

Title	人間・ 計算機と群を成す歌声合成技術の研究
Sub Title	Singing voice synthesis that can form ensemble with humans and computers
Author	高道, 慎之介(Takamichi, Shinnosuke) 中村, 友彦(Nakamura, Tomohiko) 深山, 覚(Fukayama, Satoru) 齋藤, 大輔(Saito, Daisuke) 猿渡, 洋(Saruwatari, Hiroshi)
Publisher	
Publication year	2025
Jtitle	科学研究費補助金研究成果報告書 (2024.)
JaLC DOI	
Abstract	本研究の目的は,人と人の歌唱相互作用の計算機的再現である.人間歌手同士が重唱する際,その各人は,好き勝手に歌唱するのではなく他人の歌声と調和するように自らの歌声を制御する.すなわち,人と人の歌唱の間には相互作用があり,その相互作用が,人と人から成る群による一体的な歌声と鑑賞体験をもたらす.他方,AI歌手(歌声合成機能を持つ計算機)は,その相互作用機能を有さない.人間歌手とAI歌手から成る群が歌唱する場合,人間歌手が一方向的にAI歌手に合わせるほかない.また,AI歌手同士の群においては一切の相互作用が無い.故に本研究では,歌唱相互作用機能を持つ歌声合成技術を研究開発した. The purpose of this study is to reproduce the interaction between human singers. When human singers sing together, each singer does not sing alone but controls his/her own singing voice so that it harmonizes with that of the other singer. In other words, there is an interaction between human singers and their singing, and this interaction brings about a unified singing voice and appreciation experience by a group of human singers. On the other hand, AI singers (computers with singing voice synthesizing functions) do not have such interaction functions. When a group of human and AI singers sing together, the human singer has no other choice but to unilaterally match the AI singer. In addition, there is no interaction between AI singers and human singers. Therefore, in this study, we researched and developed a singing voice synthesis technology that has a singing interaction function.
Notes	研究種目：挑戦的研究 (萌芽) 研究期間：2023～2024 課題番号：23K18474 研究分野：音楽情報処理
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KAKEN_23K18474seika

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

令和 7 年 4 月 25 日現在

機関番号：32612

研究種目：挑戦的研究（萌芽）

研究期間：2023～2024

課題番号：23K18474

研究課題名（和文）人間・計算機と群を成す歌声合成技術の研究

研究課題名（英文）Singing voice synthesis that can form ensemble with humans and computers

研究代表者

高道 慎之介（Takamichi, Shinnosuke）

慶應義塾大学・理工学部（矢上）・准教授

研究者番号：90784330

交付決定額（研究期間全体）：（直接経費） 5,000,000 円

研究成果の概要（和文）：本研究の目的は、人と人の歌唱相互作用の計算機的再現である。人間歌手同士が重唱する際、その各人は、好き勝手に歌唱するのではなく他人の歌声と調和するように自らの歌声を制御する。すなわち、人と人の歌唱の間には相互作用があり、その相互作用が、人と人から成る群による一体的な歌声と鑑賞体験をもたらす。他方、AI歌手（歌声合成機能を持つ計算機）は、その相互作用機能を有さない。人間歌手とAI歌手から成る群が歌唱する場合、人間歌手が一方的にAI歌手に合わせるほかない。また、AI歌手同士の群においては一切の相互作用が無い。故に本研究では、歌唱相互作用機能を持つ歌声合成技術を開発した。

研究成果の学術的意義や社会的意義

人間歌手同士が重唱する際、その各人は、好き勝手に歌唱するのではなく他人の歌声と調和するように自らの歌声を制御する。本研究では、このような歌唱相互作用機能を持つ歌声合成技術を開発した。この成果は、将来、人間とAIが共創して楽曲を作成する際に、AIから人間への一方的な作用だけでなく、人間からAIへの作用を実現するものである。

研究成果の概要（英文）：The purpose of this study is to reproduce the interaction between human singers. When human singers sing together, each singer does not sing alone but controls his/her own singing voice so that it harmonizes with that of the other singer. In other words, there is an interaction between human singers and their singing, and this interaction brings about a unified singing voice and appreciation experience by a group of human singers. On the other hand, AI singers (computers with singing voice synthesizing functions) do not have such interaction functions. When a group of human and AI singers sing together, the human singer has no other choice but to unilaterally match the AI singer. In addition, there is no interaction between AI singers and human singers. Therefore, in this study, we researched and developed a singing voice synthesis technology that has a singing interaction function.

研究分野：音楽情報処理

キーワード：歌声合成 コーパス 歌唱相互作用

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

人間歌手同士が重唱する際、その各人は、好き勝手に歌唱するのではなく他人の歌声と調和するように自らの歌声を制御する。すなわち、人と人の歌唱の間には相互作用があり、その相互作用が、人と人から成る群による一体的な歌声と鑑賞体験をもたらす。他方、AI 歌手(歌声合成機能を持つ計算機)は、その相互作用機能を有さない。人間歌手と AI 歌手から成る群が歌唱する場合、人間歌手が一方的に AI 歌手に合わせるほかない。また、AI 歌手同士の群においては一切の相互作用が無い。

2. 研究の目的

本研究の目的は、人と人の歌唱相互作用の計算機的再現である。人間の相互作用をモデル化するためのデータベース、および歌声合成技術を開発する。

3. 研究の方法

①人間の歌唱相互作用を模擬する歌声合成技術。

本研究では、重唱歌声合成における歌唱者間相互作用を計算機で再現する手法を提案した。まず 2 声部同時無音検出による楽譜特徴量・歌声特徴量の時系列同期前処理を行い、各声部の楽譜特徴量系列をゼロパディングした後、特徴方向で結合して time-lag モデルおよび duration モデルに入力する。予測した time-lag と duration に基づき楽譜特徴量をフレームレベルに展開し、統合されたフレーム特徴量と対数 F0 系列を LSTM ベース音響モデルに入力して、メルケプストラム係数、帯域非周期性指標、有声/無声フラグを予測する。学習時には従来の L1/L2 損失に加え、声部間の対数 F0 差分ロス ($\Delta F0$) およびメルケプストラム 0 次元パワー差分ロスを導入し、相互作用効果を再現可能なモデルを最適化した。推論では、各声部ごとに WORLD ボコーダで独唱波形を合成後、声部間の相対ゲイン調整を行い重唱音声波形を加算によって生成する。さらに、3 声部以上への拡張方法として、全声部並列モデルとミニバッチ内乱択サンプリングの二手法を検討した。

②作用を再現するためのデータベース

本研究では、既存の jaCappella コーパス v1 の設計方針に沿って、日本語アカペラ重唱データの多様性をさらに高めることを目的にコーパスを拡張した。著作権保護期間満了済みの童謡・唱歌から、バラード、エレクトロニック・ダンス・ミュージック (EDM)、ソウル/ファンクの 3 ジャンル、各 5 曲ずつ計 15 曲を選定し、5 名のプロ編曲者が Vo (リードボーカル)、S (ソプラノ)、A (アルト)、T (テナー)、Bs (バス)、VP (ボーカルパーカッション) の 6 声部編成で重唱アレンジを実施した。譜面は PDF および MusicXML 形式で整備し、日本語歌詞に加えて発音を考慮した Hepburn 式ローマ字表記を付与した。収録は COVID-19 感染対策下で 20 名のセミプロ歌手を個別スタジオに招集し、Shure SM58 を用いてチャンネル数 1, 48 kHz/24 bit の RIFF WAV 形式で取得した。各声部の録音には仮歌や他声部、ティック音をヘッドホンで提示し、複数テイクの中からベストテイクを組み合わせた。音高が譜面から ± 1 半音以上ずれる場合は Melodyne によるピッチ補正を行い、リバーブやイコライザーなどのエフェクト処理は行わず素材の純粋性を保持した。最終的に、新旧合計 50 曲の同期済み 6 声部音源と、モノラル/ステレオ混合音を、重唱分離・合成研究向けに配布可能な形で整備した。

4. 研究成果

①人間の歌唱相互作用を模擬する歌声合成技術。

実験ではまず、LSTM ベース音響モデルと拡散モデルベースを比較する主観評価を実施し、30 名のクラウドワーカーによる 5 段階 MOS 評価で LSTM (2.82 ± 0.15) が拡散モデル (2.56 ± 0.15) を有意に上回り ($p=0.017$)、以降の実験では LSTM を採用した。独唱音声の客観評価 (MCD [dB], F0-RMSE) では、提案する MultiTrack (MT) モデルは Baseline と比較してメルケプストラム歪みがやや増加 (10.82 vs 10.58) したものの、F0-RMSE は改善 (0.10 vs 0.12) し、他声部情報の有用性を示唆した。F0 差分ロス導入モデルは F0-RMSE が悪化した一方、パワー差分ロス (Powdiff) 導入モデルは LF0-RMSE を維持しつつメルケプストラム歪みの増加は最小限に留まった。重唱音声の主観評価 (まとめ MOS) では、MT+Powdiff が 3.16 ± 0.19 で Baseline (3.13 ± 0.19) と同等以上の評価を獲得し、MT+F0diff は 2.60 ± 0.21 に低下した。これにより、声部間の音量バランスを調整するパワー差分ロスが重唱音声の一体感向上に寄与することが明らかになった。

②作用を再現するためのデータベース

jaCappella コーパス v2 の妥当性を検証するため、まず歌詞のシラブル頻度を既存の商用楽曲コレクションと比較したところ、非語彙的シラブルの出現頻度偏りが両データで一致し、特に Bs では 99% 以上を占めるなど、実データの特性を忠実に再現していることを確認した (Vo 79.5%, S 38.2% など)。次に、Vo~T の音高分布は商用データと類似し重畳領域が多い一方、Bs は低域に偏っており、声部間の音域差が重唱分離の難易度に影響することが示された。さらに、既存の重

唱分離手法 DPTNet, MRDLA, X-UMX を用いて新規サブセットの分離実験を行ったところ、未知ジャンルに対しては DPTNet が最も高い Δ SI-SDR を示し、MRDLA はコーパス v2 のデータを学習に追加すると全声部で性能向上した。一方、X-UMX は一部声部で横ばいまたは微減となり、波形領域モデルとスペクトログラム領域モデルの特性差が性能に影響することが示唆された。スペクトログラム例では、DPTNet/MRDLA に広帯域アーティファクトが、X-UMX には高域成分欠落や他声部漏洩が見られ、モデル毎の改善方針が明らかになった。これらの結果から、データ拡張による分離性能向上の有効性と、今後のアノテーション強化の必要性が示された。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計3件（うち招待講演 0件／うち国際学会 0件）

1. 発表者名 武 伯寒 , 渡邊 研斗 , 中塚 貴之 , Tian Cheng , 中野 倫靖 , 後藤 真孝 , 高道 慎之介 , and 猿渡 洋
2. 発表標題 複数のオーディオエフェクトが適用された楽音に対するエフェクトチェイン推定と原音復元
3. 学会等名 日本音響学会春季研究発表会
4. 発表年 2024年

1. 発表者名 中村 友彦 , 高道 慎之介 , 丹治 尚子 , 深山 覚 , and 猿渡 洋
2. 発表標題 jaCappellaコーパスv2：重唱分離・合成のための日本語アカペラ重唱コーパスの拡張
3. 学会等名 情報処理学会 音楽情報科学研究会
4. 発表年 2023年

1. 発表者名 兵藤 弘明 , 高道 慎之介 , 中村 友彦 , 小口 純矢 , and 猿渡 洋
2. 発表標題 歌唱者間相互作用を再現するDNN重唱歌声合成の検討
3. 学会等名 情報処理学会 音楽情報科学研究会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

—

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分 担者	中村 友彦 (Tomohiko Nakamura) (50866308)	国立研究開発法人産業技術総合研究所・情報・人間工学領 域・主任研究員 (82626)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	深山 覚 (Fukayama Satoru) (90712191)	国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員 (82626)	
研究分担者	齋藤 大輔 (Saito Daisuke) (40615150)	東京大学・大学院工学系研究科（工学部）・准教授 (12601)	
研究分担者	猿渡 洋 (Saruwatari Hiroshi) (30324974)	東京大学・大学院情報理工学系研究科・教授 (12601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------