

マルチモーダルインタラクション分析のための コーパス環境構築

來嶋 宏幸[†], 坊農 真弓^{†‡}, 角 康之[†], 西田 豊明[†]

kijima@ii.ist.i.kyoto-u.ac.jp

[†] 京都大学情報学研究科

[‡] 日本学術振興会

人と人のインタラクションの機械的な理解やコミュニケーション支援を強化するため、映像、音声、身体動作、視線といったマルチモーダルなデータを蓄積し、それらにインタラクション上の有意味なインデックスを付与したコーパスの構築を進めている。本稿では、複数種類のセンサを用いて会話の場を記録するための環境である IMADE ルーム及び、記録された大量のデータを用いた分析を支援するためのツールの試作について報告する。また、コーパス構築に向けたインタラクションの抽象的なインデックスの抽出手法について検討する。

Development of the Environment for Multimodal Interaction Analysis

Hiroyuki Kijima[†], Mayumi Bono^{†‡}, Yasuyuki Sumi[†], Toyoaki Nishida[†]

kijima@ii.ist.i.kyoto-u.ac.jp

[†] Graduate School of Informatics, Kyoto University

[‡] JSPS Research Fellow

For development of the system to understand and support human interactions, our research group construct Interaction Corpus which has a variety of multimodal data: video, audio, gesture, gaze and so on. In this paper we propose IMADE Room, which is the environment to record the sensor data in conversation area, the tool to analyze our corpus, and the method to extract the abstract-level annotation of human interactions.

1 はじめに

近年、日常的なミーティングにおける会話の音声や映像を収録し、ミーティング後のサマリなどに再利用するための研究が国内外を問わず盛んに行われている [1][2][3][4]。このような会話は、知識伝達や創造を行うための効果的な手段の一つである。我々はこのような知識の伝達や創造の強化・支援に向けて、より効果的かつ汎用性の高い会話再利用のための技術を構築したいと考えている。まず始めに我々の研究を進める上で重要となるのは、自然な会話データの中から何らかの有意義な行動データを発見し、それを検索可能にすることである。そのためには会話中の映像等のデータに対して何らかの根拠に基づいたインデックス情報を付与する必要がある。

そのためのアプローチとして、我々は人と人のイ

ンタラクションの場を複数センサを用いて記録し、それらにインデックスを与えたコーパスの構築を試みている。ここで言うコーパスとは従来の言語データの集合を指すわけではなく、人と人のインタラクションの自動認識手法を研究していくに当たって必要な研究基盤となる、マルチモーダルなデータの集合を指す。また、ここで扱うインデックスとは例えば、会話の盛り上がりや主要話者といった、人間は暗黙的に理解出来るが単一のセンサデータでは機械的には抽出し得ない抽象的な情報を指す。コーパスにインデックスを付与することにより、会話の構造をより多重的に観察可能になり、会話再利用のための有益な情報の検索・発見が期待できる [5]。

我々はこれまで展示見学の場を対象としてデータを収録し、データを階層的に解釈することでインタラクションのインデキシングを行ってきた [6]。文

献 [6] では、簡易なウェアラブルデバイスを用いることで人の出入りがあるような開放的な場所における多人数インタラクションを記録し、展示会参加者の注視対象と発話行為の組み合わせから基本的なインタラクションの意味の解釈を試みた。それによって、ある程度、展示会全体としてのインタラクションを機械的に理解することや展示見学者のハイライトシーンを見つけることが出来ることを示した。反面、会場に点在する一つ一つの会話の場の意味的解釈としては解釈する材料となるデータの種類の精度が低く、そこから導き出されるインデクスの信頼性に問題があった。

そこで今回我々は詳細にインタラクションの意味を理解し、インデキシングを行うことが出来るような環境のデザインを目指す。ポスターを参照しながら会話する予備のデータを分析した結果、身振り手振り、指差し、視線情報といった非言語情報が会話の盛り上がりといったインデキシングを行うための指標として有効であるという知見が得られている [7]。

また、コーパスを構築していくに当たり、蓄積された膨大なデータの中から分析者が興味のあるシーンを切り出してくるための仕組みや切り出したシーンのマルチモーダルなデータを同期しつつ閲覧・分析するための環境が必要であることが明確になった。我々は文献 [6] のコーパスに関して、前者のシステムとして CorpusViewer[8] を、後者のシステムとして iCorpusStudio[9] を現在試作している。

本稿ではインタラクション・コーパス構築に向けた会話周辺の非言語情報収集環境としての IMADE (Interaction Measurement Analysis and Design Environment) ルーム及びデータ閲覧や分析、コーパス構築の支援を行うために試作した iCorpusStudio について述べる。また、それらを用いたインデクスの付与手法及び会話における有意な単位の抽出方法を提案する。

2 関連研究

これまでにも、AMI[10]、NIST[11] など、会話の場をカメラやマイクロフォン等の複数センサ類を用いて記録・分析し、インタラクションのコーパスを構築する試みが多くなされてきた。これらは、対象をミーティング形式の会話に絞って大量に蓄積し、それらのデータをコーパスとして外部に提供している。同様に VACE[12] もミーティング形式の会話を収録・蓄積したコーパスであるが、AMI や NIST が音声情報を主とし、非言語情報を音声の補助的なデータと捉えているのに対して、VACE では言語情報・非言語情報を等価に扱ったマルチモーダルインタラクションの分析を行っているという違いがある。マルチモーダルインタラクションの分析を試みる点では

我々と同様のアプローチであるが、VACE ではその対象を被験者を着席させメンバーや話の内容を固定したミーティング会話に絞っている。一方我々は会話場の発生や場への人の出入りなどといった会話場創出のダイナミクスに興味がある。従って対象とすべき会話は実験室的な会話ではなく、より実世界的な会話、すなわち、ミーティングというドメインに捉われておらずある程度自由に動き回ることができ、会話への人の出入りや会話内容の変遷といった動きが見られるような会話である。

本稿ではインタラクション・コーパス構築に向けて試作した、マルチモーダルなデータを統合的に閲覧・分析するための環境を紹介する。先行研究として、記録された会話の映像・音声などのデータを閲覧したり、人手によりアノテーションを付けるためのシステム (Anvil[13] や WaveSurfer[14]、NXT[15]、AMI の JFerret、VACE の MacVisSTA[16]) などが現在開発されており、システム間のデータ共有方法についても議論が進められている [17]。これらのシステムと我々が試作した iCorpusStudio の基本性能の比較を表 1 に示す。従来の会話分析では、単一のビデオデータを見ながらの分析作業が行われてきた。そのため過去に扱われてきた Anvil や WaveSurfer は単一ビデオデータのみの対応となっている。しかし、我々が扱いたいのは単一のビデオデータだけではなく、センサから取得された時空間的に大量にあるマルチモーダルなデータである。その点、JFerret、MacVisSTA などのコーパスを構築しているプロジェクトにおいて開発されるツールは、多視点映像やそのコーパス特有のデータに対応している。JFerret は AMI の主目的の一つであるアノテーションを用いたミーティングブラウザとして開発された。そのため、可視化されるデータの種類の他に会話内容の書き起こしやミーティングで使用されたスライド、ホワイトボードのストロークと多岐に渡る。また、検索機能を搭載することで、自分の興味のある部分だけを集中的に閲覧出来るという特徴がある。VACE の MacVisSTA も、複数映像やモーショントラッキングの 3 次元座標、音声波形を閲覧しながらアノテーション付与作業を行える。

しかし我々が必要とするのは、従来のシステムが目指してきたような対象のシーンについて単にアノテーションを付与するだけでなく、大量にあるコーパスのデータの中から分析者が閲覧したいシーンを見つけ出したり、より上位にある社会関係や人間関係といった、抽象的なアノテーションを付与する方法を検討するための環境である。従って、分析者の要求に応えながら閲覧する環境や、IMADE ルーム (詳細は 3 節で述べる) 特有のデータに特化していて、かつ分析者を支援するような機能のあるシステムが

	Anvil	WaveSurfer	NXT	JFerret	MacVisSTA	iCorpusStudio
複数映像	×	×	×	○	○	○
音声波形の表示	○	○	×	×	○	○
ラベリング作業	○	○	○	×	○	○
検索機能	×	×	×	○	×	○
インデックスデータの形式	XML	CSV	XML	XML	XML	CSV

表 1: 会話分析ツールの基本性能の比較

必要である。

iCorpusStudio では分析に必要な機能を持たせつつ (表 1 右側), 特有のデータを扱う枠組みとして, 生データに対して閾値処理を行うといった, プリミティブなアノテーションの自動抽出を試みるための環境を提供する。また, 分析を支援するための試験的な機能として, イベントの共起性に着目し, ラベル間のオーバーラップを手軽に検出したり, 各ラベルの頻度の算出といった基本的なデータの演算を行える機能を実装した。

3 IMADE ルーム

3.1 システム概要

人と人のインタラクションを機械的に理解するために, まずは複数センサ群を用いて詳細に記録することを試みる。そのための環境として我々はIMADE ルームを試作した。IMADE ルームの特徴を挙げると以下ようになる。

- 人のインタラクションを構成している様々なモダリティを各種センサを用いて記録する。
- 複数種類のセンサ群を協調的に用いることで, インタラクションの場を多角的に記録する。
- 実世界的なコーパス構築のために, 一つのドメインに捉われず必要に応じて柔軟にセッティングを変更可能である。

IMADE ルームはスタジオではなく, 大学の研究室の一室 (1125 × 740cm) を利用しており, 日常的な空間にセンサを配置した形になっている。図 1 にIMADE ルームのシステム構成を示す。システムは, 環境側に設置されるセンサ群と会話参加者が装着するセンサとに大別される。環境側センサの配置を変更することで, 様々なドメインのデータ収録に対応出来る。

3.2 各種センサ

身体の動作を記録するためのモーションキャプチャシステムとして PhaseSpace 社製のものを利用した。これは赤外線を発する LED のタグを CCD センサ搭載のカメラで LED の光を受け取ることで LED タグ

の 3 次元座標を計測する方式を取っている。収録の際には 30Hz でデータを取得し, タイムスタンプを付与しつつデータベースへ記録するようにしている。

視線情報を記録するためのアイマークレコーダには, ASL 社製の MobileEye を利用した。これはデジタルビデオカメラに接続されたゴーグルを被験者が装着し, 眼球に反射した赤外線を検知することで人の視線方向を検出する。これにより, 1 人称視点映像に装着者の注視点を表すマークを重畳表示したビデオとビデオ上のどこを見ているかという 2 次元座標を取得出来る。座標データに関してはモーションキャプチャ同様タイムスタンプを付与しながらの記録を試みている。

人が装着する無線マイクロフォンには, SHURE 社製の WH30-XLR ヘッドウォーンマイクを用いた。また, ATI 社製オーディオプリアンプで増幅し, 48kHz, 16bit の PCM データとして各チャンネル独立して記録するオーディオキャプチャシステムを構築している。これにより, 部屋の中を自由に歩き回る事や各話者の音声を独立に記録する事が可能となった。

会話参加者の行動を記録するためのカメラシステムを試作している。現在は, SONY 製のパンチルト一体型カメラ 8 台を PC に 1 対 1 対応させてハードウェアエンコードを行い, Mpeg-2 形式での保存を行っている。しかし, フレーム送りの操作に弱かったり, 時間情報の付与が困難なため他センサとの同期が難しいという問題がある。これらを解決するために, IEEE1394 カメラから取得された画像を Motion-JPEG の形式で保存する枠組みを現在試作している。これにより映像の切り貼りが容易になり, 3.3 節において説明するシステムで扱いやすくなる。

以上のセンサ間の対応を取るのにタイムスタンプを用いているため, 各 PC の時刻が重要な基軸になる。そこで, 1 台サーバとして NTP(Network Time Protocol) のサーバを構築し, それを各 PC に参照させることで時間のズレを軽減している。

3.3 記録データ抽出環境の試作

データベースには各センサの時間情報の他に各実験の基本情報 (実験実施者や日付など) を登録する

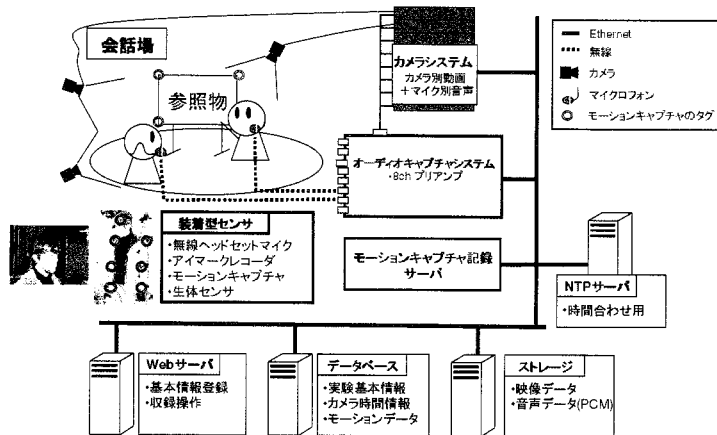


図 1: IMADE ルームのシステム構成

ことで、後からデータを取ってくる際の目安となるようにしている。しかし、データを閲覧したい人間が誰でもデータベースの扱いに慣れているわけではなく、またセキュリティの問題を考慮しても誰でも自由にアクセス出来る状態は避けたい。そこで、Web ページを通して対話的にデータベースとやり取りすることで、必要なリソースのダウンロードを行える様な仕組みを試作している。データを取得したいユーザは、複数データであれば当然時間同期されたものを必要とすると考えられる。そこでユーザからは、

1. どの実験のデータかを選択
 2. 必要とするリソースの種類 (どのカメラのビデオか、誰の音声か) を決定
 3. どの時間帯のデータが必要かを選択
- を順次フォームを通して入力することで、サーバ側はデータベースを参照後、指定された時間の分だけ抽出した映像、音声、モーションのデータを提供する。現在はテキストベースのシンプルな入力フォームであるが、将来的には、大量にあるデータの中から分析対象となるシーンを発見するための環境として、例えば映像であればストリーミングでその場で確認出来るようなより対話性の高いシステムを実装したいと考えている。

4 iCorpusStudio の試作

3.3 節で、多角的に記録したインタラクションのデータをデータベースを意識せずに取得出来る仕組みの試作について述べた。ここから取得出来るデータは時間同期の取れているマルチモーダルなデータである。2 節でも述べたが、これらを扱うには従来の会話分析を行うツールでは力不足であり、IMADE

ルームのデータを分析することに特化したシステムを構築する必要がある。今回試作した iCorpusStudio には閲覧環境 (4.1 節)、アノテーション付与作業環境 (4.2 節)、分析・データ処理環境 (4.3 節) としての 3 つの側面を持たせた。

4.1 閲覧環境

IMADE ルームで取得出来るセンサデータの内、可視化したいデータとして、環境カメラの映像、アイマークレコーダの 1 人称映像、モーションキャプチャの 3 次元座標データ、各音声の音声波形、付与されたアノテーションといったものが挙げられる。

試作した iCorpusStudio を利用している様子を図 2 に示す。下側にあるパネルがアノテーションを表示するためのパネルで、左から右に時系列表示されている。アノテーションの時間軸は映像やモーションの時間と連動しているため対応関係がわかりやすい。また、全体の中からあるアノテーションが付与されたシーンだけを見たいと考えるユーザのために、アノテーションのキーワード検索機能を実装している。

4.2 アノテーション付与作業環境

iCorpusStudio は従来の会話分析ツールと同様のアノテーション付与環境も提供している。従来のツールを使っていたユーザにも馴染み易いように、インタフェースは Anvil や WaveSurfer を参考にした。アノテーションとして付与できる情報は、ラベルの開始時間、終了時間、アノテーションの内容である。これらは図 2 下部のパネル上でグラフィカルに行うことができ、対応する場面を見ながらの作業が可能である。

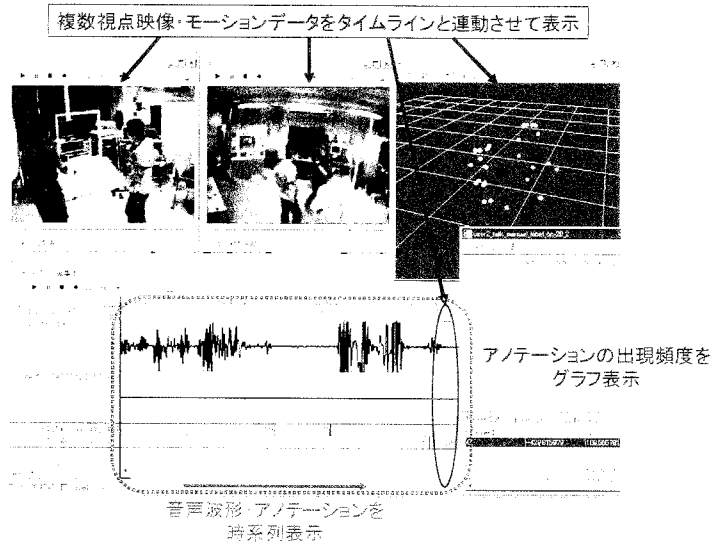


図 2: iCorpusStudio

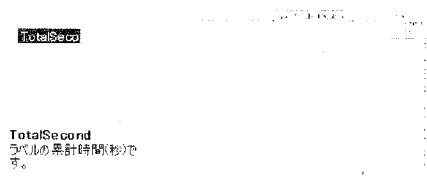


図 3: 出現頻度, ラベルの総数, 累計時間といった基本情報 (左) と出現頻度・累計時間の時間変化 (右)

4.3 分析・データ処理環境

従来のシステムではアノテーションを付与した後の分析, 例えばアノテーション同士のオーバーラップ関係を見たり, アノテーションの結果を統計分析にかけるといった操作は他のツールを利用していた. 一方我々が IMADE ルームで獲得したデータに関して行いたいのは単にラベル付けをして終わりではなく, それらを用いてインタラクション上有意なインデキシングを行うことである. 従って, iCorpusStudio には付与されたデータを用いて分析するための側面が必要である. 今回我々が iCorpusStudio に持たせた分析支援の機能には以下の様なものがある.

センサデータ処理

我々の興味はインタラクションの基礎となるプリミティブなインデクスから機械的にインタラクション上の有意なインデクスを抽出することで

ある. そのための基盤となるプリミティブなインデクスは手作業でつけたアノテーションの他にセンサデータを処理して抽出出来ることが望ましい. 今回は音声データに閾値処理及びセグメンテーションを行うことで発話のアノテーションを生成したり, モーションデータをベクトル演算することで指差し動作の自動抽出を試行出来る環境を試作した. iCorpusStudio 上ではこれらの処理に必要な閾値をユーザが自由に設定出来る. この上で様々な閾値の組み合わせで試行し, それを確認する (例えば, 手作業で付与した発話のアノテーションと比較する) ことで適切な閾値を見つけ出すことが出来れば, それらを用いてデータベースのデータを直接バッチ処理することは容易になる. 以上のように本研究は iCorpusStudio 上で半自動処理を行うこと自体が主目的ではなく, 将来的な自動処理を検討するための試行環境であることが特徴である.

アノテーションの出現頻度の算出

分析者がアノテーションを付与した結果から確認したい基本的な情報として, 各ラベルの出現頻度, 総数, 累計時間といったものが挙げられる. iCorpusStudio では, アノテーションに関してこれらの基本情報をすぐに確認出来る. また, ある一定の時間枠におけるラベルの出現頻度の変化を折れ線グラフ表示して可視化することで, そのアノテーションに関して特徴的な場面の確認も行える (図 3).

インタラクションの共起の検出

付与されたアノテーションから意味的なインデキシングを行うためのアプローチとして、インタラクションの共起性に着目し、アノテーション同士のオーバーラップの検出を可能にした。今回用意したオーバーラップの種類は図4の2種類である。オーバーラップした結果を他のアノテーション同様に扱うことで、頻度を確認したり、映像と合わせて閲覧してアノテーションに含まれる意味の考察を行うことができる。

これらの機能はまだ試験的なものであり、実際の分析の場面で利用していく中で改善していきたいと考えている。そのため現在、IMADE ルームを利用して談話構造や多人数会話の参与構造を分析している研究者に iCorpusStudio を利用してもらいながら、機能や操作性に関するフィードバックをもらっている。

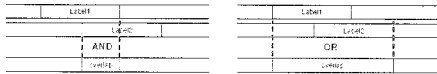


図 4: オーバーラップの例 (左 : AND, 右 : OR)

5 インタラクション上の有意味なインデクスの抽出

本節では iCorpusStudio を用いた分析手順の例と実際に IMADE ルームで収録した会話のデータについて、アノテーションの構造に着目することで見えてくるインタラクション上有意味なインデクスについて考察する。

5.1 iCorpusStudio の利用例

インタラクション・コーパス構築に向けて、収録したセンサの生データからなるレイヤー、発話・注視・指差しといった基本的な非言語情報からなるプリミティブなレイヤー、プリミティブなデータの分析から見出した上位の有意味なインデクスの付与されたレイヤーといった複数レイヤーの階層構造でインタラクションの理解を目指す。

センサの生データは単なる数値データであり、記録する際に絶対時間を付与されているだけであるが、実際に扱うには各データからアノテーションを抽出したい。例えば音声データであれば、まず音声のパワーの閾値を設定し 2 値化する。2 値化されたデータについて、ユーザが無音区間を例えば 300msec と設定してセグメンテーションを行うことで、300msec

以上 OFF が続けば、発話 OFF のアノテーションを、300msec 以内に ON があれば、開始時間を直前の ON、終了時間を次の ON とする発話 ON のアノテーションを付与する。

発話の開始・終了といったプリミティブなアノテーションがあれば、それらの共起関係を抽出することでそこに含まれる人の意図や意味を付与出来る可能性がある。例えば、会話参加者同士の発話のオーバーラップが頻出する場面は会話の盛り上がりを示す可能性があるし、お互いを見る eye contact や同じものを見る joint attention などにはなんらかの人の意図が含まれるといったことが考えられる。このような、参与者間、表現モダリティ間の時間的なオーバーラップを自動で検出して結果を可視化することにより、特徴のありそうなシーンを視覚的に特定することが出来る。

プリミティブなアノテーションの時間的オーバーラップ関係から特徴的なシーンが特定出来れば、そのシーンのデータを閲覧することでより上位のインタラクション上有意味なインデクスを付与したり、その他のデータと更に比較することで新たな特徴の発見を試みるということが可能となる。例えば、発話のオーバーラップの集中している箇所には「盛り上がっているシーン」という上位のインデクスを付与出来る可能性がある。このようにプリミティブなデータを組み合わせることで、単一のセンサでは機械的に理解し得ない、人の主観によって定まるような、抽象的なイベントさえも理解できる可能性があると考えられる。ただし、こういった手法によって抽出された、人間の主観に近く、抽象度の高いイベントは、アンケートや参与者に対するインタビューといった社会心理学的方法を用い、その妥当性を評価する必要がある。

以上の様に、IMADE ルームで収録された会話中のマルチモーダルなデータを iCorpusStudio 上で分析することで、我々がインタラクション・コーパスに盛り込みたいインタラクション上の意味のあるインデクスを自動検出する方法の研究が加速されると考えている。

5.2 事例分析

本節では IMADE ルームで収録された合意形成を志向した複数人インタラクションの実験 [18] で収録された実データについて、どのようなインデクスを付与することが可能かを検討する。

この実験の特徴は部屋の北側、南側にそれぞれ 2 枚ずつポスターを設置し、ポスター上にマグネットを配置するというタスクを与えることで被験者の立ち位置の変化が顕著に見られる点である。実験風景

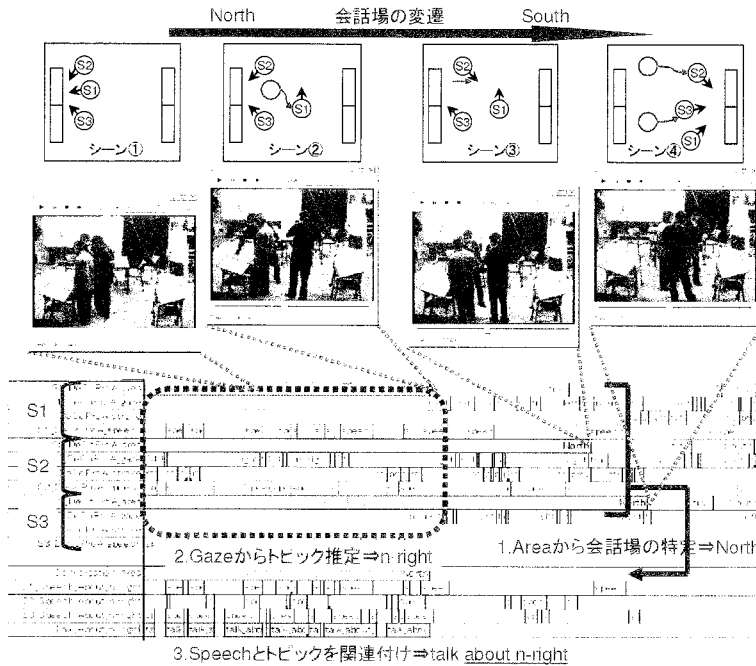


図 5: アノテーションの表示例

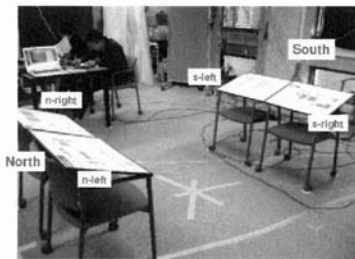


図 6: 多人数インタラクション実験風景

を図 6 に示す。タスクは来年度の研究室内の人の配置を考えることで、被験者は 3 人である。

収録したデータを元に、各個人の立ち位置 (北側に配置したポスター north, 南側に配置したポスター south), 注視情報 (North の内右側に配置した n-right, n-left...), 指差し等のストローク, 発話のアノテーションを 3 人×4 種類の計 12 チャンネル付与した。これらを iCorpusStudio 上で可視化・分析した例を図 5 に示す。

身体方向による会話場の抽出

この実験ではポスターが複数配置されているため、会話の場は各ポスター間を移動する。この移動す

る会話の場を iCorpusStudio を用いれば定量的に抽出できると考えられる。図 5 のシーン 1 では 3 人が北側を向いている。この時点での会話場は必然的に北側 (North) となる。そこから時間経過と共に一人、また一人と移動して行き全員が南側を向いた時点で、会話場は南側 (South) へと移ったと考えられる (シーン 4)。このように、会話参加者の身体位置のアノテーションがあれば、時間的なオーバーラップを取ることで会話場を抽出することが容易であることが分かる。

注視・ストロークによるトピックの推定

会話の場を抽出できたならば、次はその中身に着目する。ここで可視化された会話参加者の注視情報、ストローク情報を閲覧したり、オーバーラップを取ることで、例えば 3 人中 2 人が n-right に joint attention しているなど、n-right へ注目している情報が多ければ今現在の会話のトピックが n-right に関するものであることが推定出来る。従ってその区間の発話ラベルは単なる発話ではなく、「n-right に関する発話」という一段階意味的に解釈されたアノテーションへと変換出来る。このように非言語情報のインデックスを組み合わせることで、会話内容に踏み込むことなくある程度トピックの推定が行える。

行動シフト順序による主要参加者の推定

従来会話の場をコントロールする主要参加者の検出には発話の累積時間や発話率など、主に発話の情報を用いたアプローチが多かった。この実験では発話だけではなくそれ以外の要素、身体方向や視線情報が主要参加者の推定に関与する可能性を示すことが出来る。図5の例を見ると、最初3人が北側を向いている(シーン1)が、参加者の一人(S1)が北側から離れ南側へ行こうとする(シーン2)と、それに同調してS2が北側を離れ(シーン3)、最後にはS3も南側へとシフトする(シーン4)。この事例で観察したいのは、これ以外の会話場のシフトが今回同様S1に同調する形で成り立っていることが多いのか、それとも他の人から動き出すことの方が多いのか、そして実験後のアンケートの「主要参加者は誰だと感じたか」という設問の解答と、行動シフト順序の早い人間との相関関係が存在するかという点である。この相関関係を見出すことが出来れば、行動シフト順序の理解は会話の主要参加者を推定する有効な手法であると考えられる。

このように多人数インタラクションのデータを可視化・処理することで、インタラクションを構造的に分析することが出来る。マクロな視点からミクロな視点へとシフトしながら分析を行うことによって、プリミティブなデータにインタラクション上の有意な属性を付与するといった、より抽象的なデータの抽出が可能となると考えられる。

6 今後の課題

人と人のインタラクションを分析するための枠組みとして、マルチモーダルなデータをセンサデバイスを用いて記録するIMADEルームと、そこで蓄積されたデータを分析する環境であるiCorpusStudioについて紹介した。また、iCorpusStudioを利用してインタラクションを構造的に捉え、インタラクション・コーパス構築に向けてインタラクション上有意なインデクスを付与する手法について検討した。

今回iCorpusStudioでは、分析支援機能としてイベントの共起性に着目して可視化を試みた。今後は、更なる分析を進めるための統計的手法の統合や、自動抽出機能強化に向けたパターン認識技術を導入したいと考えている。また、研究に日常的に利用してもらうためにもインタフェース面の改善も行いたい。

謝辞

本研究は、文部科学省科学研究費補助金「情報爆発時代に向けた新しいIT基盤技術の研究」の一貫で実施された。IMADEワーキンググループの関係者の皆様による日頃の支援と議論に感謝します。

参考文献

- [1] 西田豊明. 会話情報学の構想. 人工知能学会誌, 21巻2号, pp.144-149, 2006
- [2] 坊農真弓, 高梨克也. 多人数インタラクション研究には何が必要か?-インタラクション研究の国内外の動向と現状-. 人工知能学会誌, 22巻5号, 印刷中
- [3] Steve Renals, Samy Bengio (Eds.). *Machine Learning for Multimodal Interaction, Second International Workshop, MLMI 2005. Lecture Notes in Computer Science, 3869, Springer, 2006.*
- [4] S.Renals, S.Bengio, J.G.Fiscus(Eds.). *Machine Learning for Multimodal Interaction, Third International Workshop, MLMI 2006. Lecture Notes in Computer Science, 4299, Springer, 2007.*
- [5] 坊農真弓, 高梨克也. 多人数インタラクション研究の方法-言語・非言語コミュニケーション研究のための分析単位とその概念-. 人工知能学会誌, 22巻6号, 印刷中
- [6] 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
- [7] 角康之, 熊谷賢, 瀬戸口久雄, 西田豊明. 非言語情報を利用した会話シーンの抽出と意味的インデキシング. 情報処理学会研究報告(ヒューマンインタフェース), No.2006-HI-119, pp.87-94, 2006
- [8] 角康之, 岩澤昭一郎, 間瀬健二. インタラクション・コーパス分析ツールの試作. 情報処理学会研究報告(ヒューマンインタフェース), Vol.HI104-7, 2003
- [9] 来嶋宏幸, 角康之, 西田豊明. インタラクション・コーパスの閲覧・分析環境の開発. インタラクション2006(インタラクティブ発表), 2006
- [10] S.Renals. AMI: Augmented multiparty interaction. In *Proc. NIST Meeting Transcription Workshop, Montreal, 2004.*
- [11] M.Michel, J.Ajot, and J.Fiscus. The NIST Meeting Room Corpus 2 Phase 1. In [4]
- [12] L.Chen, R.T.Rose, F.Parrill, X.Han, J.Tu, Z.Huang, M.Harper, F.Quek, D.McNeill, R.Tuttle, and T.Huang. VACE Multimodal Meeting Corpus. In [3]
- [13] Michael Kipp. ANVIL: A generic annotation tool for multimodal dialogue. In *Proc. Eurospeech 2001, pp.1367-1370, 2001*
- [14] <http://www.speech.kth.se/wavesurfer/>
- [15] J.Carletta, S.Evert, U.Heid, J.Kilgour, J.Robertson and H.Voormann. The NITE XML Toolkit: Flexible annotation for multimodal language data. *Behavior Research Methods, Instruments, and Computers 35(3) 353-363, 2003*
- [16] R.Travis, Francis Quek, and Yang Shi. MacVISTA: A System for Multimodal Analysis. In *Proc. of Int. Conf. on Multimodal Interface (ICMI), 2004*
- [17] <http://multimodal-annotation.org/>
- [18] 坊農真弓, 松村憲一. 多人数インタラクション理解のためのデータ収集デザインとその分析言語・音声理解と対話処理研究会(第48回). (JSAI SIG-SLUD). 人工知能学会研究会資料SIG-SLUD-A602 43-48, 2006