

唇形状を用いて動画閲覧を行うハンズフリーインタフェースの実装

江崎 将矢

1. はじめに

近年，感染症対策や障害者のため手を使わずにコンピュータを操作できるハンズフリーインタフェースが開発されるようになった．その中の1つにリップリーディング(読唇術)技術があるが，日本語に対して高い認識率ではない．北原らの研究[1]では，五十音に対応したデータセットを作成し，テストしているが約45%の文字認識率という結果が出ている．そこで本研究では，母音だけの簡単な唇形状を用いて動画の操作を行うインタフェースを試作し，どの程度認識率が向上するのかを調べる．

2. 唇形状を用いたインタフェース

2.1 唇形状の認識方法

本研究は python の開発環境である Spyder を使用し，カメラから取得した顔画像から母音発音時の唇形状を認識するインタフェースを作成する．各形状の判別は，形状を作った時の横幅が通常時(図1)の横幅より大きい時「あ・い・え」，小さい時「う・お」と区別し，

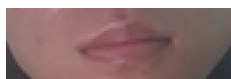


図1 通常時の唇

その後，唇の縦横の割合により判別する．通常時の座標はプログラムを実行後，最初に取得してもらう．座標取得は通常時の唇の状態動画の画面を2回押すとできる．縦横の割合は，縦÷横×100で計算し，あ：45~60%，い：20~30%，う：10~22%，え：31~40%，お：33~45%，とする．唇を動かしているときに誤操作されないために，ある形状をしてから約1秒継続した場合に唇の形状が認識される．唇の座標取得には Dlib を使用した．唇のランドマークは図2のように取得され，本稿では，61，63，65，67番の座標から唇の形状を判別する．



図2 唇のランドマーク

2.2 動画再生

唇形状と動画操作の対応は次のとおりである．[あ]は再生，[い]一時停止，[う]は早送り，[え]は

巻き戻し，[お]は最初から再生である．動画は OpenCV を使用し，指定した動画ファイルをフレームごとに読み取り，表示させる．動画はプログラムを実行すると再生された状態から始まり，動画の画面を押すと動画とカメラ映像の再生停止が切り替えられる．

3. 実験と評価

本インタフェースの認識精度を求めるために，被験者4名に「あいうえお(一時停止，再生，早送り切り替え，巻き戻し切り替え，最初から再生)」の順で操作を20回ずつ行ってもらい，認識率を求めた．結果を表1に示す．また，4人の被験者に操作性についてアンケートを行った．

表1 実験結果

項目	あ	い	う	え	お
認識率	90%	95%	80%	90%	80%

実験の結果として，表1のように簡単な唇形状だけで高い認識率を出すことができたが，「う」と「お」，「あ」と「え」の唇形状が似ているため，それぞれで誤認識されることがあった．また，唇形状を変化させたが認識しない場合があったため，縦横比率の調整の必要がある．

操作性について，操作したときのフィードバックがないため，正常に動作しているのかわかりづらいとの指摘があった．また，同じ形状を継続しすぎると操作が2重で行われることがあったため，操作後少しの間反応しない時間を作る必要がある．

4. まとめ

今回の研究で，簡単な唇形状だけでコンピュータを操作して認識率が高くなるが，多くの操作をすることが厳しく，形状ごとの処理の内容を覚える必要があることがわかった．このことから，必要操作の種類が少ないインタフェースに適しているが，複雑な操作をするためには，視線情報などほかの部位情報と組み合わせる必要があることがわかる．

[参考文献]

- [1] 北原 瑠伊，張 力峰：機械学習を用いた日本語読唇における五十音データセット作成の提案，産業応用工学会全国大会2021，2021年

[担当教員] 石原 真紀夫