

参加者のインタラクションを可視化したマルチメディア議事録の作成

横江 優貴[†] 伊藤 義道[†] 馬場口 登[†]

[†] 大阪大学大学院工学研究科 〒 565-0871 大阪府吹田市山田丘 2-1
E-mail: †{yokoe,ito,babaguchi}@nanase.comm.eng.osaka-u.ac.jp

あらまし マルチメディア議事録とは、映像と音声から構成される会議環境の記録であり、発言者の表情や参加者の反応など人同士のコミュニケーションの様子を忠実に記録する媒体として近年注目を集めている。本稿では、このような媒体の一つである MMLog (Multimedia Log) を、利用者にとってより使い勝手の良いものに改良することを目的とし、会議環境記録システムである MMLogger (Multimedia Logger) に、発話区間のタイムライン表示機能や参加者同士のインタラクションの様子を可視化する機能を新たに組み込んだシステムを提案する。本稿ではとくに、映像情報と音声情報を利用して参加者の対話の様子を可視化する手法について述べ、その有効性を実験をとおして検証する。**キーワード** マルチメディア議事録, インタラクションの可視化, 話者方向推定, 音声/映像解析

Multimedia Meeting Logs Visualizing Participants' Interactions

Yuki YOKOE[†], Yoshimichi ITO[†], and Noboru BABAGUCHI[†]

[†] Graduate School of Engineering Osaka University, Yamadaoka 2-1, suitea-shi, 565-0871 Japan
E-mail: †{yokoe,ito,babaguchi}@nanase.comm.eng.osaka-u.ac.jp

Abstract Multimedia meeting log is a record of meeting environment, which consists of video and audio data of meeting participants, and has received a lot of attention as a useful media for making a record of vivid activity of human communication. In this report, we propose a new framework of MMLog (Multimedia Log: a realization of a multimedia meeting log), by extending the functions of existing MMLogger (Multimedia Logger: a recording system that generates MMLogs). Among these functions, this report mainly focuses on a method for visualizing the interactions of meeting participants using video and audio data, evaluating the method through experiments.

Key words Multimedia Meeting Logs, Visualization of Participants' Interactions, Speaker Localization, Audio/Video Analysis

1. はじめに

近年、会議環境の記録として、従来の紙を媒体とした記録に代わり、映像や音声を主体とした記録に対する関心が高まっており、様々な研究がなされている [1]~[8]。Chiu ら [3] は、大規模で特殊な撮影環境を構築することにより、参加者の映像だけでなく、会議に使う資料や周囲の様子も記録し、記録の編集・閲覧の際に利用する方法を提案している。Nagao ら [5] は、事前に参加者に特定の行動を取らせるといった制約を設けることによって、より詳細な情報を映像に付加する方法を提案している。また、これらの研究とは異なり、大規模な設備を必要とせず、コストや可搬性の面で優れた小型のシステムも開発されている [1], [6]~[8]。Tokunaga ら [6] は、少人数による会議環境を想定して、前方位センサとマイクロフォンアレーからなる会議環境の記録システム MMLogger (Multimedia Logger) を開発し、可搬性が高くかつ、特別な事前準備を必要としないシス

テムを実現している。また、その記録を閲覧するための議事録 MMLog (Multimedia Log) も開発されている。

これらの会議環境記録システムは、いずれも音声・映像記録を用いることによって、発言内容のみを記録した従来の紙を媒体とした議事録では決して知ることができない話者の様子や表情など、会議参加者のコミュニケーションの様子を知ることができる。しかし、既存研究の多くは会議環境の記録における話者の様子やその発言内容の解析を主な対象としており、話者だけでなく他の参加者を含めた会議環境の様子を、閲覧者に対して視覚的に分かり易く提示することを主目的とする研究はあまりなされていない。閲覧者にとっては、例えば、会議全体の流れ、議論に積極的に参加しているのは誰か、誰と誰が議論しているか、会議における主要な人物は誰か、といった情報も重要であると考えられる。そこで本稿ではまず、Tokunaga ら [6] によって提案された MMLogger の機能を拡張することにより、これらの情報を分かり易く提示するための新たな MMLogger



図 1 MMLog

を提案する。

以上に述べた機能を MMLogger に実装する際、技術的に最も困難であると考えられるのは、誰と誰が議論しているかをシステムが推定することである。MMLogger で利用している全方位センサは、話者だけでなく対話相手も同時に撮影することが可能であるが、解像度の低い全方位映像のみで話者が誰に向かって話しているかを正確に判断することは困難である。そこで、本システムでは、映像情報からは、正面・左・右のおおまかな顔の向きを推定し、この結果をマイクロフォンアレーによって得られた過去の話者方向推定の結果と照合することにより対話者を推定する手法を採る。

話者の顔方向は、これまで目や口唇といった部位の画像中の位置から推定する手法 [9]~[11] や、あらかじめ、それぞれの角度の顔画像を学習サンプルとして撮影しておき、最も近いものを選ぶ方法 [12] などが提案されている。しかし、画像の解像度が低いという全方位センサの特性から、顔の細かな特徴を必要とする前者の手法では正確な推定ができない。また、会議の前にサンプル画像の撮影が必要な後者の手法は、特別な事前準備を必要としないシステムの実現を目指す本システムには適さない。そこで、本稿では顔の中でも検出が容易で個人による差異の小さい口唇に着目し、口唇周辺領域における画素値の差分を求め、その左右の偏りによって顔が正面・左・右のいずれを向いているかを推定する手法を提案する。

本稿の構成は以下のとおりである。まず、第 2 節において、会議環境の記録システムである MMLogger および閲覧者に表示するための表現形式である MMLog について述べる。次に、第 3 節において、本システムで用いるマイクロフォンアレーを利用した話者方向推定手法について簡単に述べる。第 4 節では、口唇に着目した顔方向を推定する手法と発話履歴に基づく対話相手の推定手法について述べる。第 5 節ではその手法を実験をとおして検証する。第 6 節において、得られた結果をまとめる。

2. MMLog と MMLogger

本節では、提案するマルチメディア議事録 MMLog と、会議環境を記録し MMLog を作成する MMLogger について述べる。

2.1 MMLog

MMLog (Multimedia Log) は、映像と音声からなる会議環境の記録であり、発言者の表情や参加者の反応など、人同士の

活き活きとしたコミュニケーションの様子を忠実に記録するための媒体である [6]。図 1 に、本システムで作成した MMLog を示す。図 1 に示す MMLog には、単に話者や周囲の状況の映像が表示されるだけでなく、以下に述べるような表現形式も追加されており、閲覧者にとってより使い勝手のよい記録となっている。

① 話者の正面映像：全方位映像から話者方向推定の結果に基づいて抽出した話者の正面映像が表示される。話者方向推定の手法については第 3 節で述べる。

② パノラマ映像：全方位映像から作成したパノラマ映像が表示されたものであり、話者の様子だけでなく、他の参加者の反応も閲覧することができる。

③ 対話者/注目したい人物の映像：話者の顔方向および、発話の履歴から推定した対話相手が表示される。これにより、従来の記録では得られなかった話者とその対話相手の間で交わされるインタラクションが忠実に再現される。顔方向および、対話相手の推定の方法については第 4 節で述べる。また、注目したい人物がいる場合には、パノラマ映像中におけるその人物領域をクリックすることにより③の領域に拡大表示される。

④ 発話区間タイムライン：記録しておいた話者方向推定結果の履歴を基に表示される。発話区間のタイムラインは、利用者が一見して誰がいつ発言したかを知ることができるだけでなく、会議の大まかな流れを知る上でも有用である。

⑤ 参加者のインタラクション表示図：図中で示している円の位置と大きさはそれぞれ会議環境における各参加者の位置と発言の多さを表しており、①、②、③の映像だけでは分かりにくい会議環境の状況を俯瞰的に把握することができる。また、矢印は誰と誰が対話しているかを示すものであり、会話の中心人物は誰か、頻繁に議論を交わしているのは誰かといった参加者同士のインタラクションを知る上で有用である。

⑥ アノテーション記述スペース：会議の状況を議題ごとに構造化して記述するためのスペースであり、閲覧者が自由にアノテーションを書き込むことができる。また、各アノテーションには対応する映像とタイムスタンプを割り当てることができるため、興味のある場面でアノテーションを付与しておけば、いつでもその場面にスキップすることができる。



図2 MMLoggerのセンサ部

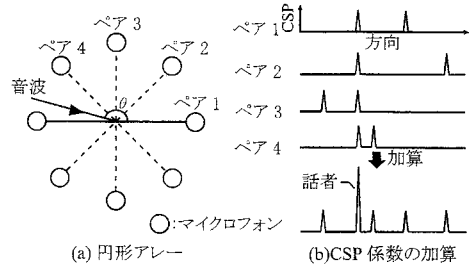


図4 円形アレーによるCSP係数の加算

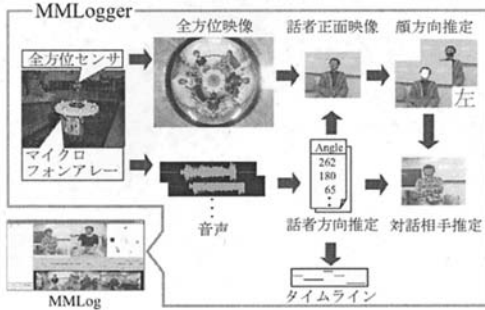


図3 MMLoggerを用いたMMLogの作成

2.2 MMLogger

MMLogger [6]~[8] は会議環境の記録システムであり、記録した映像・音声情報から MMLog を作成する。そのセンサ部(図2参照)の中央には全方位センサを設置し、そのまわりを囲むように8本のマイクを同一円周上に等間隔で配置する。全方位センサは周囲360度の映像を撮影することが可能であり、これを参加者の中央に配置することにより、周囲の映像情報を全て1台のカメラで取得することができる。このため、複数のカメラを必要とせず、可搬性の高いシステムを実現できる。また、取得した映像に対して処理を施すことにより、会議環境全体を表示したパノラマ映像や、全ての参加者の正面映像を作成することができる。また、8本のマイクで構成されるマイクロフォンアレーは、会議中の発言内容を記録するだけでなく、得られた音声データを基に話者方向を推定することが可能であり、その情報は話者の正面映像の作成や、参加者の座席位置の推定、およびタイムラインの作成に利用する。

MMLogger は以上の情報を統合することによって MMLog を作成する。図3に、MMLogger を用いて記録した情報から MMLog を作成する流れを示す。

3. 話者方向推定

本節では、全方位映像から話者の正面映像を抽出する際や、発言履歴から対話相手を推定する際に必要な、マイクロフォンアレーによる話者方向の推定手法について述べる。

本システムで用いる話者方向推定手法はCSP(crosspower spectrum phase)法[13]に基づいている。CSP法は、2本のマ

イクロフォンに音波が到来したときに生じる到来時間差に基づいて音源方向を推定する方法である。単一音源からマイクロフォン i, j に到来した信号を、それぞれ $s_i(n), s_j(n)$ とすると、マイクを結ぶ直線と音声の到来方向のなす角度 θ は以下の式を順に計算することによって求めることができる。

$$S_i(f) = \text{DFT}[s_i(n)], \quad (1)$$

$$\text{CSP}_{i,j}(n) = \text{IDFT} \left[\frac{S_i(f)S_j^*(f)}{|S_i(f)||S_j(f)|} \right], \quad (2)$$

$$\tau = \arg \max_n (\text{CSP}_{i,j}(n)), \quad (3)$$

$$\theta = \cos^{-1} \left(\frac{c \cdot \tau / f_s}{d} \right). \quad (4)$$

ただし、DFTは離散フーリエ変換、IDFTは逆離散フーリエ変換であり、 $\text{CSP}_{i,j}$ は信号 $s_i(n)$ と $s_j(n)$ から得られるCSP係数を表す。また、*は複素共役を意味し、 c, f_s, d はそれぞれ、音速、サンプリング周波数、マイクロフォンの間隔を表す。到来時間差 τ は式(3)に基づき、CSP係数が最大値を取る n 、すなわち、二つの信号の相関が最も大きい n として求められる。この値を式(4)に代入することにより話者方向 θ を推定する。

複数の参加者が同時に発言する場合、複数の音声間の干渉の影響で正しく推定できない問題が生じる。また、式(4)は多価関数である逆余弦関数を含んでいるため、前節で述べた方法は周囲360度の話者方向推定に適用した場合、方向を一意に定めることができないという問題が生じる。そこで、本システムでは、図4(a)に示すように8本のマイクロフォンを同一円周上に等間隔で配置し、4つのペアで求めたCSP係数の加算に基づくYokoeらの手法[8]を採用する。これにより、図4(b)に示すように、真の話者方向に対するCSP係数のみが強調され、複数の音声間の干渉の影響や、逆余弦関数が多価関数であるために生じる架空のピークの影響も抑えることができる。

4. 対話相手の推定

MMLogにおいて参加者同士のインタラクションを可視化するために、話者の話の受け手となる対話相手を推定する必要がある。MMLoggerは可搬性が高く、事前に特別な準備を必要としないシステムを目指しているため、複雑な処理や、事前の学習が必要な手法は適さない。そこで、本節では映像と音声の情報を協調的に用いることにより、対話相手を推定するための効

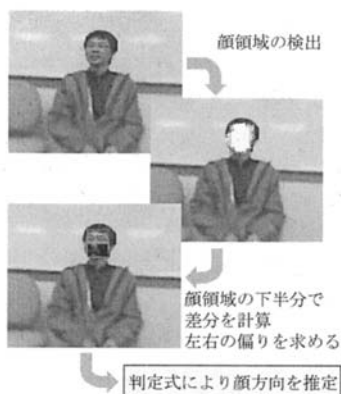


図5 顔方向の推定

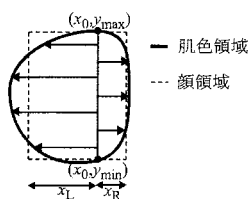


図6 顔領域の抽出

率的で簡明な手法を提案する。

4.1 顔方向の推定

本システムでは、会議参加者が MMLogger のセンサを取り囲むように座っている状況を想定し、顔方向を正面・左・右の3種類で分類する。なお、本システムにおいては、話者が正対したときの顔方向を基準としたとき、そこから話者を中心とする ± 20 度の範囲を正面と定義する。また、正面より右側の範囲を右、左側の範囲を左と定義する。以下では、これらの定義に基づく顔方向を推定する手法について述べる。

顔方向を推定する手順を図5に示す。まず、顔方向を推定するために話者の顔領域を抽出する。撮影した全方位画像から切り出した話者の正面画像を RGB 表色系から YCbCr 表色系へ変換し、Cb、Cr の値に基づいて肌色領域を抽出する。肌色領域から顔領域を抽出する方法を図6に示す。まず、抽出した肌色領域に含まれる縦方向の長さが最大となる線分を探索する。この線分上の端点の座標を (x_0, y_{\min}) 、 (x_0, y_{\max}) とする。続いて、その線分の左側にある肌色領域の境界上の各点から線分までの距離を求め、その平均 x_L を求める。同様に、右側の境界から線分までの距離の平均 x_R を求める。顔領域は、 $(x_0 - x_L, y_{\min})$ 、 $(x_0 - x_L, y_{\max})$ 、 $(x_0 + x_R, y_{\max})$ 、 $(x_0 + x_R, y_{\min})$ の4点で囲まれる矩形領域として定義する。

次に、抽出した顔領域から顔方向を推定する。顔方向は、顔領域における目や口唇等の位置によって推定できるが、全方位画像から生成した正面画像は解像度が低いため、目や口唇の明確な位置を取得することができない。そこで本手法では、検出

が容易で個人差の少ない口唇に着目し、その左右の偏りを求めることによって顔方向を推定する。口唇の領域が顔領域の左上と右下に占める割合は、顔方向によって異なること、および、口唇は縦方向の差分フィルタによって容易に検出できることを利用し、顔方向は以下のように推定する。

まず、顔領域を縦方向と横方向ともに2分割し、4つの領域に等分割する。続いて、下半分の左側および右側の2つの領域の画素をグレースケールに変換した後、それらの領域に対して各点における縦方向の差分値を計算し、それら差分値の絶対値の和を求める。なお、上半分を使用しないのは、髪や眼鏡などの複雑な形状が推定結果に影響しないようにするためである。左側、および右側の和を、それぞれ D_L 、 D_R とする。このとき、顔方向を以下の条件式に基づいて判定する。

$$\frac{D_L}{D_R + D_L} > \alpha \Rightarrow \text{右} \quad (5)$$

$$\frac{D_R}{D_R + D_L} > \alpha \Rightarrow \text{左} \quad (6)$$

$$\text{otherwise} \Rightarrow \text{正面} \quad (7)$$

ただし、 α は判定のための閾値である。

条件式 (5)、(6) における左辺の条件は、 D_L あるいは D_R が $D_L + D_R$ に占める比率を表すため、 $\alpha < 1$ である必要がある。また、条件式 (5)、(6) が同時に成立することを防ぐために $\alpha > 0.5$ とする。したがって、判定のための閾値 α は $0.5 < \alpha < 1$ の範囲で選ぶ必要がある。また、条件式 (5)~(7) より、 α を大きくすると、左もしくは右と判定される範囲が狭くなり、正面と判定される範囲が広くなるのがわかる。逆に、 α を小さくすると、左もしくは右と判定される範囲が広くなり、正面と判定される範囲が狭くなる。判定のための閾値 α の値は第5節の実験結果に基づいて決定する。

4.2 顔方向と発話履歴に基づく対話相手の推定

本節では、前節で述べた顔方向の推定結果を基に対話相手を推定するアルゴリズムについて説明する。ここでは、参加者の1人が発話を開始し終了するまでの間を発話区間とする。MMLogger は、1つの発話区間における対話相手は1人のみとし、発話区間ごとに対話相手を推定する。

まず、発話区間の各フレームに対して前節で述べた手法を用いて顔方向を推定する。次に、発話区間において最も多く推定された顔方向をその発話区間における顔方向とする。話者の位置から見て顔方向に参加者が存在しない場合、その発話区間での対話相手はないものとみなす。顔方向に参加者が一名だけ存在する場合、その参加者をその発話区間での対話相手とする。推定した顔方向に参加者が複数存在する場合、発話者の履歴を現在から順に過去に遡って探索し、顔方向に位置する参加者の中で最初に検出した参加者をその発話区間における対話相手とする。

5. 実験と評価

本節では、まず実験をととして第4.1節で述べた顔方向の判定に必要な閾値 α を決定する。次に、第4.2節で述べた対話相手の推定手法の精度を実験により評価する。以下の実験におけ

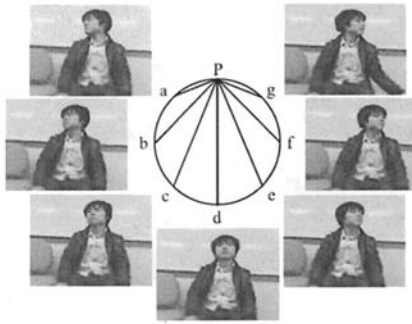


図7 顔方向推定のための実験環境

る映像のフレームレートは 30fps、音声のサンプリング周波数は 48kHz である。また、全方位映像から話者の正面映像を抽出する際の視野角は 40 度に設定する。

5.1 顔方向推定のための閾値の決定

第 4.1 節で述べた顔方向の定義に基づき、顔方向の判定に適した判定閾値 α を実験により求める。図 7 に示すように半径 1m の円周を八等分した点 P, a~g をとり、円の中心に MMLogger のセンサを設置する。点 P には椅子に座った被験者が位置し、点 a~g を向いた映像を撮影する。なお、点 a~g を向いた際の顔の傾きは、点 d で 0 度、点 c, e で 22.5 度、点 b, f で 45 度、点 a, g で 67.5 度となる。実験では 3 名の被験者に対し、点 a~g を向いた画像を、それぞれの被験者と方向につき 500 フレームずつ取得した。その後、判定閾値 α を 0.55 から 0.95 まで 0.05 ずつ変化させながら提案した手法を用いて顔方向を判定し、その結果を顔方向の定義に基づき正解率を求めた。なお、顔の傾きが 22.5 度、45 度、67.5 度での正解率は、それぞれ点 c と e、点 b と f、点 a と g の平均値としている。得られた正解率を図 8 に示す。

図 8 より、顔が左右を向いている 22.5, 45, 67.5 度では、 α の値を大きくするほど正解率が減少していく傾向が見られる。これは、第 4.1 節で述べたとおり、 α が大きくなるほど、正面と判定される範囲が広くなるためである。傾きが 45 度の場合と比較して、22.5 度で正解率が低くなる傾向があるのは顔の傾きが小さいため正面と誤って判定されるためであるが、67.5 度で低下しているのは、傾きが大きいため口唇が顔の端に寄り、口唇の一部が顔領域から外れてしまったことが原因であると考えられる。

正面と定義される範囲で最も正解率が高くなると考えられる傾き 0 度でも、 α が 0.55 のときは正解率が低くなっており、正面と定義される範囲での推定が正確でないことがわかる。逆に、0.6 のときは 1 に近い正解率が得られ、0.65 以上ではほとんど差が認められない。 α が 0.65 以上の場合、顔が左右を向いている場合の正解率が低下するため、判定のための閾値 α は 0.6 が適しているといえる。なお、閾値 α を 0.6 に設定したときの正解率の平均は 0.83 である。

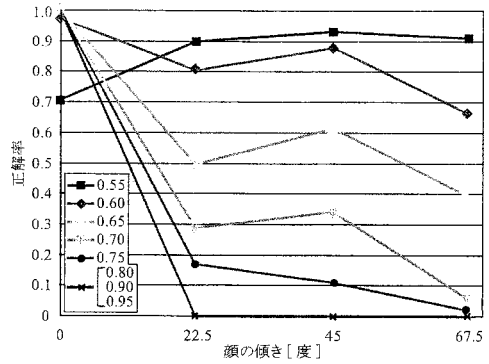


図8 判定閾値 α に対する正解率

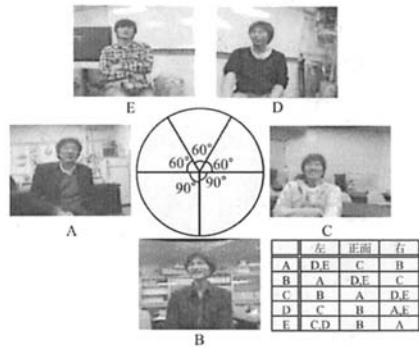


図9 対話相手の推定のための実験環境

表1 顔方向および対話相手の推定結果

		判定した数	正	誤	正解率
対話相手の推定	顔方向推定	76	60	16	0.79
	顔方向推定: 正 顔方向に参加者が1名	30	30	0	1.00
	顔方向推定: 正 顔方向に参加者が複数	30	28	2	0.93
	顔方向推定: 誤	16	0	16	0.00
	計	76	58	18	0.76

5.2 対話相手の推定実験

本実験では、記録した会話に対して提案手法を用いて各発話区間での対話相手を推定し、その有効性を検証する。本実験に用いた記録は会話時間が約 10 分 (17724 フレーム)、参加者は 5 名で、図 9 に示すように半径 1m の円周上に座っている。図 9 中の表は、顔方向の定義に基づき、それぞれの参加者に対して他の参加者がどの方向に位置しているかを示している。

会話中の笑い声やあいつちといった、そのみでは意味を持たない発話を除くため、2 秒以上の発話のみを抽出した。その結果、抽出されたフレーム数は 8024 フレーム、発話区間の数は 76 となった。これら 76 の発話区間に対し、顔方向および対話相手を推定した結果を表 1 に示す。なお、顔方向の判定のための閾値 α は 0.6 とした。

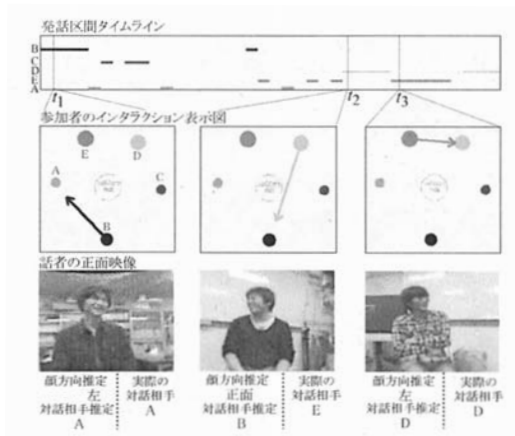


図 10 参加者のインタラクション表示図の動作

顔方向推定の正解率は 0.79 であった。この結果は前節の実験で $\alpha = 0.6$ としたときの正解率 0.83 と同程度である。顔方向推定が正しかった場合、その方向に参加者が 1 人だけ存在する場合の対話相手の推定結果は全て正解となった。一方、顔方向推定を誤った場合、対話相手の推定は全て誤りとなった。また、顔方向推定が正しかった場合、その方向に参加者が複数存在する場合の対話相手の推定は正解率が 0.93 となった。この結果は、提案した発話履歴に基づく対話相手の推定手法が有効であることを意味している。結果として得られた対話相手の推定の正解率は 0.76 であるが、この結果は顔方向推定の正解率がほぼそのままの形で影響したものと考えることができる。しかし、発話中の首振り動作を認めた実環境での結果としては、十分有効であるといえる。

本実験で得られた推定結果を基に作成された発話区間タイムラインの一部と、時間の経過とともに参加者のインタラクション表示図と話者の正面映像が変化していく様子を、顔方向および対話相手の推定結果とともに図 10 に示す。同図より、推定した対話相手に対して話者から矢印が伸びていることが確認できる。また、参加者を示す円が発話区間の長さに応じて大きくなっている。これらにより、映像や音声だけではイメージし難い参加者間のインタラクションの様子が、直感的に分かりやすい形で提示されていることがわかる。

図 10 におけるタイムライン上の時刻 t_1 , t_2 , t_3 における顔方向、および対話相手の推定結果に関しては、以下のことがいえる。時刻 t_1 では、顔方向が正確に推定され、その方向には他の参加者が 1 人しか存在しなかったため、正確に対話相手を推定できている。時刻 t_2 では、実際は右を向いているが正面と誤って推定されたため、本来の対話相手である E ではなく正面に位置する B を指している。時刻 t_3 では、推定された顔方向には 2 人の参加者が存在しているが、発話の履歴から直前に発話している D が対話相手と正しく推定できている。

6. まとめ

本稿では、会議環境において参加者同士の活き活きとしたコミュニケーションの様子を閲覧者に分かりやすく提示する機能を有する MMLogger を提案した。これによって作成された MMLog は、従来の記録とは異なり、発話区間のタイムライン表示や参加者同士のインタラクションの可視化により、閲覧者にとってより使い勝手のよい記録となっている。

また、本稿では、対話の状況を可視化するために、対話相手の推定手法を提案した。本手法は、口唇の位置に基づく顔方向推定と発話履歴に基づく推定手法を協調的に用いている。これにより、解像度の低い全方位映像でも十分な正解率が得られることを実験によって示した。また、実験により、顔方向推定の正解率が対話相手の推定結果に大きく依存することを示した。

今後の課題として、顔方向推定の精度の向上が挙げられる。具体的には、顔領域検出手法の改善により、顔方向推定の精度向上を図る。また、顔方向の推定手法を話者だけでなく他の参加者にも適用することにより、参加者の注目対象を推定するといった応用技術の実装も今後の課題となる。

文献

- [1] D. Lee, B. Erol, J. Graham, J.J. Hull, and N. Murata, "Portable meeting recorder," Proc. of ACM Multimedia 2002, pp.493-502, 2002.
- [2] B. Erol, D. Lee, and J.J. Hull, "Multimodal summarization of meeting recordings," Proc. of ICME2003, pp.25-28, 2003.
- [3] P. Chiu, A. Kapurkar, S. Reitmeyer, and L. Wilcox, "Meeting capture in a media enriched conference room," Proc. of Second International Workshop on Cooperative Buildings, pp.79-88, 1999.
- [4] W. Geyer, H. Richter, and G.D. Abowd, "Making multimedia meeting records more meaningful," Proc. of ICME2003, pp.669-672, 2003.
- [5] K. Nagao, K. Kaji, D. Yamamoto, and H. Tomobe, "Discussion mining: Annotation-based knowledge discovery from real world activities," Proc. of PCM2004, pp.522-531, 2004.
- [6] S. Tokunaga, Y. Ito, N. Nitta, and N. Babaguchi, "Meeting recording system via multimodal sensing," Proc. of JSAI Workshop on Conversational Informatics, pp.19-24, 2005.
- [7] 横江優貴, 伊藤義道, 馬場口登, "マルチメディアログ作成のための音声による話者方向推定," 画像の認識・理解シンポジウム (MIRU2006), pp.1138-1143, 2006.
- [8] Y. Yokoe, Y. Ito, and N. Babaguchi, "Audio-based estimation of speakers directions for multimedia meeting logs," Proc. of ICME2007, pp.212-215, 2007.
- [9] Y. Nakanishi, T. Fujii, K. Kitajima, and H. Koike, "Vision-based face tracking system for large displays," Proc. of ACM Ubicomp2002, pp.123-160, 2002.
- [10] 足立佳久, 中西知, 久野義徳, "顔の視覚情報処理を用いた知的車椅子," 日本ロボット学会誌, vol.17, no.3, pp.423-431, 1999.
- [11] 北島光太郎, 佐藤祥一, 小池英樹, "ウィンドウインタフェースシステムにおける頭部の位置・姿勢の有効な利用法," 情報処理学会インタラクション論文集, pp.195-202, 2001.
- [12] 大江展弘, 志村将吾, 間瀬健二, "全方位会議記録映像における注目行動判定による記録閲覧支援," 画像の認識・理解シンポジウム (MIRU2005), pp.575-582, 2005.
- [13] C. Knapp, and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol.24, no.4, pp.320-327, 1976.