

# Parrot: リアルタイム音声合成を用いたビブラートデザイン支援 インタフェースの開発

小野 雄大\* 森勢 将雅†‡

**概要.** 歌声の合成・加工技術は VOCALOID の発売から急速に発展し、現在では多くのユーザが楽曲制作に利用している。これらのソフトウェアを用いて自然に歌わせるためには、歌唱パラメータを調整する煩雑な作業（これを歌唱デザインと呼称する）が発生する。統計的手法に基づくアプローチは、人間とほぼ等価な品質の歌声を自動生成することで歌唱デザインを不要としたが、人間の歌声であれば必ずしも加工が不要であるわけではない。ユーザの好みに応じて意図的に逸脱させた表現を作り込む歌唱デザインには、一定の需要がある。本研究では、歌唱デザインの省力化を目標とし、リアルタイム音声合成技術を用いて、ビブラートのデザインと結果の確認を同時に行うインタフェース Parrot を提案する。このインタフェースについて、作業時間やユーザビリティに関する評価を実施した。実験の結果、提案した Parrot は、既存のものより作業時間を有意に短縮し、ユーザビリティに関する評価値は有意に向上することが示された。

## 1 はじめに

ニコニコ動画や YouTube において、コンピュータを用いた楽曲コンテンツが日々投稿されユーザに楽しられている。特に 2000 年代に入ってから、VOCALOID [1] や UTAU [2] といった歌声合成ソフトも開発され、合成歌唱は楽曲コンテンツにおける 1 つの地位を確立しているといえる。歌声合成ソフトウェアは、縦軸にピアノを記載し、横軸を時間とするピアノロールと呼ばれるエディタにより歌詞と譜面に相当する情報を与えることで歌声を合成できる。ただし、歌詞と譜面だけで合成された歌声は人間の歌声とは異なる機械的な印象を与えるため、歌声を構成するパラメータをユーザが調整することで品質を高める操作が必要不可欠である。本稿では、このパラメータ群を操作する一連の作業を歌唱デザインと定義する。

歌唱デザインは楽曲制作において重要な役割を担う一方、デザインに関する技能の習得が困難であり、また、その作業もストレスが大きいことが現状の課題といえる。この課題の解決に向けて、歌唱デザインの省力化を目指した研究 [3, 4] や、人間による歌唱デザインが不要なほど高い品質の歌声を合成する技術 [5, 6] など、様々なコンセプトの研究がなされている。本研究では、歌唱デザインは煩雑であるが、その行為により生成される歌声にユーザの「個性」が現れる点に着目する。楽曲コンテンツでは、同じ楽曲に歌声を加える場合でも、歌唱デザインする人により表情付けが異なり、それぞれに独特の魅力が

付与される。本研究の狙いは、作業時間と作業にかかるストレスを緩和する歌唱デザイン技術を開発することで、ユーザが歌唱デザインにより個性を作り込む作業をサポートすることである。

本稿は、主要な歌唱表現の 1 つであるビブラートに着目し、容易に付与するための技術とインタフェース開発について述べる。はじめに、関連研究と本研究との位置づけについて述べ、本研究で提案する技術のコンセプトを説明する。実装したインタフェースについて既存のインタフェースと比較し、作業時間の短縮やユーザビリティの改善ができることを示す。

## 2 歌唱デザインに関する関連研究と本研究の位置付け

ここでは、まず本研究でビブラートを扱う背景について説明し、歌唱デザインに関する関連研究について述べる。その後、関連研究に対する本研究の位置付けを説明する。

### 2.1 本研究で取り扱う歌唱パラメータ

歌声には高さの個性だけでも多様にあることが報告されている。例えば、ビブラート以外にポルタメント、オーバーシュート、プレパレーションに関する検討がなされている [7]。ポルタメントとは、ある声の高さから別の声の高さへの遷移を滑らかにする歌唱表現である。オーバーシュートは、ある声の高さから別の声の高さに遷移したときに、目的の高さを通り越し瞬時的に高く、あるいは低くなる現象である。プレパレーションは、ある声の高さから別の声の高さに遷移する前に、瞬時的に目的の高さと逆に声の高さが変化する現象である。これらの特徴を付与することで、歌声の品質が向上することが示さ

Copyright is held by the author(s).

\* 山梨大学

† 明治大学

‡ JST さきがけ

れている [7].

ビブラートは知覚する音高を維持したまま声の高さを小刻みに振動させる歌唱表現である。この振動は準周期的であるため、振幅と周波数が時間変動する正弦波により表現される [8, 9]。そのため、ビブラートは表現の幅が広く、歌手による個人差が確認されている [10]。次節で説明するように、歌声を扱う多くのソフトウェアにはビブラートを制御するための機能が搭載されていることから、ビブラートデザインは需要が高い機能といえる。ビブラートデザインをサポートする機能の実現は、歌唱デザインの省力化に寄与すると考えられる。

## 2.2 歌声合成ソフトウェアによるビブラートデザイン

波形から歌声を生成する歌声合成ソフトウェアにも、ビブラートを制御するインターフェースが搭載されている。以下では代表的な事例について述べる。

VOCALOID 4はビブラートの振幅 (vibrato extent) と周波数 (vibrato rate) それぞれについて時系列で編集するインターフェースが実装されている。また、音符の途中からビブラートをかける機能に対応するため、全体に対するビブラート区間の比率 (vibrato duration) もパラメータとなる。時系列として編集できるため、振幅と周波数を時間変動させることも可能である。VOCALOID 4によるビブラートデザインの利点は、ユーザが詳細なビブラートデザインを行える点である。一方、詳細にデザインできる利点は、作業の煩雑化に繋がるのが課題となる。

VOCALOID Vによるビブラートデザインでは、VOCALOID 4とは異なり、時系列の詳細なデザインからプリセットを用いる形に変更されている。ユーザは、多数あるプリセットから所望するビブラートに近いものを選び、その後、ビブラートの区間と振幅、周波数の調整を行う。ビブラートの振幅と周波数は、横軸と縦軸に振幅と周波数を割り当てた2軸のインターフェースにより同時指定が可能である。VOCALOID Vの利点は、プリセットの選択によりビブラートの大きな形を指定してからデザインするため、デザインにかかる時間が短くなる点にある。一方、プリセットは有限であり人間のビブラートに基づいて設定されているため、プリセットにない詳細な表現を作り込むことに対する課題が残る。VOCALOID 4と共通する問題として、エディタによるデザインから結果の確認までタイムラグが存在することが考えられる。

## 2.3 歌唱デザインの省力化・自動化

VocaListener [3] は、人間の歌から VOCALOIDのパラメータを得ることで人間の歌い方を模倣するシステムである。扱うパラメータは声の高さと声の

大きさであり、推定後手作業で調整することも可能である。ユーザが目標とする歌声を用意することで、エディタによる微調整無く所望する歌声を得ることができる点が、VocaListenerの利点といえる。しかしながら、これは、目標とする歌い方が可能な歌手を用意する必要があることを意味する。

統計的歌声合成は、歌詞と譜面の情報のみから自然な人間らしい歌声を自動的に合成する技術である。これにより、ユーザは歌唱パラメータのデザインをすることなく自然な人間らしい歌声を得ることができる。HMM (hidden Markov model) を用いた歌声合成技術 [11] を用いたものが中心であったが、DNN (Deep neural network) の利用による品質が大きく向上した。さらに、2016年に提案された WaveNet [12] を歌声合成に応用した技術 [13] では、人間とほぼ等価な品質が達成されている。これらの統計的歌声合成技術では、ビブラートのかかった歌声も得ることができる [14]。統計的歌声合成の利点は、ユーザが歌詞と譜面といった最小限の情報から人間らしい歌声を得ることができるため、歌唱デザインを行う必要がない点である。

## 2.4 ボーカル補正ソフトウェア

Auto-Tune [15] や Melodyne [16], Waves Tune [17] などは、ボーカル補正の機能を持つソフトウェアとして販売されており、ビブラートを制御・付与する機能も実装されている。

Auto-Tuneではビブラートをかけたい音符を選択し、振幅や周波数などに関する7つのパラメータの調節でビブラートデザインを行う。また、Auto-TuneのLiveパッケージ以上には、GSnap [18]にも実装されているリアルタイムビブラート調整機能が搭載されており、歌声を聴きながらパラメータを調整することができる。歌声の再生中にパラメータを操作することで、デザイン結果の確認が即時にできることが、Auto-Tuneの利点である。しかし、操作インターフェースが各パラメータに対し1つずつなため、複数のパラメータの同時制御は困難である。

Melodyneには、音の高さの揺らぎを制御する機能があり、ビブラートのデザインに適用することが可能である。ビブラートの制御をしたい音符を選択し、音の高さの揺らぎをマウスの上下操作で増減することで、ビブラートの振幅の増減を行う。Melodyneの利点は、ビブラートの振幅をマウス操作のみで調整することができ操作が容易なことである。しかし、Melodyneはビブラートデザインに特化したソフトウェアでないため、ビブラートデザインの専用エディタは搭載されていない。

Waves Tuneは、NaturalとSynthの2つの種類のビブラートをノート単位で付与することができる。Naturalでは、ビブラートの量とかける割合の2つのパラメータで自然なビブラートをかけること

ができる。Synthでは、4つのパラメータでより詳細なビブラートデザインを行う。4つのパラメータは、振幅、周波数、音符の始めからビブラートの開始までの時間、ビブラートの開始から設定したパラメータに至るまでの時間である。NaturalとSynthの2つの方法により、簡単なデザインから詳細なデザインにも対応しているのがWaves Tuneの利点である。しかし、VOCALOIDと同様に、結果を即座に確認することができない。

## 2.5 本研究の位置づけ

歌声合成ソフトウェアやボーカル補正ソフトウェアでの歌唱デザインは、大雑把に作り込める場合は繊細な作り込みが困難であり、繊細に作り込もうとすると作業の手間がかかる。デザイン結果の即時確認機能はあるものの、複数のパラメータを同時に制御することが課題となる。デザイン結果を確認するまでのタイムラグやパラメータ設定の煩雑さが歌唱デザインの難しさだと考えられる。

本研究では、エディタによる編集結果が即時出力されることと複数のビブラートパラメータを同時に制御する手法を提案することで、ビブラートデザインの省力化を図る。

## 3 ビブラートデザイン支援インタフェースの提案

本章では、まず、提案手法のコンセプトと求められる機能について述べる。その後、提案手法の実現に使用した技術と開発したインタフェースParrotについて述べる。

### 3.1 デザイン支援に求められる機能

本研究の目的はビブラートを容易に作り込める機能の実現であり、利用者はビブラートのデザインに精通していない初心者から中級者を想定している。既存のソフトウェアが有するビブラートのパラメータのうち、振幅と周波数の時系列をリアルタイムで制御する機能を実現する。編集結果を逐次確認しながらデザインすることが、歌唱デザインの省力化に寄与するという仮説を検証するためのインタフェースを実装する。このような機能の実現には、ビブラートのパラメータを制御しながら歌声を合成する技術が必要となる。

本研究では、品質の高いビブラートの実時間制御という機能を満足する音声合成技術として、音声分析合成システムWORLD [19]を利用する。WORLDは、音声の基本周波数(F0)、スペクトル包絡、非周期性指標に分解し、それぞれを制御してから再合成することで声質の変換を実現するVocoder型 [20]の音声分析合成システムである。ビブラートデザインは、基本周波数の軌跡を制御することで実現可能となる。また、WORLDは、実時間音声合成を

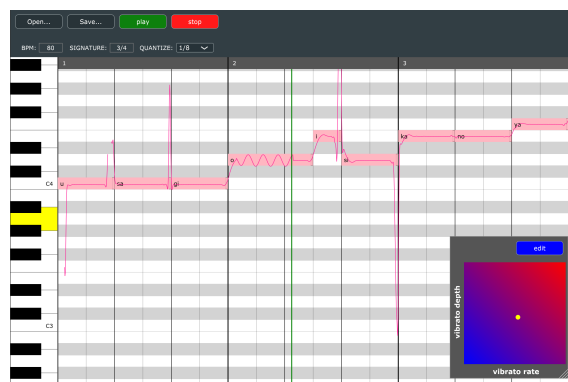


図 1. ビブラートデザイン支援インタフェース Parrot

現するための拡張が行われている [21]。この拡張したWORLDを用いることにより、歌声の高さを再生中にリアルタイムで制御することが可能となる。

### 3.2 ビブラートデザイン支援インタフェース Parrot の提案

図 1 に示す Parrot では、右下にある 2 次元のエディタを用いてビブラートのリアルタイム加工を実現する。このエディタの縦軸と横軸はそれぞれビブラートの振幅と周波数に対応する。楽曲の再生中に任意の場所をマウスやタッチペンでクリックすると、その座標から得られる振幅、周波数のビブラートの付与された歌声が得られる。この機能により、必要とされるデザイン結果を確認しながら編集するという条件を満足させ、ビブラートの振幅と周波数、複数のパラメータの同時制御が実現される。

Parrotを開発したOSは、Windows 10 Proで、統合開発環境はVisual Studio 2017、開発言語はC++とした。ParrotのGUIやオーディオ出力、ファイルの読み書きなどを実装するにあたり、ROLI社が提供するフレームワークJUICEを使用した。

### 3.3 Parrot の機能

本研究で目指している最終的な歌唱デザインインタフェースでは、VOCALOIDのように歌詞と譜面から編集元となる歌声を生成する機能を実装する予定である。試作したParrotでは、ビブラートデザインに特化して評価するため、他のソフトウェアで生成された歌声と譜面データを読み込み、デザインの対象とする。音声を読み込むと、WORLDによりF0を推定し、推定結果をピアノロール上に表示する。図1の右下に示されたビブラート編集エディタ上のeditボタンを押下すると歌声が再生される。ユーザは、歌声の再生中にエディタを操作し、ビブラートデザインを行う。デザインと同時にF0軌跡のレンダリング、ビブラートのかかった歌声の合成と再生を同時に行うことができるため、歌声の変化が即座に確認できる。

リアルタイム合成が可能なWORLDでは、F0情

報を逐次与えることで再生される歌声のF0を制御できる。ビブラートの付与にあたり、エディタからビブラートの振幅  $A$  と周波数  $f_i$  を取得する。次に、変換対象のフレーム  $n$  のビブラートを式 (1) を用いて計算し、 $F0_n$  に重畳する。

$$\Delta F0_n = A \times F0_n \times \sin \left( 2\pi T \sum_{i=1}^n f_i \right) \quad (1)$$

ここで、 $T$  はフレーム間隔を示す。本インタフェースの評価では WORLD のデフォルトパラメータである 5 ms に設定している。最後に、変換した F0 と元のスペクトル包絡および非周期性指標を用いて音声合成する。これにより、ビブラートが付与された歌声が合成される。

2次元インタフェースでは実時間の制御しか実現できないが、編集内容から振幅、周波数の時系列を得ることができる。編集結果を対象として既存のエディタと同様に詳細な作り込みを行うことも可能である。

#### 4 Parrot の評価実験

本章では、提案手法の有効性を検証するために行った評価実験の実験内容について説明する。提案手法の目標である作業時間の短縮とユーザビリティを評価するため、ビブラートデザイン作業とアンケート評価を併せて評価する。

##### 4.1 比較する従来手法

実験では、従来手法として VOCALOID 4 Editor のビブラートデザインインタフェースを使用する。最新の VOCALOID V Editor を従来手法として使用せず VOCALOID 4 Editor を使用する理由は、VOCALOID V Editor のプリセットにないようなビブラートも含んだデザインを実施するためである。ただし、作業時間や回数を正確に記録するため、ビブラートの振幅と周波数を時系列で表現する VOCALOID 4 Editor の機能に加え、それらの情報を逐次記録できるようなインタフェースを実装することとした。

##### 4.2 実験方法

被験者が自由にビブラートをデザインする場合、被験者のスキルや思い描くビブラートの差などが結果に影響することが予想される。ここでは、手本となるビブラートの音源を聴取させ、聴いたビブラートを基に手本を再現する実験を行わせることとした。評価用インタフェースには手本を再生する機能を与え、被験者は自由なタイミングで手本となるビブラートを聴取することが可能である。被験者が再現できたと判断するまでビブラートデザインを実施し、本人の申告により終了することとした。

表 1. 質問項目

1	このシステムを今後も繰り返し使いたい
2	このシステムは必要以上に複雑だった
3	このシステムは簡単に使えた
4	このシステムを使えるようになるには、技術に詳しい人のサポートが必要だ
5	このシステムは機能がよくまとまっていた
6	このシステムは一貫性に欠けるところが多かった
7	たいていの人は、このシステムをすぐに使えるようになるだろう
8	このシステムはととても使いづらかった
9	このシステムを使っていて不安を感じることはなかった
10	このシステムを使い始める前に、多くのことを学ぶ必要があった

表 2. 手本となるビブラートの条件

パラメータ	種類
Vibrato duration	3 種類 (50, 75, 100 %)
Vibrato extent	2 種類 (30, 100 cent)
Vibrato rate	2 種類 (5, 8 Hz)

比較項目は、国際規格 ISO 9241-11 [22] から、有効さ、効率、満足度とした。有効さを手本音声とデザインした音声との類似度により、効率をデザインに要した時間と回数により、満足度を質問紙による主観評価で検証することとした。質問紙の内容は、SUS (System Usability Scale) [23] の日本語訳 [24] を用いた。質問紙の内容を表 1 に示す。

##### 4.3 手本音声の条件

手本となるビブラート歌唱には、VOCALOID 4 で作成した音声を対象に、表 2 に示す条件を組み合わせた 12 種類の条件を用いた。それらに、右田らの作成した歌声データベース [10] の中から 4 人の発話者のビブラート歌唱も手本に加えた計 16 種類を最終的な手本とした。ただし、ビブラート歌唱をそのまま手本とするのではなく、歌声の F0 を推定し、推定結果を非ビブラート歌唱に転写した音声を手本として使用する。この理由は、ビブラート歌唱と非ビブラート歌唱は声の大きさなどの F0 以外の成分も異なり、提案手法と従来手法でビブラート歌唱を再現することが困難になるためである。

##### 4.4 実験手順

実験は、初めに練習セッションとして、提案手法と従来手法それぞれのインタフェースで 1 つのビブラートをデザインする。その後、順序効果を考慮し、手本となるビブラートだけではなく、従来手法・提案手法のインタフェースを使う順序まで全てランダム化した本実験を実施した。楽曲制作経験による差を考慮するため、ここでは楽曲経験のない被験者

表 3. 評価実験の実験条件

評価内容	1 音声ごとの作業時間と回数
手本音声数	16 音声
評価者数	楽曲制作経験の無い学生 15 名
評価環境	一般的な教室 (A-weighted SPL: 35 dB)
再生機材	Roland QUAD-CAPTURE SENNHEISER HD 650

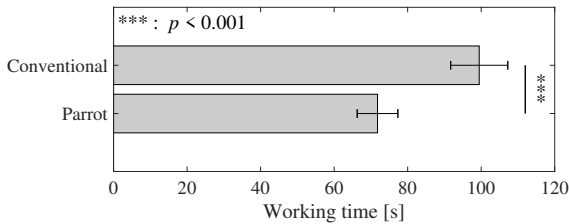


図 2. 作業時間に関する評価結果

で統一した。実験では、作業回数、作業時間を記録し、実験後に質問紙による評価を実施することとした。その他の実験条件を表 3 に示す。

#### 4.5 実験結果

作業時間、作業回数の結果を図 2, 3 に示す。図中の誤差棒は、95%信頼区間を表す。提案手法が約 72 秒、従来手法は約 99 秒で、提案手法の方がビブラートデザインに要する時間が約 27 秒短く、この差は  $p < 0.001$  で有意であった。一方、作業回数については、提案手法が約 7 回、従来手法は約 6 回で、提案手法の方がビブラートデザインに要する回数が有意に約 1 回多かった。

次いで、質問紙による主観評価として SUS 値を計算した結果を図 4 に示す。SUS 値は 100 点満点で、値が大きいほど評価が高い指標となる。提案手法の SUS 値は  $p$  値が 0.05 未満であり、従来手法よりも高評価であることが認められた。

最後に、手本音声に対するビブラートの区間、振幅、周波数の誤差について、文献 [10] を参考に分析した結果を表 4 に示す。それぞれについて有意差検定を行ったが、全てにおいて  $p$  値が 0.05 以上であり、有意差が認められなかった。

### 5 考察

ここでは、提案した Parrot の有効性について、有効さ、効率、満足度の観点から議論する。

#### 5.1 有効さと満足度に関する考察

Parrot と従来手法のインタフェースにより作られたビブラートパラメータは、有意差が認められないことから、概ね同等の特徴を持つビブラートを作ることができたといえる。ビブラートデザインを終了する基準である「被験者自身が納得する」に関しては、例えば複雑なビブラートの場合作業そのもの

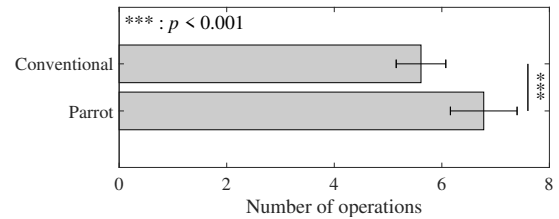


図 3. 作業回数に関する評価結果

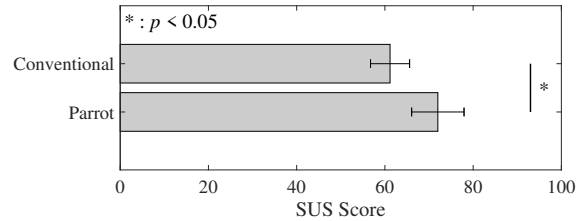


図 4. SUS 値の評価結果

を諦めるという可能性が存在する。そのため、事後アンケートを実施しているが、ビブラートデザインそのものを諦めたという報告は無く、また、実験結果に有意差が認められないことから、デザイン結果の正確性については同程度であるといえる。

満足度については、SUS の評価からも Parrot が相対的に優れているといえる。これは、再生中にビブラートを加工できるという機能が、多くのユーザーに受け入れられたことを意味する。ただし、今回の評価は楽曲制作の作業経験が無い学生を被験者としている。この評価は歌唱デザインに対する習熟度が影響するため、習熟度別に有効性を検証することが今後の重要な検討項目である。

#### 5.2 効率に関する考察

従来手法と比較すると、Parrot は作業時間を短縮しているが作業回数は増加している。短時間で作業を終えるという目標を達成していることから、Parrot はビブラートデザインの省力化を達成している。ただし、作業回数が増えていることについては、いくつかの可能性が考えられる。

1 つ目の可能性は、SUS の評価からも明らかに作業のストレスが低減しているため、繰り返しデザインすることが苦とはならなかったというポジティブな側面である。一方、もう 1 つの可能性として 1 回の作業に対して作り込めるビブラートの正確さが劣るということが示唆される。デザイン結果のビブラートのパラメータに有意な差が無いため、目的とするビブラートを正確に作り込むためには繰り返し作業する必要があった可能性がある。知覚するビブラートが手本に近いかな否かについては、各被験者がデザインした結果の主観評価により検討する必要がある。知覚するビブラートと被験者の満足度の対応が対応するかも本評価では明らかではないため、これは今後の検討項目とする。

表 4. 手本音声とデザイン結果のビブラートとの誤差

評価結果		
パラメータ	提案手法	従来手法
Vibrato duration [%]	27.6	28.2
Vibrato extent [cent]	37.7	36.8
Vibrato rate [Hz]	2.04	2.00

## 6 むすび

本稿では、ビブラートデザインを容易にするために、デザイン結果を聴取しながらデザインが可能なインタフェース Parrot を提案し、有効性を評価した。提案手法の有効性を検証するため、作業時間、作業回数、ユーザビリティに関して評価実験を行った。実験の結果、提案手法の方がデザインに要する時間が少なくなり、インタフェースのユーザビリティも高いという結果を得た。

本研究の次のステップとして、既存のソフトウェアで生成した歌声ではなく、歌詞と譜面に基づいて自然な歌声を生成する統計的歌声合成手法を組み合わせる。初期値を人間に近い歌声とし、それをユーザが望む特徴を持つよう本インタフェースで加工することにより、歌唱デザインの更なる省力化が期待される。加えて、ポルタメントなどの歌声を構成する他の歌唱表現についても加工できるよう機能を追加することを目指す。

## 謝辞

本研究は、JST さきがけ JPMJPR18J8 の支援を受けた。

## 参考文献

- [1] H. Kenmochi and H. Ohshita. VOCALOID - Commercial singing synthesizer based on sample concatenation. In *Proc. INTERSPEECH 2007*, pp. 4010–4011, 2007.
- [2] 歌声合成ツール UTAU サポートページ. <http://utau2008.web.fc2.com/>.
- [3] T. Nakano and M. Goto. VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation. In *Proc. SMC 2009*, pp. 343–348, 2009.
- [4] T. Nakano and M. Goto. A singing synthesis system able to mimic a user's singing in terms of voice timbre changes as well as pitch and dynamics. In *Proc. ICASSP 2011*, pp. 211–216, 2011.
- [5] K. Oura, A. Mase, T. Yamada, S. Muto, Y. Nankaku, and K. Tokuda. Recent development of the HMM-based singing voice synthesis system - Sinsy. In *Proc. Speech Synthesis Workshop*, pp. 211–216, 2010.
- [6] CeVIO Official Web. <http://cevio.jp/>.
- [7] T. Saitou, M. Unoki, and M. Akagi. Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis. *Speech Communication*, 46:405–417, 2005.
- [8] J. Bretos and J. Sundberg. Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos. *TMH-QPSR, KTH*, 43(1):37–44, 2002.
- [9] E. Prame. Measurements of the vibrato rate of ten singers. *STL-QPSR*, 33(4):73–86, 1992.
- [10] 右田尚人, 森勢将雅, 西浦敬信. 歌唱データベースを用いたヴィブラートの個人性の制御に有効な特徴量の検討. *情報処理学会論文誌*, 52(5):1910–1922, 2011.
- [11] 大浦圭一郎. 統計モデルに基づいた歌声合成技術の最先端. *電子情報通信学会誌*, 98(6):405–417, 2005.
- [12] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. WaveNet: A generative model for raw audio. *CoRR, arXiv preprint arXiv:1609.03499*, 2017.
- [13] M. Blaauw and J. Bonada. A Neural Parametric Singing Synthesizer Modeling Timbre and Expression from Natural Songs. *Appl. Sci.*, 7(12), 2017.
- [14] 山田知彦, 武藤聡, 南角吉彦, 酒井慎司, 徳田恵一. HMM に基づく歌声合成のためのビブラートモデル化. *情報処理学会研究報告*, 2009-MUS-80(5):309–312, 2009.
- [15] Auto-Tune & Vocal Processing Tools by Antares Audio Technologies. <https://www.autotune.mu/products/auto-tune-pro/>.
- [16] Celemony | Melodyne とは? <https://www.celemony.com/ja/melodyne/what-is-melodyne>.
- [17] Tune | Media Integration, Inc. <https://www.minet.jp/brand/waves/tune/>.
- [18] GVST - GSnap. <https://www.gvst.co.uk/gsnap.htm>.
- [19] M. Morise, F. Yokomori, and K. Ozawa. WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. *IE-ICE transactions on information and systems*, E99-D(5):1877–1884, 2016.
- [20] H. Dudley. Remaking Speech. *J. Acoust. Soc. Am.*, 11(2):169–177, 1939.
- [21] 森勢将雅. 音声分析合成システム WORLD により実時間音声合成を実現するための拡張と実装例. *情報処理学会音楽情報科学研究会*, 2016-MUS-112(20):1–6, 2016.
- [22] ISO 9241-11:2018 - Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. <https://www.iso.org/standard/63500.html>.
- [23] J. Brooke. SUS - A quick and dirty usability scale. In *P.W. Jordan, B. Thomas, B.A. Weerdmeester and I.L. McClelland (Eds.), Usability Evaluation in Industry*. London: Taylor and Francis, 1996.
- [24] 樽本徹也. ユーザビリティエンジニアリングユーザエクスペリエンスのための調査、設計、評価手法 (第2版). オーム社, 2014.