



AI時代における信頼できる音声入力の設計と産学連携によるイノベーション —技術変化の兆しを捉えたmask voice clipとハードウェア起点の価値創出—

株式会社村田製作所 技術・事業開発本部 技術イノベーション戦略部
生体エレクトロニクス備えプロジェクト課 マネージャー

もりた まさのり
森田 暢謙



1. エグゼクティブサマリー

生成AIの普及により、人とコンピュータの関係は大きな転換点を迎えている。操作の対象としてのコンピュータから、対話を通じて協働する存在へと役割が変わりつつある中で、注目を集めているのがAIエージェントである。AIエージェントとは、自然言語で指示を受け取り、状況や文脈を踏まえて多様なタスクを自律的に遂行するインテリジェントシステムである^[1]。AIの基盤モデルのコモディティ化が進むほど、その能力を左右する要因はモデル性能そのものではなく、どのような入力データを、どの程度信頼できる形で扱えるかに移っていく。

音声は、人間にとって最も自然なコミュニケーション手段の1つであり、AIエージェントがユーザーの意図や状況を理解するための有力な手掛かりとなる。一方で、騒音や第三者の声の混入、音声認識精度のばらつき、プライバシーへの配慮といった課題により、入力データとしての信頼性を確保しにくいという問題を抱えている。

本稿では、村田製作所の圧電フィルムセンサPicoleafTM^[2]を用いてマスク表面の振動を検出する装着型デバイスmask voice clipを題材として、AIエージェント時代に求められる信頼性の高いデータの入口設計を具体化した事例を示す。とりわけ、産学を往復して変化の兆しを捉えて、研究の問いを実装へと結び付けてきた過程を踏まえ、ハードウェア起点でデータが生まれる入口を押さえることの意義を考察する。

2. 導入：「信頼できる音声入力」という論点

ユーザーインターフェース（UI）は、紙と鉛筆からパーソナルコンピューター、スマートフォンへと進化し、指による画面の直接操作と情報への常時接続環境をもたらしている。大規模言語モデル（LLM）の登場によって、UIはユーザーが直接操作するものから対話を通して協働するものへ軸足を移しつつある^[3]。次世代のUIでは、ユーザーによる逐次的な操作を前提とせず、入力された指示や情報をAIエージェントが文脈情報として解釈し、適切な行動へと展開する能力が重要となる。

その入口として有力なものが音声入力である。人間のコ

ミュニケーションは本質的に「話す」という行為を中心に設計されており、AIエージェントがユーザーの環境や文脈を捉える上でも音声データは有効な手掛かりとなる。音声入力は既に多くの場面で利用されているが、利用環境によっては精度や取扱いに配慮を要する。周囲の騒音や第三者の声の混入により、話者本人の発話を明確に捉えにくい場合がある。また、周囲の会話が意図せず含まれる点は、プライバシーや情報管理の観点から留意が必要である。こうした要因はAIエージェントの判断や実用性を左右するため、「信頼できる音声入力」をいかに実現するかが重要となる。本稿でいう「信頼できる音声入力」とは、(1) 話者本人の発話のみが取得され、(2) 環境音や第三者の発話の混入を抑制し、(3) その後段の音声認識に耐えうる品質と再現性を備えた入力を指す。

3. 時代背景：AI時代の音声入力と価値の移行

Marc Andreessenが2011年に“Software is eating the world”^[4]と表現したように、ソフトウェアが価値創出の中心となる潮流は既に広く認識されてきた。現在はその次の段階として、“Data is eating the world”^[5]とも言うべき局面にきている。差別化はソフトウェアではなく、文脈データの継続獲得と活用に移りつつあり、会話や行動のデータを信頼できる形で蓄積し意思決定や行動に結び付けられるかが競争力を決める。AIエージェントは、文脈データを継続的に取得できることを前提に判断・実行するため、データが競争力を左右する時代への転換を映し出している。

その方向性を先取りする研究として、MIT Media Labが2024年に発表したMemoro^[6]が挙げられる。Memoroは骨伝導ヘッドセットを通じて周囲の会話や状況を取り込み、必要な情報を適切なタイミングで本人に提示することで、本人の記憶の想起や判断を支援する。ここで重要なのは、音声単なる操作の入力ではなく、AIが文脈を理解し、記憶を形成するための主要なデータになっている点である。

こうした研究と並行して、産業界でもパーソナルなAIエージェントの構築を前提にした動きが加速している。OpenAIのサム・アルトマンは、次のAIの前進は推論の高度化よりも、ユーザーの会話や判断履歴を踏まえて振る舞える完全な記

憶にあるという趣旨を語っている^[7]。また、Metaは2025年12月に会話の記録・要約をするペンダント型デバイスを手掛けたLimitlessを買収しており、日常の会話データを起点に文脈の理解と記憶を拡張する方向性を具体化しつつある^[8]。

これらの動きが示すのは、パーソナルなAIエージェントを実装する上で、会話データや音声による指示に関する情報を、継続的にかつ信頼できる形で取得できる手段が求められているという点である。こうしたデータが蓄積されることで、AIエージェントは単なるツールを超え、ユーザーの生活を支えるパートナーへと近づいていく。

この流れの中で、価値の源泉はハードウェアやソフトウェアから、信頼できるデータへ移行しつつある。AI時代に音声入力が注目されるのは、音声人が人の思考や状況、会話の流れを自然に含むデータであり、適切な仕組みを通じてAIエージェントの判断や行動に有効に結び付けられるためである。

4. 技術の本質：mask voice clipは何を解いたか

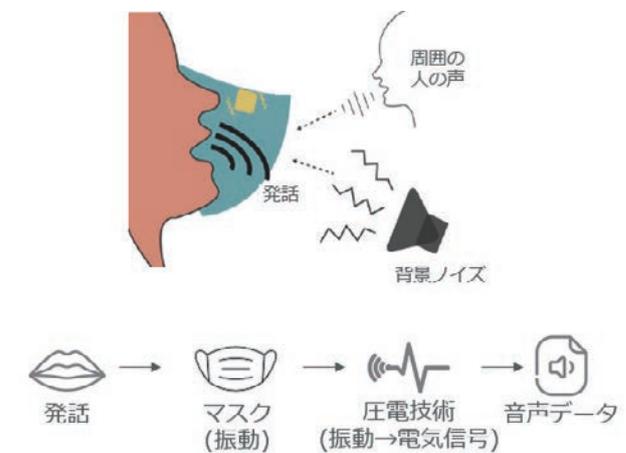
mask voice clipが解いたのは、単純な騒音環境下における音声認識精度の向上といったアルゴリズム上の問題ではない。本質は、従来のマイクロフォンが前提としてきた「空气中を伝搬する音を取得する」という構造そのものを見直し、入力の入口をハードウェア起点で再設計した点にある。

従来の音声入力、騒音や複数話者が混在する環境下で発話者の音声を安定して取り出しにくい。これに対し、指向性マイクロフォンやビームフォーミング、ソフトウェアによる音源分離など、様々な音声処理手法が提案されてきた^{[9] [10] [11]}。これらは一定の効果を持つものの、いずれも空気の振動を取得するという前提に立っており、複数の音声と環境音が混在する状況では、後処理による分離に限界がある。特に、深層学習に基づく高度な音声分離は大きな計算資源を要する傾向があり、エッジデバイスや組み込み環境でのリアルタイム利用には制約が残る。

mask voice clipが示したのは、この前提そのものを見直すアプローチである。本デバイスは、空气中の音を取得するのではなく、マスク装着者の発話に伴ってマスク表面に生じる微細な振動を直接検出する。マスク表面の振動は話者本人の発話に強く対応し、周囲の騒音や他人の声による影響を受けにくい。つまり、データを後処理で分離するのではなく、最初から混ざりにくい形で取得するという設計思想である。

この振動検出には、村田製作所の圧電フィルムセンサ

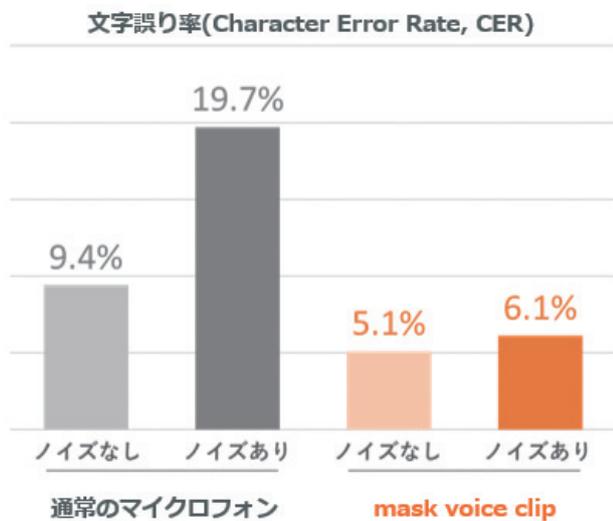
Picoleaf™を用いている。圧電材料は振動によって生じる機械的応力を電気信号に変換することができる。圧電性を有する有機フィルムで構成されたPicoleaf™は柔軟性が高く、マスクのような曲面を持つ柔らかい素材にも違和感なく配置できる。mask voice clipでは、発話に伴って生じるマスク表面の微小振動をPicoleaf™で検出し、電気信号へ変換した上で、増幅・デジタル化することで音声データとして扱う（図1）。あわせて、デバイスをクリップ型とすることでマスク本体から切り離すことができ、衛生面と現場運用性を両立した。



■ 図1. mask voice clipの音声検出メカニズム

性能評価では、オフィスや工場などの作業環境を想定し、背景騒音に加えて複数話者の発話が同時に存在する条件を構築した。その上で、話者の発話に対する音声認識結果と正解テキストを比較し、文字誤り率（CER）を算出した。その結果、mask voice clipのCERは6.1%（無雑音下5.1%）であったのに対し、通常マイクロフォンでは19.7%（同9.4%）と大きな差が確認された（図2）^[12]。mask voice clipは騒音環境下における性能劣化が小さく、入力データの信頼性を環境条件に左右されにくいことを示している。

mask voice clipの本質は、空气中の音を取得する限り避けられない構造的課題を、マスク表面の物理振動を直接検出することで回避し、入力データの信頼性を確保した点にある。AIエージェントの時代において入力の品質がデータの価値に影響するという着眼点を起点に、その課題に対する具体的なソリューションとして提案された技術事例である。



■図2. 音声認識性能の比較

5. 開発の舞台裏：音声入力・サイレントスピーチ研究という未踏領域に踏み出せた理由

暦本純一教授（東京大学大学院情報学環）は、人間とコンピュータの関係を捉え直し、計算機による能力拡張を志向したHCI（Human-Computer Interaction）の研究を進めている。研究領域は多岐にわたるが、その1つに「サイレントスピーチ（無音発話）」がある^[13]。これは、声としての音を前提とせず、口の動きや生体信号から発話意図を推定し、入力やコミュニケーションにつなげる研究である。

声を出さずにつぶやくだけでコンピュータへの入力が成立するという発想は、一見すると未来的な技術に映るかもしれない。我々にとっても、この研究の意義や価値について最初から明確に理解できるものではなかった。しかし、調査や暦本教授とのディスカッション、更に生成AIが急速に進化していく現状を踏まえる中で、本研究が技術進化の中でどのような価値を持ち、なぜ必要とされるのかが、次第に明確になっていった。

我々は、サイレントスピーチ研究を音声入力の重要性が増していく技術変化の兆しの1つとして位置付け、音声入力を注目すべき技術領域として定めた。AIの進展により人とAIエージェントが対話的に協働する未来が現実化する中で、信頼性の高い音声入力が基盤技術になると考えた。またその探索過程で、自社の圧電技術がこの要件に応え得る技術的特性を備えているという仮説に至った。

村田製作所のPicoleaf™は、これまでスマートフォン等で振動やタッチ入力を検知する用途で用いられてきたが、発話に伴う微細な振動を捉えて音声入力に応用する試みはな

かった。「音声認識に必要な情報量を十分に取得できるのか」という問いから始まり、社内のエンジニアと試作を行い、研究室での評価を踏まえて要件を更新し、改善を重ねてきた。その往復を通じて、実装要件と性能要件のすり合わせを進めた。この過程で、音声に起因する微小振動に合わせてセンサ特性の改良・調整を進め、音声の帯域を捉えられる見通しが得られ、正式に共同研究を開始した。キックオフでは研究室の学生も交えてアイディエーションを行い、複数のフォームファクターを検討した。平城裕隆氏（当時博士課程3年）らの過去検討も踏まえ^[14]、マスクへの搭載検討を進めた結果、mask voice clipとして新たな音声入力デバイスの機能が発現し、論文化に至った。本研究成果は、平城氏を中心として複数の論文及び学会発表として発信され^[8]^[15]^[16]、学生奨励賞の受賞を含む形で国内外から高い評価を得ている^[17]。

mask voice clipは無音発話を目的としたサイレントスピーチではなく、音声を伴う入力デバイスである。本研究の価値は、サイレントスピーチ研究を含む音声入力の研究を手掛かりに、音声入力の重要性が高まる兆しを捉え、その要求を技術要件へと具体化し、試作と評価を通じて実装に結び付けた点にある。

また、村田製作所ではマスク装着に依存しないデバイスへの応用やサイレントスピーチへの展開可能性についても検討を進めており、CEATEC AWARD 2025 審査委員会からは、こうした将来展開を含めた研究の広がりに対する期待が示された^[18]。

6. 産学を往復する探索の力とイノベーション：「橋渡し役」が果たす役割

AIの基盤モデルや開発ツールの汎用化により、研究成果を社会実装へとつなぐスピードは大きく加速している。とりわけHCI領域では、体験仮説を試作と評価を短いサイクルで反復して評価・検証を行うアプローチをとるため、産学の距離が縮まりやすい。この領域では、完成度の高い解を最初から求めるよりも、仮説検証を通じて方向性を見極めていく姿勢が重要となる。初期段階の不確実性を受け入れ、試作・評価・学習を重ねることで、従来の延長にない領域へと探索を広げていくことが可能となる。

このとき鍵となるのが、産学を往復しながら、研究の発想を実装可能な設計課題に変換する「橋渡し役」である。アカデミア、とりわけHCI分野における問いは、「どのような体験や関係性が成立し得るか」という将来像を描き、そ

こからインタラクションや技術要件を遡って設計するバックキャスト的な要素を持つことが多い。一方で、その問いは企業が普段から慣れ親しんだ具体的な制約条件とは必ずしも一致せず、そのままでは噛み合わないことが多い。

橋渡し役が担うのは、アカデミアの問いに内在する意義や将来ビジョンを読み解き、それが社会実装されたときの価値とインパクトを、企業が意思決定できる形へと翻訳することである。あわせて、その構想から実現可能な技術要件へ落とし込み、開発・知財といった関係者を横断して、試作と評価が回る探索の場を構築する。さらに、そこで得られた知見をアカデミアへ還流し、企業側の文脈と結び付けながら問いを更新していく。いわゆる“Boundary Spanner”とは、単なる調整役ではなく、産学の往復を通じて変化の兆しをいち早く捉え、探索の仕掛けどころと技術開発の方向性を見極める存在である。

こうした産学連携の在り方は、特定の研究領域に限られた話ではない。技術の潮流が急速に変化する時代において重要なのは、変化が顕在化してから追従することではなく、研究の現場に現れる兆しを手掛かりに、次の価値創出の起点を先に仕込むことである。

7. 製造業が見据えるべき次の起点：兆しを先読みする産学連携とハードウェアからの価値転換

AI時代の競争力は、基盤モデルの性能だけで決まるものではない。価値の源泉がハードウェアからソフトウェア、そしてデータへ移る局面では、信頼できるデータを継続的に獲得できるかが、ソリューションの実用性と差別化を左右する。日本の製造業がこれからの時代に向き合う上で、重要な観点は次の2点にまとめられる。

第1に、アカデミアとの連携を単発の共同研究ではなく、探索フェーズを継続的に回す構造として強化することである。研究の現場に現れる変化の兆しを捉え、潮流を先読みし、次の手を仕掛けていくには、産学を往復し、研究の問いを実装可能な形へと翻訳する役割が欠かせない。AI・デジタル領域の競争力が、成果物の移転そのものよりも探索の継続性に左右されることは、みずほ銀行産業調査部「みずほ産業調査79号『日本産業の中期見通し』」でも指摘されている^[19]。不確実性の高い段階から試作・評価・対話を重ねられるかどうか、結果として技術潮流の早期把握と、時代に合った技術開発の方向性の決定につながる。

第2に、センサやデバイス、インフラストラクチャといったハードウェアを起点に、データが生まれる入口を設計し、

継続的なデータ獲得につなげることである。価値の源泉がデータへ移るほど、信頼できるデータを蓄積できなければ、競争力を高めていくことは難しい。また、ハードウェアの供給までにとどまれば、付加価値はアプリケーションやサービスを提供する側に帰属しやすく、価値配分で不利になりやすい。こうした構造を踏まえ、ハードウェアを起点にデータ活用を含む価値提供まで視野に入れる重要性は、経済産業省のデジタル経済レポート「データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」^[5]でも指摘されている。

以上の議論から、AIエージェント時代における競争力は、モデルや個別技術の優劣だけでなく、研究の兆しを捉えながら探索を継続できる体制構築や、データが生まれる入口をどのように設計できるかといった点に大きく左右されることが示唆される。日本の製造業においては、産学連携の進め方やハードウェアの役割を改めて捉え直し、自社の強みをどの領域で発揮できるのかを見極めていくことが、今後ますます重要になる。

8. 結論：入力設計から始まるデータ価値

mask voice clipは、本稿で述べてきた2つの論点——アカデミアの研究から技術変化の兆しを捉えること、そしてハードウェア起点で信頼できるデータの入口を設計すること——を具体の形に落とし込んだ取組みである。サイレントスピーチ研究が投げかけた問いを出発点に、将来像を描きながら技術要件へと落とし込み、試作と評価を往復する過程を通じて、未踏領域への一歩が具体的な形になった。さらに、入力的设计がデータの信頼性を左右し、その先の価値創出に影響することを示す事例でもある。本稿で述べたのは、その第一歩である。

参考文献

- [1] Bill Gates 「AI-powered agents are the future of computing」 (2023)
<https://www.gatesnotes.com/ai-agents>
- [2] 村田製作所 圧電フィルムセンサ (Picoleaf™)
<https://www.murata.com/ja-jp/products/sensor/picoleaf>
- [3] 文部科学省 有識者検討会 (第3回) 配付資料/安宅和人氏提出「これからの人材育成を考える」 (2023)
https://www.mext.go.jp/content/20230328-mxt_kyoiku01-000028726_02.pdf
- [4] Andreessen Horowitz 「Why Software Is Eating the World」 (2011)



- <https://a16z.com/why-software-is-eating-the-world/>
- [5] 経済産業省 若手新政策プロジェクトPIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」(2025)
https://www.meti.go.jp/policy/it_policy/statistics/digital_economy_report.html
- [6] W. D. Zulfikar, S. Chan, P. Maes. 2024. Memoro : Using Large Language Models to Realize a Concise Interface for Real-Time Memory Augmentation. CHI '24.
<https://doi.org/10.1145/3613904.3642450>
- [7] Alex Kantrowitz「Sam Altman : How OpenAI Wins, ChatGPT's Future, AI Buildout Logic, IPO in 2026?」(2025)
<https://www.youtube.com/watch?v=2P27Ef-LLuQ>
- [8] TechCrunch「Meta acquires AI device startup Limitless」(2025)
<https://techcrunch.com/2025/12/05/meta-acquires-ai-device-startup-limitless/>
- [9] Benjamin B. Bauer. 1962. A century of microphones. Proc. IRE, 50 (5), 719–729.
- [10] J. Heymann, L. Drude, R. Haeb-Umbach. 2016. Neural network based spectral mask estimation for acoustic beamforming. ICASSP.
<https://dl.acm.org/doi/10.1109/ICASSP.2016.7471664>
- [11] C. Subakan et al. 2023. Exploring self-attention mechanisms for speech separation. IEEE/ACM TASLP, 31, 2169–2180.
<https://dl.acm.org/doi/pdf/10.1109/TASLP.2023.3282097>
- [12] Hirotaka Hiraki, Jun Rekimoto. 2025. MaskClip : Detachable Clip-on Piezoelectric Sensing of Mask Surface Vibrations for Real-time Noise-Robust Speech Input. AHS2025.
<https://dl.acm.org/doi/10.1145/3745900.3746088>
※本文中では「mask voice clip」と記載するが、論文タイトルにおける正式名称は「Mask Clip」である。
- [13] Internet of Brains「無言のコミュニケーションとAIで社会が変わる ～サイレントスピーチが実現するAIと人間の融合～」(2024) <https://brains.link/activity/3318>
- [14] Hirotaka Hiraki et al. 2024. WhisperMask : a noise suppressive mask-type microphone for whisper speech. AHS '24.
<https://doi.org/10.1145/3652920.3652925>
- [15] Hirotaka Hiraki, Jun Rekimoto. 2024. Piezoelectric Sensing of Mask Surface Waves for Noise-Suppressive Speech Input. UIST Adjunct '24.
<https://dl.acm.org/doi/10.1145/3672539.3686331>
- [16] 「マスク表面の振動計測による着脱可能なリアルタイム音声強調手法の提案」第84回UBI研究会
<https://ipsj.ixsq.nii.ac.jp/records/240579>
- [17] 第84回UBI研究会 学生奨励賞
<https://sigubi.ipsj.or.jp/excellent/>
- [18] 一般社団法人電子情報技術産業協会JEITA「CEATEC AWARD 2025 総務大臣賞・経済産業大臣賞・デジタル大臣賞 新設のモビリティ賞を含む部門賞がそれぞれ決定」(2025)
<https://www.jeita.or.jp/japanese/topics/2025/1007award.pdf>
- [19] みずほ銀行「みずほ産業調査 Vol.79 | 日本産業の中期見通し 一向こう5年(2026–2030年)の需給動向と求められる事業戦略」(2025)
<https://www.mizuhobank.co.jp/corporate/industry/sangyou/pdf/1079.pdf>

ITUが注目しているホットトピックス

ITUのホームページでは、その時々ホットトピックスを“NEWS AND VIEWS”として掲載しています。まさに開催中の会合における合意事項、ITUが公開しているICT関連ツールキットの紹介等、旬なテーマを知ることができます。ぜひご覧ください。

<https://www.itu.int/en/Pages/default.aspx>