

視線推定と顔認識を用いたグループディスカッションにおけるユーザ間注視状況可視化手法の検討

福井 智瑛^{1,a)} 中瀬 勇太^{1,b)} 井垣 宏^{1,c)}

概要: コミュニケーション能力やプレゼンテーション能力の向上を目的としたワークショップやディベートといったグループでの活動が様々な教育機関で実施されるようになりつつある。特にコミュニケーションについては、アイコンタクトやあいづちといった聴く行動スキルや積極的傾聴といった外部から観測可能な振る舞いについての演習が実施されている。本研究では聴く行動スキルの内、アイコンタクトや応答に関する行動に着目し、360度カメラの画像のみからグループディスカッション中のユーザ間注視状況の分析を行った。結果、6人でのグループディスカッションにおいて、最高40%程度の精度で注視状況の推定が可能であることが分かった。

1. はじめに

グループでの活動を通じたコミュニケーション能力やプレゼンテーション能力の修得を目標とする教育プログラムが様々な教育機関で実施されるようになりつつある。日本学術会議の大学教育の分野別質保証委員会がまとめている情報学分野における教育課程編成上の参照基準 [6] では、情報学を学ぶ学生が獲得すべきジェネリックスキルの一つとしてコミュニケーション能力を挙げており、PBL (Project-based Learning) のようなプロジェクト学習やワークショップにおけるディベートなどによる育成を薦めている。実際にコミュニケーション能力の育成を目的とした教育プログラムも実施されている。文部科学省による enPiT プロジェクトでは、チーム作業を円滑に進めるためのスキルの一つとしてファシリテーションスキル授業が実施されており、その中で積極的傾聴といったコミュニケーションに係るスキルの演習が行われている [11]。

特に積極的傾聴に代表される聴く行動スキルは高校生、大学生向け教育でも重要視されており、尺度化なども実施されている [7], [8]。本研究では、聴く行動スキルと言われているスキルのうち、アイコンタクトや応答に関するグループディスカッション中のユーザ行動をユーザに負荷を与えずに計測する手法について検討する。我々の提案する手法では、円形あるいはそれに類するテーブルの周りに6

名までの話者が座っており、特定のテーマに従ってディスカッションすることを想定している。テーブルの中央には360度カメラが配置されており、1秒単位で話者の認証や顔の向きを画像から推定し、誰が誰の方向を向いており、誰と誰が向かい合っているかをログとして記録する。話者側には椅子に座ること以外の制約を設けずどの程度の精度で顔の向きを用いた視線推定が可能であるかを評価実験を通して分析する。

以降2節では、既存の聴く行動スキルに対する尺度分析や評価手法について述べる。3節では、我々が検討している360度カメラを用いた個人特定と視線推定を用いた注視状況可視化手法について説明する。4節では評価実験として実施した人狼と呼ばれるコミュニケーションによって目的を達成するゲームとそのゲーム中の話者の注視状況の計測結果について詳述し、5節において考察を行う。

2. 対話時の行動スキル尺度と既存の評価手法

対人コミュニケーションにおいて、話を聴くこと、話すことは重要な手段とされており、様々な分野において積極的傾聴等の聴くスキルや話すスキルに関する研究が行われている。藤原ら [8] は聴くことを内的な情報処理過程としての認知面と身体の運動に関わる行動面の両面からとらえて、高校生・大学生向けに尺度化を行っている。石井ら [5] は訓練によって身に付けやすいような行動面に焦点をあて、聴き方・話し方スキルの尺度作成とスキルによる外的適応、内的適応との関係について検討している。

対話時のコミュニケーションの記録・評価については赤外線デバイスやヘッドセットを利用したものが提案されて

¹ 大阪工業大学情報科学部情報システム学科
〒573-0171 大阪府枚方市北山1丁目79-1

a) f.tomo.262@gmail.com

b) e1b15072@oit.ac.jp

c) hiroshi.igaki@oit.ac.jp

いる。蜂巢ら [10] は複数人による双方向の対面行動を計量し、330 ミリ秒という非常に短い時間で顕在化する頭部装着型デバイスを開発している。岡田ら [4] は視線、頭部位置や加速度、指向性ヘッドセットマイクによる音声、俯瞰映像や顔映像といった多様なデータからマルチモーダル特徴を抽出し、各話者の総合的なコミュニケーション能力を推定する研究を行っている。これらの研究はグループディスカッションにおける各話者のコミュニケーション能力を評価する際に非常に有益なものである。一方で、ヘッドセットタイプのデバイスが話者に与える装着負荷の影響は無視できないという課題も存在する。

本研究では、大学等の授業で複数回実施されるグループディスカッションにおいて、学生らがコミュニケーション時の行動スキルにどの程度習熟しているかを客観的に評価するための枠組みを検討する。複数回の実施を念頭におくため、計測やフィードバックがより容易であることが期待される。本研究では、訓練によって身につけやすいことや外部から観測可能であることを念頭に、対話時の行動スキルのなかでも相手の顔を見て対話を行うといった行動の計測を目指す。

3. 360度カメラによる視線推定と顔認識を用いた注視状況可視化システムの提案

ディスカッションでは常に話し手と聞き手の関係が成り立っている。そのため、話し手が個人に対して話しているのか、全体に向けて話しているのか、そのときに相手とちゃんと向き合っているか、また聞き手も話し手と向き合っているかは対話時の行動スキルとして非常に重要である。本研究では、話し手あるいは聞き手が対話対象の顔を見て対話することができているかを評価するために、360度カメラの画像を用いて個人の識別及び視線推定を行う。なお、ユーザに装着負荷を与えないようにするため、利用する機材はグループディスカッションを行うテーブルの中央に置く360度カメラのみとする。

ディスカッションはテーブルを囲む椅子に座って行われる。各椅子はカメラの向きに対応した位置に設置される。グループディスカッションの参加者はその椅子を動かさずにディスカッションを行う。なお、グループディスカッションの参加者数は6人までとする。360度カメラで撮影した画像から人物の名前、顔の角度を取得するまでの流れを図1に示す。図が示すように、本システムはディスカッションの各参加者が誰であるかと、各参加者がいつどの向きを向いているかを取得し、記録する。以降では、個人の特定と視線推定によりディスカッション中の注視状況を可視化する手法について説明する。

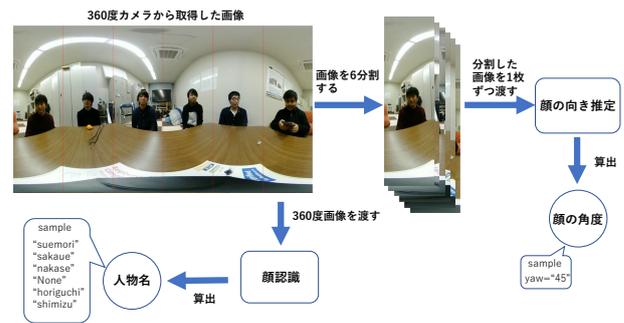


図1 画像から情報取得の流れ

3.1 個人特定

360度カメラで撮影した画像から抽出した顔画像に基づいてグループディスカッションの参加者を特定する。本機能で特定した情報からディスカッションにおいて誰がどの席に座っているのかを決定する。本研究では顔認識に Azure の FaceAPI*1 を使用する。Azure とは Microsoft 社が提供しているクラウドサービスで、様々なサービスをインターネット経由で提供している。本研究では Azure で提供されているサービスのうち、FaceAPI といわれる顔画像から様々な情報を取得できるサービスを利用する。取得できる情報としては、画像内に移る顔の位置や、予測年齢、感情や登録されている他の顔画像との類似度などがある。

FaceAPI を利用して顔画像からの人物の特定を行うためには、1人に付き5枚以上の顔画像が必要となる。そのため、グループディスカッション参加者の顔画像を事前に1人につき5枚以上用意しておき、その顔画像群を Azure に登録しておく。登録後、FaceAPI のサービス上で顔画像群毎に個人名を ID として設定しておく。

グループディスカッション中は一定時間ごとに図1が示すとおり、360度カメラ画像から6人分の画像が切り出される。切り出された画像それぞれについて、FaceAPI を利用して個人を特定する。具体的な手順を以下に示す。

- (1) 360度カメラ画像から切り出された識別対象の画像を FaceAPI に POST する。
- (2) すでに登録されている複数の顔画像群と POST された画像との類似度が算出され、類似していると思われる顔画像群の個人 ID とその類似度のリストを FaceAPI から受け取る。
- (3) 個人 ID と類似度のリストの中から類似度が 60% 以上で、かつ最も高い顔画像群の個人 ID を選ぶ。
- (4) 識別対象の画像に写っているグループディスカッション参加者の名前を選ばれた個人 ID から取得し、設定する。

これを繰り返すことで、顔写真から個人の特定を行う。(3) で類似度が 60% 以上としているのは、個人特定の際に誤認

*1 <https://azure.microsoft.com/ja-jp/services/cognitive-services/face>

する確率を下げる為である。また、最も高い類似度が60%未滿だった場合は、別の360度カメラ画像を対象に個人特定をやり直す。

図2にグループディスカッション中の参加者の位置を示す。本システムでは、360度カメラ画像から1人ずつ切り出した時点で、対象の画像に0~5のナンバーを割り当てておき、FaceAPIから返ってくる画像ごとの個人IDに従って、どの席に誰が座っているかを特定する。全座席の情報を組み合わせることで、6人分の参加者情報を0~5まで順番に{'suemori', 'sakaue', 'nakase', 'murai', 'horiguchi', 'shimizu'}のようなリストとして得ることができる。この例では、0番の席に座っている人が'suemori', 1番が'sakaue'と判定されたことを示している。以降では、ここで得られた情報にもとづいて、誰が誰の方向を向いたかを推定する。

3.2 視線推定

対話時の行動スキルとして視線を話し手や聞き手に向ける際、目だけでなく顔全体が向いていることが望ましい。そのため、本研究では顔の向きを視線としてみなすものとする。360度画像からディスカッション参加者の顔の向きを角度として算出し、その角度から誰を見ているかを推定する。顔の向き算出にはDlibを用いる。Dlibとは機械学習、画像処理、データマイニングなど幅広い処理に対応したライブラリである[1]。

360度カメラから取得した画像をディスカッション参加者ごとの画像に分けるため6分割する。分割した画像一枚ずつに対してDlibの画像処理ライブラリを用いて目や鼻といった顔のパーツの検出を行い、Mallick[3]やKwanHua[2]の手法を元にパーツ情報を利用した顔の角度の算出を行う。この手法では、画像中の顔のパーツの偏りを利用して、写真中の人物の顔がどの方向を向いているかを推定する。結果として、顔画像を入力するとpitch=[179.59392877], yaw=[-33.43570043], roll=[4.06803656]のように顔がどこを向いているかを得ることができる。ここでxyzで構成される3次元座標があったとき、yawはz軸の周りの回転角を表し、roleはx軸周り、pitchはy軸周りの回転角を表

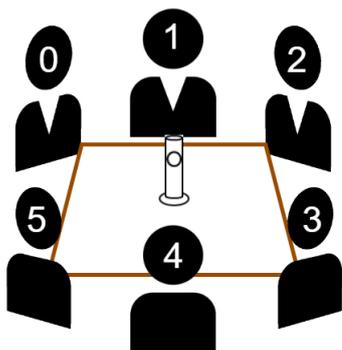


図2 グループディスカッション中の参加者の位置

表1 視点先推定基準

yaw	視点先
-15~15	正面の人物
15~47	正面右隣の人物
-15~-47	正面左隣の人物
47~90	右隣の人物
-47~-90	左隣の人物

している。本研究ではyawのみを使用し、得られた顔の角度からユーザの注視先を推定する。推定方法の基準を以下の表1に示す。なお、表の範囲を超えていた場合や顔の向き推定が行えなかった場合は誰も見ていないと判定する。例えば、うつむいている場合や空を見上げている場合、過度に右や左を向いている場合などが該当する。

個人の視点に着目した個人視点ログと向かい合っている人物に着目した向き合いログの2種類を記録する。

3.2.1 個人視点ログ

時刻、対象者名、対象者の視点先にいる人物名の3つの情報を用いていつ誰が誰を見ていたかをディスカッション参加者一人ひとりのファイルに分けてログとして記録する。個人視点ログでは視点の移り変わりの流れを一目で確認することが出来る。以下の図3はシステムが実際に記録した個人視点ログの一部である。

3.2.2 向き合いログ

3.2.1の個人視点情報からディスカッション内で向かい合っているペアが存在するかを判定する。該当するペアが存在する場合、その人物名と時間を向き合いログに記録する。以下の図4は実際に記録した向き合いログの一部である。

以降では提案システムを利用して実際にグループディスカッション中のユーザ間注視状況を集集し、評価を行う。

2019-1-22 17:14:35 suemori が nakase を見ていた
 2019-1-22 17:14:36 suemori が sakaue を見ていた
 2019-1-22 17:14:37 suemori が murai を見ていた
 2019-1-22 17:14:38 suemori が nakase を見ていた
 2019-1-22 17:14:39 suemori が sakaue を見ていた
 2019-1-22 17:14:41 suemori が nakase を見ていた

図3 個人視点ログ

2019-1-22 17:14:26 suemori と murai が向かい合っていた
 2019-1-22 17:14:27 suemori と murai が向かい合っていた
 2019-1-22 17:14:33 suemori と murai が向かい合っていた
 2019-1-22 17:14:37 suemori と murai が向かい合っていた
 2019-1-22 17:14:38 suemori と nakase が向かい合っていた

図4 向き合いログ

4. 評価実験

人狼と呼ばれる対話型ゲームを題材として6名の被験者にディスカッションを行ってもらい、そのときの注視状況を提案システムによって取得した。人狼ゲームは参加者が人狼側、村人側のいずれかになり、誰がどの役割をやっているかを伏せた状態で、毎ターン参加者が1人ずつ減っていく中で対話により誰が人狼かをあてる（村人側勝利条件）、あるいは隠し通す（人狼側勝利条件）ゲームである。人狼ゲームは人工知能化や教育現場への適用など様々な観点で研究・教育に利用されている [9]。

本実験では、被験者6名のうち1人が人狼、他5名は村人となり、4ゲームを行った。実施時間は1ゲームあたり約9～14分程度、合計約45分となった。また、最初の2ゲームは何も指示せずにゲームを実施し、後半2ゲームについては、話すときや聴くときには相手の顔を目だけでなく顔をちゃんと向けた状態でよく見るように伝えてゲームを開始した。提案システムは360度カメラによって1秒毎にディスカッション中の画像を取得し、個人の特定制と視線の推定を行い、ログとして記録した。

4.1 実験の準備

FaceAPIを利用して話者の特定を行うため、実験前に一人当たり5枚分の画像をFaceAPIに学習させた。画像の撮影には実験で使用する360度カメラではなくwebカメラを用い、画像内に顔が一つのみ含まれるものを対象とした。また、正面からの画像だけでなく、左右斜めからも撮影を行った。提案システムとは別に実際に被験者が何を見ているかを特定するため、6名の被験者のうち2名を選び、目の動きや顔の動きを別のビデオカメラで録画した。

実験の環境として、図5のように中央に机を用意し机の中心に360°カメラを設置した、図内の被験者の上部に表示された数字は席番号であり、また文字の色は座席番号と対応している。

人狼の開始と同時に360度カメラ及び提案システムと2名の被験者の実際の目の動きや顔の動きを記録するための



図5 実験の全体図

表2 顔認識結果

席番号	1回目	2回目	3回目	4回目
0	○	○	○	○
1	×	×	○	×
2	○	○	○	○
3	○	○	○	○
4	○	○	○	○
5	○	○	○	○

表3 視線推定結果 (秒)

	1回目 (840秒)	2回目 (648秒)	3回目 (593秒)	4回目 (599秒)
0	574	440	408	411
1	4	13	41	9
2	338	239	187	180
3	574	422	385	383
4	293	211	177	200
5	396	291	338	324

ビデオカメラによる録画を開始し、人狼が終了次第システムと記録用の録画も終了する。

4.2 顔認識による個人特定の結果

実験中の顔認識による個人特定結果を表2に示す。

横軸は実験回数で縦軸は席番号であり、○は正しく認識できたこと、×は誤認識したことを示す。表2からわかる通り1番に座っている人のみ正しく認識できていないことがわかる。

本システムによる視線推定結果を表3に示す。横軸は実験回数で括弧内は実験全体の秒数を示している。表中の数字はある被験者が自分以外の誰かを見ていると推定された秒数が表示されている。例えば、0番の被験者は1回目に574秒1から5番の誰かを見ていたことを示している。

1番に座っている被験者は視線の推定がほとんどできていない。また時間別の視点ログを図6から図9に示す。図は実験一回分の記録を表し横軸が時間で、6つの帯は上から順に0～5までの座席位置に座っている人の視線先の座席を示しており。色は前節で述べた通り図5によって座席番号と対応している。例えば図6の一番上の列は0番目の席に座っている人の視線情報を表し、その左端が緑色であるため実験開始後1秒後に0番席の人が3番席の人を見ていることを表す。4枚すべての図からわかることとして、実験回数に関わらず同じ参加者ごとに同じ色が多いことがわかる。特に3番席に座っている人は赤、つまり0番席に座っている人をよく見ていると判定されたことがわかる。

表 4 向きあい認識結果

	1 回目	2 回目	3 回目	4 回目
0	312	160	175	162
1	0	2	4	4
2	66	54	39	55
3	304	177	174	148
4	34	37	35	30
5	92	102	69	73

表 5 Recall/Precision 結果

実験回数	1 回目		2 回目	
	Recall	Precision	Recall	Precision
3	33.857	26.307	2.703	3.448
4	29.195	43.793	13.158	19.048
実験回数	3 回目		4 回目	
	Recall	Precision	Recall	Precision
3	36.957	30.990	20.923	17.755
4	27.757	41.477	14.620	25.000

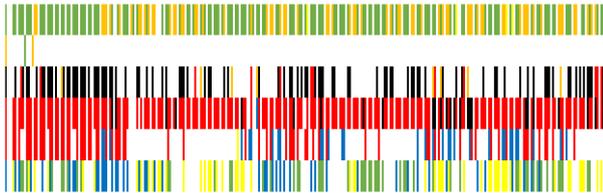


図 6 1 回目視線推定結果

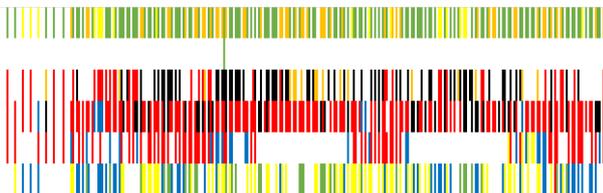


図 7 2 回目視線推定結果

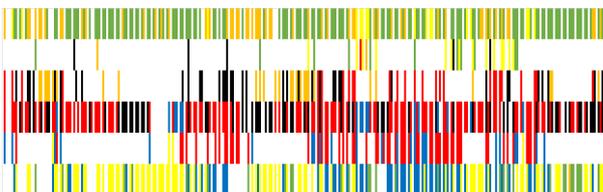


図 8 3 回目視線推定結果

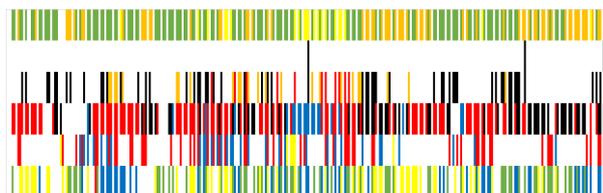


図 9 4 回目視線推定結果

向きあい認識の結果を表 4 に示す。0 番席と 3 番席の検出回数が多く、前半と後半で向きあいの検出回数に差がないことがわかる。

提案システムの精度を測定するために行った録画との比較結果を述べる。システムとは別で録画した動画は 3 番と 4 番席の人を対象としており、動画を用いて目視で顔の向きだけでなく、どこを見ていたかを確認し、システムのログと比較した。その結果得られた再現率 (Recall) と適合率 (Precision) を表 5 に示す。ここで Recall は目視による

表 6 アンケート回答項目

回答番号	回答内容
1	常に意識していた。
2	意識していたがたまに忘れる事があった。
3	たまに意識している時があった
4	あまり意識していなかった
5	忘れていた

表 7 アンケート結果

席番号	回答番号
0	2
1	1
2	1
3	3
4	3
5	3

結果とシステムログの結果が一致した秒数を目視によって誰かを見ていると判定した秒数で割ったもので、Precision は Recall と同じく目視による結果とシステムログの結果が一致した秒数をシステムログが誰かを見ていると判定した秒数で割ったものとなっている。

また後半 2 回の実験において相手の顔を目だけでなく顔をちゃんと向けた状態で見えるように伝えたが、実験後参加者にどの程度意識して行えたかアンケートを取った。5 段階の回答項目を表 6 にまた結果を参加者の座席番号とともに表 7 に示す。

5. 考察

実験の結果から顔の認識、および視点記録の精度についての考察を記述する。

5.1 個人特定の精度

個人特定の精度としては 4.2 節で述べたように 1 番席以外は正しく認識された。この 1 番席が認識されない問題の原因として一番席に座っていた人の身長が原因と考えられる。視点推定の結果でも同等の問題が発生した為、詳しくは 5.2 節で述べるがこの問題を解決することが可能になれば、Azure を利用した顔認識は一人につき 5 枚以上顔写真を登録、学習させることで誤検出を行うことなく十分な精度で利用できると思われる。

5.2 視点推定の精度

実験の結果、1番席の人物の認識回数が極端に少なかった。この問題の原因は1番席の人物の身長が高く、写真の上部に顔が写ってしまい360度画像特有の歪みから顔の検出が正しくできていないことが分かった。改善案としてカメラ配置の高さの調節や360度画像に対し水平補正を行い歪みを取り除いた画像から視点推定を行う方法が挙げられる。また、表5から、実験2回目のRecall/Precisionの結果が非常に悪くなっている。今回の実験で行った人狼ゲームでは、各参加者の役割をスマートフォンを利用して決定している。そのため、ディスカッション中にスマートフォンを見るタイミングがあり、目視による推定では、その間は誰も見ていないものとしている。しかしながら、撮影時の動画とシステムによる推定結果を確認したところ、第2回目のみスマートフォンの見方や位置の問題か、正面あるいはその隣の人物を見たものとして判定されてしまっており、精度の低下に繋がっていたことがわかった。

図6～図9の実験結果から、参加者ごとに同じ人物を見ている割合が非常に多いことが読み取れる。またその視点先はどの人物も正面の座席に座る人物を見ていると判定されている。表5の内再現率が最も高かった3回目の3番席の時間別比較データを図10に示し、同様に適合率が最も高かった1回目の4番席の時間別比較データを図11に示す。目視による結果との比較においても正面の人物をよく注視していることが確認できる一方で、目視では誰も見ていないと判定されているときにも、正面あるいはその隣の人物を見ているとシステムが判定していることが多いこともわかる。このような結果が得られる理由としては、ディスカッションにおいて正面の人物をよく見る傾向にあるというだけでなく、顔のyaw角のみで視線推定をしていることにも原因があると考えられる。今後はpitch情報を追加することで、上や下を向いている場合に誰も見ていないと判定するといった改善を行っていきたい。

また、3,4回目の試行では顔全体で注視相手を見るように指示をしていたにもかかわらず、システムの結果も目視による結果も注視状況に大きな変化はなかった。表7から読み取れるように、人狼のような考えて話すトークゲームではあらかじめ伝えていても意識し続けるのは難しく、通常行っているより積極的に相手を見ながら対話することは容易でないことが分かった。



図10 3回目3番席の時間別比較データ



図11 1回目4番席の時間別比較データ

6. おわりに

本研究ではディスカッションにおける行動スキル評価のため、360度カメラ1台でユーザーの注視状況を記録するシステムを開発した。また、コミュニケーションを通して目的を達成する対話型ゲームの人狼を実施し、目視による注視状況判定とシステムによる推定結果の比較を行った。結果として顔の向きから推定した注視状況は20%~40%程度に収まった。今後は、360度画像の歪みの補正や、yawだけでなくpitchも利用した視線推定による精度の改善を目指していきたい。

謝辞 本研究はJSPS科研費JP17K00500の助成を受けたものである。

参考文献

- [1] King, D. E.: Dlib-ml: A machine learning toolkit, *Journal of Machine Learning Research*, Vol. 10, No. Jul, pp. 1755-1758 (2009).
- [2] Lee, K.: head-pose-estimation, Learn OpenCV (online), available from (https://github.com/lincolnhard/head-pose-estimation/blob/master/video_test_shape.py) (accessed 2019-02-22).
- [3] MALLICK, S.: Head Pose Estimation using OpenCV and Dlib, Learn OpenCV (online), available from (<https://www.learnopencv.com/head-pose-estimation-using-opencv-and-dlib>) (accessed 2019-02-22).
- [4] 岡田将吾, 松儀良広, 中野有紀子, 林佑樹, 黄宏軒, 高瀬裕, 新田克己: マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定, *人工知能学会論文誌*, Vol. 31, No. 6, pp. AI30-E.1-12 (2016).
- [5] 石井琴子, 新井邦二郎: 聴き方スキル・話し方スキル尺度の作成ならびに適応との関係について, *東京成徳大学臨床心理学研究*, Vol. 17, pp. 68-77 (2017).
- [6] 大学教育の分野別質保証委員会: 大学教育の分野別質保証のための教育課程編成上の参照基準情報学分野, *日本学術会議 (オンライン)*, 入手先 (<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-h160323-2.pdf>) (参照 2019-01-14).
- [7] 藤原健志, 三宅拓人, 濱口佳和: 改訂版聴くスキル尺度の大学生への適用の検討, *筑波大学心理学研究*, No. 47, pp. 65-75 (2014).
- [8] 藤原健志, 濱口佳和: 高校生用聴くスキル尺度改訂版の作成, *心理学研究*, Vol. 84, No. 1, pp. 47-56 (2013).
- [9] 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, 松原仁ほか: 人狼知能プロジェクト, *人工知能*, Vol. 30, No. 1, pp. 65-73 (2015).
- [10] 蜂須拓, 潘雅冬, 松田壮一郎, 鈴木健嗣ほか: 複数人による双方向の対面行動を計量する頭部装着型デバイス, *電子情報通信学会論文誌 D*, Vol. 101, No. 2, pp. 320-329 (2018).
- [11] 毛利幸雄, 細合晋太郎, 鶴林尚靖, 福田晃: enPiTにおけるファシリテーションスキル授業の実践と評価について, *日本ソフトウェア科学会大会論文集*, Vol. 32, p. 11p (2015).