

# 装着型機器を用いた指文字の 音声翻訳による対人コミュニケーション支援

河原 圭佑<sup>†</sup> 鈴木 健嗣<sup>‡</sup>

筑波大学理工学群工学システム学類<sup>†</sup>

筑波大学システム情報系/JST<sup>‡</sup>

## 1 はじめに

現在、国内に聴覚障がい者は約 45 万人いる。その一部は自身の発話に対し聴覚を通じてフィードバックを得ることができないため、発話が困難である。そのため、聴覚障がい者が言語を用いてコミュニケーションを実現する手段である手話が広く利用されている。しかしながら、一般に手話の習得は容易でなく、手話を習得していない相手とのコミュニケーションは難しい。そこで機械により手話を認識する試みが行われてきた[1-3]。既存の手法は手話動作を取得するために、手にセンサを装着する必要や、カメラを机などに固定する必要があるため、日常的に使用することは容易ではない。そこで我々は、屋外で使用可能とすることを考慮し、装着性、携帯性に優れた聴覚障がい者のための対話支援機器を提案する。本稿では 3D モーションセンサを用いたネックレス型対話支援機器の試作について報告する。

## 2 提案手法

本提案のような装着型機器を利用する例として、帽子のつばに設置されたカメラにより手話を認識する研究がある。Starnier ら[4]は、連続した手話の動作に対し、隠れマルコフモデルを用い認識を行っている。ここでは指文字の認識の実現により、さらに利便性が向上するだろうと述べられている。

また近年、3D モーションセンサの精度向上により、画像認識では困難である細かい手指動作の認識が可能となってきている。そこで前述した指文字の認識が技術的に実現可能であるかを検証する。提案手法では、まず 3D モーションセンサを用い、手形状推定を行い、指の関節角を特徴量とし実時間で指文字の認識を行う。認識した指文字動作に基づき文字列を生成し首元のスピーカにより音声として提示することで、支援機を通じ、手話や指文字を理解しない他者との音声対話を支援することを目指す。システムの概要図を図 1 に示す。

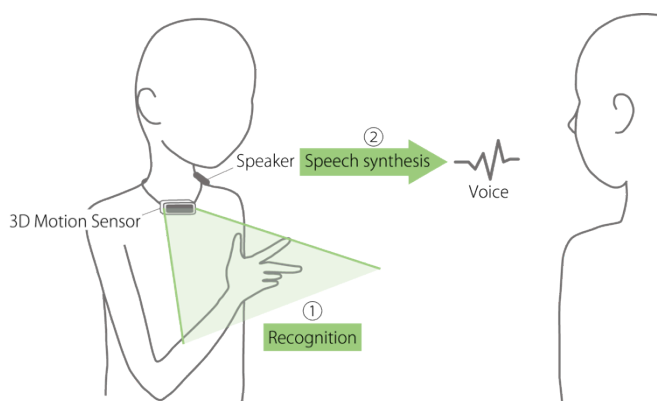


図1 システムの概要図



図2 試作機



図3 装着した様子

## 3 システム構成

### 3.1 ハードウェア

3D モーションセンサには米 Leap 社の Leap Motion を用いる。本センサは、広角レンズを用いており認識領域が広く、センサ自体が W76×H13×W13[mm]と小型であるため、装着型機器を構成するために適切であるといえる。

制作した試作機を図 2 に、実際に装着した様子を図 3 に示す。

現在の試作機では、3D モーションセンサを手提げ型のバッグ内のコンピュータに接続し処理を行わせたが、組込コンピュータを利用することで小型化することは可能であると考えられる。

### Voice synthesis from fingerspelling for hearing impaired by a wearable device

Keisuke Kawahara<sup>†</sup> and Kenji Suzuki<sup>‡</sup>

<sup>†</sup>College of Engineering Systems, University of Tsukuba, ktansai.keisuke@gmail.com

<sup>‡</sup>Faculty of Engineering, Systems and Information, University of Tsukuba/JST, kenji@ieee.org

### 3.2 ソフトウェア

ここでは指文字動作の認識と、発話音声列の生成を行う。これは JavaScript のコマンドラインランタイムである Node.js 上で動作する。

### 3.3 認識アルゴリズム

事前に認識する対象の指文字動作を示し、関節位置や関節角からなる動作テンプレートをデータベースに記録する。認識率を向上させるため、まず指文字動作における手の姿勢に着目し、手の平の向きから事前に6グループに分類するこの分類の例を図4に示す。

認識時はまずセンサにて手の姿勢を取得し、入力かどのグループに属するかを手の平の向きのみで判定する。その後一致したグループ内のテンプレートに対し、各指の関節角を特徴量とし k-NN 法を適用する。安定性を確保するため、認識のフレームレートを 15fps とし、0.5 秒以上同じ結果が得られた時に指文字と認識し、音声合成を行う。

また、今回は日本語の指文字 46 パターンのうち、動きのない 41 パターンを対象とした。

## 4 システム評価実験

システムによる指文字動作の認識精度を検証するため、評価実験を行う。1名の被験者に対し41パターンの各指文字動作を15回ずつ行ってもらい、正しく認識する回数を記録する。全体平均で70.1%の認識精度が得られた。認識速度は速いもので0.5秒、遅いものでも1秒以内に認識することが確認された。また、前述した手の姿勢による分類において、グループAに属するパターンは特に対話相手に見せることを前提とした指文字であり、指文字動作を行う本人からは指の形状を確認することが難しく、認識精度が平均的に低かったことも確認された。



図4 手の平の向きに基づく指文字の分類 (中央法規出版“新・手話辞典”より画像を引用)

## 5 アンケート調査

日常的に手話/指文字を使用している聴覚障がい者(1名)に協力してもらい事前に書面にて同意を得た後、開発したシステムを使用する。実験後に自由回答形式のアンケートにより、コメントを得た。ここでは「精度がもう少し向上するとよい」、「手話にも個人差があるので、様々な手話に対応したほうが良い」、「認識された文字を確認する必要がある」などの意見が得た。また「発話することのできない聴覚障がい者には良いと思う」などの肯定的な意見も得られた。

## 6 まとめと今後の展望

本稿では対人親和性を考慮したネックレス型対話支援機器を提案し、提案手法の試作に関して報告した。本研究は、手話の認識、音声合成に関するものであるが、聴覚障がい者のコミュニケーションを考える上では、逆に健聴者の発話を聴覚障がい者に伝えることも必要である。これは、音声認識により発話を文字列にし、メガネ型機器などの視覚提示機器により提示することで、聴覚障がい者が他者の発話を理解することを支援できると考えている。

また現在は認識する毎に1音ずつ音声合成を行っているが、単語単位で音声合成を行うことで、より伝わりやすくなると考えている。この際、単語中の指文字の一部を誤認識した場合でも、隠れマルコフモデルやDPマッチングなどの手法を用い、補完することが可能であるため、より自然かつ円滑な対話支援に寄与するものと考えている。

**謝辞** 本研究は筑波大学先導的研究者体験プログラムの支援を受けて行ったものである。

## 参考文献

- [1] T. Starner, “Visual Recognition of American Sign Language Using Hidden Markov Models”, Massachusetts Inst of Tech, Dept of Brain and Cognitive Sciences, 1995
- [2] Kramer, J.P., Lindener, P., and George, W.R. “Communication system for deaf, deaf-blind, or non-vocal individuals using instrumented glove”, U.S. Patent No. 5,047,952, 1991
- [3] 渡辺 賢, 岩井 儀雄, 八木 康史, 谷内田 正彦 “カラーグローブを用いた指文字の認識” 電子情報通信学会論文誌. (D-II), vol. J80-D2, p44-45, 1997
- [4] T. Starner, J. Weaver, and A. Pentland, “Real-time American Sign Language Recognition Using Desk and Wearable Computer Based Video”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, No. 12, p. 1371-1375, 1998.