

手指信号と非手指信号を組み合わせた手話単語認識の検討

岡田 治樹[†]

† 法政大学大学院理工学研究科応用情報工学専攻

赤松 茂^{††}

†† 法政大学理工学部応用情報工学科

1. はじめに

手話会話において、聴覚障がい者は手の動きだけでなく顔情報を必要としていることが分かった[1]. また顔情報を用いることで特定の単語ごとに文法的判別を行えることが明示されている[2]. 本研究では、5つの手話単語に対する肯定・否定・Yes/No 疑問形の分類を目指して、手指と腕の情報を表す手指信号と、顔の表情を表す非手指信号を組み合わせた手法を提案する.

2. キーフレーム抽出

本研究では、OpenPose を用いてカラー動画より身体の座標点を抽出する[3]. 取得した首と両手首の加速度を用いて、各動画画像から目視で選択した手話単語を表すキーフレームとの相関性を式(1)のピアソンの積率相関係数を用いて算出し、相関性が高いフレームを抽出する.

$$r_{xy} = \frac{\sum_{k=1}^K \sum_{i=n}^m (x_k - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=k}^c (x_i - \bar{x})^2} \sqrt{\sum_{i=n}^m (y_i - \bar{y})^2}} \quad (1)$$

3. 特徴量

2次元座標点を用いて手指信号かつ非手指信号の特徴量を算出した. 手指信号として両手首の加速度の他に、左右の腕と各指の方向ベクトルを、非手指信号として頭部の加速度の他に、頬の面積を特徴量とした. 図1に特徴量を示す. 算出したスケールが異なる特徴量をそれぞれ標準化および正規化した.

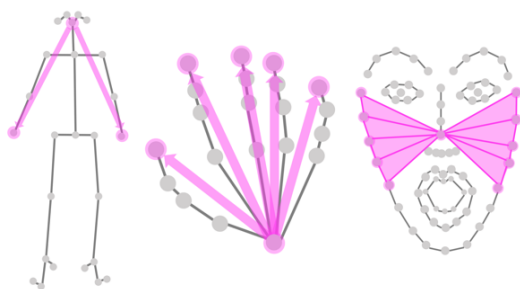


図1 MSとNMSの特徴量

4. 実験

4.1 実験環境

本研究では、手話動画を見て手話を学んだ健聴者を準ネイティブサイナとして定義し、被験者10名によるデータセットを構築した. 5つの単語の各文法において10回ずつ撮影を行った.

4.2 識別手法

撮影した被験者10名分のデータセット全体に対してk分割交差検証による評価実験を行った.

手指信号と非手指信号の組み合わせた手法として、特徴ベクトルを結合してPCAを行う手法を提案する.

各動画画像で標準化または正規化処理を行った特徴ベクトルを

$$x_i = [x_{i1}, x_{i2}, \dots, x_{im}, \dots, x_{nm}]^T \quad (2)$$

と表す. ここで、 n はフレーム番号を、 m は特徴ベクトルの総数を表す. また、1つの動画データを

$$X_f = [x_1, x_2, x_3, \dots, x_i]^T \quad (3)$$

と表す. ここで、 i は動画番号、 f は動画の総数を表す. 各単語の文法ごとの動画画像データに対して、主成分分析を行い、CLAFIC法による評価実験を実施した. μ は平均ベクトル、 φ_l は正規直交基底ベクトルとする.

$$d^2 = \|X - \mu\|^2 - \sum_{l=1}^L \{(X - \mu)^t \varphi_l\}^2 \quad (6)$$

5. 結果

文法的判別の実験結果を図2に示す. 単語ごとの文法の平均認識率は69.6%に達した. また文法ごとの認識精度を比較すると、どの特徴量においても否定形が高かった.

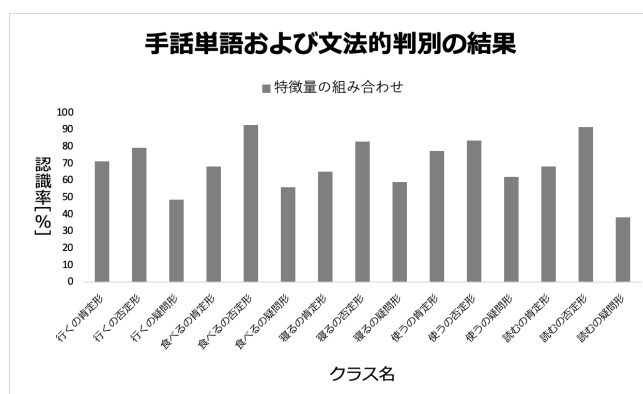


図2.5 単語の文法的判別結果

6. 今後の課題

被験者のデータセットを拡充し、新たな特徴量および組み合わせ手法の改良の提案を行う.

7. 謝辞

本研究の一部には、科学研究費補助金(基盤(C)19K12188)の助成を得た.

参考文献

- [1] 神田和幸, “ドラえもん手話の実例とNMSの情報伝達”, 第32回可視化情報シンポジウム, Vol.24, Suppl.No.1, pp.277-278, 2004
- [2] 岡田治樹, 新井駿太, 赤松茂, ”手話における顔情報を用いた文法的判別の検討”, DIA2020, 2020
- [3] C.Zhe, S.Tomas, W.Shih-En, and S.Yaser, “Realtime Multi-person 2d Pose Estimation Using Part Affinity Fields”, Proc.of the 2017 IEEE Cong. On CVPR, pp.1302-1310(2017)