Vol.2024-MUS-140 No.24
Vol.2024-SLP-152 No.24
2024/6/14

任意事例数多属性時変音声・歌唱モーフィングのWORLDへの移行とインタフェースの設計について

河原 英紀[1,a)]　森勢 将雅[2,b)]

概要：著者らが開発し公開してきた信号処理に基づく VOCODER である STRAIGHT は 1997 年に公開されて以来、音声知覚、音声合成、音声分析などの研究に広く用いられてきた。深層学習に基づく音声処理技術が人間に匹敵する音声の生成と認識を可能にした現在、信号処理に基づく VOCODER は、物理的な音声生成過程と心理的な音声知覚過程とそれらの基盤である生理学的機構の関係を理解するための手段としての役割が重要になっている。その役割を支援するために、STRAIGHT の上に構築されてきた様々なツールを、信号処理に基づく上位互換の VOCODER であり open source として公開されている WORLD の上に再構築する試みを進めている。ここでは、その最新の試みについて紹介する。具体的には、過剰な自由度と概念の複雑さなどが利用する際の障害となってきた拡張音声モーフィングを、より簡単に使うことができるようにするインタフェースの幾つかの提案について、議論して頂きたい。

Migration to WORLD vocoder and interface design of time-varing multi-attributes and multi-instances morphing for speech and singing

1. はじめに

著者らが開発し公開してきた信号処理に基づく VOCODER である STRAIGHT（以下、legacy-STRAIGHT） [1] は 1997 年に公開されて以来、その後継である TANDEM-STRAIGHT [2] とともに、音声知覚、音声合成、音声分析などの研究に広く用いられてきた。その後、信号処理に基づく上位互換の VOCODER である WORLD [3] が open source として公開されて広く用いられるに至っている。また、それらのために専用に設計された音源情報抽出方式（いわゆるピッチ抽出器）である、legacy-STRAIGHT 用の NDF [4] と WORLD 用の Harvest [5] も広く用いられている。

深層学習に基づく音声処理技術が人間に匹敵する品質の音声の生成と音声の認識能力を可能にした現在 [6,7]（原稿提出直前に発表されたこのビデオ [8] は必見）、これら信号処理に基づく音声処理に求められる役割は変わりつつある。音声は様々な生理学的・物理的制約を有する人体から生成され、環境による影響を受けまた環境に存在する様々な音と混合され、同様に物理的・生理学的制約を有する器官を持つ人体によって受け取られる。これらの生理的・物理的・心理的属性に関わるパラメタを対話的に操作し、実時間で音声として結果を確認できる環境を用意する上で（現時点では）信号処理に基づく VOCODER は最も近い位置にいる。WORLD は、信号処理に基づく VOCODER として広く使われ始めており、他の言語への移植や深層学習での利用も進んでいる。これらを背景に、これまで STRAIGHT の上に構築されてきたツールなどを WORLD に移植するとともに、インタフェースを更新することとした [9–11]。

以下では、まず信号処理に基づく VOCODER の構造と用いられるパラメタを簡単に説明したのち、VOCODER パラメタに基づくモーフィングを、歴史的経緯を踏まえながら説明し、任意事例数の音声から求められた VOCODER パラメタの各属性を時変の重みをつけてモーフィングできる拡張音声モーフィングを紹介する。次に、モーフィングを操作するツールに求められるものを議論し、試作した 3 事例モーフィングに基づくツールと、実験用の刺激連続対を作るツールについて説明する。なお、VOCODER パラメタを探索的に理解するためのツールについては、資料 [9–11] と GitHub のリポジトリ [12] で公開されている MATLAB による実装を参照されたい。

研究会のポスター発表では、これらのツールを実演を交えて紹介する。今後の進め方や拡張の方向などについて議論をお願いしたい。

[1] 和歌山大学
Wakayama University, Wakayama, Wakayama 640–8510, Japan
[2] 明治大学
Meiji Uniersity, Nakano, Tokyo 164–8525, Japan
[a)] kawahara@wakayama-u.ac.jp
[b)] mmorise@meiji.ac.jp

ⓒ2024 Information Processing Society of Japan

2. VOCODER の構造とパラメタ

信号処理に基づく VOCODER は、入力された音声信号をフィルタと駆動情報の二種類のパラメタに分解する。STRAIGHT では、分解されたパラメタから元の信号を再合成することを目的としない。STRAIGHT で目的とするのは、分解されたパラメタから生成された信号が、人間にとって同じものとして知覚されることである [1]。

駆動情報は、さらに周期成分と非周期成分に分解される。有声音などの周期的駆動は、時間的に変化するフィルタ情報を音波として伝搬できるものにするための組織的な標本化の手段であると解釈している。非周期成分には、周期成分として分解された残りがランダムであるとみなしたものが放り込まれている。破裂音を表現するための特別のパラメタを用意してはいない。別の表現をするなら、「位相情報は積極的に廃棄している」のである。さらに乱暴に言えば、「インチキだけれど本物らしく聴こえる」ようにする仕事を非周期成分に押し付けている。

これらのパラメタは、具体的には以下のように求められている。フィルタ情報は、短時間フーリエ変化によって求められるスペクトログラムから、周期的駆動によって生じる構造を選択的に除去したものである。選択的除去による操作がスペクトログラムの概形に与える影響を補償する仕組みが加えられている。周期成分は、基本周波数の（フレームごとの）値として表現されている。NDF や Harvest での音源情報の抽出では、瞬時周波数やイベントの時間間隔など複数の情報が併用されている。非周期性は、周期成分と非周期成分のパワーの比を表すスペクトログラムとして表現されている。非周期性の抽出法は、legacy-STRAIGHT [13]、Tandem-STRAIGHT [14]、WORLD（非周期成分抽出法は D4C [15]）のそれぞれで異なったものが用いられている。

3. 拡張音声モーフィング

STRAIGHT により求められる VOCODER パラメタに基づく音声モーフィングは、二事例間の補間としての定式化から始まり [16]、属性別で時変かつ外挿が可能なモーフィングに拡張された [17]。しかし、それらの定式化では、用いる事例の扱いが非対称であった。この非対称性の問題を解消し、さらに任意事例数多属性時変モーフィングへと一般化したものが拡張音声モーフィングである [18–20]。なお、STRAIGHT の上位互換 VOCODER である WORLD（と既定値で Harvest）によって求められる VOCODER パラメタは、STRAIGHT と同じ概念に基づいているため、拡張音声モーフィングは容易に WORLD に移植することができる [9–11]。

3.1 拡張音声モーフィングの処理

拡張音声モーフィングでは、前の節で説明した VOCODER パラメタに加えて、時間軸（bin の関数としての時刻）と周波数軸（bin の関数としての周波数）を加えたパラメタを処理の対象とする。具体的には、(1)**tx** 時間軸、(2)**fx** 周波数軸、(3)**fo** 基本周波数、(4)**sl** スペクトルレベル（周期性を除去されたスペクトログラム）、(5)**ap** 非周期性である。

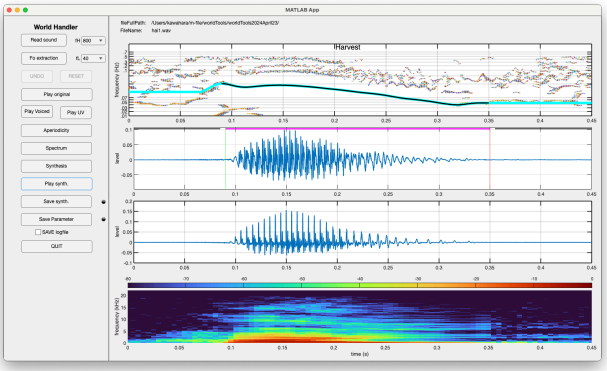


図 1 WORLD 用ツールを用いた VOCODER パラメタの抽出
Fig. 1 Snapshot of VOCODER parameter extraction using WORLD-based analysis GUI.

ここで、k 番目の事例音声（全体で K 事例あるとする）のある属性の値を $\mathbf{x}_k$ と表し、次式を用いてモーフィングされた属性の値 $\mathbf{x}_{\text{morph}}$ を求める。

$$\mathbf{x}_{\text{morph}} = \mathcal{T}^{-1}\left[\sum_{k=1}^{K} w_k \mathcal{T}[\mathbf{x}_k]\right], \text{ where } \sum_{k=1}^{K} w_k = 1, \quad (1)$$

ここで w_k は、k 番目の事例に対応する重みであり、$\mathcal{T}[\mathbf{x}_k]$ は値域に制限のある属性 $\mathbf{x}_k$ の値域を実数 $\mathbb{R}$ 全体に変換する関数を表し、$\mathcal{T}^{-1}[\,]$ は、その逆関数を表す。このようにすることにより、外挿してもモーフィングされた値を用いた生成が破綻しないことを保証している [18]。なお議論を複雑にしないよう、ここでは属性に対する重み w_k を定数としている。これを時変重みに拡張することは容易である。

3.2 モーフィング用の素材の準備

まず WORLD の分析用ツール（worldHandler）を用いて、モーフィング用の音声事例の素材を作る。図 1 に分析終了時のツールの GUI のスナップショットを示す。これは、男性が特別な感情表現を加えず（平静）に発声した「はい」という音声の分析結果である。ここで GUI の左側のパネルにある Save Parameter ボタンをクリックすることで、VOCODER パラメタをファイルに書き出すことができる。資料 [9,10] には、このツールの詳細と使用法が説明されている。また、著者の YouTube Channel（YouTube で “Hideki Kawahara” で検索する）にチュートリアルビデオを用意した。

次に行うことは、分析されて求められた異なる音声事例の VOCODER パラメタの対応づけである。対話的なツール（morphingAligner）を用いて対応付けを行う。図 2 に、二種類（平静と怒り）の感情表現で発声された「はい」の対応付けを行なった例を示す。

このスナップショットは対応付けが終わった状態である。右上のスペクトログラム表示と右下の音声波形表示では、二種類の感情表現音声を分析して求められた VOCODER パラメタが半透明（アルファチャンネルの値：0.5）で重ね書きされている。対応付けの目標は、この二つの表現の表

ⓒ2024 Information Processing Society of Japan

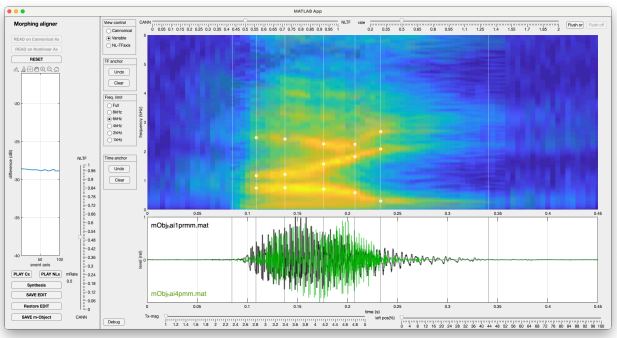


図 2 異なる音声事例の VOCODER パラメタの対応付け
Fig. 2 Snapshot of two sets of VOCODER parameter alignment using morphing alignment tool.

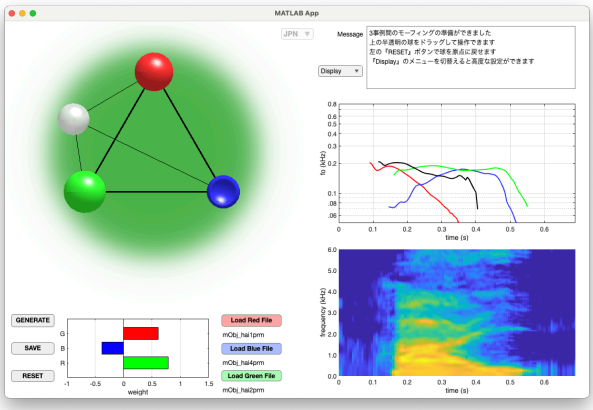


図 3 探索的研究用ツールの GUI。3 事例の VOCODER パラメタを操作している。
Fig. 3 Snapshot of three-way morphing demonstration using an interactive tool for generalized morphing for explorative investigations.

示の不一致が（スペクトログラム表示上で）最も少なくなるように調整することである。そのための手段として、図中に縦線で示された時刻手がかりと、上の図中に白丸で示された周波数手がかりを用いる。資料 [9,11] には、対応付ツールの詳細と使用法が説明されている。また、YouTube Channel にはチュートリアルビデオを用意した。

3.3 3 事例のモーフィング

前の節の手続きにより、拡張音声モーフィングに用いる素材を作ることができた。重みについての制約 $\sum_{k=1}^{K} w_k = 1$ を満たすように重み w_k を設定すれば、望むだけの数の音声素材からモーフィング音声を生成することができる。ここでは探索的研究に用いることを目的とした対話的ツール（図 3）と、その結果から主観評価実験用の刺激連続体を作るツール（図 4）を紹介する。

3.3.1 対話的探査ツール：threeItemMorpherGeneralUI

図 3 に、対話的探査ツール（threeItemMorpherGeneralUI）の GUI のスナップショットの例を示す。右上の文字が書かれた領域には、次に行うべき操作やツールの状態を説明するメッセージを表示する。（学生実験などでの経験から、

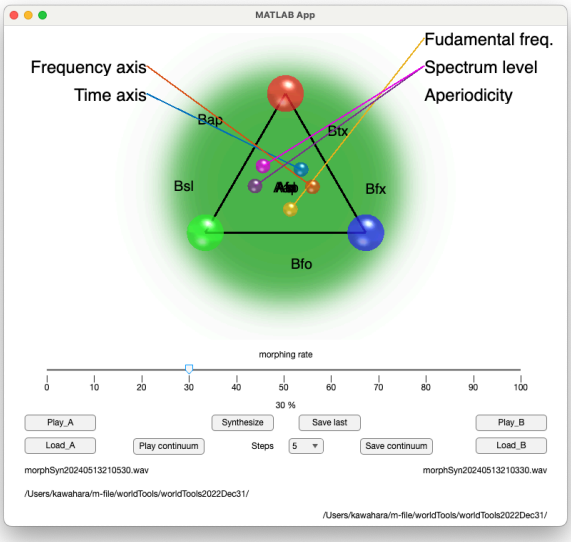


図 4 刺激連続体作成用ツールの GUI。全ての属性と事例に同じ重みを設定した条件 (A) と、全ての属性に異なった重みを設定した条件 (B) の間で、(A) から 30%離れた位置でモーフィングしている。
Fig. 4 Snapshot of stimulus continuum generation example. Condition (A) sets all instances and attributes to have the same weight. Condition (B) sets all attributes to have different weights. The snapshot shows morphing point is 30% from (A).

最近開発するツールでは、GUI に文章でメッセージを表示するようにしている [21]。）

図 3 の中央下部にある Load Red File、Load Blue File、Load Green File のボタンを用いて画面左上にある 3 個（赤、青、緑）の球体に、モーフィング用の素材を読み込むことで、拡張音声モーフィングを対話的に探索することができる。この 3 個の球体に連結されている半透明の灰色の球体をドラッグすることで、それぞれの事例の重みを設定する。それぞれの球体の体積は、重みに比例している。重みが負の場合には、球体は凹んだ椀の形に変形する。左下の中央の棒グラフは、設定された重みを表している。これらの表示と、モーフィングされた結果を表している右側の基本周波数とスペクトログラムの表示は、球体の移動に応じて更新され続ける。ドラッグを終えて球体を解放すると、その時の重みを用いてモーフィング音声が生成され再生音が提示される。

ここで左側の SAVE ボタンをクリックすると、生成された音声が WAV 形式でファイルに書き出される。このとき、モーフィングに用いた素材と重みなど再利用に必要な情報が WAV ファイルのメタデータとして、同時に書き出される。図 5 に、書き出されたメタデータの例を示す。ここで、Comment: 欄には 5 個の VOCODER パラメタ（**tx**、**fx**、**fo**、**sl**、**ap**）のそれぞれについて、（この場合は）3 個の事例（Red、Blue、Green）の重みが記入されている。また、Artist: 欄には、対応付けのツール（morphingAligner）から出力された 3 事例に対応するファイル名がフルパスで記入されている。この対話的探査ツール（threeItemMorpherGeneralUI）

ⓒ2024 Information Processing Society of Japan

```
Filename: '/Users/kawahara/m-file/worldTools/worldTools2022Dec31/morphSyn20240513210530.wav'
CompressionMethod: 'Uncompressed'
NumChannels: 1
SampleRate: 44100
TotalSamples: 25579
Duration: 0.5800
Title: 'source:,hai1.wav,hai4.wav,hai2.wav'
Comment: '[0.333  0.333  0.333],[0.333  0.333  0.333],[0.333  0.333  0.333],[0.333  0.333  0.333],[0.333  0.333  0.333]'
Artist: '/Users/kawahara/m-file/worldTools/worldTools2024April23/mObj_hai1prm.mat,/Users/kawahara/m-file/worldTools/worldTools2024April23/mObj_hai4prm.mat, ...
         /Users/kawahara/m-file/worldTools/worldTools2024April23/mObj_hai2prm.mat'
```

図 5 モーフィング音声ファイルに記録されるメタデータの例。Artist: 欄は文字数が多いため、ここでは折り返して表示している。

Fig. 5 Metadata stored in WAV format file of morphed sound. The actual Artist: field is one line. Folding is to dispaly in this Figure.

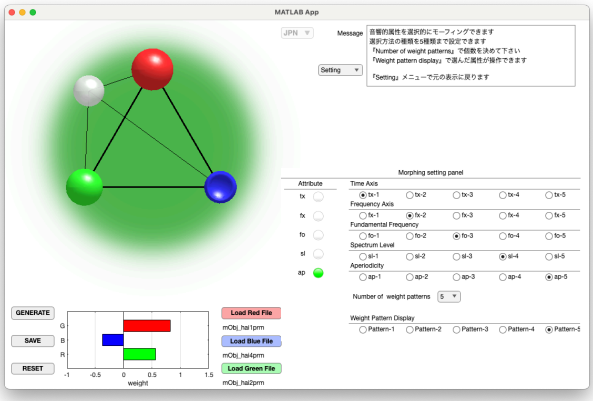


図 6 全ての属性に異なった重みを設定した例

Fig. 6 Snapshot of setting different weights for individual attributes.

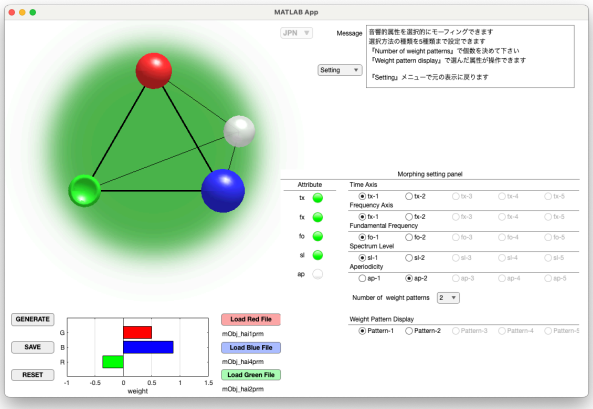


図 7 重み設定を二種類とした例

Fig. 7 Snapshot of setting weights for two patterns.

では、Artist: 欄の情報を利用して、モーフィングに必要な VOCODER パラメタを読み込む。

半透明の灰色の球体の位置を、VOCODER パラメタに対する重みに変換する際には、面積座標を用いている（ここでは文献 [22] を引用している。適切な教科書があればそちらを引用したい）。三原色の球を結ぶそれぞれの辺と灰色の球体で作られる三角形の面積が、辺に対向する頂点にある球に対応する事例の重みを与える。なお、RESET ボタンをクリックすることで、すべての事例が同じ重みでモーフィングされ、半透明の球体は三角形の中央に戻る。

既定値では、すべての属性に同じ重みが用いられている。メッセージの表示領域の隣にある Display と書かれたプルダウンメニューを Setting に切り替えると、図 6 に示す様に、それぞれの属性に異なった重みを設定することができるようになる。設定できる重みの個数の上限は、属性の数である。Weight Patter Display と書かれたボタングループから一つのボタンを選択し、左側の半透明の球をドラッグすることで選択された属性に重みを設定する。このスナップショットでは、5 個の属性全てに異なった重みを設定している。

図 7 には、重みの設定を二種類のパターンとした場合の例を示す。ここでは、Pattern-1 に重みを設定している。選択されたパターンでどの属性が設定されるかは、中央の Attribute と書かれたパネルのラジオボタンに緑色が表示されることで示される。

3.3.2 刺激連続体作成ツール：morphContinuumGen

刺激連続体作成ツールは、前の節で紹介したツールが

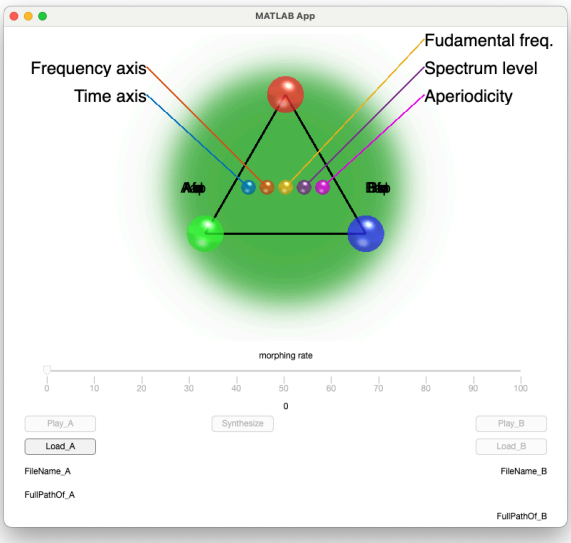


図 8 刺激連続体作成ツールの初期画面。

Fig. 8 Snapshot of the initial state of the continuum generation tool.

出力した WAV 形式のファイルを 2 個読み込んで、それらを両端とする刺激連続体を作る。図 8 に、このツール（morphContinuumGen）を立ち上げた直後のスナップショットを示す。このツールは操作が複雑ではないので、メッセージを表示せずに操作可能な GUI ツールだけを活性化することにした。この例では、モーフィング設定 (A) の音声ファイルを読み込むための Load_A ボタンだけが操作可

ⓒ2024 Information Processing Society of Japan

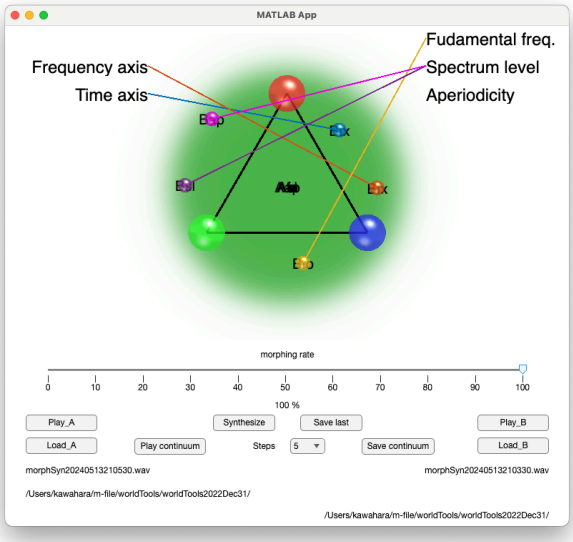


図 9 刺激連続体作成ツールでモーフィング率を 100%とした例。
Fig. 9 Snapshot of setting morphing rate 100%.

表 1 モーフィング音声 (B) に用いられた重みの設定
Table 1 Weight setting of morphed speech (B).

	Red	Blue	Green
time axis	0.737	0.456	-0.193
frequency axis	0.322	0.899	-0.221
fundamental frequency	-0.217	0.709	0.508
spectrum level	0.340	-0.303	0.963
aperiodicity	0.817	-0.376	0.558

能であり、活性化されている。

図 9 に、全ての属性と事例に同じ重みを設定したモーフィング音声 (A) と、全ての属性に異なった重みを設定したモーフィング音声 (B) を読み込んで、モーフィング率を 100%に設定したときのスナップショットを示す。図 3 に示したメタデータはモーフィング音声 (A) のものである。表 1 にモーフィング音声 (B) に設定された重みを示す。

モーフィング率を 0%に設定した場合には、それぞれの属性の位置を示す小さな球が全て（Atx、Afx、Afo、Asl、Aap が重ね書きされている）中央に集まる。モーフィング率を 100%に設定した場合には、図 9 に示すように、(B) のそれぞれの属性が設定された位置（Btx、Bfx、Bfo、Bsl、Bap）に、それぞれの小さな球が位置している。なお、モーフィング率は morphing rate と書かれた水平のスライダーで設定できる。設定されたモーフィング率は、スライダーの直下に数値で表示される。スライダーをドラッグしている間は、モーフィング率の変化に応じて表示が変化する。スライダーを解放してモーフィング率が確定すると、モーフィング音声が生成されて再生音が提示される。

対話的な予備検討で刺激連続体の仕様が決まったら、Steps のプルダウンメニューで段階数を決めて Play continuum ボタンで音を再生して確認する。良ければ Save continuum ボタンで刺激連続体の音をファイルに記録する。

GUI 上のその他のボタン類は、途中で音を再生して確認したり、記録するために用意してある。また、モーフィング音声のファイル名と格納されているディレクトリの情報をツールの最下部に表示している。

4. 課題

拡張音声モーフィングの WORLD への移植の際に、MATLAB と新しいアプリケーションの開発環境である App Designer を使うことにより、STRAIGHT に基づく従来のツール類 [19] よりも使い易いものとすることができた。このような改良したツールを用いても、モーフィング用の素材を用意するための対応付けは、時間を要しかつ難しい作業であり、自動的な支援方法を検討する必要がある。対応付けの目的はモーフィングする際の不一致を少なくするように関数を定義することである。現在の実装で用いている時間および周波数の対応点を用いる方法は、理解と実装が容易なことから出発点として選択した以上の意味はない。

拡張音声モーフィングを応用して 3 事例のモーフィングのためのツール類を紹介した。N 次元空間内に $N+1$ 個の頂点を持つ単体の頂点に事例を配置することで、面積（体積）座標を用いて合理的に重みを設定することができる。しかし、事例数が 5 以上の場合には、このような合理的で直接的な方法を用いることができない。どのような方法が可能か、議論をお願いしたい。

拡張音声モーフィングでは、時変重みを用いることができる。3 事例のモーフィングのためのツール類のインタフェースにタイムラインを追加して設定を配置する方法を考えている。歌唱合成・加工ツールなどの優れた例（多数のため網羅できないが、例えば [23,24] など）を参考に、検討したい。

深層学習に基づく音声処理技術が人間に匹敵する品質の音声の生成と音声の認識能力を可能にしたことを背景として [6–8]、信号処理に基づく VOCOER の位置付けが音声の生成と知覚の（暗黙知を含む）知識獲得の支援に向かうとして一連のツールの開発を進めてきた。（広い意味での）応用に対してはマルチモーダルな end-to-end 型の深層学習に基づくシステムの性能は圧倒的である。従来の分轄統治と個別最適化に向かう知識としてではなく、深層学習システムに介入する際の適切な界面の実現を支える基盤とする形での知識獲得が必要ではないかと考えている。

5. おわりに

ここでは、信号処理に基づく VOCODER である WORLD の上に、これまで STRAIGHT に基づいて開発されてきたツール類を移植する最新の試みと、これからの課題についての議論を紹介した。ここで紹介したツール類は著者の GitHub リポジトリで公開しており [12]、ビデオチュートリアルも YouTube Channel で公開している。音声生成と知覚、物理現象との関連を体験的に把握することに役立てて頂けることを期待している。

謝辞 ここで紹介したツールの以前の版を試用して様々なフィードバックを頂いた、和歌山大学 入野俊夫 教授、豊橋技術科学大学 松井 淑恵 教授、東海大学 程島 奈緒 教

ⓒ2024 Information Processing Society of Japan

授と学生の方々に感謝します。本研究は、科学技術振興機構の科研費 JP20H00291, JP21K19794, JP21H00497, and JP21H04900 の支援を受けて行われた。

参考文献

[1] Kawahara, H., Masuda-Katsuse, I. and de Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds, *Speech communication*, Vol. 27, No. 3-4, pp. 187–207 (1999). DOI: 10.1016/S0167-6393(98)00085-5.

[2] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, *Proc. ICASSP 2008*, pp. 3933–3936 (2008). DOI: 10.1109/ICASSP.2008.4518514.

[3] Morise, M., Yokomori, F. and Ozawa, K.: WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications, *IEICE transactions on information and systems*, Vol. 99, pp. 1877–1884 (2016). DOI: 10.1587/transinf.2015EDP7457.

[4] Kawahara, H., de Cheveigné, A., Banno, H., Takahashi, T. and Irino, T.: Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT, *Proc. Interspeech 2005*, pp. 537–540 (2005). DOI: 10.21437/Interspeech.2005-335.

[5] Morise, M.: Harvest: A High-Performance Fundamental Frequency Estimator from Speech Signals, *Proc. Interspeech 2017*, pp. 2321–2325 (2017). DOI: 10.21437/Interspeech.2017-68.

[6] Cooper, E., Huang, W.-C., Tsao, Y., Wang, H.-M., Toda, T. and Yamagishi, J.: A review on subjective and objective evaluation of synthetic speech, *Acoustical science and technology / edited by the Acoustical Society of Japan*, Vol. (Advanced online publication), p. e24.12 (2024). DOI: 10.1250/ast.e24.12.

[7] Tan, X., Chen, J., Liu, H., Cong, J., Zhang, C., Liu, Y., Wang, X., Leng, Y., Yi, Y., He, L., Zhao, S., Qin, T., Soong, F. and Liu, T.-Y.: NaturalSpeech: End-to-End Text-to-Speech Synthesis with Human-Level Quality, *IEEE trans. PAMI*, Vol. (Advanced online publication) (2024). DOI: 10.1109/TPAMI.2024.3356232.

[8] OpenAI: Live demo of GPT-4o realtime translation, OpenAI YouTube channel (2024). https://www.youtube.com/watch?v=c2DFg53Zhvw, (visit: 14/May/2024).

[9] 河原　英紀，森勢　将雅：音声分析合成基盤 WORLD の GUI 実装と見えてきた課題，信学技報, 音声研究会，Vol. 122, No. SP-81, pp. 3–6 (2022). [信学会：ダウンロードリンク].

[10] Kawahara, H. and Morise, M.: Interactive tools for making vocoder-based signal processing accessible: Flexible manipulation of speech attributes for explorational research and education, *Acoustical science & technology*, Vol. 45, No. 1, pp. 48–51 (2024). DOI: 10.1250/ast.e23.52.

[11] Kawahara, H. and Morise, M.: Interactive tools for making temporally variable, multiple-attributes, and multiple-instances morphing accessible: Flexible manipulation of divergent speech instances for explorational research and education, *arXiv cs.HC*, Vol. eprint: 2404.13418, pp. 1–5 (2024). https://arxiv.org/abs/2404.13418.

[12] 河原　英紀：WORLD GUI tools for explorational research/education of speech science and technology, GitHub (2023). https://github.com/HidekiKawahara/worldGUItools, (visit: 13/May/2024).

[13] Kawahara, H., Estill, J. and Fujimura, O.: Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT, *Proc. 2nd MAVEBA 2001*, Florence, Italy, isca-speech.org, pp. 59–64 (2001). [pdf on ISCA archive].

[14] Kawahara, H., Morise, M., Takahashi, T., Banno, H., Nisimura, R. and Irino, T.: Simplification and extension of non-periodic excitation source representations for high-quality speech manipulation systems, *Proc. Interspeech 2010*, pp. 38–41 (2010). DOI: 10.21437/Interspeech.2010-6.

[15] Morise, M.: D4C, a band-aperiodicity estimator for high-quality speech synthesis, *Speech Communication*, Vol. 84, pp. 57–65 (2016). DOI: 10.1016/j.specom.2016.09.001.

[16] Kawahara, H. and Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation, *Proc. ICASSP 2003*, Vol. I, IEEE, pp. 256–259 (2003). DOI: 10.1109/ICASSP.2003.1198766.

[17] Kawahara, H., Nisimura, R., Irino, T., Morise, M., Takahashi, T. and Banno, H.: Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown, *Proc. ICASSP 2009*, pp. 3905–3908 (2009). DOI: 10.1109/ICASSP.2009.4960481.

[18] Kawahara, H., Morise, M., Banno, H. and Skuk, V. G.: Temporally variable multi-aspect N-way morphing based on interference-free speech representations, *Proc. APSIPA ASC 2013*, pp. 1–10 (2013). DOI: 10.1109/APSIPA.2013.6694355.

[19] 河原　英紀：音声の実時間表示とモーフィングで探る声の多様性，音声研究， Vol. 18, No. 3, pp. 43–52 (2014). DOI: 10.24467/onseikenkyu.18.3_43.

[20] Skuk, V. G. and Kawahara, H.: Chapter 31: Voice Morphing, *The Oxford Handbook of Voice Perception* (Frühholz, S. and Belin, P., eds.), Oxford University Press, pp. 685–706 (2018). DOI: 10.1093/oxfordhb/9780198743187.013.31.

[21] 河原英紀，　矢田部浩平，　榊原健一，　水町光徳，　北村達也，　牧勝弘，　坂野秀樹：多様な測定信号を用いた音声の収録および聴取環境の測定例について，日本音響学会 2024 年春季研究発表会 3–P–21，pp. 721–724 (2024).

[22] Li, Q. Y., Zheng, J. R., Liao, G. and Jin, Y.: Approach on area coordinate, volume coordinate and their application in true 3DGIS, *Journal of Earth Science and Engineering*, Vol. 1, pp. 52–59 (2011).

[23] Misra, A., Wang, G. and Cook, P.: Musical tapestry: Re-composing natural sounds, *Journal of new music research*, Vol. 36, No. 4, pp. 241–250 (2007). DOI: 10.1080/09298210801984839.

[24] Dreamtronics: Vocoflex - Real-time Voice Morphing Plugin, Dreamtonics YouTube channel (2024). https://youtu.be/ORiIsU2D0Sk?si=hHYI8GfQXF7E_ozE, (visit: 14/May/2024).

ⓒ2024 Information Processing Society of Japan