4ZL-01

# 表情認識を用いた感情付加による手話翻訳

土屋　結華　　小河　誠巳
東京電機大学理工学部理工学科情報システムデザイン学系

## 1　はじめに

聴覚障害者には,コミュニケーションの手段の一つである手話という言語がある.手話は,手指信号である手や指の動きと非手指信号である顔の動きを使って表現する視覚言語である.その中でも,非手指信号の一部を担う「顔」は,重要な役割を担っている.普段私たちがコミュニケーションをする際に顔を見て話をするように,手話による会話でも手や指の動きよりも顔の表情や口の動きをみることが多い[1].

従来の手話認識モデルは,手の形と動きにのみ焦点を当てる認識技術が多いが,コミュニケーションをする上で,顔の表情は,感情を効率的に伝えられる場所である.よってここでは,手話を翻訳しテキスト化する際に,感情の付加を行うことで心情伝達の効率化を考えた.

## 2　目的

本研究では,手話を翻訳する際に顔の表情から感情を読み取り,テキストと顔文字による感情の付加を行い,心情の視認性の精度を計ることを目的とする.

## 3　関連研究

本研究の関連研究として,表情認識を用いたリアルタイム手話通訳システムがある.手話動作の認識において動作者の表情も重要なパラメータになることから,腕の動きと手の形,表情の３つの状態を認識し,それぞれ,Arms Motions, Hand States, Facial Expressions パラメータとして動作データの取得を行い,翻訳を行う.[2]

手話の認識と表情の認識を組み合わせた研究はあるが,感情に着目して翻訳を行う研究はない.

本研究では,表情認識を用いて手話動画から感情を抽出し,翻訳に追加することで心情の視認性の精度を計る.

## 4　手法

### 4.1　使用データ

本研究では,話者一人の3～10秒程度のあらかじめ翻訳された手話動画を使用する.動画をフレーム画像に変換し,最後のフレームのみ表情認識を行い感情を抽出する.出力された感情を翻訳されたテキストの後ろに英語表示と顔文字表示の二種類の動画を作成する.

心情伝達の効率化を計る評価方法としては,手話がわからない,２０代８人と５０代２人の合計１０人の健常者を対象とする.感情の英語表示と顔文字表示の動画を見てもらい,伝わりやすい・わからない・伝わりにくいの三段階で Web 上でアンケートを行う.

### 4.2　表情認識モデル

表情認識モデルとしてMulti-task Efficient-B2モデル[3]を使用した.このモデルは,マルチタスク学習とファインチューニングによって表情認識精度を向上させ、タスク間の共通点や相違点を利用しながらモデルの汎化性能も向上させる表情認識モデルである.図１にMulti-task EfficientNet-B2モデルの全体の流れを示す.

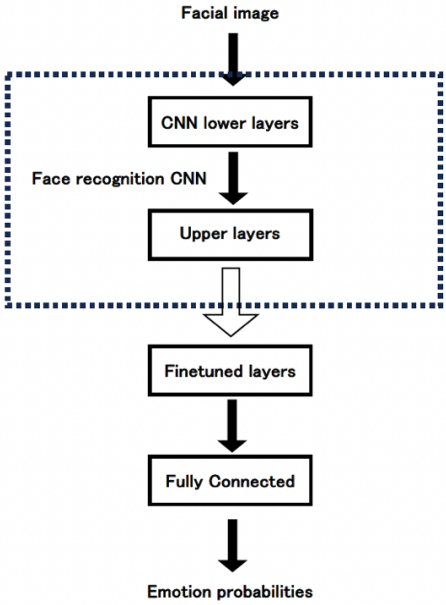


図１：Multi-task EfficientNet-B2 モデル

## 5　結果

表情認識のラベルは,　怒り(Anger)・軽蔑(Contempt)・嫌悪感(Disgust)・恐れ(Fear)・幸

福（Happiness）・普通（Neutral）・悲しみ(Sadness)・驚き(Surprise)の８クラス分類のモデルを使用した。表１にはそれぞれの顔文字ラベルを示す.

表１:顔文字ラベル

| | |
|---|---|
| 怒り | (ꐦ=　^　=) |
| 軽蔑 | (- _ -`) |
| 嫌悪感 | (- ^ -) |
| 恐れ | (;｜_｜;))) |
| 幸福 | (^ _ ^) |
| 普通 | (｜_｜) |
| 悲しみ | (;^;) |
| 驚き | Σ(｜_｜;) |

予め翻訳された７個手話動画（フジテレビ系「silent」）から,最後の表情認識結果のフレームを表情認識し, テキストと顔文字で表示した. 絵文字表記とテキスト表記のアンケート結果を表２,３に示す.

表２:テキスト表記

| | 伝わりやすい | わからない | 伝わりにくい |
|---|---|---|---|
| sumple1 | 9 | 1 | 0 |
| sumple2 | 10 | 0 | 0 |
| sumple3 | 10 | 0 | 0 |
| sumple4 | 10 | 0 | 0 |
| sumple5 | 8 | 2 | 0 |
| sumple6 | 3 | 7 | 0 |
| sumple7 | 1 | 7 | 2 |

表３:顔文字表記

| | 伝わりやすい | わからない | 伝わりにくい |
|---|---|---|---|
| sumple1 | 2 | 4 | 4 |
| sumple2 | 7 | 3 | 0 |
| sumple3 | 9 | 1 | 0 |
| sumple4 | 10 | 0 | 0 |
| sumple5 | 10 | 0 | 0 |
| sumple6 | 5 | 2 | 3 |
| sumple7 | 0 | 6 | 4 |

## ６ まとめ

本研究では,表情認識モデルを用いてテキスト翻訳された手話動画に感情の付加を行った.画像からの表情認識の精度は高いが,人によって感じた感情と出力された結果に違いがあった.アンケート結果により,顔文字よりテキストの方が伝わりやすいと答えた人数が多かった. sumple６,７は認識した結果と感じ取れる感情の違いがあり,テキスト表記ではわからないが多かったが, sumple６では,顔文字の方がテキストよりもわかりやすいと回答した人が多かった.全体的にテキストよりも顔文字がわかりにくいのは,ラベルの種類が関係していると考えられる.怒りや幸福,悲しみなど喜怒哀楽の感情ラベルは見やすいが,軽蔑や嫌悪感は伝わりにくい傾向が見受けられた.

今後の課題としては,sample データを増やしていくことと,出力される感情が一定の数値以下であれば,感情を普通のラベルにするにすることで,抽出される感情の誤差を減らしていきたい.また,手話認識モデルと合わせることで,手話を行う上で大切な表情の可視化を行い,翻訳精度の検証も行っていきたい.

## 参考文献

[1] 市川熹,長嶋裕二,寺内美奈 (2005), ”手話における「顔」の働き”,社団法人 情報処理学会 研究報告

[2] 眞田 慎, 岡田 志麻(2017), “表情認識を用いたリアルタイム手話翻訳システムの開発”,55Annual 巻 3PM-Abstract 号 p.224

[3] Andrey V. Savchenko(2021)”Facial expression and attributes recognition based on multi-task learning of lightweight neural networks”, Cornell University

[4] 池田尚志,松本忠博 (2011),”点字と手話と自然言語処理”, IEICE Fundamentals Review Vol.4 No.4

[5] 神田和幸,原大介,神谷昌明,木村勉,片岡由美子,”手話における文法、感情、意図の視覚化”, 可視化情報 Vol.23 Suppl. No.1(2003年7月)