固有声変換法に基づく声質変換ソフトウェアの開発
―あなたの声がアニメのキャラクターの声に―

1. 背景

声質変換は言語情報を保ちながら話者性などの非言語情報を変換する技術である。その応用例は幅広く、例えば、ある人の声を特定のキャラクターの声へと変換することが可能となるため、日本人声優の声質による外国語の音声吹き替えも夢ではなくなる。従来の声質変換技術では、入力話者と出力話者による同一内容発声データ対（50文程度：時間にして約3～5分）を用いて事前に学習された変換モデルにより、入力話者の音声から出力話者の音声への変換が実現される。日本語で学習された変換モデルを、他言語における変換に適用することも可能である。非常に有用な技術である一方で、入力・出力話者による同一内容発声からなる学習データが必要という大きな制約がある。そのようなデータの収録が不可能であれば、この技術は使用できない。

2. 目的

我々が提案している固有声に基づく声質変換法では、予め収録された多数話者の音声データを事前情報として活用することで、任意の話者に対する変換モデルを容易に構築することができる。この固有声変換技術に基づき、1）任意のユーザーの音声から特定話者の音声への変換（多対一変換）と、2）特定のユーザーの音声に対する声質手動制御（一対多変換）を実現するソフトウェアを開発する。

3. 開発の内容

任意のユーザーの音声をアニメのキャラクターのような声質へと変換する多対一固有声変換ソフトウェア、及び特定ユーザーの音声に対する声質手動制御を実現する一対多固有声変換ソフトウェアを開発した。また、通常発声及びアニメのキャラクターのような発声からなる音声データベースを構築した。

【音声データベースの構築】

本ソフトウェアで使用している事前収録話者の音声データベースである。収録用読み上げテキストとして、50文からなる音素バランス文を用いている。37 名の話者（男性14名、女性23名）が各々通常の発話スタイル及びアニメのキャラクターのような発話スタイルで発声した音声（16 kHz サンプリング、16 bit 量子化）が収録されている。

【多対一固有声変換ソフトウェアの開発】

任意のユーザーの音声をアニメのキャラクターのような声質へと変換する多対一固有声変換ソフトウェアである。本ソフトウェアはGUIで実装されており、Windows 上で機能する。概観を**図1**に示す。各箇所の機能は以下の通りである。

① 各種メッセージを表示する。
② データベースフォルダの場所を指定する。
③ 選択されているキャラクター声を表示する。
④ 変換先となるキャラクター声を選択する。
⑤ 音声分析パラメータを設定する。

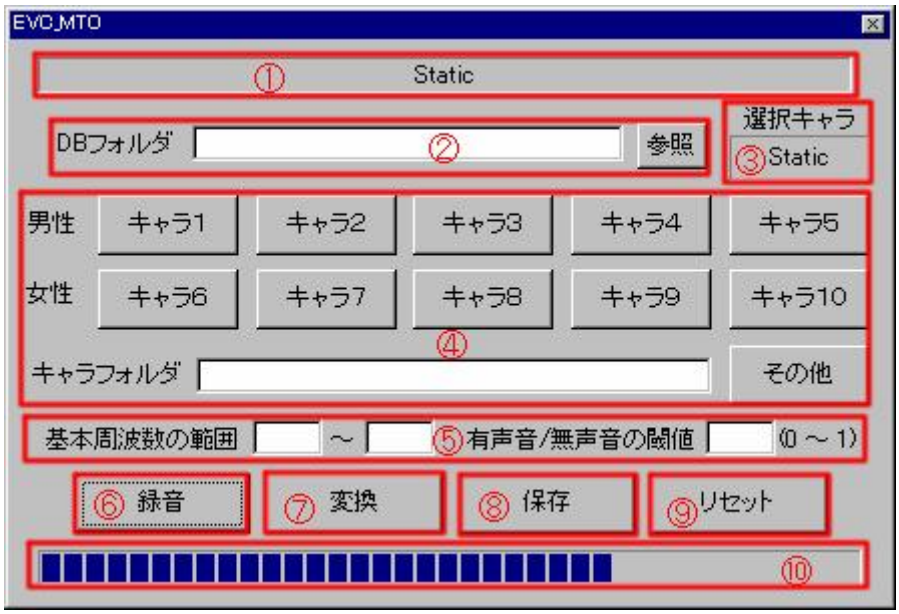

図１．多対一固有声変換アプリケーションの概観

⑥ ユーザー音声を録音する。
⑦ 録音した音声を指定のキャラの声質に変換する。
⑧ 変換音声を保存する。
⑨ リセットしてキャラクター声の選択からやり直しをする。

【一対多固有声変換ソフトウェアの開発】

特定ユーザーの音声を様々な声質へと変換する一対多固有声変換ソフトウェアである。本ソフトウェアはGUIで実装されており、Windows 上で機能する。声質表現語スコアにより変換音声の声質を制御できる。特定ユーザーから提供される音声データと事前収録話者音声データに基づき、ユーザー専用変換モデルを学習する機能も持つ。概観を**図2**に示す。各箇所の機能は以下の通りである。

① 初期データベースフォルダの場所を指定する。
② 特定ユーザー専用の固有声変換モデルの学習処理を選択する。
③ 入力音声の声質制御処理を選択する。
④ 学習データセットを作成する。
⑤ 特定ユーザー専用の固有声変換モデルの学習を行う。
⑥ 変換したい音声を入力する。参照ボタンにより収録済みの音声ファイルを入力できる。サンプル用として、男性１名及び女性１名の音声ファイルが用意されている。また、録音ボタンにより入力音声を収録できる。
⑦ 使用する固有声変換モデルのフォルダを選択する。サンプル用として、男性１名及び女性１名に対する2つの固有声変換モデルが用意されている。
⑧ 変換する際のパラメータを変更する。各スライダーを操作することで変換音声の声質を制御することができる。
⑨ 各スライダーの値を初期値に戻す。
⑩ 音声分析パラメータを設定し、選択された音声を分析する。
⑪ 音声を変換する。

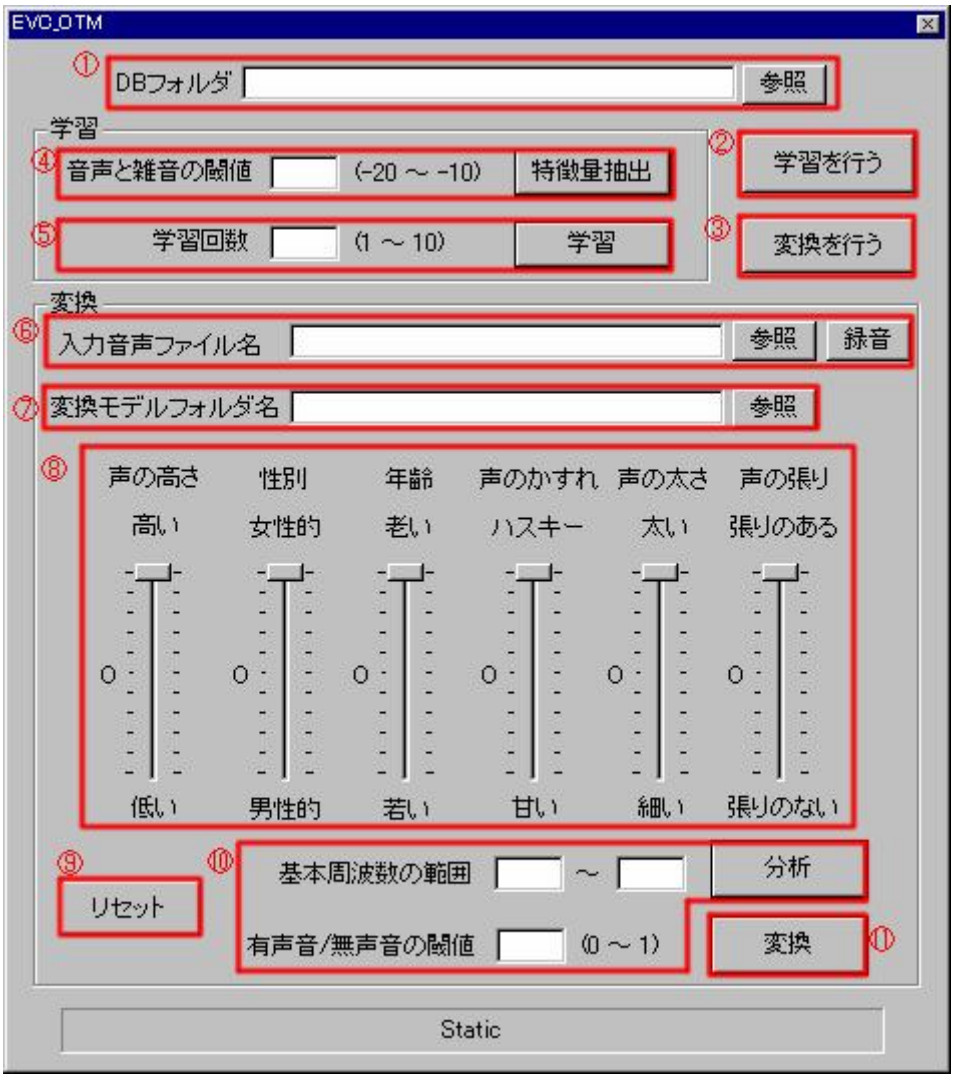

図２．一対多固有声変換アプリケーションの概観

４．従来の技術（または機能）との相違

特定話者の音声を合成する一手法として、テキストから音声を合成するテキスト音声合成技術が挙げられる。テキスト音声合成システム構築のためには、一般的に数時間規模の音声収録が必要となる。一方で、本ソフトウェア（一対多変換）では、変換モデル構築用に必要となる音声データ量は50文のみ（時間にして3～5分程度）である。

本ソフトウェアは、従来の統計的声質変換法と比べて様々な利点を持っている。一つは、事前収録話者の情報を活用することで、任意のユーザーの音声を特定キャラクターの声質へと変換できる点である（多対一変換）。従来の変換法では、入力話者は必ず学習データを発声する必要があるが、本ソフトウェアではその必要がない。また、変換時に言語情報を必要としないため、外国人話者が発声した異なる言語の音声さえも、指定したキャラクターの声質へと変換できる。従来の変換法では同一内容発声による学習が不可欠であるため、このような処理はバイリンガル話者でない限り不可能である。

これまでにも、学習処理を行わずに音声を変換する音声モーフィング技術が実現されているが、一般的に単純な変換処理しか行えず、ある特定話者の声質へと変換する事は不

可能に近い。それに対して、本ソフトウェアは学習処理なしで特定話者への変換（多対一変換）が可能である。さらに、学習データを提供したユーザーに対しては、声質表現語を用いた変換音声の声質手動制御（一対多変換）が可能であり、このような処理を実現した音声変換装置はこれまでに存在しない。

5．期待される効果

近年、様々な国との文化交流が盛んに行われており、映画やアニメなどの海外輸出及び輸入が多く見られる。こういった文化交流の中で、言語の違いは大きな壁として立ちはだかっており、音声吹き替え技術は極めて重要な役割を担っている。しかしながら、翻訳時には当然話者が変わるため、言語情報を伝えることはできても、話者性などの非言語情報を伝えることはできない。これに対して、本ソフトウェアを用いることで、言語の壁を超えた話者性制御が可能となるため、非言語情報も伝達できる音声吹き替え技術の実現が大いに期待される．それ以外でも、自分の声質を自在に制御して所望の声質へと変換できる防犯用ボイスチェンジャーを実現したり、声帯を失った発声障害者による人工的な声質を元の自分のような声質もしくは好みの声質へと変換することで、身体的な音声発声能力を回復させたりという応用例も考えられる。本ソフトウェアにより、従来の枠組みでは適用不可能であった様々な応用例において声質変換が利用可能となるため、言語や身体障害の壁を越えた、よりユニバーサルな音声コミュニケーションの実現が期待される。

本プロジェクトで構築された音声データベースは、同一話者による通常発声とアニメのキャラクター声のような発声を含んでいる。本データベースは研究用途においてフリーで使用可能であり、音声合成のみでなく音声分析や音声知覚、音声認識の分野においても、非常に興味深い研究開発リソースになると期待される。

6．普及（または活用）の見通し

本ソフトウェアにより、声質変換に関する研究開発が大いに促進される事が期待される。本ソフトウェアの一部である基盤声質変換プログラムは、世界標準のテキスト音声合成フリーソフト Festival 及び FestVox に提供されており、既に組み込まれている。固有声変換ソフトウェアに関しても、同様に提供していく方向を検討する。

声質変換技術を世の中に広めるための第一歩として、奈良県生駒市の施設に既に設置されている公共案内音声対話システムに対して、本ソフトウェアを組み込む予定である。自分の声を有名人の声へとゲーム感覚で変換する新たなアミューズメントの開拓も視野に入れ、さらなる開発を進めていく予定である。

7．開発者名（所属）

戸田智基（奈良先端科学技術大学院大学）
大谷大和（奈良先端科学技術大学院大学）
関本英彦（奈良先端科学技術大学院大学）
中村圭吾（奈良先端科学技術大学院大学）

（参考）開発者URL

http://www.hcilab.jp/MITOU/
http://spalab.naist.jp/~tomoki/index_j.html