

拡張現実感技術を用いた発話可視化システム MIERUKENの開発

長野 優一朗^{†1} 吉野 孝^{†2}

近年、拡張現実感技術の発展により、様々な情報を実空間上で可視化・共有できるようになってきた。本研究では、私たちに最も身近な情報発信手段である“発話”に着目し、身の回りの発話を可視化して周囲の人々に提供することで、日常生活の様々な場面を支援できると考えた。そこで、拡張現実感技術を用いて身の回りの発話を可視化するシステム MIERUKEN の構築を行った。本稿では、MIERUKEN の発話の可視化手法および試用実験の結果について報告する。

MIERUKEN: Speech Visualization System Based on Augmented Reality

YUICHIRO NAGANO^{†1} and TAKASHI YOSHINO^{†2}

As the spread of the Augmented Reality(AR) technology and service, we are getting at sharing and visualizing various information on real environment. In this study, we focus on “speech” that used to transmission of information on daily life. We think that speech visualization can support various situations in daily activities. We have developed speech visualization system MIERUKEN based on AR. In this paper, we present the result of a trial experiment and discuss the evaluation of three methods for visualized speech.

^{†1} 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University

^{†2} 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University

1. はじめに

ARToolKit や PTAM などの拡張現実感 (Augmented Reality: 以下, AR) 技術を用いた研究が活発化してきている^{1),2)}。その波は研究分野だけに留まらず、セカイカメラ^{*1}や ARis^{*2}などに代表されるように、一般的な利用へと拡大しつつある。また、Web カメラや GPS、ヘッドマウントディスプレイ (以下, HMD) といった AR に関連する機器も安価となり、多くの人が AR と触れ合う機会が増加してきている。

AR 技術の発展により、もたらされた流れの一つとして挙げられるのが、情報の可視化である。これまでディスプレイ上でしか見ることはできなかった情報が、実世界と関連付けられて表示・利用されるようになった。これにより、遠隔地間における空間情報の共有や実空間の物体へのアノテーションなど、実世界指向での情報利用が可能となった³⁾。今後ますます、AR を利用した可視化サービスが身近になってくるものと思われる。

これまで、日常生活における会話などの発話情報に関しては、プライバシーへの配慮からあまり積極的に可視化されていない。しかし、私たちは普段の生活の中で、何気なく耳にした会話から、新しいアイデアが思い浮かんだり、求めていた情報を得ることができたりと、様々な利益を得る機会が多々ある。もし、日常生活における発話が適切な環境下で可視化され、周辺の人々に提供されたならば、より多くの人々が、この利益を享受できる機会が増えると考えた。

そこで、本研究では、日常における人々の発話に注目し、収集した発話を AR 技術を用いて可視化・共有するシステム MIERUKEN を開発した。MIERUKEN は、3 つの可視化手法を用いて発話の可視化を行い、発話者の位置情報に基づいて情報を提供する。AR 技術を用いて発話を可視化することで、利用場面に応じて様々な支援が可能であると考えている。

2. 関連研究

2.1 発話の可視化

発話の可視化を用いた研究事例として、インタラクティブアートやパフォーマンスアートなどがある^{4),5)}。

Levin らは、雲や泡など様々なメタファを用いて発話を可視化することで、発話の可視化

*1 セカイカメラ: <http://sekaicamera.com/>

*2 ARis: <http://www.geishatokyo.com/jp/ar-figure/>

によるインタラクティブアートとしての可能性を見出した⁴⁾。また、Lewisらは、開発したテキストライブラリを用いて発話をフレキシブルな文字で可視化し、パフォーマンスをより引き立てることができた⁵⁾。

一方で、可視化した発話を用いてコミュニケーションを支援する例は多くはない。清水らが開発したしゃべりカスは、その一つの事例で、「発話の可視化による、対面コミュニケーションの支援」をコンセプトに、胸部につけたディスプレイを介して自分が発話した内容を表示するシステムである⁶⁾。発話から抽出した品詞の種類に応じてテキストの色を変化させるなどして、コミュニケーションが活性化された状態を継続させることを目的としている。

日常生活における発話を対象として、その発話を用いてユーザ支援に応用した事例は少ない。その最たる要因として、プライバシーの問題が挙げられる。しかし、普段の生活の中で、何ら関係の無い会話などの情報から、求めていた情報が得られたり、新しいアイデアが浮かんだりするなどの利益を得られることが多々ある。日常生活における発話を可視化することで、より多くの人々がこの利益を享受できると考えた。

本研究では、発話者から収集した発話情報を利用して、周辺の人々の支援に役立てる。また、利用場面に応じて可視化方法を適切に制御することによって、プライバシーに配慮した発話の利用を目指す。

2.2 インフォーマルな場におけるコミュニケーション支援

本研究では、発話の可視化を用いた支援の一つに、インフォーマルな場を対象としたコミュニケーション支援を考えている。

これまでも、インフォーマルな空間において、コミュニケーションを支援する研究が行われている^{7),8)}。松田らの研究では、共有スペースにおいて、あらかじめ登録しておいた“今必要としている情報”を付近のディスプレイ上に提示することで、情報提供者を見つけ出し、提供者と登録者のコミュニケーション支援を試みている⁷⁾。また、武田らは、ICカードを用いて、イベントに参加しているユーザの氏名や所属情報、行動履歴等を管理し、必要に応じてこれらの情報をもとに Web から情報を収集・提供するなどして、大勢の人々が行き交う空間でのコミュニケーション支援を行っている⁸⁾。

しかし、従来の研究では、支援を受ける前提として、事前に決められた情報を入力・設定しておく必要があった。

そのため、近年では、MySpace^{*3}や mixi^{*4}などのソーシャルネットワーキングサービス（以下、SNS）が一般的になってきたことを受け、SNS から取得した情報を用いてコミュニケーションを支援する事例が増えてきている^{9),10)}。村上らは、国内最大の SNS である mixi に登録されている情報からお互いの共通点を提示することで、初対面の人同士のコミュニケーションを支援している⁹⁾。

しかし、SNS を利用する場合においても、互いに共通したサービスを利用しているなど、いくつかの条件が存在する。そのため、実際に支援を受けるまでの条件を整えることが困難であった。

インフォーマルコミュニケーションの発生が偶発的であり、相手も話題もコミュニケーションが開始されるまで不明であることを考えると¹¹⁾、事前の準備なしに、その場ですぐに支援を提供できることが望ましい。

そこで、本研究では、その場で既に展開されている会話中の発話をリアルタイムに可視化し、会話参加のためのアウェアネス情報として周囲の人々に提供することで、会話の輪の中に加わるきっかけ作りを支援する。これにより、その場におけるコミュニケーションがより活発化できるだけでなく、初対面の人同士のコミュニケーションも促進できると考えている。

3. 発話可視化システム MIERUKEN

3.1 設計方針

MIERUKEN は、日常生活における様々な発話を AR 技術を用いて可視化することで、利用場面に応じた支援を提供することを目標としている。

以下に、本システムの設計方針を示す。

(1) 日常会話の利用

発話は、私たちにとって最もシンプルなコミュニケーション手段であり、私たちは発話を用いて様々な情報を発信している。発話により発信される情報は、会話をしている当事者間以外の人々に対しても様々な利益をもたらす。例えば、普段の生活において偶然耳にした会話の内容から、新しいアイデアが思い浮かんだり、探していた情報を入手できたりといったことなどが挙げられる。発話を利用することで、多種多様な情報を取得することができるだけでなく、可視化によって、直接発話が聞こえない人々に対しても、その利益を提供できるようになると考えた。

本研究では、発話をアウェアネス情報として捉え、日常生活における発話を可視化することで、利用場面に応じた様々な支援を提供する。

*3 MySpace: <http://www.myspace.com/>

*4 mixi: <http://mixi.jp/>

(2) プライバシへの配慮

可視化の対象として、日常生活における発話を利用するため、プライバシーの問題は無視できない。今回の試用実験で行ったアンケート結果からも、プライバシー侵害の恐れから発話の可視化に抵抗を感じている被験者が多く見受けられた。プライバシーへの配慮は、システムを設計する上で重要な要素の一つである。

本研究では、発話者との距離に応じて発話の可視化方法を変化させ、ユーザに提示する情報量を制御することで、プライバシー問題への対処を試みる。これにより、発話者のプライバシー配慮しながらも、周辺の人々に会話参加のための適切な情報を提供することが可能になると考えている。

3.2 システム構成

図 1 に発話可視化システム MIERUKEN のシステム構成を示す。MIERUKEN では、発話ユーザから収集した発話情報は、管理サーバを介して各 AR ユーザのもとへと送信される。

マイクから収集した音声データは、連続音声認識ソフト Julius¹²⁾ を用いて文字列情報（以下、発話データ）へと変換される。その後、GPS から取得した発話者の位置座標とともに、管理サーバへと自動的に送信される。管理サーバは、発話データを受信すると、発話ユーザの位置情報と定期的に受信している AR ユーザの位置情報とを比較し、特定範囲内にいる AR ユーザにのみ発話データを提供する。AR ユーザ側では、発話者の位置情報に基づいて発話者と発話データとを関連付け、各可視化手法に基づいて可視化された発話データ（以下、発話オブジェクト）を HMD 上に表示する。

なお、拡張現実感の実現にはマーカベースの AR ライブラリである ARToolKit¹⁾ を、発話オブジェクトの描画には OpenGL を用いている。また、管理サーバ、発話収集処理および通信処理を C# で、可視化処理を C++ でそれぞれ実装した。

図 2 に AR ユーザのハードウェア構成を示す。可視化された発話データは、HMD を介してユーザに提示される。HMD には、Viuzix 社製 iWareVR920 を利用しており、内蔵されている 3 軸加速度センサを用いて、ユーザの頭の動きに合わせて発話データを表示している。Web カメラには Microsoft 社製 LifeCam Show を用いており、本システムでは、フレームレートを毎秒 30 フレーム、解像度を 800x600px に設定して利用している。GPS および電子コンパスは、実空間上のユーザの位置座標および正面の方位座標の取得を目的としており、これらの情報を用いて実空間上における発話ユーザと AR ユーザとの位置関係を補足している。

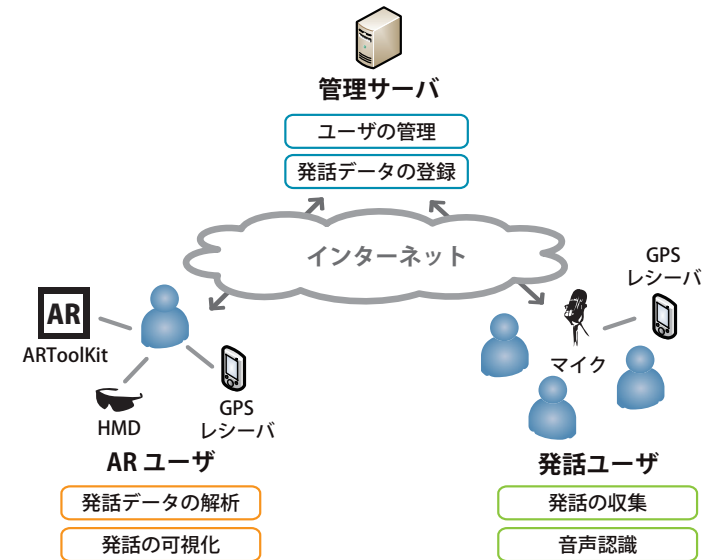


図 1 システム構成図

Fig. 1 The configuration of MIERUKEN.

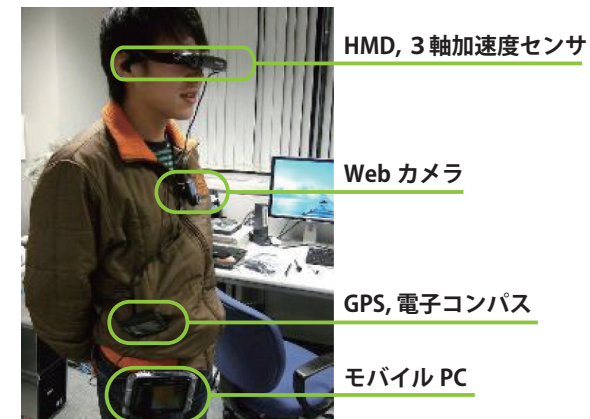


図 2 AR ユーザのハードウェア構成

Fig. 2 The configuration of AR user hardware.

3.3 MIERUKEN の機能

(1) 3つの可視化手法

MIERUKEN では、3つの可視化手法を用いて、取得した発話情報を可視化する。図3は、マイクから取得した会話者間の発話データを、それぞれセンテンス、ワード、イメージの3つの可視化手法を用いて可視化した結果である。可視化した発話オブジェクトは、発話者の頭の上に表示され、時間経過とともに上空へと移動する。

センテンス方式では、音声認識を用いて音声データを文字列データへと変換し、その結果を利用して発話内容を表現している。

ワード方式では、取得した音声認識結果を、形態素解析を用いて品詞に分解し、名詞のみを利用して発話内容を表現している。

イメージ方式では、形態素解析により抽出した名詞をクエリとして画像検索を行い、Web上から取得した画像を用いて発話内容を表現している。なお、形態素解析には Mecab¹³⁾ を、Web 画像検索には Bing API^{*5)} を利用している。

(2) 発話遡り機能

HMD 上に表示されている発話オブジェクトは、時間経過とともに画面上部へと移動する。そのため、一定時間が経過するとオブジェクトは画面から消え、AR ユーザの視界に入らなくなる

発話遡り機能は、一度画面から消えた発話オブジェクトを追跡し、再度画面上に表示させる機能である。図4に、発話遡り機能を利用して、過去の発話を閲覧する方法を示す。HMD に内蔵されている3軸加速度センサを用いて、画面上部へと移動したオブジェクトを頭を上下に動かすことで追跡し、再度画面に表示させる。過去に表示された発話オブジェクトを遡って閲覧できることで、会話の一連の流れを把握する手助けとなったり、会話の内容を再確認する機会を与えることができると考えている。

また、操作インターフェースとして、見上げるという単純な動作をメタファに用いており、ユーザは簡単に本機能を利用できると考えている。

4. 試用実験

4.1 実験の概要

本研究では、システムによる支援の提供場面の一つとして、インフォーマルな場における会話参加支援を考えている。そこで、試用実験に際し、次の2つのシチュエーションを設定した。



図3 3つの可視化手法

Fig. 3 The three kinds of visualization method.

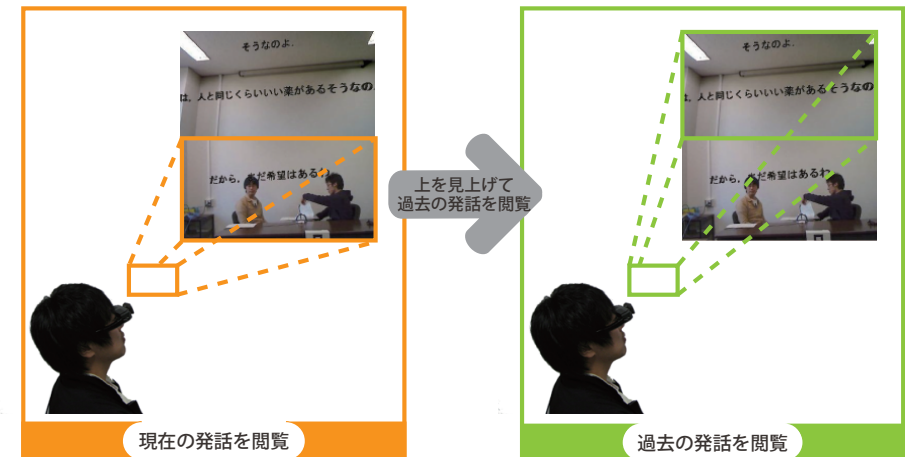


図4 発話遡り機能

Fig. 4 The function of look back to a past speech log.

- a. 会話の内容が聞こえる状態
- b. 会話の内容が聞こえない状態

会話者の話し声は、インフォーマルな空間においてコミュニケーションを図る上で重要な要素であり¹⁴⁾、会話の話し声の有無によっては、求められる要件が大きく異なることが予

*5 Bing API: <http://www.bing.com/developers>

想される。そのため、今回の実験では、会話が聞こえる状態（以下、音声有）と聞こえない状態（以下、音声無）の2つのシチュエーションを設定し、ヘッドフォンを用いて疑似的に想定している利用場面を再現した上で、本システムの試用評価を実施した。

今回の試用実験では、会話の内容把握と発話者のプライバシーの2つの観点から、2つのシチュエーションに基づいて3つの可視化手法の検証を行う。なお、発話通り機能の効果についても併せて評価を実施する。また、今回は、マイク入力による音声認識は用いず、事前に用意した対話形式のテキストを音声認識後の発話データとして用いることで、正確な音声認識がなされていることを前提に実験を実施した。これは、今回の実験の目的が、3つの可視化手法による会話の内容把握効果についての検討、および発話通り機能の評価であり、音声認識利用による精度評価に関しては評価の対象としていないためである。さらに、本実験は屋内で実施したため、今回はGPSおよび電子コンパスによる発話者と発話データとの関連付けは行わず、ARToolKitによるマーカー認識を用いて関連づけを行った。

図5に実験で用いた会話文の一例を示す。会話文には、日本語能力試験の聴解問題より、2級レベルに相当する会話文6種類を引用した^{15),16)}。また、引用した会話文にはそれぞれ、会話内容を問う四択式の問題が一問備わっており、この問題を会話の内容把握に関する調査に利用した。

被験者は、和歌山大学に所属する日本人学生13名である。うち、3名は用意した会話文を読み上げる会話者として実験に参加した。

4.2 実験手順

被験者を表1の実験パターンに従って、case1またはcase2に割り振る。次に、割り当てられたパターンに基づいて、センテンス・ワード・イメージの順に用意した会話文を可視化して、被験者にシステムを利用してもらう。実験は、上記の手順を音声有と音声無を切り替えて行う。例として、case1に割り当てられた被験者は、まず、音声有の状態ではセンテンス、ワード、イメージの順に会話文を見てもらい、次に、音声無の状態ではセンテンス、ワード、イメージの順に会話文を見てもらうこととなる。なお、今回の実験において、音声無の状態はヘッドフォンを用いて再現した。

本実験では、主な調査手段として、会話の内容把握テストおよび5段階評価によるアンケート調査を用いた。会話の内容把握テストは、各会話文を見せた直後に実施し、会話文の内容を問う四択問題への解答、会話中に登場したと思われる単語の書き出し、および意見・感想の記入を行ってもらった。アンケート調査は、全会話文を見終わった後に実施し、各可視化手法や発話通り機能に関する質問、発話可視化に関する意識調査など、実験全般を対象

会話文

F: 新しいスポーツセンターどう？料金は高めって聞いたけど。

M: 高いけど、いいよ。設備も最新で駅からもすぐ。

F: じゃ、人気あるでしょ。

M: うん。たくさんの人が利用してるし、みんなは満足しているみたいだよ。

F: みんなは？

M: しっかり使っている人はね。でも入会したとたんに、残業、残業の僕としては損している気分だよ。これじゃ、全然やせられないなあ。

問題

どうして男の人は不満ですか？

選択肢

1. 料金が安いからです。
2. あまり行く時間がないからです。
3. 全然やせないからです。
4. 利用する人が多いからです。

図5 実験に用いた会話文および問題の一例

Fig.5 Example of speech texts and a question.

表1 実験のパターン

Table 1 The experiment patterns.

	会話文					
	Text1 (センテンス)	Text2 (ワード)	Text3 (イメージ)	Text4 (センテンス)	Text5 (ワード)	Text6 (イメージ)
case1	音声有			音声無		
case2	音声無			音声有		

とした調査を行った。また、5段階の評価尺度には、リッカートスケールを用いた。

5. 実験結果

表2に、音声有および音声無の状態において、各可視化手法別に行った会話の内容把握テストの結果を示す。

Q1は、会話文の内容把握を問う四択問題の結果を正解を○、誤答を×として示している。音声有の場合、ワード方式の100%が最も高く、次いで、イメージ方式の90%、最も低かったセンテンス方式は40%であった。次に、音声無の場合、センテンス方式が80%と最

表 2 会話の内容把握テスト結果

Table 2 The result of a content understanding test.

被験者	音声有						音声無					
	センテンス		ワード		イメージ		センテンス		ワード		イメージ	
	Q1	Q2	Q1	Q2	Q1	Q2	Q1	Q2	Q1	Q2	Q1	Q2
user01		0.83		0.94		0.80		0.91		0.97	×	0.17
user02		0.73		0.86		0.79	×	0.90		1.00	×	0.57
user03	×	0.76		0.90	×	0.92		0.93	×	0.94	×	0.20
user04	×	0.89		0.97		0.96		1.00	×	0.88	×	0.11
user05	×	0.83		0.92		0.86		0.83	×	0.94	×	0.17
user06		0.89		0.88		0.96		0.83	×	0.92		0.67
user07		1.00		0.88		0.75		0.81		0.83	×	0.42
user08	×	0.80		0.73		0.73		0.86	×	0.89		0.40
user09	×	0.96		0.92		0.83	×	0.83	×	0.83	×	0.40
user10	×	1.00		0.88		0.83		0.94	×	1.00	×	0.13
平均	0.4	0.87	1.0	0.89	0.9	0.84	0.8	0.89	0.3	0.92	0.2	0.32
標準偏差	0.52	0.30	0.00	0.27	0.32	0.34	0.42	0.24	0.48	0.24	0.42	0.47

Q1: 会話文の内容把握を問う四択問題 (正解: , 誤答: ×)

Q2: 会話に登場したとして書き出された単語を, 各会話文ごとに, 会話者 3 名に「0:会話文と全く関連しない」「0.5:会話文と関連する単語」「1:会話文に登場する単語」の 3 段階で評価してもらい, 評価した単語数および評価者数で平均した値

も高く, 次いで, ワード方式の 30%, 最も低かったイメージ方式は 20%であった.

Q2 は, 被験者に会話中に登場したと思われる単語を書き出してもらった結果を, 会話者 3 名に「0:会話文と全く関連しない」「0.5:会話文と関連する単語」「1:会話文に登場する単語」の 3 段階で評価してもらい, その後, 単語数と評価者数で平均した結果である. 音声有の場合, すべての可視化手法において, 書き出された単語のうち, 80%以上の単語が会話文と関係があるとされた. 次に, 音声無の場合, センテンスおよびワード方式に関しては, 80%以上の単語が会話文と関係があるとされたが, イメージ方式に関しては, 会話文と関係があるとされた単語が 30%と少ない.

表 3 に, 5 段階評価によるアンケート結果を示す.

(1), (2) は可視化手法別の支援効果について, (3), (4) は発話を可視化することへの意識について, (5), (6), (7) は発話通り機能についてのアンケート結果である.

(3), (4) の結果から, 音声の有無によって, 発話の可視化に対する意識に変化が見られる. 両者を比較すると, 音声有の場合には, 可視化に対する抵抗意識が音声無の場合に比べて低くなる傾向にある.

表 3 5 段階評価アンケート結果

Table 3 The result of a questionnaire survey using a five-point Likert item.

質問項目	可視化手法	音声	評価平均	標準偏差
(1) 会話の内容把握の手助けになると感じた.	センテンス	有	4.2	0.79
		無	4.8	0.42
	ワード	有	4.0	0.82
		無	3.7	0.82
	イメージ	有	1.9	0.74
		無	1.9	0.99
(2) 会話参加への手助けになると感じた.	センテンス	有	3.4	0.84
		無	4.0	0.82
	ワード	有	3.6	0.84
		無	3.7	1.16
	イメージ	有	2.4	0.97
		無	2.3	1.57
(3) 自身の会話の内容が可視化されることで, プライバシーが侵害されると感じる.	センテンス	有	1.9	0.74
		無	3.8	0.92
	ワード	有	1.6	0.84
		無	3.3	0.95
	イメージ	有	1.7	1.06
		無	2.7	1.06
(4) 自身の会話の内容が可視化されることに, 抵抗を感じる.	有	2.2	0.92	
	無	3.4	0.70	
(5) 過去の発話内容を遡って見ることで, 会話内容を把握する手助けになると感じた.	有	3.6	0.70	
	無	4.2	0.92	
(6) 過去の発話内容を遡って見る行為は, (操作性の観点から) 容易だった.			4.4	0.70
(7) 遡って見ることが可能な発話の情報量は, 適切だった.			3.0	0.94

評価平均とは「1:強く同意しない」「2:同意しない」「3:どちらでもない」「4:同意する」「5:強く同意する」の 5 段階の評価基準による評価結果の平均.

6. 考 察

6.1 3 つの可視化手法

各可視化手法別に, 音声有と音声無の場合において, Q1 の正答率に違いが生じた要因について考察する.

● センテンス方式

音声無の場合での正答率 80%に対して, 音声有の場合では正答率 40%と, 40 ポイント低い. 実験後のアンケートからも, 音声有の場合において, 「情報量が多く見づらい」や「要点がわかりにくい」などの意見が多く寄せられた. しかし, 表 3 のアンケー

ト項目では、音声有および音声無の両方において、センテンス方式による可視化は、会話の内容把握の手助けになるとの評価を得ている。

この要因として、会話を聞きながら表示された文章を見たことで、注意が散漫し、内容把握が曖昧になってしまったと考えられる。センテンス方式において、被験者に提示したテキストは、事前に用意された会話文であり、会話者が話している内容と同一のものである。そのため、音声認識を利用したときのような、話している内容と表示されたテキストとが一致しないことで、被験者が混乱するような状況が発生したとは考え難い。

- ワード方式

音声有の場合での正答率 100%に対して、音声無の場合では正答率 30%と、70 ポイント低い。実験後のアンケート結果からも、音声無の場合において、「一部（問題を解答するのに）必要な情報が出ていなかった」「話の全体像がつかめない」などの意見が多かった。しかし、表 3 のアンケート項目では、音声有および音声無の両方の状況において、ワード方式による可視化は、会話の内容把握において有用であるとの評価を得ている。

この要因としては、ワード方式によって表示された単語は、会話の内容理解ではなく、聞き取った内容を再確認するための用途として利用されているからではないかと考えられる。そのため、音声無の場合においては、表示された単語のみから会話の内容を推測しなければならず、結果的に問題の解答に求められているレベルまで会話の内容を把握できていないものと推測する。

- イメージ方式

音声有の場合で正答率 90%に対して、音声無の場合では正答率 20%と、ワード方式と同じく 70 ポイント低い。実験後のアンケート結果からも、「画像だけでは会話の内容は全くわからない」「（音声無の場合）ないよりはましだが、精度が悪くて必要ない」など、画像だけでは内容理解に苦しむといった内容の意見がほとんどであった。また、表 3 のアンケート項目から、イメージ方式による可視化は、会話の内容把握の手助けとなっていないことがわかる。

音声有の場合において、正答率が高かった理由として、被験者が会話の聞き取りを主として内容把握を行ったからと考えられる。イメージ方式においては、センテンス方式とは異なり、会話の聞き取りを妨害することはないなかったと考えられる。これは、会話の聞き取りを妨害する要因として、画面上での情報量が多く、かつテキストデータの場合にのみ妨害が発生するものと考えられる。そのため、テキストデータではない画像

は、画面上の情報量が多くても、テキストデータでないため、音声有で正答率 90%を確保できたと思われる。

また、実験後に、会話者 3 人に対して、実験時にイメージ方式で試用した 2 つの会話文の画像を、会話内容に沿っている・いないの 2 段階での評価を依頼した結果、会話内容に沿った画像と判断されたのは全体の約 2 割程度であった。

6.2 発話遡り機能

実験時に発話遡り機能が利用された回数を HMD 上に表示された画面のキャプチャデータから解析を行った。その結果、発話遡り機能の利用回数は 45 回であり、そのうち約 24%に当たる 11 回は、終了間際の利用であった。ただし、実験施行回数 60(10 名×6) 回のうち、15%に当たる 9 回分の実験において、データが取得できていなかった。そのため、今回の実験においては、アンケート結果を中心に評価を行う。

表 3 の (5) より、音声有の場合は 3.6、音声無の場合は 4.2 と、音声の有無に限らず、本機能が内容把握の手助けになるとの評価を得た。自由記述には、「（会話内容を）思い返す時に利用できる」や「センテンス方式の場合に多く利用した」などの意見が挙げられていた。また、操作性の面に関しても、表 3 の (6) より、4.4 と高い評価を得ている。その理由として、「直観的でわかりやすい」や「見上げるだけでなので容易だった」などの意見が挙げられており、ユーザに負担を与えることなく有効に機能していると考えられる。しかし、表 3 の (7) より、遡ることができる情報量の適切性に関しては、3.0 と低い評価であった。その理由として、「少し前の情報なら遡れるが、あまり前の情報はみることができない」など、表示される情報量が少ないとの指摘が多くあった。本システムでは、処理速度確保のために、描画開始から一定時間が過ぎた発話オブジェクトに関しては、自動的に描画対象から除外されるよう実装している。そのため、一部の被験者においては、意図した過去の発話データまで遡れず、そのことが結果に影響したと考えられる。

7. おわりに

日常における人々の発話に注目し、収集した発話を AR 技術を用いて可視化・共有するシステム MIERUKEN の構築を行った。本論文では、MIERUKEN の試用実験の結果および会話可視化による効果について考察を行い、以下のことが分かった。

(1) 3 つの可視化手法

- センテンス方式：

音声有では、利用により会話の内容把握を著しく低下させる。

音声無では、会話内容を把握する有効な手段である。

- ワード方式：

音声有では、会話内容の再確認のために利用された。

音声無では、会話の内容把握についてあまり効果を得ることができなかった。

- イメージ方式：

音声有では、会話内容の把握にはほとんど貢献しないが、センテンス方式とは異なり、内容把握を妨害することはない。

音声無では、会話内容の把握にはほとんど貢献しない。

(2) 発話通り機能

発話通り機能を用いることで、容易に過去の発話を遡ることができる。

会話の内容把握を支援する可能性がある。

今回の試用実験の結果より、発話の可視化を用いてインフォーマルな場での会話参加支援を行う場合、話し声が聞こえる距離にいるユーザに対しては、ワード方式もしくはイメージ方式による支援が、話し声が聞こえない距離にいるユーザに対しては、センテンス方式による支援が、適していると考えられる。ただし、話し声が聞こえない距離でのセンテンス方式による支援の提供は、発話者のプライバシーの問題があるため、検討が必要である。

今後は、試用実験の結果をもとにシステムの改良を進めるとともに、実際にインフォーマルな場においてシステムを利用してもらい、コミュニケーション支援の効果について評価を行う。

謝辞 本研究は、日本学術振興会科学研究費基盤研究(B)(19300036)の補助を受けた。

参 考 文 献

- 1) Kato, H., Billingham, M.: Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, Proc. IEEE and ACM International Workshop on Augmented Reality (IWAR '99), pp.85 (1999).
- 2) Klein, G., Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces, Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07), pp.1-10 (2007).
- 3) Stafford, A., Piekarski, W. and Thomas, B.: Implementation of God-like Interaction Techniques for Supporting Collaboration Between Outdoor AR and Indoor Tabletop Users, Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '06), pp.165-172 (2006).

- 4) Levin, G., Lieberman, Z.: In Situ Speech Visualization in Real-Time Interactive Installation and Performance, Proc. International symposium on Non-photorealistic animation and rendering (NPAR '04), pp.7-14 (2004).
- 5) Lewis, J., Assogba, Y.: Taking sides: dynamic text and hip-hop performance, Proc. ACM international conference on Multimedia (MULTIMEDIA '06), pp.744-747 (2006).
- 6) 清水大悟, 安村通晃: シャベリカス: 発話を視覚化するウェアラブルインタフェース, 情報処理学会研究報告, 2009-HCI-132, pp.1-8 (2009).
- 7) 松田 完, 西本一志: 談話の杜: インフォーマルスペースにおける実世界での出会いを利用した効率的な情報共有システム, 情報処理学会研究報告, 2002-GN-31, pp.109-114 (2002).
- 8) 武田英明, 松尾 豊, 濱崎雅弘 ほか: イベント空間におけるコミュニケーション支援, 電子情報通信学会誌, Vol.89, No.3, pp.206-212 (2006).
- 9) 村上豊聡, 吉野 孝: ユーザのコミュニティ情報を用いた SNS ユーザの出会い支援システム DI の開発, マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO '07), pp.1510-1513 (2007).
- 10) 嶋田陽介, 加藤貴之, 廣嶋拓也 ほか: 共通の趣向を持つ利用者を発見するソーシャルネットワークシステム, 情報処理学会第 67 回全国大会, 第 3 分冊, pp.157-158 (2005).
- 11) Fish, R.S., Kraut, R.E. and Chalfonte, B.L.: The VideoWindow System in Informal Communications, Proc. ACM conference on Computer-supported cooperative work (CSCW '90), pp.1-11 (1990).
- 12) 河原達也, 李 晃伸: 連続音声認識ソフトウェア Julius, 人工知能学会誌, Vol.20, No.1, pp.14-19 (2005).
- 13) Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, Proc. Conference on Empirical Methods on Natural Language Processing (EMNLP '04), pp.230-237 (2004).
- 14) 渋谷昌三: パーソナル・スペースの形態に関する一考察, 山梨医大紀要, 第 2 巻, pp.41-49 (1985).
- 15) 銅直信子, 原やす江, 木下康利 ほか: あなたの弱点がわかる! 日本語能力試験 2 級模試 × 2, 日本語テキスト研究会 (2005).
- 16) 松村節子, 佐久間良子, 野呂ケイ ほか: 日本語能力試験 これで大丈夫 聴解問題 2 級, ユニコム (2005).