

非言語インタラクションに着目した チュータリング会話の計測と評価

辻本 海成^{1,a)} 角 康之^{1,b)}

概要: 本研究では、チュータリング中の非言語情報を計測し、その情報からチュータリング中の重要なシーンを推定する手法を提案する。非言語情報の中でも、チューターやチューティのチュータリングへの参加度や発話交代、頭部方向を手がかりとしてチュータリングの分析を行うことで、効率よく重要なシーンの推定ができると考える。あるチュータリングを題材に予備検討を行い、チュータリング中のチューターとチューティの頭部方向や参加度からチュータリングのシーン推定をすることの可能性を議論する。

Measurement and Evaluation of Tutoring Conversation Based on Non-verbal Interactions

KAISEI TSUJIMOTO^{1,a)} YASUYUKI SUMI^{1,b)}

Abstract: In this paper, we propose a method to capture non-verbal interaction data in tutoring and estimate important scenes in tutoring from the data. We think that it is possible to estimate important scenes efficiently by analyzing tutoring with tutor and tutee participation, turn-taking and head direction as a clue. We discuss the possibility of estimating the tutoring scene from the head direction and participation of the tutor and tutee in tutoring by a preliminary examination.

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、チュータリングと示す）の試みが注目されつつある。椿本らは学生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 [1]。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。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。

そこで本研究では、Kinect v2 センサー（以下、Kinect と示す）を用いて、チュータリング中の非言語インタラクション（非言語情報）を計測しチュータリング会話における特徴的なふるまいの議論や評価を非言語情報の観点から行う。また、会話における重要なふるまいをラベル化し、チュータリング振り返り支援になるシステムや、チュータリングを採点するシステムを目指す。チュータリング中に発生するインタラクションにおける非言語情報を分析し、

¹ 公立はこだて未来大学
Future University Hakodate

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

^{b)} sumi@acm.org

議論することによって、将来的に人とコンピュータのインタラクションへ応用する際の手がかりが得られると期待する。

2. 関連研究

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。聞き手の視線に注目し、その人が会話に参加しているかどうかの推定を試みたもの [2] 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 [3] 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 [4]。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 [5]。

会話における非言語情報を取得し議論する試みとして Sumi らは IMADE (Interaction Measurement, Analysis, and Design Environment) ルームを提案した [6]。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記録するために、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。

本研究では、モーションキャプチャシステムやアイマークレコーダなどを備えた IMADE ルームではなく Kinect を利用する。Kinect は IMADE ルームに比べて、低コストでインタラクションが計測可能である。またインタラクションを計測する際に、被験者へ一切センサ類を取り付けずに計測することになる。これにより、被験者へ変な非日常感を与えることなくインタラクションが計測できると考える。

3. 非言語インタラクションに注目したチュータリング会話の評価

本研究では、Kinect を用いてチュータリングを計測し、Sumi らが提案した iCorpusStudio [6] をもとに非言語情報の観点から議論を行う。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)が話しがちになる。しかし、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、非言語

情報から良いチュータリングのパターンが判別できると考える。非言語情報の中でも特に、頭部方向や発話交代、発話量、頭部や手の位置に注目する。注目した非言語情報を手がかりに、チュータリングにおける重要なシーンの抽出を目標とする。また抽出できた重要なシーンからチュータリング振り返りシステムやチュータリングの良い、悪い場面を推定するシステムやチュータリング終了時にチュータリングの質についてのスコアが判定されるシステムの実現を目指す。

ほかにも、2者の手や頭の動きに注目し発話交代との関連性を探る。特に手や頭がチュータリングの作業をしている空間(紙やラップトップPCが置かれている空間)に入りこんだ回数や時間、タイミングに注目する。本研究ではこの空間を Shared Space と呼ぶ。

3.1 Shared Space

本研究では Shared Space という言葉を定義する。Shared Space とはチュータリングにおいて、チューターとチューティの間の机の上の空間のことを指す(図1の紫色で表した空間)。またチュータリング中、チューターやチューティの手や頭が Shared Space 内へ入ることをコミット(Commit)すると呼ぶ。この空間は、Edward Hall が提案した Intimate Space [7] と共通するところが多いと考える。Intimate Space は、ある人から 45cm までの距離の空間であり、手で相手に触れられるぐらいの距離感をもつ空間である。チュータリング中は通常、お互いが近くに座るが、お互いが触れ合うような位置には座らない。しかし、チュータリング中において相手に向けて手を出すことや頭を前に出すことが全くないわけではない。つまり、普段は Personal Space にいる相手がチュータリング中に Intimate Space へ侵入してくる場面が存在することになる。このような行為は社会的な活動においては一般的に異例とされ、チュータリング中においても同様であると考えられる。よって、Shared Space へコミットすることは、コミットを行った人のチュータリング参加度が上がったと考えられる。ただし本研究ではチューターとチューティの2者の共同作業という点に注目するため、Edward Hall が提案した Intimate Space をそのまま利用せず、Shared Space を定義した。

4. チュータリング計測法と分析法

4.1 チュータリング設定環境

被験者は、2人とする。メタ学習ラボと同様、PC やメモ用紙、教科書等を被験者2人の前に置く。また被験者2人の距離もメタ学習ラボと同様に設定する。Kinect は被験者2人の正面に向かって設置する。設置する高さは、机に向かって座った被験者の目線程度とする。



図 1 Shared Space のイメージ
Fig. 1 Image of Shared Space

4.2 非言語情報計測システム

Kinect と Microsoft 社が提供する SDK を利用して、被センシング者のスケルトン情報と顔の方向、音声方向を計測するシステム（以下、計測システムと示す）を実装した。

計測システムで計測する非言語情報の詳細は次のとおりである。スケルトン情報は、各関節の 3 次元空間座標の集まりである。顔の方向は、Kinect の SDK に含まれる Face を利用して計測する。Face で得られる情報には、顔パーツ（目、鼻、口角）の位置や顔の状態（笑顔、眼鏡の有無、目の開閉など）などが含まれているが、本研究では Face に含まれる顔の向きを利用する。Face で得られる顔の向きはクォータニオンで表されている。ただし、後述する CSV ファイルに書き出す際はオイラー角に変換する。音声方向は、Kinect の正面を 0 度として左右の方向を -50 度から 50 度の範囲を 1 度刻みで計測する。

計測システムで計測したデータは、次項で述べる iCorpusStudio モーションデータユーティリティで利用するために CSV 形式で出力する。出力するデータは頭部や首、手を表す 3 次元空間座標と頭部方向である。頭部方向は roll, pitch, yaw で表されたオイラー角の形で出力する。また、音声方向データは同じく Kinect で取得できる音圧データとともに、発話している時間とその方向を表した CSV 形式で出力する。この情報から発話交代のタイミングが得られる。またいずれのデータも時間情報とともに出力する。

4.3 iCorpusStudio モーションデータユーティリティ

前項で得られた、CSV ファイルと Sumi らが提案した iCorpusStudio を利用し、3 次元空間内に頭部や手の位置と頭部方向のモデルを出現させる。具体的には iCorpusStudio のモーションデータユーティリティプラグインでこの機能は実装されている。しかし、Sumi らが提案したものでは Kinect に対応しておらず、また頭部方向を読み込む機能が実装されてない。そこで、Kinect に対応したプラグインを実装した。

本研究で追加実装したモーションデータユーティリティプラグインのイメージを図 2 に示す。これは Kinect の前でチュータリングを行う 2 人（図 3）の頭部や手の位置と

頭部方向を 3 次元空間上で示したものである。チュータリングを行っている 2 人はそれぞれ黄色と緑色の球体と円柱で表される。頭を表す球体から伸びている赤色の矢印は頭部方向を表す。また 2 人それぞれの手の位置も白色の球体で表されている。

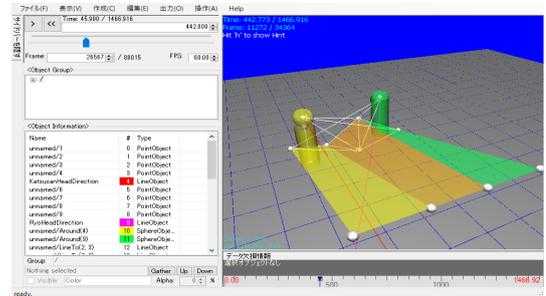


図 2 Kinect に対応したモーションデータユーティリティ
Fig. 2 Kinect-compatible Motion Data Utility



図 3 図 2 時点での現実空間の様子
Fig. 3 Context of real space at the time of Fig.2

頭部方向を表すために、プラグインでは回転を表すオイラー角から方向ベクトルを導く処理を行う。方向ベクトルを導く処理は、頭部座標からの相対座標を表した方向ベクトルを $\mathbf{v} = (x \ y \ z)^T$ とし、ラジアンで表されたオイラー角 pitch と yaw を用いて、

$$x = \sin(\text{yaw}) \times L$$

$$y = \sin(\text{pitch}) \times L$$

$$z = \cos(\text{yaw}) \times \cos(\text{pitch}) \times L$$

とする（ただし L は定数）。頭部座標と得られた \mathbf{v} を用いて頭部方向を表すベクトルを描画する。

iCorpusStudio のモーションデータプラグインでは、各オブジェクト（前述の球体や円柱など）の衝突判定やオブジェクト同士の距離などを計算する機能がある。それを利用し、共同注視や Shared Space へのコミットを自動的に検出する。それらの機能を利用して Shared Space へのコミットの検出を試みたものを図 4 に示す。

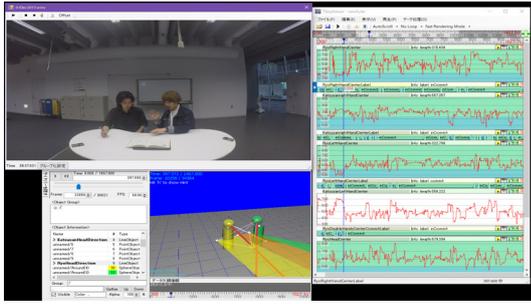


図 4 iCorpusStudio の利用例

Fig. 4 Application example of iCorpusStudio

5. チュータリング計測とシーン推定

5.1 データ収集

予備実験として、メタ学習ラボのチューターではない被験者にチュータリングを行ってもらい、提案した計測手法で計測した(図 5)。被験者はいずれも 20 代男性であった。チューターは Kinect に向かって左側、チューティは Kinect に向かって右側に座ってもらった。チューティがチュータリングへ持ち込んだ相談内容はプログラミングに関するものであり難易度は初級であった。相談内容の問題は紙に印刷されたものであった。チュータリングの長さは 30 分ほどであった。



図 5 実験環境：手前がチューター，奥がチューティ
左手前に Kinect が置かれている

Fig. 5 Experiment environment: the front person is tutor, the back person is tutee and Kinect is put on left front side

5.2 頭部方向からチュータリングのシーン推定

チューターとチューティそれぞれの頭部方向ベクトルの向きを表したグラフを図 6 に示した。図 6 の中で 2 つあるグラフのうち上のグラフはチューティの頭部方向を表しており、下のグラフはチューターの頭部方向を表している。グラフの緑色領域は、頭部方向ベクトルの要素が正の値であることを示し、紫色領域は負の値であることを示す。また、頭部方向ベクトルの x 成分(図 6 の赤線)は yaw 方向の回転を表し、頭部方向ベクトルの y 成分(図 6 の緑線)は pitch 方向の回転を表す。つまり、頭部方向ベクトルの x 成分が正の値をとるとは、Kinect に向かっている

状態で右方向に頭を向けたことになり、頭部方向ベクトルの y 成分が正の値をとるとは、Kinect に向かっている状態で上方向に顔を向けたことを表す。



図 6 頭部方向を表したグラフ

赤色の線が yaw 方向，緑色の線が pitch 方向を表す

Fig. 6 Head direction data in time series: the red line represents the yaw direction, and the green line represents the pitch direction

チューティのグラフに注目すると、おおむね x 方向 y 方向ともに負の値をとっている。このことから、このチューティはおおむね左下方向を見ていることがわかる。つまり、チューターの手元方向や Shared Space へ注視をおおむね向けていたと考えられる。また図 6 の区間 A で示した場面ではチューターが 1 人で問題を解いており、問題の紙がチューターの前に置かれていた。その時チューティは特に何もしないシーンであった。図 6 の区間 B で示した場面では、問題の紙がチューターとチューティの間に置かれていた。この時はチューターとチューティでお互い説明したり、紙に書き込んだりしたシーンであった。このように、頭部方向ベクトルから大まかなチュータリングの場面推定ができるようになる。

5.3 Shared Space へのコミット抽出

図 7 にチュータリング中に発生したコミットの時系列データを示す。各ラベル行が表すコミットは上から、チューターの右手、チューターの左手、チューターの両手(右手と左手の AND をとったもの)、チューティの右手、チューターの頭、チューティの頭である。赤色のラベルがコミットを表し、黄色のラベルがコミットを行っていない状態を表す。

今回 Shared Space へのコミットを図 8 で示した青線の長さで検出した。青線の長さはチューターとチューティそれぞれの手の座標から図 8 上の赤い円までの距離である。赤い円はチューターとチューティを結ぶ線分の垂直二等分線(図 8 の黄線)上の適当な位置に設定した。青線の長さが一定の閾値以下になるとコミットしたという様に検出した。

チューターがチューティに対して説明を行うシーン(図 9)では、チューターのコミットが多いが、一方でチューティの

会話の主導権にも注目したい。沈黙を破る発話の量や、断続的に発話する量、沈黙から発話までのタイミングに注目し、会話の主導権をどちらが握っているかを推定していきたい。

今後は、さらにチュータリングの計測を重ねて、共通してみられる特徴的なふるまいや個人差がみられるふるまいを発見したい。それらから、さらなるチュータリング中における重要なシーンの抽出や抽出精度を上げていきたい。またメタ学習ラボのチューターにも協力をお願いし、チュータリングの計測を行いたい。そこから、メタ学習ラボのチューターとそうでない人の教え方に差異があるかどうかなどの新たな知見も獲得できると予想する。

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Edward T. Hall. *The hidden dimension*. Doubleday & Company, Inc., 1966.
- [8] 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.