

論文内容の要旨

博士論文題目 Acoustic modeling and speech parameter generation for high-quality statistical parametric speech synthesis

(邦題：高音質な統計的パラメトリック音声合成のための音響モデリング法と音声パラメータ生成法)

※ 論文題目が外国語の場合はワープロ等を用いること。また、その邦文を論文題目の下に（ ）で記入すること。

氏名 高道 慎之介

要旨

音声は、人間にとって基本的なコミュニケーションツールのひとつである。音声合成は、コンピュータにより音声を人工的に合成する技術であり、本論文では、任意のテキストから音声波形を合成するテキスト音声合成技術と、入力音声波形を異なる音声情報を持つ音声波形に変換する声質変換技術を指す。この技術は、音声対話や音声翻訳等の音声コミュニケーションシステムを初めとする広い応用を見据え、活発に研究されている。

機械学習と計算環境の発達により音声合成に対する多様な統計的手法が提案される中、本論文で取り扱う隠れマルコフモデル (HMM) に基づくテキスト音声合成 (HMM 音声合成) と混合正規分布モデル (GMM) に基づく声質変換 (GMM 声質変換) は、数理モデルとしての強力性・柔軟性から、強く支持されている。一方で、これらの手法において合成される音声の音質は、自然音声と比較して著しく劣化する。その要因は、分析・生成部におけるパラメータ表現のエラー、学習部における不正確な音響モデリング、また、生成部における過剰な平滑化の3要因に分類されるが、本論文では特に、後者の2要因に焦点を当て、合成音声の音質改善に取り組む。

学習部における音響モデリングの問題点の一つは、統計モデリングにおける平均化処理により、個々の音声波形の情報が消失する点である。これに対し本論文では、直接的に音声波形を利用する素片選択合成法の考えを導入する。提案法では、各音声波形の情報を、未知の入力情報に対して頑健な統計モデルとして保持し、更に、複数の統計モデルから一つの混合モデルを構築する。これにより、従来の HMM 音声合成と GMM 声質変換の柔軟性を保持しつつ、より高音質な音声合成可能となる。

次に、生成される音声パラメータの過剰な平滑化は、合成部における音質劣化の主要因である。平滑化現象を定量化する特徴量を生成パラメータから抽出し、それを自然音声パラメータの特徴量に近づく様に補償することで、この過剰な平滑化の問題は緩和される。本論文では、過剰な平滑化を一層効果的に定量化する新たな特徴量として、変調スペクトル (MS) を導入し、生成パラメータ系列の MS を補償するポストフィルタを提案する。このフィルタ処理は、通常の HMM 音声合成・GMM 声質変換から独立した処理であるため、MS による音質改善効果のみならず、他の音声合成方式への容易な移植性をもたらす。

我々は更に、ポストフィルタ処理において個別に考慮されていた、音響モデルと MS の統計モデルの基準を、同時に最適化するアルゴリズムを提案する。ここでは、生成部における過剰な平滑化を直接的に緩和する音声パラメータ生成アルゴリズムと、高音質かつ高速な音声合成を可能にする音響モデル学習法を提案する。実験的評価により、(1) 提案するパラメータ生成法は、従来のパラメータ生成法を超える音質改善効果をもたらすこと、(2) 提案する学習法は、高速な音声合成能力を保持しつつ従来の学習法を超える音質改善効果をもたらすことを示す。

(論文審査結果の要旨)

音声合成は、コンピュータにより音声を人工的に合成する技術であり、近年統計的な音声のモデリングにより長足の進歩を遂げたが、合成音の品質には不自然さが残されていた。本論文では、任意のテキストから音声波形を合成するテキスト音声合成と、入力音声波形を異なる音声情報を持つ音声波形に変換する声質変換において、この不自然さを取り除く技術の研究を行った。現在の音声合成技術として、隠れマルコフモデルに基づくテキスト音声合成と、混合正規分布モデルに基づく声質変換が、数理モデルとしての柔軟性から、強く支持されている。本論文では、音質劣化の主要因として知られる、学習部における不正確な音響モデリングと、生成部における過剰な平滑化に焦点を当て、合成音声の音質改善に取り組んだ。提案法では、音声素片を直接的に使用する素片選択合成法のアイデアを導入し、各素片に対応する統計モデルを構築することで、従来の柔軟性を保持しつつ、より高音質な音声を合成可能にする。実験的評価により、従来の音響モデリングよりも高音質な音声を合成できることを示した。生成部における過剰な平滑化現象は、平滑化現象を定量化する特徴量を合成音声パラメータから抽出・補償することで、緩和される。本論文では、新たな特徴量として変調スペクトル (MS) を導入し、従来の特徴量である系列内変動 (GV) よりも効果的に、平滑化現象を定量化できることを明らかにした。更に、MS を考慮した音質改善法として、(1) ポストフィルタ：音声合成方式に依存せず使用可能な手法、(2) パラメータ生成：従来の合成基準と MS 基準の同時最適化で、高音質な音声を生成する手法、(3) 音響モデル学習：従来の学習基準と MS 基準の同時最適化で、高音質かつ高速な音声合成を可能にする手法を提案し、実験的評価により、従来の GV よりも高音質な音声を合成できることを示した。

これらの成果は、従来技術では本質的に解決困難であった問題に対して、解決策を示すばかりでなく、合成品質国際コンペにおいて第一位の性能を達成し、IEEE 件を含む 4 編の学術論文、9 編の査読付き国際会議論文として発表し、さらには 14 編の国内発表を行っていることから、非常に高く評価できる。