

ハンドアノテーション分析からのメディアデザイン —立食多人数パーティーにおける孤立者支援を例として—

山内 賢幸[†] 坊農 真弓^{††} 相原 健郎^{††} 西本 一志[‡]

Media Design Based on Hand Annotation Analysis: An Example of Supporting Isolated Participants in Multiparty Interaction

YOSHIYUKI YAMAUCHI[†] MAYUMI BONO^{††} KENRO AIHARA^{††} KAZUSHI NISHIMOTO[‡]

1. はじめに

1.1 背景

情報学にはメディアをデザインするという考え方があり、例えば、何らかの集団において孤立している人物を支援するメディアを開発する場合、いったいどのような状態の人が「孤立者」として定義でき、またセンシング技術などによってそれらの人々を検出できるのだろうか。従来、各種センサーを利用して位置情報を得たり、マイクから発話情報を得たりといったインタラクション理解の研究が数多くなされてきた。特に角ら[1][2]のように人に複数のセンサーをつけ、そこから記録された大量のデータからインタラクションのコーパスを構築する手法などがあげられる。しかしながら、ハードウェアを利用したセンサー主体のインタラクション理解の発想では、実際の人間の会話や行動、ジェスチャーをセンサーの能力内で測定することになる。その結果、センサーの能力外の自然なインタラクションのふるまいを見逃してしまう危険性がある。特に、人が孤立するといった、機械が読み取ることが難しい状況は、その場に参与している人々にとっては「孤立」していることが自明のことであっても、何によって「孤立」といったラベルが会話参与者間で振られるのが未だ不明である。

1.2 目的

そこで本稿においては、デジタルビデオカメラで撮影した立食形式パーティーの映像を使い、それにア

ノテーションを付与し、その情報から孤立した人を判別する方法を提案する。このことにより、実際の多人数インタラクションの複雑さ、センシング技術に加担することで見落としてきた現象を指摘する。本論文の試みを通し、「センシング技術の発展」→「インタラクション収録・理解」→「メディアデザイン」という従来の方法ではなく、「ハンドアノテーションからメディアデザインを導き出す」という新たな方法を提案する。

2. 提案：「孤立者検出」

本節では、人間の目では明らかな現象でも機械だとその状態を読み取る事が困難なインタラクションをどういったラベルにするのかについて、「孤立者検出」を例にあげて検討する。単純に考えて、孤立者を検出するには、単に、一定時間発話が行われていない者を検出すればよいかもしれない。しかしながら、実際には「傍参与者(side participant)」(坊農ら[3])といった、会話を傍らで聞いている者もいるため、その方法では「傍参与者」よりさらに会話から距離のある「孤立者」をうまく検出することが難しい。そこで本研究では、まず会話場が形成されているのかどうか、また誰が参与しているのかをアノテーションデータから判断する。こうして会話の中心部分を探し出すことにより、誰がいつ孤立しているのかわかるようになると考えられる。つまり逆転の発想で、複数のラベルから得られた会話集団の情報から、会話の中心的なアノテーションが振られない人物を「孤立者」として解釈するのである。

3. データ収録実験

実際に孤立者が存在しているのか、また孤立者を

[†] 北陸先端科学技術大学院大学 知識科学研究科

Japan Advanced Institute of Science and Technology

^{††} 国立情報学研究所コンテンツ科学研究系

National Institute of Informatics

[‡] 北陸先端科学技術大学院大学 知識科学教育研究センター

Japan Advanced Institute of Science and Technology

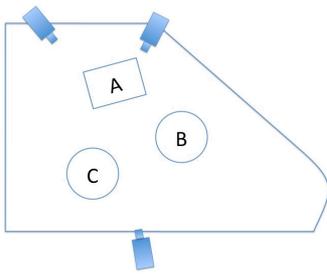


図1 会場とテーブル、カメラの設置位置

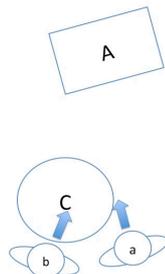


図2 身体方向の例

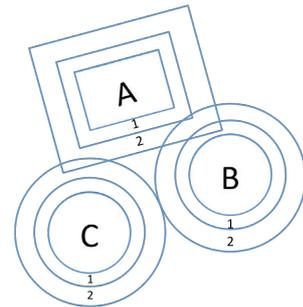


図3 身体位置のレベルの例

検出するためのハンドアノテーションをどのようなデザインにすればいいのか検討する上で、実験の動画を検証していく必要がある。そこで以下のような実験を行った。

3.1 内容

実験はデジタルビデオカメラ4台を使用して映像と音声の撮影を行った。実験参加者は大学院生10人を2組の合計20名。1組目は3つの研究科からそれぞれ4, 3, 3人。2組目は、全て同じ研究科から、修士1年生が5人、修士2年生が5人である。なお1組目と2組目の実験参加者が被らないようにしてある。また各グループには必ず全く知らない人が1人はいるように設定した。時間は2時間、それぞれ別の日にちに行った。設定した実験内容は「親睦会」とし、最初に自己紹介をしてもらい、その後はひたすらフリートークをもらった。それ以外の制約は、実験をカメラ撮影しているのだから、実験会場からなるべく外に出ないようにしてもらっただけである。その他の制約は一切かけていない。

3.2 会場

図1のようにテーブルを3つ設置し、それぞれのテーブルを撮影するためにカメラを3台設置した。そして人の頭が点で見えるように撮影するために、吹き抜けの2階部分からポールで固定したカメラ1台を使用して撮影を行った。

4. 作業環境

2時間の動画のうち、孤立者がいると思われる3分間を抜き出し、アノテーションを行った。そこで本節ではアノテーションが終わったデータから、孤立者がどのような条件でラベル付け出来るのか検討する。

4.1 分析の環境

使用した環境はMacOSX10.5.8、使用したソフトウェアはELAN3.8.0(2)である。今回はカメラ4台で撮影を行ったので、ELAN上で同時に4つの動画を同期させた。

4.2 ハンドアノテーション

レイヤーは4つ用意した。1. 発話単位, 2. 身体方向, 3. 身体位置, 4. 視線である。

[発話単位] 発話と無音区間の時間長を計る, IPU 検出の手法を用い(伝ら[4]) 発話のスタートとエンドをラベル付けする¹。

[身体方向] 図1のように配置された机を手掛かりに、両足が地面についていてかつ身体がその机の方向を向いていた時にそれぞれの机の記号(A-C)をラベリングする。また次の方向へ身体が移動し始めた時を身体移動の準備期間とし「t」というラベルを用意した²。

図2を例に身体方向のラベルの付け方を説明する。はじめに実験参加者b周辺のインタラクションを取り上げる。bは単純にCの方向を向いているので「C」と記述する。一方、aは「A」とも「C」ともとれる方向を向いている。この場合は最寄りの机の記号を当てはめる。aは現在「C」に近い、そして両肩の方向が「C」を向いているので「C」とラベルする。以上のようにラベリングすることにより現在実験参加者がどの方向を向いているのか検出可能になる。

[身体位置] 身体位置についても机を手掛かりとする。これも身体方向と同じで、両足がついて安定している状態の時の身体位置を記述する。身体移動の際のスタ

¹今回は詳細な音声データを取得していなかったので正確な発話単位を算出はしていない。単純に音声が始まった点と終わった点をスタートポイント、エンドポイントとした。

²発話単位以外の身体に関するアノテーションは共通した方法(準備区間(t)と具体的なラベルX)でアノテーションする。これはジェスチャー研究における「ジェスチャー単位」(細馬, 2008)の考え方を利用したアノテーションで、身体が安定している状態を基準として、そして次の姿勢に移動する間を準備期間とし、姿勢が安定したところからまた別のラベルに移行するという風にラベルする。身体方向, 身体位置は、両足がついているのを安定した状態とし、視線は、視線が物体から離れない状態が安定した状態である。

ートポイントは身体方向と同じで、片足が動き始めた時である。エンドポイントは両足がついて身体が安定した時である。このラベルには2段階のレベル（図3）がある。1レベルは、机にもっとも接近している状態を表している。机に寄りかかりながら喋っている状態や机の上に置いてある飲み物や食べ物を取りにきた時に接近した状態である。レベル2は、レベル1の状態以外の時である。また、図4のちょうど真ん中の辺り（机 A,B,C の間）に実験参加者がいる場合は「C2+A2+B2」のように記述する。わかりやすいように先頭には必ずその実験参加者の身体が方向づけられたテーブルを記述する³。

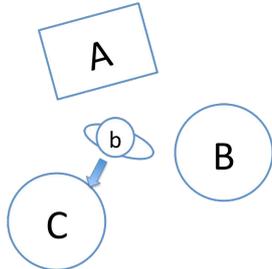


図4 身体位置のラベル付けの例

4.2.1 視線

視線は、動いていない時を基準とし、「人」と「机」に注目してラベリングする。そして視線が移動している時を準備期間とし「t」とラベルする。机と人以外に視線がいくのは至極当然のことであるが、その時は机と人ではなく、具体的に視線がどこを見ているのかを記述する。このように視線をラベリングしていくことによって、その人が一体今何に注目しているのかがわかる。身体方向と身体位置のラベルを組み合わせることによって人が今どの会話場に注目しているのかということがわかる。

5. 分析

孤立者の特徴をアノテーションデータから検出するのは非常に困難である。そのため本稿では、孤立者を直接見つけるのではなく、会話集団を探し、その集団からもれてしまった人を探すこととした。そしてそのもれている人物が孤立者と定義できるのではないかと提案した。そこで本節では、アノテーションを行ったデータから、会話集団をどのように見つけるのかを検討する。

5.1 会話集団の条件

まず始めに「発話のラベルを」基準とする。次に

その発話を行った話者の「視線のラベル」を見る。これにより、今誰が話しをしていて、誰に向かって話しているのかがわかる。そしてその話者とその視線を向けられている相手を会話集団のコアとする。次に、その会話集団の中にいる人々を判定する。基準となっている発話のスタートポイントからエンドポイントまでの間にある「視線のラベル」を参照する。これにより会話集団のコアを見ている人とそうでない人がわかる。そして次に「身体位置のラベル」を参照する。これにより、会話集団を見ている人が今どの位置から参加しているかがわかる。この時、会話集団のコアと同一のラベル上にいたならば会話集団に加わっているものと見なすことが出来る。表1で実際の会話集団を決定づける例を記述する。まずテーブル A の近くにいる a が発話を開始する。その視線の先には d がいる。次にその発話中に b, c, d は発話者の a を見ていた。そして最期にそれぞれの身体位置をみる。シーン5までで a, b, c, d が会話集団であることがわかる。

シーン	1 発話中の視線の例
1: 開始点	
2: a が発話開始	
3: a の視線	
4 会話集団のコアに対する視線	
5 身体位置の確認	

表1 会話集団決定までのアニメーション

しかしこれは会話集団形成を単純化した1つのモデル

³ 身体方向・位置の図示は、Kendon [6]を参照。

である。1 発話内ではもっと複雑に視線が交錯している。実際の表 1 の場面でのアノテーションデータを図 5 にした。

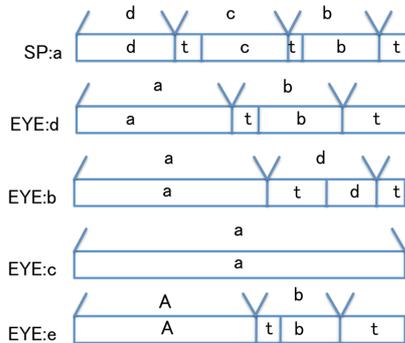


図 5 実際の発話時間と視線のアノテーションデータ

SP は発話者である a の発話の長さ、その発話時間内にどのように視線が移動したのかを表したグラフである。時間軸は左から右に進んでいる。「t」は視線の移動の準備期間である。本研究においては「t」+「視線の相手 (X)」の区間を視線の 1 単位としている。EYE は発話者以外の視線である。まず a の視線を軸に見ていく。a は 1 発話以内に d, c, b, に視線を移動させていることがわかる。a が d を見ている時、d も a に視線を合わせている。この時点で a と d が会話集団のコアである。そしてその時に同じ身体位置にいる他の視線をみても b, c が a を見ている。これらのことから a, b, c, d は同じ会話集団であると判断する。このように発話、身体位置、視線のアノテーションデータを使用して会話集団の特定を行うことは可能であると考えられる。今回のデータでは e が会話集団の中に入っていないことがわかるが、これだけで「孤立者」とであると判断するのは難しい。この前後の発話や視線の情報などを連続的に見てから判断する事が必要であると考えている。今後はより多くのデータをハンドアノテーションし、そのデータからより明確に会話集団と孤立者をわけるルールを示していきたいと考えている。特に発話時の視線の移動、ターンテイキング時の視線の移動などに重点を置いて、会話集団の条件を見つけていきたいと考えている。これにより孤立者のラベルの条件を明らかにしていけると考えている。

6. まとめ：メディアデザインへの応用

本稿において、多人数での立食形式パーティーで孤立してしまう「孤立者」を例にあげて実際の動画データを使い、アノテーション方法とその分析について議論してきた。今までの情報学はセンサーの機能の範囲

内で捉えられる現象からメディアデザインを行ってきた。しかし今回は、ハードウェア側からのアプローチではなく、発話や身体位置、身体方向などの人間の自然な振る舞いから、いかに「孤立」している状況を浮かび上がらせるのか提案してきた。これにより多人数インタラクシオンの複雑さやセンサーの機能のみでデザインを考えた場合に、拾いきれなかった現象があることを提示することが出来た。本稿を通し、人間の振る舞いを知る上でアノテーションというものがいかに有用であるかを示す事を試みた。本稿の分析はまだ十分なものではないが、今後のメディアデザインはハンドアノテーションデータ分析の結果から進めていくべきであるということを理解していただければと思う。

謝辞

本研究の一部は、国立情報学研究所平成 21 年度グランドチャレンジ「情報環境を支える日常的インタラクシオンデータ収録のためのプラットフォーム構築」(代表者：坊農真弓)の一環でなされた。データ分析のための実験に際して、実験参加者となってくれた方々、実験準備を手伝ってくれた西本研究室の皆様、実験参加者が全然集まらず困っていたところを助けてくれた北陸先端大の大森靖之さん、大木新一さん、アノテーション方法・分析に関するアドバイスをくださった千葉大学大学院の菊地浩平さんに深く感謝いたします。

参考文献

- 1) Sumi, Y., Ito, S., Matsuguchi, T., Fels, S., Iwasawa, S., Mase, K., Kogure, K., and Hagita, N. Collaborative capturing, interpreting, and sharing of experiences, *Personal and Ubiquitous Computing*, Vol.11, No.4, pp.213-328, (2007).
- 2) 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オンの記録と解釈, *情報処理学会論文誌*, Vol.44, No.11, pp.2628-2637, (2003).
- 3) 坊農 真弓, 高橋 克也: 多人数インタラクシオンの分析手法 (知の科学), オーム社, ISBN : 978-4-274-20732-7, (2009).
- 4) 伝康晴・小磯花絵・丸山岳彦・前川喜久雄・高梨克也・榎本美香・吉田奈央: 対話研究にふさわしい発話単位の認定に向けて. *人工知能学会研究会資料, SIG-SLUD-A802*, pp.27-32, (2008).
- 5) 細馬宏通: 非言語コミュニケーションのための分析単位 - シェスチャー単位 -. *人工知能学会誌*, 23, 390-396. (2008).
- 6) Adam. Kendon, : *Conducting Interaction : Patterns of Behavior in Focused Encounters*, Cambridge Univ Pr, (1991).