【パートII：情報分野研究者のためのオンリーワン共有イノベーションプラットフォーム】

5. IMADE：会話の構造理解とコンテンツ化のための実世界インタラクション研究基盤

角　康之[*1]　西田　豊明[*1]
坊農　真弓[*1]　來嶋　宏幸[*1]

＊1　京都大学情報学研究科

実世界インタラクションの理解と支援

Webやモバイル情報機器に代表される情報通信技術の発展により，我々は空間や時間の制約を越えて対話したり知識を共有することができるようになった．世界中の人が発信した膨大な情報から自分の興味に関係のある情報を探し当てることが曲がりなりにもできるようになったのは，言語的情報の蓄積や検索の技術の発展に負うところが大きい．そして，それを可能にした要因に，コンピュータで扱える言語の辞書や文法が充実したこと，そして，Webの浸透によって膨大な言語的資源が手に入るようになったことがある．

一方，現在インターネット上を行き来しているのは言語に依存しすぎた情報であり，それぞれの情報発信者や利用者の人格や状況とは切り離されて，言葉だけが飛び交っている状態とも言える．それに対して我々の実際の社会生活は，他の人と互いに顔を合わせ，さまざまな状況（時間や空間，背景にある暗黙知など）を共有しながら営まれている．そこでのコミュニケーションは，言語情報だけではなく，身振り手振り，視線，立ち位置や姿勢，発話の強弱や間，表情といった多くの非言語的な情報に支えられている．

近い将来，コンピュータが従来のデスクトップ型のものから姿を変えて，情報家電，ロボット，センサネットワークといった形で我々の社会生活に入り込んでくることを考えると，上述したような非言語情報も含んだ実世界インタラクションを理解し支援する情報学の確立が，今後10年の急務であると考える．そして，言語情報学の発展が言語的情報資源の研究基盤に支えられてきたように，実世界インタラクションに関する研究にも研究者が共用できる実世界インタラクションのデータ（インタラクション・コーパス[1]と呼ぶ）と，それを処理するための方法論やツールなどの研究基盤[2]が不可欠である．

ここで「インタラクション・コーパス」とは，人のインタラクションにかかわる現象，つまり，いつ，どこで，誰と，何を（行った，話した，見た），といった状況を表すマルチモーダルなデータの集合である．具体的には，同時計測されたビデオ，音声データ，身体動作や視線移動を表す3次元座標データなどの集合となる．我々の目的は，これらの観測データに対して，人のインタラクションの理解を助ける構造情報（インタラクションの要素やそれらの出現パターン）を付与したインタラクション・コーパスを構築することである．

IMADE：実世界インタラクション計測・分析環境

筆者らは，「情報爆発時代に向けた新しいIT基盤技術の研究」の一環で，実世界インタラクションの計測と分析を行うための研究基盤IMADE（Interaction Measurement, Analysis and Design Environment）を開発している．具体的には，京都大学の一室（約80平方メートル）をIMADEルームと呼び，各種センサ類を導入して，グループによる協調活動における実世界インタラクションの計測・分析・支援の研究開発を行うための環境構築を進めている．

IMADEルームにおいて我々が分析・支援のターゲットとしているのは，**図-1**に示したような，知識の伝達・創造を目的とした参加者同士の会話と身体動作が伴うインタラクションである．たとえば，3〜5人程度のミーティングやポスター発表会，また，身体動作を伴う共同作業などを題材としている．そこでは，発話の意味内容だけでなく，発話者の交代，発話に伴う身振り手振りや視線の移動，インタラクションにおける参照物の役割，会話場の発生や成長といった現象の理解に焦点を当てている．

IMADEの構築は，筆者らの研究グループを中心に，画像，音声，映像の研究グループと共同で進めている．システム構成図（**図-2**）を参照しつつ，現在までの取り組みを概観する．

動作計測　光学式モーションキャプチャシステムとして，Motion Analysis社のMAC 3Dシステム（8カメラ）とPhaseSpace社のシステム（10カメラ）を

図-1　会話を通した知識の共有と創造

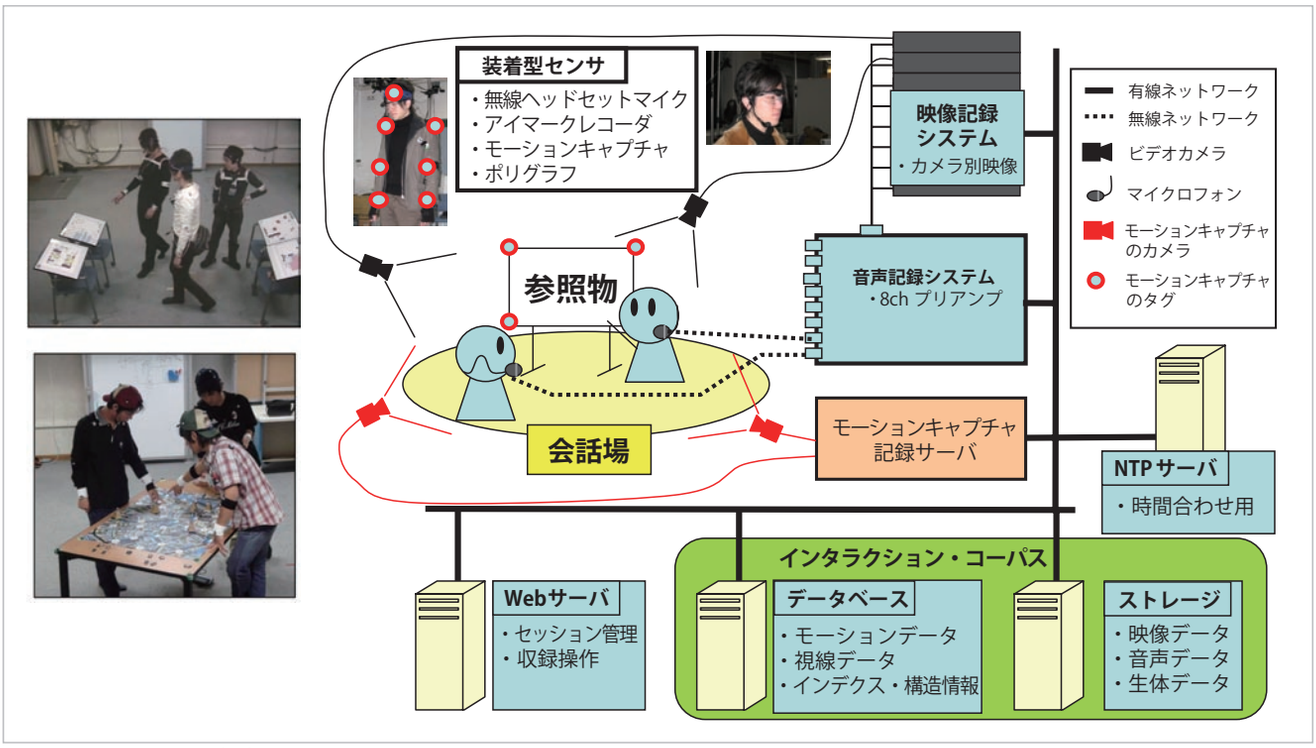

図-2　IMADE ルームの概念図

導入した．これらによって，複数人（3～5人程度）の動作や移動，対象物（パネルや操作対象機器など）の位置などを計測することができる．

視線計測　ビデオ型アイマークレコーダとして，ASL社のMobile Eye（ゴーグル型の可搬型眼球運動計測機器）を2セット導入した．これにより，個人の一人称映像と，その中での視線移動を記録することができる．

多視点映像と音声の記録　実世界インタラクションの自動解釈や分析を試みる際に，映像や音声は重要な参照データとなる．同一シーンをさまざまな視点から閲覧可能にするため，また，コンテンツとしての再利用性を高めるために，複数のカメラ（ネットワークカメラとパンチルト可能なカメラ）と，個人装着型マイク（ワイヤレスマイクと8チャネル音声記録機器）を導入した．

生体データの計測　上記まではインタラクションを映像・音声的に外部から観測するデータである．我々が対象とするのは人同士の会話や共同作業であり，参加者の心的状況を内面から計測することにも興味がある．その一歩として，ティアック社のPolymateを2セット導入した．これは，心拍，呼吸，皮膚抵抗，脳波，筋電などを計測可能である．

データ統合と閲覧　さまざまな異種センサによるデータが蓄えられるので，NTP（Network Time Protocol）による時間同期や各データの時間伸縮を吸収するための後処理が必要である．また，複数センサデータ間の空間統合，たとえば，モーションキャプチャで計測された頭部の位置・方向のデータと，アイマークレコーダによって得られる相対座標系の視線データを統合して，絶対座標系の視線データを生成する必要がある．

InTriggerの活用　上記のような大量のデータの一次ストレージとして，また，インタラクションパターンの

解析・発見の大規模計算に InTrigger プラットフォーム[3] を活用すべく，基本的なソフトウェア環境を構築中である．

iCorpusStudio：マルチモーダルデータ分析のためのソフトウェア

IMADE ではさまざまな組合せのマルチモーダルデータが取得可能であり，それらのデータをどう使うかは，ユーザ（研究者）によって異なる．たとえば筆者らの研究グループは，特定のパターンを持つ類似シーンを網羅的に集めて会話分析を行うとか，シーンごとの会話参加者と空間的構造の変化を分析し，そこから会話状況の自動認識ルールを抽出し，その知見をコミュニケーションロボットや知的環境のデザインに利用することに興味がある．社会心理学の研究者であれば，たとえば，映像・音声と生体データを計測し，それらに詳細なアノテーションを行い，分析を行うことになる．画像理解，音声理解，生体データの解釈などに興味のある研究者にとっては，評価用データの構築に利用することも可能であろう．

このように，研究動機はそれぞれ異なったとしても，マルチモーダルデータの計測と分析には共通したソフトウェア要求がある．IMADE では，そういった要求に応えるため，iCorpusStudio と呼ばれるソフトウェア環境を開発している[4]．**図 -3** は iCorpusStudio の動作画面例である．

これまでにも，会話分析のためのビデオや音声データのラベリングツールがいくつか存在していた（たとえば，Anvil[☆1] や WaveSurfer[☆2] など）．それらに対して iCorpusStudio では，映像，音声に加えて，モーションデータ，視線データ，生体データなどのマルチモーダルデータを扱うとともに，複数視点（複数チャネル）の映像，音声を同時に扱う必要がある．どういったセンサデバイスを利用するかは実験状況によって異なるため，iCorpusStudio 本体はデータの読み書き管理とラベリング記述のみを行うコンパクトなシステムにし，各種センサデータを読み込むためのソフトウェアモジュールは，プラグインとして必要に応じてインポートすることとした．また，ラベリングされたデータの分析を支援するために，従来のラベリングツールに比べて，ラベル間の演算やラベルに基づいたシーン検索の機能を強化している．

これらの特徴をより詳細に説明するために，以下に iCorpusStudio の機能を示す．

☆1 http://www.anvil-software.de/

☆2 http://www.speech.kth.se/wavesurfer/

図 -3　iCorpusStudio：インタラクション・コーパスの閲覧・分析・ラベリング環境

- 各種センサから得られるデータ（複数ビデオや音声，モーションデータ，生体データなど）を時間同期させながら同時閲覧することができる．
- ラベリング（時間幅のあるアノテーションの付与）をすることができる．ラベルは，発話の書き起こし，相槌などのパラ言語，指差しなどのジェスチャなどのインタラクション要素を記述するために利用される．
- 記述されたラベルデータは，他のツールとの互換性を保つために，単純な CSV（Comma-Separated Values）形式を採用している．
- さまざまな種類のセンサデータの取り込みに対応できるように，センサデータをインポートするためのモジュールは iCorpusStudio の本体とは切り離されて開発され，プラグインとして利用する．
- データの解釈・分析を行うための機能モジュールをプラグインとして利用する．たとえば，ラベル間の演算（重なりや境界の検出など）を行い，その結果を新たなラベルとして出力するプラグインなどを開発してきた．
- ラベルの検索や，ラベル系列の DP（Dynamic Programming）マッチングによる類似シーンの検出機能をプラグインとして提供している．
- 簡単な統計処理（ラベルの出現頻度の算出やグラフ可視化）とそのグラフ表示を行うためのプラグインを提供している．

上記の通り，iCorpusStudio は単なるラベリングツールではなく，研究者の仮説を試し，評価し，さらに他の仮説を試すというサイクルを支援するラピッドプロトタイピングの環境でもある．それぞれ異なる観点を持つさ

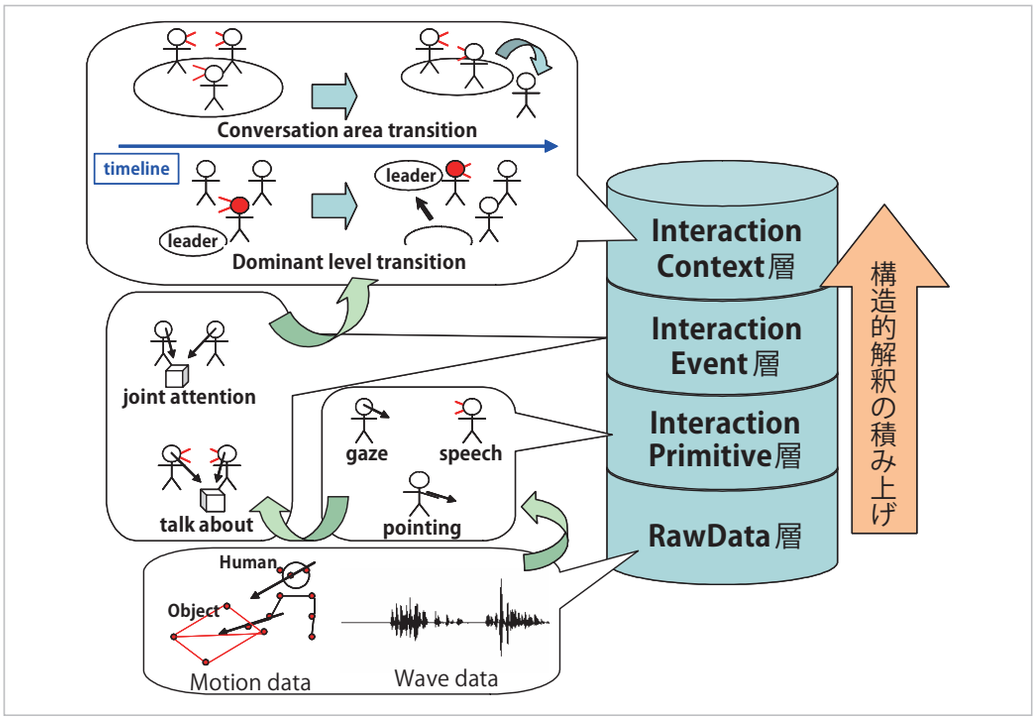

図-4　インタラクションの階層的解釈モデル

まざまな研究者が同一のデータを計測・分析・利用することを考えると，iCorpusStudio は彼らの共同作業を支援するグループウェア的役割を果たすことになる．今後，そのための機能，つまり，ラベルの登録･管理の共有支援，プロジェクト管理と作業進捗の共有促進，データ解釈ルールのマクロ化と再利用を促す機能を実現していきたい．

非言語情報による会話状況の構造理解

ここでは，IMADE を利用した研究事例紹介として，筆者らの試みを紹介する．複合的なセンサデータから会話状況を構造的に理解するために，我々は階層的な解釈モデルの構築を進めている．ここではその一部として，上半身の単純な動作データと発話有無の検出データの組合せから，ボトムアップ的に会話状況を解釈するモデルの例を**図-4** に示す．

ここでの目標は，会話の内容や作業対象に関する意味的な理解に深入りせずに，会話状況のダイナミクスを理解することである．会話参加者の発話交代や身振り手振り，そして視線の移動や共同注意といった非言語的な現象から，話題の転換点を検出したり，各参加者の会話への参与度の変化をとらえたい．解釈の元となるデータは，モーションキャプチャシステムやマイクから得られるが，そういったデータから一足飛びに抽象的な解釈を得ることは難しいので，ここでは 4 つの階層による体系的な解釈の積み上げを試みる．以下，図-4 の各層について説明する．

RawData 層　ここでは，モーションキャプチャシステムやアイマークレコーダから時々刻々得られる座標データや，音声や生体反応（脈拍，呼吸など）に関する波形データから，身体動作や発話に関する要素を取り出すための準備を行う．たとえば，モーションキャプチャデータから頭の向きや腕の向きを得るには，体の形状や動きを単純なモデルに近似し，いくつかのタグから仮想的なベクトルを形成する必要がある．また，会話における参加者間の発話交代の様子を知りたい場合には，各参加者が身に着けた接話マイクのパワー変化から発話の有無を判定する．そういった判定ルールを iCorpusStudio 上のプラグインとして実装する．

Interaction Primitive 層　ここでは，RawData 層で得られたベクトルデータや発話の有無に関するデータから，インタラクション要素の抽出を行う．たとえば，頭の向きや腕の向きを表すベクトルが空間内の対象（ポスターや会話相手のジェスチャ空間など）を貫く現象を検出して，ある対象を「見た」，「指差した」というインタラクション要素を推定する．これらのベクトル演算は iCorpusStudio の基本機能として提供されている．ただし，現実的には，人体や動的に発生するジェスチャ空間のモデリングには近似が行われるし，視線ベクトルとその対象物の衝突を検出するには検出範囲のマージンを設定する必要が生じる場合が多い．また，ベクトル衝突や発話有無の検出結果は離散データになるので，それらを時間方向にクラスタリングする必要が生じる場合が多い．このように，モデル

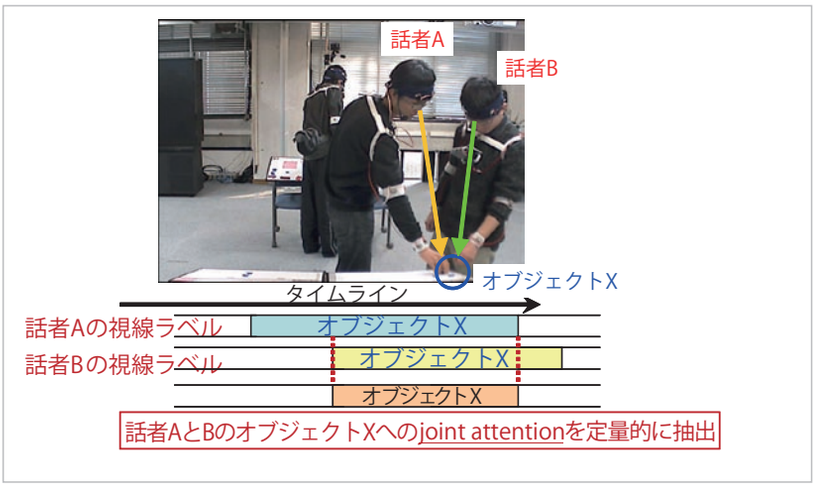

図 -5　ラベル間の重なりから Interaction Event「共同注視」を抽出している例

化には多くのパラメータが発生し，その調整が分析者ごとの仮説モデルにほかならず，そういった仮説や作業プロセスが iCorpusStudio に外在化され，研究コミュニティで再利用されることになる．

Interaction Event 層　この層は，Interaction Primitive 層で得られたインタラクション要素を組み合わせて，複数人による社会的現象の検出を試みる層である．たとえば，複数人が同一の対象物に視線を向ける（共同注視），複数人がある物を参照しながら会話をする，といった現象を検出する（例を**図 -5** に示す）．検出は，iCorpusStudio 上でラベル間の重なりを自動検出して行う．ただし，単純な AND 検索で満足な検出ができることは稀であり，実際は，重なり検出にマージンを持たせたり，他のモダリティのラベルを考慮した文脈をモデルに組み込んだりする必要があり，その作業こそが研究者の興味の対象となる．

Interaction Context 層　この層は，これまでの層の要素を組み合わせて，より「大胆な」仮説の検証を試みる層である．たとえば，複数人の発話，視線方向といったインタラクション要素の時空間的なクラスタリングから，会話場の発生・消滅のダイナミクスの定式化を試みた[5]．また，発話量の変化，視線移動，対象物への作業行為といったさまざまなインタラクション要素の組合せから，会話に参加するメンバの会話支配性（dominance）の変化を定量化することも試みている[6]．

こういった分析作業は，大まかに言うと，会話状況のデザイン，会話データの計測，センサデータの手直し，解釈ルールの試作と評価といった一連の作業を行うことになる．通常は 1 回のサイクルで満足のいく結果が得られることは稀なので，このサイクルを何度か繰り返すことが望まれる．しかし，従来の会話分析はデータの取得や解釈ルールの分析に多くの人的コストと時間をかけてきたため，これらのサイクルを繰り返すことは実際は困難であった．それに対して IMADE では，計測のためのハードウェア基盤と，分析のためのソフトウェア基盤を整備することで，これらのサイクルの高速化と再利用性の向上を目指している．実世界インタラクションの研究に興味のある読者の皆様にも，インタラクション・コーパスの構築や，そのための作業知識の体系化にご参加いただきたい．

参考文献
1) 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 (Nov. 2003).
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，Vol.22, No.5, pp.703-710 (Sep. 2007).
3) 田浦健次朗：InTrigger：オープンな情報処理・システム研究プラットフォーム，情報処理，Vol.49, No.8, pp.939-944 (Aug. 2008).
4) 來嶋宏幸，坊農真弓，角　康之，西田豊明：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，情報処理学会研究報告（ヒューマンコンピュータインタラクション），Vol.2007, No.99, pp.63-70 (Sep. 2007).
5) 高橋昌史，角　康之，伊藤禎宣，間瀬健二，小暮　潔，西田豊明：時系列イベント発見のためのグラフクラスタリング手法の提案，情報処理学会論文誌，Vol.49, No.6, pp.1942-1963 (June 2008).
6) 中田篤志，來嶋宏幸，角　康之，西田豊明：移動・動作に関するセンサデータによる多人数会話の解釈，第 22 回人工知能学会全国大会 (June 2008).

（平成 20 年 7 月 14 日受付）

角　　康之（正会員）：sumi@i.kyoto-u.ac.jp
1990 年早稲田大学理工学部電子通信学科卒業．1995 年東京大学大学院工学系研究科情報工学専攻修了．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（現在は准教授）．博士（工学）．研究の興味は，知識や体験の共有を促す知的システムの開発や，人のインタラクションの理解と支援にかかわるメディア技術．

西田　豊明（正会員）：nishida@i.kyoto-u.ac.jp
1977 年京都大学工学部卒業．1979 年同大学院修士課程修了．1993 年奈良先端科学技術大学院大学教授，1999 年東京大学大学院工学系研究科教授，2001 年東京大学大学院情報理工学系研究科教授を経て，2004 年京都大学大学院情報学研究科教授．会話情報学，原初知識モデル，社会知のデザインの研究に従事．日本学術会議連携会員．本会理事．2008 年度から人工知能学会副会長．

坊農　真弓：bono@ii.ist.i.kyoto-u.ac.jp
2005 年神戸大学大学院総合人間科学研究科コミュニケーション科学専攻博士後期課程修了．博士（学術）．2002 ～ 06 年 ATR メディア情報科学研究所学外実習生，研修研究員，研究員．2006 ～ 07 年京都大学情報学研究科研究員を経て，2007 年より日本学術振興会特別研究員（PD）．人と人との間で交わされる言葉やジェスチャを用いたインタラクションの研究に従事．

來嶋　宏幸：kijima@ii.ist.i.kyoto-u.ac.jp
2006 年京都大学工学部卒業．2008 年同大学院情報学研究科修士課程修了．現在，KDDI（株）勤務．