平成25年度 情報工学コース卒業研究報告要旨

大西 研究室	氏　　名	飯 沼　嗣 業
卒 業 研 究 題 目	計算機を用いた読唇による母音推定	

背景と目的

工事現場などの高雑音環境下では，計算機による音声認識の精度は低い．そのため人間が聞き取りに用いている視覚的情報を利用することで，計算機での聞き取り精度を向上しようという試みがある．

本研究では，母音推定で通常用いられている唇の特徴量のほかに顎及び歯の特徴量を加えて識別を行い，推定精度の差の検証を行った．また，複数人での識別と個人のみでの識別の結果を比べることで，特徴量の個人差による影響について調査した．

推定手法

「あいうえお」の母音及び「ん」を強調して発音をした顔画像を用意する．唇の上下の幅と左右の幅，唇の下端と顎の先端までの幅，及び歯の可視不可視及び口領域の画素数を特徴量として抽出する．その後，それらのデータを「ん」のデータの値で割り，正規化を行ったのち，多層パーセプトロンで学習を行い，母音の推定を試みた．

実験と結果

実験１　1母音に対し5枚，一人当たり30枚の画像を，4人分取得した計120枚の画像を用いて実験を行った．特徴量は唇の上下の幅(A)と左右の幅(B)，唇の下端から顎までの幅(C)，歯の可視/不可視(D)および口領域の画素数(E)の5つを用いた．4ホールドクロスバリデーションで，表1の5パターンの特徴量の組み合わせで実験を行った．パターン1は全パターン中最低の50%の正答率であり,「い」と「え」,「う」と「お」の区別が失敗する誤識別が目立った．顎を用いたパターン2では正答率59%であり，パターン1より10%ほど良くなった．特にパターン1における「お」の正答率40%に対し65%となり「う」と「お」間の誤識別が緩和された．パターン3の正答率は56%であり「え」及び「お」のデータの正答数に改善が見られた．パターン4では正答率66%であり全体的に正答率の向上が見られた．パターン5では57%と，正答率がパターン4に比べが落ちた．

表1：利用特徴量及び正答率

	利用特徴量	実験1 正答率(SD)	実験2 正答率(SD)
パターン1	A+B	50%(2.5)	61%(0.8)
パターン2	A+B+C	59%(1.3)	81%(1.0)
パターン3	A+B+D	56%(2.3)	68%(0.8)
パターン4	A+B+C+D	66%(1.3)	85%(0)
パターン5	A+B+C+D+E	57%(1.5)	83%(2.3)

実験２　同一の人物で1母音に対し5枚，1セット30枚として，時間を変えて4セット取ったデータを用いて，実験1と同様の5パターンを用いて識別した．実験1と同様に，パターン2では「お」のデータがパターン1に比べ大きく改善されており，正答率が40%から90%になった．それ以降のパターンにおいても実験1と同様の傾向が見られたが，パターン5においてはパターン4に比べ劣るものの，実験1の-9%に比べると減少はかなり少なく-2%であり，誤差の範囲だと考えられる．

まとめ　以上のことから唇の下端から顎までの距離および歯の可視/不可視の特徴量は，母音推定において有用であることが分かった．また，複数人での実験と同一人物による実験との比較によって，歯の可視/不可視の特徴以外は個人差が大きいことが確認された．これらのことから母音推定において個人差による影響は大きく，個人差を緩和した特徴量の設定を行う，あるいは学習データを個人ごとに保持しておき，その個人のデータに重みを大きくして他のデータと合わせ識別を行うといった方法を取ることによってより良い判別結果を得ることができると考えられる．