

発話交代と作業空間への介入に着目した チュータリングの質評価

辻本 海成^{1,a)} 角 康之^{1,b)}

概要：本稿では、1対1で行われるチュータリング中に交わされるインタラクションから、チュータリングを行う人のチュータリングへの参与度推定のための特徴的なインタラクションについて議論する。チュータリングは主観的なリフレクションや評価を行うことが多いが、発話交代やジェスチャからもチュータリングの評価ができると考える。そこで、本研究ではチュータリングを、発話交代と共同作業空間への介入に着目し参与度推定を行い、チュータリング参加者の参与度が高いと学習効果が高いという仮説のもと、チュータリングのスコアを算出するシステムを目指す。いくつかのチュータリングデータから、チュータリング中に発生する特徴的なインタラクションとして、チュータリング中は発話と同様に共同作業空間の主導権の取り合いが行われていることや発話と共同作業空間への介入には関連があることが示唆された。

Evaluation of Tutoring Session Based on Turn Taking and Intervention in Work Space

Kaisei Tsujimoto^{1,a)} Yasuyuki Sumi^{1,b)}

1. はじめに

人が行う会話は、知識伝達や意思疎通などの手段であり、最も盛んに行われているコミュニケーションの1つである。会話の中で我々は言語情報はもちろん、視線、ジェスチャ、うなずき、あいづちといった非言語情報によってさまざまな意図を伝える。我々が交わしている非言語情報には一定のパターン（アイコンタクトのタイミングや量、手の位置、発話のタイミングなど）がある。さらに、非言語情報をお互いに無意識のうちに利用し、会話の制御を行ったり、自身の気持ちを伝えたりする。知識伝達で生じるインタラクションが相手に与える印象はについて、福島らはテキストデータと音声データからプレゼンテーションが聴衆者へ与える印象の推定を行っている [1]。ただ、テキストデータや音声データだけでなく、非言語情報も相手に与える印象に大きく関わることは容易に想像できるだろう。

知識伝達の形態として、近年ではピア・チュータリング（以下、チュータリング）の試みが注目されつつある。椿本

らは学生を中心とした学習支援組織とそのための空間である「メタ学習ラボ」を構築し、メタ学習を目的としたチュータリング活動を行った。その結果、学習者にはチュータリングによるメタ認知の促進効果や、学習意欲の向上効果が得られる可能性を示唆した [2]。メタ学習ラボでは、大学2年生以上の学生チューターがチューティ（主に大学1年生）に対して1対1のチュータリングという形で学習支援を行っている。チューターは教える側の人、チューティは教えられる側の人のことを指す。椿本らは、メタ学習ラボにおけるチュータリングにおいて、学習意欲の向上効果などがあることを、言語情報ベースで示している。しかし、チュータリング中の共同注視や発話交代のタイミングなど非言語情報の観点からは議論されていない。

先述のメタ学習ラボではチュータリングを評価する客観的な手法はあまりない。現在メタ学習ラボではチュータリングの評価を、自身のリフレクションや他人からの評価によって行っている。もちろん、チュータリングのリフレクションや他人のチュータリングを評価することは、メタ学習の観点からも評価者のチュータリング技術向上につながる。しかし、それぞれの評価は評価者の主観的な評価とな

¹ 公立はこだて未来大学

^{a)} k-tsujiimoto@sumilab.org

^{b)} sumi@acm.org

る。チュータリングの個性は非常に重要で尊重されるべきものであるが、客観的な共通した評価の枠組みがあることで、チューターの育成の省力化や効率化が望めると考える。

本研究では、1対1で行われるチュータリング中に交わされるインタラクションから、チュータリングのスコアを算出するシステムを目指す。チューティへの学習効果からチュータリングスコア算出するものとし、チューティのチュータリングへの参加度が高いほど学習効果が高くなるという仮説をもとに、チュータリングへの参加度を推定する方法について議論する。本稿では、発話交代と作業空間への手の入り込みから、参加度の推定に向けたチュータリングにおける特徴的なインタラクションについて議論する。

2. 関連研究

2.1 非言語インタラクションに関する研究

会話の中で交わされるインタラクションに含まれる非言語情報を対象とした研究は、これまでもいくつかなされてきた。会話の中におけるジェスチャー研究を行っている人物として Kendon が知られている [3]。会話参加者に注目し、その人が会話に参加しているかや会話への参加度、会話の主導権の推定を試みたもの [4][5][6] や医者と患者の2者間の会話を対象とし、そこで行われるインタビューを要約するために、対話データから複数の対話コーパスを生成し、ボトムアップ的に分析する方法を提案したもの [7] がある。

複数人で行われる会話を対象とした研究もなされてきた。大塚らは、会話の構造の推論を行うための確率的な枠組みを、会話参加者の視線、頭部方向および発話の有無に基づいて提案した [8]。Nakano らは、会話優位性を自動推定するために、他人から集めた視線の量、互いに注目した視線の量、発話量、沈黙を破る発話を行った量の4つのパラメーターに着目したモデルを提案した [9]。Inoue らは、リアルな人型ロボットと人との1対1会話において、聞き手の会話への参加度へ与える要因について分析し、結果として笑い、あいづちを打つこと、うなずきが参加度と関連することを示唆した [10]。

2.2 チュータリングにおける非言語行動に着目した研究

Grafsgaard は、チュータリング内での非言語行動の分析やモデル化をマルチモーダルな観点から行った [11]。この研究では、オンラインで行われるチュータリングにおいて、チュータリングを受ける人（本研究でいうチューティのこと）の非言語行動に注目されていた。注目した非言語行動は、顔の表情や姿勢、ジェスチャーであった。

2.3 非言語情報を取得し議論する試み

会話における非言語情報を取得する試みとして Sumi らは IMADE (Interaction Measurement, Analysis, and Design

Environment) ルームを提案した [12]。IMADE ルームには非言語情報を含むインタラクション行動を計測するために、環境カメラやマイク、モーションキャプチャ、アイマークレコーダ、データ統合と閲覧用サーバが設置されていた。IMADE ルーム内においてインタラクションを計測する際は、被験者（被インタラクション計測者）は、それぞれ自分の体にモーションキャプチャ用トラッカーやアイマークレコーダ、マイクを装着していた。

Sumi らは IMADE ルームと同時に会話の構造分析を行うソフトウェア環境 iCorpusStudio を提案した [12]。この iCorpusStudio の特徴は、映像・音声の閲覧やラベリング作業を可能にするだけでなく、非言語情報間の時間構造分析を行うためのラベル間演算や身体動作、視線移動などの数値データ間計算を容易にし、非言語構造パターン解釈のための仮説検証を支援する点である。

3. 基本アイデア

本研究では、1対1で行われるチュータリング中に交わされる非言語インタラクションから、チュータリング参加者のチュータリングへの参加度推定を行う。また、得られた参加度からチュータリングのスコアを算出するシステムを目指す。最終的には、インタラクションに基づいたチュータリング評価の枠組みを提案する。

チュータリングなど人が人に何かを教える環境において、教える側（チューター）がつつい話しがちになる傾向がよく見られる。しかし、教えられる側（チューティ）が主体的に話したり作業をしたりすることが望ましく、そのことがチューティの学びにつながると考える。このことから、チュータリングの発話内容や作業内容まで分析を行わなくても、チュータリング中の非言語情報をもとに参加度推定を行うことでチュータリングの評価が可能であると考えられる。

3.1 チュータリングについて

チュータリングには様々な種類の形態が存在するが、本研究では1対1で行われるピア・チュータリング（本研究では、単にチュータリングと呼んでいる）を対象とする。本研究におけるチュータリングでは、教える側の人をチューター、教えられる側の人をチューティと呼ぶ。

3.2 着目するインタラクション

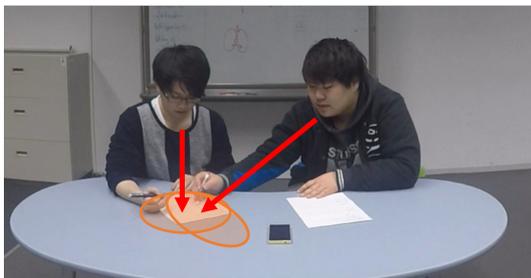
本研究で着目するインタラクションは、発話交代と作業空間への手の入り込みである。発話交代は、複数人会話の中で、会話参加者の会話への主導権を推定する際によく利用されるモダリティである。さらに、本研究では発話交代と同様に、チュータリングという場面においては、2人の共同作業空間への手の入り込みも、チュータリングの主導権を推定するために重要であると考えられる。

3.3 Shared Space

共同作業を行う複数人の間に生まれる空間に注目した研究は以前にもなされている。共同作業において Scott は、テーブル上の領域には personal, group, storage の 3 つの領域が存在することを提案した [13]。前項で述べた共同作業空間を本研究では Shared Space と定義する。Shared Space とは、人が会話をしている際に相手との間に発生する共同で所有する空間のことを指す。特にチュータリングという状況では、チューターとチューティの間の机の上の空間を指す。ただし、この空間は静的なものではなく、2 者の共同注視が行われている机上の空間を指す (図 1)。さらに、チュータリングは学習支援という性質上、一般的な会話 (立ち話や会議) に比べてインタラクション参加者の作業量が多いと考える。そこで本研究では作業空間への介入を定量的に評価し、チュータリングへの参加度を測るために Shared Space を提案する。



Shared Spaceが発生していないシーン



Shared Spaceが発生しているシーン

図 1 Shared Space の定義

また、今後 Shared Space へ手が入り込むこと、つまり作業空間へ介入することを Commit (コミット) と呼ぶ。

4. データ計測

チュータリングのデータ計測は、図 2 に示した環境で行う。記録に利用するカメラは GoPro HERO4 とする。チュータリングを行う 2 名には、チュータリングを 20 分から 30 分ほどで行うように教示する。またチュータリングの題材は自由なものとするが、ラップトップ PC を利用しないように教示する。

計測したデータは以下の観点でラベル付けを行う。

- チューターのコミット

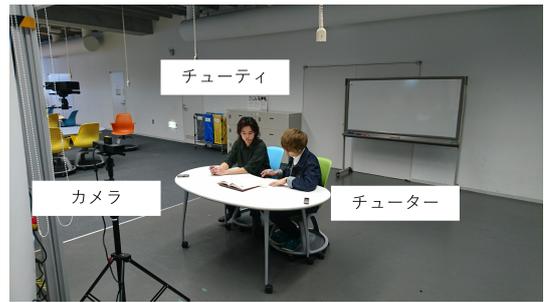


図 2 データ計測環境

- チューティのコミット
- チューターの発話
- チューティの発話

作成したラベルデータは、Sumi らが提案した iCorpusStudio [12] へ読み込む。そこで、動画と同期したラベルデータを閲覧しながら、チュータリングのインタラクションについて議論を行う。得られたデータから作成した、ラベルデータの例を図 3 に示す。

5. チュータリングでの特徴的なインタラクション

本稿では 5 件のチュータリングに対して行った分析について述べる。それぞれのチュータリングは 20 分から 25 分の長さで行われ、チュータリングに参加した人はいずれも 20 代前半の大学生であった。また、各チュータリングデータセットの概要は以下に述べる。

データセット A

チューター、チューティはともに同じ学年であり性別は男であった。

データセット B

チューター、チューティはともに男性であり、チューターが年上であった。

データセット C

チューター、チューティはともに同じ学年であり、チューターは女性でチューティは男性であった。

データセット D

チューター、チューティはともに同じ学年であり性別は男であった。

データセット E

チューター、チューティはともに男性であり、チューティが年上であった。

5.1 Shared Space の取り合い

図 3 に示したラベルのコミットに関する部分を (下の 2 行) を見ると、1 つの Shared Space に対して、チューターとチューティは交互にコミットしていることがわかる。つまり、チュータリング中の 2 人は、Shared Space の取り合い、もしくは譲り合いが行われていると考えられる。

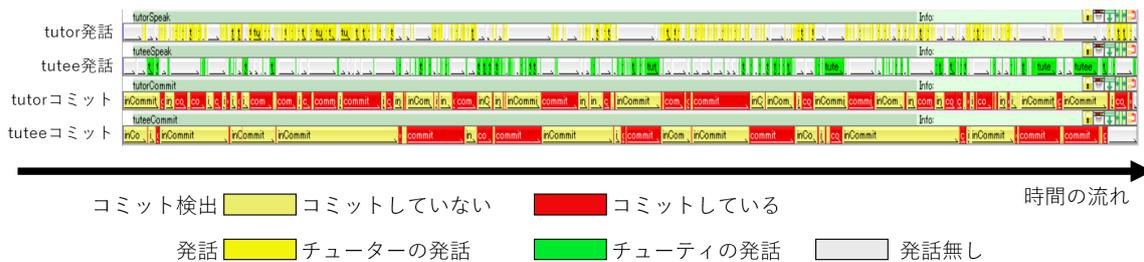


図 3 作成したラベルデータの例

このことは各データセットに共通する事項であった。各データセットのチューターとチューティが同時にコミットした割合を表 1 に示す。いずれのデータセットにおいても、全チュータリング時間での同時にコミットした割合が低くなっていることがわかる。データセット E に関しては、若干割合が高いが、それに関する考察は後述する。表 1 からわかることは、チュータリングという場面において、インタラクションを行う 2 人は、共同作業空間 (Shared Space) をお互いに取り合い、また譲り合うことである。つまり、チュータリングでのインタラクションは 2 人が交互に共同作業空間へ介入する特徴があるということである。言い換えれば、チュータリング中に同時に Shared Space へコミットすることが発生した場合、同時コミットはチュータリング中における注目すべきイベントである。

表 1 チューターとチューティが同時にコミットした割合
同時コミット量 (sec) 同時コミット割合 (%)

データセット	同時コミット量 (sec)	同時コミット割合 (%)
データセット A	2.344	0.165
データセット B	2.735	15.033
データセット C	191.416	14.644
データセット D	18.273	1.534
データセット E	300.364	21.528

5.2 発話量とコミット量の比較

各データセットの総発話量と総コミット量を示したグラフを図 4 に示す。データセットそれぞれに特徴があり、全体を通しての傾向をつかむにはデータセットの数が必要だと考える。しかし、一部に絞って考察するとわかることがいくつかある。1 つ目は、チューティの性質はチューターに左右されないことである。データセット C のチューティとデータセット E のチューティは同一人物である。一方で、データセット C のチューターはチューティと同学年の女性、データセット E のチューターはチューティの年下の男性である。異なる性質のチューター 2 つのチュータリングを受けても、データセット C とデータセット E のチューティは、チューターよりも多くコミットし、発話量はコミット量に比べて少なくなっている。したがって、現時点では、できるだけチューティの重なりが無いようにデー

タを計測していたが、同じチューティに継続的にチュータリングを受けてもらい、そのデータを計測する必要性が示唆されたといえる。

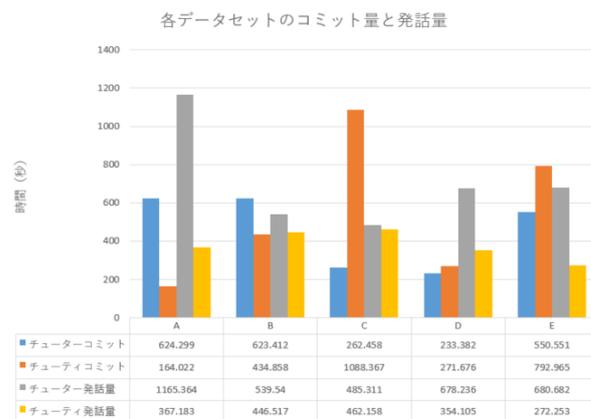


図 4 各データセットの総発話・コミット量

5.3 コミットと発話の関連性

チュータリングにおいては、発話者がただ発話するだけでなく、発話とともに作業を行うことが多いと考える。この現象は、作業を行うこと (つまりコミットをすること) で発話が誘発されると考えることもできる。チューターとチューティが行った、発話を伴うコミットの割合の各データセットを平均値、最大値、最小値を表 2 に示す。表 2 で示す割合は、発話を伴うコミット時間 / 全コミット時間を表す割合である。

表 2 発話が伴ったコミットの割合
チューター チューティ

	チューター	チューティ
平均値 (%)	81.9	51.9
最大値 (%)	97.2	77.8
最小値 (%)	64.4	24.1

表 2 から、チューターはチュータリング中、高い割合で発話が伴うコミットをしていることがわかる。このような結果が得られる原因として、チューターはチューティに対して問題を説明しながら作業することが多いことが挙げられる。また、自分の作業をチューティに注目させるために

チューターが作業をする際に発話をしているとも考えられる。実際、多くのチュータリングの場面でチューターが、「ほらここが」「次はこっち」などの発話が作業中に見られた。言い換えると、チューティに共同注視を促すためにチューターは発話を行い、結果的にチューターが発話を伴うコミットを多く行う結果になったと考えられる。一方で、チューティはチューターほど高い割合で発話が伴うコミットをしているわけではない。これは、チューターと違ってチューティは説明をしながら作業をすることが多くなく、問題を解くといった作業を多くするために得られた結果であると考えられる。ただ、チュータリング中チューティが自ら説明し、言語化することが学習効果が高くなると言われている。つまり、チュータリング中の発話を伴うチューティのコミットが多くなると、学習効果が高くなるということが予想できる。

5.4 消極的なコミット

Shared Space の取り合いの項で少し触れたが、表 1 のデータセット E では、他のチュータリングに比べて同時にコミットする割合が多い。これは、チューターが消極的なコミットを行ったため、チューターとチューティが同時にコミットをする場面が増えたと考える。消極的なコミットとは、図 5 に示したように、チューター（写真右側）が Shared Space に利き手ではない手を残しながら、利き手を Shared Space から引き、チューティへ Shared Space への介入の権利を渡すことである。このようなコミットをする理由は、チューターが説明を行い一区切りをつけたところで、チューティへ質問や疑問点を投げかけるチャンスを与えるために利き手を引くが、まだ説明は終わっておらず、引き続き説明する意思があるからであると考えられる。この消極的なコミットを作業をしているコミットと同じコミットとして考えるべきなのか、異なる性質を持つものとして考えるかを今後議論する必要があると考える。

6. 自動化に向けて

本稿では、チュータリングのデータを計測した後、手作業でラベリングを行い特徴的なインタラクションについて考察した。しかし、本研究で目指すゴールの 1 つとして、チュータリングのスコアを算出するシステムを挙げているため、現状の手作業で行うラベリング作業を自動化する必要がある。ラベリングの自動化に向けて、Shared Space へのコミットを自動検出する必要があると考える。そのためには、チュータリングを行う 2 人の頭部方向と手の位置の情報を取得する必要がある。そこで、Microsoft 社が提供する Kinect v2 センサーを利用し、Shared Space へのコミットの自動検出を試みた。本稿では、データセット D に対して Shared Space へのコミットの自動検出と発話の自動検出を試み、ラベルデータを作成した。その後、手作業で作

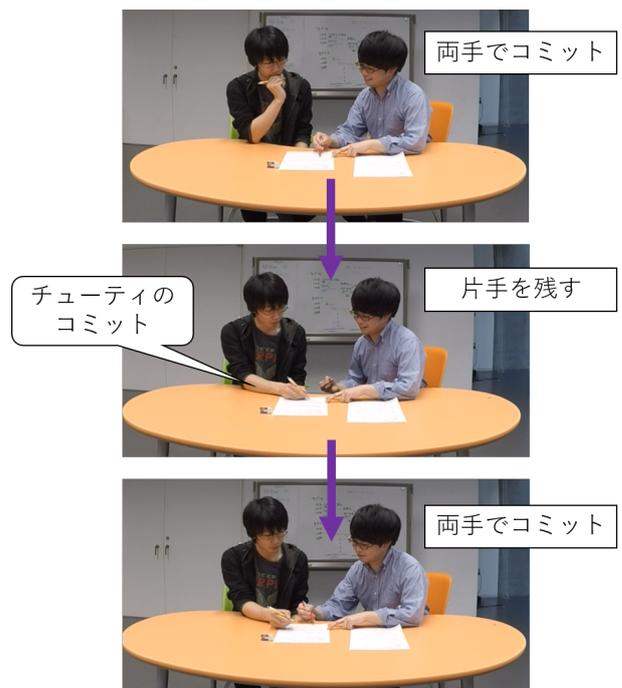


図 5 消極的なコミットを行うチューターの例

成したラベルデータと自動で作成したラベルデータの適合率と再現率を求め、その結果を表 3 に示す。

データセット D	チューター	チューティ
コミット適合率	0.343	0.359
コミット再現率	0.517	0.798
発話適合率	0.837	0.502
発話再現率	0.062	0.043

現状、Kinect v2 センサーでチュータリングを計測し、自動でラベリングを行うことは難しいと考える。その理由は 2 つあると考える。1 つ目は、Kinect v2 センサーの頭部方向検出精度がチュータリングという場面では著しく低くなるからである。Kinect v2 センサーに向けて顔を向けている場合は頭部方向が検出できるが、チュータリング中は作業をすることが多く、顔が下に向きやすい。その結果、頭部方向がほとんど検出できない状況が続いたと考える。2 つ目は、チュータリングを行うテーブルと腕の判別が難しいからである。Kinect v2 センサーは赤外線センサーを利用して深度情報を取得し、スケルトンの検出を行っているが、チュータリングを行う机と机の上で作業をする腕の判別精度は低くなる現象が見られた。その結果、腕の位置が本来の位置とは異なる場所で認識されることが多くなり、正常にコミットの検出ができなくなったと考える。今後、Kinect v2 センサーに代わるものとして Openpose[14] の利用を検討している。

7. おわりに

本稿では、1 対 1 で行われるピア・チュータリング中に

交わされるインタラクションを、発話交代と共同作業空間への介入の観点から分析した。はじめに、共同作業空間への介入を定量的に評価するために、チューターとチューティの間の机の上に2人の共同注視から発生する空間である Shared Space を定義し、Shared Space へ介入することをコミットすると定義した。また、データ計測の着目点をチュータリング中の発話とコミットとした。つぎに、発話交代と共同作業空間への介入から分析を行い、チュータリングで見られる特徴的なインタラクションについて考察した。チュータリングという環境では、多くの場面でチューターとチューティが交互にコミットすることがわかり、逆に同時にコミットしたシーンは注目すべき点であることが示唆された。また、コミットを行う際は高い割合で発話を伴うことも確認した。発話を伴うコミットはチュータリングにおいて、説明しながら作業をするシーンであることが多いことから、チューティが多く発話を伴うコミットをすることでチュータリングのスコアを高く評価できることを示唆した。さらに、コミットにも種類があり、消極的なコミットの扱いについて議論した。最後に、今後本研究で目指すチュータリングのスコアを算出するシステムに向けて、チュータリング中の発話とコミットの自動検出について議論した。

今後、チュータリングのデータセットを増やし、さらなるチュータリングの着目点や特徴的なインタラクションを発見していきたい。また、ラベルデータ作成の自動化に関する検討や、チュータリングへの参加度を推定するモデルの検討を行う必要があると考える。

参考文献

- [1] 福島悠介, 山崎俊彦, 相澤清晴. 文書と音声解析に基づくプレゼンテーション動画の印象予測. *電子情報通信学会論文誌 D*, Vol. 99, No. 8, pp. 699–708, 8 2016.
- [2] 椿本弥生, 大塚裕子, 高橋理沙, 美馬のゆり. 大学生を中心とした持続可能な学習支援組織の構築とピア・チュータリング実践. *日本教育工学会論文誌*, Vol. 36, No. 3, pp. 313–325, 2012.
- [3] A. Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004.
- [4] Yukiko I. Nakano and Ryo Ishii. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, IUI '10, pp. 139–148, New York, NY, USA, 2010. ACM.
- [5] Ryo Ishii and Yukiko I. Nakano. An empirical study of eye-gaze behaviors: Towards the estimation of conversational engagement in human-agent communication. In *Proceedings of the 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction*, EGIHMI '10, pp. 33–40, New York, NY, USA, 2010. ACM.
- [6] Misato Yatsushiro, Naoya Ikeda, Yuki Hayashi, and Yukiko I. Nakano. A dominance estimation mechanism using eye-gaze and turn-taking information. In *Proceedings of the 6th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction*, GazeIn '13, pp. 13–18, New York, NY, USA, 2013. ACM.
- [7] Kenji Mase, Yuichi Sawamoto, Yuichi Koyama, Tomio Suzuki, and Kimiko Katsuyama. Interaction pattern and motif mining method for doctor-patient multi-modal dialog analysis. In *Proceedings of the ICMI-MLMI '09 Workshop on Multimodal Sensor-Based Systems and Mobile Phones for Social Computing*, ICMI-MLMI '09, pp. 6:1–6:4, New York, NY, USA, 2009. ACM.
- [8] 大塚和弘, 竹前嘉修, 大和淳司, 村瀬洋. 複数人物の対面会話を対象としたマルコフ切替えモデルに基づく会話構造の確率的推論. *情報処理学会論文誌*, Vol. 47, No. 7, pp. 2317–2334, jul 2006.
- [9] Yukiko Nakano and Yuki Fukuhara. Estimating conversational dominance in multiparty interaction. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, pp. 77–84, New York, NY, USA, 2012. ACM.
- [10] Koji Inoue, Divesh Lala, Shizuka Nakamura, Katsuya Takamashi, and Tatsuya Kawahara. Annotation and analysis of listener's engagement based on multi-modal behaviors. In *Proceedings of the Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*, MA3HMI '16, pp. 25–32, New York, NY, USA, 2016. ACM.
- [11] Joseph F. Grafsgaard. Multimodal analysis and modeling of nonverbal behaviors during tutoring. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI '14, pp. 404–408, New York, NY, USA, 2014. ACM.
- [12] Yasuyuki Sumi, Masaharu Yano, and Toyoaki Nishida. Analysis environment of conversational structure with nonverbal multimodal data. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, pp. 44:1–44:4, New York, NY, USA, 2010. ACM.
- [13] Stacey D. Scott, M. Sheelagh T. Carpendale, and Kori M. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pp. 294–303, New York, NY, USA, 2004. ACM.
- [14] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *CoRR*, Vol. abs/1611.08050, 2016.