

自閉スペクトラム症者のソーシャルスキル指導のための人工社会知能の構築と応用

研究代表者	大塚和弘	横浜国立大学・大学院工学研究院・教授
共同研究者	神山 努	横浜国立大学 教育学部・准教授
研究協力者	望月大輝	横浜国立大学・大学院理工学府・大学院生

1 はじめに

本研究は、子どもや障がい者の「対話的・協働的な学び」を支援するため、学びの場で生じる社会的相互作用を認識・理解可能な「人工社会知能(ASI: Artificial Social Intelligence)」を構築し、その解析結果を活用することで、個別最適な学びの場の設計原理を見出すことを目的として計画された。社会において、人は他者との協働や関係性を構築するため、他者の思考や感情、行動とその背後にある動機を理解するといった「社会的知性」を発揮する。人工社会知能は、そのような人の能力を模したものであり、その構築と活用のため本研究では、人の社会的知性が創発・育成される「対話的・協働的な学びの場」に焦点を当てた。このような学びの場においては、指導者と学習者の間、及び、学習者間において多様な社会的相互作用が生じる。本研究では、人工社会知能を用いてそのような社会的相互作用の解析を試みた。加えて、その結果に基づき、学びの質や効果の向上のため、学習者一人ひとりに最適化された教育指導法の設計の指針を得ることを長期的ゴールとして設定した。

本研究では、このような社会的相互作用が大きく関与する対話型の教育場面として、特に自閉スペクトラム症(ASD)児童に対する「ソーシャルスキル指導(SST)」に焦点を当てた。ASD は、社会的相互作用における非言語コミュニケーション行動の欠損などにより特徴付けられる発達障がいであるが、個々人によってその障がいの実態は様々である。そのため、定量的に障がいの程度を把握することは困難であり、最適な個別指導を計画することが難しい。そこで本研究では、SST において、各人の言語・非言語行動を画像・音声データとして収録し、人工社会知能を用いて、学習者・指導者間、及び、学習者間で生じる社会的相互作用を解析し、障がいの程度や指導の効果を把握するというアプローチを策定した。また、この人工社会知能による分析結果を利用することで、ASD 者の実態に応じて、より精度の高い個別指導計画を立案し、実践するための方法論を見出すことも長期的な目標とした。

1 年目は、非言語モダリティの中でも特に頭部運動に着目し、指導者の頭部運動を用いた働きかけに対して児童側がどのように応答するか、指導者・児童間において現出する機能の組とその応答時間の分析を実施し、ASD 児と TD 児の間には顕著な差異が見られることを確認した。この分析に用いるデータを収集するため、まず、ASD 児を対象とした実験的なソーシャルスキル指導 (SST) の状況を収録した。具体的には、ASD 児として、小学生の男児 2 名、女児 1 名を対象とし、1 回 10 分の SST のセッションを合計 19 回、一定期間において実施した。また、定型発達児 (TD 児) として、小学生の女児 1 名を対象として、同様の SST セッションを合計 3 回実施し、収録を行った。続いて収録された映像データより、畳み込みニューラルネット (CNN) を用いて、頭部の動きがもつコミュニケーション機能 10 種 (話し掛け、強調、反応確認、思考、相槌、応答、理解、肯定、正の感情表出) それぞれの有無を時刻毎に検出した。その後、指導者の頭部運動機能が生じたとき、児童側の頭部運動機能が生じるまでの反応時間と反応の頻度を、指導者・児童間での各機能の対において求めた。結果、2 名の ASD 児は、TD 児よりも反応が遅い機能組が多く見られた。反対に 1 名の ASD 児は、TD 児よりも反応が早い機能組が多く確認された。このような結果より、様々な社会的相互作用の種別とその量、反応時間という観点から、ASD 児毎の特性に紐づいた分析と考察が可能であることが分かった。

前年度の研究に引き続き、1 年間の延長となる 2 年目の成果について以下に概要を述べる。この 2 年目の取り組みとしては、前述の研究を 2 つの観点から拡張を行った。一つ目は、非言語モダリティを、頭部運動と顔表情の 2 モダリティに拡張した点で、これらモダリティ間の連携による幅広い非言語機能を分析の対象とした。2 つ目は、「相乗機能スペクトラム解析」(sFSA)を導入した点にある。sFSA は、非言語行動の特徴である機能多重性と解釈曖昧性を捉えることを主眼として、複数モダリティに跨る複数機能の知覚強度分布を表現した「機能スペクトラム」を入力とし、その因子分解により、対話において生じる主要な機能群を発見することができる分析枠組みである。

以下、2. にて研究の背景を述べ、その後、3. にて研究の方法と結果について詳細を述べる。

2 研究の背景

近年、教育分野において、持続可能な社会の創り手を育成するため、教師が児童生徒へ知識を授ける従来型の学習法から、児童・生徒の主体性を引き出し、児童・生徒間や児童・生徒・教師間の対話や協働を介した学び方への変革が強く求められている。また、情報通信技術を活用することで、場所や時間を問わず、より効率的な学習が促進できることが強く望まれている。これまで教育心理学の分野において、発達障害者の社会性支援のため、様々な行動分析法が開発・実践されており、個別最適な学習支援法が開発されてきた[Nagayama2023, 丹治 2021]。しかしながら、そのような「学び」を導く指導法の設計や、学びの効果を定量的に把握することは容易ではないという問題も指摘されている。また、インクルーシブ社会を形成するため、発達障がい者／児への特別支援教育の促進も強く望まれているが、障がいの様相は極めて多様であるため、個々の障がいの状況を適切に把握し、最適な個別指導を計画することも容易ではないことが知られている。

例えば、ASD 者など発達障がい者の社会性支援の方法として、ソーシャルスキル指導(SST)が効果的であることが示唆されている。従来の SST では、専門家が支援対象者の行動観察に基づき、個々人に応じて具体的な指導目標及び指導計画を立ててきた。その際の評価手法は、心理尺度と専門家による対象者の行動観察の併用が主であった[Hood2017, Ko2019]。しかしながら行動観察は、十分に訓練を受けた専門家が十分に時間をかけて行う必要があるため高コストであり、その問題が SST の支援現場への実装困難へとつながっている。そのため、より客観的かつ効率的な分析法の実現が望まれている。さらに、従来の SST では、主に会話内容の選択や頷きなどの応答、話し手と聞き手の会話交代（話を振る）ことなどが指導目標の中心であったが、実際の会話では目線移動、発話の抑揚といった微細な非言語行動・表出が相手への伝達に大きく影響する。こうした側面は、従来、心理専門家による行動観察では適切に分析・評価することは困難であった。それに対して本研究では、微細な非言語情報の計測・定量評価が可能な ASI 技術に着目し、SST への適用を目指す点にて、最先端の取り組みであると位置づけられる。

このような観点から、教育を支援するための情報通信技術や人工知能技術の導入・活用が望まれているが、現在、ChatGPT などに代表される人工知能技術は、言語理解・運用能力において著しい進歩を遂げているものの、「知・情・意」と称される人の知能の内、「情」や「意」に対応する社会的知能の実現は立ち遅れている。そのため人々の行動や心理を理解し、その知的活動を支援することが可能な「人工社会知能(ASI: Artificial Social Intelligence)」の開発が次世代のフロンティアとして注目されている。その一環として、近年、人と人との対話場면을対象として、画像や音声データより、人物行動や場の状態を認識する人工知能研究が盛んに行われている。例えば、就職面接場면을対象とし、応募者の雇用可能性[Nguyen2014]やコミュニケーションスキル[岡田 2016]を推定する研究や、グループディスカッションやグループワークにおける対話者本人の主観的印象（やる気、雰囲気の良いさなど）[Otsuchi2021]、集団の結束力[Walocha2020]などを推測するモデルも提案されている。これら研究の一環として、我々の研究室では、対話者の非言語行動が発揮する「機能」に着目した研究を進めている。その一環として、対話者の表出する複数モダリティの非言語行動（頭部運動や顔表情など）を対象とし、それらが担う多様な「機能」を自動的に認識するモデルを提案している[Otsuka2020]。また、その検出された機能に基づいて、対話者の主観（満足度など）が高精度に推定可能であることも示している[Otsuchi2021, Otsuchi2023]。

本研究では、人工社会知能の構成と、学習者の行動理解・支援という2つの課題に対して、文理の知の相互還流による有機的な連携を目指した。その成果は、情報通信と教育という2つの分野へインパクトをもたらし、文理融合型研究のモデルケースとして、今後、様々な研究開発プロジェクトの創出・発展へも貢献することが期待される。

3 実験、及び、分析

3-1 学びの場でのデータ収集（ソーシャルスキル指導(SST)の実施）

ASD 児を対象とした実験的なソーシャルスキル指導(SST)を実施し、その状況を収録した(注1)。SST は、図1のように、1名の成人指導者（健常者）と1名の対象児童が対面にて対話を行う状況を対象とした。約10分間の対話を一つのSSTセッションとした。SST は、自由対話形式とした。この対話を、各々の正面顔、

及び、二者の側面を捉えるカメラ（図1の画角）により、収録した。動画を撮影するカメラとして、GoPro Hero 10 を使用し、収録の解像度はフルHD、フレームレートは 29.97[frame/sec]と設定した。また、両者の胸部には、音声を収録するためにマイクロフォン（Rode 社 Wireless Pro）を使用した。音声収録のサンプリング周波数は、48[kHz]とした。また、同様の SST セッションを、定型発達（TD）児を対象として実施した。表1に参加した児童とその特徴を記述する。



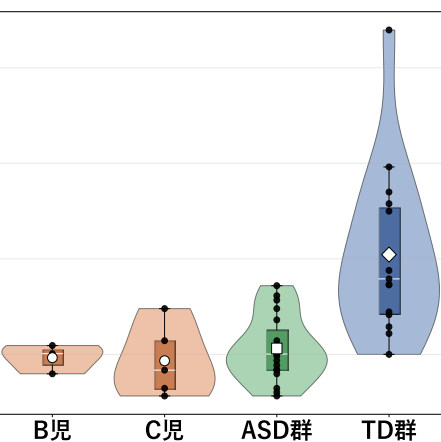
図1：ソーシャルスキル指導（SST）の様子

表1：各参加児の特徴

実験参加児童の詳細と実施セッション数

分類	名称	性別	学年	セッション数
ASD	A児	男	小学4年	11
	B児	男	小学5年	3
	C児	女	小学4年	5
TD	D児	女	小学3年	3
	E児	男	小学3年	3
	F児	男	小学3年	3
	G児	男	小学3年	3
	H児	男	小学4年	3

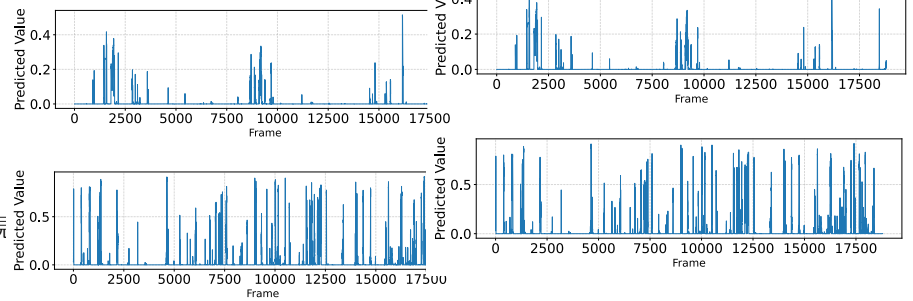
発話テンポ基底



ラム解析(sFSA: synergistic functional spectrum analysis)

解析[Imamura2025]とは、非言語行動の特徴である機能多重性と解釈曖昧性を捉えるモダリティに跨る複数機能の知覚強度分布を表現するベクトルである「機能スペクトラム」の時系列を入力とし、その因子分解により、対話において生じる主要な機能群を発見する方法である。今回、対象とする児童の対話における非言語行動の機能を分析する準備段階として、成人健常者一般の対話において、頻出する主要な非言語行動の機能群を特定し、その機能群を対話映像より自動推定するための深層学習モデルを構築した。

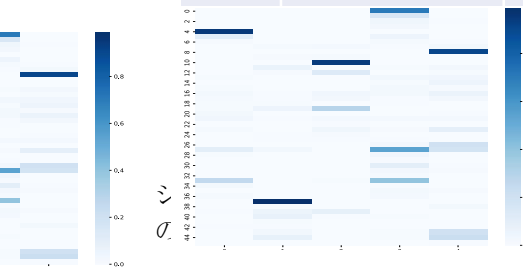
まず、成人健常者同士の対話データを対象として、顔表情 27 種（話し掛け、反応確認など）、及び、頭部運動 19 種（リズム取り、強調など）について、映像フレーム毎に 3 人の外部評定者が各機能の有無を非排他的にラベリングしたデータに着目した。このラベリングデータに基づいて、フレーム毎の機能スペクトラムを求め、その時系列を並べた行列に対して、sFSA を適用した。sFSA では、この入力行列を、準直交非負値行列分解により、基底行列（相乗機能基底と呼ぶ）と係数行列の積に分解する。この基底行列は、顔と頭の 2 モダリティがどのように共起してコミュニケーション機能を果たすかを表す。この基底行列は、表 2 に示すように、「 $\mathbf{b}_1$: 話し手の思考」、「 $\mathbf{b}_2$: 話し手の話し掛け」、「 $\mathbf{b}_3$: 聞き手の相槌」、「 $\mathbf{b}_4$: 聞き手の思考」、「 $\mathbf{b}_5$: 親和感情」に対応する機能群を表す列ベクトル（基底ベクトル）から構成される。続いて、対話映像から、これら



の基底が表す機能群が各時刻で発現
[Imamura2025]を構築した。

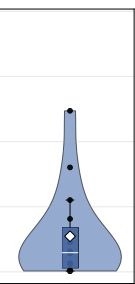
表 2：相乗機能スペクトラム解析により得られた機能基底

基底	解釈名	主要な機能	機能区分
b_1	話し手思考	話し手の思考（顔表情）、話し手の思考（頭部運動）	発話関連
b_2	相槌反応	相槌（頭部運動）	反応関連
b_3	聞き手思考	聞き手の思考（顔表情）、退屈・不注意（顔表情）	反応関連
b_4	話し掛け	話し掛け（顔表情）、リズム取り（頭部運動）、話し手の思考（頭部運動）	発話関連
		親密さ（顔表情）、笑い（顔表情）、正の感情表出（頭部運動）	その他

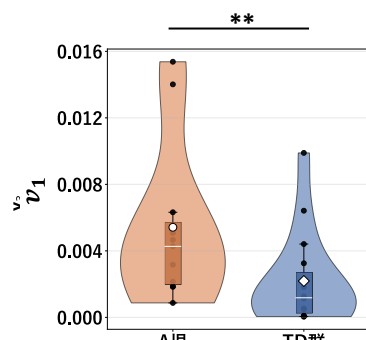


今回新たに収録した児童を含む定型発達児童 5 名それぞれによる対話セッション単位に、5 個の機能群の発現度合いを CNN にて推定した。その後、各児童の平均強度を計算し、比較を行った。

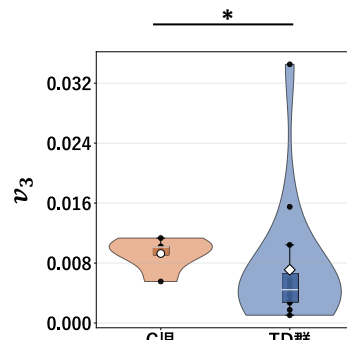
思考基底



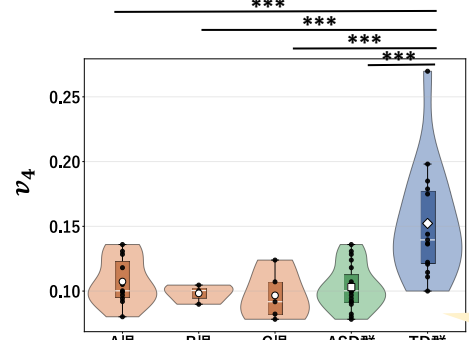
b_1 ：話し手思考基底



b_3 ：聞き手思考基底



b_4 ：話し掛け基底



分類
ASD

図 2：ASD 児と TD 児との間で、機能基底の平均強度の間に統計的に有意差が見られた機能基底

図 2 は、セッションごとの、機能基底に対応する機能群が現出した平均強度をバイオリンプロットで表現したものである。ASD 児と TD 児との間で、機能群の平均強度の間に統計的に有意差が見られた機能基底のみ表示している。図 2 において、●は、各セッションにおける平均値、○は、個人における平均値、□は、ASD 群全体の平均値、◇は、TD 群全体の平均値をそれぞれ表す。また、*は、ブレンナーマンソール検定において得られた p 値の範囲を示す (* $p < .05$ 、** $p < .01$ 、*** $p < .001$)。

図 2 の話し手思考基底 b_1 より、A 児は、話し手思考の機能群の平均強度が、TD 群より全体的に高位置かつ広く分布していることが読み取れる。つまり、A 児は、話すときに思考をしているように見えた度合いが高かったことを示唆する。また、聞き手思考基底 b_3 より、C 児は、聞き手思考の機能発現の平均強度が TD 群より高位置に分布し、かつ分布幅が狭い。これは傾聴児に思考関連の機能の表出が TD 児よりも活性化しており、情報処理や理解の負荷が高い可能性を示唆している。加えて、話し掛け基底 b_4 については、ASD 児全員、及び、ASD 群の平均強度の分布が、TD 群と比較して、低い位置に分布していることが分かった。これより、ASD 児が相手に向かって話し掛ける時、相手からみてその話し掛けの様子が伝わりにくい可能性が示唆される。この結果は、ASD 児は、話し掛けにおいて、相手の注意や理解を引き出すための社会的な手がかりが乏しく、会話相互作用が立ち上がりにくいという一般的な知見とも整合するものである。

4 まとめと今後の課題・展望

本研究では、ASD 児を対象としたソーシャルスキル指導 (SST) の場面を対象とし、そこで生じる ASD 児の

非言語行動や社会的相互作用を、AI を用いた自動認識技術により検出・分析することが可能であるか検証し、その結果、これら技術の有効性・可能性を確認した。この技術は、人工社会知能の重要な構成要素として将来的に広く用いられると期待される。特に本研究では、頭部の運動と顔の表情に着目し、その生起の頻度や出現する機能の種別などの分析を行い、ASD 児特有の性質を明らかにした。加えて、本研究は、工学系の研究者・学生と教育学部との文理融合型の研究プロジェクトであり、文理の障壁を超えて連携を行うための体制作りにおいても一定の効果が認められたと考える。以下に今後の課題について述べる。

まず、今後、より多くの ASD 児、TD 児を募集し、ソーシャルスキル指導 (SST) の収録を行う予定である。本研究において対象とした ASD 児は 3 名と少数であり、TD 児も 5 名と少数であった。分析の結果、ASD 児と TD 児との間にて特定の機能群の平均強度にて、差異が確認された一方、個人性が強いことも示唆された。今後、ASD 児と TD 児と差異について、より一般性を持った主張をするためには、ASD 児、TD 児ともに実験参加者を増やすことが最重要の課題としてあげられる。

また、本研究は、頭部運動と顔表情という 2 つの非言語モダリティのみに着目して、分析を進めた。しかし、従来、ASD 児に特有の性質として、アイコンタクトなどの視線相互作用が欠如するという性質もよく知られている。今後、頭部運動と顔表情に、視線行動も加え、より多数のモダリティの行動を分析することで、ASD 児特有の行動の性質がより明確になると目される。これまで我々の研究室では、頭部運動機能、顔表情機能の他に、視線機能の認識モデル[Tashiro2023]も構築してきた。今後、これらの技術を活用することで、ASD 児の視線の分析を行い、より多様な非言語行動を介した指導者と児童間の社会的相互作用の分析へと駒を進める予定である。それにより ASD 児特有の社会的相互作用の様相をより深く解明することができると期待される。

加えて、技術的な課題としては、頭部運動や顔表情などの機能を検出するための機械学習モデルの性能向上もあげられる。今回の SST 対話の分析に用いた機械学習モデルは、定型発達の成人同士の対話において生じた行動を用いて訓練されたモデルであった。上述の実験・検討により、このモデルに一定の有効性が認められた一方、このモデルでは捉えきれない子どもや ASD 児特有の行動も多く存在することが示唆されている。よって、今後は、これら児童の非言語行動について、より深く吟味することが求められる。

最後に、これらの分析の結果に基づき、個別最適な学びの場を設計するための指針の策定も並行して行う予定である。

【参考文献】

- K. Nagayama, T. Kamiyama, and K. Wakabayashi, "Video modeling and systematic instructions to teach graphing skills to elementary school students with intellectual and hearing disabilities," *The Journal of Special Education*, Vol. 57, No. 4, pp. 248-255, 2023.
- 丹治敬之, 小路一直, 内田佳那, 神山 努, 涌井 恵: 知的障害特別支援学級における機能代替アプローチによる意欲的な読み書き学習をめざした ICT 活用実践. LD 研究, Vol. 30, No. 4, pp. 307-313, 2021.
- S. A. Hood, K. C. Luczynski, and D. R. Mitteer, "Toward meaningful outcomes in teaching conversation and greeting skills with individuals with autism spectrum disorder," *Journal of Applied Behavior Analysis*, Vol. 50, Issue 3, pp. 459-486, 2017.
- J. A. Ko, A. R. Miller, T. W. Vernon, "Social conversation skill improvements associated with the Social Tools And Rules for Teens program for adolescents with autism spectrum disorder: Results of a randomized controlled trial," *Autism*, Vol. 23, No. 5, pp. 1224-1235, 2019.
- L. S. Nguyen, D. Frauendorfer, M. S. Mast, and 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, 2014.
- 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, 涌井 恵: 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. A130-E.1-12, 2016.
- K. Otsuka, and M. Tsumori, "Analyzing multifunctionality of head movements in face-to-face conversations using deep convolutional neural networks," *IEEE Access*, Vol. 8, pp. 217169-217195, 2020.

- S. Otsuchi, Y. Ishii, M. Nakatani, and K. Otsuka, “Prediction of Interlocutors’ Subjective Impressions Based on Functional Head-Movement Features in Group Meetings,” in Proceedings of 23rd ACM International Conference on Multimodal Interaction (ICMI2021), pp. 352–360, 2021.
- F. Walocha, K. Maman, M. Chetouani, and G. Varni, “Modeling Dynamics of Task and Social Cohesion from the Group Perspective Using Nonverbal Motion Capture-based Features,” In Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI ’20 Companion), pp. 182–190, 2021
- S. Otsuchi, K. Ito, Y. Ishii, R. Ishii, S. Eitoku, and K. Otsuka, “Identifying Interlocutors’ Behaviors and its Timings Involved with Impression Formation From Functional Head-Movement Features and Linguistic Features,” in Proceedings of 25th ACM International Conference on Multimodal Interaction (ICMI2023), pp. 336-344, 2023.
- T. Backer van Ommeren, H. M. Koot HM, A.M. Scheeren, and S. Begeer, “Sex differences in the reciprocal behaviour of children with autism,” Autism, Vol. 21, No. 6, pp. 795-803, 2017.
- M. Imamura, A. Tashiro, S. Kumano, and K. Otsuka, “Synergistic Functional Spectrum Analysis: A Framework for Exploring Multifunctional Interplay among Multimodal Nonverbal Behaviours in Conversations,” IEEE Transactions on Affective Computing, vol. 16, no.2, pp. 1056-1073, 2025.
- A. Tashiro, M. Imamura, S. Kumano, and K. Otsuka, “Analyzing and Recognizing Interlocutors’ Gaze Functions from Multimodal Nonverbal Cues,” in Proceedings of 25th ACM International Conference on Multimodal Interaction (ICMI2023), pp.33-41, 2023

(注 1) 本実験は、横浜国立大学 人を対象とする非医学系研究倫理審査委員会による承認(非医-2024-09、非医-2024-78)を受け実施された。また、実験に先立ち、本人及び保護者に研究内容を説明し、参加の同意を得た。

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
頭部運動機能検出を用いた自閉スペクトラム症児の対話特徴分析	日本特殊教育学会 第 63 回大会	2025 年 9 月 13 日発表
自閉スペクトラム症児の会話における顔表情と頭部運動の非言語機能の特徴分析	電子情報通信学会 HCG シンポジウム 2025	2025 年 12 月 12 日発表