

グループディスカッションにおける発言者の言語/非言語の表出と 評価者評価の関係の分析

武川 直樹^{†a)} 中山 知大^{††} 徳永 弘子[†] 大和 淳司^{†††}
山下 直美^{†††}

Analysis of Verbal / Nonverbal Expressions of Speakers and Evaluators' Evaluations in Group Discussions

Naoki MUKAWA^{†a)}, Tomohiro NAKAYAMA^{††}, Hiroko TOKUNAGA[†],
Junji YAMATO^{†††}, and Naomi YAMASHITA^{†††}

あらまし 近年、採用選考の場面でグループディスカッション（以下 GD）を課す企業が増えている。GD は、就職希望者の論理性、知識、協調性が同時に評価できるためと考えられる。しかし、就職希望者にとってこのような多面的な能力を獲得し、採用場面で発揮することは難しい。また、採用側にとっても能力を合理的に評価することが難しい。本研究は、採用選考時の GD において意見を伝える能力の評価に焦点をあて、評価者の主観的な評定値を正解とする従来方法の問題を検証するために、客観的に評価した論理性とパラ言語/非言語の多寡・評価者評定の関係を仮説に基づいて分析・検証して明確な合理性を導入することを目標にしている。そのため模擬的な GD を実施し、評価法を実験的に検証した結果、評価基準を構成する主要な評価項目、及びその関連性を明らかにした。また、学生、企業の両者にとって合理的な選考基準により論理性を評価するための指針を提案した。

キーワード グループディスカッション、言語/非言語、採用評価、評価基準、論理性、発話機能

1. ま え が き

就職活動の採用選考過程においては、しばしばグループディスカッション（以降 GD と称す）が導入されている。GD において企業の採用者は学生らに議論

テーマを与え、学生個人の振舞いが GD の場にどのように貢献し、またグループ全体として与えられたテーマに対してどのような結論を導くかを観察して、グループに貢献できる力を測るものである。採用選考における GD では議論の結論が評価されるよりも、ディスカッションの過程における論理の構成力、協調性、議論の進め方の統率力など多面的な要素が評価対象とされる。採用者側は、採用の選考過程における代表的な評価項目をこれらの多様な観点で評価し、総合的に高い評価を得た学生を次の選考ステップに進めているものと思われる。しかし、GD の評価の定義、評点の基準を合理的に記述することは難しい。その結果、採用される側の学生も具体的にどう評価されるか、自分のどのような能力を向上させておくべきかが、明確に意識されていないと考えられる。更に企業の採用担当採点者も GD が筆記テストと異なり、学生のふるまいをもとに印象評価に基づく評価となるため、採点者の価値観、能力、センスによって評価が左右され、評価の根拠、基準を明確にしたうえで評価しているとは必

[†] 東京電機大学システムデザイン工学部，東京都
Department of System Design and Technology, Tokyo Denki
University, 5 Senjuasahi-cho, Adachi-ku, Tokyo, 120-8551
Japan

^{††} 東京電機大学大学院情報環境学研究所，印西市
Information Environment Technology, Graduate school of
Tokyo Denki University, 2-1200 Muzai Gakuendai, Inzaishi,
270-1382 Japan

^{†††} 工学院大学情報学部，八王子市
Faculty of Informatics, Kogakuin University, 2665-1
Nakano, Hachioji-shi, 192-0015, Japan

^{††††} 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，京都府

NTT Communication Science Laboratories, NTT Corporation,
2-4 Hikaridai, Seika-cho, "Keihanna Science City",
Kyoto-fu, 619-0237 Japan

a) E-mail: mukawa@mail.dendai.ac.jp

DOI:10.14923/transinfj.2017HAP0016

ずしもいえない。

本研究では、特に GD における「自己表現」「意思疎通」などのスキルとして、発言に根拠があるか、一貫性があるのか、議論を明確に整理できているかなど意見を伝える力 [1] に着目する。既存の意見を伝える力の評価に関する研究は、人の個人特性を言語・非言語情報を用いてパターン化、モデル化、推定するアプローチのことが多い。これらの研究では、採点者の下す評価が究極の標準 (gold standard) であることを前提と考えている。しかしながら、採点者は、採用時の GD に立ち会い、同時に複数の学生に対して、短時間のうちに多数の項目の評価をしなければならない。採点者には評価項目それぞれを独立に評価することが規定されているにもかかわらず、ある評価項目の評価が他の項目の評価に無意識に影響を与えたり、目立ちやすい特徴に引きずられて評価項目の評価がゆがめられるハロー効果を生じたりすることがある。したがって、多数の項目を評価してその総合評点で採用判断をすることがよいとは限らない。

そこで、本研究では、採点者の下す評価を gold standard と考えず、被評価者の発言時の言語、非言語的表出に基づいて、意見を表出するスキルに関する、より合理的な基準から評価の構造を分析することを試みる。学生、企業の両者にとって合理的な選考基準を構築することにより、学生が自分のディスカッションの能力を向上させるための自己評価基準や、企業が学生を評価する評価基準を明確にできれば、採用に関わる学生、企業の両者にとって大きな寄与になると考えられる。なお、GD における意見表出スキル以外の、傾聴能力、協力的な行動スキル、議論のマネジメント能力も重要であるが、これらの評価は今後の課題とする。

2. 関連研究

GD 評価に関する研究として、対話者の言語・非言語情報をパターン化して多人数対話のメカニズムを解明するもの [2]～[4]、言語・非言語情報から発話者の個人特性の推定を試みたものがある。個人特性を推定する研究には、発話者の言動から個人の性格を推定するものや [5], [6]、GD での個人の役割を推定するもの [7]、コミュニケーション能力など個人の能力を評価するものが挙げられる [8], [9]。本研究は、発話者のコミュニケーション能力を評価する研究に位置づけられる。

近年、GD に限らず、発話者のコミュニケーション能力を評価する研究 [8]～[10]、更に評価を発話者にフィー

ドバックすることによってコミュニケーション能力の向上を目指す研究が活発に行われている [11], [12]。これらの研究は、就職活動の面接場面において採点官から高得点を得ることを目的としており、採点官の評価を gold standard として扱っている。例えば、岡田らは、マルチモーダル情報を用いて発話者のコミュニケーション能力に対し、採点官の評価値を gold standard として機械学習し、評価値推定モデルを構築している [8], [9]。しかし、採点官のつける評価が必ずしも正確であるとは限らない。特に GD では、ある発話者の発言が論理性に欠いていたとしても、自信をもって大きな声で発言し、他のメンバーの反応が良ければ、採点官は惑わされ、高く評価してしまう可能性がある。他のメンバーの発言内容を踏まえて自身の考えを論理的に展開できる能力を評価する場合には、他者の反応に惑わされずに、発話者の発話内容が根拠をもって論理的に表現されているか否かで評価されるべきである。

既存研究と異なり、本研究は、発話者が自身の考えを論理的に展開・整理できる能力 (コンセプチュアルスキル) の評価に焦点を当てる。そのため、GD における発話内容を書き起こし、文字情報のみから発話者の論理性を評価し、これを比較の標準 (reference standard) とする。その上で、採点官による評価値と reference standard との乖離を調査し、採点官が GD において、どのような言語・非言語情報の影響を受けやすいかを分析する。分析の結果、採点官がより精確に学生を評価するための指針を与えることが可能になり、評価される学生側には、よりその場に適した自己アピールをするための発言時のふるまい指針が得られることが期待できる。

3. アプローチと評価手順

3.1 アプローチ

本研究では、採点者の評価項目と被採点者の発言の表出の関係性、採点者評価の合理性を解明するために、GD における発言の言語的な論理構造や非言語的ふるまいの多寡などの客観的に評価可能な指標を構築し、その指標に基づき、採点者が評価する複数の評価項目と発言者の言語/非言語表出の対応関係を分析する。そのため、通常採用評価時に採用者が行う GD 全体を通しての採点ではなく、短いひとまとまりの発言場面を分析単位として言語的/非言語的ふるまいを分析対象とし、採点者評価の妥当性の有無を明らかにする。

まず、後述する発話機能の基準に照らして発言者の

発言の論理を評価し、その評価を基準に発言者の表出するパラ言語/非言語的なふるまいと採点者の評価の関係を分析する。論理構成が適切であっても非言語的表出が不適切であれば、発言者の内容は伝わらず、更に誤解される可能性がある。逆に内容が伴わなくても非言語的なふるまいに影響され、採点者が高評価を与えてしまう可能性がある。

そこで、本研究では以下の仮説を立て、分析する。

仮説 1: 発言の論理評価項目と発言者のパラ言語/非言語的なふるまい評価項目との相関は高い

仮説 2: 発言に論理性がなくてもパラ言語、非言語が多いと、評価者の「論理性」に関わる評価項目の評価点が向上する。

分析手順を以下に示す。

(1) 採用選考時を模擬した GD を実施し、映像を収録する。収録した会話映像を編集し、編集映像から発言者の発言を書き起こす。発言の論理性に着目するため、発言の要旨を抽出した要約テキストデータを作成し、主張のポイントを構成する「推論」、「憶測」の発言を抽出し、ラベルを付与する。

(2) 付与したラベルをもとに論理性あり/なしのシーン、また非言語的なふるまいの多寡に基づき、非言語動作の多い/少ないシーン、合計 4 シーンを選択する。更にそれらをテキストのみ、テキスト + 音声、テキスト + 音声 + 映像の 3 種類の評価用刺激データセットを作成する。

(3) 評価実験を実施する。有職者、有職経験者を実験協力者として論理性・説得性などに関わる項目、パラ言語/非言語表出の効果に関わる複数の項目に関して評価を行う。

(4) 評価者による多数の言語評価項目とパラ言語/非言語の評価項目の結果を因子分析して評価項目の主要因子を明らかにする。GD の論理性に関わる評価因子に対して、採点者がどのような言語・非言語情報の影響を受けやすいかを明らかにする。続いて構造方程式モデリング (SEM) により評価項目間の因果関係を分析する。言語の評価とパラ言語/非言語評価の因果関係、論理と相関の高い非言語評価項目、評価項目に影響を及ぼす非言語情報を明らかにする。

3.2 GD の会話映像の収録

採用選考場面を想定して GD を実施した。参加した学生は全 15 名で、4 名による GD を入れ替わりながら 1 セッション 15 分の GD を 6 セッション実施した (一部の学生はセッションに複数回参加している)。ま



図 1 分析用編集映像

Fig. 1 A scene from edited video for analysis.

た、各セッションにおいて採用を判断する採点担当の教員 3 名が同席し、セッション終了後に各参加者を評価した。なお、議論の開始時に議論テーマを指示し、事前の個人検討時間を 5 分与えた後、議論を開始し、15 分以内に 4 人で一定の結論に達するよう議論を進めるよう指示、最後に 1 分程度で結論を報告させた。また、議論に先立ち、採用試験時の GD の重要性、目的、目標について 10 分間のレクチャを行った。このように GD 実際の採用時の GD を模擬するものとして妥当な環境を構築している。また、会話は映像収録するために 4 人の参加者の正面にそれぞれ 1 台のカメラ、俯瞰用に 2 台のカメラを設置した。音声は各参加者にピンマイクを装着して収録した。各セッションの会話テーマは、「成人年齢の 18 歳化に賛成か反対か?」「就職活動における会社選びのポイントはなにか?」「10 年後の携帯電話はどうなっているか?」「成人年齢の 18 歳化に賛成か反対か?」「就職活動における会社選びのポイントはなにか?」「10 年後の家電業界はどうなっているか?」である。参加者にはディスカッション中の会話内容、及び顔写真が学会などで公表されることを説明し、同意書に承諾を得た。会話収録後、謝金を支払った。

3.3 発話の書き起こし

収録した映像を映像編集ソフト Adobe Premiere Pro により図 1 に示す配置に編集し、得られた会話映像をアノテーションソフト ELAN [13] を用いて参加者全員の発話を時系列に沿って書き起こした。0.3sec 以上の無音区間を発話の区切りとする間休止単位 [14] で発話を書き起こした。参与役割は、当該の参加者が他の参加者らに一定の発話権を承認されている参加者を「話し手」とし、それ以外の参加者を「聞き手」と認定した。

表 1 発言の要約テキストとラベル/タグ付与の例
Table 1 Examples of summarized text and its speech function labels and evaluation tags.

(a)論理的と評価された発話の例 (a) An example of logical utterances		
要約テキスト	機能ラベル	評価タグ
高校卒業後，専門学校や大学に進んでいく	一般論	
しかし高校卒業までだと年上の人の意見が肌で感じられない	一般論	
それに対し上の学校にいくことで自分より年上の人の意見に触れる機会がある	一般論	
したがって高校生を卒業して成人するのは経験不足	推論	根拠明確
(b)非論理的と評価された発話の例 (b) An example of illogical utterances		
要約テキスト	機能ラベル	評価タグ
大学生になるにしても夢に向かって志を持っている人は少ない	憶測	
18になって社会に出たり，大学に入ったからといってすぐに意識が変わるとは思えない	憶測*	
ある程度猶予期間が必要	推論	根拠曖昧*

* 当初2名が一致し3名が合議の上決定した

3.4 発言内容の要約テキスト化

書き起こされた発言に対してフィラーや副詞などを削ぎ落とし要約テキスト作成する。具体的には、「あの一、」や「えーと」などのフィラー，意味的に余分な「～について」などの発言部分を除去する。その結果，表1のように論理構造が明確な要約テキストを得る。

3.5 発言の論理構造化

発言者の論理性を評価する reference standard とする発話を抽出することを目的に，発話機能ラベル [15] を用いて発言者の主張における論理的構造を分析する。ここでは，論理的主張に関わる「意見」の表出における「推論」「憶測」の機能ラベルを上記要約テキスト中に付与する。「推論」は「一般論」「例示」「体験」やそれまでの「推論」に基づき，発言者にとって自身の考えを，根拠をもとにアピールする重要な発言であり，論理性を評価するために有効であると考えられる。一方，

表 2 2 グループの主張に付与された根拠評価結果
Table 2 Number of tags of with/without evidence.

タグ	タグの個数
推論の根拠明確	5
推論の根拠曖昧	4
推論の根拠なし	5
憶測	8
評価者の不一致	6

「憶測」は，根拠があいまいなまま考えを述べるものである。つづいて，「推論」「憶測」ラベルについて van Emmeren の分類 [16] に基づいて，表2のように「根拠明確」「根拠曖昧」「根拠なし」の3段階で評価する。

以上の評価ルールのもとで，6 グループの GD から一人が発言し続ける（複数発話が連続する）話者交替を含まない 14 シーンを予備的に抽出し，発言者の意見主張場面の要約テキストデータから，「推論」「憶測」ラベル，及び根拠の3段階評価を GD に関与しない第3者のラベラー3名により付与した。始めにラベラーは評価対象の発言の1～2分前からの音声聞いて議論の文脈を理解する。音声は評価すべき発言の直前になると合図が鳴るように作成されている。評価シートは，実際の発言テキスト，要約テキストが記入されている。ラベラーは評価対象の発言に「推論」「憶測」ラベルが付与されるかを判断し，「推論」ラベルを付与した発言には更に，「根拠明確」，「根拠曖昧」，「根拠なし」のいずれかの評価を記入する。表1の右にラベラーの付与したラベル，タグを示す。

「推論」「憶測」ラベルと発言の根拠の明確さ評価の結果を表2に示す。「根拠明確」の「推論」は5，「根拠曖昧」「根拠なし」の「推論」と「憶測」の合計は18であった。すなわち発話機能ラベルから形式的には論理展開されているとみなせるが，実際には根拠がなく，正当性が乏しい発言が多いことがわかる。

4. 発言者の発言の論理性とふるまいの分析

4.1 評価テストの目的

前節で検討した「推論」ラベルとその根拠の評点は，発言者の主張の論理性についての客観的な reference standard として利用できると考えられる。この基準で学生の個々の発言を評価した結果，発言には推論の根拠に乏しいものが多く含まれていることから，本基準は実用性もあると考えられる。一方，この基準は，現状の採用の場において，採点者が論理性判断のために

積極的に使っているとはいえない。採点者の印象評価にゆだねられている評価結果は、GD 参加者の表出する非言語行動の多寡、あるいは良否により影響され、合理的でない可能性が考えられる。そこで、以下では、前節で得られた推論の根拠評価を reference standard とし、GD 参加者のパラ言語/非言語表出と評価者の論理評価との関連を分析し、どの非言語的ふるまいが論理評価に影響を与えるかを明らかにする。

4.2 評価用刺激の作成

収録した全 GD 映像から論理の有無、非言語的ふるまいの多寡に基づき典型的な発言シーンを 4 シーン抽出する。具体的には、言語の論理の有無に基づく(「論/非論」と、「非言語行動」の多寡(「動/静」)の観点から「論&動」「論&静」「非論&動」「非論&静」を選定した。発言シーンは 25~40 秒程度の長さである。シーンの選択に当たって、論理の有無については前節の根拠評価を元に選定した。非言語行動の多寡については発言中の視線先が向けられていた方向(人か人以外かの注視時間)や人へのアドレッシングの回数、出現した例示的ジェスチャ(発言内容の説明を支援する:多いと積極的)、適応的ジェスチャ(緊張などの自己接触に見られる:多いと消極的)などの回数を基準に総合的に判断した。詳細を表 3 に示す。

なお、選択された 4 シーンの論理の根拠評価タグのコーダー間一致率は 0.83 であった。不一致のシーンについてはコーダー間で話し合い、全員合意のもと最終的なタグを決定した。また、4 シーンは、一人の複数発言が連続する話者交替を含まないものを選択しているが、前の発言者の発言の論理の誤りを指摘する次話者の発言、訂正する発言など、発言者をまたがる発言についての分析は今後の課題である。

次に、仮説 2 に関連して、発言に論理性がなくてもパラ言語、非言語が伝わることにより評価者の「論理性」に関わる評価点に変化することを調べるため、選定した各発言シーンについて、情報メディアが異なる

3 種の評価用刺激、「要点テキスト」、「音声」、「映像」を作成した。「要点テキスト」は、発言内容の要点や論理などの意味的な部分に対する評価が行える。「音声」の評価は、「要点テキスト」で評価した意味の評価に加え、言い淀み、間、フィラーなどのパラ言語に対して評価が行われる。更に、「映像」では「音声」で観察できる行動に加え、視線や顔の向き、ジェスチャなどの非言語行動に対して評価が行われる。映像刺激は、発言者だけを写したバストショット映像である。フィラーや言い淀み、視線移動、ジェスチャは、発言者の意見の主張をより効果的に印象づける役割を果たし、採点者の評価に影響を与えると考えられる。

4.3 発話・ふるまいの評価者評価の手順

評価者には、著者らの所属する大学に通う学生の父母(有職者)14 人に依頼した。評価にあたって、評価者には、「企業で新入社員の採用に関わっており、来年度より一緒に働く新入社員を選ぶ」立場であると教示した。

評価は、以下の 15 項目を 5 段階にて行った。まず、言語的スキルの項目は、①理解:話者である学生の言いたいことは理解できる、②納得:学生の言っていることに納得できる、③相応:学生の言っていることは議論のテーマとしてふさわしいと思う、④明瞭性:学生の発言の内容は要点が明瞭である、⑤根拠:学生の発言内容には根拠(具体例、一般論など)がある、⑥順序:学生の発言内容は順序立てられて説明されている、⑦説得力:学生の発言には説得力がある、⑧表現力:学生の発言に表現力がある、を用意した。また、音声、映像を含むメディアの評価については、磯ら[17]の研究を参考に、⑨流暢性:学生の話し方は流暢である、⑩活力:学生の話し方に活力がある、⑪言い淀み:学生は言いよどみしていた、⑫話の間:学生の話の間のとりかたはうまい、更に、⑬視線:学生の視線が聞き手に向いている、⑭自信:学生の表情に自信がある、⑮仕草:学生の仕草が気になる、により評価する。提示の順番効果を排除するため、評価刺激の提示順はカウンターバランスをとった。

5. 結 果

5.1 評価項目の因子分析

発言者の言語情報における論理性表現と非言語情報表現の違いに注目し、言語情報の論理性と非言語情報の多寡の組み合わせによる「論&動」「論&静」「非論&動」「非論&静」の四つの GD シーンを合わせ、映像

表 3 各刺激の非言語行動回数/時間率
Table 3 Number/rate of nonverbal behaviors in stimuli.

非言語行動	非論 & 静	非論 & 動	論 & 静	論 & 動
聞き手を見る時間率 (%)	25	54	4	22
聞き手を見る回数 (回)	7	7	2	13
ジェスチャー (回)	0	3	0	2
	1	0	2	0

刺激の15項目の評価結果に対して因子分析（最ゆう法，バリマックス回転）を行った。

その結果，固有値1以上の因子が三つ認められた。固有値は，第1因子から順に8.2，1.9，1.1であった。累積寄与率は第3因子までで75.1%であった。第1因子は，⑩活力「学生の話し方に活力がある(0.90)」，⑭自信「学生の表情に自信がある(0.72)」，⑨流暢性(0.72)，⑬視線(0.70)が高い因子負荷を示しており，活力のあるふるまいに関連した内容の項目群といえる（ただし，かっこの数値は因子負荷量である。以下も同様）。そこで第1因子を「自信・活力」の因子と解釈する。第2因子は，⑤根拠「学生の発言内容には根拠がある(0.88)」，⑥順序「学生の発言内容は順序立てられて説明されている(0.80)」，⑦説得力「学生の発言には説得力がある(0.71)」，②納得「学生の言っていることに納得できる(0.69)」が高い因子負荷を示しており，「根拠・説得力」の因子と解釈する。この因子は，テキスト・音声・映像に共通する①から⑧までの項目に高い因子負荷量をもっているため，テキスト・音声・映像の刺激の違いに関し，第2因子を用いて「論理性」の評価を行うことは妥当であると考えられる。第3因子は，⑪言い淀み：学生は言いよどみしていた(-0.99) ⑮仕草「学生の仕草は気になった(-0.75)」の2項目が高い因子負荷を示しており，「パラ言語/非言語表現力」の因子と解釈する。以上のように，15の評価項目は3因子として解釈できた。

次に，各因子の因子負荷量を見ると，第1因子「自信・活力」，第3因子「パラ言語/非言語表現力」ではパラ言語/非言語に関わる評価項目の負荷量が高く，第2因子「根拠・説得力」では言語的な評価項目の負荷量が高い。第1因子と第2因子の相関は低いと考えられるから，仮説1：発言の論理評価項目と発言者のパラ言語/非言語的ふるまい評価項目との相関は高い，は支持されなかった。すなわち，自信や活力など発言者の積極的なふるまいが多いからといって評価者が，発言を根拠や説得力があると評価するとは限らない，といえる。

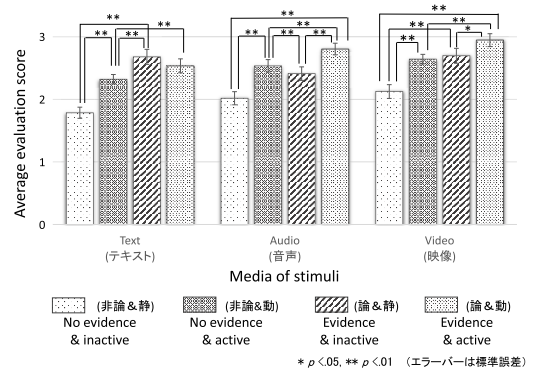
5.2 発言者の言語/非言語表現とテキスト・音声・映像のメディアによる論理性伝達

参加者の発言の論理性と非言語情報の多寡に基づいて「論&動」「論&静」「非論&動」「非論&静」の4水準を用意し，それらに伝達メディアの要因（要点テキスト・音声・映像の3水準）を組み合わせた12刺激を評価者に評価させることにより，因子「根拠・説得力」

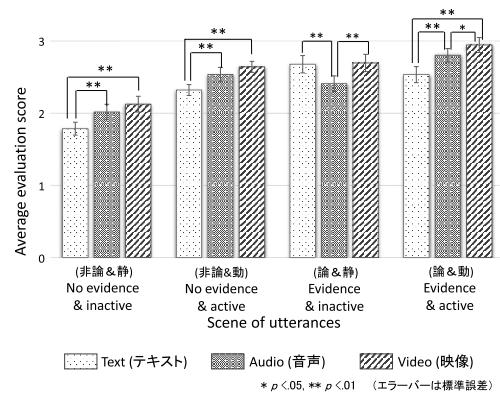
の得点を分散分析した。ここで「根拠・説得力」の因子の因子負荷量が65%以上の四つの評価項目を選択し，それらの評点が「根拠・説得力」を構成するとみなす。その結果，1刺激に対する評価データ数は14人×4項目で56となる。

2要因被験者内分散分析を行った。始めに，F値を求めて有意差検定を行った結果，伝達メディアの主効果，発言者の表現（論理性表現と非言語情報表出の多寡）の主効果，が有意であった，伝達メディア： $F(2, 55) = 14.77$ ， $p < 0.01$ ，発言者の表現： $F(3, 55) = 40.01$ ， $p < 0.01$ 。交互作用： $F(6, 55) = 3.12$ ， $p < 0.01$ 。交互作用について下位検定を行った結果，2×3の全ての組み合わせで有意であった。

多重比較検定を行った結果を図2に示す。評価者が映像を見る場合には「非論&静」<「非論&動」，「非論



(a) 論理／非言語表現条件間
(a) Between logical/nonverbal conditions



(b) メディア条件間
(b) Between media conditions

図2 因子2「根拠・説得力」の評価得点
Fig.2 Evaluation scores of factor 2 “evidence and persuasiveness”.

&静」<「論&静」,「非論&静」<「論&動」で有意差があった ($*p < .05$, $\alpha' = 0.0083$). すなわち, パラ言語/非言語表出が多い発言者は, 発言内容の根拠があいまいであっても「根拠・説得力」が高く評価されることが分かった.

次に, 提示メディアの違いに着目する. 「論&静」の発言に対しては伝達メディア (要点テキスト・音声・映像) が変わっても「根拠・説得力」の評価には有意差がなかった. 一方, 「非論&静」においては, 「テキスト」<「音声」, 「テキスト」<「映像」が有意であった. 「非論&動」においても, 「テキスト」<「映像」が有意であった. 同様に「論&動」においても「テキスト」<「音声」, 「テキスト」<「映像」であった. すなわち「論&静」以外の発言者の表現に関しては, テキストよりも音声, 映像を見て評価するほうが高い評価を得ていた. 以上の結果から, 「仮説 2: 発言に論理性がなくてもパラ言語, 非言語が多いと「論理性」に関わる評価項目の評価点が向上する.」は支持された.

6. 考 察

6.1 評価項目の因子分析

評価実験において評価者に 15 項目の評価項目を用意して評価したが, 評価点に基づく因子分析では 3 因子, 「自信・活力」「根拠・説得力」「パラ言語/非言語表現力」のみが抽出された. 日常的に使われる言い方にすれば, 発言が論理的で話しぶり・身振りの表現力があり, 元気と自信があることが高く評価されるという妥当な結果と考える.

6.2 発言者の表現とテキスト・音声・映像のメディアによる論理性伝達

発言者の表現の要因と伝達メディアの要因による分散分析の結果, 実際の採用時の GD の場に近いと考えられる映像を見て評価する場合には, パラ言語/非言語表出が多い発言者の方が, 「根拠・説得力」が高く評価されることが分かった. 論理は, 要点テキストにまとめられている方が, 話し方や身振りに惑わされることなく合理的な評価が可能であると考えられるが, 評価者は, パラ言語/非言語表出が多い発言者を高く評価する傾向があることが分かった. この結果は, 採用者の評価を Gold standard として盲目的に採用することに問題があることを示唆している.

次に, 発言者が論理的であってもパラ言語/非言語表出が不十分な場合 (「論&静」) には, 伝達メディア (要点テキスト・音声・映像) が変わっても「根拠・説

得力」の評価には有意差がなく, 評価者は映像を見ても論理が向上したとは評価しない一方, 「非論&動」の発言に対しては, 伝達メディアが音声, 映像になるにつれて論理的になると評価される. 映像により論理が改善することはないにもかかわらず, 評価者は映像から発言者のふるまいをみて論理性があるように感じてしまうと考えられる. 例えば, 自信のある表情をして発言しているので, 言語的に内容が伴わない発言でも説得力があると思ってしまう, 言いよどみがないので主張に根拠があるように感じてしまう可能性がある. 更に, 評価者がある部分を高く評価すると直接関係のない他の部分の評価も高くなるハレーション効果も考えられる. したがって, 学生の論理を合理的に評価したい場合には, 学生を見ずに論理を確認する評価も必要になるかもしれない.

6.3 SEM による評価項目の構造分析

評価結果から, 評価者は, 必ずしも reference standard に従って「根拠」を評価するのではないことが示唆され, 評価者評価がどのような構造となっているかを明らかにすることが課題として残った. そこで, 更に構造方程式モデリング (SEM) により評価項目の因果関係について検討し, 評価者評価の構造の解析を試みる [18]. 例えば, 「説得力を感じたので恐らく根拠があるはずである」などと評価した可能性を明らかにすることができると考えられる.

15 の評価項目に対して抽出された三つの因子に対して最も大きな負荷量をもつ項目の一つずつ選択し, これらの関係について分析する. 論理性に関わる評価項目のうち最も総合的評価に近い意味合いをもつものとして「⑦説得力 (persuasiveness)」を選択し, SEM によるパス解析で 3 因子から選んだ代表的な項目「⑤根拠 (evidence)」, 「⑩活力 (energetic)」, 「⑮仕草 (gesture)」との関係を分析する. 図 3 に示すとおり, 「説得力 (persuasiveness)」には「根拠 (evidence)」の係数が 0.68 となり最大の影響をもっている. また図中右端の○は残差分を表す. データ数が十分ではなく残差も 0.30 と大きいため, モデル適合度は十分に高いとは言えず, 有意なパスは根拠から説得力を説明するパスのみであったが, 意味内容的には適切な関係性が示唆された.

また, 四つのシーンを「動&静」, 「非論&論」の 2 種類の分割法でそれぞれ 2 グループに分割して同様の分析を行った. 図 3 と同様のパス図が四つ描かれるがその係数のみをまとめた結果を表 4 に示す, 「静」よりも

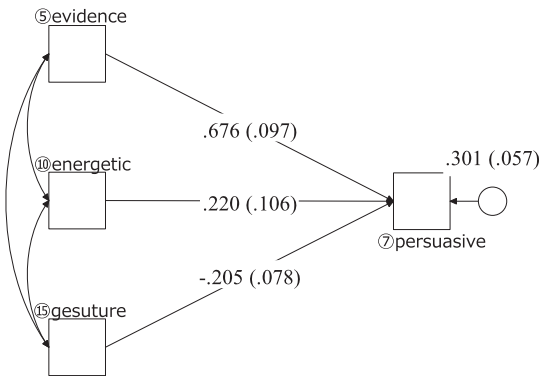


図 3 各要因の主な項目と説得力の関係 () 内は標準誤差
Fig. 3 Effects on “Persuasiveness” from dominant items of each factor. Numbers in parentheses are Standard Error.

表 4 シーンの特徴と説得力への係数比較
Table 4 Persuasiveness from evidence and scene conditions.

説得力への係数	比較 1		比較 2	
	動	静	非論理	論理
⑤ 根拠	0.75	0.536	0.042	0.364
⑩ 活力	0.206	0.439	-0.354	0.641
⑮ 仕草	-0.255	-0.243	0.344	-0.347

「動」の方が「根拠」から「説得力」へのパスの係数が大きく、また、「非論」よりも「論」の方が「根拠」から「説得力」へのパスの係数が大きい結果が得られた。これらもその意味内容的に妥当なものと考えられる。

次に、「根拠」の有無が適切に評価されているかどうかを他の項目との関係性から分析した。15 項目のうち客観的計測項目として捉えやすいと考えられ、かつ「根拠」への影響の大きいものとして「明瞭」、「順序」、「話の間」を取り上げる。話の順序が根拠の有無の評価に影響するのは合理的であるが、一方、発話の明瞭性や話の間のとり方は話者の表現上のスキルであり、本来、根拠の有無に直接影響するとは考えられない。それにもかかわらず評価者が影響を受けている可能性がある。そこで仮説として「評価者は説得力があると感じた (= 自分が説得された) 発話については根拠があると感じる」を設定し検証する。これは根拠があるから説得力を感じるという因果と異なり、合理的とはいえないが、評価者が発言者の表現上の特性に依存して説得力を感じ、納得した場合に、その原因を発話に根拠があったからに違いない、と妥当といえない合理化をしてしまうことは起きうることだと考えられる。

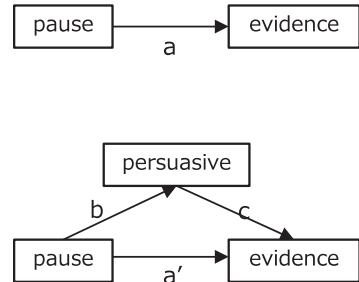


図 4 媒介分析 (「話の間」と根拠の関係を例示)
Fig. 4 Mediation analysis (Example of “evidence” on “pause” case).

表 5 媒介分析の結果
Table 5 Results of mediation analysis.

項目名	直接	間接			間接/直接
	a	b	c	a'	b c / a
④ 明瞭	0.69	0.60	0.50	0.39	0.43
⑥ 順序	0.97	0.84	0.19	0.81	0.17
⑫ 話の間	0.52	0.59	0.66	0.12	0.76

検証のため、ここで取り上げた「明瞭」、「順序」、「話の間 (pause)」の 3 項目から「根拠 (evidence)」への直接の影響と、「説得力 (persuasiveness)」を媒介とした影響とを比較する (図 4)。直接効果を想定した場合のパス係数 a と、間接効果を想定した場合のパス係数 b, c, 及びこのときの直接効果 a' を表 5 にまとめた。

表 5 の「話の間」から「根拠」への効果は「説得力」を媒介とした成分が大きく貢献していることがわかる。「順序」については、間接効果が直接効果に比べて小さく、「明瞭」は両者の中間であることがわかる。「話の間」や「明瞭性」が「根拠」の有無に影響しているのは「説得力」を介した間接的な効果であることが確認できた。

以上の SEM による分析結果は、それぞれの効果の大小順序を含め、各項目の意味内容から考えて合理的に説明できるものとする。これは、評価者が「説得力を感じたが、それはきっと根拠があるからだろうからに相違ない」、という本来の因果関係を逆転させた思考を無意識に行った結果であるとの解釈が可能である。これは評価項目の選択、網羅性、合理性、更に、評価者への評価法の指示、ひいては個別要素の評価に基づく全体の総合評価という評価スキーム自体の妥当性についての課題を提起していると考えられる。

7. む す び

本研究は、採用選考時のグループディスカッションに焦点を当て、GD 参加者が発言するときに表出する言語、非言語的ふるまいに基づく、合理的な評価基準を明確にすることを目的とし、実験的に検証した。

採点者の下す評価ではなく、文字情報のみから発言者の論理性を評価し、これを reference standard とした。その上で、採点者による評価値と reference standard との乖離を調査し、採点官が GD において、どのような言語・非言語情報の影響を受けやすいかを分析した。

評価実験の結果、評価者の「論理性」の評価は、言語から解釈できる論理性だけでなく、発言者のジェスチャー、話の間の使い方などに大きな影響を受けることが明らかになった。テキストよりも音声、映像とメディアが伝達する情報が増えるに従い、論理性の評価が高くなることもこれを説明しているものと考えられる。媒介分析の結果、評価者は、必ずしも「根拠があるから説得力がある」という因果関係に従って評価するのではなく、「説得力を感じたので恐らく根拠があるはずである」と発言者の論理性を評価した可能性が示唆された。

本研究の結果、これまでの評価項目による評価基準には合理性の観点からは様々なバイアス（メディア依存性、シーン特性依存性）があり問題があることが明らかになった。論理性の観点からは、「根拠」と「説得力」の評価を合理的な妥当性をもって評価する基準の開発が必要である。一方、この種のバイアスは実際の業務におけるコミュニケーションシーンにも同様に存在するものと考えられる。例えば、入社後に気持ちよく一緒に仕事をしたい人物を選ぶ、という観点からはバイアスの存在自体は問題ではなく、バイアスの存在を前提として、評価側も被評価側も GD における評価の目的と意味、その有効性の範囲を認識すべきであろう。そのうえで、学生、企業の両者にとって合理的で納得できる評価基準を構築することにより、学生が自分のディスカッションの能力を向上させるための自己評価基準や、企業が学生を評価する評価基準を明確化する必要がある。

また、本研究では、「発話者が自身の考えを論理的に展開する『意見を伝える能力』の評価に焦点を当てたが、今後は、「聞く力・協調・尊重」「本質を理解する力」に対する合理的な評価法の確立を課題としたい。

謝辞 本研究の一部は科学研究費助成事業基盤研究(C)26330233 による助成を受けた。

文 献

- [1] “若年者就職基礎能力修得のための目安委員会報告書,” July 2004. <http://www.mhlw.go.jp/houdou/2004/07/dl/h0723-4h.pdf>
- [2] L. Chen and M.P. Harper, “Multimodal floor control shift detection,” Proc. ACM ICMI, pp.79–86, Cambridge, MA, USA, Nov. 2009.
- [3] 中田篤志, 角 康之, 西田豊明, “非言語行動の出現パターンによる会話構造抽出,” 信学論 (D), vol.J94-D, no.1, pp.113–123, Jan. 2011.
- [4] R. Ishii, K. Otsuka, S. Kumano, M. Matsuda, and J. Yamato, “Predicting next speaker and timing from gaze transition patterns in multi-party meetings,” Proc. ICMI, pp.79–86, Sydney, Australia, Dec. 2013.
- [5] O. Aran and D. Gatica-Perez, “One of a kind: Interring personality impressions in meetings,” Proc. ICMI, pp.11–18, Sydney, Australia, Dec. 2013.
- [6] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro, “Multimodal recognition of personality traits in social interactions,” Proc. ICMI, pp.53–60, Chania, Crete, Greece, Oct. 2008.
- [7] D. Sanchez-Cortes, O. Aran, M.S. Mast, and D. Gatica-Perez, “A nonverbal behavior approach to identify emergent leaders in small groups,” IEEE Trans. Multimedia, vol.14, no.3, pp.816–832, June 2012.
- [8] 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, AI30-E, Nov. 2016.
- [9] S. Okada, Y. Ohtake, Y. Nakano, Y. Hayashi, H. Huang, Y. Takase, and K. Nitta, “Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets,” Proc. ICMI, pp.169–176, Tokyo, Japan, Nov. 2016.
- [10] L. Nguyen, D. Fraendorfer, M. Mast, D. Gatica-Perez, “Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior,” IEEE Trans. Multimedia, vol.16, no.4, pp.1018–1031, June 2014.
- [11] H. Tanaka, S. Sakti, G. Neubig, T. Toda, H. Negoro, H. Iwasaki, and S. Nakamura, “Automated social skills trainer,” Proc. IUI, pp.17–27, Atlanta, Georgia, USA, March 2015.
- [12] E. Hoque, M. Courceon, B. Mutlu, C. Martin, and W. Picard, “MACH: My automated conversation coach,” Proc. UbiComp, pp.697–706, Zurich, Switzerland, Sept. 2013.
- [13] <https://tla.mpi.nl/tools/tla-tools/elan/>
- [14] 坊農真弓, 高梨克也 (共編) / 人工知能学会 (編) 知の科学 多人数インタラクションの分析手法, pp.153–171, オーム社, 東京, 2009.

- [15] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人, “発話単位タグ標準化案の作成,” 人工知能学会誌, vol.14, no.2, pp.251–260, March 1999.
- [16] F.H. van Eemeren and R. Grootendorst, “A systematic theory of argumentation,” The Pragmadiological Approach, Cambridge University Press, Cambridge, 2004.
- [17] 磯友輝子, 木村昌紀, 桜木亜季子, 大坊郁夫, “発話中のうなずきが印象形成に及ぼす影響—3 者間会話場面における非言語行動の果たす役割,” 信学技報, HCS2003-25, Oct. 2003.
- [18] 小杉考司, 清水裕士, M-plus と R による構造方程式モデリング入門, 北大路書房, 2014.

(平成 29 年 5 月 1 日受付, 9 月 5 日再受付,
11 月 6 日早期公開)



武川 直樹 (正員：フェロー)

1974 早大・理工・電子通信卒. 1976 同大大学院修士課程了. 同年日本電信電話公社 (現 NTT) 入社. NTT データ, NTT コミュニケーション科学基礎研究所を経て, 2003 東京電機大学. 画像符号化, 画像処理, 画像認識, ヒューマンインタフェースの研究開発に従事. 教授. 工博. 2012 本会ヒューマンコミュニケーショングループ運営委員長. IEEE, ACM, 情報処理学会, 人工知能学会, 日本顔学会各会員.



中山 知大

2015 年東京電機大・情報環境学部卒, 2017 同大大学院修士課程修了. 同年株式会社アイティフォーに入社. 在学中にグループディスカッションにおける言語・非言語の表出と印象評価の研究に従事.



徳永 弘子

1987 年昭和女子大学家政学部生活美学科卒. 2006 年よりヒューマンコミュニケーションの研究に従事. 現在東京電機大学システムデザイン工学部研究員. 博士 (工学). 日本認知科学会, 日本食生活学会会員.



大和 淳司 (正員：フェロー)

1988 年東大・工・精密機械卒. 1990 同大大学院修士課程了. 同 NTT 入社. 1998 MIT 電気工学及びコンピュータ科学科修士了. NTT ヒューマンインタフェース研究所, NTT コミュニケーション科学基礎研究所等を経て, 2016 工学院大学情報学部教授. 画像認識, インタクション等の研究開発に従事. 2010 年度本会論文賞, 喜安善市賞受賞. IEEE, ACM, 人工知能学会各会員.



山下 直美

1999 年京都大学工学部情報工学科卒業. 2001 年京都大学大学院情報学研究科数理工学専攻修士課程修了. 同年, 日本電信電話 (株) コミュニケーション科学基礎研究所入所. 博士 (情報学). CSCW, HCI の研究に従事.