

# 実時間音声合成技術を活用した 歌唱デザインインタフェースの試作

森勢 将雅<sup>1,2,a)</sup> 渡邊 優介<sup>1</sup>

**概要:** 本稿では、歌唱表現の1つであるビブラートを対象に、実時間音声合成技術を活用して作り込む作業（本稿ではこの作業を「歌唱デザイン」と呼称する）を支援するインタフェースについて述べる。従来の歌唱デザインでビブラートを制御する場合、まずはエディタにより音高を編集し、編集結果に基づいて合成された歌声を聴取することでデザイン結果の妥当性を確認する必要がある。提案するインタフェースでは、楽曲の再生中にユーザが高さを実時間で制御することができる機能を実装することで、デザインと確認を同時に行うことが可能となる。ここでは、iPad上にプロトタイプを実装し、VOCALOID4を模倣したインタフェースを対象に比較実験を実施した。実験の結果、作業にかかる時間が約63%に短縮されたほか、作業に関するストレスを低減することができた。

## 1. はじめに

音楽コンテンツ制作において、VOCALOID [1] や UTAU [2] などの歌声合成ソフトウェアは、すでに広く使われる存在となった。これら音楽創作に使われるソフトウェアは、楽曲の譜面と歌詞のテキスト情報によって歌声を合成することが可能であり、誰でも容易に歌わせることが可能である。最新の技術は、専門的な技術を持たずとも人間の歌唱に匹敵する品質の歌声合成を達成しつつある [3]。しかしながら、人間の歌唱においても歌い方について調整することはあるため、人間の歌い方を達成する技術に加え、歌声を加工する歌唱デザイン技術も必要となる。特に、ユーザが望む歌い方を容易に作り込むためのインタフェースは、クリエイタが個性を演出するためにも重要な役割を担う。

既存のソフトウェアでは、簡単な操作により、ビブラートやグルウルなどの歌唱表現を付与することができる。その中でも、VOCALOID5では、ソフトウェア側が用意したプリセットにより、歌唱デザインが容易にできるよう支援するインタフェースを採用している。一方、ソフトウェアのプリセットによって用意された歌唱表現だけではなく、ユーザがイメージする歌声に近づけるためには、職人芸的な技術が必要とされる。現在の歌声合成・歌唱デザインソフトウェアで歌唱表現を作り込むためには、何度も試行錯

誤的に歌唱表現をデザインして再生し、聴覚的に確認する煩雑な作業が必要である。

歌唱デザインに関する作業が煩雑になる要因として、作業中にどのような歌唱が合成されるか確認できないことが挙げられる。デザイン中でも即座に結果をフィードバックする機能を実現することは、ユーザがイメージする歌唱を作り込むデザイン作業の効率改善に繋がる。本研究では、ビブラートの作り込みを対象とし、音声の高さに相当する基本周波数 (F0) を実時間で加工できるインタフェースを提案する。本インタフェースを用いることで、歌唱デザイン結果を聴きながら実施することが可能になり、デザインと確認作業を統合することが可能となる。したがって、短時間で多くの試行が可能となり、ユーザの求める歌唱表現のイメージを短時間で作り込めることが期待できる。

## 2. 関連研究と本研究の位置づけ

ここでは、歌唱デザインに関する関連研究を示した後、それらと比較した本研究の位置付けについて述べる。

### 2.1 VOCALOID

VOCALOID は、合成音声による歌声合成ソフトウェアとして世界的に利用されている。本ソフトウェアは、譜面と歌詞の情報を入力するだけで歌わせることができる。歌唱表現の付加については、一般的なビブラートやポルタメントから、グルウルなどまで様々な表現を扱うことができる。VOCALOID をより表情豊かに歌わせる取り組みとして、ユーザの「歌い方」における F0 と音量・音色を真似

<sup>1</sup> 山梨大学  
University of Yamanashi, Yamanashi 400-8511, Japan  
<sup>2</sup> JST さきがけ  
<sup>a)</sup> mmorise@yamanashi.ac.jp

るよう歌声合成するシステム VocaListener が提案されている [4,5]. VocaListener は, VOCALOID で制作した歌唱に対してユーザの歌い方をお手本として真似させる機能により, 歌唱デザインの自動化を実現した技術である.

2018 年に発売された VOCALOID5 では, 音楽ジャンルにフィットさせるためのスタイルプリセットを用意し, 歌い方を自動で調整する機能が追加された. ビブラートなどの歌唱表現についても, ソフトウェア側が既存の歌手から抽出したプリセットを選択するだけで良く, 初心者でも容易に歌声を作り込むことが可能となりつつある. ただし, VOCALOID に歌詞と譜面を入力することで合成される歌声は, 無調整の場合人間の歌声に近いとは言いがたい. ユーザが作り込むことにより自然性を向上させる作業が必要となるが, 一方でこれは, ユーザの個性を持たせることにも繋がる.

## 2.2 統計的歌声合成

統計的歌声合成技術は, 細かな歌唱デザインを不要にするため, 歌詞と譜面から人間とほぼ等価な歌声を出力することを目指して日々進歩している. 製品例としては CeVIO [6,7] があり, VoiceText [8] でも人間の歌声に匹敵する品質の歌声合成を実現している. CeVIO では, VOCALOID と同様に譜面と歌詞の情報を入力するだけで自然な歌声で歌わせることが可能である. また, 歌声の感情表現や声質を設定するパラメータを有する.

## 2.3 人間の歌声を加工するソフトウェア

人間の歌声を加工することを目指したソフトウェアでは, Auto-Tune [9] や Melodyne [10] が利用されている. これらのソフトウェアは, 入力された歌声がすでに人間のものであるため, 歌手の個性を残したまま歌唱表現の工夫をすることや, 意図的に電子楽器的にするなど, 演奏表現に幅を持たせることが可能である. 特に, 人間の歌唱が有する揺らぎを持たない電子楽器的な表現も, 現在では歌唱表現の 1 つとして認められ, コンテンツに利用されている.

## 2.4 本研究の位置付け

歌声合成ソフトウェアが要求する機能は, 歌詞と譜面から歌声を合成する機能と, 入力された歌声を加工する歌唱デザイン機能に大別されるといえる. VOCALOID や CeVIO は両者の機能を併せ持ち, Auto-Tune や Melodyne は後者に特化したソフトウェアと位置付けられる. 本研究では, 歌唱デザインの支援を目的としており, 歌唱デザインを効率改善を目指したインターフェース開発に取り組む. ただし, 基本的には歌唱表現が無い音源を入力とし, 歌唱デザインにより表現を付与することを目指している.

歌唱デザインで扱うことが可能な音声パラメータは複数存在するが, 本稿では比較的应用範囲の広い F0 の制御に

特化したインターフェースを開発する. F0 は声の高さに相当するため, ビブラートやポルタメント等多くの歌唱表現の実現に利用できるパラメータである. その中でも, 音高を準正弦波的に振動させるビブラートは代表的な歌唱表現であるため, デザイン対象をビブラートに絞って検討する. 歌唱デザインと結果の聴取を同時に行えるインターフェースを実現することで, イメージした歌唱表現を容易にデザインすることが可能であるか検証する.

## 3. 提案する歌唱デザインインターフェース

ここでは, 本研究で実装したビブラートを実時間で制御する歌唱デザインインターフェースについて述べる. インタフェースに対する要求について示した後, 具体的な実装について説明する.

### 3.1 インタフェースに対する要求

歌唱表現のデザインにおける煩雑さを緩和するインターフェースを実現するためには, 直感的な操作性と歌唱デザイン内容が即座にフィードバックされる機構が不可欠である. 例えば, Melodyne の F0 制御システムでは, 録音された歌声に対し, F0 を変化させた際の音が即座にフィードバックされることで, 操作が結果に与える影響を予想できる. しかし, これは F0 の音高を補正することが目的であり, 本研究で目的とするビブラートの制御にそのまま応用することはできない. 実時間歌唱デザインを目的としたインターフェースの実現では, 以下に示す機能が要求される.

- ユーザの操作と合成結果が直結させるための実時間音声合成技術
- 容易にビブラートを加工することが可能となる歌唱デザインインターフェース
- デザイン結果に基づき合成結果を再現する機能

これらの機能を実装することで, ユーザは操作に直結したフィードバックにより歌唱表現をデザインすることができ. 操作と確認と同時に行えることは, 作業時間の短縮に繋がる. ユーザは, 同等の作業時間でより多くのパターンを試すことができるため, 歌唱デザインに必要な総作業時間の短縮が期待できる. また, デザイン結果を記録・再現する機能は, 歌声合成ソフトウェアの実現のためには必要不可欠である.

### 3.2 実装環境とライブラリ

本研究では, 実時間音声合成技術に, Vocoder [11] に基づく音声分析合成方式 WORLD [12] (D4C edition [13]) を用いる. WORLD は, 音声から F0 だけではなく, 音色に相当するスペクトル包絡, かすれの程度に相当する非周期性指標を取り出す技術と, 3つのパラメータから音声波形を生成する技術から構成される. GitHub 上で C++のソース

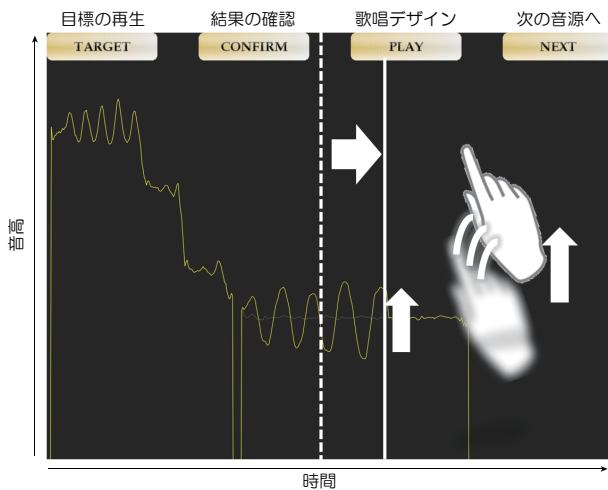


図 1: 本インターフェースの実行画面。縦軸が F0, 横軸が時間に対応する。本インターフェース上で上下にタップしたまま動かすことで、再生中の F0 を実時間で制御することができる。

コードが公開されており\*1, 3パラメータから波形を実時間合成する仕組みも提供されている [14]。本インターフェースの開発 OS は, macOS Sierra Ver. 10.12.3, 統合開発環境は Xcode Ver. 8.3.2 であり, 開発言語は Swift 3.1 である。本インターフェースの要となる F0 を制御しつつ合成・再生を同時に実施するために, iOS で動作するソフトウェアフレームワークである AVFoundation を用いた。

### 3.3 インターフェースの実装

本節では, インターフェースの実装について説明する。実装したインターフェースの実行画面を図 1 に示す。図 1 で示している操作画面の横軸は時間, 縦軸は音高に対応する F0 である。再生中, タップ位置を上にはずらせば, 対象となる時間の F0 は高くなり, 下にはずらせば低くなる。タップ位置を起点に上下に振動させれば, ビブラートの歌唱を表現することが可能である。

図 1 に上部には, 4つのボタンを設置している。TARGET ボタンと NEXT ボタンは, 後述する本インターフェースの評価実験のために実装したボタンである。TARGET ボタンは, 実験において被験者にデザインさせる目標となるビブラートを再生するボタンである。TARGET ボタンを押すと, 伴奏となる音楽とともに, デザインさせるビブラートの含まれた歌声が再生される。

PLAY ボタンは, インターフェースの目的であるビブラートのデザイン用ボタンである。PLAY ボタンを押すと, 伴奏となる音楽とともに加工前の歌声が再生される。再生と同時に, 図 1 に示す白線がタイムスライダーとして表示される。PLAY ボタンを押した後から再生が終わるまでの間に画面上をタップしたまま上下に動かすと, 白線の位置に

\*1 <https://github.com/mmorise/World>

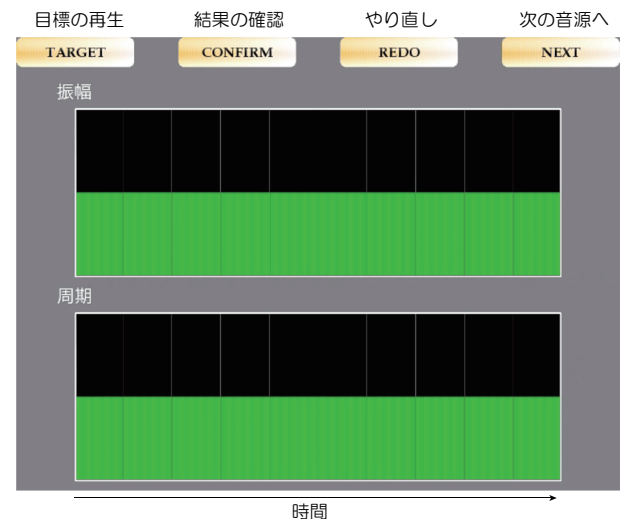


図 2: VOCALOID4 Editor を模擬した評価用インターフェース。

対応した歌声の F0 を変更することができる。その時の音の高さは, タップした位置を起点とした相対位置に対応して変動する。PLAY 中は, F0 操作と同時に操作後の音の高さを聴くことができるため, 合成結果を聴きながらビブラートをデザインできる。

CONFIRM ボタンは, 歌唱デザインにより得られたビブラート軌跡に基づいた歌声を再現する機能である。本機能により, ユーザは作られたビブラートの自然さについて再度確認することができる。NEXT ボタンを押すことで歌唱デザインは終了し, 次にデザインする目標となるビブラートに移行する。本ボタンは, ビブラートデザインに要した時間と PLAY ボタンを押した回数を記録する機能も併せ持つ。

## 4. インターフェース評価実験

提案法の有効性を評価するため, インターフェースの評価実験を行う。ここでは, まず比較対象となる従来法のインターフェースについて説明する。その後, 提案法の補助インターフェースについて説明し, 最後に本評価実験の実験方法と条件の詳細について述べる。

### 4.1 比較対象となるインターフェース

本実験では, 2.1 節で紹介した VOCALOID4 Editor を対象に比較実験を行う。ただし, 制御内容を記録・検証するため, 図 2 に示すように, 同様の機能を有するインターフェースを iPad 上に実装して用いることとした。本インターフェースには, ビブラートの振幅と周期に対応する 2つの時系列グラフが表示されている。2つのグラフを用いて, ビブラート振幅とビブラート周期を加工することでビブラートをデザインする。本インターフェースの使用感については, 事前に筆者らが VOCALOID4 editor と比較して同程度であることを確認済みである。

表 1: 実験条件

被験者数	13 名 (男性 11 名, 女性 2 名)
評価環境	防音室 (A-weighted SPL 18 dB)
使用したイヤホン	ETYMOTIC RESEARCH ER4SR
加工音源数	8 音源
測定項目	1 音源ごとの作業時間と加工回数

上部にある 4 つのボタンのうち, TARGET, CONFIRM, NEXT は前節で説明したボタンと同様の機能を有する. REDO は, 編集したグラフを初期状態に戻す, やり直しに相当する機能である.

#### 4.2 実時間基本周波数変換インタフェース

本実験では, 提案法や従来法によるインタフェースとは別に, 提案法を操作するうえで歌唱デザインの補助を行うインタフェースも用いた. ここでは, この歌唱デザイン補助インタフェースについて説明する. 提案法は, スワイプ操作による音声の直感的な実時間変換を実現したインタフェースである. このような補助インタフェースを併用することで, 最初から歌唱デザインインタフェースを使うよりも簡単にデザイン技術を習得できることが期待される.

直感的な実時間合成機能については, 提案法と同様である. 本インタフェースではさらに, 自由な操作性を提供するために, 横軸時間方向へのスワイプも可能となっている. 楽曲への取り込みはできないが, 本機能による自由な操作は歌唱表現デザインへのイメージの手助けとなることが期待できる.

#### 4.3 実験方法と条件

本評価実験では, 事前に用意したビブラートを目標音源として被験者に呈示し, 被験者が満足するまで目標のビブレートとなるようデザインさせるタスクを課す. 本評価実験では, ビブラートをデザインするための楽曲として, 知名度の高い国歌「君が代」と童謡の「ふるさと」を用いる. 表 1 に実験条件を示す.

2 曲について, 同一のフレーズを対象に大きく特性が異なる 2 種類のビブラートをデザインさせることとした. ビブレートには, 振動の速さや深さだけではなく, ビブレートがフレーズの途中から始まるようにするなど, いくつかの条件が存在する. 本実験では, 速さや深さだけではなく, フレーズの途中からビブレートが始まる条件を含めることで, ビブレート開始時刻をデザインする幅を持たせている. 被験者数は, VOCALOID 等による音楽編集経験のない男性 11 名, 女性 2 名の計 13 名とした. 加工音源数は, 前項で示した 4 種類のビブラートを 2 つのインタフェースでデザインさせるため, 計 8 音源である.

次に, 実験の流れについて説明する. 被験者は, 2 台の iPad を用いて評価実験を実施する. 片方の iPad には提案

表 2: アンケート内容

番号	質問項目	回答
1	補助インタフェースの有用性	1:有用ではない 5:有用である
2-1	提案法の直感的操作性	1:とても悪い 5:とても良い
2-2	提案法のストレス	1:感じない 5:感じる
3-1	従来法の直感的操作性	1:とても悪い 5:とても良い
3-2	従来法のストレス	1:感じない 5:感じる
4	実時間音声合成の有用性	1:有用ではない 5:有用である

法とその練習用アプリ, 補助インタフェースがインストールされており, もう一方の iPad には従来法とその練習用アプリがインストールされている. 本実験では, 本番前に各インタフェースの練習フェーズを設け, 本番前に歌唱デザインについて慣れさせることにした.

本実験では, デザインさせるビブラートを聴取してから満足するまでの時間, および加工回数を記録する. ビブレートデザインそのものに対する慣れの影響を勘案し, 実験はデザインするビブレートだけではなく使用するインタフェースまで含めて無作為に決定した. この提示は, ユーザの正面には指示用の PC が設置されており, 2 台ある iPad のどちらを手取るか指示することで実現している.

全てのビブラートをデザインし終えた後には, アンケートによるインタフェースの主観評価を行う. アンケートの内容は, 表 2 の設問に対して 5 段階で行った. また, 本実験において, アンケートの最後に自由記述の欄を設けた.

#### 4.4 作業時間と加工回数

実験結果のうち, まずは各ビブレートのデザインに要した作業時間について, 提案法と従来法の比較結果を図 3 に示す. それぞれの値は, 全被験者における全ビブレートの平均作業時間を示している. 「\*」は, 提案法と従来法の各作業時間に対してウィルコクソン符号付き順位検定を有意水準 1%で行った結果, 有意な差が認められたことを表している. 誤差棒は, 95%信頼区間を示している. 結果より, 全被験者における全ビブレートの平均作業時間は提案法の方が従来法に比べ有意に減少し, 約 63%の時間で作業を完了していることが分かる.

次いで, 各ビブレートの加工に要した加工回数について提案法と従来法の比較結果を図 4 に示す. それぞれの値は, 全被験者における全ビブレートの平均加工回数を示している. 本結果は, 全被験者における全ビブレートの平均加工回数は, 従来法・提案法ともに同程度であることを示している. また, 両方の結果を併せると, 1 回あたりの試行時間については, 提案法のほうが短いといえる.

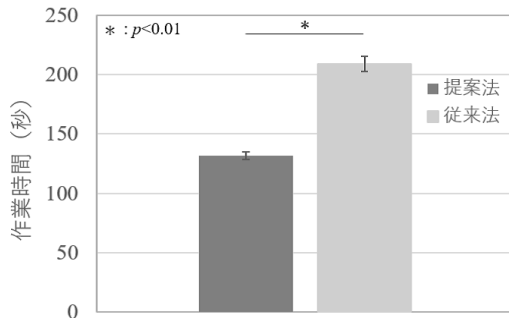


図3: 提案法と従来法における作業時間の比較. 誤差棒は95%信頼区間を示す。「\*」は、有意水準1%で有意な差が認められたことを表している。

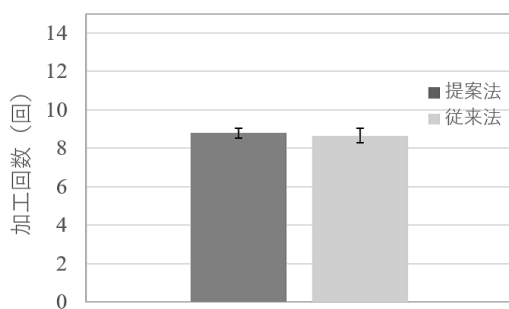


図4: 提案法と従来法における加工回数の比較.

#### 4.5 アンケート結果

図5に、本実験の最後に行ったアンケートの結果を示す。縦軸は、全被験者の各質問に対する平均スコアを、図中の誤差棒は95%信頼区間を示す。それぞれのスコアは、表2に示す質問番号に対する値を示している。

補助インタフェースが提案法の音をイメージするうえで有用であるかについて、5段階評価で平均4.5の評価が得られた。提案法と従来法における直感的操作性を問う質問2-1、3-1については、提案法は4.0、従来法3.2であった。提案法と従来法における操作のストレスを問う質問2-2、3-2については、提案法は2.5、従来法は3.1であった。最後に、提案法の特徴である実時間歌唱デザイン機能の良さを問う質問4のスコアは4.8であり、大半の被験者にとって容易に歌唱デザインできるインタフェースであること示した。

### 5. 考察

前節では、提案法の有効性を検討する評価実験の結果について述べた。ここでは、作業時間と試行回数の比較について考察し、その後、提案法と従来法のストレスの違いから提案法の有効性について述べる。

#### 5.1 作業時間と加工回数の比較

本実験で用いた4種のビブラートは、それぞれ異なるF0

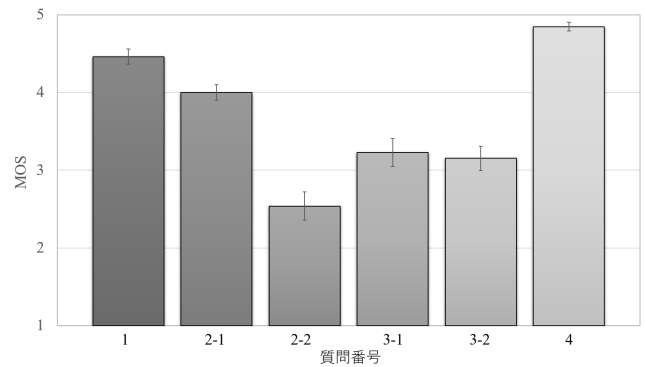


図5: アンケート結果. 縦軸はMOS, 横軸は表2に示す質問番号に対応している。

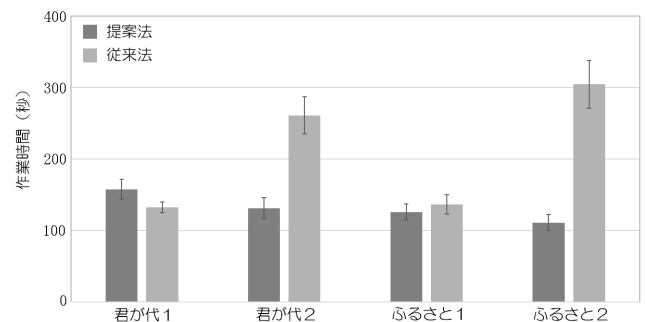


図6: 提案法と従来法におけるビブラートごとの作業時間の比較. 縦軸はそれぞれのビブラートデザインに対する作業時間を表す。

の変化となるよう設定した。ビブラートのF0変化の特徴ごとに、インタフェース間における作業時間に違いがあることが考えられる。そこで、それぞれの条件におけるビブラートデザインに要した作業時間を図6に示す。

図6より、従来法では君が代2とふるさと2について大きく作業時間が増加していることがわかる。これらのビブラートは、編集範囲全体に対して途中からビブラートがかかる特徴がある。提案法では、操作内容が結果に直結するため編集位置が明らかであることに対して、従来法は編集範囲がわかりにくく、顕著に差が出たのではないかと考えられる。また、アンケートの最後に設けた自由記述欄に記述された内容から、従来法はグラフから合成されるビブラートのイメージがしづらいことによってデザインが難しく感じる被験者がいることが確認できた。それに対して提案法では、実時間歌唱デザインによってビブラートのタイミングが明確になり加工しやすくなったことで、これらのビブラートについて作業時間が短くなったのだと考えられる。

#### 5.2 アンケート結果による分析

インタフェースごとのストレスの度合いの調査は、実験の最後に行ったアンケートで実施した。提案法と従来法のストレスの度合いの差の平均を求めた結果を図7aに示

す。図中の右に示している数値は、提案法と従来法の差分の平均値と、括弧内は平均の95%信頼区間を示している。値が負であることから、提案法のほうが有意にストレスが少ないことを示している。同様の手順で求めた、提案法と従来法における直感的操作性の差の平均を求めた結果を図7bに示す。本結果は、提案法の方が従来法に比べ直感的操作性が有意に高いことを示す。

両方の比較結果を併せると、提案するインタフェースが既存のものよりも直感的、かつ低ストレスで歌唱デザインができることを示唆する。アンケート最後の自由記述欄より、ビブラート振幅やビブラート周期についてどの程度加工しているかイメージがつきやすいことが示されていた。特に、歌唱デザイン中即座に結果が得られていることが、途中からビブラートをかけるようなデザインを想定する場合に有効であるといえる。以上の結果から、即座に歌唱デザイン結果が得られるインタフェースは、1回あたりの試行時間の短縮やビブラート開始時刻との対応が容易であることなどが、ストレスの軽減に繋がったと考えられる。

### 5.3 デザイン結果のビブラート類似性

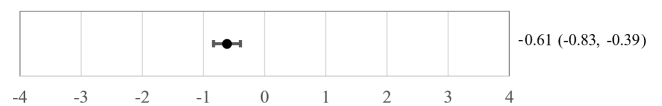
今回の実験では、被験者が納得できるまでビブラートを作り込むことをタスクとしている。アンケート結果には「目標に近づけることができなかった」というコメントが無かったため、全被験者が最低限納得できるビブラートを作り込んでいるといえる。この際、被験者が納得できるビブラートであることは間違いないが、聴取者によって妥当な結果であるかについては検証する必要がある。しかしながら、準正弦波的な振動の類似性について、F0軌跡そのものの誤差を評価すると位相のずれの影響が大きく、主観的な印象とは大きく異なる結果になる可能性がある。

ビブラートの客観評価を行う場合、ビブラートの速さや深さ、および開始時刻を計測し、真値とのズレを評価する方法が考えられる。客観指標と知覚的類似性について必ずしも一致するとは限らないため、客観評価に加えて、知覚的類似性を主観評価により検証する必要がある。これらについては、今後の重要な検討項目となる。

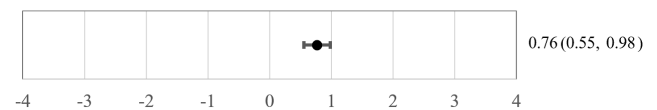
## 6. おわりに

本研究では、ビブラート加工に特化した実時間音声合成による歌唱デザインインタフェースを提案した。本インタフェースの有効性を調査する評価実験を行った結果、提案法は従来法に比べ作業時間が有意に減少し、ストレスも緩和できることが示された。

今後の展望として、ビブラート以外にも存在する様々な歌唱表現を同様の実時間制御でデザインするためのインタフェースの実現が挙げられる。また、高さだけではなく音色を変化させる表現も存在することから [15]、音色を含む実時間制御についても今後の重要な検討課題である。



(a) 提案法と従来法のストレスの度合いの差の平均を求めた結果。



(b) 提案法と従来法の直感的操作性の差の平均を求めた結果。

図7: アンケート結果の差の平均による分析結果。図中の右に示している数値は提案法と従来法の差分の平均であり、括弧内は95%信頼区間を示している。

謝辞 本研究は、科研費 JP16H05899, JP16H01734, 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. 4009–4010, 2007.
- [2] UTAU <<http://utau2008.web.fc2.com/>> (2019–1–8)
- [3] M. Blaauw and J. Bonada, "A neural parametric singing synthesizer," in Proc. INTERSPEECH 2017, pp. 4001–4005, 2017.
- [4] 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.
- [5] T. Nakano and M. Goto, "VocaListener2: 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. 453–456, 2011.
- [6] 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, vol. 98, no. 6, pp. 460–466, 2015.
- [7] CeVIO <<http://cevio.jp/>> (2019–1–8)
- [8] VoiceText <<http://voicetext.jp/voiceactor/>> (2019–1–8)
- [9] Auto-Tune <<http://www.autotune.mu/products/auto-tune-8/>> (2019–1–30)
- [10] Melodyne <<http://www.celemony.com/>> (2019–1–8)
- [11] H. Dudley, "Remaking speech," J. Acoust. Soc. Am., vol. 11, no. 2, pp. 169–177, 1939.
- [12] M. Morise, F. Yokomori, and K. Ozawa, "WORLD: a vocoder-based high-quality speech synthesis system for real-time applications," IEICE transactions on information and systems, vol. E99-D, no. 7, pp. 1877–1884, 2016.
- [13] M. Morise, "D4C, a band-aperiodicity estimator for high-quality speech synthesis," Speech Communication, vol. 84, pp. 57–65, 2016.
- [14] 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, vol. 2016-MUS-112, no. 20, pp. 1–6, 2016.
- [15] 森勢将雅, 村主大輔, 馬場隆, 片寄晴弘, "奄美大島民謡風の歌唱デザインを支援するシステム: グインレゾネータ," 情報処理学会論文誌, vol. 54, no. 4, pp. 1244–1253, 2013.