論 文 の 内 容 の 要 旨

論文題目　　複数人歌唱楽曲に対する音楽情報処理に関する研究

氏　　名　　須田　仁志

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，娯楽や宗教など様々な目的で演奏および鑑賞されている．情報技術の発展によって音楽を情報として扱うことができるようになり，様々な音楽情報処理技術が提案，研究されるようになった．音楽鑑賞においては，蓄音機や音楽のデジタル化によって，かつてその場で演奏される中で耳を傾け鑑賞されていた音楽は，好きな時間や場所で様々な目的で聴くことができるようになった．また，音楽サブスクリプションサービスの発展によって，膨大な音楽ライブラリをいつでもどこでも楽しむことができるようになった．音楽から様々な情報を抽出する技術は音楽理解技術とよばれ，音楽の検索や鑑賞に役立てられている．たとえば，コード進行や楽曲のジャンル，ムードなどといった音楽的情報の抽出技術は，膨大な楽曲に自動でメタデータを与え，曲名やアーティスト名だけでない様々なクエリによる音楽検索を可能にする．自動採譜技術は，特別な訓練なしに音楽を採譜することを可能にし，採譜された楽譜をもとに演奏を楽しむことを可能にする．音楽理解技術は音楽の可視化にも応用されており，楽曲を視覚的に楽しむことのできるインタフェースや，楽曲間の関係を可視化するインタフェースなどが提案されている．音楽を合成したり創作したりする技術も音楽情報処理に含まれる．たとえば，自動作編曲や自動作詞技術によって，音楽的な素養や訓練なしに好みの音楽を制作できる．また，歌声合成技術によって好みの歌手に既存の楽曲や制作した楽曲を歌わせることができ，音楽制作をより自由に豊かにしている．

複数人が歌唱する楽曲に対する音楽情報処理としては，各歌唱者やパートに分離する音源分離技術や，自然な同時歌唱を合成するための歌声合成技術などが研究されてい

る．とくに複数人が楽曲を歌唱する際には，歌唱者が入れ替わりながら歌唱するパート割り構造を持つことも多く，それを対象とした研究もなされている．本論文ではとくに，パート割りのある楽曲から「いつ誰が歌唱しているか」を事前知識なく推定する歌唱者ダイアライゼーションとよばれる技術について検討する．これまでの歌唱者ダイアライゼーション研究では，会話音声を対象とした話者ダイアライゼーション技術を利用している．しかし，歌声は話声と比べて音高の幅が広く，音素継続長が長く，また複数人が同時に歌唱している区間が長い．したがって，話者ダイアライゼーション手法をそのまま用いては，高い精度のダイアライゼーションを実現できないことが懸念される．そこで本研究では，話者ダイアライゼーション手法を拡張し，高い精度で歌唱者ダイアライゼーションを行える手法を提案する．提案法には，各時刻で同時に何人が歌唱しているかを推定する同時歌唱者数推定技術や，ArcFaceとよばれる埋め込み表現抽出法を用いた歌唱者表現を導入する．とくに同時歌唱者数推定では，Cosacorrスコアとよばれる新たに提案した音響特徴量を用いることで，高い精度での推定を可能にする．評価にはコンパクトディスク（compact disc；CD）に収録された日本のアイドルソングの音響信号をそのまま用い，現実的な条件で評価した．実験により，既存の話者ダイアライゼーション手法と比較して，提案する歌唱者ダイアライゼーション手法の有効性が示された．また，CosacorrスコアやArcFaceの有効性についても確認された．

本論文では，歌声の声質を別の歌唱者の声質に変換する声質変換法についても検討する．声質変換は，入力話者の発話から抽出された音響特徴量を変換モデルによって変換し，変換された音響特徴量を用いて合成することで達成される．古典的な声質変換手法においては，入力話者と出力話者が同じ内容を発話したパラレルデータを必要とする．しかし，パラレルデータが利用できない条件では変換モデルを学習することができない．そこで，パラレルデータを必要とせず，入出力話者が異なる内容を発話した音声を用いて学習可能なノンパラレル声質変換法が広く検討されている．ノンパラレル声質変換法の多くは，音声から話者情報と言語情報を分離し，分離された話者情報のみを置換することで実現される．これまでのノンパラレル声質変換法は，膨大な話者の発話から構築した背景知識や，入出力話者による多数の発話を必要としており，システム全体を構築するための学習コーパスの用意が容易ではない．本研究では，非負値行列因子分解（non-negative matrix factorization；NMF）とよばれる行列分解法を利用して，背景知識なしに音声の話者情報と言語情報を分離することで，少量の入出力話者の発話のみを必要とする変換モデルの学習法を提案する．NMFは，スペクトログラムをはじめとした非負の物理量を，テンプレートである基底と，そのテンプレートの利用状態を表す生起状態に分解する手法である．本論文では，NMFの生起状態が，基底と音響特徴量間の連続的なアラインメントであることに着目する．このアラインメントは，INCAアルゴリズムとよばれるノンパラレルデータ間でアラインメントを得る手法と同様の手順で得られ，ノンパラレルな条件下でも変換モデルの学習に必要な生起状態を得られる．INCA

アルゴリズムと異なり，得られるアラインメントが離散的ではなく連続的であるため，高い自然性を持つ変換音声が合成できることが期待される．話声を用いた評価実験により，既存のノンパラレル声質変換法であるINCAアルゴリズムやCycleGAN-VCと比較して，提案法はより自然な変換音声を合成できることが確認された．本論文では，入力話者に関して事前に学習する必要のないone-shot変換についても検討し，提案法を用いてone-shot変換が実現できることを確認した．

複数人歌唱の楽曲を楽しむことができるインタフェースの一例として，本論文ではVocalRemixerとよばれるWebインタフェースを提案する．VocalRemixerは，楽曲からパート割りを推定する歌唱者ダイアライゼーション技術と，歌声を変換するノンパラレル声質変換技術の両方を利用する．このインタフェースでは，パート割りのある楽曲に対して自由にそのパート割りを編集することができ，特定の歌唱者にソロで歌唱させたり，まったく異なるパート割りで歌唱させたりすることができる．被験者にインタフェースを利用させて行った評価実験では，インタフェースの新規性や魅力について高い評価を得た．また，被験者のコメントから，インタフェースをさらに拡張することで，より魅力的なインタフェースを構築できる可能性が示された．

本論文では，本研究における究極の目的である複数人歌唱楽曲に対する音楽情報処理において，とくに音楽理解技術を起点とした技術に着目して述べる．音楽情報処理には様々な方向性の研究が含まれており，それぞれの技術が相互に作用している．これは，複数人歌唱の楽曲に対する音楽情報処理においても同様である．本論文で述べる技術は複数人歌唱楽曲に対する音楽情報処理のほんの一端であり，他の様々な音楽情報処理や未検討の音楽情報処理と相交わることで，音楽情報処理がより発展するものである．