AI主導の対話を高SNR MEMSマイクロフォンにより実現

Dr. Gunar Lorenz、インフィニオン テクノロジーズ

はじめに

インフィニオンでは、長年にわたり、優れたオーディオ技術が民生機器のユーザー体験を向上させるのに不可欠だと信じてきました。アクティブノイズキャンセリング、外音取り込み機能、スタジオ録音、ビームフォーミング、その他関連技術において目覚ましい進歩をもたすため、様々なことに取り組んできました。インフィニオンは、XENSIV™ MEMSマイクロフォンの主要サプライヤーとして、MEMSマイクロフォンの音質向上を目指してきました。これにより、完全ワイヤレスイヤホン (TWS) やオーバーイヤーヘッドホン、ノートパソコン、タブレット、会議システム、スマートフォン、スマートスピーカー、補聴器、さらには自動車に至るまで、さまざまな民生機器において優れたユーザー体験を実現しています。

今日、私たちはAIが日常生活に革命をもたらし、ChatGPTのようなツールが直感的なテキストや音声によるやり取りを通して、生産性の概念を再定義する、刺激的な時代に生きています。AIを活用したシステムが進化を続けるなか、従来のビジネスモデルや 信念、そして前提が揺らいでいます。新たに形成されつつあるAIエコシステムにおいて、音声はどのような役割を果たすのでしょうか？ 私たちはビジネスリーダーとして、これまでの信念を見直す必要があるのでしょうか？ 生成AIの台頭によって、高品質な音声入力の重要性は低下してしまうのでしょうか、それとも高品質な音声入力こそがAIベースのサービスやパーソナルアシスタントの普及に欠かせないものとなるのでしょうか？

AI - 助けになるアシスタントから親友に

質問の内容だけでなく、その質問の仕方に合わせて、人が反応を変えるのは自然なことです。人間の声には、年齢、性別、社会的/文化的背景、そして、感情を判断するためのさまざまなヒントが含まれています。さらに、空港、オフィス、交通機関や、ランニングなどの身体的活動など、置かれている状況を認識することで、質問者の意図を把握し、それに応じて回答や会話を調整できるようになります。

AIの性能が大幅に向上したにもかかわらず、依然としてAIベースのアシスタントシステムは、人間の質問の意図や、特定のメッセージがどのように受け取られるのか正確に予測するのは難しいという認識が根付いています。人間と機械との対話を向上させるには、AIがどのような話し方や表現方法を選ぶのかにおいて、聞き手に関する知識、聞き手の感情、状況による文脈の変化という、3つの重要な要素を考慮しなければなりません。

多くの場合、必要な情報を抽出し、適切に応答するには音声を聞くだけで十分です。一例として、会ったことのない相手と通話や電話会議をする場面を考えてみてください。さらに、直接会う機会が一度もないまま、会話を繰り返すことで、人が相手に対して抱く認識がどのように変化していくか考えてみてください。

最近の研究によれば、AIの言語応答スタイルがわずかに変化するだけで、AIの社会的能力や人格に対する認識に、顕著な変化をもたらすことが示唆されています。将来のAIシステムは、適切なレベルで音声を入力すれば、効果的なコンパニオンとして機能でき、人間の友人のように質問をしたり、真摯に答えを聞いたり、時には静かに耳を傾けて判断を控えるといった行動ができると考えられます。

人間は音声をどのように聞いているのでしょうか？

言語によるあらゆるコミュニケーションと同じように、音声メッセージも言葉で感情や考えを伝えます。さらに、声のピッチ、速度、音量、環境ノイズといった他の要素も、音声全体の知覚に影響を及ぼします。

科学的な観点では、人間の耳は周波数と音圧レベルの2つの要素によって音声信号を感知します。音圧レベル (SPL) はデシベル (dB_{SPL}) で表され、周囲の大気圧を基準にした音圧の振幅を示します。100 dB_{SPL}の音圧レベルは、芝刈り機やヘリコプターなど、きわめて大きな音に相当します。

SPL範囲の最下点 (0 dB) は、20 µPaの音圧振幅と関連しており、これは良好な聴力を持つ若く健康な人の1 kHzにおける聴覚閾値に相当します。会話におけるあらゆる人間の声は、100 Hzから8 kHzの周波数帯域に属します。

図 1に、ISO 226:2023 ［2］ による人間の聴覚閾値を示します。

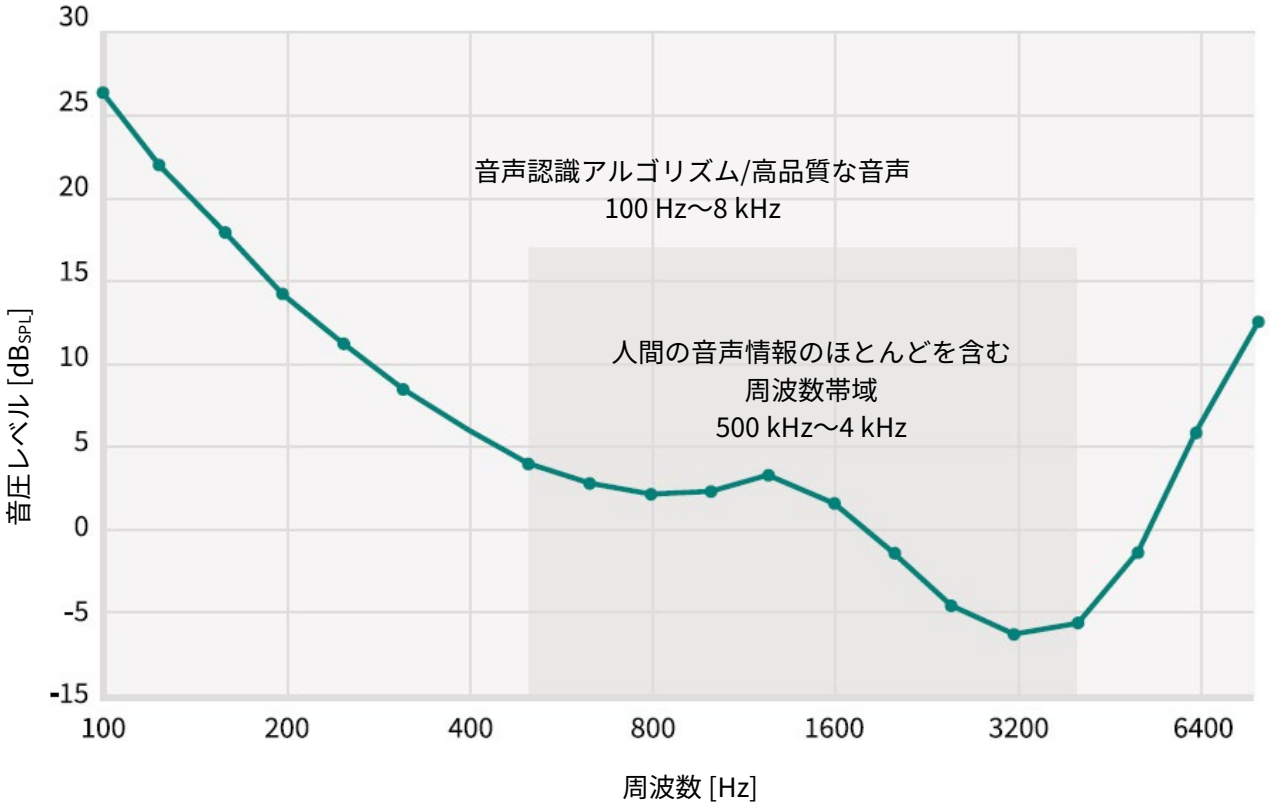


図 1: 音声閾値: 反復試験において被験者が応答の50%を正しく検出する音のレベル（ISO 226:2023）

図 1からわかるように、人間の聴覚はとりわけ500Hzから6kHzの範囲の周波数に対して敏感です。この周波数帯において、相対的に問題があると、声や楽器の聞こえる音の品質に重大な影響を及ぼす可能性があります。500Hzから4kHzの周波数帯には、人間の声から得られる情報の大部分が集中しており、声の明瞭度に影響を与え、特に2kHz帯域は重要です。

また、5kHzから10kHzの周波数は音楽が集中する帯域で、音に「生き生きとした感じ」や「明るさ」を感じさせます。しかし、これらの周波数に含まれる音声由来の情報は比較的少なく、「ship」「chip」「zip」といった単語の頭にあるような歯擦音のみです。

図 2に示すように、人間の聴覚閾値は年齢とともに低下します。

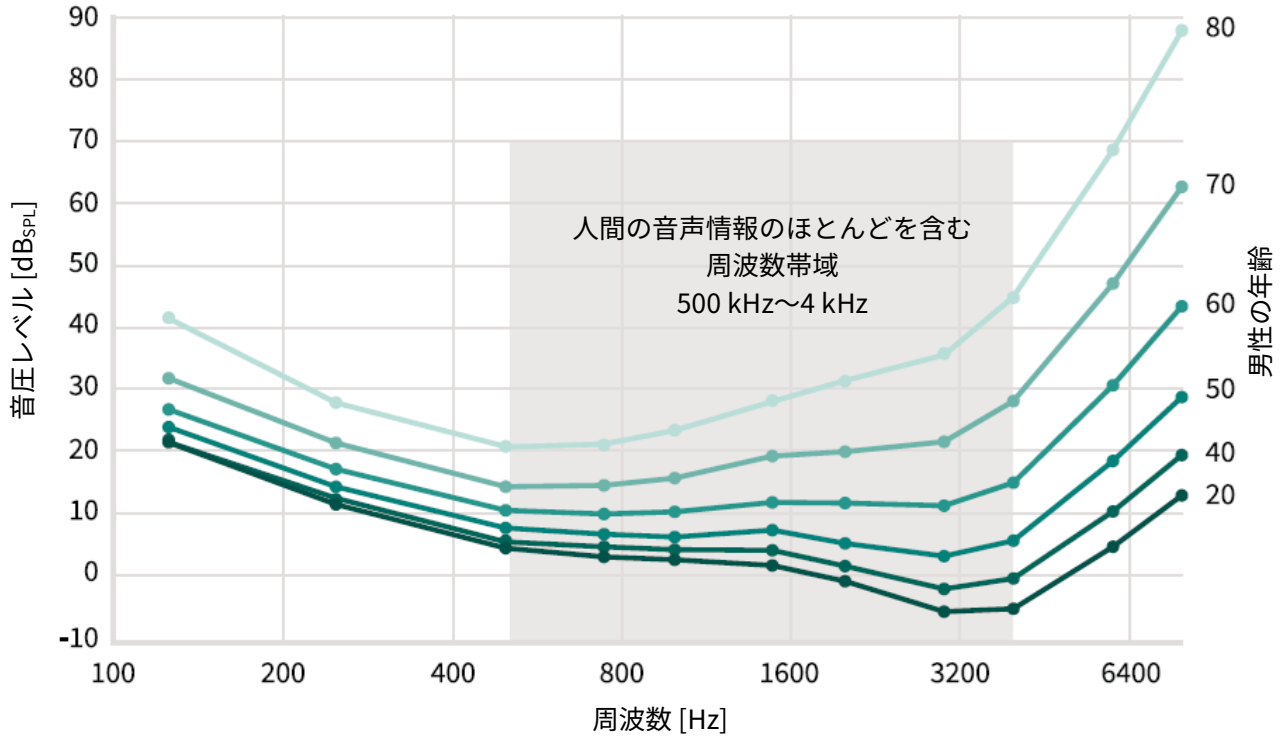


図 2: 正常な聴覚を持つ、幅広い年齢の男性が、モノラルイヤホンで音を聞いたときの聴力閾値の低下を示しています。なお、女性についても同様実験が行われており、年齢に伴う聴力低下がやや低いことが示されています (ISO 7029:2017)

40歳から50歳の間の多くの人が経験する軽度難聴でさえ、日常生活に重大な影響を及ぼす可能性があります。たとえば、軽度難聴の人は、騒がしい環境において、集団の中で会話についていくのは難しいかもしれません。 また、警告信号や警報といった重要な聴覚情報を聞き逃す可能性もあります。

現在のオーディオハードウェアは将来のAIに適応できるのでしょうか？

人間がどのように音声信号を知覚するのか理解できたところで、現在、そして将来のAIが、人間と区別がつかないほど高いレベルで機能するのに必要な音声入力の品質について、最初の質問に立ち返って考えてみましょう。

現在、市場に出回っている多くの消費者向け機器と同様に、音声信号はMEMSマイクロフォンを使って録音されます。MEMSマイクロフォンは、AI搭載アシスタントツールの主要な音声キャプチャ技術であり、現在、AIを用いたツールは市場で入手可能になりつつあります。

MEMSマイクロフォンによって得られる音声記録の品質は、マイクロフォンのダイナミックレンジに依存します。ダイナミックレンジの上限は最大入力音圧レベル (AOP) によって定義され、これは高い音圧レベルを入力した際のマイクロフォンの歪み特性を表します。自己雑音は、ダイナミックレンジの下限を表します。マイクロフォンの自己雑音に対する従来の指標は信号対雑音比 (SNR) ですが、これはマイクロフォンの自己雑音と、基準信号との比率を意味します。しかし、SNRの数値は、人間が音をどのように認識するかを前提として、A特性の重み付けをしているため、本議論の目的とは少し異なります。

録音された信号の受信者がAIである場合、マイクロフォンに関連するパラメーターの一つである「等価雑音レベル (ENL)」の方が、人間の聴覚に基づいた情報を無視するため、マイクロフォンの性能を評価するうえで、より適しています。ENLは、外部音源が存在しない状態でマイクが生成する信号を指します。

マイクロフォンの周波数帯域におけるENLは、マイクロフォンの聴覚閾値に最も近い値と考えられます。ただし、これは単純化されたものであり、通常、音声チャンネル、追加の防水対策、音声処理チェーンなど、他にも多くの要素があることに注意が必要です。

2つのMEMSマイクロフォンのENL曲線と人間の聴覚閾値との比較を視覚的に表している図 3をご参照ください。

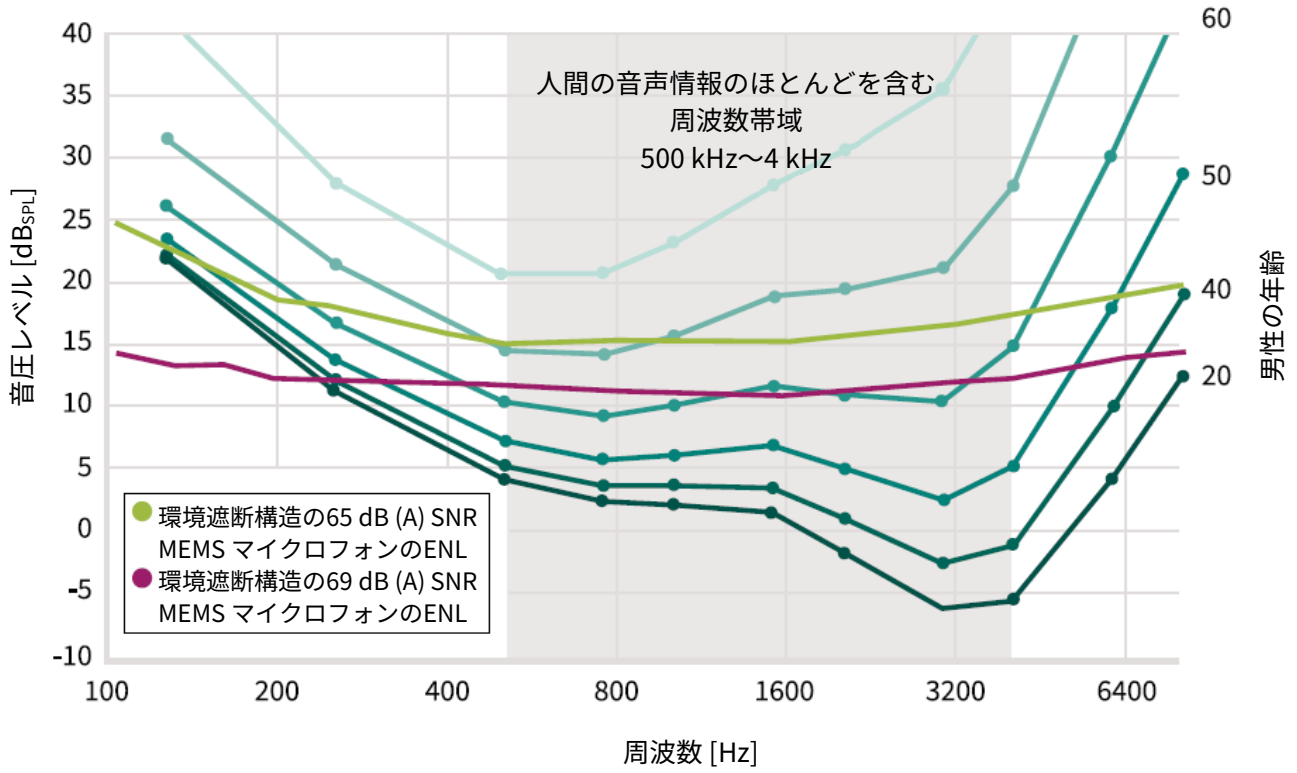


図 3: 中音域と高音域のMEMSマイクロフォンの音響1/3オクターブバンドの自己雑音 (ENL)
一般男性の聴覚閾値との比較結果

赤い線は、環境保護用ダストバリアを内蔵したSNR 65 dB (A) のマイクロフォンのENL曲線を示しています。これに該当するMEMSマイクロフォンは、現在、さまざまなメーカーのハイエンドモデルのスマートフォンに採用されています。

紫色の線は、インフィニオンの最新ハイエンドデジタルマイクロフォンのENL曲線を示しています。このマイクロフォンは革新的な環境バリアを採用し、粒子と湿気から保護しています。このマイクロフォンは現在の最先端技術を表し、2024年にハイエンドタブレットに搭載される形でリリースされました。2024年には、同様の性能のマイクロフォンを搭載したスマートフォンが数多くリリースされました。注目すべきは、マイクロフォンの自己雑音を5~10 dBSPL削減することで、音圧の対数スケールをでは重要な成果効果である点です。

インフィニオンはハイエンドMEMSマイクロフォンの自己雑音低減において著しい進歩を遂げましたが、人間の耳と比べると、低音圧レベルの検出に関しては、人間の耳には及びません。特に2kHz帯域は、人間の耳が音声をはっきりと認識する上で、きわめて重要です。若者の聴覚能力とインフィニオンの最先端マイクロフォンの性能には、12 dBSPLを超える差があります。更に、現在、ハイエンドのスマートフォンで使用されているマイクに比べると、17 dBSPLというきわめて大きな差があります。繰り返しになりますが、この評価はMEMSマイクロフォンの自己雑音のみを対象としており、他の雑音は考慮されていませんが、周辺雑音により、全体的な性能がさらに低下しかねません。

現時点でのMEMSマイクロフォンの限界は、人間の音声情報の大部分を含む周波数帯域 (500 Hz～4 kHz) において最も明らかです。現在、市場で最も高性能なMEMSマイクロフォンであっても、60歳の人の聴覚レベルに相当します。
公開されているデータによると、最新のMEMSマイクロフォンを使用したAIベースのバーチャルアシスタントは、騒がしい環境や遠距離での会話を行わなければならない状況においては特に、難聴を患っている高齢者と同等の聴覚能力になる可能性が高いと予測されます。

まとめと展望

AIの急速な進歩はその速度を落とすことなく、むしろ高SNR MEMSマイクロフォンの普及を更に加速させるでしょう。現時点で最新のMEMSマイクロフォンは人間の聴力にはいまだ及びませんが、インフィニオンがマイクロフォンの自己雑音低減を掲げて歩んできた経験は、現在の、そして未来のAIに恩恵をもたらします。音声処理技術のさらなる改善は、環境の識別、文脈の理解、感情の認識、話者の識別、複数話者ダイアライゼーションといったAI機能の強化に不可欠なものです。高品質な音声入力により、AIは人間の聴力と同等、あるいは凌駕し、人間と自然な対話ができるようになるでしょう。

さらに、人間と機械との対話レベルが向上することで、AIを使った新たなユースケースやサービスを実現できます。例えば、将来、Microsoft Copilotが、チームの会議を要約するだけでなく、会議全体の雰囲気を評価するのを想像してみてください。将来のAIは人間の発する音声と音響情報をもとに、議論されたアクションアイテムを強調したり、優先順位をつけたりできるようになるかもしれません。更に、AIを使用したコーチング機能を追加し、今後の会話をより望ましく進行できるように、ユーザーの役に立つアドバイスを与えられるかもしれません。

AIを使用した就職希望者の一次面接や、音声のみで話者識別を行う機能が、オンラインショッピングに十分なセキュリティレベルで実現されるのを想像してください。
これらは、人間と同等もしくはそれ以上の聴力を備えた将来のAIが実現可能な例の一部に過ぎません。インフィニオンの強化されたMEMSマイクロフォン ソリューションと共に、エキサイティングな旅のスタートを切ることを誇りに思います。

購入に関して

インフィニオンの販売拠点または販売パートナーまでお問い合わせください:
www.infineon.com/WhereToBuy

サービスホットライン

インフィニオンはフリーダイヤルの**0800/4001**サービス ホットラインを代表番号として提供しており英語、中国語、ドイツ語で24時間利用可能です。

ドイツ	0800-951-951-951 (ドイツ語/英語)
中国本土	4001-200-951 (北京語/英語)
インド	000-800-4402-951 (英語)
米国	1-866-951-9519 (英語/ドイツ語)
その他の国	00*-800-951-951951 (英語/ドイツ語)
ダイレクトアクセス	+49-89-234-0 (相互接続料金、ドイツ語／英語)

*ご注意: 一部の国から上記国際電話番号にアクセスする際「00」以外のコードをダイヤルする必要があります。
詳細は https://www.infineon.com/service/ をご覧ください。

発行元:
インフィニオン テクノロジーズ ジャパン株式会社

© 2025 Infineon Technologies AG. All rights reserved.

Public

Version: V1.0_JA Date: 11/ 2025

Stay connected!

Scan QR code and explore offering
www.infineon.com

Please note!
This Document is for information purposes only and any information given herein shall in no event be regarded as a warranty, guarantee or description of any functionality, conditions and/or quality of our products or any suitability for a particular purpose. With regard to the technical specifications of our products, we kindly ask you to refer to the relevant product data sheets provided by us. Our customers and their technical departments are required to evaluate the suitability of our products for the intended application.

We reserve the right to change this document and/or the information given herein at any time.

Additional information
For further information on technologies, our products, the application of our products, delivery terms and conditions and/or prices, please contact your nearest Infineon Technologies office (www.infineon.com).

Warnings
Due to technical requirements, our products may contain dangerous substances. For information on the types in question, please contact your nearest Infineon Technologies office.

Except as otherwise explicitly approved by us in a written document signed by authorized representatives of Infineon Technologies, our products may not be used in any life-endangering applications, including but not limited to medical, nuclear, military, life-critical or any other applications where a failure of the product or any consequences of the use thereof can result in personal injury.