

手の認識空間を拡張するヘッドマウントディスプレイ 一人称視点手話認識システムの提案

乾竜躍^{†1} 加藤恒夫^{†1} 田村晃裕^{†1}

概要: 聴覚障害者と手話を知らない健常者とのリアルタイムコミュニケーションを可能にするため、筆者らはヘッドマウントディスプレイ(HMD)上で一人称手話認識と音声合成・認識技術を組み合わせたシステムの構築を目指している。しかし、HMDの手の認識空間はユーザの前方に限定されており、頭上や耳付近に手指を近づける手話の認識は難しい。そこで、HMDに手話者のアバターを投影し、アバターを用いて手話の認識空間を拡張する手法を提案する。Hololens2を用いてプロトタイプを試作し、こめかみを指さす手話や顎を触る手話を認識できることを確認した。

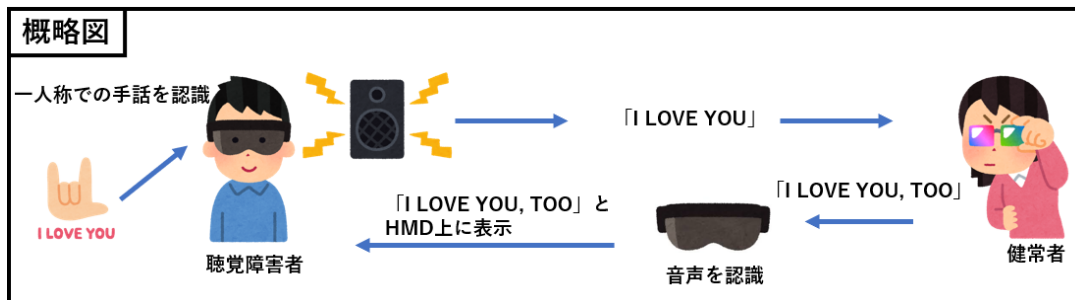


図1 システム全体図

1. はじめに

聴覚障害者にとって手話は大切なコミュニケーションの手段である。しかし、手話を知らない健常者とのコミュニケーションは簡単ではない。筆者らは、手話認識と音声合成・認識技術を用いて聴覚障害者と健常者のリアルタイムコミュニケーションを可能にする図1に示すシステムの構築を目指している。

手話とは主に、顔の表情、口の動き、手の動きなどの上半身の動きで表現され、音声からは独立した言語である。様々な手話の中でも日本手話は国内で主に聴覚障害者がコミュニケーションに利用している、手話認識は古くから研究され、一般的には手話者を正面からカメラで捉える。しかし、この方法ではカメラの設置や撮影者が必要になる。これに対して Fujimoto らはヘッドマウントディスプレイ(HMD)を用いる一人称視点での手話認識システムを提案した[1]。Hololens2の手の姿勢推定機能を活用して指文字の認識を可能にした。しかし、認識できる手話はHMDのカメラと深度センサで認識できる範囲に限定され、手がHMDの認識範囲からはみ出す手話は認識することができない。

そこで、今回HMDの手の認識範囲の制約を受けない一人称手話認識システムを提案する。手話の中には自身の頭を指さすなどの手話があるが、手話者自身のアバターをディスプレイに表示し、それを使って手話認識を行う方法を提案する。ディスプレイ上のアバターの頭を指さすことで手話者自身の頭を指さしているように認識させる。

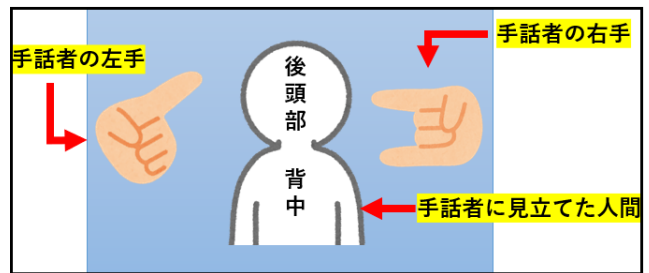


図2 手話者のHMDに映るアバターと手のイメージ

2. 手話者自身のアバターを用いて認識対象空間を拡張する一人称手話認識

手話者のHMDに映る自身のアバターと手のイメージを図2に示す。手話者自身と同じ向きを向いた半透明のアバターが両手で触れる位置に投影される。手話者からは正面にアバターの後頭部が見える。アバターの頭頂部、左右のこめかみ、両耳、鼻、顎の位置にタッチするためのポイントを用意しておく。このポイントに手指で触れることで、アバターを利用する手話の認識が開始される。

実際には、アバターを用いない通常モードとアバターを用いる拡張モードの2種類の動作モードを切り替える。通常モードではアバターは関係なく手話の認識を行う。アバター上のポイントを手指でタッチすると、拡張した空間で手話認識が行われる。拡張モードで手話が認識されると、通常モードに戻る。いずれの動作モードでも一定時間の手指特徴点データが蓄積されると、手話認識が行われる。

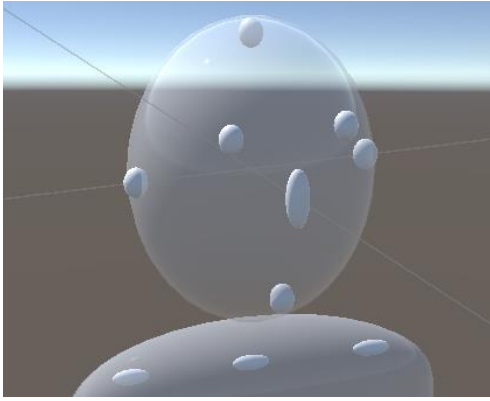


図 3 unity 上に作成したアバター



図 4 「いいよ」認識時の HMD 画面

3. プロトタイピング

3.1 システム構成

HMD として Hololens2 を用いた。Hololens2 の手指姿勢推定機能を用いて、両手指の座標を取得することができる。手話の認識は PC 上で行う。Hololens2 と PC は WebSocket で接続し、Hololens2 からは手指の関節の座標情報を送信し、PC から手話の識別結果を受信する。

3.2 手話認識

手話認識は、1.5 秒間（間引きした 45 フレーム）の両手指 25 点の 3 次元座標を用いて行う。先ず HMD 画面上の位置依存性とサイズ依存性をなくすために、フレームごとに各点の座標から手首の座標を減算し、さらに x 軸、y 軸、z 軸それぞれの両手の範囲で正規化する。このベクトルを用いて、予め記録しておいた各手話クラスのプロトタイプとコサイン類似度を測る。コサイン類似度に対して、各手話クラス個別に設定した閾値で検出判定を行う。手話が検出されなかった場合は特徴点座標にシフトウィンドウを適用する。また、複数の手話クラスが同時に検出された場合には、最後に判定された手話を出力する。

手指の座標情報を送信する合図を、左手で握りこぶしを作ったときとした。合図をしている間、手指の座標が PC に送信される。そのため、今回は右手による手話に限定されることになった。

具体的には以下の手話[2]をシステムに登録した。

- 「す」、「な」、「ね」、「ふ」、「へ」、「ま」を除く五十音の指文字：Hololens2 の手指の姿勢推定機能は手を下に向けたときに認識できないため、手を下に向けた上記の 6 文字を認識対象から除外した。
- 「いいよ」：立てた右手小指の指先で顎を 2 回触る。
- 「思う」：人差し指の指先をこめかみにあてる。
- 「印象」：右手こぶしをこめかみにあてる
- 「感想」：右手人差し指の指先をこめかみにあて、5 指を上に向けて頭の脇から右を少し上げる。

- 「イメージする」：右手人差し指の指先をこめかみにあて、指を開いた右手のひらを前に向けて円を描くように回す。
- 「思い出」：右手人差し指の指先をこめかみにあて、頭の脇で右手の指を揺らせながら、斜め後方へ動かす。
- 「アイデア」：こめかみにあてた右手人差し指を斜め上へ上げる。
- 「意見」：立てた右手小指のこめかみにあて、回しながら前へ出す。
- 削除：右手でチョキのポーズを作る。

空間拡張モードにおける HMD 画面の例を図 4 に示す。画面 1 行目の「入力文字：」以降に認識された手話、2 行目には削除された文字数、3 行目にはアバターにタッチした指とポイントが表示されている。

4. おわりに

場所を選ばず聴覚障害者と健常者がリアルタイムコミュニケーションを行えるシステムの構築を目指し、HMD を用いた一人称手話認識において手話の認識空間を拡張する方法を提案し、Hololens2 を用いてプロトタイプを試作した。HMD 画面上に現れる手話者のアバターを用いることで、こめかみを指さす手話や小指で顎を触る手話を認識できることを確認した。

現在の試作では、手話認識のトリガを左手のポーズで与えているため、両手の加速度などをトリガとするシステムを開発するとともに、手話認識の精度評価を予定している。

参考文献

- [1] T. Fujimoto, T. Kawamura, K. Zempo, S. Puentes, "First-person view hand posture estimation and fingerspelling recognition using HoloLens", Proc. IEEE GCCE 2022, pp.323-327, 2022.
- [2] 一般財団法人全日本ろうあ連盟：わたしたちの手話 学習辞典 1, 2015.