

話者継続・交替時における対話行為と視線行動に基づく 共感スキルの推定

石井 亮^{1,a)} 大塚 和弘^{2,†1} 熊野 史朗² 東中 竜一郎¹ 青野 裕司¹

受付日 2020年4月13日, 採録日 2020年10月6日

概要: 多人数対話における話者継続・交替時の現話者の発話の意図を示す対話行為と視線行動の組合せ情報が、個人の共感スキルレベルによって異なることを初めて明らかにした。また、個人の共感スキルレベルを推定するために、話者継続・交替時の対話行為、および視線行動を利用することが有用であることを初めて明らかにした。具体的に、4人対話コーパスデータを分析した結果、話者継続・交替時に、情報提供、自己開示、共感、そのほかの4つの対話行為カテゴリを含む発話にともなう視線行動は、個人の共感スキルレベルに応じて異なることが示唆された。これらの知見から、視線行動と対話行為情報を用いた個人の共感スキルレベルの推定器を構築し、その性能を評価した。その結果、話者継続・交替時の視線行動と対話行為の両方の情報を使用することで、1~4点の個人の共感スキルのEC値を平均誤差0.183という高精度に推定することが可能であることが示唆された。

キーワード: スキル推定, 共感スキル, 話者交替, 対話行為, 視線行動

Estimation of Personal Empathy Skill Level Using Dialogue Act and Eye-gaze during Turn-keeping/changing

RYO ISHII^{1,a)} KAZUHIRO OTSUKA^{2,†1} SHIRO KUMANO² RYUICHIRO HIGASHINAKA¹ YUSHI AONO¹

Received: April 13, 2020, Accepted: October 6, 2020

Abstract: We explored gaze behavior that occurs towards the end of utterances and dialogue acts (DAs) during turn-keeping and changing to estimate empathy skill levels in multi-party discussions. We collected data on empathy skill levels, utterances that include five DA categories, and gaze behavior from participants in four-person discussions. The results of an analysis suggest that gaze behavior accompanying utterances that include the DA categories during turn-keeping and changing differs in accordance with people's empathy skill levels. The implementation and evaluation results of our proposed model for estimating empathy skill levels suggest that using both gaze behavior and DAs during turn-keeping and changing is effective for estimating an individual's empathy skill level in multi-party discussions.

Keywords: skill-estimation, empathy skill, turn-changing, dialogue act, gaze behavior

1. 序論

近年、人間の言語・非言語的行動を手掛かりに、コミュ

ニケーションスキルを推定する技術に大きな注目が集まっている。たとえば、演説スキル [1], [2], 説得力 [3], 面接での会話スキル [4], リーダーシップ [5], グループディスカッションでの会話スキル [6] を定量的に推定するモデルが提案されている。

我々は、多人数対話における個人の「共感スキル（共感特性）」を定量的に推定するモデルの構築に取り組んでいる [7], [8]。他者の感情を理解し共有する能力である共感スキルは、最も重要な社会的スキルの1つであり、心理学分野では長い間、研究がなされている [9], [10]。これまで、我々

¹ 日本電信電話株式会社 NTT メディアインテリジェンス研究所
NTT Media Intelligence Laboratories, NTT Corporation,
Yokosuka, Kanagawa 239-0847, Japan

² 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corpora-
tion, Atsugi, Kanagawa 243-0198, Japan

^{†1} 現在、横浜国立大学大学院工学研究院
Presently with Faculty of Engineering, Yokohama National
University

^{a)} ryo.ishii.ct@hco.ntt.co.jp

は対話中の話者継続・交替時の発話末付近の視線行動の情報をを用いて、個人の共感スキルレベルを推定する試みを行ってきた [7], [8]. 既存のスキル推定研究 [1], [2], [3], [4], [5], [6] では、主に会話全体から得られる発話、視線行動、韻律などの全体量を推定のための特徴量として用いていた。これに対して、我々は会話の重要な局面である話者継続・交替時の発話末における 1 秒程度のごく短い時間区間の行動に焦点を当てた [7], [8]. その理由として、話者継続・交替を施行するために、発話言語だけでなく、視線行動を用いていることが広く知られており [11], [12], [13], [14], [15], [16], このような視線行動が個人のコミュニケーションスキルと関連性があると考えたからである。推定器の構築・評価を行った結果、話者継続・交替時に起こるミクロな視線行動のみから、既存研究で用いられる会話全体から得られる特徴量を用いるよりも、より高精度な共感スキル推定器を構築できることを明らかにした [7], [8].

本研究では、より高精度な推定器の実現を目指して、我々の先行研究で利用した視線行動 [7], [8] に加えて、話者継続・交替を円滑に行うために重要な役割を行っている、発話の意図を示す対話行為の情報を焦点を当てる。対話行為とは、話者が何かしらの意図を持って発話することであり、その意図の種類について分類がなされている [17], [18]. 会話参加者にとって、会話中の対話行為の情報は、発話の継続や交替を円滑に行うための重要な情報源であると考えられている。たとえば、もし現話者が質問や確認の意図を持つ発話（本研究では、このような発話の対話行為を“ターン委譲”と定義する）を行えば、対話相手は次に発話を行うことを現話者から期待されていることを明示的に理解可能である。このような発話による明示的なターン委譲の意図が現話者から示されれば、話者交替のための視線行動によるターン授受のための緻密なインタラクションの必要はない可能性が考えられる。一方で、発話の意図が、ターン委譲ではなく、自己開示や共感を示すものであった場合であっても、話者交替は生じる。その場合、発話情報のみから明示的に現話者の話者継続・交替の意思を、対話相手は理解できないと考えられる。このような際には、適切に視線行動を用いて、話者継続・交替を円滑に行うのではないかと考えられる。このように、対話行為と視線行動は、話者継続・交替時において、複合的に同時に利用されている。そして、現話者の対話行為に応じて、話者継続・交替時の現話者や非話者の視線行動は大きく異なる可能性が考えられる。

このような背景から、本研究では、これまで利用されていた視線行動のみ [7], [8] だけではなく、発話の対話行為の情報にも焦点を当て、2つの情報がどのように利用されているかについて着目する。我々は、発話の対話行為に共起する視線の利用方法において、共感スキルレベルによって差異が見られるのではないかと仮説を立てた。もし、このような差異を見い出せば、その情報は先行研究で利用さ

れている視線行動のみの特徴量よりも詳細なパラメータであり、それを利用することで、対話行為を用いずに視線行動の特徴量のみを用いた推定器 [7], [8] よりも、より高精度な共感スキル推定器を構築できる可能性が考えられる。本研究では、個人の共感スキルレベルと話者継続・交替時の対話行為と視線行動の関係性を分析的に検証し、その結果を基に対話行為と視線行動を用いたスキル推定モデルを構築する。本研究の取り組みは、視線行動と対話行為の組合せ情報が、個人の共感スキルとどのような関係性を持つのかを検証する新たな試みである。

本研究では、4 人対話での参加者の Davis の対人反応性指標 (IRI: Interpersonal Reactivity Index) [19] (最も広く利用されている個人の共感スキルを測る指標)、視線行動、および 5 種類の対話行為カテゴリ (情報提供、自己開示、共感、ターン委譲、そのほか) の分類を含む発話情報を収集する。次に、対話行為カテゴリごとに、個人の共感スキルレベルに応じた話者継続・交替時の視線行動の差異について分析を行う。最後に、分析結果を基に、視線行動と対話行為情報を用いた個人の共感スキルレベルの推定器を構築し、その性能を評価する。

本稿は以下のように構成されている。まず 2 章では、多人数対話のコーパス構築について述べる。3 章では、話者継続・交替時の対話行為ごとの視線行動と共感スキルとの関係性の分析、4 章では、推定モデルの実装と評価について述べる。5 章では、分析結果、推定モデル、システム化の可能性についての議論を行い、6 章で、本研究から導かれる結論と今後の展望を述べる。

2. 対話コーパス

対話コーパス収集のための対話の参加者は初対面の同世代 (20 代~30 代) の女性 4 人であり、異なる参加者が 4 グループ参加した (参加者は計 16 人)。各グループの参加者は、向かい合うように着席し、与えられた議題に対して意見を交わして 10 分以内に 1 つの結論を出すことが教示された。与えられた議題は、“男性と女性はどちらが得であるか”、“恋愛と結婚は一緒であるか別であるか”、といった参加者間で意見の分かれやすいものを採用した。発話の収録のために各被験者の胸につけられた指向性ピンマイク (SONY ECM-88) を用いた。対話状況の全体的な外観や参加者の様子の収録として、ビデオカメラ (FLIR rasshopper2) を用いた。映像は 30 Hz で収録された。撮影されたビデオの一例を図 1 に示す。1 つの視点の映像のみから、参加者全員の挙動の観察およびアノテーションを可能とするために、参加者の座席は半円座とした。各グループの参加者を区別するために、座席の左から順に P1, P2, P3, P4 と呼ぶ。収録した音声、ビデオデータおよび、共感スキルを測るアンケート用紙の結果から下記のコーパスデータを収集した。



図 1 複数人対話の様子を収めたビデオの一例

Fig. 1 Example video showing multi-party meeting.

- 発話：熟練したアノテータが ELAN [20] を用いて、音声情報から発話内容の書き起こしを行った。発話単位として、定めた長さ以上の無音区間によって区切られた一連の音声区間である間休止単位 IPU (Inter-Pausal Unit) を利用した。本研究では 200 ms 以上の無音区間を区切りとした。なお、発話の開始・終了時刻は視線情報の時間分解能と一致させるために、30 Hz の時間分解能で付与された。IPU を抽出した後、あいづち [21] を除外し、同一の人物による継続した IPU を 1 つの発話ターンとした。そして、時間的に連続した IPU のペアを作成し、話者継続および話者交替時の IPU の組を作成した。作成された IPU の組は、話者継続時で 1,227、話者交替時で 129 であった。
- 対話行為：各発話 (IPU) の発話内容を基に対話行為を付与する。対話行為は、発話文の単語、文字の N-gram、語彙体系の情報を用いた推定手法 [17], [18] を利用して、自動取得した。この手法では、33 の細かな対話行為カテゴリが推定される。本研究で扱うデータ数に対して、33 のカテゴリ数は多すぎるため、これらのカテゴリを少数の主要なカテゴリに統合して扱うこととした。本研究で扱うような参加者が自由に発話を行う対話において、“情報開示”、“自己開示”、“共感”、“ターン委譲”を明示する“質問”に関する対話行為カテゴリに該当する発話の出現頻度は高く、主要なカテゴリであることが知られている [17], [18] ため、これらのカテゴリを利用することを考えた。また、本研究が着眼する話者交替に関連するターン委譲に類するカテゴリは、前述した“質問”のカテゴリ以外に、発生頻度が必ずしも高くないものとして“確認”、“提案”のカテゴリが存在する。これらのターン委譲に関する発話カテゴリを網羅的に扱うために、“質問”、“確認”、“提案”の 3 つのカテゴリを統合して、“ターン委譲”のカテゴリを新たに設定した。さらに、これらの 4 つのカテゴリに含まれないものは、出現頻度があまり高くなくことが知られているため、“そのほか”のカテゴリとして統合した。以上より、本研究では下記の 5 つの対話行為

表 1 共感スキル評定値測定時の各因子における質問文の一例
Table 1 Example of questionnaire sentence for measuring score of each indicator of empathy skill.

因子	質問文の一例
共感的関心 (EC)	自分より不運な人たちを心配し、気にかけることが多い。
視点取得 (PT)	何かを決める前には、自分と意見が異なる立場のすべてに目を向けるようにしている。
個人的苦痛 (PD)	非常事態では、不安で落ち着かなくなる。
想像性 (FS)	小説に登場する人物の気持ちに深く入り込んでしまう。

カテゴリを利用することとした。

- 情報提供：情報を提供する
- 自己開示：自身の情報を開示する
- 共感：共感したことを伝達する
- ターン委譲：次の発話を促す（質問、提案、確認のカテゴリを統合）
- そのほか：上記 4 つのカテゴリに分類されないもの（挨拶、あいづち、非同意、繰返し、いい換え、承認、感謝、謝罪、フィラー、感嘆、そのほかのカテゴリを統合）
- 視線：NTT MM-Viewer [22] を利用すると、3 次元空間内に参加者全員の顔のアップ映像をそれぞれの座席位置に合わせて 3 次元的に配置し、その映像を自由な視点から観察することが可能である。これを利用して、アノテータが動画 (30 fps) の各画像フレームに対して注視対象のラベリングを行った。ラベリングされた注視対象は、“4 人の各参加者 (P1, P2, P3, P4)” および“人以外”の 5 種類である。
- 共感スキル評定値：すべての参加者の IRI の評定値 [19] を、所定の質問紙によって取得した。個人の共感性を「共感的関心 (EC : Empathic Concern)」、 「視点取得 (PT : Perspective Taking)」、 「個人的苦痛 (PD : Personal Distress)」、 「想像性 (FS : Fantasy Scale)」の 4 つの側面 (因子) から測定する。具体的に、4 つの各因子に対して 7 項目 (計 28 項目) の質問を用いて評定値を測定する。各因子における質問文の一例を表 1 に示す。EC は同情などの他者指向的感情の喚起されやすさ、PT は他者の視点に立って、その他者の気持ちを考える程度、PD は他者の苦痛の観察により自己に生じられる不安や恐怖にとらわれてしまう程度、FS は物語などのフィクションの登場人物に自分を置きかえようと想像する傾向、といった側面をそれぞれ測定する。表 2 に、取得した 16 人の共感スキル評定値の各因子の平均値と分散、表 3 に、因子間の相関係数を示す。表 2 を見ると、4 つの因子の平均値は、2.59~3.10 であり、1~4 点の評定値の平均値 2.50 に比べて極端に大きな差異は見られなかった。

表 2 共感スキル評定値の各因子の平均値と分散

Table 2 Mean and variance of score of each indicator of empathy skill.

因子	平均値	分散
共感的関心 (EC)	2.59	0.64
視点取得 (PT)	3.10	0.50
個人的苦痛 (PD)	2.52	0.59
想像性 (FS)	2.81	0.48

表 3 共感スキル評定値の因子間の相関係数

Table 3 Correlation coefficient between scores of indicators of empathy skill.

因子	相関係数
共感的関心 (EC)–視点取得 (PT)	−0.14
共感的関心 (EC)–個人的苦痛 (PD)	0.37
共感的関心 (EC)–想像性 (FS)	0.62
視点取得 (PT)–個人的苦痛 (PD)	−0.11
視点取得 (PT)–想像性 (FS)	−0.10
個人的苦痛 (PD)–想像性 (FS)	0.13

また、分散は 0.48~0.64 であり、参加者の評定値の分散が極端に小さいということはなかった。よって、これらの評定値を用いて分析を行うこととした。また、表 3 を見ると、EC と PD の相関係数は 0.37 で弱い正の相関、EC と FS の相関係数は 0.62 でやや強い正の相関が見られた。そのほかの組合せでは、相関は見られなかった。

3. 視線行動と共感スキルの分析

3.1 分析方法

日本人は互いの意志や感情を尊重し、他者との争いや過度な議論を避けようとする傾向にあるため、IRI の 4 つの共感スキル尺度のうち、共感的関心 (EC) が日本人の対話において最も重要なスキルの 1 つであると考えられている [23], [24]。このような背景から、共感スキルと視線行動、対話行為の関連性を検証する最初の試みとして、本研究では特に EC に着目をする。

図 2 に 16 人の参加者の EC の評定値のヒストグラムを示す。分析にあたって、EC の評定値の高低に応じて、各対話行為カテゴリに分類される発話に共起する視線行動の差異を調査することとした。具体的に、EC の評定値に基づいて参加者をスキル高、スキル低の 2 群に分け、2 群間で対話行為と視線行動の組合せにどのような発生頻度の差があるかを分析した。2 群のサンプル数に偏りが出ないように、評定値の閾値は EC の全体の平均値 2.50 に設定し、高・低群それぞれ 8 名とした。

分析に使用する発話の対話行為情報は、前章で示した 5 種類の対話行為カテゴリを用いる。次に、分析に使用する視線行動について説明する。本研究では、任意の時間

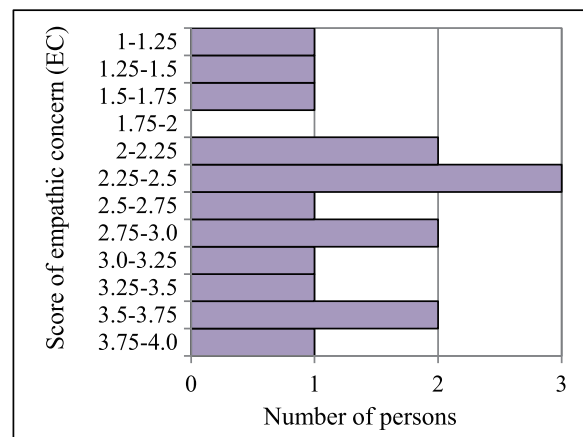


図 2 16 人の参加者の EC 評定値の分布

Fig. 2 Distribution of EC scores for 16 participants.

区間内で、現話者、非話者、人以外の注視対象に対して、どのような順序で視線を向けたか、また人物を注視した際に相互注視が起きたのかといった視線行動の時系列的な遷移を N-gram パターンで表現した視線遷移パターン (GTP: Gaze Transition Pattern) を利用する。我々の先行研究において、発話末の 1,000 ms 前から、200 ms 後までの 1,200 ms の時間区間における GTP が話者交替の次話者や発話タイミングを予測する手がかりとして有用であった [14], [15], [16]。また、そのような背景から、共感スキルの推定においても同一区間内の視線行動から生成される GTP を特徴量として用いたところ、共感スキル推定に有用であった [7], [8]。以上の理由から、本研究においても GTP を視線行動の分析パラメータとして採用した。なお、発話時間が 1,000 ms 未満の発話については、発話開始時刻を分析区間の開始時刻とした。また、発話末 200 ms 後よりも前に発話が始まった際は、分析区間の終端は発話開始直前までとした。これに該当するデータは約 10% 程度であった。GTP に含まれる視線行動は、下記の 7 つに分類される^{*1}。

- *S*: 現話者を注視し、相互注視が起きなかった。
- *SM*: 現話者を注視し、相互注視が起きた。具体的に、現話者を注視していた区間で、1 度でも非話者と相互注視が起きれば、その視線行動は *SM* ラベルとして扱う。
- *L1~L3*: 非話者を注視し、相互注視が起きなかった。なお、*L1*, *L2*, *L3* は異なる人物を示し、1,200 ms の分析区間内で、注視した非話者に対して順番に番号が割り当てられる。たとえば、現話者 *P1* が、非話者 *P2*, 非話者 *P3*, 非話者 *P2* の順に注視した際や、非話者 *P4*, 非話者 *P2*, 非話者 *P4* を順に注視した際に、いずれの注視区間内においても相互注視は起きなかったとすると、注視遷移パターンはどちらも *L1-L2-L1* となる。
- *LM1~LM3*: 非話者を注視し、相互注視が起きた。

^{*1} *S* は Speaker, *L* は Listener, *N* は Next speaker, *M* は Mutual gaze の頭文字であり、GTP の構成要素はこれらの注視対象および相互注視の記号の組合せによって表される。

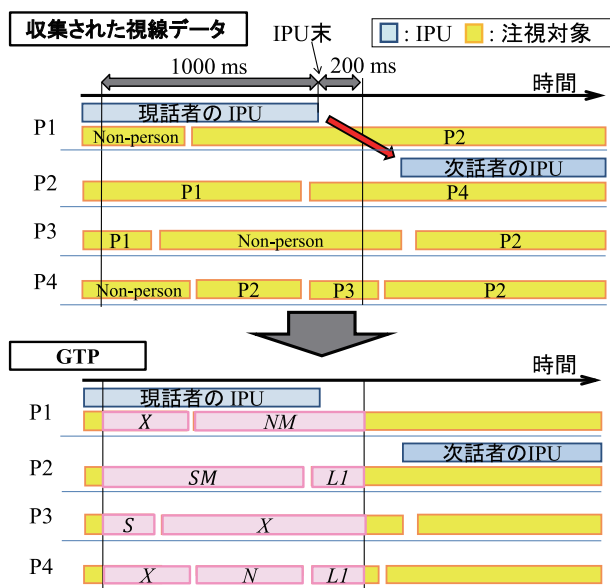


図3 話者交替時のGTPの生成例

Fig. 3 Example of GTP generation during turn-changing.

具体的に、非話者を注視していた区間で、1度でも非話者と相互注視が起きれば、その視線行動はLMラベルとして扱う。LM1, LM2, LM3は異なる人物を表す。

- N: 話者交替時に次話者を注視し、相互注視が起きなかった。
- NM: 話者交替時に次話者を注視し、相互注視が起きた。具体的に、次話者を注視していた区間で、1度でも次話者と相互注視が起きれば、その視線行動はNMラベルとして扱う。
- X: 人以外を注視。たとえば、床や天井など。

図3は、P1からP2への話者交替が発生した際のGTPの生成例を示したものである。発話末付近の分析区間1,200msにおいて現話者P1は、人以外を注視した後、次話者P2を注視している。P1がP2を注視していた区間に、P2もP1を注視しており、相互注視が起きている。よって、P1のGTPはX-NMとなる。次話者P2は、現話者P1と相互注視を行った後、非話者P4を注視した（相互注視なし）。よって、P2のGTPはSM-LIとなる。非話者P3は、現話者P1を注視してから、人以外を注視しており、P3のGTPはS-Xとなる。非話者P4は、人以外を注視した後、次話者P2および非話者P3を注視しており、P4のGTPはX-N-LIとなる。

これらのGTPは、話者継続、交替時の参与役割に応じて異なることが示されており[7], [8], [14], [15]、次節より、各参与役割時に応じて、共感スキルの高、低群で、対話行為ごとにGTPの発生頻度にどのように差異があるかを分析した結果を示す。

3.2 分析結果

話者交替時の現話者、次話者、非話者、話者継続時の現

話者、非話者の5つの参与役割時において、スキル高・低群で各対話行為にともなって出現するGTPの発生頻度どのように差異があるかを分析した*2。具体的に、5つの参与役割条件下において、各対話行為カテゴリごとに、スキル高・低群間でGTPの発生頻度に差異があるかを統計検定を用いて検証した。その結果、話者交替時の非話者、話者継続時の現話者、非話者の3つの参与役割時において、いくつかの対話行為カテゴリにともなったGTPの発生頻度に、スキル高・低群間で差異が見られた。一方、話者交替時の現話者、次話者の2つの参与役割時には差異が見られなかった。話者交替時の現話者と次話者はターン授受のために相互注視を行うなど、他の参与役割に比べて、発話末付近で視線行動に大きな特徴があることが知られている[11], [12], [13], [14], [15], [16]。そのため、スキル高・低群間で何らかの差異が生じるのではないかと期待されるが、差が見られなかったのは意外な結果であった。これについては、5.1節において議論する。

以降、本章では、差異の見られた3つの参与役割時の分析結果について、特に重要な結果が得られたものから順に、話者交替時の非話者、話者継続時の非話者、話者継続時の現話者での分析結果を報告する。

3.2.1 話者交替時の非話者における分析

話者交替時の非話者の258のデータ（129の各話者交替シーンに2名の非話者が存在）を分析したところ、全部で23種類のGTPが出現したが、19のGTPは発生頻度が全体の5%未満と少なかったため、Othersカテゴリとして統合した。

最初に、対話行為カテゴリ間で5つに分類した各GTPの発生頻度の分布に差異があるかを、スキル高群・低群別に、カイ二乗検定を行って検証した。その結果、スキル高群において有意差が認められた（スキル高群： $\chi^2 = 40.7$, $df = 28$, $p < .05$ ）。よって、対話行為のカテゴリが異なると、話者交替時の非話者のGTPの出現頻度の分布が異なることが示された。

次に、各対話行為カテゴリにおいて、スキル高・低群で、GTPに差異があるかを検証した。この際、Family-wise error rate 調整のため、Holm法を用いてp値を補正した。その結果、対話行為カテゴリが情報提供、自己開示、共感であったときに有意差または有意傾向が認められた（情報提供： $\chi^2 = 12.0$, $df = 4$, $p < .10$, 自己開示： $\chi^2 = 15.1$, $df = 4$, $p < .05$, 共感： $\chi^2 = 16.9$, $df = 4$, $p < .01$ ）。対話行為カテゴリがターン委譲、そのほかであるときは、有意差が認められなかった。よって、発話の対話行為カテゴリが情報提供、自己開示、共感であったときの話者交替時の非話者のGTPは、共感スキル高・低群で差異があることが分かった。

さらに、差異の見られた3つの対話行為カテゴリにおい

*2 対話行為およびGTPの発生頻度には、あるカテゴリのものが増えればその分、他のカテゴリのものは減るという相互依存関係がある。

て、5つに分類した GTP のうち、どの GTP でスキル高・低群で差異があるのかを検証するために、残差分析を実施した [25]。残差分析によって得られる調整済み残差から、特徴的な個所を抽出する。各対話行為カテゴリにおけるスキル高・低群でのサンプル数、各 GTP の発生頻度と残差分析の結果を、表 4、表 5、表 6 に示す。なお、5つの

表 4 対話行為カテゴリが“情報提供”であった際の話者交替時の非話者の共感スキル高・低群ごとのサンプル数 N 、GTP の発生頻度、残差分析の結果 ($\dagger: p < .10$, $*: p < .05$, $**: p < .01$)

Table 4 Occurrence frequency of GTP and residual analysis results for listeners who have high/low empathy skill level during turn-changing for dialogue act category of information provision ($\dagger: p < .10$, $*: p < .05$, $**: p < .01$).

GTP \ 共感スキル	低群 ($N = 26$)	高群 ($N = 33$)
X	0.500 (+**)	0.121 (−**)
S	0.077	0.182
$S-X$	0.077 (−*)	0.273 (+*)
$X-S$	0.038	0.091
<i>Others</i>	0.308	0.333

表 5 対話行為カテゴリが“自己開示”であった際の話者交替時の非話者の共感スキル高・低群ごとのサンプル数 N 、GTP の発生頻度、残差分析の結果 ($\dagger: p < .10$, $*: p < .05$, $**: p < .01$)

Table 5 Occurrence frequency of GTP and residual analysis results for listeners who have high/low empathy skill level during turn-changing for dialogue act category of self-disclosure ($\dagger: p < .10$, $*: p < .05$, $**: p < .01$).

GTP \ 共感スキル	低群 ($N = 33$)	高群 ($N = 72$)
X	0.394 (+**)	0.153 (−**)
S	0.030 (−**)	0.264 (+**)
$S-X$	0.121 (+†)	0.042 (−†)
$X-S$	0.061	0.042
<i>Others</i>	0.394	0.500

表 6 対話行為カテゴリが“共感”であった際の話者交替時の非話者の共感スキル高・低群ごとのサンプル数 N 、GTP の発生頻度、残差分析の結果 ($\dagger: p < .10$, $*: p < .05$, $**: p < .01$)

Table 6 Occurrence frequency of GTP and residual analysis results for listeners who have high/low empathy skill level during turn-changing for dialogue act category of empathy ($\dagger: p < .10$, $*: p < .05$, $**: p < .01$).

GTP \ 共感スキル	低群 ($N = 25$)	高群 ($N = 35$)
X	0.480 (+**)	0.057 (−**)
S	0.080 (−*)	0.343 (+*)
$S-X$	0.000	0.029
$X-S$	0.120	0.114
<i>Others</i>	0.320	0.457

GTP の発生頻度の合計は、スキル高・低群それぞれにおいて 1 となる。残差分析の結果については、有意水準 5% を採用し、5~10% については有意傾向が見られるものと解釈する。

表 4~表 6 において、 N または NM を構成要素に持つ GTP は含まれなかった。その理由について述べる。従来研究において、話者交替時に現話者と次話者が相互注視をすることは、円滑にターン授受を行うことが知られている [12], [13], [14], [15]。また、そのとき非話者は、現話者が次話者を注視したのに遅れて、次話者の発話を聞くために注視を開始する傾向があることが示唆されている [13]。GTP の分析区間末は発話末直後 200 ms までの時間であり、この時間内で、話者交替時の非話者がすでに次話者に視線を向け始めているケースはほとんどなかったためだと考えられる。

残差分析の結果から、具体的に下記の結果が得られた。

- 表 4 より、対話行為カテゴリが情報提供であったとき、話者交替時の非話者の $S-X$ の発生頻度が、共感スキル高群は低群よりも有意に高い。これに対して、 X の発生頻度が、共感スキル低群は高群よりも有意に高い。
- 表 5 より、対話行為カテゴリが自己開示であったとき、話者交替時の非話者の S の発生頻度が、共感スキル高群は低群よりも有意に高い。これに対して、 X の発生頻度が、共感スキル低群は高群よりも有意に高く、 $S-X$ の発生頻度が、高いという有意傾向にある。
- 表 6 より、対話行為カテゴリが共感であったとき、話者交替時の非話者の S の発生頻度が、共感スキル高群は低群よりも有意に高い。一方、 X の発生頻度が、共感スキル低群は高群よりも有意に高い。

まず、対話行為カテゴリ内の GTP と共感スキルの関係性に着目した際の興味深い結果として、 S と $S-X$ といった非常に似たような視線行動を行う GTP 間において、共感スキルとの関係性が逆転していることがあげられる。前述したとおり、対話行為カテゴリが自己開示であったとき、話者交替時の非話者の S の発生頻度が、共感スキル高群は低群よりも有意に高く、 $S-X$ の発生頻度は、逆に共感スキル低群は高群よりも高い傾向が見られた (表 5 参照)。 S は分析区間 1,200 ms において現話者を一方的に注視し続けていることを示す。また、 $S-X$ は現話者を一方的に注視していたのを止め、人以外に視線を向けたパターンである。すなわちこの 2 つの視線行動の違いは、非常に短い分析区間内で現話者から視線を外したか否かという微細な視線行動の差である。この結果から、対話行為ごとに分類した際の共感スキルと GTP の分析の有用性が示唆された。

次に、対話行為カテゴリ間でのスキル高・低群と GTP の発生頻度の傾向が異なる顕著な例として、 $S-X$ の発生頻度に着目する。具体的に、対話行為カテゴリが情報提供

であったときは、スキル高群の方が S - X の発生頻度が高いのに対して、対話行為カテゴリが自己開示であったときは、スキル低群の方が発生頻度が高い傾向にある（表 4 参照）。すなわち、話者交替の非話者であったときに、現話者から人以外に視線を向ける（相互注視なし）頻度が、対話行為カテゴリが情報提供であれば、共感スキルが高い人物の方が高い傾向にあり、対話行為カテゴリが自己開示であれば、共感スキルが低い人物の方が高い傾向にある。この理由については、5.2 節で詳しく議論する。

このように、現話者の発話の対話行為カテゴリが情報提供、自己開示、共感であったとき、話者交替時の非話者の GTP は、共感スキル高・低群で差異があることが分かった。これらの結果から、発話の対話行為カテゴリが情報提供、自己開示、共感であったときの話者交替時の非話者の GTP の発生頻度は、EC スコアの推定に有用な特徴量となる可能性が示唆された。

3.2.2 話者継続時の非話者における分析

3.2.1 項の分析と同様の手順で分析を行った。話者継続時の非話者の 3,681 のデータ（1,227 の各話者継続シーンに 3 名の非話者が存在）を分析したところ、全部で 66 種類の GTP が出現したが、59 の GTP は発生頻度が全体の 5% 未満と少なかったため、*Others* カテゴリとして統合した。

3.2.1 項の分析と同様の手順で、対話行為カテゴリ間で 8 つに分類した各 GTP の発生頻度に差異があるかを、スキル高群・低群別に、カイ二乗検定を行って検証した。その結果、有意差が認められた（スキル高群： $\chi^2 = 40.7$, $df = 28$, $p < .05$, スキル低群： $\chi^2 = 39.0$, $df = 28$, $p < .05$ ）。よって、対話行為カテゴリが異なると、話者継続時の非話者の GTP の発生頻度の分布が異なることが示された。

次に、各対話行為カテゴリにおいてスキル高・低群間で、GTP に差異があるかを検証した結果、対話行為カテゴリが情報提供、自己開示であるとき、スキル高・低群間で、GTP の発生頻度に有意差または有意傾向が認められた（情報提供： $\chi^2 = 16.2$, $df = 7$, $p < .10$, 自己開示： $\chi^2 = 25.5$, $df = 7$, $p < .01$ ）。対話行為カテゴリが共感、ターン委譲、そのほかであるときは、有意差が認められなかった。よって、発話の対話行為カテゴリが情報提供、自己開示であったときの話者継続時の非話者の GTP は、共感スキル高・低群で差異があることが分かった。

さらに、差異の見られた対話行為カテゴリ（情報提供、自己開示）において、8 つに分類した GTP のうち、どの GTP でスキル高・低群で差異があるのかを検証するために、残差分析を実施した [25]。対話行為カテゴリが情報提供、自己開示であった際のスキル高・低群でのサンプル数、各 GTP の発生頻度と残差分析の結果を、表 7、表 8 に示す。具体的に下記の結果が得られた。

- 表 7 より、対話行為カテゴリが情報提供であったとき、話者継続時の非話者の S , SM - X , *Others* の発生

表 7 対話行為カテゴリが“情報提供”であった際の話者継続時の非話者の共感スキル高・低群ごとのサンプル数 N , GTP の発生頻度、残差分析の結果（†: $p < .10$, *: $p < .05$, **: $p < .01$ ）

Table 7 Occurrence frequency of GTP and residual analysis results for listeners who have high/low empathy skill level during turn-keeping for dialogue act category of information provision (†: $p < .10$, *: $p < .05$, **: $p < .01$).

共感スキル GTP	低群 (N = 347)	高群 (N = 417)
S	0.424 (−†)	0.477 (+†)
X	0.231 (+*)	0.163 (−*)
SM	0.095	0.103
X - S	0.101 (+†)	0.067 (−†)
S - X	0.052	0.031
SM - X	0.012 (−†)	0.026 (+†)
X - SM	0.006	0.012
<i>Others</i>	0.081 (−†)	0.120 (+†)

表 8 対話行為カテゴリが“自己開示”であった際の話者継続時の非話者の共感スキル高・低群ごとのサンプル数 N , GTP の発生頻度、残差分析の結果（†: $p < .10$, *: $p < .05$, **: $p < .01$ ）

Table 8 Occurrence frequency of GTP and residual analysis results for listeners who have high/low empathy skill level during turn-keeping for dialogue act category of self-disclosure (†: $p < .10$, *: $p < .05$, **: $p < .01$).

共感スキル GTP	低群 (N = 482)	高群 (N = 542)
S	0.392 (−*)	0.459 (+*)
X	0.245 (+**)	0.133 (−**)
SM	0.095	0.118
X - S	0.062	0.076
S - X	0.025	0.026
SM - X	0.012	0.018
X - SM	0.041 (+†)	0.024 (−†)
<i>Others</i>	0.127 (−†)	0.146 (+†)

頻度が、共感スキル高群は低群よりも有意に高い傾向にある。一方で、話者継続時の非話者の X の発生頻度が共感スキル低群は高群よりも有意に高い。また同様に、 X - S の発生頻度が共感スキル低群は高群よりも高い傾向にある。

- 表 8 より、対話行為カテゴリが自己開示であったとき、話者継続時の非話者の S の発生頻度が、共感スキル高群は低群よりも有意に高い。また同様に、*Others* の発生頻度が、共感スキル高群は低群よりも高い傾向にある。一方で、話者継続時の非話者の X の発生頻度が、共感スキル低群は高群よりも有意に高い。また同様に、 X - SM の発生頻度が高い傾向にある。

対話行為カテゴリ内の GTP と共感スキルの関係性に着目した際の興味深い結果として、 S と X - S といった非常

に似たような視線行動を行う GTP 間において、共感スキルとの関係性が逆転していることがあげられる。前述したとおり、対話行為カテゴリが情報提供であったとき、話者継続時の非話者の S の発生頻度が、共感スキル高群は低群よりも高く、 X - S の発生頻度は、逆に共感スキル低群は高群よりも高い傾向にあるという結果であった（表 7 参照）。ここで、 S は分析区間 1,200 ms において現話者を一方的に注視し続けていることを示す。また、 X - S は人以外への注視状態から現話者に一方的に視線を向けたパターンである。すなわちこの 2 つの視線行動の違いは、非常に短い分析区間内で現話者に最初から視線を向けていたか、それとも分析区間の途中から向け始めたのかという微細な差である。この結果から、対話行為カテゴリごとに分類した際の共感スキルと GTP の分析の有用性が示唆された。

このように、現話者の発話の対話行為カテゴリが情報提供、自己開示であったとき、話者継続時の非話者の GTP の発生頻度は、共感スキル高・低群で差異があることが分かった。これらの結果から、発話の対話行為カテゴリが情報提供、自己開示であったときの話者継続時の非話者の GTP の発生頻度は、EC スコアの推定に有用な特徴量となる可能性が示唆された。

3.2.3 話者継続時の現話者における分析

3.2.1 項、3.2.2 項の分析と同様の手順で、分析を行った。話者継続時の現話者の 1,227 のデータ（1,227 の各話者継続シーンに 1 名の現話者が存在）を分析したところ、全部で 32 種類の GTP が出現したが、26 の GTP は発生頻度が全体の 5%未満と少なかったため、*Others* カテゴリとしてまとめた。

対話行為カテゴリ間で 7 つに分類した各 GTP の発生頻度の分布に差異があるかを、スキル高群・低群別に、カイ二乗検定を行って検証した。その結果、有意差が認められた（スキル高群： $\chi^2 = 37.0$, $df = 24$, $p < .05$, スキル低群： $\chi^2 = 40.6$, $df = 24$, $p < .05$ ）。よって、対話行為カテゴリが異なると、話者継続時の現話者の GTP の発生頻度の分布が異なることが示された。

次に、各対話行為カテゴリにおいて、スキル高・低群間で、GTP に差異があるかを検証した結果、対話行為カテゴリがそのほかであった際のみ有意差が認められた（ $\chi^2 = 17.1$, $df = 6$, $p < .05$ ）。対話行為カテゴリが情報提供、自己開示、共感、ターン委譲であるときは、有意差が認められなかった。よって、発話の対話行為カテゴリがそのほかであったときの話者継続時の現話者の GTP は、共感スキル高・低群で差異があることが分かった。

さらに、差異の見られた対話行為カテゴリ（そのほか）において、7 つに分類した GTP のうち、どの GTP でスキル高・低群で差異があるのかを検証するために、残差分析を実施した [25]。各 GTP における発生頻度、サンプル数、残差分析の結果を、表 9 に示す。表 9 を見ると、対話行

表 9 対話行為カテゴリが“そのほか”であった際の話者継続時の現話者の共感スキル高・低群ごとのサンプル数、GTP の発生頻度、残差分析の結果（†: $p < .10$, *: $p < .05$, **: $p < .01$ ）

Table 9 Occurrence frequency of GTP and residual analysis results for speakers who have high/low empathy skill level during turn-keeping for dialogue act category of others (†: $p < .10$, *: $p < .05$, **: $p < .01$).

GTP	共感スキル	
	低群 (N = 158)	高群 (N = 132)
X	0.430 (−*)	0.553 (†*)
$LM1$	0.158 (†*)	0.076 (−*)
X - $LM1$	0.133 (†*)	0.061 (−*)
L	0.057	0.045
X - $LM1$ - X	0.038	0.053
$LM1$ - X	0.051 (†*)	0.008 (−*)
<i>Others</i>	0.133 (−†)	0.205 (††)

為カテゴリがそのほかであったとき、話者継続時の現話者の X の場合の発生頻度が、共感スキル高群は低群よりも有意に高い。また、*Others* の場合の発生頻度が共感スキル高群は低群よりも高いという傾向にある。一方で、 $LM1$, X - $LM1$, $LM1$ - X の場合の発生頻度は、共感スキル低群が高群よりも有意に高い。

この結果の興味深い点の 1 つとして、 X , X - $LM1$, $LM1$ - X といった非常に似たような視線行動ラベルを持つ GTP 間において、共感スキルとの関係性が異なっていることがあげられる。 X は分析区間 1,200 ms において、人以外を注視し続けていることを示す。また、 X - $LM1$ は人以外から非話者に視線を向けて相互注視が起きたパターン、 $LM1$ - X は非話者と相互注視を行ってから人以外に視線を向けたパターンである。 X と X - $LM1$ の視線行動の違いは、非常に短い分析区間内で人以外を注視していた状態から、それを続けたのか、区間内の途中で非話者と相互注視をしたかの違いである。また、 X と $LM1$ - X の視線行動の違いは、人以外への注視が、分析区間内で継続的に起こったのか、それとも非話者との相互注視の後に起こったのかの違いである。このような微小な差異にもかかわらず、前述したように、話者継続時の現話者の X の場合の発生頻度が、共感スキル高群は低群よりも高く、 X - $LM1$, $LM1$ - X の場合の発生頻度は、共感スキル低群が高群よりも高いという差異が見られた。この結果から、対話行為ごとに分類した際の共感スキルと GTP の分析の有用性が示唆された。

以上より、発話の対話行為カテゴリがそのほかであったとき、話者継続時の現話者の GTP の発生頻度に、共感スキル高・低群で差異があることから、発話の対話行為カテゴリがそのほかであったときの話者継続時の現話者の GTP の発生頻度は、EC スコアの推定に有用な特徴量となる可能性が示唆された。

4. 共感スキルの推定モデル

4.1 推定モデルの構築と評価方法

分析結果から、話者継続・交替時の参与役割および現話者の対話行為に応じた GTP が、共感スキルの高低によって異なることが明らかとなった。次に、共感スキルの高・低群間で、発生頻度の差異が見られた対話行為ごとの GTP の発生頻度の情報を利用して、参加者の共感スキル（EC の評定値）の推定が可能であるかを検証した。具体的に、分析で差異の見られた、話者継続時の現話者・非話者、話者交替時の非話者の 3 つの参与役割の状況下で、対話行為ごとの GTP の発生頻度を各参加者ごとに算出し、その頻度を特徴量とする共感スキルの推定モデルを構築した。この推定モデルの比較モデルとして、従来研究で利用されているような、会話全体での発話量や注視量を用いた推定モデル、対話行為を考慮せずに GTP のみを利用した推定モデル [7], [8]、対話行為の情報のみを利用したモデル、会話全体から得られる情報と対話行為ごとの GTP の発生頻度のすべての情報を利用した推定モデルの構築を行った。下記に、実装した予測モデルと使用した特徴量を示す。

- 発話量モデル：グループ内での発話の頻度を特徴量として使用。
- 視線量モデル：現話者、非話者に視線を向けていた 2 つの頻度を特徴量として使用。
- 単独 GTP モデル：各参与役割における、GTP の発生頻度を特徴量として使用。このモデルは、先行研究 [7], [8] で提案されたものと同一である。なお、データ数に対して次元数が多すぎるため、貢献度を基に 10 次元まで次元削減を行った。
- 単独 DA モデル：各参与役割時の現話者の対話行為の発生頻度を使用。
- GTP+DA モデル：各参与役割において、3.2 節の分析で有意差のあった対話行為に応じた GTP の発生頻度を特徴量として使用。なお、データ数に対して次元数が多すぎるため、貢献度を基に 10 次元まで次元削減を行った。
- ALL モデル：発話量モデル、視線量モデル、GTP+DA モデルで使用した特徴量のすべてを使用した。なお、データ数に対して次元数が多すぎるため、貢献度を基に 12 次元まで次元削減を行った。

推定モデル構築には、データマイニングツール Weka [26] に実装された SMOreg アルゴリズム [27] を利用した Support Vector Machine for Regression (SVR) を用いた。コストパラメータ C とハイパーパラメータ γ は、最適な値をグリッドサーチで探索した。目的変数として、各個人の EC 評定値を設定した。また、モデルの構築・評価を行う手順として、Leave one-person out cross validation を採用した。これは、15 名の参加者のデータで学習を行い、1 名

表 10 共感スキルの推定モデルの性能

Table 10 Performance of models for estimating empathy skill.

モデル	平均絶対誤差 (MAE)
発話量モデル	0.840
視線量モデル	0.629
単独 GTP モデル	0.220
単独 DA モデル	0.552
GTP+DA モデル	0.183
ALL モデル	0.188

のデータでテストを行う 16 分割交差検定法である。

4.2 推定モデルの評価結果

各モデルにおいて、推定された EC 評定値と真値の平均絶対誤差 (MAE) を表 10 に示す。結果を見ると、まず、発話量モデルの誤差 0.840、視線量モデルの誤差 0.629 に比べて、単独 GTP モデル [7], [8] は誤差 0.220、単独 DA モデルは誤差 0.552 であり、誤差が小さかった。この結果から、参与役割に応じた GTP および対話行為の発生頻度は、会話全体から得られる発話量や視線量を用いた推定モデルよりも共感スキル（EC の評定値）の推定に有用であることが示唆された。

GTP+DA モデルは、誤差 0.183 であり、すべてのモデルの中で最も性能が高かった（16 分割交差検定結果を用いた、GTP+DA モデルと単独 GTP モデル [7], [8] の対応のある t 検定の結果： $t(15) = -6.28$, $p < .01$, GTP+DA モデルと単独 DA モデルの対応のある t 検定の結果： $t(15) = -77.4$, $p < .01$ ）。また、GTP+DA モデルの誤差 (0.183) と ALL モデルの誤差 (0.188) に有意な差は見られなかった。よって、各参与役割条件下での対話行為に基づく GTP の発生頻度情報は、共感スキルの推定に有用であることが示唆された。

5. 議論

5.1 話者継続・交替時の対話行為・視線行動と個人の共感スキルレベルの分析結果全体について

3 章での分析結果から、話者継続・交替時の現話者の発話の意図を示す対話行為と視線行動の組合せ情報が、個人の共感スキルレベルによって異なることを初めて明らかにした。具体的に、話者交替時に現話者の発話の対話行為カテゴリが情報提供、自己開示、共感であった際の非話者の視線行動 (3.2.1 項参照)、話者継続時に現話者の発話の対話行為カテゴリが情報提供、自己開示であった際の非話者の視線行動 (3.2.2 項参照)、話者継続時に現話者の発話の対話行為カテゴリがそのほかであった際の現話者の視線行動 (3.2.3 項参照) が個人の共感スキルレベルによって異なることを明らかにした。

一方で、話者交替時の現話者、次話者の 2 つの参与役割

時には差異が見られなかった。話者交替時の現話者と次話者はターン授受のために相互注視を行うなど、他の参与役割に比べて、視線行動に大きな特徴があることが知られている [11], [12], [13], [14], [15], [16]。そのため、スキル高・低群間で何らかの差異が生じるのではないかと期待されるが、差が見られなかったのは意外な結果であった。1つの見解として、話者交替時の現話者と次話者では、相互注視をするといった視線のやり取りのルールがある程度、厳守されるため、共感スキルによる行動の差異が出にくい可能性が考えられる。

これとは逆に、話者交替時の非話者は、ターン授受の当事者ではなく、会話の中心的な参与役割ではない。むしろ、そのようなより周辺的な参与役割のときにこそ現れる視線行動が共感スキルに大きく関連しているのではないかと考えられる。

5.2 対話行為と視線行動の共起関係の分析の効果

次に、本研究が着眼した、対話行為と視線行動の共起関係の分析の効果について議論する。対話行為カテゴリと視線行動の共起関係の分析によって、共感スキル高・低群間で差異の見られた顕著な結果の一例として、3.2.1 項における、現話者の対話行為カテゴリが自己開示と情報提示の際、話者交替時の非話者の *S-X* の GTP の発生頻度の分析結果に着目する。

共感スキルが高い人物は低い人物に比べて、話者交替時の非話者であったとき、現話者の対話行為カテゴリが情報提供である際は、現話者への注視状態から視線を人以外に外す (GTP は *S-X*) 傾向にあった (表 4 参照)。GTP の分析区間は、発話末の 1,000 ms 前から、発話末の 200 ms 後までの 1,200 ms であったが、この分析区間の後から、次話者の発話が始まるまでの区間内で、およそ半数のケースで非話者が次話者に視線を向けていた。従来研究において、話者交替時に現話者と次話者が相互注視をすることは、ターン授受を行うための相互行為であることが知られている [12], [13], [14], [15]。また、そのとき非話者は、現話者との相互注視を避けるために現話者から視線を外したり、現話者が次話者を注視したのに続いて、次話者の発話開始に注意を向けるために次話者を注視したりする傾向があることが示唆されている [13]。本研究で得られた結果は、このような従来知見で報告された視線行動の傾向と非常に類似している。よって、共感スキルレベルの高い人物は、話者交替時の非話者であった際に、現話者が情報提供を行う発話をした際に、このような従来研究で示されているような典型的な視線行動を行う傾向が高いのではないかと考えられる。

一方で、現話者の対話行為カテゴリが自己開示である際は、この傾向は逆転し、共感スキルが低い人物の方が高い人物に比べて、現話者への注視状態から視線を人以外に外

す (GTP は *S-X*) 頻度が高い傾向にあった (表 5 参照)。

このような逆転が起きた理由の1つの見解として、2つの対話行為カテゴリの性質に基づいて考察する。まず、自己開示カテゴリの発話は、現話者自身の自己の情報を開示することである。一方で、情報提供は単に何かしらの情報を他者に伝達する。したがって、自己開示カテゴリの発話は情報提供カテゴリの発話よりも、より個人に関する重要な発話である性質を持つ。一方で、現話者の発話中に視線を向けることは現話者の話に興味を持っていることを示すなど、重要なフィードバック機能を持っていることが知られている [12]。我々が分析で扱った、IRI の EC スコアは、他者指向的な感情の喚起されやすさを示しており、このスコアが高い人物は、より感情的側面の強い自己開示の発話に対して積極的に聞く姿勢があるのではないかとと思われる。これを裏付けるものとして、表 5 の結果では、共感スキルの高い人物は低い人物に比べて、現話者を注視し続けていた (GTP は *S*)。

以上より、共感スキル EC の高い人物は、話者交替時の非話者であったとき、現話者の発話の対話行為カテゴリが情報提供といった単なる情報伝達の際は、話者交替規則に従って、現話者とのターン授受を行わないために、視線を現話者から外すことを重視する可能性が考えられる。一方で、現話者の発話の対話行為カテゴリが感情的側面の強い自己開示であった際には、話者交替規則を重視して現話者から視線を外すことよりも、現話者に視線を向け続けて、現話者の自己開示の発話に興味を持っていることを示すフィードバックを行っているのではないかと考えられる。

この結果から、同じ参与役割における GTP に着目した際に、共起している現話者の対話行為カテゴリが異なることで、異なった (逆の) 共感スキル上の効果をもたらすことがありうることが示唆される。

なお、本分析において利用したデータサンプルは、話者継続シーンで 1,227、話者交替シーンで 129 であり、話者継続時に比べて話者交替時で少量であった。話者交替時の分析については、より多くのデータで検証を行うことで、新たな傾向の発見が見られる可能性もある。今後は、データの拡充と詳細な追加検証を実施し、本研究で得られた結果の信頼性をより高めることにも取り組みたい。

さらに他の結果として、現話者の対話行為カテゴリがターン委譲の際は、共感スキルレベルに応じた視線行動の差異はいっさい見られなかった。この結果から、現話者の発話によるターン移譲は、非常に明示的な話者交替を行うシグナルであるため、共感スキルレベルによらず、どの人物も同様の視線行動を行うのではないかと考えられる。

5.3 話者継続・交替時の非話者の視線行動と共感スキルレベルの関係性の総括

さらに、3章の分析結果に関して、各参与役割ごとに独立

表 11 現話者の発話が情報提供・自己開示カテゴリであった際の話者継続・交替時の非話者の視線行動において、スキル高・低群間の比較によって有意な差が見られた GTP（および視線行動）のまとめ、GTP と視線行動の説明の右側の括弧内は、各条件下における出現頻度を示す。2 群間で頻度が高かった群側に GTP を掲載

Table 11 Summary of listener's GTP (a.k.a gaze behaviors) showed significant difference in occurrence probability when dialogue act category was information provision/self-disclosure during turn-keeping/changing. Numbers in parentheses to right of explanation of GTP and gaze behavior indicate occurrence frequency under each condition. GTP is posted on group side where frequency was high between two groups.

<情報提供>		
	スキル低群	スキル高群
話者交替	X：人以外を注視 (0.500)	S-X：現話者⇒人以外を注視 (0.273)
話者継続	X：人以外を注視 (0.231)	S：現話者を注視 (0.477)
	X-S：人以外⇒現話者を注視 (0.101)	SM-X：現話者と相互注視⇒人以外を注視 (0.026)
<自己開示>		
	スキル低群	スキル高群
話者交替	X：人以外を注視 (0.394) S-X：現話者⇒人以外を注視 (0.121)	S：現話者を注視 (0.264)
話者継続	X：人以外を注視 (0.245)	S：現話者を注視 (0.459)
	X-SM：人以外を注視⇒現話者と相互注視 (0.041)	

的に分析して得られた結果をまとめて、共感スキル高群・低群間における視線行動の傾向差を俯瞰的な視点で考察する。3 章での分析結果では、参与役割が話者継続時と交替時の非話者であった際に共通して、情報提供と自己開示の対話行為カテゴリ下で、スキル高・低群間での視線行動に有意な差が見られた。そのため、ここでは特に、現話者の対話行為カテゴリが情報提供と自己開示であった際に、参与役割が話者継続・交替時で非話者であった 2 つの状況下で、共感スキル高群・低群における視線行動の差について着目して包括的に議論する。

まず、より全体的な傾向を掴みやすくするため、3.2.1 項、3.2.2 項における、情報提供と自己開示のそれぞれの対話行為カテゴリ下で、話者継続・交替時で非話者であったときに、共感スキル高群・低群における視線行動で差があった GTP についての結果を表 11 のようにまとめた。すなわち、表 11 は、表 4、表 5、表 7、表 8 にて、スキル高・低群間で見られた GTP の差異をまとめたものである。なお、頻度の少ない GTP が統合された *Others* カテゴリの差については、解釈が困難であるため割愛する。また、表 11 に示された GTP は、スキル高・低群間で有意差または有意傾向が見られたものについて、頻度が高かったスキル群側にのみ結果を掲載している。

まず、スキル低群の人物は、2 つの対話行為カテゴリ、話者継続・交替時の 2 つの状況のいずれにおいても、スキル高群に比べて X（人以外を注視）が多い傾向にあることが分かる。すなわち、スキル低群の人物は高群に比べて、対話行為カテゴリが情報提供、自己開示の違いにかかわら

ず、また、話者継続・交替の状況の違いにかかわらず、非話者であった際には現話者に視線を向けない傾向が高いことが考えられる。前述したとおり、現話者に視線を向けることは現話者の話に興味を持っていることを示すなど、重要なフィードバック機能を持っていることが知られている [12]。我々が分析で扱った、IRI の EC スコアは、他者指向的な感情の喚起されやすさを示している。この EC スコアが低い人物は、発話末に視線を対話相手に向けない傾向にあることから、対話相手の感情を含む様子を視線に向けて積極的に読み取ろうとしていない、また対話相手へ積極的なフィードバックを行わない傾向にある可能性が考えられる。

一方、スキル高群の人物は、どちらの対話行為カテゴリにおいても、話者継続・交替時を通じて、スキル低群に比べて S（現話者を注視）が全体的に多い傾向にあることが分かる（ただし、情報提供カテゴリにおける話者交替時においては、S に差は見られない）。これはスキル低群とは対照的に、発話末においてつねに現話者に注視している傾向が強いことを示唆する結果である。EC スコアが高い人物は、話者継続・交替時に非話者である際は、一貫して対話相手の感情を含む様子を視線を向けて積極的に読み取り、また対話相手へ積極的なフィードバックを行う傾向にある可能性が考えられる。

5.4 分析における発展的課題やコミュニケーションロボット・エージェントへの知見の応用性

本研究では、視線行動を行った本人のみのスキルの高低

を分類して、現話者の対話行為にともなう視線行動を分析した。本人だけではなく対話相手のスキルの高低との組合せによって、対話行為と視線行動に差異がでる可能性も考えられる。このような、本人と他者の共感スキルと話者継続・交替時の振舞いを検証することも興味深い。

また、本研究では、IRI の中でも EC の評定値を用いて分析を実施した。EC と PD の相関係数は 0.37 で弱い正の相関、EC と FS の相関係数は 0.62 でやや強い正の相関がみられたことから（表 3 参照）、EC の評定値を基にした視線行動の分析結果と、PD、FS を基にした視線行動の分析結果は同様の傾向を示す可能性も考えられる。これについては、今後の検証が必要である。

また、コミュニケーションロボットやエージェントが話者交替を行うときは、従来知見 [12], [13], [14], [15], [16] を基に、多くの場合で視線を対話相手に向けることが採用されている。本研究の知見を考慮すると、コミュニケーションロボットやエージェントが話者交替を行うときは、現話者の対話行為カテゴリを考慮して、視線行動を制御することでより人間らしい視線行動の表出が可能になる可能性も考えられる。

5.5 共感スキル推定器の評価結果

話者継続・交替時の発話末の視線遷移パターン（GTP）と対話行為（DA）を単体で用いた共感スキル推定器はともに、会話全体から得られる発話量や視線量を用いた推定器よりも有用であった。また、GTP の方が DA よりも単体で有用であった。また、GTP と DA を組み合わせた推定器は、単独 GTP モデルを上回る高精度な推定が可能であった。よって、本研究で着目した DA と GTP の組合せ情報が個人の共感スキルを推定するのに有用であることが明らかとなった。

また、GTP+DA モデルと ALL モデルで性能に差がなかったことから、我々が着目する GTP と DA の組合せのみで非常に有用な特徴量であることが示唆された。よって、発話量モデルや視線量モデルといった会話全体から得られる言語・非言語情報よりも、話者交替といった会話で重要なイベントから得られるミクロな振舞いこそが、人物の会話スキルを推定するための特徴量であることが示唆された。また、本研究では、IRI の中でも EC の評定値を扱って推定器の構築を検討した。IRI の EC 以外の 3 つの評定値 PT, FS, PD において、同様の手順で GTP+対話行為モデルを構築した結果、平均絶対誤差が PT で 0.26, FS で 0.17, PD で 0.21 であり、高精度な推定が可能であることが示唆された。このようなことから、話者継続、交替における発話末の視線行動と対話行為の情報は多様なコミュニケーションスキルを推定するための特徴量として有用である可能性が考えられる。

今後、本研究で用いた共感スキル IRI 指標以外のスキル

評定値と、話者継続・交替時の発話末の対話行為、視線行動の関係性の分析を実施することや、スキル推定器を構築して、対話行為と視線行動の特徴量の有用性を広く検証することは非常に興味深い。

5.6 自動推定システム構築の可能性

最後に、本推定技術を用いて、実対話から自動的に共感スキルを推定可能なシステム構築の見込みについて議論する。本研究では、発話区間（IPU）検出、発話テキスト抽出（音声認識）、注視対象（視線）検出を、音声および映像から人手アノテーションにより実施した。よって、これらが自動化できれば、対話中の音声、映像情報のみから、共感スキルを自動推定できるシステムの構築が可能となる。発話区間検出および音声認識は、近年、飛躍的に技術革新を遂げ [28]、雑音環境下でも高精度で検出が可能であり、音声認識アプリがスマートホン上でも広く利用されている。また、顔画像情報から注視対象を推定する技術においても、すでに高精度な注視対象推定が実現されている [29]。よって、自動推定システムの実現可能性は高く、これらの技術を用いた実計測データから、どの程度スキルを正確に推定できるかを検証することも将来課題の 1 つである。

6. 結論と展望

本研究では、多人数対話における話者継続・交替時の現話者の発話の意図を示す対話行為と視線行動の組合せ情報が、個人の共感スキルレベルによって異なることを初めて明らかにした。また、個人の共感スキルレベルを推定するために、対話行為、および視線行動の情報を利用することの有用性を初めて明らかにした。

4 人対話コーパスのデータ分析の結果として、話者継続・交替時に、情報提供、自己開示、共感、そのほかの 4 つの対話行為を含む発話にともなう視線行動は、個人の共感スキルレベルに応じて異なることが示唆された。顕著な結果の一例として、共感スキルが高い人物は低い人物に比べて、話者交替時の非話者であったとき、現話者の対話行為カテゴリが情報提供である際は、現話者から視線をより逸らす傾向にあった。一方で、対話行為カテゴリが自己開示であった際は、この傾向が逆転した。またそのほかの結果として、現話者の対話行為カテゴリがターン委譲の際は、共感スキルレベルに応じた視線行動の差異は見られなかった。

これらの知見から、視線行動と対話行為情報を用いた個人の共感スキルレベルの推定器を構築し、その性能を評価した。その結果、話者継続・交替時の視線行動と対話行為の両方の情報を使用することで、1～4 点の個人の共感スキルの EC 値を平均誤差 0.183 という高精度に推定することが可能であることが示唆された。

本研究によって、会話でのスキル推定研究において、会話全体の言語・非言語行動だけではなく、話者交替時といっ

た会話における重要な場面におけるミクロな言語・非言語行動を用いることが重要である可能性が示唆された。従来研究が扱っている対話場面の様々なスキル推定 [4], [5], [6] において、参与役割に応じた話者継続・交替の視線行動と対話行為の情報は有効に利用できる可能性も考えられる。今後、共感スキル以外の様々なスキル推定へ適用されることを期待したい。また、これまで個人の共感スキルを測るためには、質問紙を用いて測定する必要があったが、本推定技術を利用することで、今後は会話の状況をセンシングするだけで共感スキルを自動測定することが可能となる。

今後の取り組みとして、話者継続・交替時の視線行動以外の非言語行動 [15], [30], [31], [32], [33] も扱い、共感スキルの関連性の解明および、推定器の構築を行う予定である。

参考文献

- [1] Wortwein, T., Chollet, M., Schauerte, B., Morency, L.-P., Stiefelhagen, R. and Scherer, S.: Multimodal public speaking performance assessment, *Proc. ACM ICMI*, pp.43–50 (2015).
- [2] Ramanarayanan, V., Leong, C.W., Chen, L., Feng, G. and Suendermann-Oeft, D.: Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring, *Proc. ACM ICMI*, pp.23–30 (2015).
- [3] Park, S., Shim, H.S., Chatterjee, M., Sagae, K. and Morency, L.-P.: Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach, *Proc. ACM ICMI*, pp.50–57 (2014).
- [4] Nguyen, L., Frauendorfer, D., Mast, M. and Gatica-Perez, D.: Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior, *IEEE Trans. Multimedia*, Vol.16, No.4, pp.1018–1031 (2014).
- [5] Sanchez-Cortes, D., Aran, O., Mast, M.S. and Gatica-Perez, D.: A nonverbal behavior approach to identify emergent leaders in small groups, *IEEE Trans. Multimedia*, Vol.14, No.3, pp.816–832 (2012).
- [6] 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, *Proc. International Conference on Multimodal Interaction*, pp.169–176 (2016).
- [7] Ishii, R., Kumano, S. and Otsuka, K.: Analyzing Gaze Behavior During Turn-taking for Estimating Empathy Skill Level, *Proc. 19th ACM International Conference on Multimodal Interaction, ICMI 2017*, pp.365–373, ACM (online), DOI: 10.1145/3136755.3136786 (2017).
- [8] 石井 亮, 熊野史朗, 大塚和弘: 話者継続・交替時における参与役割に応じた視線行動に基づく共感スキルの推定, ヒューマンインタフェース学会論文誌, Vol.20, No.4, pp.447–456 (2018).
- [9] de Waal, F.B.M.: The antiquity of empathy, *Science*, Vol.336, pp.874–876 (2012).
- [10] Decety, J. and Svetlova, M.: Putting together phylogenetic and ontogenetic perspectives on empathy, *Developmental Cognitive Neuroscience*, Vol.2, No.1, pp.1–24 (2012).
- [11] Sacks, H., Schegloff, E.A. and Jefferson, G.: A simplest systematics for the organisation of turn taking for conversation, *Language*, Vol.50, pp.696–735 (1974).
- [12] Kendon, A.: Some functions of gaze direction in social interaction, *Acta Psychologica*, Vol.26, pp.22–63 (1967).
- [13] 榎本美香: 日本語における聞き手の話者移行適格場の認知メカニズム, ひつじ書房 (2009).
- [14] Ishii, R., Otsuka, K., Kumano, S., Matsuda, M. and Yamato, J.: Predicting Next Speaker and Timing from Gaze Transition Patterns in Multi-Party Meetings, *Proc. International Conference on Multimodal Interaction*, pp.79–86 (2013).
- [15] Ishii, R., Otsuka, K., Kumano, S. and Yamamoto, J.: Predicting of Who Will Be the Next Speaker and When Using Gaze Behavior in Multiparty Meetings, *The ACM Trans. Interactive Intelligent Systems*, Vol.6, No.1, p.4 (2016).
- [16] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Analysis and Modeling of Next Speaking Start Timing based on Gaze Behavior in Multi-party Meetings, *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp.694–698 (2014).
- [17] Meguro, T., Higashinaka, R., Minami, Y. and Dohsaka, K.: Controlling listening-oriented dialogue using partially observable Markov decision processes, *International Conference on Computational Linguistics*, pp.761–769 (2010).
- [18] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T. and Matsuo, Y.: Towards an open-domain conversational system fully based on natural language processing, *International Conference on Computational Linguistics*, pp.928–939 (2014).
- [19] Davis, M.H.: A Multidimensional Approach to Individual Differences in Empathy, *JSAS Catalog Sel. Doc. Psychol.*, Vol.10 (1980).
- [20] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. and Sloetjes, H.: ELAN: A Professional Framework for Multimodality Research, *Proc. 5th International Conference on Language Resources and Evaluation (LREC '06)* (2006).
- [21] Jovanovic, N., op den Akker, R. and Nijholt, A.: Addressee identification in face-to-face meetings, *Proc. Conference of the European Chapter of the Association for Computational Linguistics* (2006).
- [22] Otsuka, K., Araki, S., Mikami, D., Ishizuka, K., Fujimoto, M. and Yamato, J.: Realtime meeting analysis and 3D meeting viewer based on omnidirectional multimodal sensors, *Proc. International Conference on Multimodal Interfaces and Workshop on Machine Learning for Multimodal Interaction*, pp.219–220 (2009).
- [23] 登張真穂: 多次元的視点に基づく共感性研究の展望, 性格心理学研究, Vol.9, No.1, pp.36–51 (2000).
- [24] 菊池章夫: さらに/思いやりを科学する—向社会的行動と社会的スキル, 川島書店 (2014).
- [25] Haberman, S.J.: The analysis of residuals in cross-classified tables, *Biometrics*, Vol.29, pp.205–220 (1973).
- [26] Bouckaert, R.R., Frank, E., Hall, M.A., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H.: WEKA—Experiences with a Java Open-Source Project, *The Journal of Machine Learning Research*, Vol.11, pp.2533–2541 (2010).
- [27] Keerthi, S.S., Shevade, S., Bhattacharyya, C. and Murthy, K.K.: Improvements to Platt’s SMO Algorithm for SVM Classifier Design, *Neural Computation*, Vol.13, No.3, pp.637–649 (2001).

- [28] Nassif, A.B., Shahin, I., Attili, I., Azzeh, M. and Shaalan, K.: Speech Recognition Using Deep Neural Networks: A Systematic Review, *IEEE Access*, Vol.7, pp.19143–19165 (2019).
- [29] Liu, G., Yu, Y., Mora, K.A.F. and Odobez, J.-M.: A Differential Approach for Gaze Estimation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, p.1 (2019).
- [30] Ishii, R., Kumano, S. and Otsuka, K.: Predicting Next Speaker Using Head Movement in Multi-party Meetings, *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp.2319–2323 (2015).
- [31] Ishii, R., Kumano, S. and Otsuka, K.: Analyzing mouth-opening transition pattern for predicting next speaker in multi-party meetings, *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp.209–216 (2016).
- [32] Ishii, R., Otsuka, K., Kumano, S. and Yamamoto, J.: Using respiration to predict who will speak next and when in multiparty meetings, *The ACM Trans. Interactive Intelligent Systems*, Vol.6, No.2, p.20 (2016).
- [33] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Analysis of Respiration for Prediction of Who Will Be Next Speaker and When? in Multi-Party Meetings, *Proc. International Conference on Multimodal Interaction*, pp.18–25 (2014).



石井 亮

2008 年東京農工大学大学院工学府情報工学専攻修士課程修了。同年日本電信電話株式会社入社。現在, NTT メディアインテリジェンス研究所主任研究員。2013 年京都大学大学院情報学研究科博士後期課程修了。2011～

2013 年成蹊大学客員研究員。2019 年よりカーネギーメロン大学客員研究員。コミュニケーション科学, マルチモーダルインタラクションの研究に従事。ACM 15th Int. Conf. Multimodal Interaction (ICMI2014) Outstanding Paper Award, 他受賞。電子情報通信学会, 人工知能学会, ヒューマンインタフェース学会各会員。博士 (情報学)。



大塚 和弘 (正会員)

1995 年横浜国立大学大学院工学研究科博士課程前期修了。2007 年名古屋大学大学院情報科学研究科博士後期課程修了。1995 年日本電信電話株式会社入社。以後, NTT ヒューマンインタフェース研究所, 2001 年 NTT コ

ミュニケーション科学基礎研究所, 2010 年スイス Idiap Research Institute 特別招聘研究員, NTT コミュニケーション科学基礎研究所特別研究員/主幹研究員等を経て, 2020 年 4 月より横浜国立大学大学院工学研究院知的構造の創生部門准教授。コミュニケーション科学, コンピュータビジョンに興味を持つ。ACM Int. Conf. Multimodal Interfaces (Interaction) Outstanding Paper Award (2007, 2012, 2014 年), 電子情報通信学会 H22 年度論文賞・喜安善市賞, 経産省 Innovative Technologies 特別賞 (2012 年), 他受賞。電子情報通信学会シニア会員, IEEE 各会員。博士 (情報科学)。



熊野 史朗

2009 年東京大学大学院情報理工学系研究科博士課程修了。同年日本電信電話株式会社入社。2016～2017 年 University College London での Honorary Research Associate を経て, 現在, NTT コミュニケーション科学基

礎研究所主任研究員, 2020 年より筑波大学システム情報系客員准教授。感情計算論 (Affective Computing) に関する研究に従事。Asian Conference on Computer Vision (ACCV2007) Honorable Mention Award, 他受賞。IEEE, 電子情報通信学会, AAAC 各会員。情報理工学博士。



東中 竜一郎 (正会員)

2001 年慶應義塾大学大学院政策・メディア研究科修士課程, 2008 年博士課程修了. 2001 年日本電信電話株式会社入社. 2020 年より名古屋大学大学院情報学研究科教授. NTT メディアインテリジェンス研究所・NTT コミュニケーション科学基礎研究所客員上席特別研究員. 2004 年から 2006 年まで英国シェフィールド大学客員研究員. 質問応答システム・対話システムの研究に従事. 著書に『質問応答システム』(コロナ社), 『人工知能プロジェクト「ロボットは東大に入れるか」: 第三次 AI ブームの到達点と限界』(東大出版), 『Python でつくる対話システム』(オーム社) 等. 人工知能学会, 言語処理学会, 電子情報通学会各会員. 博士 (学術).



青野 裕司 (正会員)

1999 年大阪大学大学院基礎工学研究科博士後期課程修了. 同年日本電信電話株式会社入社. NTT サイバースペース研究所, NTT アドバンステクノロジー, NTT 研究企画部門を経て, 現在, NTT メディアインテリジェンス研究所心理情報処理プロジェクトマネージャー (主幹研究員). 音声認識, 音声合成に関する研究に従事. 電子情報通学会会員. 工学博士.