

審査の結果の要旨

氏名 須田 仁志

本論文は「複数人歌唱楽曲に対する音楽情報処理に関する研究」と題し、全6章からなる。音楽は、有史以前から続く人類の重要な文化の1つであり、娯楽や宗教などの様々な目的で演奏及び鑑賞されている。情報処理の発展によって様々な音楽情報処理技術が提案、研究されるようになってきたが、本研究では複数人が歌唱する楽曲に対する音楽情報処理のうち、とくに音楽理解技術を起点とした技術に着目して論じている。

第1章「序論」では、本研究の背景となる音楽そのものの学術的な立ち位置から出発し、音楽情報処理の現在までの流れを概観するとともに、本論文の位置づけを明確にし、研究の目的を明らかにするとともに本論文の構成を示している。

第2章では、本論文の前提知識となる音声の分析合成に関する基礎知識が述べられている。人間の発声機構、音声から抽出される音響的特徴量について説明し、その後、音楽情報処理に欠かせない基本周波数の推定手法について種々の手法を述べている。

第3章では、複数人が歌唱する楽曲から「誰がいつ歌唱しているか」を推定する技術として、「歌唱者ダイアライゼーション」という技術について論じている。複数人の会話から「誰がいつ発話しているか」を推定する話者ダイアライゼーション技術との対応関係や関連研究に触れたうえで、高い精度で歌唱音声に対してダイアライゼーションを行う技術を提案している。本論文の提案法は、特定の話者表現を持つ話者が各フレームで発話しているかを推定する **TS-VAD** と呼ばれる機構を持つ話者ダイアライゼーションを基礎とし、各歌唱者の歌唱状態を個別に推定する **target singer VAD** という機構を導入し、加えて同時歌唱者数の推定を高精度に行う **Cosacorr** スコアと呼ばれる新たな音響特徴量を用いている。最終的に **Spleeter** による音源分離、**Arcface** による歌唱者表現と組み合わせて総合的な提案法を構成し、市販の **CD** に収録されているパート割りのある楽曲に対する評価実験において、既存のダイアライゼーション手法と比較して高い精度を実現している。

第4章は、「非負値行列因子分解を用いたノンパラレル声質変換」と題し、歌唱者を自由に入れ替えることのできる音楽インタフェースを目的として、汎用的なノンパラレル声質変換技術である **Soft INCA** アルゴリズムを提案している。声質変換モデルの学習には、従来、入力話者と出力話者が同一の発話内容を発話したパラレルデータが必要であった。一方、同一でない発話を学習に利用するノンパラレル声質変換は多数提案されているが、膨大な別データに依拠した背景知識や入出力話者の発話が大量に必要であった。本論文の提案法では、非負値行列因子分解の時変成分と時不変成分に分解する特徴を利用して、入力話者による発話と出力話者の音響モデルの連続的なアラインメントを得ることで、アラインメントの不一致に伴う品質劣化を回避し、少量の入力話者の発話でノンパラレル声質変換を実現している。

第5章では、複数人が歌唱する楽曲を加工し鑑賞できるインタフェースである **VocalRemixer** を提案している。**VocalRemixer** では複数人が歌唱しているパート割りのある楽曲に対して、各歌唱者を相互に変換することにより、各歌唱者があたかもソロで歌唱した歌声を合成する。主観評価実験において、提案するインタフェースの使いやすさ、新規性、魅力を確認している。

第6章は「結論」であり、本論文で提案する各手法についてまとめており、また複数人の歌声を扱う音楽情報処理の今後の展望について述べている。

以上これを要するに本論文は、複数人歌唱楽曲に対する音楽情報処理のうち、特に音楽理解技術を起点とした技術について、高精度な歌唱者ダイアライゼーション技術および非負値行列因子分解に基づく汎用的なノンパラレル声質変換技術という二つの要素技術を提案し、**VoiceRemixer** という新たな歌唱インタフェースを通してこれらの技術を結合することで、従来の音楽情報処理では適切に論じられてこなかった、複数人歌唱楽曲の分析、加工、提示技術を実現しており、情報工学、特に音楽情報処理に寄与するところが少なくない。

よって本論文は博士（工学）の学位請求論文として合格と認められる。