

タブレット PC の前背面カメラを活用した AR による会話特徴量の可視化

安達 寛之^{1,a)} 明神 聖子^{1,b)} 島田 伸敬^{1,c)}

概要: 現代においてコミュニケーションの機会が増加し重要なものとなっているが、人により発話の多寡があり発話の少ない人からの情報が提供されない場面が考えられ、会話ロボットによる介入など、それを補助する様々な方法が模索されている。我々は、多人数会話におけるコミュニケーションについて、タブレット PC の前面および背面の 2 台のカメラを用いて、発話の対象者や発話の頻度などの会話の特徴量を取得する。そして、AR を利用して、カメラによる実世界の映像の中に会話特徴量を可視化表示するシステムを実装する。さらに、情報をそれぞれの人向けに調整したうえで提示できるという AR の特性を活用し、各参加者の状況に応じた情報を提示することで支援を行い、コミュニケーションの活性化を目指す。

AR Visualizing Features of Conversations by using Cameras on Front and Back of Tablet PC

HIROYUKI ADACHI^{1,a)} SEIKO MYOJIN^{1,b)} NOBUTAKA SHIMADA^{1,c)}

Abstract: Nowadays, opportunities for communication are increasing and important. However, sometimes, ideas and topics by people who do not speak frequently are not utilized sufficiently for communication. Therefore, many previous works has been exploring methods to support people who do not speak frequently by using conversation robots assistances etc. We get the features of conversation such as 'conversational partner' and 'utterances frequency' by using cameras on front and back of a tablet PC in conversations by mutiple people. We develop the system that shows these features to visualize them in real worlds by AR. In addition, AR can show information which individually fixed based on user-specific view. We aim to support for individuals according to each user's situation by showing the information and activate communication.

1. はじめに

拡張現実感 (Augmented Reality:AR) とは、カメラなどを通して見た現実の映像にコンピュータを用いて情報を付加し、それが現実存在するかのように提示する技術である。近年、カーナビに AR 機能が搭載されたり携帯型ゲーム機でも AR を体験できたりするなど、AR は徐々に身近なものとなってきている [1]。特にスマートフォンやタブレット PC といったモバイル端末の普及に伴い、セカイカメラ [2] などに代表される、AR 技術を利用したアプリケー

ションも多く開発されるようになった。

一方で、現代においてコミュニケーションは非常に重要なものであるが、頻繁に発言する人同士のみで会話が進められ、あまり発言しない人が持っている情報が提供されないような場面が考えられる。このような人から積極的に情報を引き出すよう働きかけることで、新たな話題を提供しコミュニケーションを活性化させたい。

コミュニケーション支援の例として、松原らによる「サイバー囲炉裏」[3] がある。これは囲炉裏をメタファとした語らいの空間を用意し、会話のきっかけになるような情報を表示することで、インフォーマルコミュニケーションを触発するシステムである。タッチパネル式の水平ディスプレイに表示される泡をさわること、それにリンクしたホームページが表示され、話題が提供される。このように

¹ 立命館大学
Ritsumeikan University
a) adachi@i.ci.ritsumei.ac.jp
b) seiko@i.ci.ritsumei.ac.jp
c) shimada@ci.ritsumei.ac.jp

システムがあらかじめ用意しておいた情報を表示するだけでなく、状況に応じた適切な情報を提示する例として、木村ら [4] は適切な音量で発話できるようなコミュニケーション支援のために発話音量を視覚的に提示する方法を検討している。また、大島ら [6] はコミュニケーションの背景にある雰囲気の変化など参加者の拮抗した関係性から生まれるダイナミクスをリアルタイム CG で可視化する TableTalkPlus と呼ばれるシステムを用い、会話の場をデザインする試みを行っている。TableTalkPlus では、発話者のもとにクリーチャーと呼ばれる光が集まり、話終わると周囲に拡散していくなど会話の移り変わりを表現することで、会話を促進させたり性質を変容させたりする効果が示されており、新しいコミュニケーションツールとしての可能性がうかがわれるものである。

本研究では、タブレット PC の前面カメラで会話の参加者の顔を映す一方、背面カメラで他の参加者を映すというように、2 台のカメラを同時に利用して発話の対象者や発話の頻度などの会話の特徴量を取得し、AR による可視化を行う。TableTalkPlus では参加者全員が同じ映像を見ているが、本システムでは各参加者がタブレット PC を持つことで、それぞれに異なる情報の提示を可能とする点で異なる。また、AR を用いたコミュニケーション支援の例として Almeida ら [5] によるものなどがあるが、本システムでは単純に提示するのではなく、各参加者によって異なる別の参加者との関係性の提示など、情報を個別に調整したうえで提示するという点が特徴である。これを利用して各参加者の状況に応じた支援を行うことで、コミュニケーションの活性化を目指す。

以下、2 章で本研究のコンセプト、3 章で開発するシステムについて述べる。4 章でシステムの実装、5 章で実験とその結果、6 章でまとめと今後の課題を述べる。

2. コンセプト

本研究のコンセプトは、多人数会話におけるコミュニケーションについて、タブレットの前背面カメラを活用してその特徴量を取得し、頻繁に会話している人とそうでない人といった参加者同士の関係性など目には見えない情報を AR によって可視化することで、コミュニケーションの活性化を支援するシステムを実現することである。そして、可視化する情報の違いがコミュニケーションに及ぼす影響や、何をどう提示すればコミュニケーションの活性化につながり、参加者を支援できるかを明らかにすることを目指す。

AR によって可視化される会話特徴量とは例えば次のようなものである。

- 誰に対して話しているか
- どの程度話しているか
- 話題に対する反応はどうか

これらの情報は会話をしている瞬間のものと同様までの履歴の 2 種類がある。瞬間の様子は直接その場でわかるが、それに加えて映像として提示することでさらにそれを強く意識することができる。履歴については、例えば自分がある人には頻繁に話しているが別の人にはあまり話していないという情報が提示されることで、その人と会話してみようというような働きかけの効果が期待される。

また、コミュニケーションの活性化支援とは、話題に対する各参加者の反応を共有したり、それぞれの状況に応じて異なる情報を提示したりすることで、コミュニケーションを活性化させようというものである。特に、後者については本研究の特徴的な取り組みであり、ある参加者から見たほかの参加者との関係性と、別の参加者から見た関係性とは異なることから、それを区別して提示する。これは各参加者が端末を所持することに加え、個別に情報を調整し提示できるという AR の特性が活用されることによるものである。例えば、ある参加者との間でどれくらい話し合ったかなどの 2 者関係は参加者ごとに異なるので、そのような情報を AR を用いて各参加者ごとに調整して提示し、各自に適した支援を行う。このような個別の情報に対し、いま誰が誰に話しているかというような、参加者の関係に因らない情報は参加者全体で共有される。

3. システム

3.1 システムの概要

開発するシステムの概要を図 1 に示す。テーブルを囲んでの多人数会話におけるコミュニケーションの基本的なモデルとして 3 人での会話について考える。参加者はそれぞれタブレット端末を所持した状態で会話を行う。各端末では取得した会話特徴量をサーバへ送信し、サーバ上で統合された結果に基づき可視化を行う。また、テーブルには位置検出の基準となるマーカを設置する。

特徴量の可視化の様子を図 2 に示す。ここでは参加者 B が参加者 C に対して発話している様子が参加者 A の視点で可視化されている。個別の情報として各参加者に対する発話頻度の差が星形の数で表現されるなど、これまでの履歴に基づいて自分と相手との関係性が提示されることに加え、全体で共有される情報として矢印による発話者とその対象者の表現や、頭上の音量表示の増減で発話の強度の表現などが行われる。

3.2 システムの機能

システムの基本機能として、「顔の方向の推定」と「発話頻度の計測」の 2 つがあり、参加者ごとにこの情報が計測される。

- 顔の方向の推定

タブレット端末の前背面 2 台のカメラを用いてマーカ基準での参加者の位置と顔の向きを求める。

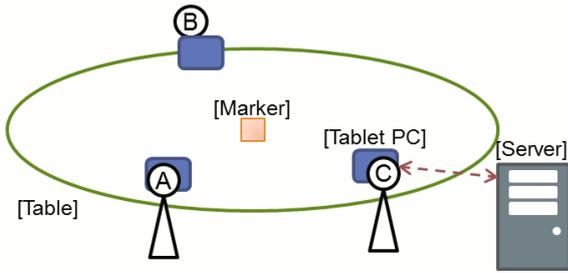


図 1 システムの構成
Fig. 1 System structure.

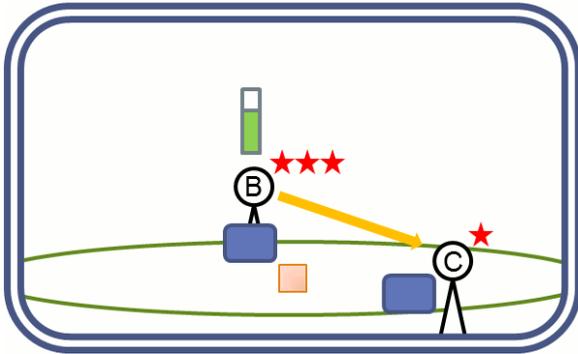


図 2 特徴量の可視化の様子
Fig. 2 Visualization of features on the screen.

- 発話頻度の計測

タブレット端末のマイクを用いて参加者がどの程度発話しているか計測する。さらに、顔の方向の推定結果からマークを基準とした各方向に対する発話頻度のヒストグラムを作成する。

そして、それらの計測結果から「会話特徴量の可視化」および「参加者の状況に応じた支援」が実現される。

- 会話特徴量の可視化

各参加者の位置・顔の向き・発話頻度などの情報をサーバに送信し、サーバ上で情報を統合することで、誰が誰にどの程度話しているのかの履歴を記録する。それにより取得された発話の様子や参加者間の関係性などの会話の特徴量を各参加者に可視化する。

- 参加者の状況に応じた支援

参加者同士の関係性はそれぞれ異なったものであり、頻繁に話している相手とそうでない相手がいる。また、自分から積極的に話す人とあまり積極的には話さない人がある。まだ話し合っていない人に注意を向けるといように、参加者の状況に応じて提示する情報を変化させることで、そうした人たちとも会話をする機会を増やすなどの支援を行う。

4. 実装

本システムはタブレット PC として Sony VAIO Duo 11(Windows 8) を用い、サーバとして一般的なデスクトップ PC(Windows 8) を用いた。また、プログラムは C++ 言語を用いて開発した。

以下、顔の方向の推定、発話頻度の計測、特徴量の可視化についてその実装の詳細を述べる。

4.1 顔の方向の推定

顔の方向の推定のため、タブレット端末の前面カメラで顔検出、背面カメラでマーカ検出を行う。本方式のようにタブレット端末の両面のカメラを用いたシステムとして富岡ら [7] によるものがある。顔検出にはオムロン社の画像センシング技術 OKAO Vision[8] を用い、マーカ検出には ARToolkit[9] を用いた。それぞれの検出処理で得られる座標変換行列および背面カメラから前面カメラへの座標変換行列を合成することで、マーカに対して参加者がどの位置にいるか、どの方向を向いているかを推定する。

ある参加者とその参加者のタブレット端末、基準となるマーカの位置関係を図 3 に示す。ここで、行列 aT_b は a を基準座標系として b の座標系へ変換する同次変換行列である。このとき、背面カメラからマーカへの座標変換行列 ${}^{bc}T_m$ は、背面カメラの映像から検出されたマーカの位置と姿勢を ARToolkit の機能で取得することで求められ、この結果から ${}^mT_{bc} = ({}^{bc}T_m)^{-1}$ を計算する。また、タブレット端末の前背面のカメラはそれぞれ逆向きに備わっており、位置もずれている可能性がある。特定のタブレット端末についてその位置関係は既知であるから、背面カメラから前面カメラへの座標変換行列として ${}^{bc}T_{fc}$ を定めておく。前面カメラから顔への座標変換行列 ${}^{fc}T_f$ は、顔の回転成分と並進成分 (X, Y, Z) から成る。回転成分は OKAO Vision の機能で取得することができる。並進成分は、特定のカメラについて実世界での長さが既知の物体とカメラ映像上での物体の長さ、実世界での物体までの距離とカメラ映像上での距離の関係性をあらかじめ計算しておき、それに基づいて実世界での顔までのおおよその並進量を計算することで求められる (図 4)。実世界での物体の長さが L 、その物体までの距離が D のとき、カメラ映像上での物体の長さを l とすると、カメラ映像上での物体までの距離 d は式 (1) で求められる。この関係性を長さが既知である実世界での人の両目の間隔 $L' \cos \theta$ およびカメラ映像上での両目の間隔 l' に対して適用すると、式 (2) のようにして顔までの奥行 Z が求められる。さらに、画像中の顔の x, y 座標から並進成分 X, Y が式 (3) で求められる。したがって、マーカから顔への座標変換行列 mT_f は式 (4) で得られる。

$$d : D = l : L \Rightarrow d = \frac{Dl}{L} \quad (1)$$

$$d : Z = l' : L' \cos \theta \Rightarrow Z = d \frac{L' \cos \theta}{l'} \quad (2)$$

$$X = x \frac{Z}{d}, \quad Y = y \frac{Z}{d} \quad (3)$$

$${}^mT_f = {}^mT_{bc} {}^{bc}T_{fc} {}^{fc}T_f \quad (4)$$

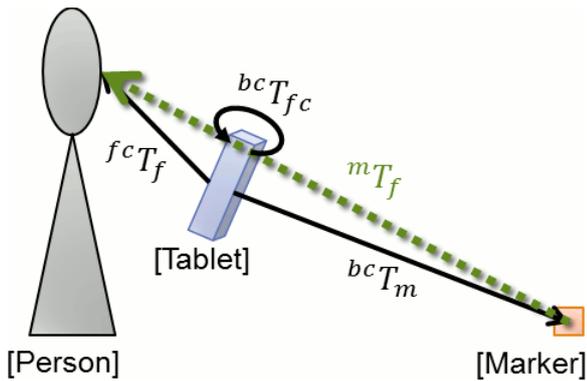


図 3 参加者・タブレット端末・マーカの位置関係
 Fig. 3 The position relation of user-tablet-marker.

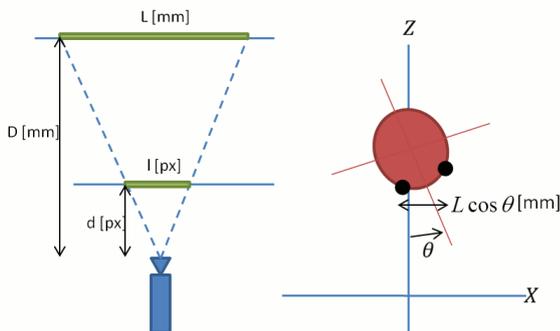


図 4 物体の実サイズと投影サイズの関係に基づく奥行推定
 Fig. 4 Estimation of face depth based on the relationship between the actual size of measured object and the projected size of it.

4.2 発話頻度の計測

参加者がどの程度発話しているか計測する。タブレット端末のマイク入力を利用し、一定以上の音量で音声が発出されたときを発話として扱う。音声を 8 ビットモノラルの 44.1kHz で計測し、1 秒間隔で過去 1 秒間の信号の RMS (二乗平均平方根) を計算する。タブレット端末のマイク音量を最大にして入力レベルの増幅はしない設定において、顔から 30cm 程度離れた状態で発話をしていないときの RMS 値が 3 前後であり、発話をしているときの値がおおむね 10 以上であったことから、後述の今回の実装では 10 を閾値とした。

また、顔の方向の推定機能と組み合わせることで、発話しているときの顔の位置と向きを取得することができる。顔が向いている方向に対して会話していると想定し、視線に最も近いところに位置している参加者を発話の対象者とする (図 5)。各参加者について、その人に対して話しているとき、その人から話しかけられているとき、別の人と話しているその人の様子を見ているときの 3 つの状態が考えられ、加えて誰の方も見ていないという状態 (その他) がある。各参加者とその他について会話の時間のヒストグラムを作成する。例として図 6 に参加者 A の会話のヒストグラムを示す。この図では参加者 B に比べ参加者 C と会

話していることが多く、またどちらの参加者も見えていない時間が 10 秒程度あったことが示されている。

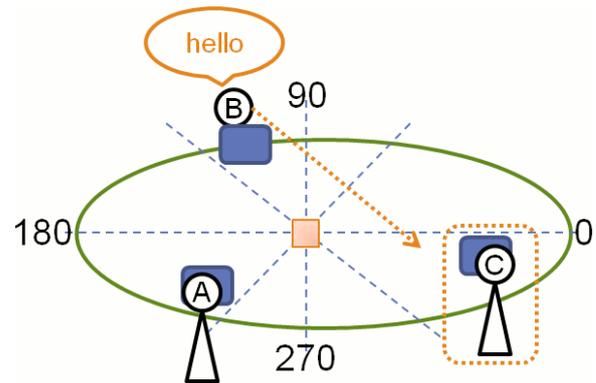


図 5 発話の対象者の推定

Fig. 5 Estimation of the conversational partner.

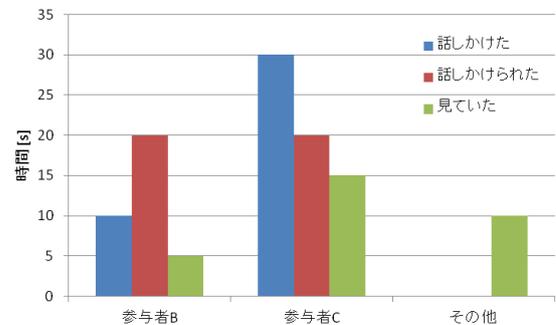


図 6 参加者 A の会話ヒストグラムのイメージ

Fig. 6 An image of the histogram of A's conversation.

4.3 特徴量の可視化

各参加者から取得された顔の向きや発話頻度をもとに会話特徴量の可視化を行う。クライアント・サーバ方式の通信を行い、各タブレット端末では取得した特徴量データをサーバに送信する。サーバ側では各参加者の位置と顔の向き、発話の状態から誰が誰に話しているのかといった全体で共有される情報を作成し、参加者全員に送信する。また、ある参加者とある参加者の関係性など個別の情報に関しては、それぞれの参加者に送信する。サーバには Node.js[10] を用い、通信ライブラリには Socket.IO[11] を用いる。可視化の手順を以下に示す。

- (1) 参加者の位置や顔向き、発話頻度や強度などの特徴量データを取得
- (2) 特徴量データをサーバへ送信
- (3) サーバで各端末からのデータを共有し参加者間の関係性を求める
- (4) 関係性に基づく情報を各端末へ送信
- (5) 受信データに基づいて可視化

5. 発話方向の推定実験

5.1 実験条件

開発したプロトタイプシステムを図 7, 図 8 に示す. これは誰に対して話しているかを求める前段階として, マーカ基準でどの方向に向かって発話しているかを推定し, マーカの周囲を 8 等分した領域に対する発話頻度のヒストグラムを計測するシステムである. ヒストグラムは画面の右上に提示される. 発話方向の推定および発話頻度の計測機能について, 発話状態の検出とそのときの顔の向きが正しく取得できているか調べる. 円形のテーブルがあり参加者がそれを囲んで会話している場面で, 発話者がマーカ基準で 270 度の方向にいる状態を想定する. この状態でマーカ基準で 45 度と 100 度の方向の対象に発話する場合 (図 9) と, 同じく 0 度方向と 90 度方向の対象に発話する場合 (図 10) の 2 通りの状況について, 発話方向の推定実験を行った. それぞれの状況で各対象について 10 秒程度発話を行うものとする.

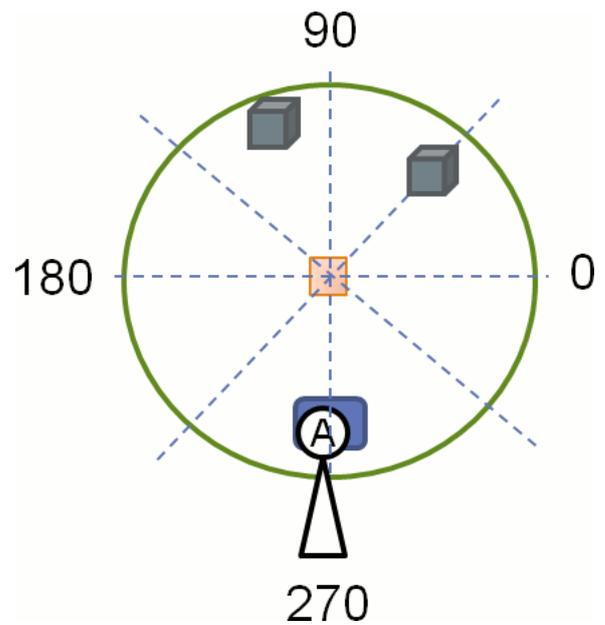


図 9 マーカ基準で 45 度と 100 度の方向の対象に発話する場合
Fig. 9 An example of a user's speaking in the direction of 45 degrees and 100 degrees from the marker.



図 7 ユーザによるプロトタイプシステムの使用例

Fig. 7 An example of a user's operating the prototype system.

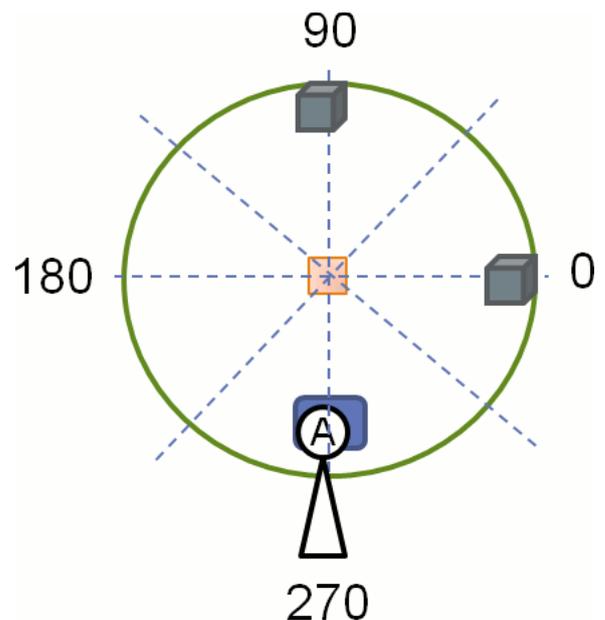


図 10 マーカ基準で 0 度と 90 度の方向の対象に発話する場合
Fig. 10 An example of a user's speaking in the direction of 0 degree and 90 degrees from the marker.

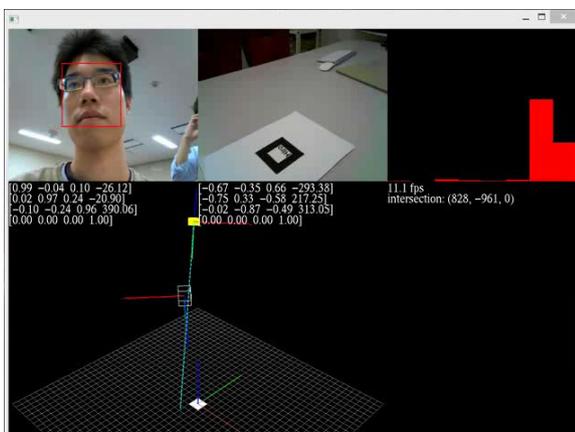


図 8 動作中のプロトタイプシステムの画面例

Fig. 8 An example of the screen of the prototype system.

5.2 実験結果

マーカ基準で 45 度と 100 度の方向の対象に発話した場合の結果を図 11 に示す. 発話頻度のヒストグラムにはマーカ基準で 0-45 度, 45-90 度, 90-135 度の方向に発話している様子が記録されており, 正しく発話方向を推定できていることがわかる. 次にマーカ基準で 0 度方向と 90 度方向の対象に発話した場合の結果を図 12 に示す. 発話頻度のヒストグラムにはマーカ基準で 45-90 度, 90-135 度の方向に発話している様子が正しく推定できているが, 0 度

方向についてはほとんど記録されておらず正しく推定できていない。これはタブレットのカメラでマーカと顔の両方が検出されていないと発話方向が推定できないことが原因である。ある程度の範囲はタブレットや体の向きを変えることで対応できるが、今回の条件では顔の向きまたはマーカの検出が難しく、そのため0度方向の対象についてはうまく視線推定ができなかったからであると考えられる。

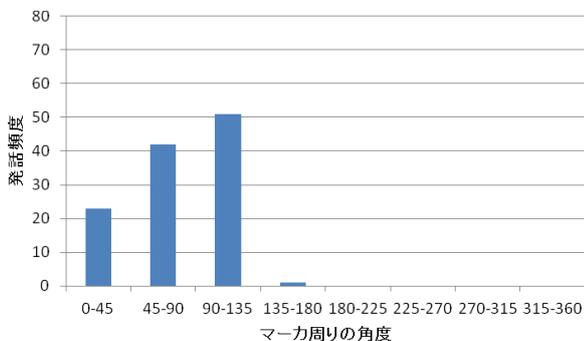


図 11 マーカ基準で 45 度と 100 度の方向の対象に発話したときのヒストグラム

Fig. 11 The histogram of utterances frequency of the user's speaking in the direction of 45 degrees and 100 degrees from the marker.

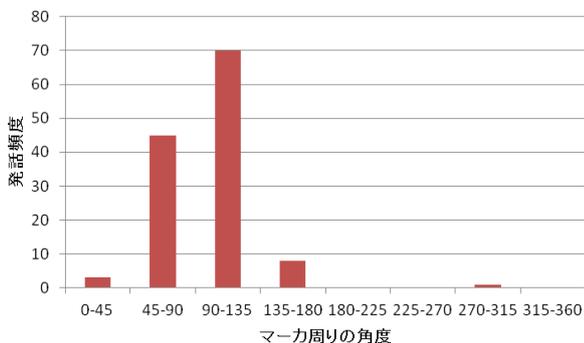


図 12 マーカ基準で 0 度と 90 度の方向の対象に発話したときのヒストグラム

Fig. 12 The histogram of utterances frequency of the user's speaking in the direction of 0 degree and 90 degrees from the marker.

6. おわりに

本稿では、タブレットの前背面カメラを利用して発話対象者や発話頻度などの会話特徴量を取得し、ARを用いて各参加者向けに調整したうえで提示することで、コミュニケーションの活性化を支援するシステムを提案した。プロトタイプとして顔の方向と発話頻度を推定し、発話方向のヒストグラムを計測するシステムを実装し実験を行った。その結果、正しくヒストグラムが得られた場合と顔の検出に失敗したなどの原因で正しく得られなかった場合があった。今後は、今回の問題点の改善を行うことに加えて、カ

メラ映像中へのARによる可視化や複数人への対応、参加者の支援機能を盛り込み、実際の多人数会話の場において評価実験を行う予定である。

参考文献

- [1] いまさら聞けない AR (拡張現実) の基礎知識, <http://www.atmarkit.co.jp/ait/articles/1109/26/news136.html>, 2013/11/17 閲覧
- [2] Sekai Camera Web, <http://sekaicamera.com/>, 2013/11/17 閲覧
- [3] 松原孝志, 白杵正郎, 杉山公造, 西本一志, 言い訳オブジェクトとサイバー囲炉裏:共有インフォーマル空間におけるコミュニケーションを触発するメディアの提案, 情報処理学会論文誌, Vol.44 No.12, pp.3174-3187, 2003
- [4] 木村篤信, 井原雅行, 小林稔, 発話音量制御支援のための視覚フィードバック提示方法の検討, 電子情報通信学会論文誌, Vol.J89-D No.10, pp.2183-2193, 2006
- [5] Igor de Souza Almeida, Jun Miyazaki, Goshiro Yamamoto, Makoto Fujisawa, Toshiyuki Amano, Hirokazu Kato, AR based Co-located Meeting Support System, The 21st international Conference on Artificial Reality and Telexistence, November 2011
- [6] 大島直樹, 岡澤航平, 本田裕昭, 岡田美智男, TableTalk-Plus: 参加者の共同性や社会的なつながりを引き出すアーティファクトとその効果, ヒューマンインタフェース学会論文誌, Vol.11 No.1, pp.105-114, 2009
- [7] 富岡誠, 池田聖, 佐藤宏介, カメラ内蔵タブレット型拡張現実感のための Homography 変換を用いた利用者視点画像生成, 第 16 回画像の認識・理解シンポジウム (MIRU), SS4-25, July 2013
- [8] OKAO Vision — OMRON Global, http://www.omron.com/r_d/coretech/vision/okao.html, 2013/11/17 閲覧
- [9] ARToolKit Home Page, <http://www.hitl.washington.edu/artoolkit/>, 2013/11/17 閲覧
- [10] node.js, <http://nodejs.org/>, 2013/11/17 閲覧
- [11] Socket.IO: the cross-browser WebSocket for realtime apps., <http://socket.io/>, 2013/11/17 閲覧