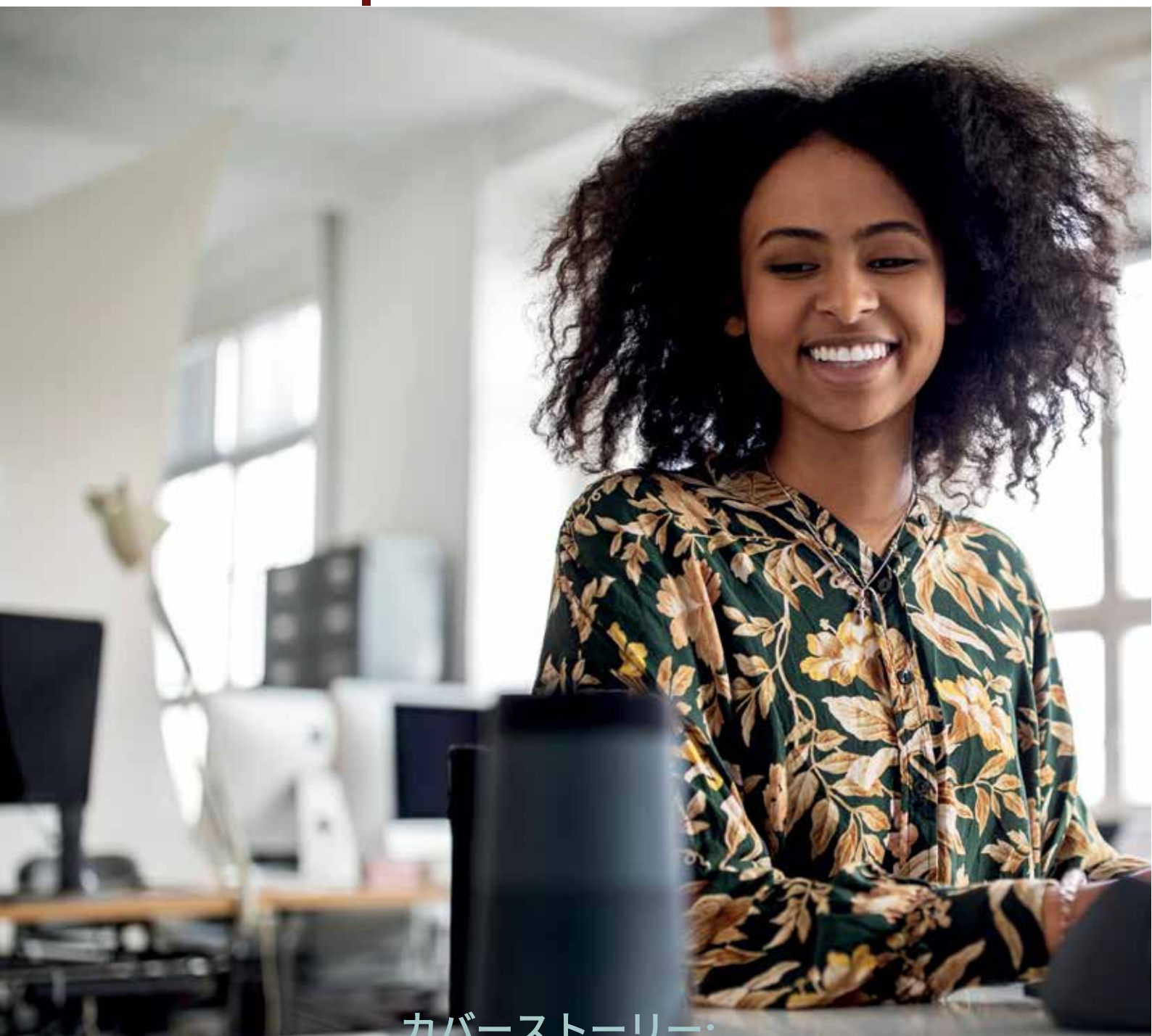


MARCH 2024

EE Times

EUROPE



カバーストーリー：
AI生成音声の進歩: 高SNR MEMSマイク
ロフォンの重要な役割

PAGE 6



カバーストーリー

AI生成音声の進歩： 高SNR MEMSマイクロフォンの 重要な役割

BY SOFIA CUCALON インフィニオン テクノロジーズ、コンシューマーセンサー
JULIAN KORNPROBST MEMSセンサーシステムコンピテンスチーム ビジネスリーダー

驚くほどリアルなコンテンツを作り出す文章生成ツールや画像生成ツールなど、人工知能 (AI) の統合が日常生活を大きく変えたことは間違いありません。しかし、AIの影響は視覚や文字メディアにとどまらず、音声テキスト変換 (STT) や自然言語処理 (NLP) などのオーディオアプリケーションにまで及んでいます。オーディオ アプリケーションにおける新たな品質レベルに貢献できるのは、最新の大規模言語モデルベースのAI世代だけなのでしょうか。それとも、こうした開発においてハードウェアは、なおも重要な役割を果たしているのでしょうか。具体的には、私たちの日常生活を変えようとしているこの新しい品質のヒューマン マシン インターフェースに、高いS/N比 (信号対雑音比) のMEMS (微小電気機械システム) マイクロフォンは、どのような貢献をしているのでしょうか？この記事では、こうした疑問を探り、音声生成 (TTS) やNLPのような最先端オーディオ アプリケーションの開発において、高いS/N比のMEMSマイクロフォンが果たす重要な役割について考察します。

ク アルコムของผู้调查レポート「State of Sound」の2023年度版によると¹、一日のうちヘッドフォンを装着する時間は増加しています。

カフェなど公共の場で仕事をする人は増え続け、それと共に静けさを求めたり、オンライン会議を行ったりするために、周囲の雑音を遮断しようとヘッドフォンを利用する人も増えています。空き時間には、同じヘッドフォンでゲームを楽しみ、音楽やオーディオブックを聴き、また友人との会話を楽しみます。このように装着時間が長くなったことで、快適さだけでなく、音質もいまや重要な購入基準になりつつあります。調査では、ヘッドフォンを購入する際に、空間オーディオ、クリアな音声通話、低遅延オーディオといった「プレミアムオーディオ機能」に関心を持つ人が増えていることがわかりました。回答者の73パーセントがオーディオ機器購入のたびに、音質が向上することを望んでいると答え、この数値は昨年の67パーセントから増加しています。

自動車だけでなく、家電製品においても重要

なオーディオ機能は、音声認識と音声生成です。数年前から、SiriやAlexaなどに代表される音声アシスタントは、音声コマンドを使ったスマートホームの制御などにより操作を簡素化し、新しいアプリケーションも実用化されてきました。現在、さまざまな機器が音声アシスタントを搭載し、スマートフォン (図1) やヘッドフォンからスマートTV、スマートスピーカー、スマートホーム機器、ノートPC、タブレットまで多岐にわたります。自動車に搭載される音声アシスタントも増加しており、ドライバーはハンドルから手を離すことなく、さまざまな機能を制御できます。SARは、音声アシスタントが搭載された機器の市場は、2028年までに販売数が年間30億台に増加し、CAGR (年平均成長率) は5%に成長すると予測しています。²

オーディオにおけるAI利用の展望

しかし、現在のシステムはいまだ完璧とは言えない状況です。音声認識はアクセント、言語的不完全性、ちょっとした周囲の雑音な

どでエラーが発生します。音声出力は、いまなお非常に専門的で、実際の音声とは明らかに異なります。

最新世代のAIは、あらゆる人間と機械の相互作用における、技術的な革命に他なりません。AI生成音声による利点は、音声アシスタントや、それにより人間の意図をより正確に理解することにとどまりません。本当の人間の声とほとんど区別がつかない人工音声を生成することで、視覚障害者のアクセシビリティを向上させることなどができます。様々なデジタルプラットフォームにおけるユーザー体験を向上させ、エンターテインメント分野やカスタマーサポートにおける新たな可能性を提供します。

AI生成音声の主要アプリケーションは、音声テキスト化するSpeech-to-Text (STT) です。AIを使用すると高速、高精度です。STTは、音声生成と同様に、ノートパソコンやスマートフォンなどの家電製品、音声アシスタントの搭載、会議の自動文字起こしなど、さまざまなアプリケーションへの応用が期待さ

AI生成音声の進歩: 高SNR MEMSマイクロフォンの重要な役割



図 1: 高SNR MEMSを使用した音声認識の利点を活かした、スマートフォン搭載の音声アシスタント

れます。会議では、AIを使用したアプリケーションが、誰がどのような発言をし、どのような指摘がなされたかを要約し、会議出席者ごとの異なる視点を把握できます。

NLPと表現力豊かな音声生成

NLP (自然言語処理) は生成音声AIの基本的な構成要素です。アクセント、口語的表現、不明瞭な発音、話し言葉と書き言葉の違いに関係なく、話し言葉の意味を理解することが目的です。話す速さ、イントネーション、声のトーンに基づいて意見や感情を認識することも、NLPの一部です。人間の声の音域は非常に広いので、NLP用の音声録音では、周囲の雑音や雑談などの外部からの影響を最小限に抑え、純粋な声を可能な限り正確にとらえなければなりません。つまり、マイクと信号処理は、NLPの品質を左右する大きな一因となるのです。

優れた音声認識を実現するには、可能な限り多くの人間の声を録音して、AIを訓練しなければなりません。そうやって初めて、音声の微妙なニュアンスに対応し、話し言葉を理解できるのです。

音声AI用MEMSマイク

NLPに関して言えば、音声AIは適切なハードウェアが導入されて初めて、最適な仕事ができます。すべては、人間の音声を作り出す音波を電気信号に変換することから始まります。この変換の完成度が、録音された信号の理解度に影響します。損失や劣化があれば、STTの精度に影響します。

オーディオの流れにおける最初の部品であるマイクロフォンは、音声AI機器を設計する上で重要な役割を果たします。MEMSマイクロフォンは他の追随を許しません。きわめてコンパクトなサイズでありながら、高性能、低消費電力を実現しており、さまざまな機器に容易に組み込むことができます。

MEMSマイクロフォンは3つの構成要素からなります(図2)。まず、実際のセンサー部品であるマイクロエレクトロメカニカルシステムです。音波がメンブレンを動かし、メンブレンはバックプレーンとコンデンサを形成します。それにより生じる静電容量の変化が電気信号を生成します。第二の構成要素であるASICには、メンブレン用のチャージポンプ、アンプステージ、クリーンな電源用の低ドロップアウトレギュレーター(LDO)、キャリブレーションロジックが搭載されています。これらの部品は3番目の構成要素であるパッケージに統合されます。パッケージは部品を保護し、遮蔽し、音響バックボリュームを形成します。

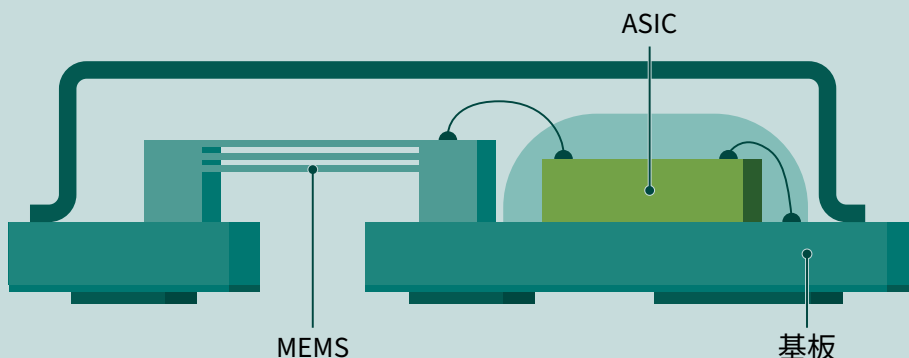


図 2: MEMSマイクロフォンのブロック図

AI生成音声の進歩: 高SNR MEMSマイクロフォンの重要な役割

周囲の雑音やアクセントがあったり、話者とマイクロフォンとの距離が最適でなかったり、と難しい条件下でも音声の微妙なニュアンスを認識するために、重要なマイクロフォンの特性となるのはSNRであり、マイクロフォン固有の自己雑音と標準的な基準信号との差を表します。

音声AI用XENSIV™ MEMSマイクロフォンの利点

前述のように、音声AI機器には正確な音声認識を行うために高SNRマイクロフォンが必要です。インフィニオンには、これまで高性能MEMSマイクロフォン³を開発してきた実績があります。SDM (Sealed Dual Membrane) は、インフィニオンの革新的なMEMSマイクロフォン技術で、2枚のメムレンと帯電したステータを利用して、密閉された低圧空洞 (図3) と差動出力信号を作り出します。このアーキテクチャは、超高SNR (最大75dB SNR) と超低歪みを実現し、マイクロフォンレベルで高い防塵保護 (IP57) を実現します。

インフィニオンのXENSIV™ IM73A135は、MEMSマイクロフォンとして業界最高レベルのSNR値73 dBを実現し、要求の厳しい音声AIなどのアプリケーションに最適です。また、小型パッケージ (4×3mm²) で音声キャプチャユニットが小型化されているため、ノートPCや会議用電話からスマートスピーカーやスマートフォンの至るまで、幅広い機器に音声AI技術を実装できます。

XENSIV™ MEMSのもう一つの利点は低消費電力です。異なるモードで動作させることによる節電で、最終製品での電力効率を高めるのです。生成音声AIを搭載した機器の多くはポータブルでバッテリー駆動のため、長いバッテリー寿命を実現することが重要です。

コンパクトでコスト効率が良く、低消費電力であるため、複数のマイクロフォンを一つの危機に搭載できます。これにより周囲の雑音を検知/低減し、精度の高い音声認識が実現されます。また、ビームフォーミングアルゴリズムを採用することにより、周囲の雑音から特定の話者の音声を分離/キャプチャし、優れた音声認識を実現します。

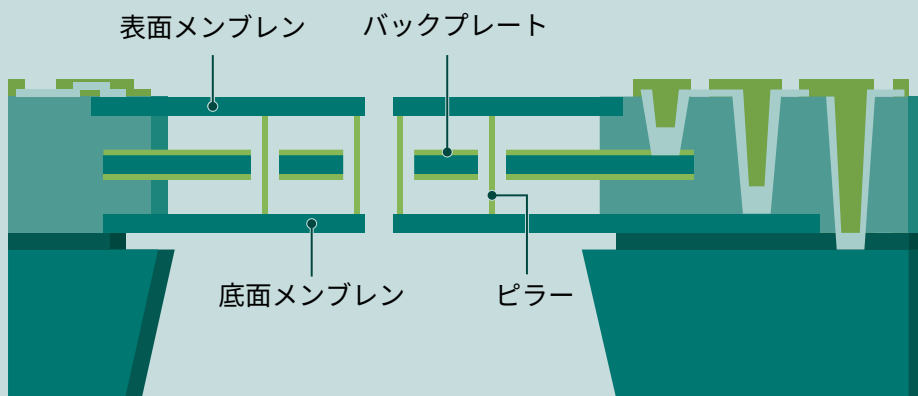


図3: SDM技術は2枚のメムレンと帯電したステータを利用して、密閉された低圧空洞と差動出力信号を作り出すことで、超高SNRと超低歪みを実現

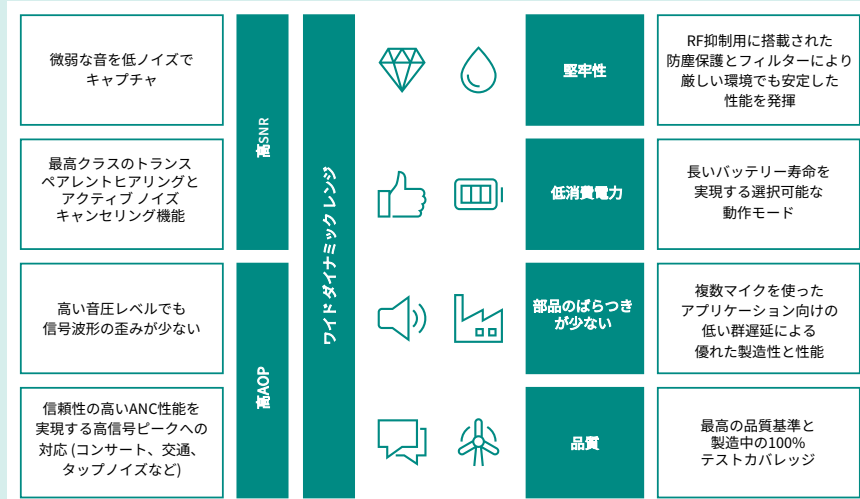


図4: XENSIV™ MEMSマイクロフォンのキーバリューインジケーター (KVI)

詳細はこちら: www.infineon.com/mems

高音質を重視する世界において、MEMSマイクロフォンの利点は市場での数字にも表れています。高SNRのMEMSマイクロフォンの市場は、低SNRのマイクロフォンよりも著しく急速に成長しています。例えば、Omdiaは、SNR64 dB以上のMEMSマイクロフォンの民生分野におけるCAGRを8.7%と予想しており、2027年までに30億近い販売数を見込んでいます⁴。

インフィニオンは、こうした傾向を予測してきており、音声AIアプリケーション向けに、ますます高性能になるMEMSマイクロフォンの開発に取り組み続けています。

現時点でSNR 73 dBという驚くべき値を実現していますが、さらに高いSNRや消費電力が低減された製品が間もなく登場します。

まとめ

AI生成音声の領域では、高SNR MEMSマイクロフォンの搭載がとても重要な役割を果たします。AIによりSTTなどのオーディオアプリケーションが進化する中、MEMSマイクロフォンは表現力のある音声データを取得することでそうした変化に貢献しています。この進歩により、音声認識がより自然なものになり、家電製品から視覚障害者のためのアクセシビリティ機能まで、さまざまな領域で適用できるようになります。優れたMEMSマイクロフォンの利点により、今後数年間で音声AIは、ボイスクローン、感情認識など、さらなるアプリケーションを開拓することでしょう。

インフィニオンテクノロジーズは、MEMSマイクロフォンのすべての構成要素を自社で開発/製造しています。また、MEMS、ASIC、パッケージの最適な組み合わせを見つけて、あらゆるアプリケーションで最高の性能を実現します。これにより、インフィニオンは進化する音声AIにおいて、ユーザー体験の向上と幅広いアプリケーションへの道を切り開いていきます。

参考文献

¹Qualcomm Technologies Inc. (2023). "The 2023 State of Sound Report."

²SAR Insight & Consulting. Voice assistant platform forecasts, 2023.

³インフィニオンテクノロジーズ. www.infineon.com/mems

⁴Omdia. (2023). "MEMS Microphone Report."