From Noise to Nuance: MEMSマイクロフォンとAI理解の進化

インフィニオン テクノロジーズ、Dr. Gunar Lorenz

テクノロジーに満ち溢れた現代社会において、高音質な音声は私たちの日常にかかせないものです。音声を捉えて送信することは、バーチャル アシスタントからビデオ通話にいたるまで、私たちの生活に大きく関係します。しかし、日常生活というノイズの多い環境において、高音質な音声キャプチャは、技術的に容易ではありません。進化したインフィニオンのMEMSマイクロフォンは、より明瞭で詳細な音声をキャプチャし、AI搭載デバイスに人間のような交流を可能にします。

MEMS (Micro Electro Mechanical Systems; 微小電子機械システム) は、微小な機械部品と電子回路を1つのチップに組み込んだ小型デバイスです。自動車のエアバッグからスマートフォンのセンサーにいたるまで、MEMSは数えきれないほど多くのアプリケーションに使用されています。なかでも非常に小さなMEMSマイクロフォンは、小型であるにもかかわらず、きわめてクリアな音質を実現します。最新のスマートフォン、イヤフォン、スマートスピーカーが周囲の状況を“聞く”ことができるのは、このMEMSマイクロフォンが搭載されているからです。

AIがさらに日常生活に浸透しつつある現在、MEMSマイクロフォンが“聞く”音声の品質は、さらに重要になっていきます。これは、AIが受信した音声データをもとに、人間の音声、感情、意図を理解しているからです。音声が明瞭であるほど、AIはより正確に私たちを理解できます。そのため、AIの未来を思い描くうえで、進化し続けるAI技術に後れを取らないためには、MEMSマイクをどのように進化させるべきなのかを考える必要があります。

AIの進化に伴い音声が重要になる理由

私たち人間が話すとき、伝えられる情報は言葉だけではありません。声のトーン、声の共鳴、声の大きさ、話すスピード、イントネーションなど、話し方がコミュニケーションの重要な部分を占めており、それらによってメッセージの受け取られ方が変わってきます。一般的に、言葉によるコミュニケーションは全体の7 %程度に過ぎず、残りの93 %は声のトーン (38 %) やボディランゲージ (55 %)[1]など、非言語的なものだと言われています[1]。声は、私たちの感情や意図、さらには周囲の環境に関する微妙なニュアンスを伝えます。例えば、慌ただしく大きな声は緊急性を伝え、柔らかく安定したトーンからは落ち着いている様子が伝わります。人間がこうしたニュアンスを自然に解釈するように、高度なAIはこうしたニュアンスを理解し、直感的かつ“人間的”と感じられるようになることを目指しています。

Siriや協働マシンのようなAIアシスタントが、言葉によるコミュニケーションの7 %程度だけではなく、私たちが求めるものをきちんと理解できる状況を想像してみてください。例えば、怒っているとき、私たちはただ誰かに話を聞いてほしいだけなのかもしれません。別に頼んでもいない解決策が欲しいわけではなく、ただ一緒に怒ってほしいだけなのかもしれません。(共感力のある) 人は、経験的にこうした状況に対応できます。しかし、機械はこのような状況をどのように“見て”、“感じる”のでしょうか？

AIが きちんと“聞ける”ようになれば、機械はデジタルな友達やデジタルな同僚、または話すwikiから、効果的にあなたに寄り添う友人に変貌するでしょう。AIがこのレベルに到達するには、感情、トーン、緊急性といった声の持つニュアンスを認識する必要があります。例えば、誰かが騒がしいカフェや通りを歩いている状況で電話をかけてきたとき、AIは周囲の雑音を取り出し、それに応じて反応できなければなりません。空港などでは、「迅速な回答が必要である」というような、質問者のニュアンスを理解することが求められます。AIはこうした理解によって、単なるヘルパーを超えた存在となり、状況に合わせて応答を切り替えられるような、頼りになる存在となりえます。

しかし、最大の課題の一つとして、AIの周囲の雑音や妨害の多い現実環境下における、音声理解が難しいことが挙げられます。ほとんどの民生用機器において音を捉える主要技術として用いられているMEMSマイクロフォンは、改良され続けています。しかし、マイクロフォンと人間の聴覚が捉えられる音には、まだ隔たりがあります。AIが人間のように私たちを自然に理解できるようになるには、こうした隔たりをなくすことが不可欠です。

ヒアリング技術:
人間はいかにして音声を認識するか

人間の聴覚は驚くべき進化を遂げ、騒がしい部屋で会話に集中したり、音楽を十分に堪能したりできます。私たちの耳は、自然と雑音を遮断しつつ、重要な音に集中することができます。例えば、私たちは特定の周波数、なかでも音声情報が多く存在する500 Hzから6 kHzの周波数に対して特に敏感です。更に、もう一つの重要な要素に、音圧レベル (SPL)、すなわち音量があります。SPLは、基準音圧 (通常は人間の聴覚の閾値) に対して、音波によって変化した音圧の変化を振幅[dB(デシベル)] で表したものです。

しかし、人間の聴力にも限界があります。特に年齢を重ねるにつれ聴力は徐々に下がり 、高音が聞き取りにくくなったり、騒々しい環境下での会話を理解しにくくなったりすることが多くあります。MEMSマイクロフォンにも同様の限界があります。MEMSマイクロフォンはさまざまな音を聞くことができますが、声のニュアンスや小さな音、特に騒がしい場所で音を捉えるという点では、人間の耳には及びません。人間の聴覚と比較した場合、現在使用されているほとんどのMEMSマイクロフォンの性能は、特にスピーチの詳細まで検出する点においては、60歳の男性[2] の聴覚能力と同等のレベルです。

どのように音が認識されているのか理解することは、現在のAIシステムが直面する課題を理解する助けとなります。AIがお年寄りのようにしか“聞く”ことができないのであれば、感情的なニュアンスや文脈といった重要な情報を聞き逃してしまうことになります。AIが本当の意味でアシスタントとして、注意深く耳を傾け、言葉の意図を理解し、適切に応答するには、最高音質の音声入力が必要です。AIに高品質なデータを提供するためには、未来のMEMSマイクロフォンが、可能な限り音声を明瞭かつ詳細に捉えられるようにすることが重要です。

MEMSマイクロフォンが未来のAIにとって重要な理由

前述したように、人間の耳は周波数と音圧レベルに基づいて音声を検出しますが、その感度は年齢を重ねるにつれて低下します。高品質のMEMSマイクロフォンには、大きな音でも小さな音でも歪みなく捉える、広いダイナミックレンジが必要です。マイクロフォン自身がもつ自己ノイズが小さいほど、捉えられる音は明瞭になります。これはAIデバイスにとって特に重要で、より明瞭な音はAIシステムがより正確に音を捉えやすくします。

ここでMEMSの技術について触れると、たとえ最新のマイクが人間の耳にはまだ及ばないとしても、AIによる人間と機械との高度な相互交流に欠かせないのは「高いSNR(Signal to Noise Ratio)」のマイクロフォンです。SNRは、対象となる音 (声など) と周囲のノイズの強さの比率を表したものです。高SNRマイクロフォンはノイズの占める割合が低いことを意味するため、高い明瞭度が得られ、音声認識に最適です。

このように、高いSNRのマイクロフォンは、短いコマンド認識で使用されるさまざまな生成AIモデルの強化においてその実力を発揮します。いわゆるウェイクワード検出のような単純作業では、「Alexa」や「Hey Siri」のように特定の単語によってデバイスを起動させますが、高SNRでははっきりとした信号が提供され、迅速かつ確実な起動が可能となります。複雑な作業の場合、音声アシスタントに搭載されている「大規模言語モデル (Large Language Models: LLM)」が文脈から低品質の音声を解釈します。膨大な言語データで訓練されたLLMは、テキスト、オーディオ、ビジュアルを統合し、さらに文脈を用いて、音声のテキスト化をより安定したものにします。LLMは、音声が不完全でも、言葉の意図を認識するのに優れています。最後に、ローカル デバイスのエッジAIモデルもまた、コマンドを明確に理解するうえで高SNRが必要です。

AIとMEMSの未来: 未来にはいったい何が？

AIアシスタントが、単にコマンドに従うだけでなく、声のトーンを理解し、会議の内容を要約し、緊急性によって次の行動の優先順位まで付ける状況を想像してみてください。インフィニオンのMEMSマイク技術の進歩により、感情や気分の検出、会話中にいつ誰が話したのかを結びつける (話者ダイアライゼーション)[3]ことが、ますます可能になってきています。音声入力の向上により、AIは応答を最適化し、感情的なニュアンスを理解し、本当の意味で助けとなる対話ができるようになります。MEMS技術が進化するにつれ、AIは私たちの声に耳を傾け、理解し、自然と共感できるようになり、そして私たちの日常生活に欠かせない存在となるでしょう。その先にはきっと明るい未来が広がっています。遠隔医療から、特別なニーズを持つ人々へのサポート、メンタルケア、言語学習まで、さまざまな分野においてその可能性は無限に広がっていきます。

インフィニオンのマイクロフォン設計は、標準的なMEMS技術の枠を超え、シングル バックプレート (SBP) 技術およびシールド デュアル メンブレン (SDM) 技術を中心とした高いSNRマイクロフォンが特長で、幅広いアプリケーションに対応します。また、システム設計最適化は、MEMSの可能性を最大限に引き出す重要な鍵となります。

高いSNRを持つMEMSマイクロフォンを搭載した、今後のAIアプリケーションは、高度な環境認識、話者ダイアライゼーション[3]、音の区別を実現することで、AIデバイスの対話方法は世界に革命をもたらすことでしょう。AIデバイスは、かつてない高精度で重要な情報をフィルタリングし、優先順位をつけるようになります。例えば、スマートフォンは音楽のジャンルに基づいてサウンドプロファイルを自動調整し、イヤフォンは緊急のサイレンを検知し、スマートグラスは機械の異常を音で識別できるようになるかもしれません。ヘルスケア分野では、AIが医師からの指示を正確に理解することで、会話パターンや呼吸の微妙な変化を検知し、患者のモニタリングをサポートできます。教育分野では、生徒の声のトーンや学習意欲のレベルを認識するで、個人に最適化された学習を提供することができます。こうした進歩により、単純な応答を超えたAIの新たな基礎が築かれます。AIは私たちの周辺環境の変化や言葉のニュアンスを理解し、人間との対話に近いレベルの関わりを可能にするでしょう。そして、私たち人間のニーズに対応し、私たちの生活を変え、かけがえのないパートナーとなることでしょう。

参考文献

[1] online.utpb.edu/about-us/articles/communication/how-much-of-communication-is-nonverbal
[2] www.eetimes.com/ai-driven-interactions-powered-by-high-snr-mems-microphones
[3] en.wikipedia.org/wiki/Speaker_diarisation

Published by
Infineon Technologies AG
Am Campeon 1-15, 85579 Neubiberg
Germany

© 2024 Infineon Technologies AG.
All rights reserved.

Public

Version: V1.0_JA
Date: 12/2024

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.