一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2012-167(2013-2)

［招待講演］ スマートポスターボード：ポスター発表における場のマルチモーダルなセンシングと認識

河原　達也†

† 京都大学 学術情報メディアセンター
〒 606-8501 京都市左京区吉田本町
E-mail: †kawahara@i.kyoto-u.ac.jp, http://www.ar.media.kyoto-u.ac.jp/crest/

あらまし　学会等で一般的に行われているポスター発表における会話（ポスター会話）は、マルチモーダルな多人数インタラクションに関する様々な興味深い研究テーマを提供してくれる。本稿では、著者らが進めているポスター会話のマルチモーダルなセンシング・分析・認識に関するプロジェクト、及び構築しているシステムの概要を紹介する。我々は特に、視線配布や相槌などの聴衆の反応（聞き手行動）に着目し、発話パターンの予測・検出を行い、さらに興味・理解度を推定することを目指している。そのために、ポスター会話をマルチモーダルに収録した上で、会話参加者の行動を捉え、会話の場を視覚化するシステム「スマートポスターボード」を設計・実装している。
キーワード　マルチモーダル, 行動信号処理, 会話分析, 話者交替, ポスターボード , マイクロフォンアレイ

[Invited Talk] Smart Posterboard: Multi-modal Sensing and Recognition of Poster Presentation

Tatsuya KAWAHARA†

† Kyoto University, Academic Center for Computing and Media Studies
Sakyo-ku, Kyoto 606-8501, Japan
E-mail: †kawahara@i.kyoto-u.ac.jp, http://www.ar.media.kyoto-u.ac.jp/crest/

1. はじめに

マルチモーダルな情報処理は、主としてヒューマンマシンインターフェースの高度化を目指して行われてきた。これは携帯端末やキオスク端末等のディスプレイだけでなく、人間型ロボットも含む。一方、画像処理や音声処理が高度になり、上記のようなインターフェースを意識しない人間の自然なふるまいも扱えるようになり、いわゆるアンビエントなシステムを目指した研究開発も可能になっている。実際に、ミーティング [1] や自由会話 [2] などの人間どうしの会話を対象とした研究も行われている。

我々はポスターセッションにおける会話（＝「ポスター会話」）に焦点をおいたプロジェクトを進めている [3], [4]。ポスターセッションは、学会やオープンラボなどで一般的になっているが、未だに情報通信技術 (ICT) の導入がほとんどなされておらず、ICT 分野の会議でも紙のポスターを用いることが多い。一部液晶ディスプレイや携帯プロジェクタを用いる場合もあるが、センサを備えた環境は世界的にも前例がないと思われる。講演や講義の映像・音声が収録・配信されることが一般的になっているのに対して、ポスターセッションを収録して分析した研究も皆無に近い。

ポスター会話は、講演と会議の中間的な形態と捉えることができる。すなわち、発表者が自身の研究内容について少人数の聴衆に説明する一方、聴衆の側も相槌や頷きなどでリアルタイムにフィードバックし、時折質問やコメントも行う。また会議と違って、参加者は立っており、動くこともできるので、マルチモーダルなインタラクションを行うことが多い。さらに、ポスター会話を扱う利点としては、話題や他の参加者に対する親近性を制御しながら、（研究者を集めてくれば）自然でリアルなデータを収集することが非常に容易であることが挙げられる。

本プロジェクトの目標は、人間どうしのインタラクションの信号レベルのセンシングとより高いレベルの認識である。認識のタスクとしては、視線・話者・発話区間などの検出に加えて、聴衆の理解・興味度の推定などを考えている。これらは会話アーカイブに対する新たなインデキシングの枠組みを提供することが期待される。例えば、自身あるいは同僚のポスター

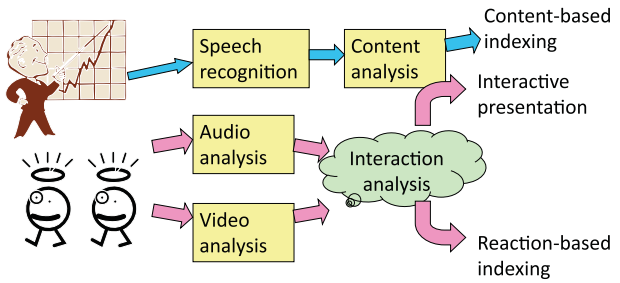

図 1　マルチモーダルなインタラクション分析の概要

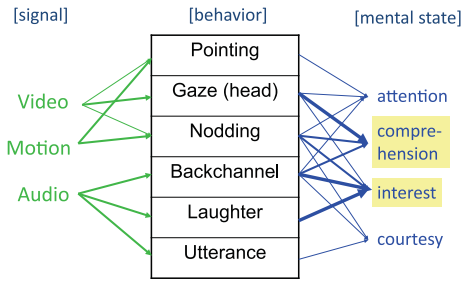

図 2　マルチモーダルなセンシングと分析の概要

セッションが終わった後で、どのような質疑・コメントが行われたか、どこに興味を持ってもらえたか、どこが聴衆にとってわかりにくいところであったか、といった要求に応えることができる。また、将来自動でプレゼンテーションを行う知的な会話エージェントの基盤にもなる。

これまで、発話者の音声の認識や言語解析などに基づいた、内容に基づく (content-based) インデキシングのアプローチが研究されてきたが、これに対して、聞き手の反応に着目した (interaction-based) アプローチを提案する。具体的には、相槌・頷き・視線配布などの非言語情報に注目する。現状の音声認識・言語解析システムと比べて、人間の聴衆の方がはるかに説明内容を理解しているのは自明であるので、その反応を捉える方が合理的と考える。この枠組みを図 1 に示す。

そのために、多人数会話のマルチモーダルなセンシングと分析を行うための基盤を構築する。処理の概要を図 2 に示す。音響信号からは、発話とともに笑い声や相槌を検出する。これらに加えて、視線・頷き・指示動作の情報を検出する。「正解」のコーパスを構築する際には、モーションキャプチャシステムや視線計測装置などの特殊な装置を使用するが、最終的なシステムは、カメラと遠隔マイク、さらには Kinect のような簡易センサのみで実現することを想定している。

その上で、これらの情報を組み合わせることによって、興味・理解度の推定を行う。動画投稿サイトなどの事例からもわかるように、我々は他の人が興味を持ったものを視聴したくなるのが普通であるので、このようなアノテーションは有用であると考えられるが、一方でアノテーションの基準や評価を含めて、これらを明確に定義するのは非常に困難である。そこで、これらに関係すると考えられ、より客観的に定式化できるマイルストーンをいくつか設定する。

本稿では、まず 2 章でセンシング環境と収集したコーパスの説明を行う。3 章では、図 2 の左側の部分、すなわち信号レベルから行動レベルの処理について述べる。4 章と 5 章では、図 2 の右側の部分、すなわち行動レベルから心的状態レベルの処理について述べる。具体的には、視線や相槌の情報を用いることで、聴衆の誰がいつどのような質問をするか予測するモデルの構築を試みる。これらの分析によって、ポスター発表が理解されたのか、興味を持たれたのかといった「場の認識」を行うための手がかりを得られることを期待している。6 章では、現在構築しているスマートポスターボードシステムの紹介を行う。

2.　ポスター会話のマルチモーダルコーパス

2.1　収録環境

我々は、ポスター会話における音声・映像・動作・視線などのマルチモーダルな情報を収録するための環境（IMADE ルーム）の構築を進めてきた [5], [6]。音声に関しては、各参加者にワイヤレスのヘッドセットマイクを装着してもらうとともに、ポスターボードの上に設置するマイクロフォンアレイを設計した。映像に関しては、参加者全員とポスターをカバーできるように、6〜8 個のカメラを部屋に設置した。この部屋には、モーションキャプチャシステムも設置されており、各参加者の位置と動きを把握するために使用した。各参加者には、このマーカに加えて、頭部に加速度センサと視線計測装置を装着してもらっている。このように多数の装置を装着してもらうのは参加者にとって負担となるが、大半の装置は帽子とウエストバッグに納めており、あまり気にならないように工夫している(注1)。実際のポスター会話の収録の様子を図 3 に示す。

2.2　コーパスとアノテーション

上記の環境を用いて、これまで 3 ヶ年度にわたって合計 31 セッションのポスター会話を収集してきた。ただし、いくつかのセンサデータが欠損したものも含まれる。以降の章の分析では、アノテーションを整備した 4 つのセッションを用いている。これらのセッションにおける発表者と聴衆の組合せはすべて異なっている。各セッションにおいては、1 名の発表者（A と表記）が自身の研究に関する発表を、2 名の聴衆（B,C と表記）に対して行う。聴衆は、発表者についても研究内容についても初めて接する設定となっている。セッションの長さは制御しているわけではないが、おおむね 20 分程度である。

ヘッドセットマイクで収録された音声データは、ポーズで区切られた発話単位 (IPU) に分割し、時間と話者ラベルを付与した上で、『日本語話し言葉コーパス』(CSJ) と同じ基準で書き起こしを行った。ただし、フィラー以外に相槌と笑いに対してもアノテーションを行った。

視線情報は、視線計測装置とモーションキャプチャシステムのデータを用いて、視線ベクトルと他の参加者やポスターの位置との交差判定に基づいてアノテーションを行った。頷きについては、頭部の加速度センサのデータに基づいて自動アノテーションを行った。

（注1）：実際には視線計測装置等のキャリブレーションが大きな負担である。

図 3　ポスターセッションの収録の様子

3.　会話参加者の発話・視線等の自動検出

図 2 の左側の部分、すなわち音声や映像の信号から、各会話参加者の行動を検出する処理については、順次研究開発を進めている。ここでの目標は、参加者にマイクなどの装置を一切装着してもらうことなく、ポスターボード等に設置したセンサだけで処理を行うことである。具体的には、マイクロフォンアレイで収録される音声信号、及び複数台のカメラ（Kinect センサ）で得られる映像信号・距離情報を用いる。

3.1　視線（頭部方向）推定

視線推定には様々な方式が提案されているが、本研究では Kinect センサを用いて、ポスター会話に適した実用的な視線推定を実現した。なお、画像の解像度及び眼鏡等の影響により眼球そのものを安定に撮影することは困難であるため、視線方向を頭部方向で代用する。視線と頭部方向のずれは平均 10 度程度で、ポスターを注視する状況ではさらに小さくなる傾向がある [7]。処理は以下の手順で行っている。

（1）　正面顔検出

Kinect センサで撮影したカラー画像及び距離画像から、Haar-like 特徴を利用した物体認識法を用いて正面顔探索を行う。同時に複数人を処理することが可能である。

（2）　頭部モデル獲得

正面顔の検出結果に従って、距離画像から頭部の 3 次元形状を、カラー画像からその色情報を計算する。計算結果はポリゴンとテクスチャ情報に変換し、頭部モデルとする。

（3）　頭部追跡

頭部方向の推定を、画像への頭部モデルのフィッティングとして行う。具体的には、頭部を剛体とみなし、頭部の 3 次元位置と姿勢を表す 6 変数をパーティクルフィルタによる追跡処理で逐次計算する。さらに、得られた 6 変数を初期値とした最適化処理を行うことで、頭部の 3 次元位置と姿勢の推定の高精度化を行う。

（4）　注視対象推定

頭部追跡処理で得られた 6 変数から、3 次元空間で視線に対応する半直線を求める。この半直線と、ポスターボードや他の参加者との交差判定を行うことで、注視対象を決定する。

上記の処理は、GPU を使うことで、オンライン・リアルタイムに行うことも可能である。

3.2　頷きの検出

頷きは、上記頭部モデルの上下方向の動きとして捉えられる。

3.3　マイクロフォンアレイによる音声の分離・強調

音声の分離と強調は、ブラインド空間的サブトラクションアレイ (BSSA) [8] によって実現する。これは、マイクロフォンアレイで得られる信号に対して、遅延加算 (Delay-and-Sum) 型ビームフォーミングを行うとともに、独立成分分析 (ICA) に基づいて各会話参加者の音声と背景雑音を分離し、目的信号以外の抑圧を行うものである。ポスター会話の設定では、発表者・聴衆・背景雑音の 3 つの成分への分離を行う（聴衆間の分離は行われない）。その際に、画像処理によって得られる各参加者の位置情報を用いることで、ICA のフィルタ計算の高速化を実現している。これは、画像処理と音声処理の統合と位置づけられる。この処理を逐次的に行うことで、参加者が移動しても追随できるようにしている。

19 チャネルのマイクロフォンアレイを用いる場合は、高い品質の音声強調ができるが、リアルタイムには処理できない。Kinect センサ内蔵の複数のマイクロフォンを用いる場合は、音質は低下するが、リアルタイム処理が可能である。

分離された各音声チャネルに対して発話区間検出を行う。発話区間検出には、目的音声以外を徹底的に抑圧した信号を用いる。ただし、この信号は歪みが大きいので、人間の聴取や音声認識などのために、抑圧の程度は小さいが歪みも小さい信号も用意する。

3.4　相槌と笑い声の検出

相槌と笑い声は、聴衆の非言語的な音声による反応として重要であり、著者らは Podcast などの会話コンテンツを対象に、検出方法の研究を行ってきた [9]。これは、GMM によるモデル化と BIC によるセグメンテーションを組み合わせた方法である。

ただし相槌は、通常の音声と周波数特徴が類似しているため、韻律的特徴を併用する。聴衆の相槌のうち、特に重要な反応と考えられるのは、「へー」「あー」「ふーん」などの非語彙的で引き延ばし型のものであるので、持続長が長く、基本周波数や周波数包絡が一定時間平坦なものを抽出する。

笑い声は GMM による単純な方法で、再現率・適合率とも約 70%の検出が可能であるが、相槌は明瞭でないものが多いので、再現率約 30%・適合率約 80%となっている。ただし、聴衆の明白な反応を検出できれば十分と考えている。これにより、聴衆の興味を捉えられることを示している [9]。

4.　視線と相槌の情報に基づく聴衆の発話の予測

図 2 の右側の部分、すなわち聴衆の行動と興味・理解度などの心的状態の対応のモデル化と解明（データ収集や実験の方法）は容易でない。ただし一般に、聴衆は当該発表に興味を持った時に質問やコメントを行うと考えられるので、聴衆による発話は興味と関係があると期待される。

表 1　視線の継続時間（秒）と発話交替の関係

	発表者 A による	聴衆による発話	
	ターンの保持	B	C
発表者 A が聴衆 B を注視	0.220	**0.589**	0.299
発表者 A が聴衆 C を注視	0.387	0.391	**0.791**
聴衆 B が発表者 A を注視	0.161	0.205	0.078
聴衆 C が発表者 A を注視	0.308	0.215	0.355

多人数における会話において話者交替は自明でなく、ターン（発話権）が誰に譲られ、取得されるかをモデル化・予測することは、会話分析において重要なテーマとなっている [10]。また、多人数を会話相手とする音声対話エージェント・ロボット [11], [12] や、発話者に対してカメラのズームやマイクのビームフォームを制御するようなシステムの研究開発においても重要である。２名の対話における話者交替に関する研究はこれまで数多くされているものの、多人数会話において話者交替を予測する工学的なモデル化はきわめて少ない [13], [14]。ポスターセッションにおける会話は、先行研究で取り組まれたミーティングや自由会話などと異なり、発表者が大半の時間でターンを保持し、発話数も圧倒的に多い。しかしながら、聴衆の質問やコメントの発話の方がより重要と考えられる。したがって本研究では、聴衆によるターンの取得を予測する問題を設定する。具体的には、いつ聴衆の誰が発話するかの予測を行う。

発話交替において視線配布が重要な役割を果たしていることが知られている [11], [14], [15]。しかしながらポスター会話では、参加者の視線の大半はポスターに注がれる。地図やコンピュータなどを用いた会話でも同様と考えられる。そこで本研究では、相槌などの情報の利用も考える。なお本章（及び次章）で述べる結果は、コーパスに付与された「正解」のアノテーションに基づくもので、前章で述べた視線や発話の自動検出に基づいて行われたものではない。これらの統合は今後の課題である。

4.1　視線対象と発話交替の関係

発表者の発話 (IPU) 終了時におけるすべての参加者の視線対象を同定した。視線対象は、ポスターか他の参加者（もしくは「なし」）のいずれかである。発表者の発話終了直前の 2.5 秒間における視線の継続時間も計測した。2.5 秒間に設定したのは大半の発話が 2.5 秒以下であるためである。発話交替との関係を表 1 に示す。発表者が聴衆にターンを譲る場合には、他の場合に比べて、その聴衆の方をより長く見ていることがわかる。

4.2　両者の視線状態と発話交替の関係

次に、発表者と聴衆の両者の視線配布に基づく視線状態を表 2 のように定義する。この表では「聴衆」と表記しているが、実際にはこれらの状態は聴衆の各人について定義される。例えば、“Ii” は発表者と聴衆の 1 人との相互注視である。また、“Pp” はポスターに対する共同注意に相当する。

発表者の発話末におけるこれらの視線状態の頻度と発話交替の関係を表 3 に示す。ここでは、各々の頻度は 2 名の聴衆に対して合計をとったものとなっており、発話交替に応じて分類している。発話交替は、当該視線状態に関係した人がターンを取得した場合 (self) とそうでない場合 (other) に分類している。

表 2　発表者と聴衆の視線状態の定義

視線配布者	発表者		
聴衆	視線対象	聴衆　**(I)**	ポスター **(P)**
	発表者 **(i)**	**Ii**	**Pi**
	ポスター **(p)**	**Ip**	**Pp**

表 3　発表者と聴衆の視線状態と発話交替の関係

	発表者による	聴衆による発話		合計
	ターンの保持	(self)	(other)	
Ii	125	17	3	145
Ip	320	**71**	26	417
Pi	190	11	9	210
Pp	2974	147	145	3266

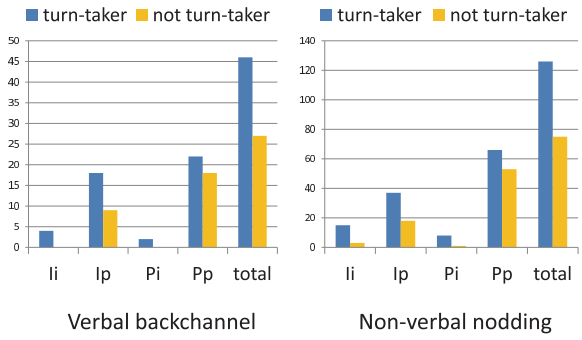

図 4　相槌・頷きと視線状態及び発話交替の関係

相互注視の状態 “Ii” が発話交替に最も関係が深いと考えられたが、その頻度は多くない。聴衆が発表者を見ている状態 “Pi” の頻度も多くない。最も有用な情報と考えられるのは、発表者が聴衆を見ている状態 “Ip” であり、この場合にその聴衆にターンが譲られる場合が多い。この結果は前節の分析とも符合する。

4.3　相槌と発話交替の関係

発表者の発話終了直前の 2.5 秒間における相槌と頷きの頻度と発話交替の関係を図 4 に示す。視線状態との関係も示している。聴衆が発話しようとする際 (turn-taker) には相槌・頷きとも多くなっている傾向が見られ、特に “Ii” や “Ip” といった発話交替との関係が深い視線状態においてより顕著である。

4.4　聴衆からの発話の予測実験

前節までの分析に基づいて、聴衆によるターンの取得を予測することができるか実験を行った。これは、話者交替の予測と次話者の予測の２つのタスクに分割して行った。話者交替の予測タスクでは、ターンが発表者から聴衆（の誰か）に譲られるかの予測を行い、次話者の予測タスクにおいて聴衆の誰が実際に発話するのか予測する。これらの予測は、発表者の各発話 (IPU) の終了時、すなわち話者交替・次話者の発話が行われる前に、それまでの情報を用いて行う。

話者交替の予測タスクでは、発表者の発話の韻律特徴をベースラインとして用いた。具体的には、予測時点の直前の発表者の発話の F0（平均・最大値・最小値・レンジ）とパワー（平均・最大値）を計算した。聴衆の相槌と頷きについては、予測

表 4　話者交替の予測結果

特徴	再現率	適合率	F 値
韻律	0.667	0.178	0.280
相槌	0.459	0.113	0.179
視線	0.461	0.216	0.290
韻律+相槌	0.668	0.165	0.263
韻律+視線	**0.706**	0.209	0.319
韻律+相槌+視線	0.678	0.189	0.294

表 5　次話者の予測結果

特徴	予測精度
視線	66.4%
相槌	52.6%
視線+相槌	**69.7%**

時点の直前の各々の頻度を求めた。視線配布に関する特徴は、視線対象と視線状態の両方に関して、出現頻度と継続時間の両方で求めることができる。ただし実際には、これらの 4 通りの組合せの間で予測性能に有意な差は見られず、併用する効果もなかった。

話者交替を予測するモデルとして SVM を用い、4 セッションでクロスバリデーションによる学習・評価を行った結果を表 4 に示す。話者交替、すなわち聴衆によるターンの取得に関する再現率 (recall)・適合率 (precision)・F 値（=再現率と適合率の調和平均）を求めている。発表者の発話の大半は継続され、話者交替が発生する割合は 11.9%に過ぎないので、その予測はかなり困難である。そこで、話者交替の再現率を最も重視する。韻律特徴が最も高い再現率を得ており、視線特徴が適合率と F 値で最高になっている。これらの 2 つの特徴を組み合わせることにより、再現率・適合率ともに改善することができた。これに対して、相槌・頷きを用いる効果は見られなかった。

次に、発話交替が起こった場合に、次話者の予測を行った。多人数会話において、（実際に次の話者が発話する前に）次話者を予測することは非常に困難なタスクであり、これまでの研究例も皆無に近い。この場合、発表者（現発話者）の韻律情報には次話者を示唆する情報は含まれないので、聴衆の相槌・頷きと視線配布の特徴を用いた。これらの特徴は、話者交替の予測では聴衆全体について求めていたが、次話者の予測では聴衆の各人毎に求める。予測結果を表 5 に示す。今回は、相槌・頷きの特徴を用いる効果が見られ、視線配布の特徴と組合わせることにより、約 70%の精度が得られた。

以上を総合すると、発話権の管理は発表者が主に行っているものの、聴衆がターンを取得するためには、視線や相槌などでフィードバックを行う必要があることが推察される。

5.　聴衆の聞き手行動と質問の種類の関係

次に、聴衆の聞き手行動とその後にどのような種類の質問を行うのかの関係を調べた。本研究では、質問の種類を確認質問と踏み込み質問の 2 つに分類した。確認質問は、発表者の説明に対する理解を確認するためのもので、「はい/いいえ」のいずれかで答えることができる(注2)。これに対して踏み込み質問は、発表者の説明に含まれなかったことに関して質問を行うもので、通常「はい／いいえ」のみで答えられるものでなく、何らかの補足説明が必要となる。

表 6　相槌の頻度（秒当り）と質問の種類との関係

	確認	踏み込み
質問者	0.034	**0.063**
非質問者	0.041	0.038

表 7　頷きの頻度（秒当り）と質問の種類との関係

	確認	踏み込み
質問者	0.111	0.127
非質問者	0.109	0.132

表 8　発表者を注視する時間割合と質問の種類の関係

	確認	踏み込み	説明中平均
質問者	**0.023**	**0.201**	0.094
非質問者	0.071	0.118	0.094

この質問の分類と対応する発表者の説明区間のアノテーションは、発表者がポスターに沿って説明している間は比較的容易であるが、発表を一通り終えて全体的な質疑に入ってしまうと非常に困難になる。そこで、質疑に入った後は除外して、発表者の説明中に行われた質問についてのみアノテーションを行った。また、視線や相槌などの聞き手行動について、前章では発話 (IPU) 毎に特徴を求めていたが、ここでは発表者の複数の発話をまとめた文を人手で定義して求めた。これらの点が前章の分析との違いである。

5.1　相槌・頷きと質問の種類の関係

相槌と頷きについて、説明区間の長さ（秒）で正規化した出現頻度と質問の種類との関係を表 6 と表 7 に示す。実際に質問を行った聴衆（質問者）とそうでない聴衆（非質問者）の比較も行っている。踏み込み質問を行う際には、相槌の頻度が有意に多くなっていることがわかる。しかしながら、頷きに関しては特徴的な傾向が見られなかった。相槌には理解・納得の程度を示す働きがあるのに対して、頷きについては必ずしもそうではないことを示唆している。

5.2　視線配布と質問の種類の関係

聴衆の視線配布について、発表者 (A) を注視している時間の割合と質問の種類との関係を表 8 に示す。参考までに、説明中の平均値（質問者・非質問者共通）も示している。全体的な質疑などの会話状態を除いているため、表 1 よりも全般に値が小さくなっている。平均値及び非質問者と比較して、確認質問を行う際には発表者を注視している時間が短く、踏み込み質問の場合は長いことがわかる。確認質問を行うのは、ポスターの内容が十分理解できていない場合であるので、ポスターを注視している時間が長くなると考えられる。一方、踏み込み質問の場合は、ターンを積極的に得ようとしているものと推察される。

(注2)：発表者が実際に「はい／いいえ」のみで答えたとは限らない。

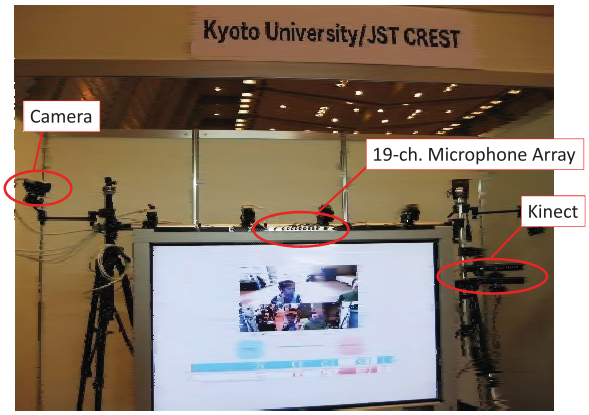

図 5　スマートポスターボードの概観

6. スマートポスターボード

以上述べた処理・分析に基づいて、「スマートポスターボード」の設計と実装を進めている。これは、ポスターセッションを収録し、参加者のふるまいをセンシングし、インタラクションのアノテーションを行うものである。2 章で述べたコーパス収録時のように、参加者にヘッドセットマイクや視線計測装置などの特殊な装置を装着してもらうのは現実的でなく、また部屋に装置を設置するのも可搬性の点で問題となるので、すべてのセンサはポスターボード（65 インチ液晶ディスプレイ）に装着することにした。その概観を図 5 に示す。ディスプレイの上部に 19 チャネルのマイクロフォンアレイを設置し、併設する三脚にカメラ（6 台）と Kinect センサを設置している。

3 章で述べた信号レベルから行動レベルの処理は、ほぼ実装ができている。4 章・5 章で述べた行動レベルから興味・理解度レベルの処理は、まだ統合できていないが、行動レベルのアノテーションを視覚化するだけでも、議論が活発であるか、会話が盛り上がっているか、聴衆の関心が散漫になっていないか、などを示すことができる。

このように、ポスター会話を収録した映像・音声とあわせてアーカイブ化・視覚化したブラウザを構成することで、セッションに立ち会えなかった人（あるいは発表者自身）がセッションの様子（例えば、どこに興味を持たれたか、どこが理解しにくかったか）を把握したり、重要なやりとりを効率的に視聴（ブウジング）することができる。ポスターセッションは通常長時間にわたるので、これらの実現は有用と考えている。

また、これらの技術を発展させることにより、聴衆の反応に応じて提示コンテンツを切り替える自動プレゼンテーションも実現できると期待される。理解が困難な箇所には補足説明を行ったり、興味が高い箇所には追加説明やコメントを付加するなどの機能を考えている。これは、著者らが研究開発を進めてきたプロアクティブな情報案内システム「情報コンシェルジェ」[16] の発展形と位置づけている。

謝　　辞

本研究は、JST CREST「人間調和型情報環境」領域ならびに科学研究費補助金の支援を受けて実施されたものである。特に、スマートポスターボード（2 章・3 章・6 章）の研究開発は、CREST プロジェクトに参画して頂いている京都大学の吉本廣雅研究員、Tony Tung 特定助教と、奈良先端科学技術大学院大学の猿渡洋准教授をはじめとする多くの方々の貢献によるものである。また、4 章・5 章の研究は、京都大学の高梨克也特命助教と大学院生の岩立卓真君、土屋貴則君、林宗一郎君におうものである。

文　　献

[1] S.Renals, T.Hain, and H.Bourlard. Recognition and understanding of meetings: The AMI and AMIDA projects. In *Proc. IEEE Workshop Automatic Speech Recognition & Understanding*, 2007.

[2] K.Ohtsuka. Conversation scene analysis. *Signal Processing Magazine*, Vol. 28, No. 4, pp. 127–131, 2011.

[3] T.Kawahara. Multi-modal sensing and analysis of poster conversations toward smart posterboard. In *Proc. SIGdial Meeting Discourse & Dialogue*, pp. 1–9 (keynote speech), 2012.

[4] 河原達也. ［招待講演］ スマートポスターボード: ポスター会話のマルチモーダルなセンシングと認識. 電子情報通信学会技術研究報告, SP2012-51, 2012.

[5] 瀬戸口久雄, 高梨克也, 河原達也. 多数のセンサを用いたポスター会話の収録とその分析. 情報処理学会研究報告, SLP-67-6, 2007.

[6] T.Kawahara, H.Setoguchi, K.Takanashi, K.Ishizuka, and S.Araki. Multi-modal recording, analysis and indexing of poster sessions. In *Proc. INTERSPEECH*, pp. 1622–1625, 2008.

[7] 矢野正治, 中田篤志, 福間良平, 角康之, 西田豊明. 非言語マルチモーダルデータを用いた会話構造の分析のための環境構築. 情処学研報, 2009-UBI-22-12, 2009.

[8] Y.Takahashi, T.Takatani, K.Osako, H.Saruwatari, and K.Shikano. Blind spatial subtraction array for speech enhancement in noisy environment. *IEEE Trans. Audio, Speech & Language Process.*, Vol. 17, No. 4, pp. 650–664, 2009.

[9] 河原達也, 須見康平, 緒方淳, 後藤真孝. 音声会話コンテンツにおける聴衆の反応に基づく音響イベントとホットスポットの検出. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3363–3373, 2011.

[10] 坊野真弓, 高梨克也（編）. 多人数インタラクションの分析手法. オーム社, 2009.

[11] D.Bohus and E.Horvitz. Models for multiparty engagement in open-world dialog. In *Proc. SIGdial*, 2009.

[12] S.Fujie, Y.Matsuyama, H.Taniyama, and T.Kobayashi. Conversation robot participating in and activating a group communication. In *Proc. INTERSPEECH*, pp. 264–267, 2009.

[13] K.Laskowski, J.Edlund, and M.Heldner. A single-port non-parametric model of turn-taking in multi-party conversation. In *Proc. IEEE-ICASSP*, pp. 5600–5603, 2011.

[14] K.Jokinen, K.Harada, M.Nishida, and S.Yamamoto. Turn-alignment using eye-gaze and speech in conversational interaction. In *Proc. INTERSPEECH*, pp. 2018–2021, 2011.

[15] A.Kendon. Some functions of gaze direction in social interaction. *Acta Psychologica*, Vol. 26, pp. 22–63, 1967.

[16] 河原達也, 川嶋宏彰, 平山高嗣, 松山隆司. 対話を通じてユーザの意図・興味を探り情報検索・提示する情報コンシェルジェ. 情報処理, Vol. 49, No. 8, pp. 912–918, 2008.