジュレータの追加

「せんだいドドンパ節」(2010)は、せんだいメディアテーク10周年記念プロジェクトの一環で制作された。東北地方の民謡に取材し、「高音キン」と名付けた架空の民謡歌手の声で複雑な「こぶし回し」の民謡歌唱に挑戦した。この時、民謡や演歌においていわゆる「唸り」と称される独特な喉の使い方(都はるみの演歌歌唱に誇張された形で聞くことができる)をシミュレートしようとした。これは声帯緊張度パラメータでは表現することができず、試行錯誤のなかから振幅変調(AM)のモジュレータを矩形波にし、周波数を80Hz程度でランダムに変調する方法を発見した。このデプスをパラメータ化して実装したのが「HARUMIモジュレータ」である。同曲をリアルタイムに演奏する場合は、これをMIDIキーボードのアフタータッチやフット・コントローラにアサインしてコントロールする。

## 5 兄弟式発声コントロール

こうした音声合成エンジンをコントロールして発話／歌唱させるためには、最低でも音高(ピッチ)、音量、音素ラベルの3つの要素をリアルタイムに与えて行かなければならない。F兄弟の作品制作における技術的な挑戦の骨子は、これら3要素をひとりの演奏者がコントロールする方法を開発することである。

2003年にF兄弟が自らパフォーマンスを行った音響作品「兄弟 de ビザ注文」では、三輪と佐近田が1台のMIDIキーボードを連弾演奏し。三輪はキーボードの低音域を使って音素ラベルの指定を、佐近田は高音域を使って音高+音量をリアルタイムにコントロールした。この時、三輪が考案したMIDIノートメッセージと音素ラベルの変換規則を整理し、ピアノ式鍵盤のコンビネーションですべての日本語音素を指定する「規格」として発表したのが「兄弟式日本語鍵盤音素変換標準規格(Brother's Keyboard-to-Phoneme Transfer Standard for Japanese language: BKPTSJ)」である。

BKPTSJ規格は、一般に市販されているピアノ式鍵盤のMIDIキーボードで兄弟式音声合成エンジンをコントロールするための変換プロトコルである。任意にキースプリットを設定し、それより低い音域で音素ラベルを指定し、高い音域で音高(ピッチ)と音量をコントロールしながら、ひとりの演奏者が両手の演奏で発話／歌唱をコントロールすることができる。

Brother's Keyboard-to-Phoneme Transfer Standard for Japanese language

### 兄弟式 日本語鍵盤 音素変換標準規格

きょーだいしき にほんごけんぱん おんそへんかん ひょーじゅんきかく

kyo da i si ki ni ho n go ke n ba n o n so hen ka n hyo zyu n ki ka ku

yA yU yO N wA

図14: 兄弟式日本語鍵盤音素変換標準規格 (BKPTSJ)



まず音素ラベルの指定から説明しよう。図14にある通り、ピアノ式鍵盤の黒鍵には母音、白鍵には子音が割り当てられている。例えば/a/と発音させたいのであればG♭の鍵盤を、/so/であれば、DとB♭の鍵盤を同時に押せば良い。ヤ行、ワ行については、/yo/→/i/+/o/、/wa/→/u/+/a/と分解し、黒鍵の二和音で指定する。こうして日本語の五十音表のすべての音は、単音あるいは二和音で指定が可能になる。濁音については「子音鍵盤の三度上の白鍵を同時に押す」という規則に従う。例えば/zo/であれば、/s/=Dの三度上のFを同時に押すことになり、D+F+B♭の三和音で指定する。ハ行の半濁音/p/音については例外的に四度上の白鍵を同時に押すことで指定する。キャ、シュ、ビョなどの拗音についてはモーラを構成する2つの母音を同時に押す。例えば、/kyo/であれば、/k/+/i/+/o/となり、C+D♭+B♭を同時に押せばよい。

このようにBKPTSJ規格は、論理的かつ効率的に、濁音や拗音を含めたすべての日本語音素(音節)の指定を1オクターブ内の鍵盤コンビネーションで解決する。これにより、演奏者は片手(左手)だけで任意の音素ラベルを指定し音声合成エンジンに送ることが可能になる。またキースプリットより低い音域内であれば、和音の転回形と同様の考え方で、任意の鍵盤のオクターブをシフトさせて「替え指」で演奏することも可能である。

次に音高(ピッチ)の指定について説明しよう。BKPTSJ規格の2003年バージョンにおいては、メロディや抑揚の指定は「キースプリットより高い音域を右手で演奏する」という想定のみでピッチの詳細なコントロール法は規格に盛り込まれていなかったが、「NEO都々逸——六編」(2009年)の制作において、日本の伝統歌唱における微細な「こぶし回し」をリアルタイムに音声化するために、「和音平均化アルゴリズム」を追加した。これは「同時に押さえた鍵盤の音高を単純平均する」という規則をアルゴリズム化したものだ。例えば、隣り合うBとCの鍵盤を同時に押せば、その単純平均(つまり中間)の音高が得られる。MIDIノート・ナンバーの値で言えば、71と72を同時に押さえることで71.5の音高を得るという意味である。同時に押さえる鍵盤は最大4つまで可能である。この結果、12音平均律を前提としたピアノ式鍵盤を使いながら、全音程の1/4、1/6、1/8の単位で音高が一意に指定できるようになった。次の譜例は「NEO都々逸——六編」の第1曲である。譜例の下段で歌唱の音素ラベルを指定し、上段で音高(メロディ)を同時に演奏するが、こぶし回しのビブラート部分を三音のコンビネーションにより微分音程で演奏していることが分かる。

今日の稽古は

Formant Bros.

図15:「NEO都々逸——六編」第1曲

## 6 インターフェイスの身体論

この楽譜に記されているものが歌唱の「メロディと伴奏」ではなく、機械に対する「発声の指示」であることに再度注意を喚起したい。BKPTSJ規格におけるF兄弟の試みが持つ音楽史上の意義は、西洋音楽で伝統的に用いられて来たピアノ式鍵盤と五線譜の体系をそのまま用いながら「発声の指示」を初めて確定的に記譜できる、新しいコードを提示したことにある。文字を使うことなく音符のみで言葉(歌詞)を記述し、従来五線譜で表すことが不可能であった半音以下の微分音程を1/8音程の精度で記譜できることは、世界中の音楽文化において民族固有の歌唱法とともに歌い継がれて来たさまざまな歌唱を確定的に記譜する道を拓き、さらには未聞の歌唱法に基づく新たな音楽を構想する可能性をも切り拓くことにつながると考える。

F兄弟がことさら「五線譜での記譜」を強調する理由は、それを実際に読めて演奏できる「訓練された身体」が、この社会に多数存在するからである。総務省統計局による平成21年全国消費実態調査「主要耐久消費財に関する結果」では、二人以上の世帯の25.1%がピアノを所有している(総務省統計局 2010)。周知の通り、明治以降のわが国の音楽教育は西洋クラシック音楽を前提に行われて来た。その過程でピアノは単なる楽器のひとつではなく、西洋クラシック音楽の文化的価値、ひいては西洋的価値を体現する象徴的な存在となっている。数ある楽器のなかでも特に高価で巨体を有するピアノがこれだけの世帯に普及している理由は、決して「音楽の楽しみ」だけでは説明できない。こうした歴史文化的状況のなかで、五線譜に書かれていれば10本の指

が指定どおり正確に動く「訓練された身体」が現在もお多数生み出され続けている。

岡田暁生は、19世紀のヨーロッパ社会においてピアノが普及して行く過程を、近代的軍隊や工場などにおける身体の規律化とパラレルな問題として論じている。

ある楽器の習得のために長期間にわたる不断の訓練が必要なのは、ピアノに限らずあらゆる時代、文化において変わらない。しかし、その訓練法や背後にある身体観において19世紀のヨーロッパは異質な光彩を放っている。ピアノは当時の人びとにとって「複雑なメカを搭載した黒塗りの高性能鋼鉄マシーン」と考えられていたし、実際、中産階級の家庭に広く普及して行くにともない、工場の流れ作業によって大量生産される歴とした工業機械製品になった。その過程で「幾多の部品を頑健に組み上げた近代工業文明の傑作としてのピアノを弾きこなす者は、機械に比肩するような完璧な技術人であるべきだ」(岡田 2008, p.96)との思潮が支配的となる。例えば、ピアノ教本として有名な練習曲集「ハノン」は1837年に出版されたが、音楽的表現の練習ではなく機械的な指体操のドリルであり、指を体系的に均質化しようとする発想に貫かれている(p.89)。片手の5本の指は身体構造的にそれぞれ特性があり本来は均質ではあり得ない。しかし、それをすべて均質にし、楽譜の指示通りにオートマティックに動く「機械」へと変貌させるのだ。この種の思想を端的に表すのが、同時期にさかんに普及した矯正器具であろう。1836年にH. Hertzが考案したDACTYLIONは、まさに人間をサイボーグ化するかのような指筋肉トレーニング器具である(p.124)。

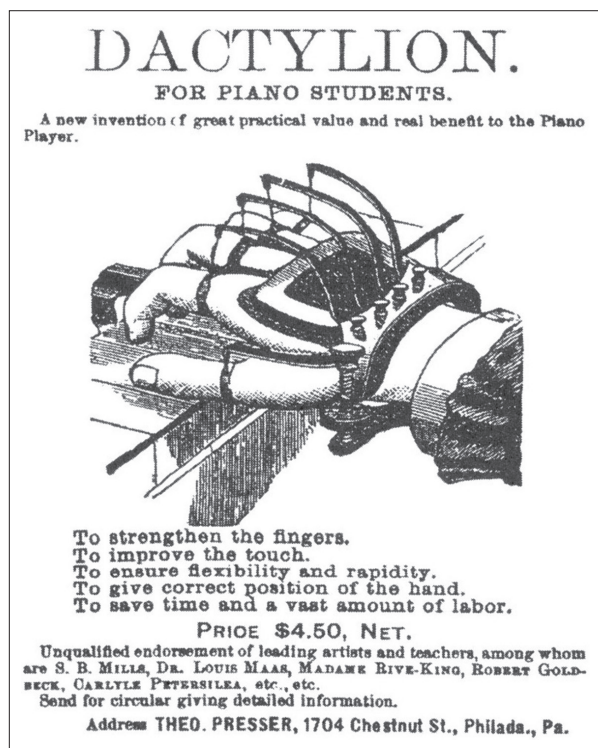


図16: ピアノの指訓練器具DACTYLION新タイプの広告(岡田 p.128)

19世紀のヨーロッパでは、この種の矯正器具や装置がさかんに考案され、それを使った指訓練のための練習曲集とあわせて中産階級に普及して行った。さらには訓練を積んでも構造的に単独では高く持ち上がらない薬指について、あえて腱を切断してまで解放するピアニストのための外科手術が推奨されたりもした(p.142)。

しかし、このような身体の均質化はピアノの訓練に限った話ではない。並行して、近代的軍隊の教練、近代スポーツの競技化、体育／体操の一般化、工場のシステマティックな労働管理などさまざまな場面において、自然な身体の流れるような動きを複数の単純操作に分解し、それぞれを反復し強化する発想のもと、身体の規律化が大々的に展開される(p.206)。

岡田の議論を敷衍して言えば、ピアノが表現するものは、音楽にとどまらず生活領域全般の「近代性」であり、ピアノが弾けるようになることは、この「近代性」に身体を適合させる巨大な文明的潮流の過程の一部だと言っても過言ではない。本稿で「近代性」と呼んでいるものは、機械テクノロジーがあらゆる生活領域に進展するなかで、人間の思考、身体、社会、文化が、ということは「人間」そのものが、否が応にも「機械との関係」に立たされ再配置される歴史的様相の総体を意味している。「機械との関係」に立たされた人間の反応はもちろん単純ではない。機械テクノロジーの進展と人類の進歩を素朴に結びつける進歩主義、人知を超えた自然のなかに真実を求めた自然主義、とりわけ、19世紀ヨーロッパの思想と芸術を席卷したロマン主義思想は、ある時は機械を蔑視し嫌悪する一方でそれを理想化し、自身の身体をオートマティックな機械へと無化することで、その分だけ人間的で純粋な精神の解放を夢想した(p.206)。こうした矛盾に満ちたアンビバレントな態度が「近代性」の複雑な様相を示している。

したがって、五線譜が読めてピアノが弾ける近代的に訓練された身体をこれまでとは全く異なった仕方でも扱う試みについて、F兄弟は、これら所与の「近代性」をひとたび外在化し組み替える創造行為のひとつだと考えている。それゆえ、BKPTSJという「規格を制定すること」についても、単なるプログラミング上の工夫に留まらず、ひとつの「作品」として、しかも広義の政治性を伴う「作品」として位置づけている。

当然ながら、こうした音声合成エンジンをコントロールするために「VOCALOIDキーボード」のような専用のインターフェイスを新たに開発する発想はあるだろうし、F兄弟もその魅力と意義を否定しない。ただし、インターフェイスを考えることは、単にテクノロジーの問題ではなく、機械と身体の接点を思考することである。そこには暗黙のうちにひとつの歴史的な身体観が前提とされる。先述のVODERの専用インターフェイスをただひとり操作できたHelen Harperの特殊な身体を思い返してみれば、それはいつの世にも突発的に現れる例外的な「アクロバティックな身体」の範疇に留まり、同じく「訓練された身体」といっても、ピアノのような歴



史・社会・政治・文化的な広がりの中で身体の制度化とその組み替えを思考する契機にはならないのではないだろうか。

## 7 鍵盤からアコーディオンへ

このようなピアノ式鍵盤インターフェイスによる数々の実験を経て、近年のF兄弟は「アコーディオン」というインターフェイスに関心を寄せている。そして、2011年度より3カ年で日本学術振興会科学研究費補助金研究「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」(基盤研究(C) 研究課題番号 23520175)を進めている。

合成音声のコントロールにアコーディオンを採用することは、ある意味で直感的で自然な発想である。蛇腹を広げたり縮めたりしながら空気をコントロールして発音させるメカニズムは、構造的にも身体感覚的にも呼吸や呼気による発声と非常に近いからである。ちなみに近代アコーディオンの原型は、1829年にオーストリアとイギリスでほぼ同時期に発明されたが、イギリスにおいてコンサーティーナ“concertina”と命名されたボタン式のアコーディオンを発明したのは、すでに本稿で紹介したC. Wheatstone(1835年にvon Kempelen型の機械式スピーチマシンを制作した)その人であった(渡辺 1993)。アコーディオンとスピーチマシン。Wheatstoneのなかで両者は類似した発想、共通した技術だったのだろう。

2004年にROLAND社はMIDI対応のアコーディオン型シンセサイザーV-ACCORDIONシリーズを発表する。本稿執筆時点では大小のピアノ式/ボタン式をあわせて計11種類のモデルが販売されており、F兄弟もこれをインターフェイスとして使用することにした。V-ACCORDIONの左手ボタンは、主としてベースノートとコードを演奏するために配置され、その数は本体の大きさによって72、96、120個とさまざまである。F兄弟の発想は、この左手ボタンによって音素ラベルを指定し、右手のピアノ式鍵盤でメロディ/抑揚を演奏し、蛇腹(ペローズ)で音量や声帯緊張度をコントロールするというものだ。以下は72ボタン(12列6段)のピアノ式鍵盤FR-1モデルを中心に説明して行く。



図17: ROLANDの「V-ACCORDION」FR-1

## 兄弟式日本語ボタン音素変換標準規格

アコーディオンの左手ボタンで日本語音素を指定するために2012年に制定されたのが「兄弟式日本語ボタン音素変換標準規格(Brother's Button-to-Phoneme Transfer Standard for Japanese language: BBPTSJ)」である。

V-ACCORDIONの各ボタンはMIDIノート・ナンバーを単音あるいは和音(コード)で出力するが、ROLANDの製品仕様により異なったボタンが同一のMIDIノート・ナンバーを出力するなど、MIDIメッセージの受信側では全ボタンを一意に識別することが不可能である。このため本研究の目的から実際には使用できないボタンがあり、規格もそれに合わせて考えることを余儀なくされた。そのうえで、まずは各単独ボタンに次のような音素ラベルをアサインした。

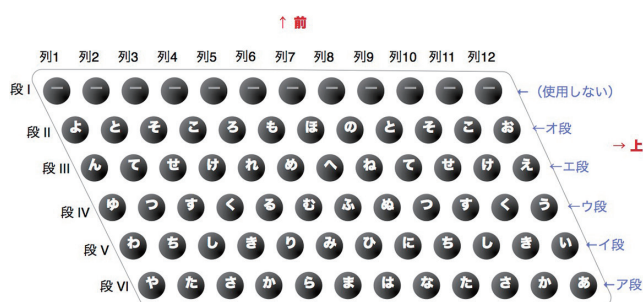


図18: 単独ボタンにアサインした音素ラベル(FR-1)

濁音については「起点となるボタンの右隣(奏者から見て上隣)のボタンを同時に押す」という規則に従う。ハ行の半濁音/p/音については例外的に2つ右隣のボタンを同時に押すことで指定する。また/ya/, /yu/, /yo/, /wa/については列12の母音ボタンを2つ同時に押すことでも指定可能だ。例えばやは、/ya/→/i/+/a/と考え「い」と「あ」の2ボタンを同時に押せばよい。

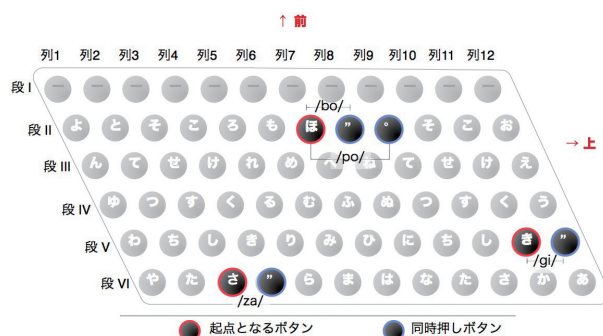


図19: 同時ボタンにより濁音を指定



ピアノ式鍵盤のためのBKPTSJとの大きな相違点は「時間差ボタン」の考え方を導入したことである。日本語で頻出する任意の母音から/n/への移行は、「ん」ボタンを用意するだけではなく「任意のボタンを押さえながら、その左隣(奏者から見て下隣)ボタンを、時間差をあけて押せば/n/へと移行する」ことで解決した。同時ボタンか時間差ボタンかを識別するスレッシュOLD時間は自由に設定可能であるが、現在は50ミリ秒としている。



図20: 時間差ボタンにより/n/に移行

/kya/, /shu/, /byo/などの拗音、あるいは現在発音中の母音から任意の他の母音への滑らかな以降についても「時間差ボタン」の考え方を採用し、「任意のボタンを押さえながら、近くのア、イ、ウ、エ、オ段ボタンを、時間差をあけて押す」規則で実現した。ここでは時間差ボタンはひとつに特定されず、起点ボタンから指が届く範囲の複数のボタンを「段」として捉えている。



図21: 異なる母音、拗音への時間差移行

右手で演奏するピアノ式鍵盤については、和音平均化アルゴリズムをそのまま適用させている。

以上のBBPTSJ規格をアルゴリズム化し、兄弟式リアルタイム音声合成演奏システムに実装して作曲したのが「夢のワルツ」(2012年)である。演奏については、アコーディオン奏者でありV-ACCORDIONの広報・普及活動も行なっている長坂憲道氏、「NEO都々逸」をはじめこれまで数多くピアノ式鍵盤でF兄弟作品を演奏して来たピアニスト岡野勇仁氏がアコーディオン演奏のトレーニングを積んで実現した。

アコーディオン最大の特徴である「蛇腹(ベローズ)」には、音量と声帯緊張度の2つのパラメータをスケーリングしてアサインした。予想通り、これによる合成音声の表現力の向上は劇的なものであった。蛇腹の操作で出力されるのは単にMIDIコントロールメッセージのエクスペッション・データのみであり、理屈のうえではフット・コントローラでも任意のスライダーやツマミ等でも代用が可能はずだ。にも関わらず、蛇腹という特殊なインターフェイスのおかげで、演奏行為の身体性と合成音声の表現性のあいだに、奏者にとっても聴衆にとってもごく自然な対応関係性が成立し、ピアノ式鍵盤では決して得られなかったダイナミックで豊かな表現力が獲得できた。ちなみにこの作品は、すべての必要機材がバッテリー/電池で駆動可能なため、モバイル「流し」スタイルで演奏可能であり、初演のサントリー・ホール前カラヤン広場、同年の京阪電車プリペアド・トレイン車内での演奏など、自由な移動という演奏スタイルの面でも新しい展開を見た。

ピアノ式鍵盤のためのBKPTSJと比較すれば、アコーディオン用のBBPTSJは、V-ACCORDIONの製品仕様に合わせてボタンと音素ラベルを対応させているため、規格の体系性、演奏上の合理性の両面でまだ発展途上であり、今後の改訂はあり得る。また五線譜による記譜の面でも、元来アコーディオン楽譜のボタンの指定はコードネームや和音による指定で、ボタンを一意に記述する記譜法ではない。この点でも「発声の指示」を確定的に記譜し、それが読めて演奏できる訓練された身体により発話/歌唱を行うというF兄弟の基本的コンセプトにアコーディオンは多少とも整合しない点がある。しかしながら、発声行為とアコーディオンの類似性は決して表面的なものではなく、インターフェイスと身体性の問題をさらに深く考察して行くうえでも、引き続き豊かな示唆を与え続けてくれるだろう。

(※本研究はJSPS科研費 23520175の助成を受けたものである)

## 【注釈】

- (1) 「録楽 roku-gaku」とは三輪による造語で、録音技術に代表されるメディアテクノロジーによってはじめて創造・伝達・聴取が可能になった音楽の形態を意味している。これまで「録音された音楽」を「音楽一般」から区別する言葉がなく、両者を安易に同一視して来た文化状況への批判から案出された。
- (2) 1937年ニューヨーク万国博覧会でHarperがVODERを操作している音声／映像記録をYoutubeで視聴することができる。  
(<http://youtu.be/0rAyrmm7vv0>)。ちなみに1960年代にDudleyがベル研究所を退職するに当たりVODERがはなむけの言葉を送る演出があったが、その時に操作を行ったのも彼女だったという(Gold, et al. 2011, p.13)。
- (3) 映画「2001年宇宙の旅」において、HAL 9000が破壊される時に最後に歌われる曲も「デイジー・ベル」であった。ただしその声は合成音声ではなく声優Douglas Rainのものである。

## 【引用文献】

- Dudley, H. and Tarnoczy, T.H. (1950), “The Speaking Machine of Wolfgang von Kempelen”, The Journal of the Acoustical Society of America, vol.22(2), pp.151-166.
- Fant, G., Liljencrants, J. and Lin, Q.(1985), “A four-parameter model of glottal flow”, STL-QPSR vol.26(4), pp.1-13.
- Flanagan, J.L., Allen, J.B. and Hasegawa-Johnson, M.A. (2008), “Speech Analysis Synthesis and Perception, Third edition”,  
<http://hear.ai.uiuc.edu/ECE537/PDF/main-all.pdf>
- Font, J. O.(2001), “Musical and Phonetic Controls in a Singing Voice Synthesizer”, Master Thesis, Polytechnics University of Valencia,  
<http://mtg.upf.edu/node/2248>
- Gold, B., Morgan, N. and Ellis, D. (2011), “Speech and Audio Signal Processing: Processing and Perception of Speech and Music”, Wiley-Interscience.
- Groundwater, J.(2005), “Alexander Graham Bell: The Spirit of Invention. Calgary”, Alberta, Canada: Altitude Publishing.
- Kelly, J. and Lochbaum, C.(1962). “Speech Synthesis”, In Proceedings of the 4th International Congress on Acoustics, paper G42, pp.1-4.
- Kempelen, W. V. (1791), “Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine”,  
<http://drl.ohsu.edu/cdm/compoundobject/collection/hom/id/3031>
- Klatt, D. H. (1980), “Software for a Cascade/Parallel Formant Synthesizer”, J. Acous. Soc. Amer. 67(3), pp.971-995.
- Klatt, D. H. (1987), “Review of text-to-speech conversion for English”, J. Acous. Soc. Amer. 82(3), pp.737-793.
- 赤川達也, 岩野公司, 古井貞熙 (2007), 「HMMを用いた話し言葉音声合成のためのモデルの検討」, 電子情報通信学会技術研究報告. SP, 音声107 (77), pp.13-18.
- 岡田暁生 (2008), 「ピアニストになりたい! 19世紀 もうひとつの音楽史」, 春秋社.
- 加々見翔太, 濱野桂三, 柏瀬一輝, 山本和彦 (2012), 「リアルタイム日本語歌唱鍵盤楽器“VOCALOIDキーボード”の開発」情報処理学会インタラクティブシンポジウム2012研究報告.
- 剣持秀紀, 大下隼人 (2008), 「歌声合成システムVOCALOID: 現状と課題」, 情報処理学会研究報告, 音楽情報科学2008(12), pp.51-56.
- 総務省統計局(2010), 平成21年全国消費実態調査「主要耐久消費財に関する結果」,  
<http://www.stat.go.jp/data/zensho/2009/taikyu/gai-menu.htm>
- 渡辺芳也 (1993), 「アコーディオンの本」, 春秋社.

## 【本文中で引用したフォルマント兄弟作品】

- 「兄弟deビザ注文」(2003), “Ordering a Pizza de Brothers!” , 初演: 2003年11月22日, 演奏: フォルマント兄弟 (MIDIキーボード).  
Festival Beyond Innocence, BRIDGE (大阪).
- 「フレディの墓／インターナショナル」(2009), “Le tombeau de Freddie/L’internationale”.
- 「NEO都々逸——六編」(2009), “NEO DO-DO-I-TSU - six Japanese folk songs -”, 初演: 2009年12月8日, 演奏: 岡野勇仁 (MIDIキーボード), 田中悠美子 (三味線), 同志社大学寒月館 (京都).
- 「せんだいドドンパ節」(2010), “Sendai Dodonpa Bushi”, 初演: 2010年8月28日, 演奏: 岡野勇仁 (MIDIキーボード), サントリー芸術財団サマーフェスティバル2010<MUSIC TODAY21> 芥川作曲賞創設20周年記念ガラコンサート, サントリーホール (東京).
- 「夢のワルツ」(2012), “Yume no Waltz”, 初演: 2012年8月26日, 演奏: 岡野勇仁 (MIDIアコーディオン), 佐近田展康 (ギター), サントリー芸術財団サマーフェスティバル25周年記念「ジョン・ケージ: ミュージサーカス」, サントリーホール (東京).

