

# 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ゴールとして設定した。

特に本研究では、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、自閉スペクトラム症(ASD)者に対する「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が、個人によってその障がい実態は様々である。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、SSTにおいて、各人の言語・非言語行動を画像・音声データより計測し、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。

## 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、個人に依って具体的な指導目標及び指導計画を立ててきた。そのこの評価手法は、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とって行う必要があるため高コストであり、それが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、実際の会話では視線移動、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 への適応を目指す点にて、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、話者の非言語行動が發揮する「機能」に着目した研究を進めている。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を示している[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た。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 者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)。具体的には、ASD 者（ASD 児）として、小学生の男児 2 名（A 児、B 児と記す）、女児 1 名（C 児）を対象とした。また、定型発達児（TD 児と記す）として、小学生の女児 1 名を対象とした。SST は、1 回 10 分のセッションを単位とし、合計 20 回、一定期間において実施した（A 児 11 回、B 児 3 回、C 児 6 回、D 児 3 回）。SST は、自由対話形式とした。実施場所は、X 大学のプレイルーム、または、児童の自宅にて行った。図 1 に SST の様子を示す。図 1 に示すように、SST は、成人男性のトレーナー（図 1 左）と対象児童（図 1 右）とが対面した状態で行われ、各々の正面顔及び、人物側面を捉えるカメラ（図 1 の画角）により動画の撮影を行った。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。表 1 に各児童の特徴を記述する。



図 1：ソーシャルスキル指導（SST）の様子

Table 1 各参加児の特徴

対象児童	学年	性別	診断	WISC-IV結果
A	小4	男	ASD, ADHD, LD	FSIQ90, VCI91, PRI100, WMI85, PSI91
B	小5	男	ASD	-
C	小4	女	ASD, ADHD, LD	FSIQ91, VCI115, PRI74, WMI68, PSI110
D	小2	女	診断なし	-

### 3-2 行動の計測と分析

前述の SST において収録された映像を用いて、対象児童、及び、指導者との間に生じた社会的相互作用の分析を行った。本研究では、トレーナー側での頭部運動を用いた働きかけに対する児童側における頭部運動を用いた応答の様相を分析した。頭部運動に着目した理由としては、頭の動きは、コミュニケーション上の役割を有し、精神的及び感情的な情報を表すことが知られ、また、自閉スペクトラム症 (ASD) 児の非言語コミュニケーション能力を知るうえでも重要な役割を持つことが示唆されている [Zhao2021]。しかし、非言語コミュニケーションの研究分野では、頭の動きに着目した研究は、視線や表情を用いた研究に比べ少なかった。その要因としては、頭の動きは、従来、頷きが肯定、首振りが否定、傾げが疑問というように、明確なジェスチャとしてみなされることが一般的であり、自発的かつ微細な動きに含まれる多様な意図や感情などを、運動のカテゴリに分類することは困難であるとされていたことがあげられる。近年、深層ニューラルネットなど技術の進歩により、そのような微細な頭部運動とそこに含まれる豊富な意味合いを教師あり機械学習の枠組みで捉えて、自動認識が可能な技術が実現されている [Otsuka2020]。本研究は、そのような新しい行動認識技術の応用として、発達障がいを抱える児童 (特に ASD 児) の教育の場面を対象とする。

表 2：頭部運動機能の定義

区分	記号	機能名 (英)	機能名 (日)	機能定義
発話関連	hs1	rhythm	リズム取り	発話に従って自然と現れる頭の動き
	hs2	emphasis	強調	発話を強調する際に表出される顕著な頭の動き
	hs5	confirm response	反応確認	聞き手の反応を伺う際や反応を促す際に現れる頭の動き
	hs8	think to speak	思考[発話時]	発話時に思考したり、言葉を選ぶ際に現れる頭の動き
反応	hr1	back-channel	相槌	話を聞いていることのサインとして現れる縦方向の頭の動き
	hr2	response	応答	他者の質問・呼びかけに対して応答する際に現れる縦方向の頭の動き
	hr5	thinking	思考[受け手]	話者の発言を理解しようと考えているときの (ことを示す) 頭の動き
	hr6	understanding	理解	話者の発言が理解できたことを示す、またはその振りをする際の縦の頭の動き
	hr11	affirmation	肯定	肯定・同意・承諾・許可・承服などの返答をする際に現れる縦方向の頭の動き
他	hc1	positive emotion	正の感情表出	共感・同情・好感・関心・喜びなどの正の感情表出に伴う頭の動き

#### (1) 頭部運動機能の検出

まず、SST の状況を収録した映像より、OpenFace と呼ばれる顔画像追跡手法を利用し、児童とトレーナーそれぞれの画像上での顔の位置・姿勢の計測を行い、各時刻 (映像のフレーム) における頭部姿勢角を算出

した。また、音声データより発話区間を検出した。次に計測された頭部姿勢角、及び、発話区間を入力情報とし、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）それぞれの有無を映像のフレーム単位に検出した。表 1 にこれら 10 機能の定義を示す。この頭部運動機能 10 種は、主に話し手の発話に関連する 4 機能（話し掛け、強調、反応確認、思考）、反応に関する 5 機能（相槌、応答、思考、理解、肯定）、及び、その他の 1 機能（正の感情表出）の区分に分けられる。なお、これらの機能は同時に複数表出されることも踏まえて、非排他的に定義され、検出も各機能について独立に行われる。

この検出のためには、本研究室で開発された畳み込みニューラルネット（CNN）を用いた[Otsuka(2020)]。この CNN は、入力として、約 1 秒間の窓をかけた頭部の姿勢角（azimuth, elevation, roll の 3 自由度）の角速度の時系列、並びに、発話の有無の情報の時系列（時間平均フィルタにより平滑化を利用）を用いた。図 2 にこの CNN の構造を示す。図 2 で示すように、この CNN は、3 層の畳み込み層（conv1, conv2, conv3）を持ち、各層において、入力された 4 チャンネルの信号は、それぞれ独立に処理され、特徴抽出が行われる。4 層目の全結合層（fc4）によりこれら 4 チャンネルの信号が統合され、最後の 5 層目（fc5）により、対象機能の有無（2 値）が出力される。CNN は、頭部運動 10 機能のそれぞれについて別個に訓練される。訓練データは、成人女性 4 名を一組、2 グループによるグループディスカッション、合計約 30 分のデータについて、3 名の外部評定者が機能の有無を非排他的に評定し、3 名の結果を論理和により統合したものをを用いた。この論理和による統合は、3 名の内、一人でも機能の存在を認めた場合に、最終結果にその判断が採用されるという方式であり、できる限り多様な頭部運動機能を敏感に捉えるという観点からこの統合法を用いた。この学習済みの CNN を、今回、トレーナと児童のデータに適用した。なお、先行研究によりこのような学習済み CNN の利用の妥当性が確認されていたため、今回も CNN のファインチューニングなどは実施しなかった。なお、事後に検出結果の妥当性は、目視により確認している。

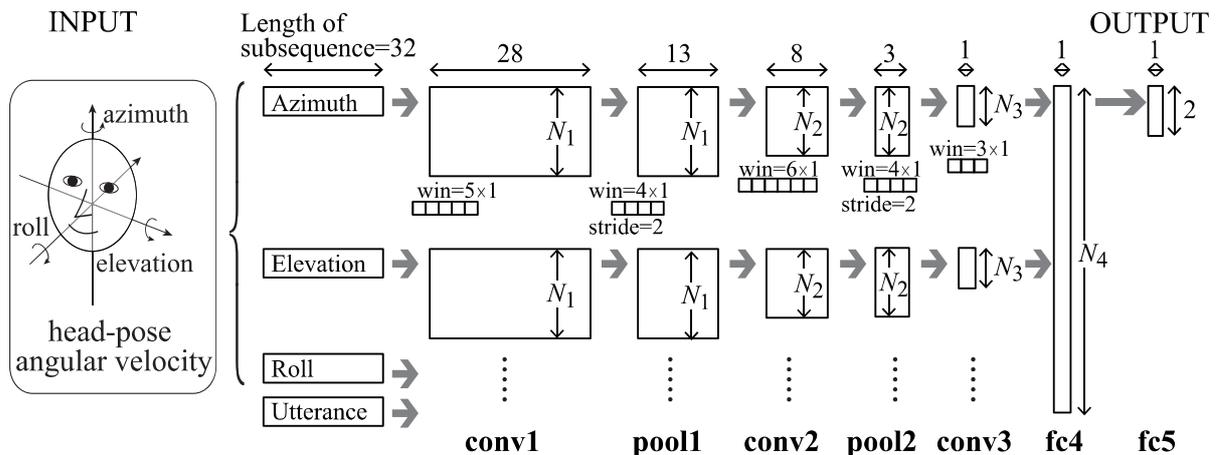


図 2: 頭部運動機能検出のための畳み込みニューラルネット (CNN) の構造。文献[Otsuka2020]からの引用・改変。

## (2) 反応時間の測定

続いて、本研究で狙いとする、指導者・学習者間における社会的相互作用の質や量を分析するため、トレーナ側での頭部運動を用いた働きかけに対する児童側において頭部運動を用いた応答の様相を分析した。具体的には、トレーナ側のリズム取り、強調、反応確認、発話時の思考の各々が表出されたときに、児童側にて、相槌、応答、思考、理解、肯定、正の感情表出のそれぞれが表出される回数や、表出までの反応時間を調査した。本研究では、ASD 児と定型発達児との 2 群間での差異を知ることを主眼とするのではなく、ASD は、その障害の程度や種類の幅が広く（故に「スペクトル」と称される）、ASD 児との間の差異が非常に大きいことから、ASD 児の各々と、TD 児との個人間にて比較を試みた。具体的には、トレーナ側での表出機能と児童側での表出機能の組を考え、それぞれの機能組について、反応時間が 3 秒未満の場合の平均値、及び、反応時間が 3 秒以上の割合を計算した。その後、3 名の ASD 児それぞれについて、TD 児との間にて反応時間の差があるか検証した。具体的には、平均反応時間には Mann-Whitney の U 検定を、3 秒超の割合には Fisher

の正確確率検定を用いた。なお、有意水準は5%とした。今回は、これら双方に差があることをもって、各 ASD 児と TD 児との間の反応時間に差があると定めて、両者間の多重比較により、双方に有意差が認められた機能組を特定した。この多重比較のため Bonferroni の方法による有意水準の補正を行った。

表 3：ASD 児（A 児、B 児、C 児）と D 児の反応時間に有意差が認められた機能組。機能を示す記号は、表 2 に従い、機能組は、発話関連機能（hs1, hs2, hs5, hs8）の各々と各児童の機能の各々が対となったものを指す。

発話関連機能	A 児	B 児	C 児
リズム取り hs1	hr1, hr6	hr11, hc1	hr1, hr2, hr5, hr6, hr11, hc1
強調 hs2	-	hr11, hc1	hr1
反応確認 hs5	hr1, hr5	hr2, hr6, hr11, hc1	hr1, hr5, hr6
思考(発話時) hs8	hr1, hr5, hr6	hr2, hr5, hr6, hr11, hc1	hr1, hr2, hr5, hr6, hr11, hc1

## （2）結果と考察

結果、2名の ASD 児（A 児、C 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（B 児）は、TD 児よりも反応が早い機能組が多く確認された。表 3 にこれらの機能組を示す。

A 児は、トレーナ側のリズム取りや反応確認、思考の 3 種類の発話関連機能に対して、相槌や思考、理解といった反応機能を表出するまでの時間が D 児よりも有意に遅かった。その要因として、A 児は言語理解能力が高くない（WISC-IV: FSIQ90, VCI91, PRI100, WMI85, PSI91）が、それにより、トレーナの意図を理解して自らの発話表現をするまでに、時間を要したと考察される。一方、B 児は、多くの機能組で有意に D 児よりも反応が速く、D 児よりも有意に反応が遅い機能組はみられなかった。その背景として、B 児には自身が興味のある話題を一方的に話す傾向があり、それが影響したものと思われる。さらに C 児は、トレーナ側での発話関連機能 4 種類を含む機能組において、D 児よりも有意な遅れが見られた。C 児の場合、映像からは、視線の逸れや、不自然な笑顔が時折見受けられたが、全体的に顕著な特徴はなかった。先行研究では、ASD のある女兒には、周囲に適応しようと自身の特性を隠す傾向があることが示唆されている [Backer van Ommeren2017]。この傾向により、相槌や思考などの社会的相互作用を示す頭部運動機能がすぐには表出されず、遅れを伴って現れた可能性が考えられる。

このように、頭部運動という非言語行動がもつ様々な機能に着目し、トレーナ・児童間の社会的相互作用の種別とその量、反応時間を定量化し、ASD 児の個人毎の特性に紐づく分析と考察ができる可能であることが分かった。

## 3 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、自動認識技術を用いて検出・分析することが可能であるか検証し、これら技術の有効性を確認した。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も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 児特有の性質が明らかとなった。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、分野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。このような成果を踏まえて、本研究にはさらなる発展の余地があることから、今後も継続する予定である。以下にそのため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 児は 1 名とさらに少数であった。分析の結果、ASD 児と TD 児との間にて反応時間の差異が確認された一方、個性が強いことも示唆された、一般的な ASD 児と TD 児と差異について主張できる統計的な知見が見出されたとはいえない。そこで対象とする参加者を ASD 児、TD 児ともに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、頭部運動という一つの非言語モダリティのみに着目して、分析を進めた。その結果、頭部運動機能の表出の反応時間により、ASD 児の社会的相互作用が特徴付けられる見込みが得られた。一方、従来、ASD 児に特有の性質は、他のモダリティにも現れることが指摘されている。例えば、表情の表出が少

ないことや、アイコンタクトなどの視線相互作用が欠如するといった性質が知られている。頭部運動にこれらの顔表情、視線を加えた、複数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。我々の研究室では、今年度注目した頭部運動機能の他に、これまで顔表情機能、並びに、視線機能の認識モデルを構築してきた。今後、これらの技術を活用することで、ASD 児の顔表情・視線の分析が可能となり、マルチモーダルな非言語行動を介したトレーナ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頭部運動機能を検出するための機械学習モデルは、定型発達の成人同士の対話において生じた頭部運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、これら児童の頭部運動の機能について、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を策定する予定である。

## 【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. AI30-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.

A. Zhao, A. Zhu, X. Zhang, H. Tang, J. Xing, X. Hu, J. Lu, Q. Peng, X. Qu, “Atypical head movement during face-to-face interaction in children with autism spectrum disorder,” *Autism Research*, Vol. 14, Issue 6, pp. 1197-1208, 2021.

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.

(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～15 日