

論文の内容の要旨

論文題目 音響情報・言語情報に基づく調音運動の復元

氏 名 内田 秀継

音声の生成過程における舌や口唇の運動を調音運動と呼ぶ。調音運動によって音声の音響特性が特徴付けられ、それが音韻性として知覚される。つまり、調音運動は、音声に言語的情報を付加する役割を持つ。調音運動に関する情報は、音声生成過程における音韻性の表現として、音声認識や声質変換、発音トレーニングなど様々な分野での応用が期待されている。

その一方で、調音運動の大部分は口腔内でおこなわれるため、ビデオカメラなどを用いた体外からの観測が難しいという問題がある。調音観測システムと呼ばれる特殊な機材を用いることで、発話中の話者の調音運動を測定することが可能だが、機材や測定に必要なコストは非常に高いものとなっている。多くのユーザーを想定する音声認識や声質変換、発音トレーニングなどのシステムに、高コストな調音観測システムを組み込むことは難しい。そこで、音声から調音運動の情報を推定（復元）する音声-調音マッピングが検討されている。

音声-調音マッピングは調音観測システムに頼らずに調音運動情報を獲得する技術として注目を集めている。音声-調音マッピングは、音声と調音運動の同時測定データ（音声-調音パラレルデータ）を用いて、音響空間と調音運動空間の対応（変換）関係を機械学習によってモデル化する。このとき、特定の話者（モデル話者）の音声{調音パラレルデータを用いて構築した変換モデルは、その話者の専用モデルとなり、モデル話者以外の音声を入力した場合、推定精度は著しく悪化する。つまり、従来の音声-調音マ

ッピングは、入力としてモデル話者の音声（音響情報）が与えられた場合にしか適用できないという制約がある。この適用範囲の狭さが音声-調音マッピングを実用的な応用から遠ざけている一因となっている。そこで、本研究では、音声-調音マッピングの適用範囲を拡大するために、様々な情報から調音運動を復元することを試みる。まず、モデル話者以外の音声、つまり、任意の話者の音声から調音運動を復元する手法として話者正規化音声-調音マッピングを提案する。話者正規化音声-調音マッピングでは、声質変換を用いて任意話者の音声をモデル話者の音声に変換することで、入力音声と専用モデルの間での音響的ミスマッチを軽減し、高い精度で調音運動を推定することが可能になる。次に、音響情報以外の情報から調音運動を復元する手法を提案する。本研究では、音声の構造的表象によって抽出される言語的特徴量に注目し、従来の音声-調音マッピングでは困難であった、音響情報が存在しない音声に対する調音運動の復元を試みた。