

ECHOES:歌い方をライブラリ化するシステムの検討

小川 真 岡部 誠 尾内 理紀夫*

概要. 本研究では、歌声合成システムを利用するユーザが魅力的な歌声を作成するにあたり、歌声合成に必要なパラメータをコントロールしやすくする技術の開発を目指す。VOCALOID2 初音ミク [1] の発売以来、歌声合成を使用した楽曲が広まっている。VOCALOID2 はサンプリングされた音声を利用し歌声を合成するソフトウェアであり、ユーザが入力した歌詞、メロディ、演奏情報から歌声を合成する。また、フリーソフトウェアである UTAU が発表され、アマチュアによる歌声コーパスの開発などが盛んに行われ、歌声合成をより表情豊かに歌わせようと日々交流が行われている [2]。こうした既存のソフトウェアではより人間らしい歌声や、表情豊かな歌声を合成するためには歌声の編集作業に慣れた人間が時間をかけて調整する作業が必要なため、歌声に不慣れた初心者には扱いづらいものである。そこで、ECHOES では実際の人間の歌唱をパラメータ化し楽譜情報と合わせて記録することで歌い方をライブラリ化する。さらに合成時には作成したライブラリを参照することで、初心者にも容易に人間らしい歌い方を作成できると同時に、さまざまな歌手を従来よりも簡単にライブラリ化できるシステムの開発を行う。

1 研究背景

歌い方を扱う技術には、中野らによる人の歌を真似るシステム VocaListener[3] が挙げられる。VocaListener は人間の歌から音高および音量を合成歌唱に転写することでシステムを実現している。しかし VocaListener では人の歌を真似るためには合成したい歌詞および旋律で歌った人間の歌唱が必要であり手軽とは言えない。また、Sinsy[4] をはじめとする HMM 歌声合成では学習する特徴ベクトルに歌い方が含まれるため、コーパスを収録した本人の歌を再現することができる。HMM 歌声合成では歌声モデルを作成する際に大量の歌声の収録が必要であり、加えて声質を変換し別の人の声にする話者適応の技術があるとはいえ、さまざまな歌い方をライブラリ化し容易に扱えるシステムとは言えない。

そこで本研究では VocaListener の技術を応用し、楽譜情報に基づき細かく切り出された範囲において転写を行うことで上記問題点を解決する。すなわち楽譜情報とそれに対応する歌声波形から、まず楽譜情報を細かい単位に分解し対応する位置の歌声波形から分析した振幅包絡と F0 遷移を歌い方パラメータとし、対応させた状態でライブラリへと記録する。合成時には与えられた楽譜情報をやはり細かく分解し、各断片に最も適した歌い方パラメータを検索し接続することでより人間らしい歌い方を容易にライブラリ化および合成できるシステムを実現する。

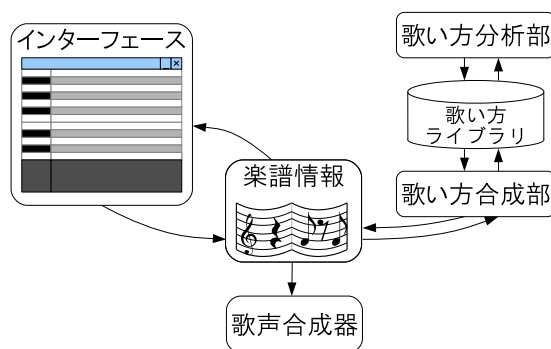


図 1. システムの概要

2 ECHOES の構成

図 1 に ECHOES の概要を示す。ECHOES では楽譜情報を音高および音長のいわゆる楽譜情報のほかに、振幅包絡などの演奏情報も含むものとする。インターフェースと歌い方合成部は楽譜情報を共有することで、合成された歌い方を容易に編集できるようにする。本章ではインターフェースとコアである歌い方分析部および歌い方合成部について説明する。歌声合成器については以前我々が開発した歌声合成器 v.Connect[5] を用いる。

2.1 インターフェース

インターフェースを図 2 に示す。ECHOES はピアノロールを採用している。上部のピアノロール上には音符情報以外に F0 遷移を表示することで、ユーザは歌詞や音符を入力する以外にも F0 遷移の編集および確認ができる。またコントロールトラックと

Copyright is held by the author(s).

* Makoto Ogawa, Makoto Okabe and Rikio Onai, 電気通信大学電気通信学部情報工学科

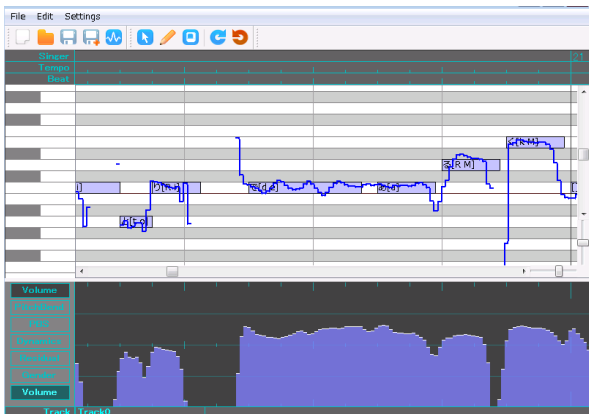


図 2. ECHOES のインターフェース

呼ばれる下部では振幅包絡をはじめ経時パラメータの編集および確認が可能である。

2.2 歌い方分析部

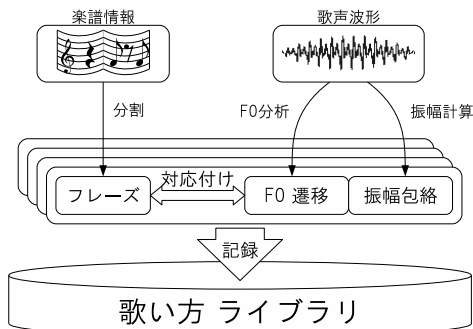


図 3. ECHOES 分析部の概要

図 3 に分析部の概要を示す。ECHOES 分析部では楽譜情報と歌声波形を歌声合成器に適した形としてデータベースに格納するため、楽譜と波形の分析を行う。まず、歌声波形は F0 推定法 DIO[6] と波形の振幅二乗平均を用いて、F0 遷移と振幅包絡のパラメータに変換される。次に楽譜情報はフレーズ単位に分割される。ある音符に対し自身の音高と音長と、先行音があれば先行音との音高差と先行音長、をフレーズ特徴量とし、対応する範囲の F0 遷移と振幅包絡の組を歌い方パラメータとして、これらに対応させ保存する。

2.3 歌い方合成部

図 4 に合成部の概要を示す。合成部ではユーザが入力した楽譜情報に適した歌い方パラメータを生成する。まず、ユーザが入力した楽譜情報からフレーズを計算し、各フレーズと最も似ているフレーズに対応する歌い方パラメータを検索する。検索された

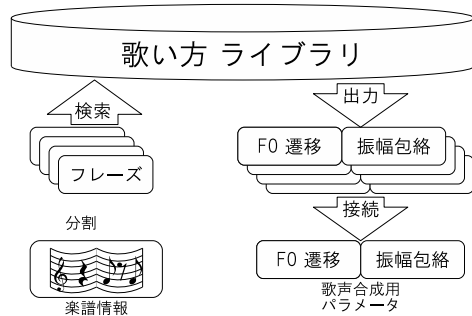


図 4. 歌い方合成部概要

各歌い方パラメータはフレーズの位置と長さに基づいて接続され、最終的な歌声合成用のパラメータとして整形し歌声合成器に出力する。

3 今後について

今後提案法を実装する段階において、まだモデルに取り入れていない発音位置の補正や、接続モデルの作成について検討を行う。また、F0 推定では誤差や推定の失敗が起こると考えられるし、より歌声に適したフレーズ化の方法を探りたい。最終的には主観評価を行い論文として発表するとともに、開発したシステムはオープンソースとして公開する予定である。

参考文献

- [1] 佐々木渉: 仮想楽器をリアルにする「未来(ミク)の記号」と、VOCALOID で注目される「人の形」「声の形」について、音楽情報科学研究会, Vol.2008, No.50, pp.57-60.
- [2] 森勢将雅, 河原英紀, 小川真: 協調的創造活動支援を目的とした歌声合成基盤技術の研究開発, 情報処理学会 研究報告 Vol.2012-MUS-94 No.9.
- [3] 中野倫靖, 後藤真孝: VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案, 情報処理学会 研究報告 2008-MUS-75, Vol.2008, No.50, pp.49-56, May 2008.
- [4] 徳田 恵一, 大浦 圭一郎 (名工大): [依頼講演] 自動学習により人間のように歌う音声合成システム — Sinsy —, 情報処理学会 研究報告 Vol.2012-MUS-94 No.1.
- [5] 小川真, 矢崎俊志, 阿部公輝: v.Connect: ユーザが声色操作可能な歌声合成器, 情報処理学会 研究報告 Vol.2012-MUS-94 No.10.
- [6] 森勢将雅, 河原英紀, 西浦敬信, 「基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法」電子情報通信学会 論文誌 D, vol.J93-D, no.2, pp.109-117, Feb. 2010.