Vol.7 No.1, 2005

装着型体験記録装置による対話インタラクションの判別機能実装と評価

伊藤 禎宣 *1 *2 岩澤 昭一郎*1 *2 土川 仁 *1 *2 角 康之 *1 *3
間瀬 健二 *1 *2 *4 片桐 恭弘 *1 小暮 潔 *2 萩田 紀博 *2

Implementation and Evaluation of Conversational Interaction Recognition System using Wearable Experience Capturing System

Sadanori Ito *1 *2, Shoichiro Iwasawa *1 *2, Megumu Tsuchikawa *1 *2, Yasuyuki Sumi *1 *3, Kenji Mase *1 *2 *4, Yasuhiro Katagiri *1, Kiyoshi Kogure *2, and Norihiro Hagita *2

Abstract – **The purpose of our wearable experience capturing system is to recognize the interaction among humans. Various interactions, such as conversations, are effective index of the captured experience. We assume that the non-verbal information, such as the relative position among humans, head direction, and talking behavior, can be used for recognition of conversational interaction. We developed a wearable device, the head worn type local positioning system, called "InfraRed ID Tracker", to sense the relative position and the head dierction. In this paper, we evaluate and discuss about accuracy of the experience capture and interaction recognition functions by use of the non-verbal information.**

Keywords : **interaction recognition, experience capturing, local positioning system, wearable devices**

1. はじめに

近年、我々が日常的に体験する情景を多様な手法で記録し、再活用する試みが幅広く行われている。この背景には、無為な日常をも常時記録可能な小型収録装置や大容量保存装置が現実的な選択肢として現れてきたことがある。カメラやビデオを操作して選択的に記録し、後から参照するという従来的な体験記録と再活用の形態に対して、装着型記録装置[20]等を用いた非選択的な常時記録と選択的参照という、より柔軟性の高い記録情報の運用が可能になりつつある。一方で、膨大な体験記録からの適切な選択的参照を支援するといった情報の有用性を高める目的で、機械可読かつ有意味なインデクスを体験記録へ自動付与する要求が高まっている[1]。

我々は、展示会場のように自由移動可能な屋内空間を対象として、体験の記録と再活用を行うシステムを開発している[36]。展示会場における有意味な体験としては、展示ブースへの滞在や自由な位置で始まるグループディスカッションといった、人対物や人対人のインタラクションが主に想定される[31]。本システムでは、このような体験場面を、参加者の視野映像や音声による視聴覚的な体験のコンテンツとして、装着型記録装置を用いて常時記録する。また同時に、対話インタラクションの場面を判別し、これをインデクスとした体験記録を生成する。

本稿では、対話の判別モデルとして、インタラクション過程の視覚的行為可能性に着目した"インタラクション・スコープ"[14]を提案する。様々な対人対物インタラクションにおいて、対話相手やポスタのような話題対象物を視認する行為は、インタラクションの開始や関係を判別するのに有用かつ特徴的なモダリティ[30]である。しかし、自由移動可能な空間で複数人の視線方向を随時計測するのは困難という問題がある。そこで、近似的に視覚的行為の起り得る被視対象物との相対位置の範囲をインタラクションスコープと規定し、これを検出する赤外線IDトラッカ装置を実装した。

本装置は、トラッカとタグから構成される。トラッカは頭部装着型で、視線対象物の認識と映像記録を目的とし、実効的な視野相当の記録画角を持つ。画角内のIDタグが付与された対象物を認識し、タグのIDと画面上の位置を取得できる。タグは被視対象物に装着され、視覚的インタラクションが可能な範囲(スコー

*1: ATR メディア情報科学研究所
*2: ATR 知能ロボティクス研究所
*3: 京都大学情報学研究科
*4: 名古屋大学情報連携基盤センター
*1: ATR Media Information Science Laboratories
*2: ATR Intelligent Robotics and Communication Laboratories
*3: Graduate School of Informatics, Kyoto University
*4: Information Technology Center, Nagoya University

プ) を、トラッカによりセンシング可能な範囲としてハードウェア的に実装している。装置は、認識の即時性や低処理負荷といった特徴を持ち、装着者と被視対象物との視覚的インタラクションが可能な相対位置関係にあるか随時判定する。

本稿では、装置が視線を頭部方向によって代替し、視覚的行為や体験を近似的に記録することの適切性について、検証する。視線方向と頭部の運動に関する生理的影響や、外的な環境要素が与える影響について検討する。また、これらの非言語情報によるインタラクション判別の適否について述べる。

次章では、対面対話時の相対位置や行動に関する心理学や環境心理学からのアプローチと、対面対話分析の工学的実装に関する既存研究について述べ、本研究の立場と新規性を明らかにする。3章では、開発した装着型体験記録装置によるインタラクション場面判別機能の実装について述べる。4章と5章では、自由移動可能な空間での対面対話状況を、参加者の視線と相対位置の両側面から分析し、体験記録装置の特性評価を行う。

2. 関連研究と本研究の位置づけ

例えば会議室の席順に人間関係をみるように、外部観測可能な非言語行動から相互行為の意味を捉えるアイデアは、我々が直観的に納得できるものである。非言語行動からインタラクションを分析する試みは古くから行われてきた。特に対面対話時の位置関係に着目した研究としては、対人関係のような心理的側面[12],[17],[29]や、対話環境のような物理的側面[2],[3]からの分析が行われている。初期の代表的研究としては、人々による対人距離の取り方をテリトリーの概念で示したSommerのPersonal Space[29]や、人間関係を対人距離にマップして親密距離 (～45cm) や公共距離 (3.5m～) といった分類を行ったHallのProxemic Theory[12]がある。SommerやHallが主に一対一関係での距離を分析したのに対して、Kendonは複数参加者で構成される対面対話過程の相対位置に着目した。彼は、参加者個人の前方には、視野や接触可能な範囲で規定される、対話活動に利用可能な空間 “Transactional Segment(操作領域)” [17]があるとしている。活動の態様に応じて共同の操作領域を維持する特定の配置 “F-formation” が現れることから、配置の観測結果から参加者の活動や役割[11]を推測する可能性が示唆されている[4]。また、Barkerは、これらの人的配置を含む相互行為に、椅子や机といったmilieu(物理的環境要素) が与える影響や制約[2]に着目し、環境要素と行為の関係を記述する枠組みとしてBehavior Setting(行動場面) の概念[3]を提唱した。

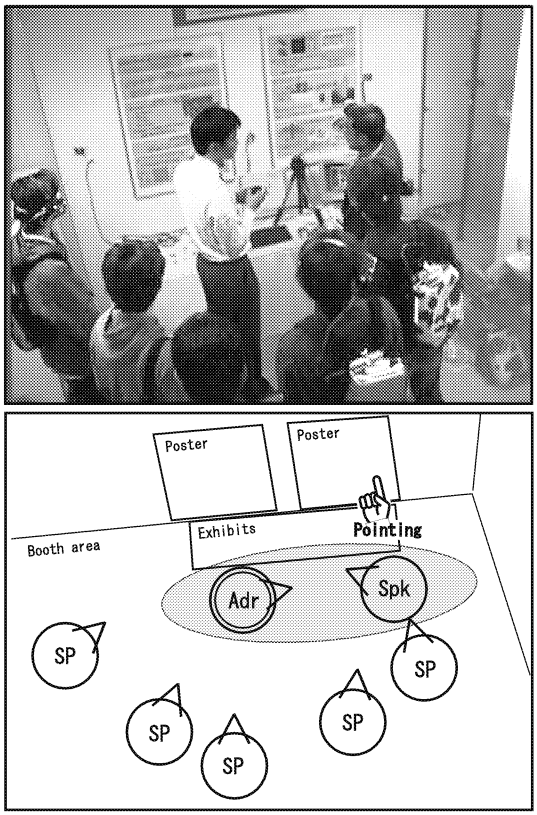


図1 展示会場の対話場面分析の例
Fig. 1 Examples of interaction analysis in the exhibition.

本稿が対象とするような、ある展示会場のインタラクション場面をKendonとBarkerの論理から分析した例を図1に示す。このように特徴的な活動場面を抽出して索引を付与することは、映像や音声による体験記録の再利用性を高める上でも必要な作業と言える。

これらの分析的研究が熟練者の手入力に依存する一方で、対面対話の非言語行動記録を自動化[23],[39]し、対話行為の分析[5],[8],[9],[30]や、対話ビデオへのインデクシング[24],[26]といった、インタラクション記録の再利用に重点をおく研究も進められている。

対話分析目的では、人々の集合を識別するための装置として、Sociometer[8]やMeme tag[5]といった赤外線送受信機やRFID(Radio Frequency IDentification)タグ[9]を使った研究がある。このようなユーザ装着型デバイスによる実装は、自由移動可能な空間での非言語行動の一つと言える、人々の集合と離散を記録できるという利点がある。しかし、これらはセンサによる数mの認識範囲内に相手が存在するか否かを判定するものであるため、人々がどちらを向き、何を見て

いるかといった、より詳細なインタラクションの判別はできない。

これに対して、既に集合した参加者による移動の無い環境で、誰が誰を見て話しているかといった、より詳細な分析と記録を行う研究もある。着席姿勢での対面会議を対象とする[30]では、設置型カメラの映像から参加者の視線方向を抽出し、対話過程での相手方を識別して、対面会議記録へのインデクス付与を行っている。映像コンテンツとしての再利用性を高めることを目的とした研究[6],[7],[18],[24],[25],[33],[39]では、対話過程の意味的内容を設置型カメラの映像から分析し、インデクシングする手法[13],[24],[33]が広く行われている。多くのRFIDや赤外線送受信機を用いた研究よりも、映像による手法は、詳細なインタラクション判別が可能とされる。しかし、これらの設置型カメラを用いた方法は、限られた空間を対象としたものであり、オクルージョンの発生といった原理的問題から、自由に移動可能な広い空間で多人数の体験を記録するという、本稿の目的には向かない。

映像のみの利用による意味理解の限界 (semantic gap) という問題から、映像以外の複数モダリティを併用[28]する手法の研究も進められている。例えば、講義やプレゼンテーションなどの特定シチュエーションでの特徴的な行動を、PC端末の操作[26]や身体動作センサ[23]から抽出し、これを基準としたインデクシングや編集を行う手法が提案されている。しかし、これらの手法も、特定場面での分析の精緻化には効果的だが、設置型装置と該当装置を撮影するカメラを用いており、対象は固有空間に制限される。

自由に移動可能な空間で、個人の体験記録として視野映像を記録する目的では、装着型のカメラ利用が提案[15],[20],[32],[35]されている。装着型カメラによる記録映像から、注視行為といったインデクスを作成する研究[32]もある。しかし、ビジョンのみによる手法は処理負荷が高く、同時に多人数が利用する環境や、インタラクション判別の結果を随時利用する用途には向かない。また、頭部装着カメラを用いたこれらの研究では、実際の視線方向と頭部方向とのズレが問題となるが、この点についての議論は、あまりなされていない。

本装置では、特定場所への滞在といった、位置に紐付け可能な情報のほか、参加者間のグループディスカッションのように、自由な空間で始まるインタラクションをとらえることも目的としている。この場合、複数グループが隣接して並存することは十分考えられるので、単に近接範囲ではなく、認識の指向性が必要である。また、体験コンテンツの再活用を考えた際には、映像内の対象物 (相手) 判定が必要であり、ビジョンによる手法が目的に適う。しかし、このようなインタラクションの発生をトリガとするガイダンスサービスなどをアプリケーションに想定した場合、判別の即時性が求められるため、ビジョンによる高処理負荷な手法での実現は困難である。

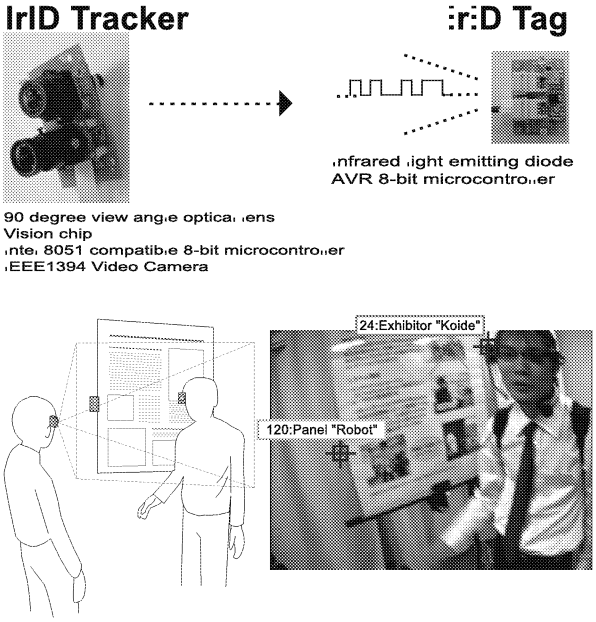


図2　赤外線IDトラッカの構成と視野記録の様子
Fig. 2　Configuration of the Infrared ID Tracker system.

本稿では、頭部に装着した視野映像記録カメラと、そこに併置した赤外線IDトラッカ装置による対象物識別機能により、低負荷、短時間での指向性ある対象物識別や相対位置取得を可能とした。

次章では、開発した装着型体験記録装置の概要について述べる。

3.　装着型体験記録装置の実装

インタラクションコーパスプロジェクト[36]は、“体験”の記録と再活用を目標として進めている。視覚的体験をカバーする視野カメラと、聴覚的体験をカバーするマイクを備えた、ユーザ体験記録用の装着型クライアントと、無線LAN経由で各記録情報を収集するデータベースサーバを基本構成とするシステムを構築した。

これら記録された映像音声の再活用には、記録状況に沿ったインデクス付与が有用である。人々の対面対話インタラクションに着目したインデクシングのため、発話パワーレベルによって切り出した話者判定結果と、赤外線IDトラッカ装置が検出した視野方向の対象物タグID番号を用いる。赤外線IDトラッカ装置は、赤外線LED(Light Emitting Diode)を備えたタグと、イメージセンサを備えたトラッカから構成される (図2)。タグは、1個または複数のLEDを持ち、タグ固有の

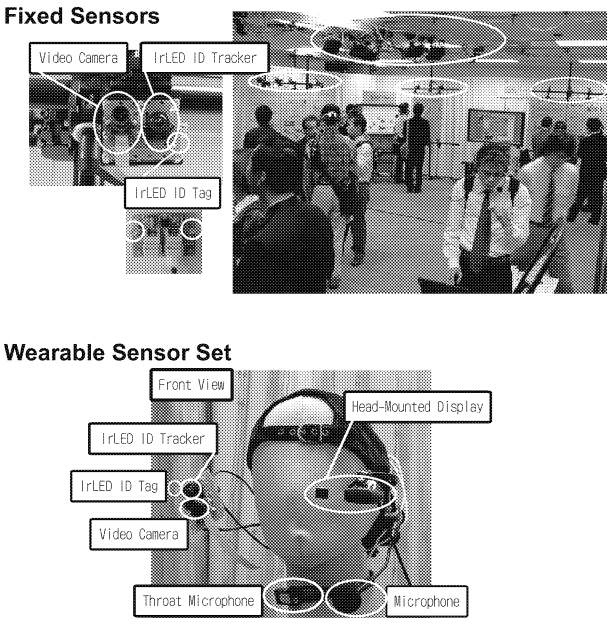


図 3 体験記録装置のハードウェア構成
Fig. 3 Hardware configuration of the experience capturing system.

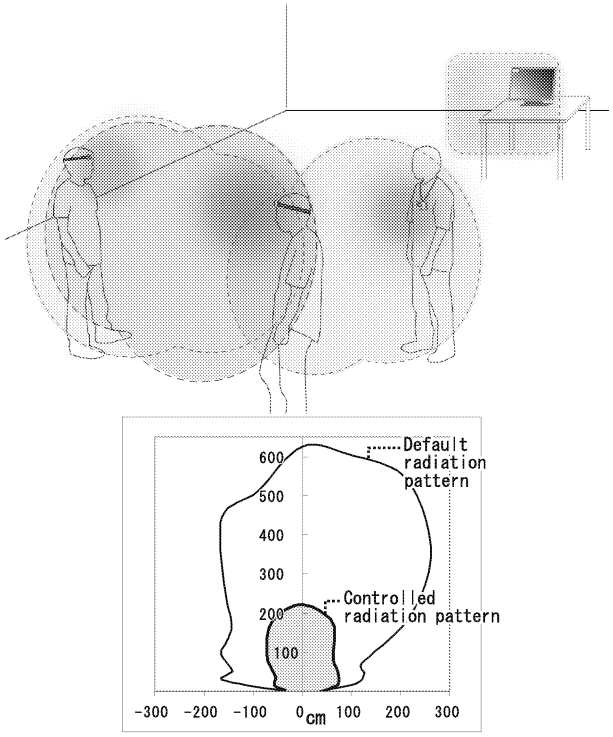


図 4 LED 配光範囲制御による Interaction-Scope のイメージ
Fig. 4 InteractionScope is assembled as a radiation pattern of the IrID Tag.

ID 番号をマンチェスタ符号化した 200Hz の明滅で発信する。トラッカは、約 90 度の画角、128 x 128pixel の解像度と 400Hz のフレームレートを持つイメージセンサと、画面上に映った明滅するタグ LED の ID をデコードし、ID 番号と画面上の X-Y 座標を出力するマイコンからなる。画面内の全タグ n 個を 150+(100 × n) msec 時間で随時出力できる。移動に対する水平垂直方向の最大追従角速度は 56.25deg/sec である。運用では、参加者の頭部側面前方にタグとトラッカを視野カメラに並置して装着する (図 3)。

トラッカが視野カメラの撮影範囲をカバーし、撮影対象となった別の参加者をタグ ID 番号から互いに判定し、相手の相対位置と併せて記録できる。これにより、複数人が集合している、向き合っているといったグループ抽出が可能になる。さらに、接話マイクによる発話判定結果から、発話のターンテイキングを考慮した、対話中のディスカッショングループを取り出すことが可能になる [21]。

展示会場の行動場面では、ポスタ、展示物や展示空間からなる環境要素があり、展示物を利用して説明する説明者、展示物や説明者周辺で情報収集する来場者、空きスペースで談話する説明者や来場者といった、定型的行動がある (図 1)。このような視野対象となる環境要素であるポスタや展示物にはタグを取り付け、滞在空間となる展示ブースなどには同エリアを検出エリアとするトラッカを取り付ける (図 3)。これにより、来場者の滞在場所や視野方向、ディスカッションといったインタラクションを展示会場の行動場面の文脈から の解釈にあてはめることができる。

タグは、取り付け対象物の種類に応じて、トラッカから識別可能な範囲を LED の発光強度と放射角により制御できる。例えば、視線方向 2m 先に大型のポスタ展示がある場合と、同距離に PC のディスプレイがある場合を同じ視覚的行為と捉えることはできない。通常、高精細なディスプレイを使った作業がより近距離で行われることを考えると、後者で視覚的インタラクションが行われる可能性は低いと考えられる。このように視覚的行為が可能な範囲は、被視対象の解像度に依存し [10]、被視対象との距離や角度で決定できる [17],[37] とされる。本研究では、この範囲をインタラクションスコープ (図 4) と呼び、赤外線タグ装置による近似的実装のための調整を行っている。

運用の過程では、視野カメラによる記録画像が、必ずしも視野対象を捉えていない、トラッカによるタグ認識の範囲と精度が、実際の対話状況と合っていないことがある、など装置仕様や性能面の問題が明らかになることもあった。現在は、経験的な改修を重ね、視野カメラの画角を 90 度とし、トラッカの識別範囲を 4m 超とした実装で運用している。

次章以降では、これらの経験的に決定した記録画角や識別範囲について、検討する。展示会場を模擬し、立位姿勢で自由移動可能な空間でのディスカッション

を対象として、視野カメラが視線方向を十分に記録できているか、トラッカ識別範囲が、発話相手を十分捉えているか、について、視線測定装置と、1mm 精度で 3 次元位置を測定可能なモーショントラッカによる計測を行い、その結果について検討する。

4. 対話過程の視野分析

装着型体験記録装置を構成する頭部装着型カメラとトラッカについて、インタラクションをとらえる装置としての有効性を検討する実験を行った。多くの頭部装着型カメラを用いた既存研究は、情報取得段階の装置仕様の適切性に関する議論が無いため、実運用上の適切性や有効性の評価が困難であった。本章では、視線方向を頭部方向が代替することについて、視線計測装置を装着した被験者による対話実験から検討する。視線方向と頭部方向の差異から、頭部装着カメラによる視覚的体験映像の記録、およびトラッカによる視野対象判別の妥当性と、適切な記録・検出の範囲について述べる。

4.1 頭部方向と視線方向

視線は、視覚的体験を伴うインタラクションの記録と判別に最適な手掛かり情報である。視力は中心視 (視線方向約 2 度) から周辺視 (左右約 160 度) へ極端に低下する。文字や形状の認知的処理が可能な有効視野は約 20 度と限られるため、多くの視覚的体験は、その範囲内で起きると考えられる [27], [38]。すなわち、相手の様子を見ながら進められる人対人、人対物のインタラクションや、話題対象物となるポスタなどの情報源を共同注視しながら進めるインタラクションを捉えるには、視線方向の観測が重要な因子であると言える。また、視覚的体験を映像として記録するには、同範囲を含む、視線対象を構図の主役として、十分な画質が維持できる程度の画角設定が望ましい。

一方で、視線を常時着装可能な装置で検出することは困難である。視線の測定方法には、コイル内蔵型コンタクトレンズによるサーチコイル法や筋電位を計測する EOG 法といった接触型と、眼球へ照射した赤外線の反射率を計測する強膜反射法や角膜反射法といった非接触型がある [22]。接触型は装着者の負荷が大きく、非接触型も装置形状や計測可能な距離に制限があるため、実験室環境での利用を超えて、自由に移動する複数人の視線を観測することは困難である。このため、視野画像として頭部方向や体方向に同調するカメラ [20], [32], [35] やセンサ [8] を代替にする研究が多い。カメラの利用では、動視野全域をカバーする 180 度以上の広角カメラを使う例 [35] や、水平画角 40 度前後の通常のビデオカメラを使う例 [20] が見られる。センサ利用では、体前方の広範囲を検出域とする赤外線センサを使う例 [8] が見られる。我々も、前章であげたように、頭方向に同調するセンサ (赤外線 ID トラッカ) を用いた検出を行っている。

しかし、これらの姿勢に同調して検出範囲を決定するセンサが、視線を代替することの妥当性について言及した研究は少ない。ここでは、人対人の対話といった視覚的体験を伴うインタラクションを捉えるために、頭部方向が視線の代替となりうるか検討する。また、視線対象を包含する体験映像として、頭部装着カメラを代替とする適切性について検討する。

4.2 視線移動と画角

視線移動は眼球運動を司る外眼筋の収縮によるもので、全方向に約 50 度の可動域がある。眼球運動により中心視可能な範囲を注視野と呼び、この範囲では頭を動かさずに対象を見ることができる。頭部装着カメラとして、注視野と有効視野全体をカバーする画角を想定すると、約 120 度必要となる。これは標準的なビデオカメラの約 3 倍広角であり、同解像度での記録では画質の低下が著しく、コンテンツとして利用するには適さない。また、肩部や胸部へのカメラ装着では、脊椎 (首) の回旋範囲 160 度がこれに加わるため、体正面方向の記録を視界相当とするのは無理がある。また、肩部や胸部の姿勢変更を伴わない視野対象の変化を検出できない、という問題がある。通常の視線移動では、外眼筋のストレスから、視線が一定時間以上停留する注視状態には、頭と体の姿勢を変えて視線を正面へ定位させる指向運動が伴う。このため、注視野全体を記録可能な広い画角は必ずしも必要ではない。ただし、視線移動が指向運動を伴う程度は、環境因子や視線対象によって異なることが考えられる。装着型カメラを用いた既存研究では、記録画角や検出範囲への言及が無いことも多いが、これらを勘案した模擬的視野角を、記録状況に応じて経験的に選択していると考えられる。例えば、卓上作業など 1m 未満の短距離での記録では広角カメラを使う例が多く、比較的長距離の屋外景観を記録する場合には狭角の例が多い。卓上作業の多くは、近距離広範囲に視線対象物が分布し、短時間で視線対象が変化するため、頭部が視線の移動に追従する指向運動が起き難いことが広角カメラを使う理由と考えられる。長距離の場合は、その逆となる。具体的に建築物の外観設計時に想定すべき来訪者の視野角を 60 度とした研究 [34] もある。

我々が展示会場の来訪者体験を記録するために使った最初の頭部装着カメラは水平画角 44 度であったが [36]、視野対象物と考えられるポスタや対話者が画面から外れることが多く見られた。体験映像としての記録を想定する屋内の展示会場やフリーディスカッションといった場面を想定し、姿勢と視線移動に関する実験を

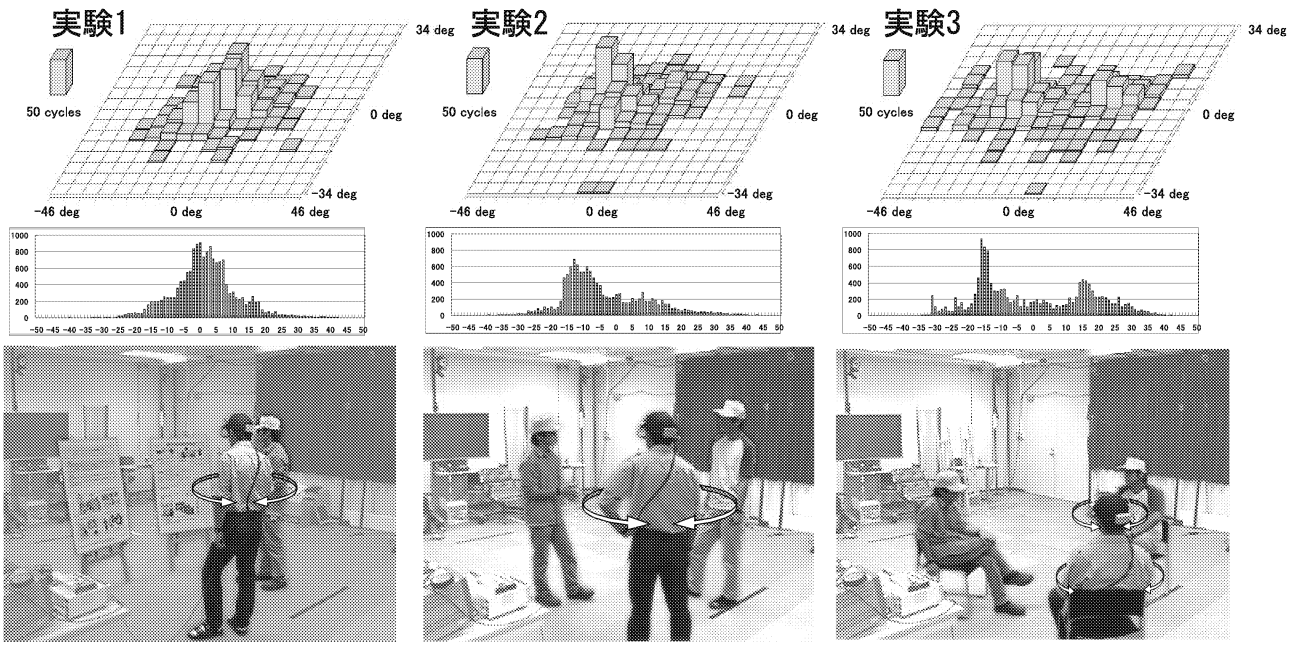


図 5 視線停留の検出結果と実験の様子
Fig. 5 Experimental results of eye fixation and experimental set-up.

行った。

4.3 実験設定

実験 1 として、展示会場を模して説明者 1 名が話題対象ポスタ 2 枚を用いて被説明者 1 名にポスタ内容の解説と議論を行う過程を記録した。実験 2、実験 3 として、フリーディスカッションを模して被験者 3 名による対話過程を記録した (図 5)。実験 1 と実験 2 は実験空間内で立ち位置や姿勢を自由に変えられる立位で行った。実験 3 は、等間隔で直径 2m の円状に配置した椅子に着席した座位で行った。実験時間は各 5 分とした。視線測定装置を装着する被験者は各 1 名である。実験 1 では、被説明者が装着した。それぞれ対話課題は、「体験記録用ウェアラブルセットの運用における問題の確認と解決」である。各実験の被験者延べ 8 名は本研究所員であり、話題に関する知識を持つ。視線測定装置としてナック製 EMR-8B [1]を用いた。瞳孔角膜反射方式により半径約 46 度の眼球運動を 0.1 度精度で検出可能である。被験者は約 250g の帽子型計測ユニットを装着する。

4.4 実験結果

各実験における視線停留回数の分布を図 5 に示す。視線の水平垂直移動を X-Y 軸方向で表し、視線停留の回数を棒グラフの高さで表す。ここでは、視線が 2 度の範囲で 100msec 以上動かない状態を視線停留とした。視線対象物を各参加者とポスタとした場合、各対象物間の視線移動は、実験 1 で 71 回、実験 2 で 80 回、実験 3 で 81 回であり、インタラクションの状態に応じた頻繁な視線対象の変更が観察された。

表 1 視線検出の結果
Table 1 Experimental results of gaze measurement.

実験	中心位置 X	中心位置 Y	標準偏差 X	標準偏差 Y	分布形状	最頻値
1	1.2	−1.5	10.1	5.4	単峰分布	0
2	−6.2	1.3	13.1	6.1	単峰分布	−13
3	−3.2	2.2	17.7	5.8	双峰分布	15, −16

実験 1 と実験 2 での視線停留は、頭部正面方向に対する中心を、それぞれ (1.2, -1.5)、(-6.2, 1.3) をとして、水平方向 10〜13 度、垂直方向 5〜6 度を中心とする水平垂直方向約 50 度の範囲の単峰型に分布した (表 1)。

視線対象物となるポスタや他参加者の相対位置は視線方向の分布に現れていない。このように、自由な指向運動が可能な立位で姿勢制限の無い条件では、視線対象物を注視野全域を使って追いかけるわけではなく、頭部を含む姿勢変更が伴う視線移動により、視野対象方向に頭部を定位させていることがわかる。しかし、その範囲は約 50 度であり、一般的な 40 度程度の狭角カメラを用いた体験映像記録[36] では、十分に視野対象を捉えているとは言えないことがわかる。

一方、座位で姿勢制限のある実験 3 では、最頻値が水平方向 15 度と-16 度の 2ヶ所に現れる双峰型に水平約 70 度の範囲内で分布した。視野測定装置を装着した被験者から、視野対象となる他参加者は視野面上、

1: EMR-8B, http://eyemark.jp/

左右30度の位置にあるので、被験者は約15度の視線移動と、約15度の姿勢変更により、視線対象物を捉えていることがわかる。この結果から、着席状態のように被験者が自由な姿勢変更をできない条件設定では、指向運動が制限されるため、狭角の頭部装着カメラによる体験映像記録は困難であると考えられる。また、このような姿勢制限のある状況を想定した場合、肩部や胸部へのカメラやセンサの装着は、検出範囲の広角化が要求され、視野方向との乖離が進むため、望ましくないと言える。

本稿では、体験映像記録場面として、比較的少人数かつ近距離でディスカッショングループが構成される、展示会場でのインタラクションを想定している。このような場面で、視線対象となるポスタや人など幅50cm～1m程度のモノは、想定されるディスカッションの距離2m～3m程度では水平画角約10～15度に相当する。視線対象物に装着されたトラッカを識別し、これを画角内におさめるのに、実験1と実験2で得た視線移動範囲に対象物相当の画角を加えると、画角80～90度程度が適切と言える。

結論として、実験結果から、体験映像記録用頭部装着カメラと対象物判定用のトラッカについて、経験的に決定した画角90度は、立位姿勢で自由な移動や姿勢変更が可能な空間では、適切であった、と言うことができる。また、姿勢への制約が無い状況であれば、視線の代替として頭部方向を用いることも可能である、と言うことができる。

5. 対話過程の相対位置分析

本章では、対話過程における話し手と聞き手の相対位置関係の分析を行う。トラッカによる視野対象検出と、検出結果を用いたインタラクションの記録と判別の妥当性について検討する。視線方向の代替である頭部方向を含む参加者間の相対位置により、対話過程の話し手や聞き手といったインタラクションをとらえることができることについて確認する。多くのインタラクションは視覚的体験を伴うものであり、発話相手の特定と視野方向などは協調的に行われるという仮説[16]のもとに、実装した。座位姿勢に限定したものであるが、実際に視野方向を用いてミーティング過程の判別を行う研究[30]はある。ここでは、高精度な設置型モーショントラッカを用いて、実際の立位対話過程での位置関係を計測し、頭部方向からの相対位置について、このような仮説を満たすものであるか、検討する。

人々が立位姿勢で行き交う自由移動可能な環境でインタラクションをとらえるには、対話過程にある話し手と聞き手を特定することが重要である。話し手は、音声パワーレベルの観測といった方法で、比較的容易に特定できる。一方で、聞き手の定義と特定には困難が伴う。対話集団が既定された実験室環境や、参加者の役割が明確な講演などの環境では、座席位置や役割などから聞き手が自明な場合も多く、特定作業も容易である。しかし、自由移動が可能な環境では、そのような固定的な情報に頼ることは出来ない。流動的に形成される対話集団の範囲を決定し、体験を共有する集団としての対話者を検出するため、聞き手の特定が重要である。ここでは、正解データとしての聞き手を、被験者である話し手自身が、話し相手と認識する対象者と定義する。

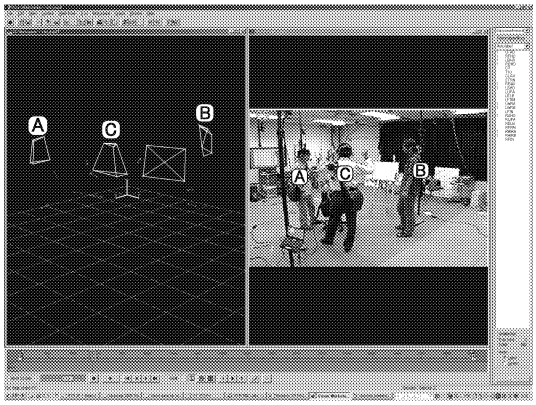


図6 Viconによる位置検出の様子
Fig. 6 Measurement of positions using Vicon.

5.1 実験設定

展示会場のように自由移動可能なディスカッションスペースでは、話題対象物としての掲示・展示物や、操作対象としてのパソコンや白板といった環境要素の存在が、通常は想定される。

展示会場を模した今回の実験では、被験者を本研究所所員3人とし、課題を「インタラクション記録用装着型ウェアラブル装置の改良」を話題とする10分間のフリーディスカッションとした。

話題対象物として、同装置を装着したマネキン1体、操作対象物として、メモ用の白板1面を準備し、被験者が自由に使えるものとした。被験者および対象物は、モーショントラッカによる位置検出用のマーカと、体験映像や音声記録用のウェアラブル装置を装着する。被験者は、室内を自由に移動でき、対象物の利用や移動もできることとした。

モーショントラッカには、Vicon Motion Capture System[2]を用いた。Viconは、7.5m × 10mの部屋外周に設置した12台の赤外線照射装置付赤外カメラと、再帰性反射材製の直径1cm球形パッシブマーカから構成される。各カメラ画像上での各マーカの2次元位置をもとに各マーカの3次元位置を再構成する。測定

2: Vicon, http://www.vicon.com/

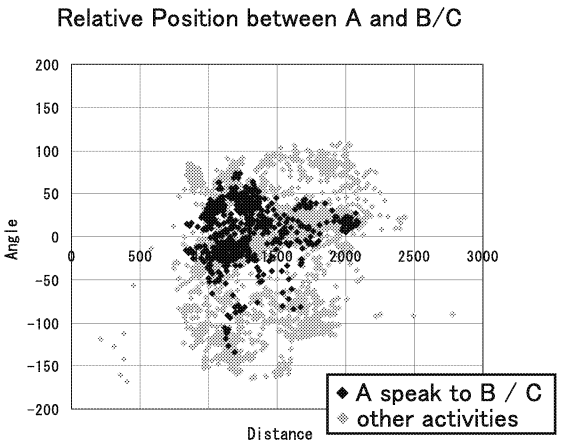


図 7 被験者 A からみた相対位置マップ
Fig. 7 Relative positional map seen from the subject A.

可能範囲は 2.5m × 3.5m 四方であり、本実験での時間分解能は 60Hz、空間分解能は約 1mm である。図 6 に示すように、マーカを対話空間で視線対象物となり得る人や物に貼り付けた。人の取り付け位置は、頭部 4 点 (帽子使用)、肩 3 点 (肩先 2 点と右肩甲骨上 1 点) である。被験者 A の頭部マーカ左前を *LFHDa*、右前を *RFHDa*、中央後を *BAHDa* とし、*LFHDa* と *RFHDa* の中点を *CFHDa* としたとき、頭部方向を

$$\vec{va} = \overrightarrow{\mathrm{BAHDaCFHDa}} \qquad (1)$$

とする。また、*CFHDa* を体の位置とした。これらの 3 次元座標情報は、サーバに直接蓄積される。各体験記録装置で記録した映像と音声は、予め NTP(Network Time Protocol) で時刻合わせしたクライアント側で時間情報を付与し、無線 LAN 経由でサーバへ蓄積する。音声のうちスロートマイクの入力は、パワーレベルでゲート処理した結果を被験者の発話区間として蓄積する。以上の機器を用いて、この実験では、10 分間の体験映像、音声、ユーザおよび環境要素の絶対位置、話者特定可能な発話記録を得る。また、実験参加後の被験者により、記録映像を参照しながら、話し手である被験者自身が話し相手と認識する対象者を聞き手としてハンドラベリングし、話し手と聞き手の関係を記録した。なお、聞き手を意識していない、複数人いる、といった発話は、ラベリングの対象外とした。ラベリング対象外となった発話割合は、時間比で全体の 5.29 %、回数比で 11.76 %に相当し、発話時間 10 秒未満の相づちなどに多かった。

5.2 実験結果

聞き手を特定可能な発話について、発話者自身によるハンドラベリングを行った結果にもとづき、被験者 A と B,C 間の相対位置を距離と角度から 3Hz でマッピングした図 7 を示す。標準偏差は、実験全時間の距離:349.11cm、角度:61.08 に対して、発話中の聞き手は距離:256.24cm、角度:31.75 に位置し、聞き手 (=話し相手) を特定した発話では、相手を頭部正面方向に定位させていることがわかる (表 2)。観測された対話過程の 88 %は、頭部装着カメラおよびトラッカの持つ画角 90 度の範囲内で起こっている。ここから、自由移動可能な空間においては、トラッカが、話し手と聞き手によるインタラクション過程を視野方向の代替となる頭部装着トラッカで識別するのに十分な性能を持っていることが確認できる。

表 2 実験時間全体と対話中の相対位置比較
Table 2 Comparison of relative position between the time of speaking and the whole experimental time.

	全時間		発話時	
	Distance	Angle	Distance	Angle
平均	1492.617	−10.0977	1270.237	5.002968
分散	121869.6	3730.571	65572.09	1006.723
標準偏差	349.1063	61.07981	256.2416	31.7501

一方で、識別範囲外となる外れ値も観測されているので、その特徴について、以下で検討する。

5.3 外れ値に関する議論

本実験時間中に外れ値として観測されたのは、A-B 間関係で 1 回 5 秒間、A-C 間関係で 1 回 4 秒間と、時間的には短い。

被験者 A-B 関係では、A から見た B の相対角度が ± 50 度の範囲にほぼ集中しているのに対して、相対距離は 70cm〜2m と広範囲に分布している。このような外れ値が観測された状況での、被験者の動きを図 8、図 10 に示す。図 8 は、相対距離が 2m 前後に広がった状況での被験者 B の動きを矢印で表し、B の視野映像の変化を表したものである。図 10 は、A-B 間の相対位置をマッピングしたもので、被験者 B の移動が観測された時間を明色で表す。被験者 B は、A-B 間対話継続中に、マネキンが装着する体験記録装置に話題が及び、これを参照し、発言に利用するため、立ち位置を大幅に変えていた。この様子は、B の視野カメラ (図 8) にも記録されている。

被験者 A-C 関係では、A から見た C の相対距離が 1m〜1.5m と安定しているのに対して、相対角度は ± 50 度に集中する一方、100 度を越える範囲にも分布しており (図 11)、眼球運動の限界を考えても、相手を全く見ずに対話を継続している時間があったことが考えられる。このような外れ値が観測された状況での、被験者の動きを図 9 と図 11 に示す。被験者 A は、C との対話と並行して、白板を使った話題の記録を行っていたことが観察されている。実際の被験者 A と被験者 C の視野カメラの様子を図 9 に示す。C が A を

図 8　AB 間の相対距離によらないインタラクションの例
Fig. 8　An example of interaction with outlier in relative distance between A and B.

図 9　AC 間の相対角度によらないインタラクションの例
Fig. 9　An example of interaction with outlier in relative angle between A and C.

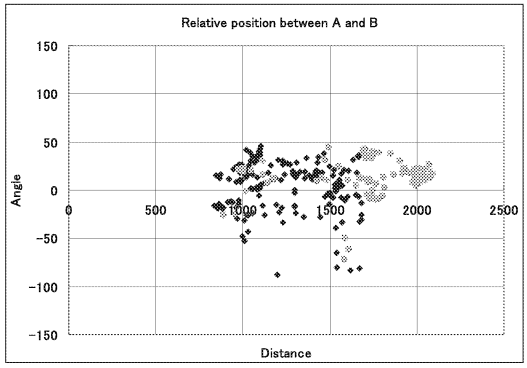


図 10　AB 間の相対距離によらないインタラクションでの相対位置マップ
Fig. 10　Relative positional map of interaction with outlier in relative distance between A and B.

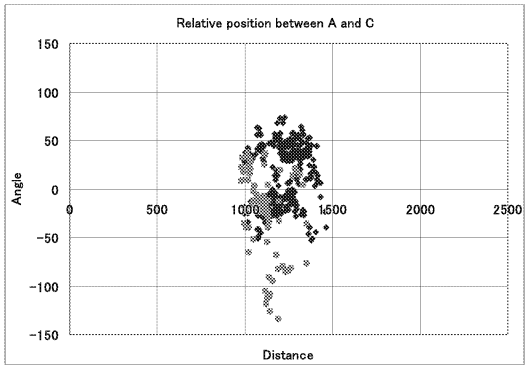


図 11　AC 間の相対角度によらないインタラクションでの相対位置マップ
Fig. 11　Relative positional map of interaction with outlier in relative angle between A and C.

見続けているのに対して、A は C と白板を交互に見ていることがわかる。

これらの実験で観測された外れ値は、共通して、対話相手との二者間関係に加えて、マネキンや白板といった環境要素が関連する状況で発生している。固定された環境要素により「自由」な移動が制限された状況とも言える。類似した状況として、第 4 章の実験 1 ではポスタ発表を模擬した被説明者の視線方向を記録している。被説明者の発話とは無関係の記録なので直接比較はできないが、ここで被説明者の視線対象物は、時間比で 1:6 と圧倒的にポスタを見ている時間が長かった (図 12)。

このような、対話過程に対話相手以外の操作対象や話題対象物が関係する場合には、モダリティの衝突や選択が行われていると考えられる。例えば、A-C 間の外れ値観測例では、白板と対話相手を同時に見ることができない状況で、白板の視覚的確認と書き込み操作、対話相手との音声対話の両方が並行して行われている。これらモダリティの意図的選択は、各参加者の行為目標に依存するが、そのような選択が可能か否か

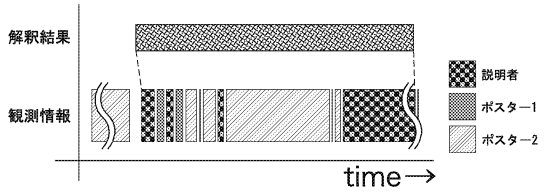


図 12 第 4 章実験 1 での視線対象の変遷と解釈の例

Fig. 12 Examples of the transition of the gaze object in the experiment 1 of chapter 4 and the interpretation result.

は、白板のような環境要素の有無で決まる。例えば、ポスタ発表の見学者がポスタではなく手元の資料を見る、といった状況は容易に想像できる。本システムでは、説明者以外のポスタや資料にタグを付けることで、そのような状況には対応できる。

外れ値が観測される状況は、体験コンテンツの記録という側面では、問題は生じない。装着型記録装置では、白板への書き込み過程を映像で記録し、対話を音声で記録できる。また、第 4 章実験 1 の状況でも、説明者の音声記録と、説明対象のポスタの映像記録は、むしろ再活用可能な体験記録として望ましいものと考えられる。

しかし、インタラクション判別の側面では、本稿でここまで述べた判別手法をそのまま適用できないため、問題が生じる。対策の一つは、「(人が) ポスタに話しかけることはない」といったヒューリスティクスにより個々の発話を判別する方法である。しかしこの状況で個々の発話を取り出した場合、人による判別すら覚束ない発話もあり、正確性は期待できない。もう一つは、外れ値状況の前後で時間的に継続する対話関係がある場合、その時間的継続性を含んで、「白板が使われた A-C 間の対話」といった、よりメタな単位でのインタラクション判別を行う方法である。発話間の継続性を判定する最大インターバル時間に閾値 (今回は約 3 秒) を設定することで、発話や白板利用といった個々のインタラクションを集合的に捉えることが可能になる。このようにメタな単位でのインデクスは、映画の 1 シーンを振り返るように体験記録を再活用する用途では、有用である。第 4 章の実験 1 のようなポスタが関係する対話状況からも、「ポスタ A とポスタ B を見ながら被説明者 A と対話した」という体験記録を容易に生成することができる (図 12)。

一方で、この判別手法では、時間的連続性の判断が必要なため、判別の即時性が無くなるという問題もある。そのため実運用上では、個々の発話判別と、集合的判別のメタ解釈レベルを、判別に必要な情報収集範囲の広さから 4 層に分割し、並行処理する方法を採っている [31]。本稿で装置特性を確認した生情報のレベルから、時間的連続性の解釈が必要なレベル、単体の装着型装置で判別可能なレベル、複数台 (他者) の情報をもって判別可能なレベル、といった分割である。

6. おわりに

外部観測可能な非言語行動を用いたインタラクション判別法として、視覚的行為可能な範囲を近似するインタラクションスコープを提案した。インタラクションスコープは、視野映像記録と視野対象認識の範囲を決定するトラッカ画角と、タグ認識範囲を決定する LED 放射角、それぞれのハードウェア実装によって、実現した。装置は、装着型実時間サービスを実現する上で重要な、認識の即時性や低処理負荷、小型といった特徴を持つ。

本稿では、経験的実装が多くあった仕様の適切性を定量的に確認した。トラッカ画角は、立位姿勢での適切性が確認された。また、着席姿勢のように身体運動を制限する環境要素によっては、適切な画角が変化することが明らかになった。タグ LED 放射角は、立位姿勢で自由移動可能な空間での対話相手特定に有効であることがわかった。また、人々が極端に密集している場合や、話題対象物を利用するための移動といった環境要素の影響がある場合は、トラッカ装置単体での即時的判別は困難になることがわかった。この場合、時間的連続性のある複数の判別結果を用いて、より大きな単位でのインタラクションとして判別できることが確認できた。

社会的に浸透するヒューマンインタフェース [36] の実現を目指す上では、本稿で提案したインタラクションスコープのように、非言語行動に着目した非侵襲なインタラクションの判別モデルが必要とされる。

今後、得た知見をもとに判別モデルの精緻化と実効的実装を進める。

謝辞

システムの開発と検証には高橋昌史、市原貴雄、山本哲史の諸氏が携わっている。鈴木紀子、坊農真弓の諸氏には、インタラクションコーパスについて議論頂いている。これらの諸氏に感謝する。本研究は、情報通信研究機構の研究委託「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」により実施したものである。

参考文献

[1] Phillippe Aigrain, HungJiang Zhang, Dragutin Petkovic: Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Re-

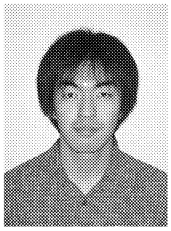
view; in Proc. of Multimedia Tools and Applications, Vol.3, No.3, pp.179–202, (1996).
[2] Christopher Alexander, Poyner Barry: Atoms of Environmental Structure; (1966), in Developments in Design Methodology, pp. 123–133, (1984).
[3] Roger Baker: Ecological Psychology: Concepts and Methods for Studying Human Behavior; Stanford University Press, Stanford, Ca., (1968).
[4] Mayumi Bono, Yasuhiro Katagiri: Interaction Analysis of Multi-party conversation; International Symposium The Origins of language Reconsiderd, Kyoto, Japan, (2003).
[5] Richard Borovoy, Fred Martin, Sunil Vemuri, Mitchel Resnick, Brian Silverman, Chris Hancock: Meme Tags and Community Mirrors: Moving from conferences to collaboration; in Proc. of CSCW ' 98, pp. 159–168. ACM, (1998).
[6] Rodney A. Brooks, Michael Coen, Darren Dang, Jeremy De Bonet, Josha Kramer, Tom´as Lozano-P´erez, John Mellor, Polly Pook, Chris Stauer, Lynn Stein, Mark Torrance, Michael Wessler: The intelligent room project; in Proc. of the Second International Cognitive Technology Conference (CT' 97), pp. 271–278. IEEE, (1997).
[7] Barry Brumitt, Brian Meyers, John Krumm, Amanda Kern, Steven Shafer: EasyLiving: Technologies for intelligent environments; in Proc. of HUC 2000 (Springer LNCS1927), pp. 12–29, (2000).
[8] Tanzeem Choudhury, Alex Pentland: Modeling Face-to-Face Communication using the Sociometer; in Proc. of the International Conference on Ubiquitous Computing, Seattle, WA., (2003).
[9] Anind K. Dey, Daniel Salber, Gregory D. Abowd, Masayasu Futakawa: The conference assistant: Combining context-awareness with wearable computing; in The third International Symposium on Wearable Computers, pp. 21–28. IEEE, (1999).
[10] J. Gibson, A. Pick: Perception of another person's looking behavior; American Journal of Psychology, 76, pp. 386–394, (1963).
[11] Erving Goffman: Frame Analysis: An Essay on the Organization of Experience; Northeastern University Press, Boston, (1974).
[12] Edward Twitchell Hall: The Hidden Dimension; Garden City, N.Y., Doubleday, (1966).(日高敏隆, 佐藤信行訳, かくれた次元, みすず書房, 1980).
[13] Stephen S. Intille, Aaron F. Bobick: A framework for recognizing multi–agent action from visual evidence; in Proc. of the Sixteenth National Conference on Artificial Intelligence, pp. 518–525, Orlando, Florida, (1999).
[14] Sadanori Ito, Shoichiro Iwasawa, Kiyoshi Kogure, Norihiro Hagita, Yasuyuki Sumi, Kenji Mase: InteractionScope: Non-fixed Wearable Positioning for Location-aware System; in Proc. of Ubicomp 2004, the Sixth International Conference on Ubiquitous Computing, (2004).
[15] Tatsuyuki Kawamura, Yasuyuki Kono, Masatsugu Kidode: Wearable interfaces for a video diary: Towards memory retrieval, exchange, and transportation; in The 6th International Symposium on Wearable Computers (ISWC2002), pp. 31–38, IEEE, (2002).
[16] Adam Kendon: Some functions of gaze direction in social interaction; Acta Psychologica, 26, pp.1–47, (1967).
[17] Adam Kendon: Conducting interaction: patterns of behavior in focused encounters; Cambridge University Press, (1990).
[18] Cory D. Kidd, Robert Orr, Gregory D. Abowd, Christopher G. Atkeson, Irfan A. Essa, Blair MacIntyre, Elizabeth Mynatt, Thad E. Startner, Wendy Newstetter: The aware home: A living laboratory for ubiquitous computing research; in Proc. of CoBuild ' 99 (Springer LNCS1670), pp. 190–197, (1999).
[19] Kurt Lewin: Principles of Topological Psychology; New York, McGraw-Hill, (1936).
[20] Steve Mann: Humanistic intelligence: WearComp as a new framework for intelligence signal processing; in Proc. of the IEEE, Vol. 86, No. 11, pp. 2123–2125, (1998).
[21] Tetsuya Matsuguchi, Yasuyuki Sumi and Kenji Mase: Deciphering interactions from spatio–temporal data; 情報処理学会研究報告（ヒューマンインタフェース）, 2002-HI102-4, (2003).
[22] Takehiko Ohno, Naoki Mukawa: A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction; in Proceedings of the symposim on ETRA 2004: eye tracking research and application symposium, pp. 115-122, (2004).
[23] Motoyuki Ozeki, Yuichi Nakamura, Yuichi Ohta: Human Behavior Recognition for an Intelligent Video Production System; in Proc. of 3th Pacific-Rim Conf. on Multimedia (PCM), pp. 1153–1160, Hsinchu, Taiwan, (2002).
[24] Patrick Chiu, Ashutosh Kapuskar, Sarah Reitmeier, Lynn Wilcox: Meeting capture in a media enriched conference room; in Proc. of CoBuild ' 99 (Springer LNCS1670), pp. 79–88, (1999).
[25] Alex Pentland: Smart rooms; Scientifc American, Vol. 274, No. 4, pp. 68–76, (1996).
[26] Maria da Graca Pimentel, Gregory D. Abowd, Yoshihide Ishiguro: Linking by interacting: a paradigm for authoring hypertext; in Proc. of Conference on Hypertext, pp. 39–48, (2000).
[27] Keith Rayner, Arnold D. Well, Alexander Pollatsek: Asymmetry of the effective visual field in reading; Perception and Psychophysics, 27, pp. 537–544, (1980).
[28] Cees G. M. Snoek, Marcel Worring: A review on multimodal video indexing; In IEEE International Conference on Multimedia and Expo, Vol. 2, pp. 21–24, Lausanne, Switzerland, (2002).
[29] Robert Sommer: Personal Space: The Behavioral Basis of Design; Englewood Cliffs, N.J., Prentice-Hall, (1969).(穐山貞登訳, 人間の空間, 鹿島出版, 1972).
[30] Rainer Stiefelhagen, Jie Yang, Alex Waibel: Modeling focus of attention for meeting indexing based on multiple cues; IEEE Transactions on Neural Networks, vol. 13, no. 4, pp. 928-938, (2002).
[31] Masashi Takahashi, Sadanori Ito, Yasuyuki Sumi, Megumu Tsuchikawa, Kiyoshi Kogure, Kenji Mase, Toyoaki Nishida: A layered interpretation of human interactions captured by ubiquitous sensors; in Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences(CARPE'04), pp. 32–38, (2004).
[32] Masashi Toda, Takeshi Nagasaki, Toshimasa Iijima, Toshio Kawashima: Structural representation of personal events; in Proc. of the ISPRS

working group V6, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV-5 W10, (2003).
[33] Jaco Vermaak, Patrick Perez, Michel Gangnet, Andrew Blake: Rapid Summarisation and Browsing of Video Sequences; In British Machine Vision Conference, BMVC'02, Volume 1, Cardiff, UK, May (2002).
[34] 芦原 義信: 外部空間の設計; 彰国社, (1975).
[35] 上岡 玲子, 廣瀬 通孝, 広田 光一, 檜山 敦, 山村 明義: ウェアラブル体験記憶装置のための体験記憶および再生についての研究; ヒューマンインタフェース学会研究報告集, Vol.3 No.1, pp. 13–16, (2001).
[36] 角 康之, 伊藤 禎宣, 松口 哲也, Sidney Fels, 間瀬 健二: 協調的なインタラクションの記録と解釈; 情報処理学会論文誌, Vol.44, No.11, pp.2628–2637, (2003).
[37] 樋口 忠彦: 景観の構造 ランドスケープとしての日本の空間; 技報堂出版, (1975).
[38] 三浦 利章: 行動と視覚的注意; 風間書店, (1996).
[39] 宮崎 英明, 亀田 能成, 美濃 導彦: 複数のカメラを用いた複数ユーザに対する講義の実時間映像化法; 電子情報通信学会論文誌 J82-D-II, No.10, pp.1598-1605, (1999).

(2004 年 8 月 12 日受付, 1 月 6 日再受付)

著者紹介

伊藤 禎宣 (正会員)



2003 年北陸先端科学技術大学院大学知識科学研究科博士後期課程修了. 同年より, (株) 国際電気基礎技術研究所 (ATR) メディア情報科学研究所客員研究員. 2004 年より, 同所および ATR 知能ロボティクス研究所専任研究員. 博士 (知識科学). 知識処理システムや協調作業支援システムに興味を持つ.

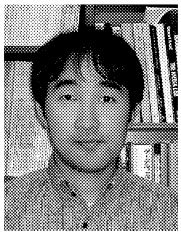
岩澤 昭一郎

2000 年成蹊大学大学院工学研究科博士課程単位取得退学. 1999 年通信・放送機構招へい研究員. 2002 年より ATR メディア情報科学研究所客員研究員, 現在同所および ATR 知能ロボティクス研究所専任研究員. 2000 年電気通信普及財団賞受賞. CG や画像処理の研究に従事. IEEE, ACM 各会員. 博士 (工学).

土川 仁

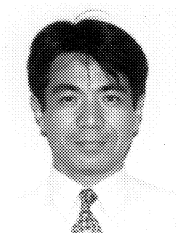
1990 年早稲田大学大学院 (機械工学) 修了. 同年, 日本電信電話 (株) (NTT) 入社. ヒューマンインタフェース研究所, サイバーソリューション研究所等において, 動画像認識, 医用画像伝送の研究に従事. 2003 年 7 月より, (株) 国際電気通信基礎技術研究所 (ATR) に出向.

角 康之 (正会員)



1990 年早稲田大学理工学部電子通信学科卒業. 1995 年東京大学大学院 (情報工学) 修了. 同年, (株) 国際電気通信基礎技術研究所 (ATR) 入所. 2003 年より, 京都大学大学院情報学研究科助教授. 博士 (工学).

間瀬 健二

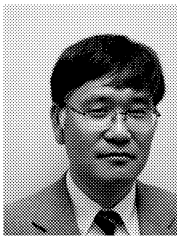


1979 年名大・工学部・電気卒. 1981 年同大大学院工学研究科修士 (情報) 課程修了. 同年日本電信電話公社入社. 1995 年 (株) ATR 知能映像通信研究所第二研究室室長. 2001 年 ATR メディア情報科学研究所第一研究室室長. 2002 年より名古屋大学情報連携基盤センター教授. 人工知能学会 1999 年度論文賞. 博士 (工学).

片桐 恭弘

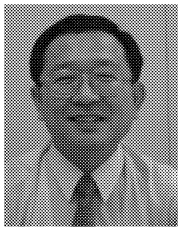
1981 年 3 月東京大学大学院工学系研究科情報工学専攻修了. 工学博士. NTT 基礎研究所を経て現在 ATR メディア情報科学研究所所長. 自然言語処理, 社会的インタフェース, マルチモーダル対話の認知科学の研究に従事. 日本認知科学会, 日本人工知能学会, 社会言語科学会, 自然言語処理学会, 情報処理学会,Cognitive Science Society, ACM, ACL, AAAI 各会員.

小暮 潔



1981 年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了. 同年, 日本電信電話公社に入社. 現在は ATR 知能ロボティクス研究所知識創造研究室室長. 博士 (工学). 自然言語処理, エージェント, ロボットなどの研究に従事.

萩田 紀博



1978 年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了. 同年, 日本電信電話公社 (現 NTT) 入社. 以来, 文字・文書・画像認識, コミュニケーション科学, インタラクション・メディア, コミュニケーション・ロボットの研究に従事. 工学博士. ATR メディア情報科学研究所長を経て、現在 ATR 知能ロボティクス研究所長. IEEE, 電子情報通信学会, 人工知能学会各会員.