

スマートグラスにおけるARを用いた 音声コミュニケーション支援手法の提案

阪田 大輔¹ 森田 健太郎¹ 佐藤 健哉²

概要:

近年, 音声認識技術の研究が盛んに行われている。音声認識は日常生活でも用いられ, その一つにコミュニケーション支援が挙げられる。音声認識を用いたコミュニケーション支援手法は, 会話内容をパソコンや携帯端末に表示するものがあり, 聴覚障がい者が視覚的に情報を取得できるようになるが, 使用場所や動作が制限されるという問題がある。そこで本研究では, スマートグラスとジャイロセンサを用いることで, 制限を払拭した円滑でバリアフリーなコミュニケーション支援手法を提案し, 実装, 評価を行った。本論文では, 既存手法と提案手法の使用環境の違いと操作性を比較し, 提案手法が有用であることを示した。

A Proposal for Supporting Speech Communication with AR on Smart Glasses

DAISUKE SAKATA¹ KENTARO MORITA¹ KENYA SATO²

1. はじめに

1.1 背景

1990年代から実用化に向けて研究されてきた音声認識が, 近年になって発展を遂げている。従来では, 単語のみの検出や認識精度が低く, 実用するには困難であった。しかし, ネットワークの発展や言語モデルの向上するにつれ, 日常生活への導入が容易になり, 様々な技術に音声認識が用いられるようになった。身近なものであれば, SiriやGoogle音声検索などが挙げられる。またロボットとの対話にも音声認識が用いられており, 会話をしたり指示したりすることができる。

最近では音声認識がコミュニケーション支援にも用いられるようになってきている [1]。聴覚障がい者や高齢者のような耳の不自由な人たちが会話をする際に, 音声認識を用いて相手の発言をテキスト化することで, 視覚的に情報を手に入れることが可能になる。テキスト化した文字を表示するデバイスにパソコンやスマートフォンなどの携帯端末を用いるため, 音声コミュニケーション支援手法を簡単

に使用でき, 自分の発言をタイピングすることもできる。また, 手話を覚えたり, 筆談器などの準備をしたりする必要がなくなるという利点がある。

しかし, 従来の手法であれば会話をする環境によっては使用できない, あるいは会話以外の動作に制限が生じるという問題がある。例えば, パソコンに会話内容を表示する場合, 机やテーブルなどのパソコンを設置できる環境が必要となる。また, 携帯端末を用いた場合, パソコンを用いたときのような設置場所の制限はないが, 端末を持つことによって手がふさがることによる, 動作制限が生じる。

本論文では, スマートグラスとジャイロセンサを用いることで, テキストを拡張現実 (Augmented Reality: AR) として扱うことで, 既存手法で生じる制限の払拭を図り, 聴覚障がい者や耳の不自由な高齢者が円滑に会話ができるバリアフリーなコミュニケーション支援手法を提案する。

1.2 目的

本研究の目的は, 従来の音声コミュニケーション支援手法における使用場所の制限, および動作制限を払拭することである。音声認識で得たテキストデータをスマートグラス上にARとして表示し, ジャイロセンサを用いて表示す

¹ 同志社大学大学院 理工学研究科 情報工学専攻

² 同志社大学大学院 理工学研究科

る会話内容を取得することで、円滑で制限のないバリアフリーなコミュニケーション支援手法を実現する。

2. 既存手法

2.1 音声認識

音声認識とは、人間の声などをコンピュータに認識させることであり、話し言葉を文字列に変換したり、音声から特徴を算出し話し手を識別する技術である。人の音声波形を音素という基本単位波形に分割し、分析した音素の組み合わせから、発声された言葉をコンピュータ内で再構築する。現在の音声認識には隠れマルコフモデルとNグラム法が用いられている。音声データが必要となるためデータ量が膨大になるが、インターネットの発達によって、クラウド上で処理をすることで、処理能力が高くないパソコンや携帯端末でもリアルタイムで音声認識をすることができる。音声認識を用いて発言内容をテキストに変換し、そのテキストをコンピュータが分析することで、人間とロボットの音声対話が可能になる。

2.2 音声コミュニケーション支援

音声コミュニケーション支援とは、音声認識や合成音声を用いて耳が不自由な人のコミュニケーションを支援したり、声を発することが困難な人のコミュニケーション支援[2]を行うことである。本論文では会話内容を音声認識によってテキスト変換し、視覚的に情報を取得できるものを指す。

この支援方法は、主に聴覚障がい者や耳の不自由な高齢者の使用を意図しており、新しく手話を覚える必要がなく、また筆談器を用いて字を書く手間を省くことができる。テキスト変換した会話内容はパソコンや、スマートフォンなどの携帯端末に表示する。以下に、パソコン、およびスマートフォンを用いた音声コミュニケーション支援手法の例を挙げる。

2.3 LiveTalk

LiveTalk[3]は、聴覚障がい者参加型コミュニケーションツールである。

会議や打ち合わせなどの複数人が情報を共有する場で、発言者の発言を音声認識し、テキストに変換後、複数のデバイスに出力する。参加している全員がリアルタイムで情報を共有することができる。

音声認識による入力だけでなく、キーボード入力やスタンピング入力、定型文入力が可能である。

本論文では、パソコンを用いた音声コミュニケーション支援手法として扱う。

2.4 こえとら

こえとら[4]は、音声認識技術や音声合成技術を用いた、

聴覚障がい者と健聴者間のコミュニケーションを支援するアプリケーションである。

アプリがインストールされたスマートフォンを相手に渡し、発言内容を音声認識し、テキストを画面に表示することでコミュニケーションをとることができる。

本論文では、携帯端末を用いた音声コミュニケーション支援手法として扱う。

2.5 既存手法の問題点

音声認識を用いたコミュニケーション支援のソフトウェアやアプリケーションは上述したもの以外にもいくつか存在する。それらの手法は一般的に、パソコンとスマートフォンなどの携帯端末に音声認識結果を表示するものとなっている。しかし、テキストを表示するデバイスによっては使用できない環境がある。

パソコンを用いた場合、パソコンを設置する必要があるため、机などのある場所での使用に限られてくる。一方、携帯端末に表示した場合、手で持つことができるため、使用する場所の制限は生じない。しかし、デバイスを持つことで、食事などの手を使った動作が難しくなり、動作制限が生じる。また、歩行中の会話においては、歩きスマホによる危険も生じてくる[5]。

既存手法は、使用目的が限定されているために、使用場所・動作に制限が生じると考えられる。聴覚障がい者や高齢者が、日常生活でコミュニケーションを取るためには、これらの制限を払拭する必要がある。

3. 提案手法

3.1 概要

提案手法では、発言を音声認識を用いてテキスト化し、得られたテキストをスマートグラスの画面にARとして表示する。ARとして表示されたテキストをジャイロセンサを用いて操作し、会話履歴や会話内容を取得することで、円滑で制限のないバリアフリーなコミュニケーション支援

手順3：画面にARとしてテキスト表示

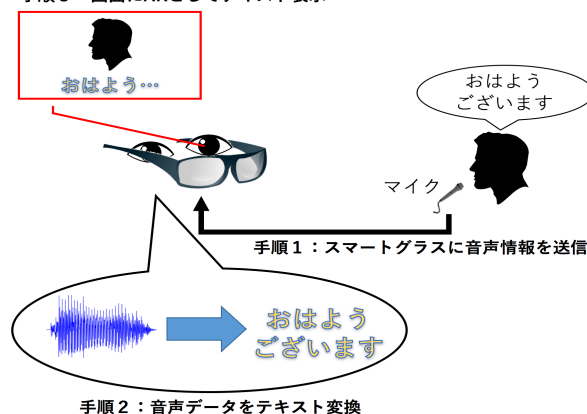


図1 提案手法の動作手順

手法の実現を図る。

図 1 に提案手法の動作手順を示す。提案手法では、マイクから得た発言内容をウェアラブル端末であるスマートグラスの画面に、AR として表示する。AR をスマートグラスの画面に表示することにより、使用場所の制限、動作制限が改善され、聴覚障がい者がより様々な状況下でコミュニケーションをとることができるようになる。

3.2 構成

提案手法における構成を以下に示す。

- スマートグラス
音声認識してテキスト化した発言内容を表示する。
- マイク
発言を取得するために用いる。
- ジャイロセンサ
スマートグラスの移動方向・移動量を計測する。今回はスマートグラスに内蔵されているジャイロセンサを用いる。
- メモリ
音声認識したテキストをテキストファイルとしてメモリに保存しておく。

3.3 音声認識結果の保存

マイクから受信した発言を音声認識する。音声認識は、会話が発せられてから終了するまでを一区切りとし、音声認識することで得られたテキストは、一回の認識毎にファイルに逐次保存していく。各テキストは一行で記述するために、新しく音声認識されると改行を挿入する。図 2 は保存されたファイルの一例を示す。

3.4 表示文字の取得

3.4.1 ジャイロセンサの設定

画面に AR としてテキストを表示し、操作にはジャイロセンサを用いる。ジャイロセンサによって、移動時の角速

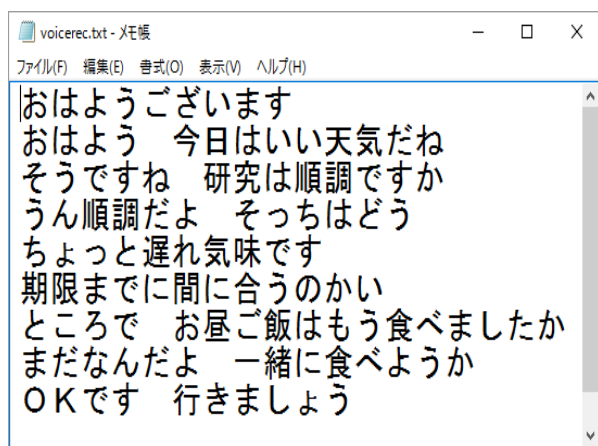


図 2 音声認識結果が保存されたファイルの一例

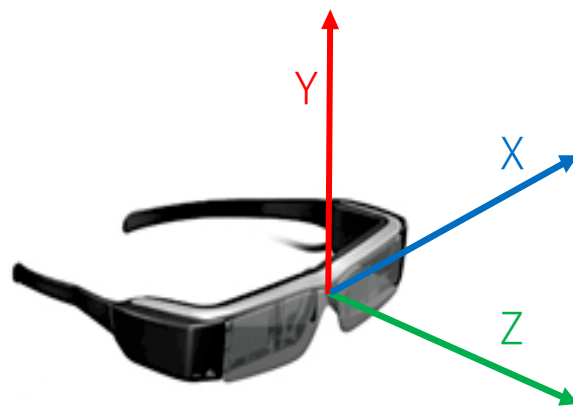


図 3 ジャイロセンサの方向軸設定

度が求まる。ジャイロセンサの値から首の動き、つまりスマートグラスの動きを求めることで表示文字の取得を行う。図 3 にジャイロセンサの方向軸の設定を示す。ジャイロセンサはスマートグラスの左右の方向に X 軸，上下方向に Y 軸，前後方向に Z 軸を設定する。提案手法において、以下のようにジャイロセンサを利用する。

- X 軸中心の移動
上下の文字を取得
- Y 軸中心の移動
左右の文字を取得

3.5 ジャイロセンサによる表示文字の取得

首を動かす際に動いたスマートグラスの移動方向・移動量をジャイロセンサの値から算出する。図 4 は画面の移動時の概要を示す。実際に見える画面までの距離を r とする。以下の手順でジャイロセンサを用いて画面の移動量 D_x ， D_y を算出する。

- (1) ジャイロセンサから移動時の角速度 ω_x ， ω_y を取得する
- (2) ω_x ， ω_y と移動時間 t から移動角度 θ_x ， θ_y を求める

$$\begin{aligned}\theta_x &= \omega_x t \\ \theta_y &= \omega_y t\end{aligned}\quad (1)$$

- (3) 画面までの距離 r と式 (3.1) から、画面の移動量 D_x ， D_y を算出する

$$\begin{aligned}D_x &= r \tan \theta_x \\ D_y &= r \tan \theta_y\end{aligned}\quad (2)$$

式 (3.2) によって得られた画面の移動量 D_x ， D_y を加算し、一定量 D を満たすと、対応する移動方向にあるテキストをファイルから読み込み画面に表示する。

図 2 のようなテキストが保存されているときに、どのように表示文字が遷移するかを図 4 に表す。図中の X 軸，Y 軸はジャイロセンサの移動軸を表している。まず正面を向いている際には図中の黒色の四角内のテキストがスマート

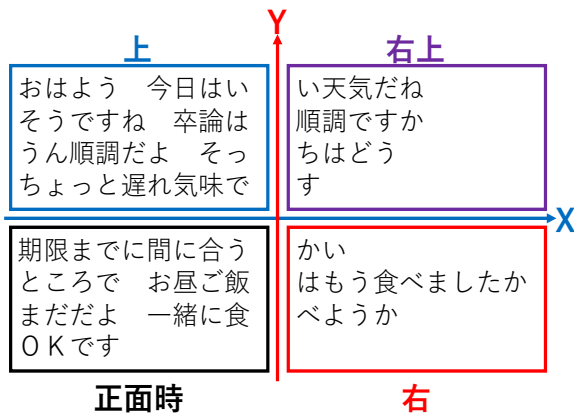


図 4 表示文字取得の一例

グラスに表示されている．このときに上にある文字を見るために，首を上へ傾けると X 軸中心の移動をジャイロセンサが検出し，式 (3.1)，および式 (3.2) から得られた移動量をもとに上にある文字を表示する．図中では，青色の四角内のテキストである．一方，正面時から右を向くと，Y 軸中心の移動をジャイロセンサが検知し，同様にして表示文字を取得する．図中における赤色の四角内のテキストである．また，上と右を同時に，つまり右上を見た際には，両方向のジャイロセンサが反応し，図中の紫色の四角内のテキストが表示されるようになる．

4. 実装

4.1 実装環境

式 (3.1) における θ_x , θ_y は十分小さいため，式 (3.2) を $D_x = r\theta_x$, $D_y = r\theta_y$ と近似し， $r = 1$, $D = 1$ とした．また，ジャイロセンサの値に閾値を与え， $\omega_x < 1$, $\omega_y < 1$ であれば移動量を増加するようにした．

次に，表 1 に提案手法の実装環境を示す．今回の提案手法を実装するにあたり，スマートグラスは EPSON 社の MOVERIO BT-200 を用いた．マイクは，ヘッドセットである plantronics 社の M70 を用いた．音声認識は，Android Speech Recognizer を用い，インターネット経由で Google のサーバ上で音声認識を行った．Java 言語を用いて，音声認識したテキストを AR 表示するアプリケーションを Android Studio で開発し，MOVERIO にインストールした．図 5 は使用した MOVERIO と M70 である．

表 1 実装環境

スマートグラス	EPSON MOVERIO BT-200
マイク	plantronics M70
音声認識	Android Speech Recognizer
OS	Android 4.0.4
CPU	Texas Instruments OMAP4460
RAM	1GB
プログラミング言語	Java



図 5 実装で使用した機材



図 6 機材を用いた会話

4.2 会話方法

図 6 は機材を用いて健聴者二人が会話を行っている様子である．提案手法の実装ではマイクにヘッドセットを用いたため，健聴者側にヘッドセットの装着をしてもらい会話を行った．スマートグラスの画面を写真に取ることが困難なため，アプリケーションをスマートフォンに導入し，会話した際に表示されるイメージ画面の正面時を図 7 に，右に向けた際の表示画面を図 8 に示す．実際にスマートグラスで見えるテキストは背景が透過されて見える．どの発言がいつの発言かを確認しやすくするため，各発言に番号を付与して表示した．

5. 評価

5.1 使用制限における評価方法

提案手法と既存手法の比較評価を行うにあたり，使用場所が変わっても会話可能であるかどうか，また，別の動作をしながら会話が可能であるかどうかを比較する．既存手法は，パソコンを用いたコミュニケーション支援手法と携帯端末を用いたコミュニケーション支援手法の 2 種類であり，第 2 章で取り上げた LiveTalk をパソコンを用いたコ

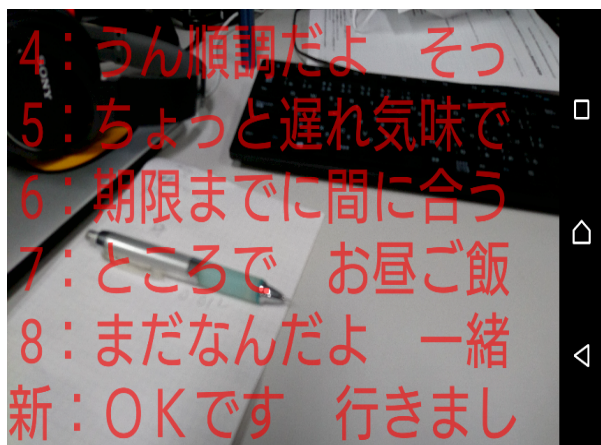


図 7 スマートフォンで実装した際の実行画面（正面）

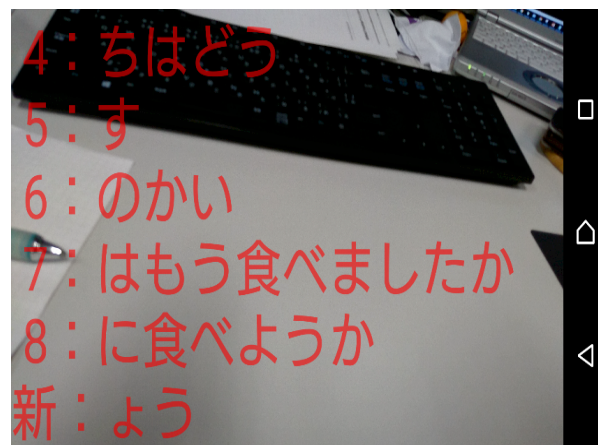


図 8 スマートフォンで実装した際の実行画面（右移動時）

コミュニケーション支援手法に、こえとらを携帯端末を用いたコミュニケーション支援手法に採用した。

使用場所の評価として、パソコンが使用できない場所でも提案手法が使用可能であるかどうかを評価するため、研究室と同志社大学の正門前で会話を行い、提案手法が使用可能であるかどうかを調べた。

動作制限における評価は、歩行中と食事の二つの動作で評価した。

5.2 評価結果

表 2 に評価結果を示す。会話が不可能であった場合、制限があるものとみなして結果を出した。

LiveTalk は、研究室ではパソコンをテーブルに置くことができたため会話が可能であったが、正門前で会話を行うには、パソコンを手を持って会話することは不可能であった。こえとらは研究室、正門前のどちらでも会話可能であった。

動作制限については、食事の会話はどちらのデバイスもテーブルに置くことで会話は可能であった。しかし、歩行中の会話は LiveTalk、こえとらの両方が、歩きスマホの危険により会話が不可能であった。

提案手法を用いた場合は、どの環境であっても会話可能であった。

5.3 操作性における評価

今回の提案手法ではジャイロセンサで表示文字を取得したが、従来のスマートグラスの操作はタッチパッドやコン

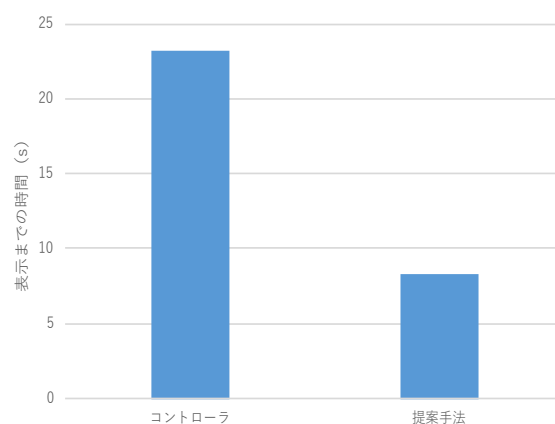


図 9 表示文字取得までの平均時間

トローラを用いる。表示文字取得におけるジャイロセンサを用いる優位性を示すために、指定の会話履歴を画面に表示するまでの時間を、ジャイロセンサとコントローラを用いた場合で比較した。図 9 は試行を 5 回行ったときの平均時間を表す。コントローラを用いた場合、約 23 秒かったが提案手法では約 8 秒であった。

6. 考察

6.1 使用制限における評価に対する考察

提案手法を用いることで、既存手法を用いた場合には会話できなかった場所でも会話できることが示せた。また、提案手法を用いて会話しながら別の動作をすることが可能であることが示せた。食事の制限の評価は、テーブルがあったため会話可能という結果になったが、テーブルがなく立食する場合、会話が難しくなる。提案手法では立食であっても会話は容易にできる。よって、提案手法が既存手法よりバリアフリーなコミュニケーション支援手法である。

6.2 操作性における評価に対する考察

コントローラを用いた場合、タッチ操作を認識しなかつ

表 2 評価結果

	使用環境	LiveTalk	こえとら	提案手法
使用場所の制限	研究室	無	無	無
	正門前	有	無	無
動作制限	歩行中	有	有	無
	食事中	無	無	無

たり、スクロールの反応が悪かったりすることから、指定文字を表示するまでに時間がかかった。提案手法では、首を動かすという単純な動作のため時間がかからなかった。このことから、提案手法は従来より円滑なコミュニケーションを実現できる。

しかし、歩行中や食事中は首が頻繁に動くため、意図せずに会話履歴を取得したり、テキストがぐらついたりすることがあった。意識的に首を動かす際の首の動きと、無意識に動かす首の動きには速度の違いが出るため、環境ごとにジャイロセンサの値に閾値を設定し、移動量の増加を抑えることで解決できる。また、加速度センサを併用し、歩行中であるという状態を判断することでテキストのぐらつきを低減できる。

7. まとめ

近年、音声認識が発展するにつれ、日常生活で音声認識を用いた技術が実用化されている。その一つに、音声コミュニケーション支援が存在する。音声認識した発言をテキスト化し、聴覚障がい者や耳の不自由な高齢者が視覚的に情報を取得するものである。

従来の音声コミュニケーション支援手法は、音声認識結果をパソコンや携帯端末に表示していたが、会話する場所や別の動作をしながらであると使用が制限されるという問題があった。

本研究では、円滑で制限のないバリアフリーなコミュニケーション支援手法を実現することを目的として、スマートグラスに発言内容を AR としてテキスト表示し、ジャイロセンサを用いて AR 情報を操作し表示テキストを取得する手法を提案した。

提案手法と既存手法の、使用場所の制限を比較し、提案手法が制限が少ないことからバリアフリーであることを示した。また、表示文字を取得するまでの時間を比較することで、スマートグラスにおける操作に提案手法を用いる優位性を示し、円滑なコミュニケーションが可能であることを確認した。

謝辞

本研究の一部は JSPS 科研費 (26540038) および文部科学省私立大学戦略的研究基盤形成支援事業の助成を受けたものである。

参考文献

- [1] 西川 俊, 高橋 秀知, 小林 正幸, 石原 保志, 柴田 邦博, 聴覚障害者のためのリアルタイム字幕表示システム, 電子情報通信学会論文誌, Vol.J78-D2, No.11, pp.1589-1597, 1995,
- [2] 中村 圭吾, 戸田 智基, 猿渡 洋, 鹿野 清宏, 肉電動人口音声の変換に基づく喉頭全摘出者のための音声コミュニケーション支援システム, 電子情報通信学会論文誌,

- Vol.J90-D, No.3, pp.780-787, 2007,
- [3] FUJITSU, Software LiveTalk, <http://www.fujitsu.com/jp/group/ssl/products/software/applications/ud/livetalk/>, (参照 2017-1-27).
- [4] FEAT, こえとらアプリサポートページ, <http://www.koetra.jp/>, (参照 2017-1-27).
- [5] 小松 史旺, 小林 吉之, 持丸 正明, 三林 洋介, 歩きスマホが反応時間および歩行動作に与える影響, 人間工学, Vol. 51 No. Supplement p. S178-S179, 2015.