

知覚年齢をリアルタイムに制御可能な歌声声質制御インタフェース

小林和弘 戸田智基 中野倫靖 後藤真孝 Graham Neubig Sakriani Sakti 中村哲*

概要. 本稿では、人が歌声を聞いた際に知覚される年齢（知覚年齢）に着目し、歌声を若くしたり年齢を重ねた歌声に聞こえるようにリアルタイムに操作出来る歌声声質制御インタフェースを提案する。歌手の表現可能な声質は身体的制約により制限されるため、一人の歌手が生み出すことができる歌唱表現は限定される。この制約を取り除くために、我々は、歌手の個性を保持したまま、知覚年齢に沿って声質を制御する技術を提案している。知覚年齢に基づいた声質制御により、歌手は自身の歌声を身体的な制約を超えて、所望の知覚年齢を持つ歌声へと変換する事が可能となる。本稿で提案するインタフェースでは、知覚年齢制御に対しリアルタイム声質制御を統合する事で、知覚年齢スライダーを操作により、ユーザの所望する最適な知覚年齢を直感的に探索する事が可能となる。

1 はじめに

歌声は、メロディーやリズムに加え言語情報である歌詞を伝えることができ、他の楽器には難しいような多彩な音楽表現が可能である。しかし、個々の歌手が表現できる声質の範囲は身体的制約により制限される。この身体的制約を超え、歌手が意のままに声質を操れる技術が実現すれば、更に豊かな音楽表現を生み出されると期待される。

歌手の声質を制御する技術として、入力歌手の声質を目標歌手の声質へと変換する統計的手法に基づく歌声声質変換 [3] が提案されている。我々は、この枠組みを拡張し、主観的声質表現語の一つである知覚年齢に基づく声質制御法を提案した [1]。本手法は、歌手の個性を保持しながら歌声の知覚年齢を制御可能であるため、歌手の過去や未来の声質を再現する事が可能である。しかし、フレーズ毎に一律の知覚年齢スコアに基づき声質制御が行われるため、動的な知覚年齢制御は出来ず、結果として、直感的に理想の声質を探求することが難しかった。

本稿では、歌声の知覚年齢制御法とリアルタイム声質変換の枠組み [2] を統合することで、知覚年齢に基づくリアルタイム歌声声質制御インタフェースを提案する。本インタフェースにより、歌手は知覚年齢制御スライダーを動的に操作しながら、歌声を所望の知覚年齢を持つ声質へと変換することが出来る。

リアルタイムにユーザが入力した何かを音へと変換するインタラクションとしては、口パクから音声 [6]、楽器の音色 [4]、歌声の音高 [5] などが提案されているが、音の音色や声質を表現語などの直感的

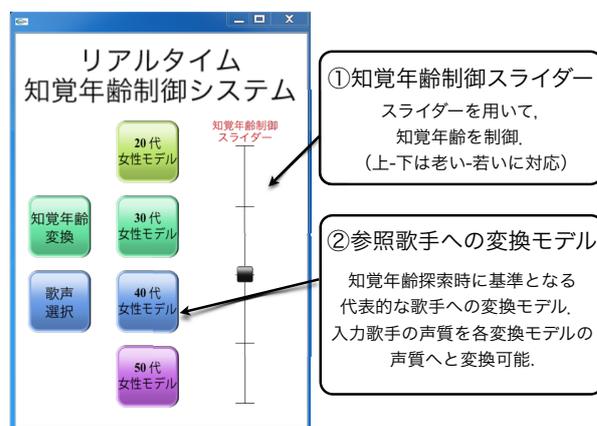


図 1. インタフェースの実行画面と代表的な機能

な指標に基づいて制御する事は出来ない。本インタフェースでは、声質表現語である知覚年齢に基づき、直感的かつ容易な歌声声質制御が可能である。

2 歌声のリアルタイム知覚年齢制御インタフェース

個性を保持した歌声の知覚年齢制御法 [1] に対し、リアルタイム声質変換 [2] の枠組みを適用することで、リアルタイムに知覚年齢を制御出来るインタフェースを提案する。

2.1 インタフェースの概要

図 1 に歌声のリアルタイム知覚年齢制御インタフェースの実行画面を示す。本インタフェースは、知覚年齢制御スライダーを“上-下”に操作する事で歌手の知覚年齢を“若い-若い”という風に制御する事が出来る。予め収録した歌声を選択し、知覚年齢変換を実行する事で、声質の変換が開始される。本インタフェースでは、変換歌声を確認しながら、知覚年齢スライダーを操作することで、よりインタラ

Copyright is held by the author(s).

* Kazuhiro Kobayashi, Tomoki Toda, 奈良先端科学技術大学院大学 情報科学研究科, Tomoyasu Nakano, Masataka Goto, 産業技術総合研究所 (AIST), Graham Neubig, Sakriani Sakti, Satoshi Nakamura, 奈良先端科学技術大学院大学 情報科学研究科

クティブな知覚年齢制御が可能となる。また、参照とする知覚年齢を持つ声質を確認するために、各年代の代表的な声質への変換モデルを保持しており、各変換モデルを用いてリアルタイムな歌声声質変換 [3] も可能である。

2.2 歌声エフェクタとしての知覚年齢制御

歌声に用いられる代表的なエフェクタとしては、Digital Audio Workstation に搭載されているコンプレッサやリバーブなどが挙げられる。一方で、歌手単体では実現不可能な歌唱表現を可能とするエフェクタは限られている。知覚年齢に基づく歌声声質制御では、歌手は自身の歌唱出来る限界を超えた声質での歌唱が可能となる。また、知覚年齢スライダーの操作のみで声質を制御出来るため、歌手は、歌声エフェクタに対する複雑な制御方法を学ぶ必要はなく、直感的に歌唱表現を広げる事が出来る。

3 統計的手法に基づく歌声の知覚年齢制御

歌声生成において歌手が抱える身体的制約について述べる。次に、身体的制約を超える歌唱を可能とする知覚年齢に基づく歌声声質制御法について述べる。

3.1 身体的制約を超えた歌唱表現の実現

身体的制約により、一人の歌手が歌える歌唱表現は制限される。そのため、自分の歌唱可能な歌声（声質）が、歌唱したい楽曲に合わない場合などに、楽曲と歌唱表現のミスマッチがしばしば発生する。もし歌手が直感的な指標に基づいて、身体的な限界を超えて様々な声質へと声質制御可能な歌声エフェクタが実現できれば、より豊かな歌唱表現の実現につながる。

3.2 知覚年齢に基づく歌声声質制御

歌声の知覚年齢を制御したい歌手が、自身の歌唱する若い声（過去の歌声）や老いた歌声（未来の歌声）を適切に記録・保存しているならば、歌唱モーフing技術により知覚年齢制御は一部達成できるしかし、自身の未来の歌声はもちろん、過去の歌声を保有している歌手は稀である。個人性を保持した歌声の知覚年齢制御法 [1] では、幅広い年代の歌手らの音響特徴量を基に、知覚年齢に依存して変化する音響特徴量を修正重回帰混合正規分布モデルにより統計的にモデル化する事で、任意の入力歌手の歌声に対して知覚年齢に基づく声質制御を実現する。

4 未来ビジョン

本インタフェースは、歌を活用したエンターテインメントの可能性を広げることができる。例えば、カラオケにおいて、昔好きだった楽曲をその時の歌声で歌うことで、懐かしさや没入感を得られる事や、

逆に自分の未来を想像しながら年上の歌手の楽曲を歌うといった新たな楽しみ方につながる。他にも、歌声合成ソフトウェアにおいて、既存の声質のみでなく、ユーザが自由にかつ直感的に声質を制御できるため、歌声合成の可能性を広げる事が期待される。

5 まとめ

本稿では、歌声の声質を知覚年齢に基づいてリアルタイム制御出来るインタフェースを提案をした。本インタフェースにより、知覚年齢制御スライダーを動的に操作することで、歌手が望む声質を持った歌声での歌唱表現が可能となる。

今後は、ユーザが一から全てを設定するのではなく、まずはインタフェースがユーザの声質と歌いたい楽曲に合った制御パラメータを自動で決定し、さらなるチューニングをユーザが行うといった、より実用的なインタラクションの実現を目指す。そのためには、特定の声質表現語のみで声質を制御するだけでなく、歌声と楽曲の関係を考慮して、最適な声質制御パラメータを自動決定するインタフェースの構築が必要である。

謝辞

本研究の一部は、JSPS 科研費 26280060, 26280061 および JST OngaCREST プロジェクトの助成を受け実施したものである。

参考文献

- [1] K. Kobayashi, T. Toda, H. Doi, T. Nakano, M. Goto, G. Neubig, S. Sakti, and S. Nakamura. Voice Timbre Control Based on Perceived Age in Singing Voice Conversion. *IEICE Trans. Inf. Syst.*, E97-D(6):1419–1428, June 2014.
- [2] T. Toda, T. Muramatsu, and H. Banno. Implementation of computationally efficient real-time voice conversion. *Proc. INTERSPEECH*, Sept. 2012.
- [3] F. Villavicencio and J. Bonada. Applying voice conversion to concatenative singing-voice synthesis. *Proc. INTERSPEECH*, pp. 2162–2165, Sept. 2010.
- [4] 山本 和彦. 身の回りのものに任意の音色を割り当てて演奏可能な電子楽器インターフェース ～Possessing Drums ～. *インタラクション 2012*, Mar. 2012.
- [5] 中野 皓太, 森勢 将雅, 西浦 敬信, 山下 洋一. 基本周波数の転写に基づく実時間歌唱制御システムの実現を目的とした高品質ボコーダ STRAIGHT の高速化. *電子情報通信学会論文誌. A, 基礎・境界*, 95(7):563–572, Jul. 2012.
- [6] 李 翔, 曆本 純一. SmartVoice: 言語の壁を越えたプレゼンテーションサポーターティングシステム. *WISS2013 論文集*, Dec. 2013.