人の対話場面における相互理解と合意形成過程の解明に向けた非言語機能の分析基盤の構築

Building a Nonverbal Functional Analytical Framework for Elucidating Mutual Understanding and Consensus-Building in Human Dialogic Interactions



大塚 和弘（Kazuhiro OTSUKA, Ph. D.）
横浜国立大学 大学院 工学研究院 教授
(Professor, Yokohama National University, Faculty of Engineering
電子情報通信学会 IEEE 他
受賞：電子情報通信学会 平成22年度 論文賞 受賞 (2011年). 同 平成22年度 喜安善市賞 受賞（2011 年), ACM 9th Int. Conf. Multimodal Interfaces (ICMI2007) Outstanding Paper Award 賞 受賞 (2017年) 他

研究専門分野：知能情報学、マルチモーダルインタラクション

あらまし

本研究は、人と人とが対面や遠隔で行う会議や対話の場面に焦点を当て、その対話の場において、対話者の非言語行動と対話者間の非言語相互作用が果たす機能を理解するための分析枠組みと計算基盤の構築を目的とした。非言語行動として、頭部運動や顔表情、視線行動に焦点を当て、それらのもつコミュニケーション上の機能を定義し、観測可能な行動から推測する機械学習モデルを提案した。また、複数の観測者による複数の非言語機能の知覚強度の分布である「機能スペクトラム」を提案し、そのスペクトラム分解に基づいて対話を構成する基盤的な非言語機能や相互作用機能を見出す「非言語機能スペクトラム解析」を提案した。さらにこれら対話中に出現した非言語機能の出現頻度や機能間のバランスなどを手がかりとして、対話者本人が対話中に抱いた主観的印象を推測する機械学習モデルを併せて構築した。

１．研究の目的

本研究では、人と人とが行う対話に焦点を当て、その対話の場において、対話者の非言語行動と対話者間の非言語相互作用を理解するための分析枠組みと計算基盤の構築を目的とした。特に非言語のモダリティとして、対話を駆動する主要な役割を担うとされる、頭部運動、顔表情、視線行動に焦点を当てた。本研究では、これら非言語行動の機能の解明、及び、それらを観測可能な対話者の行動から推測するための機械学習モデルの構築を目指した。さらに対話中に現れる各種の非言語機能や非言語相互作用機能の現出の状況が、対話者や対話そのものの質とも深く関係すると考え、対話者本人が感じた対話の質を、対話者の非言語行動から推測可能な機械学習モデルの構築も目的とした。

２．研究の背景

人と人とが行う「話し合い」や「会議」は、社会集団による意思決定のための最も基本的な営みである。近年、従来型の対面で行われる会議以外にも、新型コロナ感染症の流行を契機として、遠隔会議が急速に普及し、今では社会のインフラとしての地位を確立している。一方、遠隔会議においては、相手の顔がよく見えない、場の雰囲気が分からないなど、対面とは異なる環境に起因する問題も指摘されている。その要因の一つとして、人は、他者との対話において、単に音声を介して、言語情報をやり取りしているだけではなく、表情や身振り手振りという非言語的なモダリティを介して、意思疎通を図り、人間関係の調整を行っていることがあげられる。それを踏まえると、遠隔状況においては、離れた地点間を結ぶ、非言語の伝達チャネルが十分な帯域を有せず、対面時にやり取りされていた豊富な非言語情報が伝達されないことが、遠隔会議の違和感の要因であると考えられる。

このような遠隔会議の質を向上させるため、実際に人の対話で生じている非言語情報を定量的に把握し、対面の状況と遠隔状況とを比較して、その差異を分析することが、問題解決の鍵を握ると考えられる。しかしながら、従来、定量的に対話の質を計測するための技術は確立されていない。また、非言語情報は、定量化や記号化が困難であり、対話における機能を明確に

人の対話場面における相互理解と合意形成過程の解明に向けた非言語機能の分析基盤の構築

Building a Nonverbal Functional Analytical Framework for Elucidating Mutual Understanding and Consensus-Building in Human Dialogic Interactions

把握することは難しい課題とされている。我々は、対話において現れる非言語行動の機能の種類と量を知ることにより、遠隔状況と対面状況との差異を明確にすることができ、対話環境の性能向上や人々のコミュニケーションスキルの向上へと繋がるヒントが得られると考えた。この考えのもと。本研究では、人と人との対話や会議における、対話者の非言語行動の機能を解明するための分析基盤の構築という課題に取り組むこととした。

３．研究の方法

本研究は、前述の目的に対して、1）非言語機能の定義とラベリング、２）非言語機能の自動認識、３）対話者の主観的印象の推測、４）非言語機能スペクトラム解析、という複数の段階を設定し、研究を実施した。

まず、最初の１）非言語機能の定義とラベリングにおいて、本研究では、非言語行動のモダリティとして、頭部運動、顔表情、視線行動に着目した[1,2,3]。頭部運動は一般的な頷きや首振りなどのような「同意」、「非同意」の意思表示の手段とされるが、実際は、話し手、聞き手の間において相互作用を調整するための多種多様な機能を担っている[1]。例えば、話し手の発話のリズム取りや強調、反応確認などが典型的な発話に関連する頭部運動機能である。また、聞き手において、頷きによる発話傾聴のサインや、首の傾げによる「思考」状態の表出、また、同意・非同意などの表示も、頭部運動の機能に含まれる。また、楽しさなどの感情表出も頭部運動は担う。

顔表情は、従来、喜怒哀楽などの感情表出のデバイスとして研究がなされてきた。しかし、顔表情は、感情表出以外にも様々な機能を担う[2]。例えば、眉毛を上げる表情により発話の強調がなされる。また、聞き手の認知状態として、不確かさの表出などの機能も担う。加えて、感情表出において、基本感情以外に、不注意や退屈、焦燥など多種多様な感情の表出に顔表情が用いられている。

さらに視線行動についても、単に人を見るという機能以外にも、相互作用を調整するための様々な機能がある[3]。例えば、話し手が、聞き手を一瞥することで聞き手の状態をモニタリングしたり、また、聞き手に対してアイコンタクトを試みることにより、何等かのフィードバックを要求したり、さらに、視線を聞き手から逸らすことで、「思考」している状態を示すなど、視線行動は様々な機能を有する。また、聞き手が話し手を見ることで、話し手発話への興味を示し、発話の継続を促すことや、話し手から視線を逸らすことで、思考していることを示す。さらに相手を見ることにより親密さを表出することもできる。

我々の研究室では、これらの非言語行動の機能を列挙・整理し、外部評定者によりラベリングを行うという取り組みを行ってきた。

続いて２）の非言語機能の認識では、１）において付与された機能のラベルを正解データとして、それを自動認識するための機械学習モデルを構築した[1,2,3]。例えば、視線機能としては、頻出上位 10 種にあたる、話し手の話し掛けの機能や、話し手の思考時の目線逸らしなどの機能を対象とした[3]。ここで機械学習モデルとして、畳み込みニューラルネット(CNN)などを構築した。対話映像のフレーム毎にこれらの機能の出現の有無を非排他的に検出する CNN を構築し、性能を検証した結果、平均 F 値 0.53 を達成し、自動認識の可能性が示唆された。また、頭部運動、顔表情についても、同様に頻出上位の機能を対象として、それぞれの機能の表出の有無をフレーム単位で認識するモデルを構築した[1,2]。

続いて、３）対話者の主観的印象の推測のタスクにおいて、前述の２）で認識された非言語機能の現出の状況を手がかりとして、対話者本人が対話に対して抱いた印象を推測するための機械学習モデルを構築した[4]。例えば、Itoh らは、頭部運動機能と顔表情機能の検出結果の時系列を入力として、それらの個々の機能の出現率や出現機能間の出現割合などの統計量を特徴量として計算した（図 1)。また、話し手と聞き手の間の相互作用を捉えるため、話し手の機能と聞き手の機能のペアの出現率を特徴量として用いた。17 グループの 4 人対話を対象として、主観的印象の項目として、興味、面白さ、楽しさ、満足、集中、団結、信頼などの合計 16 項目の本人報告スコアを推測の対象とした。合意形成ディスカッションをタスクとするセッションにおいて、全項目の平均の相関係数が 0.346 を達成し、

人の対話場面における相互理解と合意形成過程の解明に向けた非言語機能の分析基盤の構築

Building a Nonverbal Functional Analytical Framework for Elucidating Mutual Understanding and Consensus-Building in Human Dialogic Interactions

提案法による対話者の主観的印象推定の可能性を確認した。

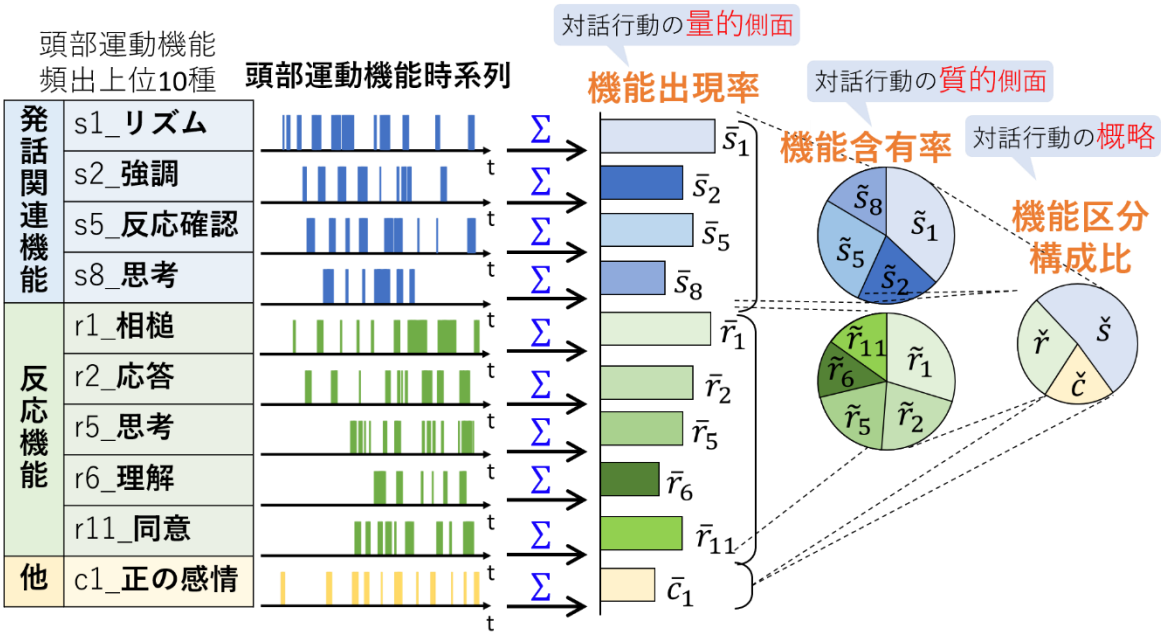


図１：頭部運動機能の特徴量の概略

さらに４）の非言語機能スペクトラム解析では、１）～３）までの方法をさらに発展させ、複数モダリティの複数機能の間の共起関係を分析し、対話を構成する主要な非言語機能群や、相互作用機能群を同定するための新たな分析枠組みを提案した。本研究では、これまで非言語行動の理解を妨げる本質的な要因として、「機能多重性」と「解釈多様性」に着目した。機能多重性とは、一つの行動が複数の機能を同時に担いうることを指し、解釈多様性は、一つの行動が見るものによって異なって解釈される可能性のことを指す。本研究では、これらの性質を表現するために、複数の外部観測者による非言語機能の知覚強度の分布として、「機能スペクトラム」という表現を提案した。この機能スペクトラムを複数モダリティについて求めて、その連結スペクトラムの時系列から、対話中に頻出し、対話を構成する基盤的な非言語機能群を同定し、それらの機能群の時間軸上の活性度を求める方法を提案した[5,6,7]。この方法を「非言語機能スペクトラム解析」と命名した。これは高次元のマルチモーダル・複数機能の次元を、低次元の空間に圧縮して表現する方法とも捉えられる。

具体的にはまず、非言語機能スペクトラム解析法として、Imamura らは、対話者個人の頭部運動機能と顔表情機能の共起関係から、両者がどのように連携して、コミュニケーション機能を果たすかを分析するための方法として、相乗機能スペクトラム解析（sFSA: synergistic functional spectrum analysis)を提案した[5]。この方法は、対話者の頭部運動の機能スペクトラムと顔表情の機能スペクトラムを連結したマルチモーダル機能スペクトラムを入力として、その時系列を束ねた行列について、非負値行列因子分解(SO-NMF)を行う方法として定式化された。図２に示すように、この方法は、入力行列を、基底行列(機能基底)と係数行列の積として近似表現するものである。機能基底行列は、対話中に頻出し、対話を構成する基盤的な機能群を表す機能基底ベクトルから構成される。係数行列は、各機能基底ベクトルが表す機能群が時間軸上で活性している度合いを表す。ここで基底ベクトル間に準直交性の制約を加えることで、各々の機能基底が、互いに異なる機能の側面を捉えるような工夫を施した。

人の対話場面における相互理解と合意形成過程の解明に向けた非言語機能の分析基盤の構築

Building a Nonverbal Functional Analytical Framework for Elucidating Mutual Understanding and Consensus-Building in Human Dialogic Interactions

また、人手で付与された機能スペクトラムを用いた結果の係数行列を正解として、これを観測可能な情報（頭部姿勢角や顔表情の活動単位の活性度）から推測するための畳み込みニューラルネットワークを提案した（図 3）。

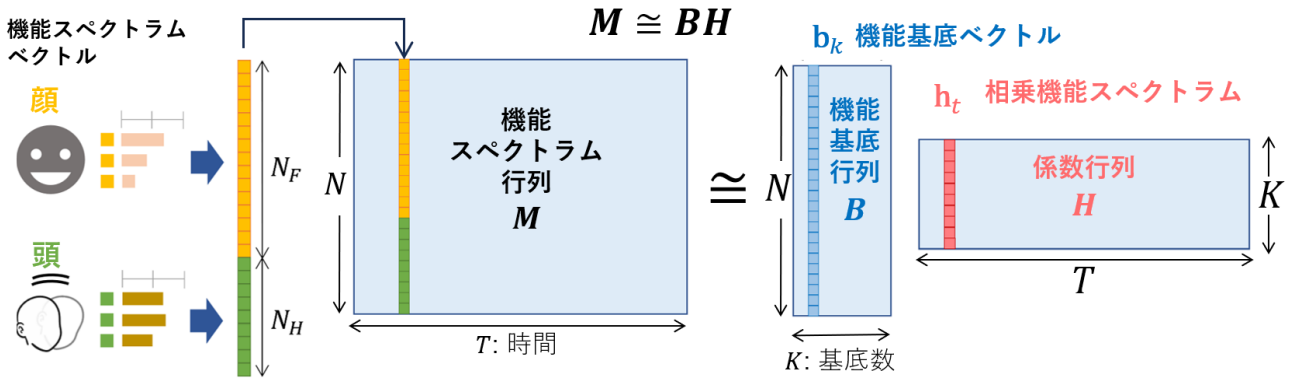


図 2：相乗機能スペクトラム解析（sFSA）の構成

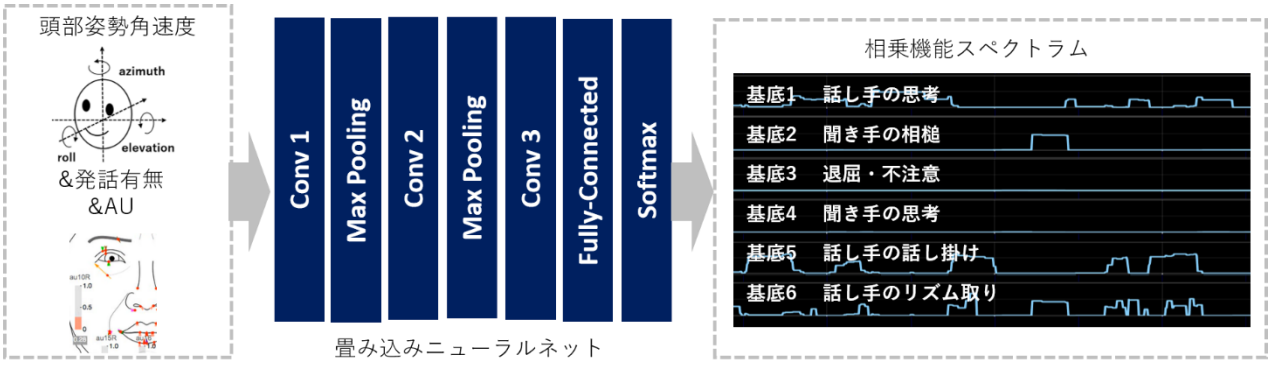


図 3：畳み込みニューラルネット回帰モデルによる相乗機能スペクトラムの推定

さらに Tajima らは、Imamura らの sFSA[5]を拡張して、対話者間の相互作用における機能現出を捉える枠組みとして、相互作用機能スペクトラム解析（i-FSA: interactional functional spectrum analysis）を提案した[6]。この方法は、頭部運動、顔表情に加えて、視線行動の機能についてのマルチモーダル機能スペクトラムを入力として、任意の複数人物の対話状況に適用するため、対話の各時刻における役割として、話し手、受けて、その他聞き手という３つの役割を想定して、それらのマルチモーダル機能スペクトラムを連結したスペクトラムベクトルを構成した。ここで、その他聞き手は、４人対話において２人いるため、この２人の機能スペクトラムベクトルは平均化され扱われた。また、話し手の振る舞いに対する受け手やその他聞き手の反応時間を考慮するため、受け手、及び、その他聞き手の機能スペクトラムは、一定の時間窓幅において平均化がなされた。その後、スペクトラムの分解を行うが、ここでは、sFSA の準直交性非負値行列分解（SO-NMF）で導入した非負値性、準直交性の制約に加えて、相互作用性の制約を付加した。ここで相互作用性の制約とは、機能基底には、話し手サイドの機能成分と、受け手・その他聞き手の側の機能成分の両方が含まれることを要請するものである。また、Tajima らは、この相互作用機能スペクトラムを観測可能な行動から

推定するための CNN を構築した。

さらに Tamura らは、Imamura らの sFSA[5]を拡張して、非言語行動の機能多重性と解釈多様性を分解して分析する方法（pFSA: perceptual functional spectrum analysis）を提案した[7]。この方法は、sFSA で見出された機能基底ベクトルの各々について、複数の外部評定者の解釈傾向を少数の代表的な解釈傾向へ圧縮して表現することにより、非言語機能群に対する代表的な解釈傾向を同定することを狙った方法である。Tamura らは、この問題を、非負値テンソル分解の方法として定式化した。また、頭部運動を対象に付与された、10 人の外部評定者のラベルを用いた実験を行った。その結果、非言語行動の機能の本質的な性質を定量的に分析するための基盤を構築することができた。

４．将来展望

本研究では、従来、定量的な取り扱いが困難であった対話者の非言語行動を、その機能的側面から分析するための基盤を構築した。この方法は、昨今、大規模言語モデル(LLM)に代表される言語的知能を補完して、社会的知能を AI に付与するためのパズルのピースとして、重要な研究領域の開拓に寄与するものと考えられる。しかし、現状、対象とする対話場面は、日本語対話に限定され、また、機能の評定も日本人の観察者に限定され、さらに全体のデータ量も比較的少量である点、提案法の一般化可能性については、今後、継続的な検討が必要な状況である。今後、対話の対象を、日本語以外の外国語に拡大し、より大規模なデータの収集、コーパス構築などが重要な今後の課題としてあげられる。また、本分析法の応用先として、遠隔会議システムの改良という工学的な応用だけではなく、精神医療や療育、教育などへの応用も考えられる。そのような場において提案法を適用して、コミュニケーション障がいの実態把握やコミュニケーションスキルの向上に向けた教育法を開発するなど、様々な展開可能性が考えられる。

参考文献

[1] Kazuhiro Otsuka and Masahiro Tsumori, “Analyzing Multifunctionality of Head Movements in Face-to-Face Conversations Using Deep Convolutional Neural Networks,” IEEE Access, Vol.8, pp. 217169-217195, 2020

[2] 今村まい, 武田一輝, 熊野史朗, 大塚和弘：「対話中の頭部運動と顔表情の相乗機能の認識」, 電子情報通信学会論文誌 A, Vol.J106-A, No.3, pp.70-87, March 2023

[3] Ayane Tashiro, Mai Imamura, Shiro Kumano, and Kazuhiro 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

[4] Koya Ito, Yoko Ishii, Ryo Ishii, Shin-Ichiro Eitoku and Kazuhiro Otsuka, “Exploring Multimodal Nonverbal Functional Features for Predicting the Subjective Impressions of Interlocutors,” IEEE Access, vol. 12, pp. 96769-96782, 2024

[5] Mai Imamura, Ayane Tashiro, Shiro Kumano, Kazuhiro Otsuka, “Synergistic Functional Spectrum Analysis: A Framework for Exploring Multifunctional Interplay among Multimodal Nonverbal Behaviors in Conversations,” IEEE Transactions on Affective Computing, vol. 16, no. 2, pp. 1056-1073, April-June 2025

[6] Momoka Tajima, Issa Tamura, Ayaka Sasaki and Kazuhiro Otsuka, “Unified Functional Spectrum Analysis: A Framework for Understanding Nonverbal Multimodal Multifunctional Interactions in Multiparty Conversations,” in IEEE Access, vol. 14, pp. 47538-47559, 2026

[7] Issa Tamura, Momoka Tajima, Shiro Kumano, and Kazuhiro Otsuka. 2025. Disentangling Perceptual Ambiguity in Multifunctional Nonverbal Behaviors in Conversations via Tensor Spectrum Decomposition. Proceedings of the 27th International Conference on Multimodal Interaction (ICMI2025), pp. 570–578, 2025

人の対話場面における相互理解と合意形成過程の解明に向けた非言語機能の分析基盤の構築

Building a Nonverbal Functional Analytical Framework for Elucidating Mutual Understanding and Consensus-Building in Human Dialogic Interactions

この研究は、令和４年度ＳＣＡＴ研究助成の対象として採用され、令和５～７年度に実施されたものです。