

## マスクの紐をインタフェース化する手法 Turning Your Mask's Strap into an Input Interface

山本 匠<sup>1) 2)</sup> 杉浦裕太<sup>1) 3)</sup>  
Takumi Yamamoto Yuta Sugiura

### 1 はじめに

マスクは装着者の耳にかけることで口と鼻を覆い、風邪等の予防だけではなく防寒・保湿などの目的で利用されてきた。COVID-19 のパンデミック以降、マスクが生活必需品となり装着することが社会的に受け入れられたことで、ウェアラブルな情報入力装置として利用される可能性が高まっている。現在、スマートウォッチやイヤホンなどのウェアラブルデバイスへの入力は、タッチスクリーンや機器に触れるのが一般的である。しかしこれらのデバイスは小型化に伴いタッチ入力のための物理的な面積が減少しており、その結果入力可能な情報の種類が限定されるなどの問題がある。これを踏まえて、タッチ入力ではなくジェスチャや音声により入力する試みが行われている。

本研究では、マスクの紐を動かすことで情報を入力する手法を提案する。デバイスには、赤外光を照射し物体からの反射光を測定するフォトリフレクタが複数取り付けられており、紐の動きをセンシングし機械学習によって複数のジェスチャを識別することができる。4 人のユーザに対して 4 種類のジェスチャの識別精度を検証した結果、79.69%の精度を得ることができた。マスクを用いた入力インタフェース [1, 2] と比べ、マスク自体にセンサを張り付ける手間がいらぬという利点がある。

### 2 提案手法

本手法は、ハードウェアデバイスに備え付けられた光反射型センサがマスクの紐の動きを捉え、取得したセンサデータからジェスチャを識別する。以下では提案ジェスチャ、ハードウェア、ソフトウェアに分けて詳細に説明する。

#### 2.1 提案ジェスチャ

本研究で提案するジェスチャの種類を図 1 に示す。本研究では、マスクの上の紐を上上げる動作 (Upper-Up)、下げる動作 (Upper-Down)、マスクの下の紐を上げる動作 (Lower-Up)、下げる動作 (Lower-Down) の 4 種類のジェスチャを用いる。

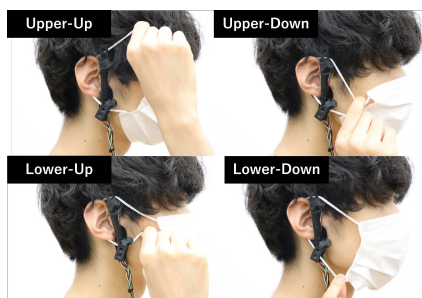


図 1 本研究で使用した 4 種類のマスクの紐を動かすジェスチャ

- 1) 慶應義塾大学
- 2) imuka06x17@keio.jp
- 3) sugiura@keio.jp

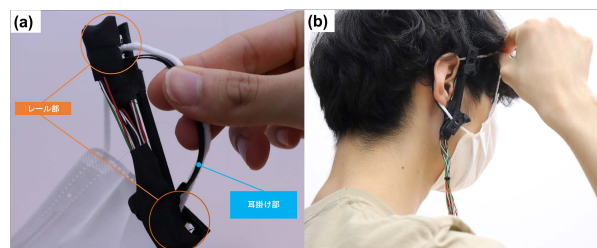


図 2 ハードウェア。(a) レール部と耳掛け部、レール部にマスクをかけている様子 (b) ユーザがハードウェアを装着した様子

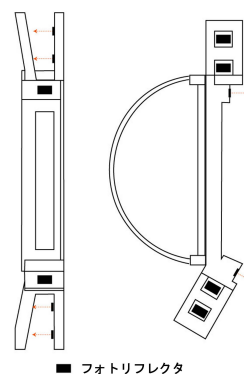


図 3 ハードウェアにおけるセンサ位置と向き

#### 2.2 ハードウェア

ユーザはハードウェアを耳にかけることで装着する。ハードウェアには耳にかけることができる部分 (耳掛け部) に加えレールのような構造をしている部分 (レール部) があり、レール部にマスクの紐をかけ (図 2(a)), マスクの紐と耳掛け部を装着することでユーザはマスクとハードウェアデバイスを耳に固定する (図 2(b))。

ハードウェアには反射型光センサであるフォトリフレクタ (SG-105, コーデンシ) が 6 つ付けられている。6 つのフォトリフレクタのうちの 4 つはレールに埋め込まれており、頭の中心部から耳側に向いている。残りの 2 つはレールの出口に固定されており、視線と同一の方向を向いている (図 3)。6 つのフォトリフレクタは有線でマイクロコントローラ (Arduino Pro mini 3.3V) を含む回路に接続され、マイクロコントローラと PC が有線ケーブルでシリアル通信することで PC 上でセンサ値を取得した。

#### 2.3 ソフトウェア

ソフトウェアは得られたセンサ値から機械学習によってジェスチャを識別する。

センサデータから識別するまでの流れを図 4(a) に示す。まず取得したジェスチャごとのフォトリフレクタのセンサ値を取得した。通信速度は約 35[fps] であり、実験者が開始を入力してから 5 秒間のデータを取得したため、合計 175 フレームを取得した。取得したデータに対

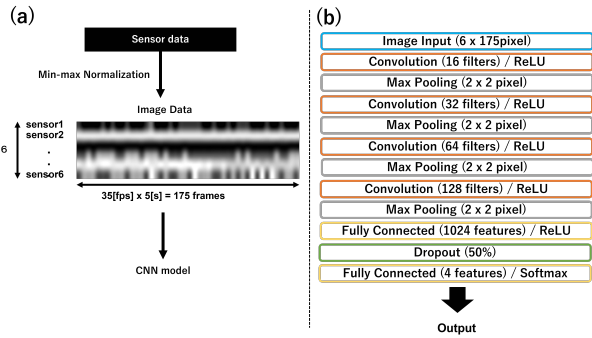


図 4 (a) センサデータの前処理のプロセス (b) CNN のアーキテクチャ

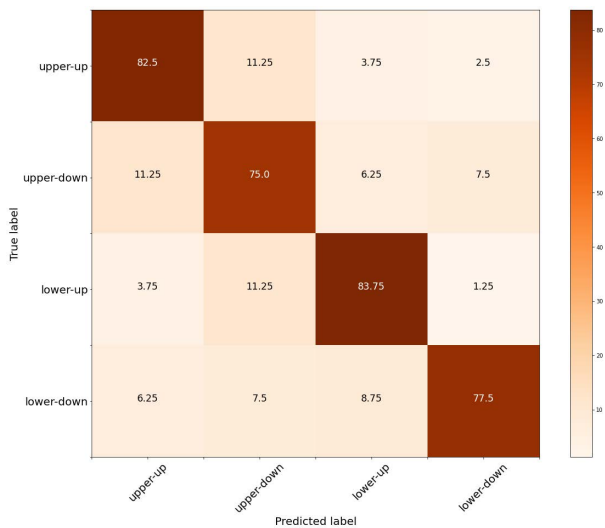


図 5 識別結果

して最大値 255, 最小値 0 となるように正規化し, センサ値を 8 ビットのグレースケール画像に変換した. 6 つのセンサデータを合わせて 1 枚の画像にし,  $6 \times 175$  のグレースケール画像を 1 枚の入力画像とした.

ジェスチャ識別には Convolutional Neural Networks (以下, CNN) を用いた. CNN のアーキテクチャは図 4(b) に示す通りであり, 4 種類のジェスチャを識別する. 実装には python を用いた.

### 3 実験

#### 3.1 概要

本研究では, 以上に示した 4 種類のジェスチャの識別精度を評価する実験を行った. 実験協力者は男性 4 名 (平均 21.75 歳) であり全員右利きであった. 実験協力者は, 本研究で提案する耳掛け型装置とマスクを右耳に装着し, 最初にジェスチャに関する説明を受け各ジェスチャを数回練習した後, 4 種類のジェスチャを各 20 回実施し, 合計 320 枚の画像データとした, マスクは同じ種類の白色のマスクを着用した.

#### 3.2 結果

得られた画像データに対して, 10 分割交差検証により精度を評価した. 実験結果を図 5 に示す. 各ジェスチャの平均識別精度は 79.69% であった. 各ジェスチャごと

の識別精度は, Upper-Up が 82.50%, Upper-Down が 75.00%, Lower-Up が 83.75%, Lower-Down が 77.50% であった.

### 4 制約と議論

本研究で用いた手法において, 黒い紐を有するマスクを使用することは困難である. フォトリフレクタは赤外光を発光しその赤外光の反射光を受光することでセンシングを行うが, マスクの紐が黒い場合は赤外光が反射せず, マスクの紐を動かすことによるセンサ値の変化が生じないためであると考えられる.

精度評価に関しては 4 人の被験者に対して平均識別精度が 79.69% であった. 今後フォトリフレクタの個数を増やすなどで, これらの精度を向上させていくことが可能であると考えられる. また, 今後はさらに多くの被験者による精度評価を行う必要もある.

本研究で使用したジェスチャは研究者自身が定義したものであったが, マスクの紐を動かすことをジェスチャとする上で, どのようなジェスチャがユーザに好まれるのかを探索する必要がある. ユーザ自身にジェスチャを考案してもらい, ユーザ定義型のジェスチャセットを決定するためのユーザ調査をすることで, ユーザが使いやすいジェスチャを定義することができると考えられる. また, 使用したジェスチャは 4 種類に留まっているが, 今後はジェスチャの種類を増やしていき多様な入力を実現できると考えられる.

### 5 終わりに

本研究では, 光反射型センサであるフォトリフレクタが埋め込まれた耳掛け型デバイスを用いてマスクの紐の動きをセンシングし, CNN を用いてジェスチャ識別を行った. 4 人の被験者に対して精度評価実験を行った結果, 平均識別精度は 79.69% であった. 今後は, さらにフォトリフレクタを増やすなどのハードウェア面での改良を加えつつ, ユーザ定義型のジェスチャセットの決定, さらなる識別精度の向上とより多様なジェスチャの識別を試みる.

#### 謝辞

本研究は JST 科研費 JPMJPR2134 の支援を受けたものです. また深層学習手法をご教授いただいた松尾佳奈氏に感謝いたします.

#### 参考文献

- [1] Yusuke Kunimi, Masa Ogata, Hiroataka Hiraki, Motoshi Itagaki, Shusuke Kanazawa, and Masaaki Mochimaru. 2022. E-MASK: A Mask-Shaped Interface for Silent Speech Interaction with Flexible Strain Sensors. In *Augmented Humans 2022 (AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 26–34. <https://doi.org/10.1145/3519391.3519399>
- [2] Suzuki, Y., Sekimori, K., Yamato, Y., Yamasaki, Y., Shizuki, B., Takahashi, S. (2020). A Mouth Gesture Interface Featuring a Mutual-Capacitance Sensor Embedded in a Surgical Mask. In: Kurosu, M. (eds) *Human-Computer Interaction. Multimodal and Natural Interaction*. HCHI 2020. Lecture Notes in Computer Science(), vol 12182. Springer, Cham. [https://doi.org/10.1007/978-3-030-49062-1\\_10](https://doi.org/10.1007/978-3-030-49062-1_10)