

平成27年度 メディア科学専攻修士論文要旨

大西 研究室	氏 名	飯 沼 嗣 業
修士論文題目	読唇つき補聴器 ～母音推定を用いた音声強調～	
<p>工事現場やカクテルパーティなどの高雑音環境下において、聴覚のみで正確に聞き取り、理解することは計算機のみならず人間においても困難である。ノイズ除去によって精度を向上させる試みもあるが、ノイズの種類によって性能が左右され、非定常雑音環境下では思うように精度が出せない場合がある。しかし、こういった環境下でも認識精度を向上させるために、人間は視覚をも用いて聞き取りを行っているという事実がある。それと同様に計算機においても視覚情報を利用することで聞き取り精度を向上しようという試みがある。</p> <p>そこで本研究では、画像の情報、特に唇の情報を新たに加え、唇の動きを捉える読唇という方法を実装することで解決を試みる。そして、唇の動きから母音を推定し、母音ごとの特徴に合致した音声強調を行うことで、ノイズの種類に依存しない音声認識精度が向上する補聴器システムを提案する。</p> <p>提案システムは、顔・唇の特徴量抽出、画像を利用した母音推定、音声強調処理の3つで構成される。顔・唇の特徴量抽出では Active Appearance Model に基づいた特徴点の取得が用いられる。画像を利用した母音推定では、多層パーセプトロンが用いられ、音声強調処理ではバターワースフィルターを用いた音声強調が行われる。なお、AAM と多層パーセプトロンは事前に学習が行われ、母音強調フィルターも事前に各母音に対応したものを作成されている。</p> <p>音声と画像それぞれに対する性能評価の実験が行われた。</p> <p>画像による母音認識精度実験では、F 値は最大でも 0.383 と低かった。これはデータの取得方法に問題があり、より自然な発話時のデータを多量に用意する必要があることが確認された。</p> <p>音声の評価実験では、機械による認識率は、路上ノイズを用いて SN 比 0dB のとき音声強調フィルター適用前が 48% であるのに対し、適用後が 40% と低い評価であった。このことから、提案手法は機械による認識に向いていないことが実証された。しかし、これはフィルター適用後の音声を利用してモデルを訓練することで解決可能であると考えられる。聴覚による単音母音判別実験における認識率は、路上ノイズを用いて SN 比 -15dB のとき、音声強調フィルター適用前が 40% であったのに対し、適用後は 90% であり、その差は 50% と大幅な精度の向上を達成した。また、カクテルパーティノイズを用いて SN 比 -15dB の場合に、音声強調フィルターを適用後の認識率は 96% であり、ノイズの種類に拠らずに精度の向上が確認された。さらに、聴覚による母音単語判別実験における認識率は、SN 比 15dB のときの音声強調フィルター適用前が 16% であるのに対し、適用後は 84% と 68% の認識率の向上が確認できた。このことから、単語においても認識精度を向上することが確認された。</p> <p>これらの実験から、提案システムは機械による認識に必要な特徴を失ってはいるが、聴覚では高い精度で認識されること、ノイズに拠らずに精度が向上することを示した。</p>		