

Voiceless Intention Log: 頸部生体情報を用いた 発話欲求推定に基づく会話参加支援システム

鳥海 優輝^{*1} 遠藤 ゆら^{*1} Lu Yongjiang^{*1} 西 隼輝^{*1}
尾崎 亮太^{*2} 大津 耕陽^{*1} 伊藤 雄一^{*1}

Voiceless Intention Log: A Conversation Participation Support System
Based on Estimating the Desire to Speak Using Neck Biometric Data

Yuki Toriumi^{*1}, Yura Endo^{*1}, Lu Yongjiang^{*1}, Toshiki Nishi^{*1},
Ryota Ozaki^{*2}, Kouyou Otsu^{*1} and Yuichi Itoh^{*1}

Abstract – In group conversations, situations arise where individuals may wish to speak but are unable to do so due to the flow of the conversation or constraints on the timing of their turn. While previous studies have proposed methods for estimating the desire to speak and turn-taking using posture changes, eye movements, and respiration, few studies have focused on physiological signals measured from the neck, which contains key structures involved in speech production. In this study, we propose a conversation participation support system that estimates physiological responses related to the desire to speak by analyzing changes in blood flow, pulse waves, and acceleration data obtained from XB-01, a compact brain activity measurement device worn on the neck. It has been reported that physiological responses preparatory to speech, such as respiratory adjustments and laryngeal muscle activity, occur prior to the onset of speech. Therefore, by comprehensively analyzing changes in blood flow and body movements in the neck, we attempt to detect states in which a person has the desire to speak but has not yet begun speaking. Furthermore, by visually signaling the estimated results to those around the user, we aim to support the coordination of speaking opportunities and facilitate smooth communication.

Keywords : Speech Intention Estimation, Turn-taking, Neck Biometric Data, Conversation Participation Support, Communication Support

1. はじめに

グループディスカッションや会議、雑談などの複数人による会話において、意見を持ちながらも、会話の流れや発話タイミングの制約により、発話できない状況に陥ることがある。先行研究によれば、議論の合意形成は発話量の多い参加者の意見に大きく影響されることが示されている^[1]。このような発話機会の制約は、発話回数の減少を引き起こすだけでなく、特定の視点への偏りを生み、多様な意見の共有を妨げる要因となる^[2]。

会話は発話の切れ目に応じて発話権を交替することで成立するため、衝突を避けつつ参入するためには、話者が切り替わる適切なタイミングを捉える必要がある^[3]。しかし、実際には、座席位置などの空間的制約に加え、上下関係・評価不安といった社会的・心理的

要因が適切な発話タイミングの判断を難しくする。

加えて、昨今のオンライン環境では、対面の会話において発話交替の手がかりとなる視線や相槌などの非言語情報が制限されるため、発話タイミングの判断はより困難になる^[4]。筑波大学による4,343名を対象に実施した調査でも、オンライン会議の課題として「発言するタイミングがつかみにくい」が挙げられており、70.4%の参加者がこの項目に肯定的な回答を示している^[5]。また、通信遅延によって発話衝突が発生することも多く、発話衝突への懸念から発話を控える参加者が存在することも報告されている^[6]。

これらの背景から、発話欲求を持ちながらも発話に至れていない参加者の存在を周囲へ伝えることは、公平な議論を促す有効な支援につながると考えられる。しかし、参加者の内面的な発話欲求は外部から直接観測することが難しく、周囲がそれを正確に認識することは容易ではない。そこで本企画では、発話開始前に現れる身体内部の生体反応に着目し、これらを検出することで発話欲求を客観的に推定することを試みる。既存研究では、発話欲求が生じる際に、脳活動や呼吸

^{*1}: 青山学院大学 理工学部

^{*2}: 青山学院大学大学院 理工学研究科

^{*1}: College of Science and Engineering, Aoyama Gakuin University

^{*2}: Graduate School of Science and Engineering, Aoyama Gakuin University

などの身体活動に特徴的な変化が現れることが示されており^{[7][8]}、これらの生体反応を検出することで発話欲求を推定できる可能性がある。

本企画では、小型脳活動計測装置 XB-01 を用いて発話欲求を推定し、発話に至れていない参加者の存在を周囲へ自然に伝達することで、円滑なコミュニケーションを支援するシステムを提案する。実装の初期段階では、XB-01 の本来の用途である脳の前頭前野の血流変化および脈波を用いた推定を試みる。ただし、XB-01 は原理上、脳以外の部位における血流変化や脈波の計測に応用できる可能性があることから、本企画では頸部のセンシングにデバイスを適用できる可能性についても検討する。発話開始前には、発声に向けた呼吸調整や喉頭・声帯の運動準備が頸部において生じることが報告されている^{[8][9]}。そのため、頸部の血流変化、脈波および加速度を計測することで、発話に関連する生体反応を捉えられる可能性がある。また、頸部への装着は、額部での計測と比べて日常会話場面で利用しやすいことから社会的受容性の高いシステムを実現できると考えられる。

本企画は、従来大型であった脳活動計測装置が小型化されたことの利点を生かし、同様の機構を身体の他部位における生体反応の検出へと展開できる可能性を探索する取り組みとして位置付けられる。企画の準備期間内では、頸部での計測可能性を検討するほか、頸部・額部を併用した計測による推定精度向上の可能性についても検討する。

2. 関連研究

2.1 会話参加支援システム

会話参加や発話タイミングを支援するために発話欲求やターンテイキングを推定する研究が行われている。既存研究では、オンライン会議における発話衝突を防ぐため、前傾姿勢や身体の動きなどの発話予備動作を検出し、発話欲求の有無を参加者間で共有する方法を提案している^[10]。また、オンライン会議参加者の発話欲求の程度を顔表情や視線から推定し、その人が少し前のめりに見える映像効果を付与することで、発話欲求を伝え、その機会の偏りを緩和する Fair Talk が提案されている^[11]。しかし、これらの手法では、マスク着用者や表情変化の少ない参加者に対しては発話欲求の推定が困難であり、さらに発話欲求を示す非言語情報には個人差が存在するため、推定精度には限界がある。

2.2 頸部における発話に関する生体情報

発話は口腔運動だけではなく、呼吸運動や咽頭運動などを伴う複雑な生体活動である。会話では、発話しようとしても他者の発話と競合することで、実際の発

話に至らない場合がある。既存研究では、複数人の会話中に生じる吸気、呼気、息止め、発話の重なりを分析した。その結果、沈黙区間だけでは捉えにくい会話に割り込みたいという意思や、発話を準備したが、実際には発話ができなかった状態を読み取れる可能性が示されている^[7]。このことは、発話欲求が音声として表出される以前から、身体内部では発話に向けた生体的な反応が生起していることを示唆している。

発話開始前には喉頭筋群や呼吸補助筋群が活動し、発声に向けた呼吸調整や声帯位置の制御が行われることが報告されているが^{[8],[9]}、これらの生体反応は発声の欲求に伴う身体が発声準備によって生じるものであると考えられている。また、頸部に装着したデバイスを用いて喉の筋肉の微細な振動や収縮、頸動脈波の信号から発話内容を推定し、脳卒中後の構音障害患者のコミュニケーション支援を行う ReVoice が提案されており^[12]、頸部が発話に関する情報を取得可能な計測部位であることが示されている。

一般に、筋活動が生じる際には酸素消費量が増加し、それに伴って局所的な血流量や酸素化状態が変化することが知られている。喉頭筋を収縮させた際には声帯組織の酸素分圧が変化することが報告されており^[13]、発声に関与する筋群の活動が局所的な代謝変化を伴うことが示されている。さらに、発話時には組織の血流量および酸素化状態が変化することも報告されている^[14]。これらの知見から、発話準備段階において活動する喉頭筋群や呼吸補助筋群においても、筋活動に伴う局所的な血流変化が生じる可能性が考えられる。したがって、頸部における血流変化を計測することで、発話欲求に起因する発話準備段階における生体反応を取得できる可能性がある。

一方で、既存研究では、頸部を発話情報の取得部位として用いる研究はあるものの、頸部の血流変化に着目し、発話欲求はあるが発話機会を得られていない状態を推定する試みは十分に検討されていない。そこで本企画では、XB-01 を用いて発話直前に生じる身体内部の血流変化と加速度情報を取得することで、発話欲求の検出ができるかを検討する。特に血流変化は、姿勢や外見的な動作とは異なり、発話に至る直前の筋活動や生体反応といった内部状態を反映する可能性がある点に特徴がある。そのため、視線や表情、姿勢といった従来の手法では捉えにくい潜在的な発話準備状態を推定できることが期待される。

3. 実装計画

本企画では、XB-01 から取得される血流変化および加速度情報を用いて、発話欲求に伴う生体反応を検出し、対面環境における会話参加を支援するシステムを

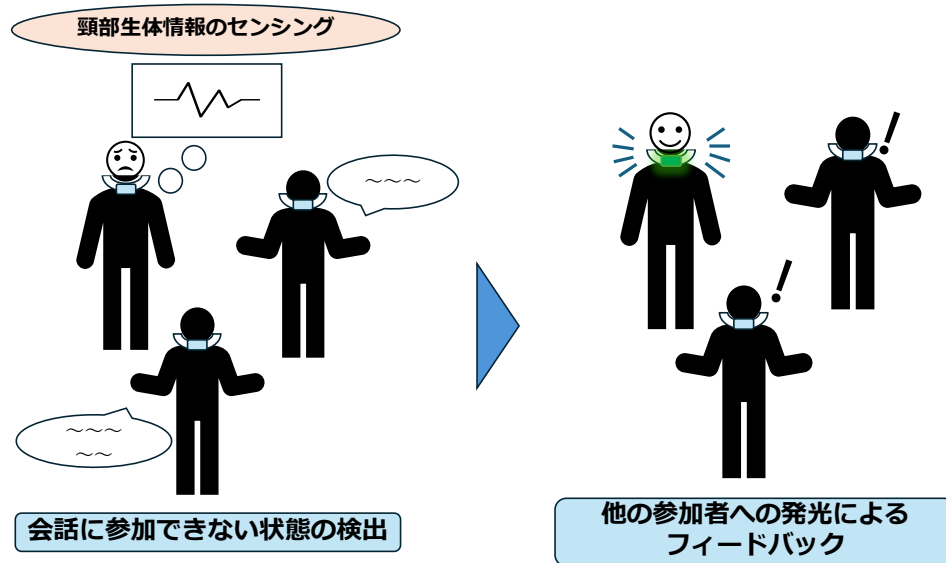


図1 提案システム利用イメージ図

提案する．1章で述べたように，実装の初期段階では前頭前野の血流変化を測定対象とするが，最終的には頸部を計測部位としたシステムの実現を目指す．本節では，最終的に実現を目指す頸部装着型デバイスの実装計画について述べる．展示では，複数人による対面会話を対象とし，会話に参加したいという意思を持ちながらも，割り込むタイミングをつかめずにいる体験者の状態を推定する．推定された状態を光によって周囲の参加者に提示することで，発話欲求を持ちながら発話に至れていない状況を共有する体験を構築する．体験者の発話や呼吸を妨げずに生体情報を取得するため，デバイスはバンドを用いて頸部に装着する．また，対面環境において状態を視覚的に提示するため，バンドの前面に小型LEDを組み込む．

本システムでは，信号処理と閾値判定を組み合わせたルールベースの推定手法を採用する．取得したデータに対してリアルタイムにフィルタ処理を行い，加速度センサから得られる顎の動きや深呼吸に伴う加速度変化と，NIRSセンサから得られる血流変化を特徴量として抽出する．これらの特徴量が事前に設定した閾値を同時に満たした場合，システムはその状態を「発話欲求に伴う生体反応が生じた可能性のある状態」として検出する．ただし，同様の生体反応は通常の発話開始時にも生じる可能性がある．そこで，本システムでは，生体反応を検出してから実際の発話が始まるまでの経過時間を計測する．発話が生じないまま一定時間が経過した場合に，発話欲求を持ちながらも発話に至れていない状態であると判定し，フィードバックを作動させる．実際の発話の有無については，外部マイクなどの別機器を使用せず，頸部に装着した加速

度センサが捉える発声時の喉の振動から推定する．

対話を支援するフィードバックについて，システムから音や言葉で明示的に通知を行うと，体験者にプレッシャーを与えるだけでなく，システムが誤検知を起こした際の不満を高めてしまうことが指摘されている^[5]．この知見を踏まえ，本企画では会話の進行を強制的に遮る介入を避け，光によるさりげない通知手法を採用する．

発話欲求を持ちながらも発話に至れていない状態が判定されると，その体験者の頸部バンドに搭載されたLEDが自動的に点灯する．この仕組みにより，誤検知時のリスクを抑えながら，会話のテンポを妨げない通知を実現する．また，体験者の心理的負担の軽減を図りながら，周囲の視覚へ緩やかに働きかけることで，対話の円滑化を目指す．

4. 体験プラン

本デモブースでは，来訪者に対してポスターやスライドを用いて，対面の会話やオンライン会議において，発話欲求を持ちながらも発話に至れていない参加者がいるという課題について説明する．合わせて，発話欲求を推定し，その状態を本人以外の会話参加者へ光で提示するシステムの仕組みと企画の目的を説明する．その後，対話の機会を得にくい状況と，発話欲求が周囲に提示される状況を体験してもらう．図1に体験のイメージを示す．体験者はバンドを用いてデバイスを頸部に装着し，希望する体験形式に応じて，以下2つのシナリオの一方を選択する．

4.1 体験シナリオ1

体験シナリオ1では，スタッフ1名が発表者役を務め，別のスタッフ1名と体験者が聴衆役として参加

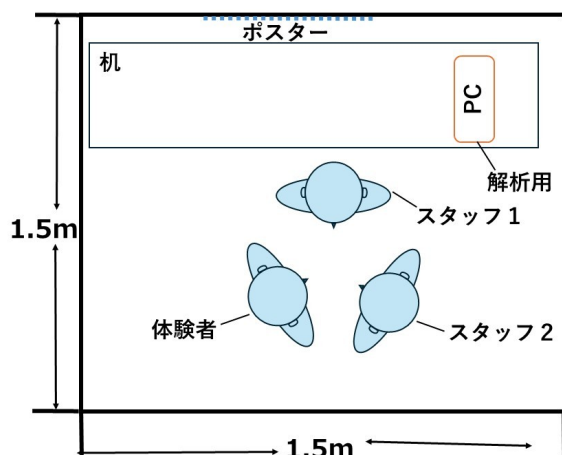


図2 フロアプラン

する。発表者は、体験者が内容を理解しやすい身近なテーマについて短い模擬プレゼンテーションを行う。体験者には発表を聞きながら質問や意見を考えもらう。聴衆役のスタッフが続けて質問を行い、発表者とのやり取りが継続する状況をつくる。これにより、体験者が質問や意見を持っていたとしても、発話を開始するタイミングをつかみにくい状況を再現する。システムが体験者の発話欲求を推定し、その状態が一定時間継続すると、体験者の頸部に装着したLEDが点灯する。発表者はその光を手がかりとして、体験者に「何か質問や意見はありますか」と発話を促す。これにより、発話機会につながる一連の仕組みを体験できる。

4.2 体験シナリオ2

体験シナリオ2では、体験者と会話相手役のスタッフ2名の計3名で、旅行や趣味などの身近なテーマについて会話を行う。体験中、スタッフ2名は1つの話題について一定時間やり取りを続ける。これにより、体験者が話題に参加したいと思っていても、発話を開始するタイミングをつかみにくい状況を再現する。システムによって体験者の発話欲求を推定し、その状態が一定時間継続すると、体験者の頸部に装着したLEDが点灯する。スタッフはLEDを手がかりとして会話に間を設けたり、体験者に会話を振る。これにより、発話機会の調整につながる様子を体験できる。

4.3 フロアプラン

図2にフロアプランを示す。本企画では、1.5 m×1.5 mのスペースを想定する。ブース内の長机には解析用PCなどの機材を配置する。体験デモにおいては、発表形式と雑談形式のいずれのシナリオを実施する場合でも、この長机の前のスペースに体験者1名とスタッフ2名の計3名が立ち、互いに顔を合わせるようにして会話を行うスタイルを採用する。

参考文献

- [1] Gerber, M.: Equal Partners in Dialogue? Participation Equality in a Transnational Deliberative Poll (Europolis); Political Studies, **Vol. 63**, No. 1 Suppl., pp. 110-130 (2015).
- [2] DiMicco, J. M., Hollenbach, K. J., Pandolfo, A. and Bender, W.: The Impact of Increased Awareness While Face-to-Face; Human-Computer Interaction, **Vol. 22**, No. 12, pp. 47-96 (2007).
- [3] Sacks, H., Schegloff, E. A., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. Language; **Vol. 50**, No. 4, pp. 696-735 (1974).
- [4] 敷田, 増田: 分散環境における話者交替のウェアネス支援; 情報処理学会論文誌, **Vol. 56**, No. 1 (2015).
- [5] 玉木, 東野, 小林, 井原, 岡田: 遠隔会議における発話衝突軽減手法; 情報処理学会論文誌, **Vol. 53**, No. 7, pp. 1797-1806 (2012).
- [6] 岡田: 新型コロナウイルス感染症拡大に伴う学生等の生活・メンタルヘルス・学習状況等に関する調査結果(速報版); 国立大学法人筑波大学働く人への心理支援開発研究センター, <https://www.human.tsukuba.ac.jp/counseling/w/wp-content/uploads/2020/10/surveyresult20201002.pdf> (2020), (参照 2026-06-13).
- [7] Włodarczyk, M., Heldner, M.: Breathing in Conversation; Frontiers in Psychology, **Vol. 11**, 575566 (2020).
- [8] Tremoureaux, L., Raux, M., Ranohavimparany, A., Morélot-Panzini, C., Pouget, P. and Similowski, T.: Electroencephalographic evidence for a respiratory-related cortical activity specific of the preparation of prephonatory breaths; Respiratory physiology & neurobiology, **Vol. 204**, pp.64-70 (2014).
- [9] Travis, L., Shiba, M.D., Dinesh, K. Chhetri, M. D.: Dynamics of phonatory posturing at phonation onset; The Laryngoscope, **Vol. 126**, No. 8, pp. 1837-1843 (2015).
- [10] 寺尾, 米村: 話者の姿勢から発話予備動作を特定するオンライン会議支援システムの検討—圧力センサーを用いた発話予備動作の検出精度に関する実験的検証—; 信学技報, **Vol. 121**, No. 363, HCS2021-54, pp. 65-70 (2022).
- [11] Yoshida, S., Iijima, R., Hashimoto, A., Ma, J.: FairTalk: Facilitating Balanced Participation in Video Conferencing by Implicit Visualization of Predicted Turn-Grabbing intention.; arXiv:2025.20138 (2025).
- [12] Kadam, S., Jikamade, A., Mattoo, P., Hole, V.: ReVoice: A Neural Network based Voice Cloning System; 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), pp. 1-6 (2024).
- [13] Tomita, M., Matsuo, K., Maehara, N., Umezaki, T., Takemoto, S.: Measurements of Oxygen Pressure in the Vocal Fold During Laryngeal Nerve Stimulation; Archives of Otolaryngology-Head & Neck Surgery, **Vol. 114**, No. 3, pp. 308-312 (1988).
- [14] Wolf, U., Scholkmann, F., Rosenberger, R., Wolf, M., Nelle, M.: Changes in Hemodynamics and Tissue Oxygenation Saturation in the Brain and Skeletal Muscle Induced by Speech Therapy – A Near-Infrared Spectroscopy Study; The Scientific World Journal, **Vol. 11**, No. 1, pp. 1206-1215 (2011).