Vol.2021-CLE-33 No.4
2021/3/25

マルチモーダル情報に基づく グループ対話における知識伝達の分析

羽山 徹彩[1,a)]　横山 翔汰[1]

概要：複数人の対話で有用な議論にならない原因のひとつとして，参加者が話題に対する知識不足や，その知識を対話のなかで活用できないことが挙げられる．そのため，対話のなかで参加者が持つ知識とそれを活用するための客観的な評価指標があれば，有用な対話に導くための技術が実現できる．しかしながら，その知識伝達過程の現象を捉えることの難しさから，定量的な調査分析はほとんど実施されてこなかった．そこで，本研究では少人数グループの対話を対象に，マルチモーダル情報分析によって効率的な知識伝達となる特徴を明らかにすることを目的とし，実施した．そのために，まず話題への知識格差がある参加者のグループを構成し，議論を実施することで，マルチモーダル情報を含んだ対話コーパスを作成した．そして，統計的手法により対話コーパスを解析することで，効率的な知識伝達となるマルチモーダル情報の特徴量を明らかにした．その結果として，対話中に出現する表層的言語に関する特徴量を用いることで，0.9203 の高精度で，参加者を話題の知識を授ける人と受け取る人の立場に分類できることがわかった．さらに，その話題の知識を授受する参加者の異なる立場によって，対話の経過時間に従った，効率的な知識伝達に対するマルチモーダル情報の異なる特徴を明かにした．

Analyzing Knowledge Transfer in Group Dialogue based on Multimodal Information

1. はじめに

複数人での対話は会議や学習などの場で，情報共有，アイデア出し，或いは学びのための有用な方法として利用されてきた [11]．その際，参加者同士が話題に対する知識を出し合い，検討し尽すことが，有用な議論に導き易くなるとされているものの，必ずしもそのような議論にはならない [3]．その原因のひとつとして，参加者が話題に対する知識不足や，その知識を対話のなかで活用できないことが挙げられる．そのため，対話のなかで参加者が持つ知識とそれを活用するための客観的な評価指標があれば，有用な対話に導くための技術が実現可能といえるものの，そのような技術はまだ実現されていないのが現状である．

これまで，対話システムやコミュニケーション支援の基礎として，計算機的アプローチに基づいた対話分析に関する研究が取り組まれてきた．そのような対話分析では言語だけでなく，発話特徴や身体特徴などの複数のモーダル情報を対象とし，取り組まれてきた [4]．従来研究では少人数のグループを対象に，合意，共感，礼儀正しさとなる対話中のマルチモーダル情報の特徴をもとに，統計的手法により分析してきた [5], [8], [9], [14]．しかしながら，従来研究よりも根元的な対話中の知識とその伝達過程に対しては現象を捉えることの難しさから，これまで対象とされてこなかった．

そこで，本研究では少人数グループの対話を対象に，マルチモーダル情報分析によって効率的な知識伝達となる特徴を明らかにすることを目的とし，実施した．そのために，まず話題への知識格差がある参加者のグループを構成し，議論を実施することで，マルチモーダル情報を含んだ対話コーパスを作成した．そして，統計的手法により対話コーパスを解析することで，話題への理解が深まったグループのマルチモーダル情報の特徴量を明らかにした．その結果として，対話中に出現する表層的言語に関する特徴量を用いることで，0.9203 の高精度で，参加者を話題の知識を授ける人と受け取る人の立場に分類できることがわかった．

[1] 長岡技術科学大学
Nagaoka University of Technology, 1603-1, Kamitomioka Nagaoka, Niigata 940-2188, JAPAN
[a)] t-hayama@kjs.nagaokaut.ac.jp

さらに，その話題の知識を授受する参加者の異なる立場によって，対話の経過時間に従った，効率的な知識伝達に対するマルチモーダル情報の異なる特徴を確認した．本研究の成果により，グループ対話のなかで知識伝達し易い状況を判断でき，効果的な対話の誘導する技術が開発できる．

2. 先行研究

社会的信号処理に関する先行研究では，マルチモーダル情報を含んだ対話コーパスに対し，リーダーシップ，性格特性，コミュニケーション能力の推定について取り組まれてきた [1], [2], [8], [9], [14], [15]．そのなかで，社会的振舞に関連する手掛かりは主に，身体的外見，身振りと姿勢，表情と視線，発話行動，および空間と環境の 5 項目に分類され，それぞれの振舞の手掛かりを情報技術を駆使し，対話のなかで取得することで対話コーパスが作成されてきた [15]．

例えば，Sanchez-Cortes ら [14] は，グループ対話のなかで創発的なリーダーを特定するために，ELEA 対話コーパスを使用し，初対面の人たちで構成されたグループの対話に対しマルチモーダル情報分析を行った．その結果，80%精度で，リーダーを特定することができ，その特徴として発話時間，発話量，発話割込みの量が多いことがわかった．Nihei ら [8] は，グループ対話のなかでファシリテーションと関連する影響力を持つ参加者を対象に，マルチモーダル情報分析した．その結果，豊富なファシリテーション経験を持つ参加者に対し，対話中の影響力のある発言数と高い相関があることを明らかにした．Beyan ら [2] はグループ対話のなかで創発的リーダーとそのスタイルを対象にマルチモーダル情報分析を行った．その結果，66%以上の精度で識別が可能であり，発話活動と注視行動が識別に大きく影響することを明らかにした．Okada ら [9] はグループ対話のなかでコミュニケーション能力を推定するために，専門家が評価したコミュニケーション評価に基づきマルチモーダル情報を分析した．その結果，コミュニケーション能力の推定には発話韻律の特徴，言語特徴，動作特徴が識別精度に大きく影響すること明らかにした．さらに，情報伝達力や明瞭な主張に対しては言語特徴，発話ターンの特徴，発話韻律の特徴が識別精度に大きく影響することも明らかにした．この研究の情報伝達力や明確な主張の評価に関しては，本研究が対象とする知識伝達に関係するものの，外部の有識者による主観的な評価に基づいており，実際の参加者間の情報伝達力ではないといえる．そのため，本研究では知識伝達に焦点をあて，対話者の理解度を調査することで，実質的な基準を用いる点で従来研究と異なる．

3. 分析方法

3.1 実験環境

マルチモーダル情報分析に関する研究は参加者の募集，実験設備の準備，或いは分析用データの整形など対話コーパスの作成に多大な労力を要するため，公開されている対話コーパスを使用する場合が多い．しかしながら，本研究が対象とする対話中の知識伝達を扱った対話コーパスが存在しない．そのため，本研究では**図 1** に示すような，マルチモーダル情報を収集するための対話データ取得環境を構築し，対話中の知識伝達を分析可能にする対話コーパスを作成する．

図 1 マルチモーダル情報を収集するための対話データ取得環境．
Fig. 1 Environment to Obtain Group Dialogue Data for Multimodal Analysis.

構築した対話データ取得環境では，参加者ごとに，頭部加速度センサ，音声取得マイク，および顔撮影用カメラを装着し，取得時間とともに逐次，それらデータが保存される．頭部加速度センサでは 3 軸加速度センサを前頭部に装着させ，加速度の x, y, z 軸の各値を 9 フレーム/秒ごとにテキストファイルに保存する．音声取得マイクでは参加者ごとに，WAV 形式の音声データファイルに保存される．また顔撮影用カメラでは参加者ごとに，顔動画を MPEG4 形式でファイル保存される．

以上のような対話データ取得環境によって，頭部動作，発声行動，表情についてのマルチモーダル情報を分析するためのデータ取得が可能となる．

3.2 利用するマルチモーダル情報

対話分析で用いられるマルチモーダル情報にはこれまで主に，発話特徴，身体動作の特徴，表情，および距離空間が分析する特徴として用いられてきた [15]．そこで本研究では 3.1 節で述べた対話データ取得環境において，マルチモーダル情報として発話特徴，身体動作の特徴，および表情を採用する．距離空間に関しては，本実験環境において定位置の座席，および参加者同士が知人でないグループ構成とするため，今回の分析に採用しないこととする．その他のモーダル情報として，先行研究 [8], [9] に用いられている，個人の性格や表層的言語の特徴も分析に含める．

各モーダル情報の詳細については，3.2.1-3.2.3 節にそれぞれ述べる．

3.2.1 発話特徴

- 発話活動

– 発話長の合計
発話区間から算出した発話時間を発話長とし，参加者ごとの発話長の合計時間を用いる．
– 発話回数
発話区間の出現回数を用いる．
– 発話長 (1 秒以上) の合計
内容を含まないフィラーや感嘆等を除くために，1 秒以上の発話長の合計時間を用いる．
– 発話長 (1 秒以上) の回数
内容を含まないフィラーや感嘆等を除くために，1 秒以上の発話の出現回数を用いる．

- 発話韻律
– ピッチの平均差および幅
発話の抑揚を特徴量として扱うために，参加者ごとのピッチの平均差，およびピッチの最大値から最小値の差を算出した幅を用いる．
– 音響インテンシティの最大値および最小値
参加者ごとの語気の強さを特徴量として扱うために，音声から音響インテンシティの最大値と最小値を算出し，用いる．
– 音響インテンシティの幅
音響インテンシティの最大値と最小値の差を算出し，用いる．

3.2.2 身体動作の特徴

- 頭部動作量
参加者ごとに前頭部に装着した 3 軸加速度センサのデータに対し，式 (1) で算出した合計値を特徴量として扱う．

$$\sum_{i=1} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2} \quad (1)$$

ここで，x_i, y_i, z_i は状態 i の加速度センサ x, y, z 軸の加速度の値を示す．

- 表情
参加者ごとに対話中の顔動画に対し，表情認識技術を適用することで，その認識結果となる表情の出現割合を特徴量として用いる．表情認識結果の例としては，怒り，軽蔑，嫌悪，恐怖，幸福，平常，悲しみ，驚き等がある．

3.2.3 その他の特徴

- 表層的言語
参加者ごとに対話中の発話テキストのなかで出現する語の品詞数を特徴量として用いる．そのなかでも，内容語および応答を表す品詞として，「名詞，動詞，フィラー，および感動詞」の出現頻度を求め，使用する．
- 性格特性
Big-Five[12], および Locus of Control[7] の 2 種類の性格特性評価指標を用いる．Big-Five は「外向性，情緒安定性，開放性，勤勉性，協調性」の 5 因子に関する個人特性のモデルであり，学術的に検証され，多くの研究で採用されてきた [13]．Locus of Control は行動への信念を表す人格変数であり，コミュニケーションに関する個人特性としても利用されてきた．

3.3 知識伝達過程の分析方法

少人数の対話における知識伝達の過程を観測することは非常に難しい．そのため，本研究では話題に関する知識の偏りを持ったグループ構成にすることで，対話での知識伝達が発生し易い対話状況を作り出す．その手順として，まず対話の話題に対し，事前知識のある参加者とない参加者が含まれるグループを構成する．次に，その話題に対し対話を行い，対話後にその話題について参加者ごとの知識量を測定する．その結果，事前知識のある参加者はある程度，話題に関する知識を持っているものの，事前知識のない参加者は対話後に知識量が多少の差が生じる．その知識量の差が生じた対話の特徴を明らかにするために，対話中のマルチモーダル情報に対し統計的に分析する．対話後に話題の知識量が多くなるグループのマルチモーダル情報の特徴が明らかになれば，それが知識伝達に有意に働く特徴量であるといえる．

以上から，本研究ではグループ対話における効率的な知識伝達となるマルチモーダル情報の特徴量を明らかにするために，以下の点を明らかにする．

- 知識伝達が比較的促進されたグループとそうでないグループの対話中のマルチモーダル情報の特徴
- 事前知識がある参加者とそうでない参加者との対話中のマルチモーダル情報の特徴
- 対話後の参加者の知識量と対話中のマルチモーダル情報の特徴量との相関関係

4. 対話コーパスの作成と分析結果

4.1 対話コーパスの作成

図 1 の実験環境において，4 人 1 組のグループで，定められた話題に対し 20 分間の話し合いを行った．4 組のグループが 2 種類の話題に対し，それぞれの対話を実施した．その際，1 種類の話題に対して 4 人中 2 人が事前学習を行い，もう 1 種の話題に対して他の 2 人が事前学習を行うといったカウンターバランスが考慮された．参加者は大学の学部 4 年生と修士学生の 16 名で，そのグループ構成には個人的な関係がコミュニケーションに与える影響に配慮し，顔見知り程度か，初対面の人たちがグループとなるようにした．話題には参加者が事前知識を持たない，「クローン技術」および「ビットコイン」の 2 種類が採用された．それら話題に関する事前学習では予め準備された関連資料を実験前日に手渡し，予習しておくことが求められた．対話後の知識量の測定には，各話題に関する選択問題 15 問の正解数が用いられた．

マルチモーダル情報の抽出には，発話ターンおよび発話韻律の特徴量について，音声分析ソフトウェア Julius[6] を用いて音声の発話区間を分割し，Praat[*1]を用いて各特徴量抽出を行った．また発話中の表層的言語についての特徴量を抽出するために，音声認識ソフトウェア VideoIndexer[*2]を用いて，その結果に人手で修正を加えた．また表情認識には，Azure Face API[*3]が使用された．

本研究で作成された対話コーパスは，延べ 8 対話 32 名の情報が含まれる．その対話コーパスに含まれるマルチモーダル情報の特徴量を図 **2** に示す．

4.2 対話コーパスの分析結果

4.2.1 知識量とマルチモーダル情報の特徴量との相関関係

参加者の知識量を独立変数，マルチモーダル情報の特徴量の組合せを従属変数として回帰分析を行った．参加者の知識量には対話後の理解度テストの得点が用いられ，回帰分析には線形のサポートベクトル回帰モデル (SVR) が用いられた．SVR のハイパーパラメータの最適化には，グリッドサーチが適用された．それぞれのマルチーダル情報の特徴量の組合せに対する，SVR を用いた回帰分析の結果を図 **3** に示す．

回帰分析の結果において，相関係数が 0.4 以上，つまり決定係数 R^2 が 0.16 以上の特徴量の組合せは，「表層的言語」，「発話ターンと表層的言語」，「発話ターンと頭部動作」，「発話ターンと頭部動作と表層的言語」，および「発話韻律と発話ターンと表層的言語」の 5 種であった．そのため，マルチモーダル情報のなかで，表層的言語の特徴量，或いはそれと発話ターン，または頭部動作と組合せた特徴量が，参加者の知識量と中程度の相関があることがわかった．一方で，決定係数 R^2 の最大値は「表層的言語」の特徴量を用いた場合の 0.1937 であった．そのため，参加者の知識量をマルチモーダル情報の特徴量から十分に推定できる回帰モデルが構築できていないこともわかった．

4.2.2 知識伝達が比較的促進されたグループのマルチモーダル情報の特徴

対話コーパスに含まれる延べ 8 グループの対話コーパスデータに対し，理解度テストの結果をもとに，知識伝達が比較的促進されたグループ対話とそうでないグループ対話に分けて，マルチモーダル情報の特徴量の違いを分析した．その際，対話コーパスに含まれている 2 種類の話題とその理解度テストの難易度を考慮し，話題ごとにグループ分けがなされた．知識伝達が比較的促進されたグループとそうでないグループとの有意差があるマルチモーダル情報の平均特徴量について，表 **1** に示す．

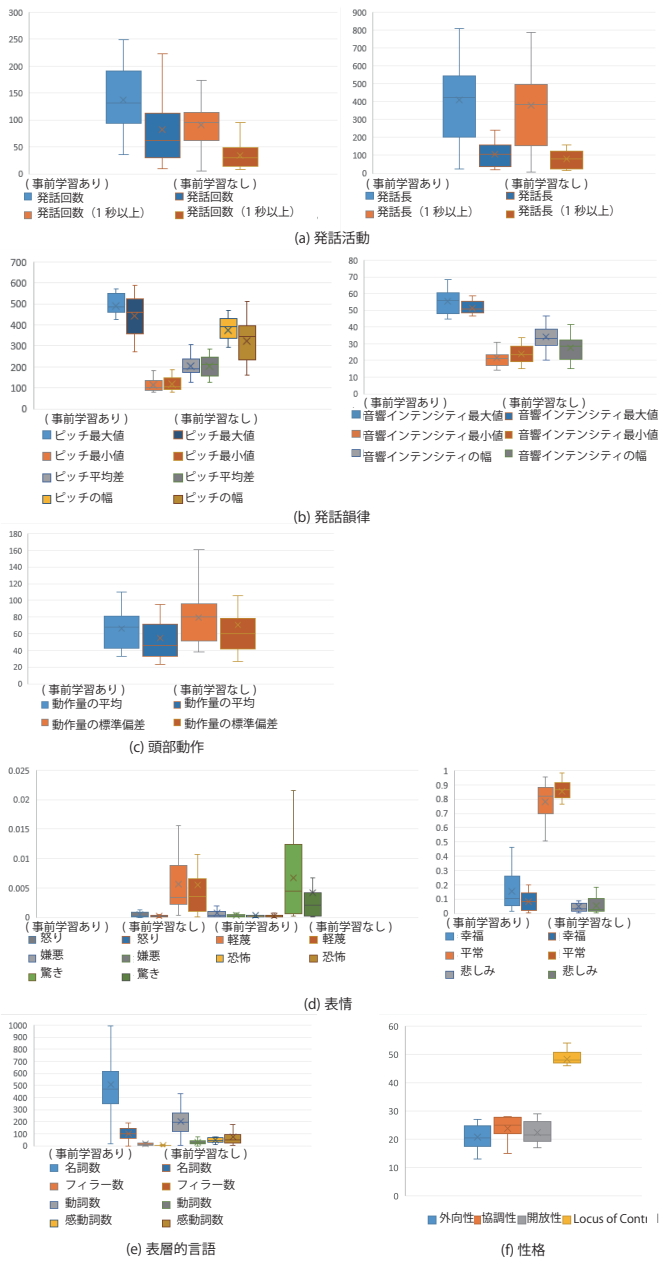

図 **2** 対話コーパスに含まれるマルチモーダル情報の特徴量
Fig. 2 Amount of Each Feature of Multimodal Information in the Dialogue Corpus

対話後の知識量の多少が異なるグループでは，事前知識がある参加者の音響インテンシティの最小値と表情（恐怖）のマルチモーダル情報の特徴量に対し，統計的に有意差が確認された．その音響インテンシティの最小値に関する特徴量では，知識量が多いグループが平均 24.11 と，知識量が少ないグループの平均 19.02 に比べ，高い結果が得られた．表情認識（恐怖）に関する特徴量では，知識量が多いグループが平均 0.0001084 と，知識量が少ないグループの平均 0.0006052 に比べ，小さい割合ながら少ない結果が得られた．その一方では，事前知識がない参加者では，知識伝達が比較的促進されたグループとそうでないグループの間で，有意なマルチモーダル情報の特徴が確認されなかっ

[*1] http://fon.hum.uva.nl/praat/
[*2] https://azure.microsoft.com/ja-jp/services/media-services/video-indexer/
[*3] https://azure.microsoft.com/ja-jp/services/cognitive-services/face/

表 1　知識伝達が比較的促進されたグループとそうでないグループとの有意差があるマルチモーダル情報の平均特徴量

Table 1　Average Amount of Multimodal Information Feature with Significant Differences between Groups with Relatively Facilitated Knowledge Transfer and Those without.

	事前知識がある参加者			事前知識がない参加者		
	高い知識量	低い知識量	$F(1,14)$ 値	高い知識量	低い知識量	$F(1,14)$ 値
音響インテンシティの最小値	24.11	19.02	$3.77(p < .10)$	24.15	23.94	$0.004(p > .10)$
表情（恐怖）	0.0001084	0.0006052	$3.83(p < .10)$	0.0001603	0.0001497	$0.010(p > .10)$

※下線は，分散分析，有意水準 10%で，高い知識量のグループとそうでないグループとの有意差ありの項目を表す．

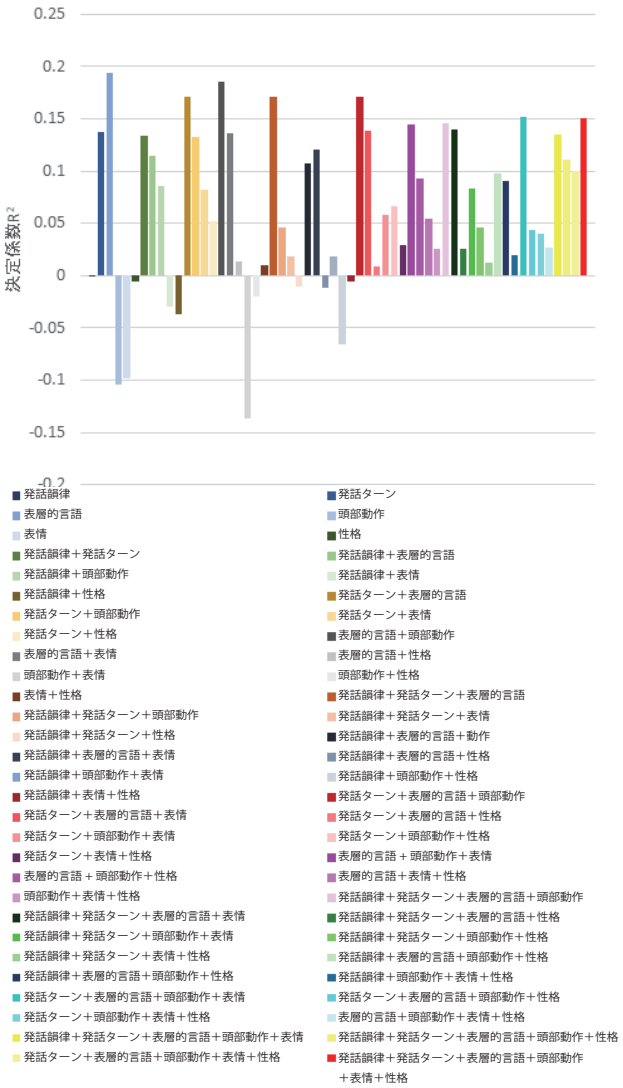

図 3　知識量を推定するための SVR を用いた回帰分析の結果（縦軸が決定係数で，横軸がマルチモーダル情報の特徴量の組合せを表す）．

Fig. 3　Results of Regression Analysis Using SVR to Estimate the Amount of Knowledge (the vertical axis and the horizontal axis represent the coefficient of determination and the combination of features of multimodal information, respectively).

た．以上から，知識伝達が比較的促進されたグループとそうでないグループのマルチモーダル情報の特徴としては，事前知識のある参加者が音響インテンシティの最小値が高く，恐怖の表情が現れ難いことがわかった．

4.2.3　知識伝達が比較的促進されるグループの特徴の時間的変化

知識伝達が比較的促進されたグループとそうでないグループ間において，経過時間に応じた，マルチモーダル情報の特徴とその変化について調査した．そのために，知識伝達が比較的促進されたグループとそうでないグループのマルチモーダル情報の特徴量の違いを経過時間 5 分間ごとに分けて，統計的に分析した．その分析結果に対し，有意差があったマルチモーダル情報の特徴量について，**表 2** に示す．

対話後の知識量の多少が異なるグループ間では，開始から 5 分間において，ピッチの平均差と幅，および表情認識（幸福）に関する特徴量に有意差が確認された．事前知識のある参加者において，ピッチの幅の特徴量では，高い知識量のグループが平均 312.85 と，低い知識量のグループの平均 400.62 に比べて，小さい結果であった．一方で，事前知識のない参加者において，ピッチの平均差に関する特徴量では，高い知識量のグループが平均 186.85 と，低い知識量のグループの平均 259.78 に比べて，小さい結果であった．また表情認識（幸福）に関する特徴量では，高い知識量のグループが平均 0.1220 と，低い知識量のグループの平均 0.0355 に比べて，大きい結果であった．以上から，知識伝達量が大きいグループでは開始から 5 分間において，知識伝達量が小さいグループに比べ，事前知識のある参加者とない参加者ともにピッチの抑揚が小さく，事前知識のない参加者の表情が明るいことがわかった．

次に，開始 5 分後からの 5 分間において，音響インテンシティの最小値，発話長の合計，および表情認識（恐怖）に関する特徴量に有意差が確認された．事前知識のある参加者において，音響インテンシティの最小値に関する特徴量では，高い知識量のグループが平均 24.65 と，低い知識量のグループの平均 18.85 に比べて，大きい結果であった．また表情認識（恐怖）に関する特徴量でも，高い知識量のグループが平均 0.0001383 と，低い知識量のグループの平均 0.00000985 に比べて，大きい結果であった．一方で，事前知識のない参加者において，発話長の合計に関する特徴

表 2 知識伝達が比較的促進されたグループとそうでないグループとの差がある 5 分ごとのマルチモーダル情報の平均特徴量

Table 2 Average Amount of Multimodal Information Feature Every 5 Minutes with Significant Differences between Groups with Relatively Facilitated Knowledge Transfer and Those without.

経過時間	特徴量	事前知識がある参加者			事前知識がない参加者		
(分)		高い知識量	低い知識量	$F(1,14)$ 値	高い知識量	低い知識量	$F(1,14)$ 値
0-5	ピッチの平均差	194.32	217.61	$0.73(p > .10)$	186.85	259.78	$6.32(p < .05)$
	ピッチの幅	312.85	400.62	$5.31(p < .05)$	262.85	327.88	$0.97(p > .10)$
	表情（幸福）	0.1961	0.1275	$1.03(p > .10)$	0.1220	0.0355	$3.97(p < .10)$
5-10	音響インテンシティの最小値	24.65	18.85	$3.54(p < .10)$	23.78	25.93	$0.91(p > .10)$
	発話長の合計	107.73	124.43	6.67×10^{-6} $(p > .10)$	26.78	21.88	$4.19(p < .05)$
	発話長（1 秒以上）の合計	102.68	113.46	$0.002(p > .10)$	21.93	13.79	$6.48(p < .05)$
	表情（恐怖）	0.0001383	0.0000985	$5.91(p < .05)$	0.0000962	0.0000960	$0.01(p > .10)$
10-15	音響インテンシティの最小値	23.95	18.70	$3.39(p < .10)$	26.84	23.56	$0.41(p > .10)$
	表層的言語（動詞数）	74.63	49.63	$0.003(p > .10)$	6.50	3.75	$5.21(p < .05)$
15-20	音響インテンシティの最小値	25.56	19.66	$6.39(p < .05)$	22.86	23.30	$0.02(p > .10)$

※下線および二重下線はそれぞれ，分散分析，有意水準 10%および 5%で高い知識量のグループとそうでないグループとの有意差ありの項目を表す．

量では，高い知識量のグループが平均 26.78 と，低い知識量のグループの平均 21.88 に比べて，大きい結果であった．以上から，知識伝達量が大きいグループでは開始 5 分から 5 分間において，知識伝達量が小さいグループに比べ，事前知識のある参加者の音響インテンシティが一定以上に強く，緊張した表情が少し程度であるが表れやすく，事前知識のない参加者がより長く発話することがわかった．

次に，開始 10 分後からの 5 分間において，音響インテンシティの最小値，および，表層的言語（動詞数）に関する特徴量に有意差が確認された．事前知識のある参加者において，音響インテンシティの最小値に関する特徴量では，高い知識量のグループが平均 23.95 と，低い知識量のグループの平均 18.70 に比べて，大きい結果であった．その一方で，事前知識のない参加者において，表層的言語（動詞数）に関する特徴量では，高い知識量のグループが平均 6.5 と，低い知識量のグループの平均 3.75 に比べて，大きい結果であった．以上から，知識伝達が大きいグループでは開始 10 分から 5 分間において，知識伝達量が小さいグループに比べ，事前知識のある参加者の音響インテンシティが一定以上に強く，事前知識のない参加者が内容語を多く含む発話が多いことがわかった．

最後に，開始 15 分後からの 5 分間において，音響インテンシティの最小値に関する特徴量に有意差が確認された．事前知識のある参加者において，音響インテンシティの最小値に関する特徴量では，高い知識量のグループが平均 25.56 と，低い知識量のグループの平均 19.66 に比べて，大きい結果であった．一方で，事前知識のない参加者において，有意な差がある特徴量が確認されなかった．以上から，知識伝達量が大きいグループでは開始 15 分から 5 分間において，知識伝達量が小さいグループに比べ，事前知識のある参加者の音響インテンシティが一定以上に強いことがわかった．

4.2.4 事前知識がある参加者のマルチモーダル情報の特徴

事前知識がある参加者とそうでない参加者に対して，マルチモーダル情報の特徴量の違いについて調査した．そのために，サポートベクターマシーン (SVM) により，マルチモーダル情報の特徴量の組合せをもとに，事前知識がある参加者とそうでない参加者との識別精度を調査した．その識別精度の結果を**図 4** に示す．

SVM を用いた識別精度は図 4 が示すように，最も高い値で 0.9203 であり，その特徴量の組合せは，「表層的言語」，「発話ターンと表層的言語」，「表層的言語と動作」，「発話ターンと頭部動作と表層的言語」，「頭部動作と表層的言語と性格特性」，「発話韻律と発話ターンと頭部動作と表層的言語」，および「発話韻律と表層的言語と頭部動作と性格特性」の 7 種類であった．表層的言語を含まない特徴量の組合せでも「発話韻律と発話ターン」，「発話ターンと表情」，「発話韻律と発話ターンと性格特性」，「発話ターンと頭部動作と表情」，および「発話ターンと頭部動作と性格特性」で 0.765 と，発話韻律と発話ターンを用いれば，ある程度の高い識別精度が得られた．以上から，事前知識がある参加者の識別は表層的言語に関する特徴量を用いれば高精度に識別でき，また発話韻律や発話ターンの特徴量を用いてもある程度高精度に識別できることがわかった．

4.3 考察

本研究の目的は，グループの対話を対象にマルチモーダル情報分析によって効率的な知識伝達となる特徴を明らかにすることである．そのために対話コーパスを分析し，対話前および対話後の知識量はともに，対話中に発話した表

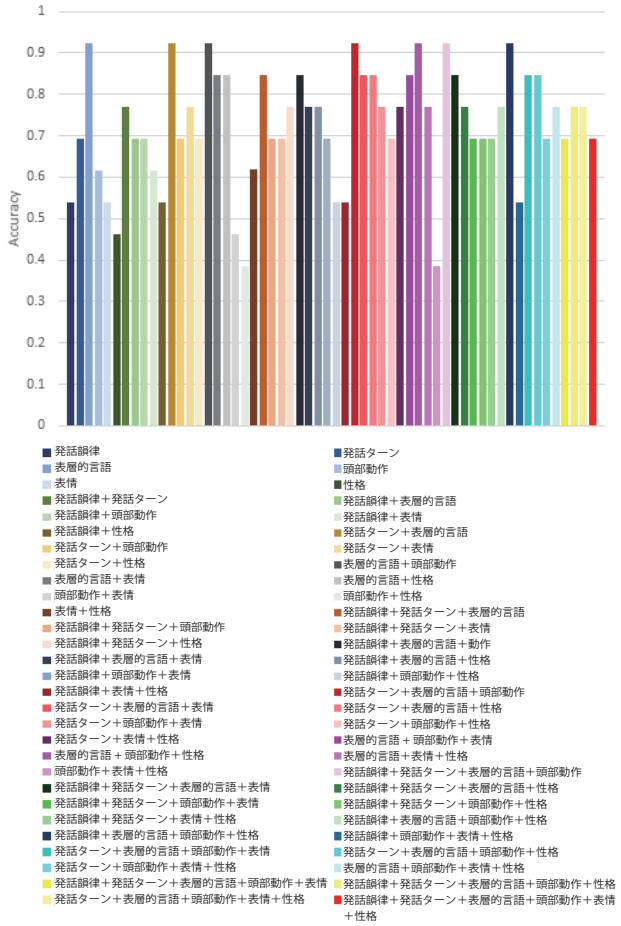

図 4 SVM を用いたマルチモーダル情報の特徴量に基づく事前知識がある参加者の識別精度.

Fig. 4 Correlation between the Amount of Knowledge of Each Participant and the Amount of Features of Multimodal Information.

層的言語に関する特徴量と相関があり，表層的言語に関する特徴量を用いることで対話前の知識量の多少が高い精度で識別可能であることを明らかにした．また効率的な知識伝達を特徴付けるマルチモーダル情報としては，話題の知識を授ける側（話題に対する知識が多い人）と受け取る側（話題に対する知識が少ない人）の立場で異なることも明らかにした．その結果として，知識を授ける人は音響インテンシティがある程度以上の大きさで，極小さな割合で緊張した表情が少ないことが明らかになった．さらに，時間経過を考えた場合には，知識を授ける人は対話開始時に，抑揚を小さく発話し，対話中ごろにやや緊張感を持って発話するとともに，それ以降音響インテンシティをある程度以上強く発話していることが明らかになった．その一方で，知識を受ける人は対話開始時に，穏やかな表情で参加し，抑揚の少なく発話し，それ以降内容を含んだ発話量を増やしていくことが明らかとなった．

以上のように，対話中のマルチモーダル情報を分析することで，話題の知識を授ける人と受け取る人の立場を識別できき，効率的な知識伝達されている対話に含まれたマルチモーダル情報の特徴が明らかになった．

5. まとめ

グループ対話のなかで参加者が持つ知識とそれを活用するための客観的な評価指標は有益な議論を促すために有用であるものの，その方法についてはマルチモーダル対話分析の研究分野ではまだ明らかにされてこなかった．そこで，本研究では少人数グループの対話を対象に，マルチモーダル情報分析によって効率的な知識伝達となる特徴を明らかにした．そのために，まず話題への知識格差がある参加者のグループを構成し，マルチモーダル情報を含んだ対話コーパスを作成した．そして，統計的手法により対話コーパスを解析することで，効率的な知識伝達となるマルチモーダル情報の特徴量を明らかにした．その結果として，対話中の表層的言語に関する特徴量を用いることで，0.9203 の高精度で，参加者を話題の知識を授ける人と受け取る人の立場を分けることが可能であった．また，各参加者の立場で，対話の経過時間に従った，効率的な知識伝達に有効なマルチモーダル情報の特徴が確認された．

今後の進展として，本研究の知見を実践的に適用することで，グループ対話の知識伝達に関する評価方法とそれを用いた支援方法の開発が挙げられる．本研究では対話での知識伝達を捉えやすくするために，参加者の知識格差が伴った環境での対話コーパスを構築し，分析した．そのため，本研究で確認してきた知識伝達の効率化に有効なマルチモーダル情報の特徴が有効であるものの，その特徴量については実践的な対話での検討を要する．その結果をもとに，対話のなかで知識伝達の効率性を評価する方法を開発する．また，それを使用し対話支援するための情報提示インタフェースについても開発していきたい．

謝辞 本研究は科研費（19K12264）および科学技術融合振興財団令和 2 年度研究補助金の助成を受けたものである．

参考文献

[1] Aran, O. and Gatica-Perez, D., “One of a kind: inferring personality impressions in meetings.” In Proceedings of the 15th ACM on International conference on multimodal interaction (ICMI ’13), pp.11–18 (2013).
[2] Beyan, C., Capozzi, F., Becchio, C. and Murino, V., “Prediction of the Leadership Style of an Emergent Leader Using Audio and Visual Nonverbal Features,” IEEE Transactions on Multimedia, vol. 20, no. 2, pp.441-456 (2018).
[3] Hayama, T., Xu, L. and Kunifuji S., “Developing an argumentation support system for face-to-face collaborative learning,” In Proceedings of 2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS), pp. 89-94 (2013).
[4] 速水 悟, 竹澤 寿幸, “マルチモーダル情報統合システムの研究動向,” 人工知能学会誌, Vol. 13(2), pp.206-211 (1998).

[5] 林 佑樹, 二瓶 芙巳雄, 中野 有紀子, 黄 宏軒, 岡田 将吾, “グループディスカッションコーパスの構築および性格特性との関連性の分析,” 情報処理学会論文誌, Vol.56 No.4, pp.1217-1227 (2015).
[6] Lee A. and Kawahara T., “Julius v4.5,” https://doi.org/10.5281/zenodo.2530395 (2019).
[7] Leecourt, H. M., “Locus of Control: Current Trends in Theory and Research.” Hillsdale, New Jersey: Lawrence Erlbaum (1976).
[8] Nihei, F., Nakano, Y. I., Hayashi, Y., Hung, H. H. and Okada, S., “Predicting Influential Statements in Group Discussions using Speech and Head Motion Information.” In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI ’14), pp.136–143 (2014).
[9] Okada, S., Ohtake, Y., Nakano, Y. I., Hayashi, Y., Huang, H. H., Takase, Y. and Nitta. K., “Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets,” In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI ’16), pp.169–176 (2016).
[10] Ponce-Lopez, V., Escalera, S. and Baro, X., “Multi-modal social signal analysis for predicting agreement in conversation settings,” In Proceedings of the 15th ACM on International conference on multimodal interaction (ICMI ’13), pp.495–502 (2013).
[11] Rabow, J., Charness, M. A., Kipperman, J., and Radcliffe-Vasile, S., “William Fawcett Hill’s Learning Through Discussion,” Sage Publications (1994).
[12] Rothmann, S. and Coetzer, E., “The big five personality dimensions and job performance. SA Journal of Industrial Psychology,” Vol.29(1), doi:https://doi.org/10.4102/sajip.v29i1.88 (2003).
[13] Rotter, J. B., “Generalized expectancies for internal vs. external control of reinforcement,” Psychological Monographs, No.80, pp.1–28 (1966).
[14] Sanchez-Cortes, D., Aran, O., Schmid Mast, M. and Gatica-Perez, D., “Identifying emergent leadership in small groups using nonverbal communicative cues,” In Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI ’10), Article 39, pp.1–4 (2010).
[15] Vinciarelli, A., Salamin H. and Pantic, M., “Social Signal Processing: Understanding social interactions through nonverbal behavior analysis,” In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp.42-49 (2009).