

次世代音声対話の展望：Audio LLM とセルフホスト AI が牽引する顧客体験の革新と実装戦略

Convergence Lab. 株式会社
代表取締役 社長 木村 優志
kimura@convergence-lab.com

1. エグゼクティブ・サマリー：パラダイムシフトの全貌

コンシューマーエレクトロニクス産業は現在、歴史的な構造転換の渦中にある。過去数十年にわたり、デバイスの進化は主にプロセッサの処理能力向上、ディスプレイの高精細化、あるいは単一機能の洗練といった漸進的なハードウェア・アップデートに主眼が置かれてきた。しかし、今日のイノベーションの中心は「機能の拡張」から「体験の統合」へと移行している。人工知能（AI）、パーソナルデータ、常時接続性、そして持続可能性が交差するこの新たな領域において、デバイスは単なる命令の実行者ではなく、ユーザーの行動を予測し、複数のプラットフォーム間でシームレスに機能する自律的なエージェントであることが求められている。

この構造転換を最も象徴的に表しているのが、音声ユーザーインターフェース（VUI: Voice User Interface）の進化である。初期の音声アシスタントは、天気予報の確認やタイマーのセットといった単純なタスクを処理する、既存のシステムに対する「プラグイン（付け足し）」の機能に過ぎなかった。しかし、大規模言語モデル（LLM）と音声処理技術の劇的な進歩により、音声対話は単なる入力手段から、人間と機械の関わり方を根本的に再定義する「アンビエント（環境適応型）AI」の基盤へと昇華しつつある。

本ホワイトペーパーは、マクロ経済的視点、最先端の技術動向、法制度とプライバシー要件、そして具体的なビジネス実装モデルという多角的な視点から、実社会における音声対話技術の現在地と未来を網羅的に分析するものである。特に、次世代のブレークスルーとして期待される音響データの直接処理モデル「Audio LLM」や、人間同士の自然な会話リズムを模倣する「全二重（Full-duplex）音声対話」、そして深刻化するプライバシーとレイテンシの課題を克服するための「セルフホスト AI」の役割に焦点を当てる。技術的進化が単なるスペックの向上にとどまらず、いかにして新たな顧客体験（CX）を創出し、企業に競争優位性をもたらすのか、その論理的な道筋を明らかにする。

2. マクロ環境と市場動向の俯瞰：音声インターフェース経済圏の拡大

音声対話技術の進化と普及を理解するためには、まずその基盤となる市場の成長軌道と、消費者の行動変容を定量的に把握する必要がある。マクロ経済分析の視点から見ると、音声インターフェースはすでにアーリーアダプター層による実験的な利用段階（キャズム）を越え、一般消費者層（マジョリティ層）の日常生活に深く浸透し、巨大な経済圏を形成しつつある。その裏付けとして、日本の会話型 AI 市場は 2024 年時点で 7 億 2,700 万米ドル、2033 年には約 31 億米ドルへ拡大する見込みであり⁸、音声アシスタント搭載デバイスは世界で約 84 億台に達して人口を上回る規模となっている¹⁰。米国では成人の約 35%（約 1 億人）がスマートスピーカーを利用し、インターネットユーザーの約 5 人に 1 人が音声で検索を日常利用するなど、定量的な浸透が明らかである¹⁰。

2.1. 国内外における市場規模と成長予測の定量分析

日本の音声対話および関連するスマートデバイス市場は、今後 10 年間で急激な成長を遂げると予測されている。この背景には、自然言語処理（NLP）の進化による人間らしいインタラクションの実現や、企業による最先端 AI ソリューションへの巨額の投資が存在する。特

に、顧客サービスの自動化、業務効率化、および政府によるデジタルトランスフォーメーション（DX）推進支援が、この市場拡大を強力に後押ししている。

以下の表は、国内外の主要な音声関連市場の現在価値と将来予測をまとめたものである。

市場カテゴリ	評価額（現在・直近予測）	将来予測評価額（到達年）	予測期間	年平均成長率（CAGR）	主要な市場牽引要因（成長ドライバー）
日本のスマートスピーカー市場	15.0 億米ドル (2024 年)	60.0 億米ドル (2032 年)	2025-2032	18.5%	デバイス機能の向上、AI によるパーソナライズされた体験の提供、スマートホーム機器のエコシステムとのシームレスな統合 ⁷ 。
日本の会話型 AI 市場	7.27 億米ドル (2024 年)	30.92 億米ドル (2033 年)	2025-2033	17.5%	顧客サービスの自動化需要、政府の DX 推進プロジェクト、ソフトバンク等による大規模なインフラ投資（例：OpenAI への投資等） ⁸ 。
日本のコネクテッド TV 需要	8.40 億米ドル (2025 年)	22.04 億米ドル (2035 年)	2025-2035	10.1%	ストリーミング需要の急増、スマート TV（市場の 58% を占有）への音声アシスタント統合によるインターフェースの進化 ⁹ 。
世界の音声ショッピング支出額	約 46.0 億米ドル (2021 年)	約 820.0 億米ドル (2025 年)	2021-2025	約 25.0%	Amazon や Walmart などの主要小売業者によるハンズフリー注文機能の拡充、製品検索や価格比較の完全な音声化 ¹⁰ 。

これらの定量データから抽出される重要な洞察は、ハードウェアの単なる普及から、プラットフォーム上でのサービスおよびコンテンツ消費への移行である。コネクテッドTVやスマートスピーカーは、もはや単なる映像・音声の再生装置ではない。それらは家庭内の情報ハブとなり、音声インターフェースを通じてEコマース、デジタルコンテンツ、さらには外部のサービスを直接的に呼び出し、消費を促進する強力なゲートウェイとして機能し始めているのである⁹。

日本市場における具体的な動きとして、ソフトバンクグループが2024年10月にOpenAIへ5億米ドルを投資したのに続き、2026年2月には同社がOpenAIへの追加投資として300億米ドル（3段階で実行、累計出資額は約646億米ドル・出資比率約13%に）を発表している⁸。このような民間企業による大型資本投下は、高度な会話プラットフォームの開発を加速させ、自然言語処理と人間のようなインタラクションにおけるイノベーションの土壌を形成している⁸。また、学術機関と民間企業のコラボレーションも活発化しており、ワシントン大学と筑波大学、カーネギーメロン大学と慶應義塾大学などが、AI人材育成と会話システムの向上を中心に強力なパートナーシップを構築している⁸。これらの動きは、日本の会話型AI市場が単なる技術輸入のフェーズを終え、独自のイノベーション・エコシステムを確立しつつあることを示唆している。

2.2. 消費者行動の変容と「スクリーンレス」へのパラダイムシフト

市場規模の拡大を牽引しているのは、消費者の行動様式の根本的な変化である。グローバルな調査データを紐解くと、2025年時点で全世界のインターネットユーザーの約5人に1人（約20～22%）が、オンラインでの情報検索に音声コマンドを日常的に使用している¹⁰。音声アシスタントを搭載したデバイス（スマートフォン、スマートスピーカーなど）の総数は、2024年までに約84億台に達し、世界人口を上回る規模となっている¹⁰。米国単独のデータを見ても、2025年には1億5350万人以上の成人が音声アシスタントを利用し、スマートスピーカーの普及率は全成人の約35%（約1億人）という驚異的な水準に到達している¹⁰。

特筆すべきは、消費者が音声インターフェースをどのように利用しているかという「文脈（コンテキスト）」である。調査によると、音声による検索クエリの約4分の3が「自分の近く（near me）」というローカルな意図（Intent）を持っている¹⁰。これは、ユーザーがPCの前に座ってじっくりと情報を収集する従来の検索パラダイムから、生活空間のなかで「何か別のタスクをしながら（ハンズフリーで）」即座に地域情報を得たり、タスクを完了させたりするパラダイムへの劇的な移行を示している¹⁰。

CES 2026の動向は、この消費者行動のシフトをさらに明確に裏付けている。業界全体のトレンドは、画面上のアプリケーションを視覚的に操作するスタイルから、インテリジェンスが環境の背後に溶け込み、視覚的依存を極限まで減らす「アンビエントAI（Ambient AI）」や「スクリーンレス・デバイス」へと向かっている⁴。ウェアラブルデバイスやスマートホームハブにおいて、音声は主要な（あるいは唯一の）インターフェースとして位置づけられつつある。CES 2026で発表された「Project AVA」は、この思想を体現したシステムであり、従来の画面に依存することなく、視覚および聴覚のコンテキストを背景で継続的に解析し、ユーザーの行動に対して適切なタイミングでのみ音声でガイダンスを提供するよう設計されている⁴。これは、AIがユーザーの「命令を待つ」受動的なツールから、「状況を察知して提案する」プロアクティブな存在へと進化している証左である。

2.3. 消費者心理の二面性：利便性の追求とプライバシーへの警戒

市場が急速に拡大する一方で、消費者の心理には強い葛藤が存在している。Deloitteが2025年6月に米国の約3,500人の消費者を対象に実施した調査「Connected Consumer study」によれば、消費者はジェネレーティブAIやその他のデジタル技術の恩恵を歓迎しつつも、技術の進歩が速すぎると感じており、十分な安全対策や透明性が欠如していることに強い懸念を抱いている¹¹。消費者はイノベーションを求めているが、同時に「透明性」「自身のデータ

に対するコントロール権」、そして「データセキュリティ」を技術提供者に対して強く要求しているのである¹¹。

PwC が実施した音声アシスタントの利用と認識に関する調査でも同様の傾向が確認されている。音声技術に対する一般的な認知度は極めて高く、回答者の 90% が音声対応製品を認知しており、そのうち 72% が実際に使用した経験を持っている³。若年層や高所得世帯（年収 10 万ドル以上）を中心に普及が進んでいる一方で、音声アシスタントの利用が依然として「基本的なタスク」に留まっている原因として、テクノロジーそのものに対する不信感と、プライバシー喪失への恐怖が最大の障壁として挙げられている³。

スマートスピーカーやスマート家電は、本質的に「常に聞き耳を立てている（Always-listening）」デバイスである。寝室やリビングルームといった最もプライベートな空間にマイクを設置することに対する消費者の警戒心を解かない限り、高度なユースケースへの移行や、市場の潜在的な最大価値の実現は不可能である。この消費者心理の二面性を理解し、技術的な解決策を提示することが、次世代コンシューマーエレクトロニクスにおける最大の至上命題となっている。

3. 音声ユーザーインターフェース（VUI）を再定義する技術的ブレークスルー

消費者の期待に応え、音声インターフェースを真のアンビエント AI へと昇華させるためには、基盤となるソフトウェア・アーキテクチャの根本的な再設計が必要である。従来の音声対話システムが抱えていた構造的な限界を突破しつつある、三つの革新的な技術パラダイムについて詳述する。

3.1. カスケード型処理の限界と Audio LLM（音声大規模言語モデル）の台頭
従来の音声アシスタントのアーキテクチャは、大きく三つの独立したモジュールから構成されていた。すなわち、「自動音声認識（ASR: Automatic Speech Recognition）による音響データのテキスト化」→「自然言語処理（NLP: Natural Language Processing）によるテキストの意図解析と応答生成」→「テキスト音声合成（TTS: Text-to-Speech）による音声の出力」という、カスケード型（直列型）のパイプラインである²。

このアーキテクチャは技術的に成熟しており、OpenAI の Whisper（2022 年リリース）や Google の Conformer アーキテクチャなど、強力な深層学習ベースの ASR モデルの登場により、多様なアクセントやノイズの多い音声でも人間と同等レベルの転写精度を達成している²。しかし、この「音声を一度テキストに変換する」というプロセスそのものが、次世代の VUI において決定的なボトルネックとなっている。なぜなら、人間の発話には、単なる文字列以上の膨大な情報が含まれているからである。話者の感情、声のトーン、ため息や笑い声といった非言語的表現、さらには背景の環境音（足音、サイレン、ガラスの割れる音など）は、テキスト化される過程で完全に削ぎ落とされてしまう¹²。長時間の会話においては、同じ「4 人でお願いします（Ok for 4）」という発話であっても、それが予約の時間を指しているのか、人数のことを指しているのかは、声のトーンや直前の文脈に大きく依存する¹³。

この根本的な課題を解決する次世代の基盤技術として急速に台頭しているのが、「Audio LLM（音声大規模言語モデル）」である。これは、LLM を外部のコントローラーとして使い複数の専門的な音声モデルを調整するアプローチ（AudioGPT や HuggingGPT など）とは異なり、LLM のアーキテクチャ自体を拡張し、音声の生の波形や音響特徴量を直接入力として受け取り、処理するモデルである¹⁴。Qwen-Audio や Pengi、LTU といった先進的な Audio LLM は、テキストを介在させることなくマルチモーダルな音響情報を直接理解する能力を持つ¹⁴。

この直接処理の利点は計り知れない。AI はユーザーの発話意図と同時に、感情状態や物理的な環境のコンテキストを即座に把握できる。例えば、高齢者向けに特化して設計された端末側で動作するスマートホームシステム「DESAMO」は、内蔵された Audio LLM を使用して生の音声を直接処理し、従来の音声アシスタントでは見逃されていた転倒などの重大な異常

イベントを高い精度で検知することに成功している¹⁶。さらに、Audio LLM の物理環境への認識能力は高度化しており、ドップラー効果を利用した周波数シフトの測定（音源が近づく際の波の圧縮と、遠ざかる際の波の膨張）によって、音源となる人や物体の相対速度や移動方向を推定する技術まで実装されつつある¹⁴。これは、コンシューマーデバイスが「耳」を持つだけでなく、音響を通じた「空間認識能力」を獲得したことを意味する。

3.2. 全二重（Full-duplex）音声対話と「割り込み」の高度なポリシー制御

Audio LLM による理解の深化と並行して進んでいるもう一つの巨大なブレイクスルーが、「全二重（Full-duplex）音声対話」の実現である。これまでの音声アシスタントとの対話は、トランシーバーを用いた通信に酷似していた。ユーザーがウェイクワードを発して話し終わるのをシステムが待ち、システムが処理を行って音声を返し終わるまで、ユーザーは再び待機しなければならない。このような「半二重（Half-duplex）」の通信方式は、人間同士の自然な会話リズムとは決定的に異なる¹²。

人間同士の会話では、相手の言葉に相槌を打ったり、相手が間違っただけを話している途中で割り込んで訂正したり、言葉を被せながら思考を発展させたりすることが日常的に行われる。Full-duplex 技術は、AI の応答生成（出力）とユーザーの音声入力（聞き取り）を同時に並行処理し、遅延のない自然なターン・テイキング（話者交替）を可能にする技術である¹²。

この技術領域において、Moshi、PersonaPlex、FireRedChat、FlexDuo といった最新のモデルが顕著な成果を上げている¹⁷。中でも、NVIDIA の研究チームが発表した「PersonaPlex」は、従来の全二重モデルが抱えていた「固定された単一の役割と声色しか持てない」という制約を打ち破った。PersonaPlex は、役割の条件付けを行うテキストプロンプトと、声質を複製する音声サンプルの両方を組み合わせたハイブリッド・システム・プロンプトを導入することで、特定の役割（例えば、親しみやすいカスタマーサポートや、厳格な専門アシスタント）と個別の声質を維持しながら、低遅延で自然な双方向コミュニケーションを実現している¹²。

さらに、全二重対話の実装において極めて重要なのが、システムが「いつ発言を譲るべきか（Yielding）」、あるいは「いつ発言を継続すべきか（Continuity）」を判断するポリシー制御の設計である¹⁸。すべての割り込みに対して AI が即座に沈黙することが最適とは限らない。研究によれば、車載音声アシスタントのように即時性と正確性が求められる環境では、ユーザーの割り込みや修正に対して AI が即座に発話を停止する「修復優先（repair-first）」のアプローチが適している。一方で、スマートホームのハブのように、生活音や他の家族の会話が飛び交う共有環境では、多少のノイズや関係のない発話の被りがあっても AI が自身の出力フローを維持する「継続優先（continuity-first）」のポリシーが効果的であることが示されている¹⁸。これらの文脈に応じた動的な制御技術により、スマートアシスタントはより流暢で人間らしい対話を提供し、エンタープライズの顧客サポートボットは複数ターンの複雑な文脈を維持した高度な対応が可能となる¹⁹。

3.3. Agentic AI（自律型 AI）への進化と UI の再定義

これらの高度な音声処理技術をバックエンドで統括するのが、「Agentic AI（エージェントック AI）」への進化である。2026 年現在、VUI（Voice User Interface）とバックエンドの AI サービスの相乗効果は新たな次元に達し、単なるタスクの自動化から自律的な行動の領域へと踏み出している⁶。従来のシステムは、あらかじめプログラマによって設定されたスクリプトや決定木に従うだけの「反応型（Reactive）」のツールであった。しかし Agentic AI は、ユーザーの曖昧な意図に対して自律的に推論（Reasoning）を行い、複数の行動計画を立案し、学習しながらタスクを実行する⁶。

Convergence Lab. が提唱する「Agentic UI」の概念は、この進化を端的に表している¹⁷。例えば、ユーザーが「週末のキャンプの準備を手伝って」と発話した際、旧来のシステムは「ページの検索結果を表示します」と単一のウェブ検索を返すのみであった。しかし Agentic AI

と統合された Full-duplex の VUI は、「キャンプ先の天気は雨の予報ですが、防水性の高いテントのレンタルを手配しましょうか？それとも行き先を変更しますか？」とプロアクティブに提案を行い、ユーザーとリアルタイムで対話（被せながらの会話）を行いながら、最終的な決済や手配までをシームレスに完結させる。これにより、VUI は単なる便利な追加機能から、機械の論理と人間の意図のギャップを埋める普遍的な架け橋へと変貌を遂げたのである⁶。

4. 実装を阻む「三重苦」：レイテンシ、ノイズ、プライバシーの技術的課題

技術の進化が著しい一方で、コンシューマーエレクトロニクスにおける次世代音声対話システムが、研究室の環境を飛び出して現実世界のビジネスに完全に定着するためには、依然として乗り越えなければならない重大な障壁が存在する。これらは主に「レイテンシ（遅延）」「ノイズ（環境音）」「プライバシー（情報の保護）」という3つの技術的・制度的課題に集約される²。これらの課題は独立して存在しているわけではなく、相互に絡み合いながらユーザーエクスペリエンス（UX）を損なう根本的な要因となっている。

4.1. 人間の認知閾値を超えるレイテンシ要件の達成

会話という行為において、遅延は致命的な UX の毀損をもたらす。人間の日常的な対話における標準的な応答時間（ターン・テイキングのギャップ）は約 150 ミリ秒であるとされている²⁰。しかし、従来のテキスト変換を介するカスケード型の音声 AI システムでは、文脈やプロソディ（韻律：発話のストレスやイントネーションのパターン）を正しく計算して自然な音声を生成するために、システムは完全な文章が入力され、処理が終わるのを待たなければならない。この結果、システムの応答には 400 ミリ秒を超えるレイテンシが日常的に発生していた²⁰。

この 150 ミリ秒と 400 ミリ秒の差は、人間にとって「機械的で不自然な間」として明確に認識され、対話のリズムを完全に破壊する。遅延が発生すると、ユーザーはシステムが正しく聞き取れたのか不安になり、言葉を不必要に繰り返すなどの摩擦（フリクション）が生じる。教育向けのテクノロジーやカスタマーサポートにおいて、このロボットのような硬直した対話テンポは、ユーザーの関心を維持できず、利用の継続を断念させる根本的な原因となっていた²⁰。

この認知的不協和を解決するための最新のアプローチが、インクリメンタル（漸進的）な音声生成モデルである。人間が言葉を発する際、文章全体の構成が確定する前から話し始めるのと同様に、最新の AI モデルは単語ごとに逐次的に音声を生成・処理するアーキテクチャを採用している。この手法により、自然さを維持したままレイテンシをわずか 25 ミリ秒レベルまで劇的に削減することが可能になりつつあり、これは人間の標準的な応答速度（150 ミリ秒）を遥かに凌駕するパラダイムシフトである²⁰。

4.2. 過酷な音響環境におけるノイズキャンセリングと前処理の重要性

家庭内、自動車内、あるいは製造現場といった現実環境は、研究開発が行われる無響室とは異なり、極めてノイズの多い予測不可能な空間である。テレビの音声、子供の泣き声、外を走る車のエンジン音、風切り音、あるいは複数の人が同時に話す「カクテルパーティー効果」の状況下では、対象となる話者の音声を正確に分離し、意図を抽出することが極めて困難になる²。

ハードウェアの進化、すなわち複数のマイクを搭載したマイクアレイ技術や物理的な音の到来方向を絞り込む指向性マイクの設計は確かに進歩している。しかし、それだけでは不十分であり、ソフトウェア側での高度な信号処理と機械学習モデルに特化した前処理が不可欠である。Convergence Lab.が実際に取り組んだ音声解析 AI の開発事例は、この前処理の重要性を如実に物語っている¹⁷。同社の事例では、メル周波数スペクトルやフィルタバンク特徴量といった音響特徴に、ノイズアウェア学習（ノイズ条件を明示的にモデルに与えたり、多様

なノイズによるデータ拡張で頑健性を高める手法)を組み合わせ、実環境に適応した認識を実現している¹⁷。マルチスケールの時間窓と周波数分解能を組み合わせた階層的な特徴抽出に加え、Transformer や Conformer へ渡す前段として適切な前処理を導入し、ノイズの周波数帯が限定的な産業ドメインではドメイン条件付け(周波数マスキングやバンド制約)により過学習を抑制しながら、認識精度とロバスト性を両立させている¹⁷。このような、音響工学の制約とデータ駆動の表現学習を統合した設計が、エンドツーエンドの認識性能を決定的に左右する。

4.3. 「Voice Privacy」の確立とコンプライアンス要件の壁

レイテンシやノイズといった純粋な技術的課題以上に、消費者の受容を阻み、企業の法的リスクを増大させている最大の壁が「プライバシーとデータセキュリティ」の問題である。前述の Deloitte や PwC の調査が示す通り、消費者はリビングルームや寝室に常に聞き耳を立てるデバイスを置くことに強い心理的抵抗を感じている。

従来のスマートホーム機器のアーキテクチャでは、デバイスがクラウド上の音声 AI と通信するため、音声コマンドや生体情報(声紋などの Personally Identifiable Information: PII)を含む生のオーディオデータが、そのままリモートのサーバーへとストリーミング送信されていた²¹。この構造は、企業にとって三つの深刻なリスクを内包している。第一に、データ転送経路やクラウドサーバー側でのデータ漏洩リスクである。過去のデータ流出事件は消費者の警戒心を不可逆的に高めており、市場成長の強力な阻害要因となっている²²。第二に、欧州の GDPR (一般データ保護規則)などを筆頭とする法規制の厳格化である。データの国境を越えた移動や長期保存に対する規制が強化される中、生の音声データをクラウドで処理し保持することは、企業にとって莫大なコンプライアンス監視コストと法的・財務的リスクをもたらす²¹。第三に、トラフィックの増大に伴うクラウド利用料の際限のない高騰である。

ここで明確に区別すべき概念が、「ノイズキャンセリング」と「ボイスプライバシー (Voice Privacy)」の違いである。Skyted 社の技術解説が指摘するように、ノイズキャンセリングはユーザーが「外部の音を遮断して快適に聞く」ための技術であるのに対し、ボイスプライバシーは「ユーザー自身が発する声が、周囲の人や不正な第三者に漏れない・分析されない」ことを保証する技術である²³。真のプライバシーは音源そのものから保護されなければならない。音響吸収材の利用から、リアルタイムで音声信号をフィルタリング・匿名化・暗号化する AI 駆動型のデジタルセキュリティ層の追加まで、多層的なアプローチが求められている²³。

5. セルフホスト AI によるアーキテクチャの革新：クラウド依存からの脱却と最適解

前章で挙げた「レイテンシ」「ノイズ」「プライバシー」という三重苦を同時に克服し、さらに爆発的に増大するクラウドコストを抑制するための決定的なパラダイムシフトが、「セルフホスト AI (Edge AI)」、すなわち端末側での高度な処理への回帰である。

5.1. セルフホスト AI による「プライバシー・バイ・デザイン」の具現化

セルフホスト AI アーキテクチャの最大の利点は、音声処理プロセスの大部分をローカルのデバイス内で完結させる点にある。Sensory 社が提供するようなソリューションに代表されるように、ウェイクワードの検出、話者の生体認証(声紋によるユーザー識別)、Speech-to-Text の変換、環境音の検知といった一連の処理がデバイスの内部で行われ、機密性の高い音声データが家から一歩も外部に出ない (Never leaves device) 仕組みが構築されている²¹。

このアプローチは、システム設計の初期段階からプライバシー保護を組み込む「プライバシー・バイ・デザイン」の理念を完全に具現化するものである。クラウドへの生の音声アップロードを行わないため、転送中のハッキングリスクやサーバー側でのデータ漏洩リスクは物理的にゼロとなる。また、GDPR におけるデータ最小化の原則に完全に準拠するため、メーカー側のコンプライアンス対応コストを劇的に削減できる²¹。家庭内に複数のユーザー

がいる場合でも、Sensory 社の技術では 1 ユーザーあたり 100KB 未満という極めて軽量のローカル・テンプレートを用いて声紋認証を行い、数秒でユーザー登録を完了させ、クラウドに依存することなく個人を識別することが可能である 21。

以下の表は、従来のクラウド依存型アーキテクチャと、次世代のセルフホスト AI アーキテクチャの特性を比較したものである。

評価項目	クラウド主導型アーキテクチャ	セルフホスト AI 主導型アーキテクチャ
プライバシーとセキュリティ要件	生の音声データ、個人識別情報 (PII)、生体情報を外部サーバーへストリーミング送信。経路・サーバーでの漏洩リスクが常に存在する。	音声データはデバイス内から一切外部へ出ない。生体情報のテンプレートもローカルに保持され、極めて高い機密性を確保 21。
レイテンシ (応答速度)	ネットワーク環境とサーバーの負荷状況に依存。通常 100ms~400ms 以上の遅延が発生。	デバイス内で即時処理されるため、60ms 以下での超低遅延応答が可能。常時オンで瞬時に反応する 21。
システムの可用性と耐障害性	安定したインターネット接続が必須要件。ネットワーク障害時やクラウドサーバーのダウン時は一切の機能が停止する。	オフライン状態でも自律的に動作可能。医療や自動車などミッションクリティカルな環境での確実な操作を保証 21。
ライフサイクルにおける運用コスト	ユーザー数と利用頻度の増加に比例して、継続的かつ莫大なクラウドサーバー費用 (通信費・コンピューティング費) が発生し続ける。	主なコストはデバイスの初期ハードウェア費用のみであり、スケールに伴う継続的なランニングコストを大幅に抑制可能 21。

5.2. エッジハードウェアの進化と消費電力の最適化

これら高度な AI 処理を端末側で実行可能にしたのが、半導体技術とエッジハードウェアの目覚ましい進化である。かつての組み込み用マイクロコントローラ (MCU) では、LLM はおろか単純なディープラーニングモデルを動かす処理能力すら不足していた。しかし現在では、厳しい消費電力制限の中で強力な AI 推論能力を発揮するチップが広く普及している。

CES 2026 における技術動向は、このハードウェアの進化を明確に示している。Arm ベースのプロセッサや NPU (Neural Processing Unit) を搭載したデバイスは、バッテリー駆動の制約下でありながら、高度な空間 AI や音声対話をローカルで実行している 5。例えば、Meta の最新の Ray-Ban スマートグラスは、Arm CPU と Ethos-U55 NPU を活用し、極小の電力制限の中でハンズフリーの音声対話と空間認識を実現している 24。また、Alif の Ensemble E8 のような超低電力ノード向けの新しいエンドポイントプラットフォームは、環境に溶け込むアンビエント AI のワークロードを、ネットワークの末端にある極小のセンサーやデバイスで直接処理する未来を提示している 5。

一方で、より高度な産業用アプリケーションや、複雑な Full-duplex の音声 LLM の実行においては、より高い演算能力を持つハードウェアが必要となる。ここで圧倒的な強みを発揮するのが、NVIDIA Jetson シリーズのような GPU 搭載のセルフホスト AI モジュールである 17。Convergence Lab. が提供するセルフホスト AI ハードウェアの完全ガイドでも言及されているように、Jetson はエッジ領域において事実上の覇権を握りつつある 17。同社の実際の受託開発事例では、コベルコ建機株式会社のクレーンに Jetson 搭載の産業用 PC を実装し、クラウ

ドに頼ることのできない過酷なオフライン環境下において、堅牢な設計のもとでリアルタイムの AI 推論（オブジェクト検出や状況認識）を行うシステムがすでに稼働している²⁵。音声処理領域においても、Jetson の計算資源を活用することで、Whisper や Zipformer による高精度な音声認識、VITS や StyleTTS2 による自然な音声合成、さらには軽量化された LLM モデルをローカルで連携させ、遅延のない Full-duplex の音声対話基盤をエッジ側で構築することが可能となっている²⁵。

5.3. 垂直統合型のハイブリッド・アーキテクチャの設計

エッジデバイスの能力が向上したとはいえ、世界中のあらゆる知識を内包する数兆パラメータ規模の巨大 LLM をコンシューマーデバイスにそのまま搭載することは、物理的にもコスト的にも不可能である。したがって、企業にとっての現実的な最適解は、エッジの俊敏性・安全性と、クラウドの膨大な知識・推論能力をシームレスに連携させる「垂直統合型のハイブリッド・アーキテクチャ」の構築である¹⁷。

このアーキテクチャにおいて、日常的な家電の操作、ユーザーの生体認証、緊急事態（転倒音やガラスの割れる音など）の即時検知、そしてプライベートな日常会話の一次処理は、すべて低遅延と高セキュリティを誇るエッジ側で実行される。その上で、最新のニュースの検索、複雑なスケジュールの調整、あるいは膨大な外部データベースとの照合が必要な高度な推論タスクに対してのみ、匿名化されたテキストデータ、あるいは抽出された「意図 (Intent)」のメタデータだけをクラウド側の API へと送信する。

Convergence Lab.の技術ブログでも論じられている通り、「汎用 LLM の限界」を突破するためには、LoRA (Low-Rank Adaptation) や FFT (Full Fine-Tuning) を用いて特定業務や特定ドメインに特化させた「垂直統合型ローカルモデル」をエッジ側に配置し、必要に応じてクラウド API を使い分けるハイブリッド戦略が不可欠である¹⁷。さらに、このアーキテクチャに RAG (Retrieval-Augmented Generation：検索拡張生成) の仕組みを組み合わせることで、社内の独自データベースや個人の過去のコンテキストを参照しながら、ハルシネーション (AI の嘘) を抑えた正確な回答を Full-duplex の音声対話を通じて即座に提供することが可能になる¹⁷。

6. 産業別ユースケース：次世代スマート家電が創出する新たな事業価値

Audio LLM、全二重対話 (Full-duplex)、Agentic AI、そしてエッジ処理というこれら四つの技術的要素が統合されることで、コンシューマーエレクトロニクスのユースケースはかつてない劇的な進化を遂げる。ここでは、2025 年から 2026 年以降の近い未来において想定され、すでに社会実装が始まりつつある具体的なシナリオを産業分野別に考察する。

6.1. カスタマーサポートおよびコンタクトセンターにおける顧客体験 (CX) の革新

次世代音声対話技術のビジネス実装において、最も早くかつ最大の投資対効果 (ROI) を生み出す領域がカスタマーサポートである。2026 年現在、会話型 AI は実験的な技術から顧客対応チームにとって不可欠なインフラへと移行しており、企業の 80% が顧客サービス向上のために AI 搭載ソリューションを利用、または導入を計画している。コンタクトセンターのリーダーの 80% 以上が AI による生産性向上を最優先課題として挙げており、この分野の変革は急務となっている。

従来のコールセンター向け AI は、応答が機械的で柔軟性に欠け、顧客にストレスを与えるケースが多かった²⁰。しかし、遅延を 25 ミリ秒レベルまで短縮する次世代の音声処理技術や²⁰、人間の自然な会話フローを維持しながら大量のコールを処理できる最新の AI 音声エージェント (VOCALLS など) の登場により、運用コストを増大させることなく顧客サービスを強化することが可能になった²⁹。また、CallBotics のようなエンタープライズ対応の

プラットフォームは、感情認識技術や CRM（顧客関係管理）システムとのシームレスな統合を実現し、より共感性の高い対話を提供している。

さらに、全二重（Full-duplex）通信技術は、複数ターンの文脈を維持したカスタマーサポートボットの構築を可能にし、より迅速で役立つサービス体験を実現する¹⁹。例えば、PersonaPlex のようなモデルは、顧客サービスシナリオにおける役割付け（ロールコンディショニング）や自然な応答において、最先端の性能を実証している¹²。Zendesk の調査が指摘するように、やがて 100% の顧客対応に何らかの形で AI が関与する世界が到来すると予測されており、音声 AI によるコンタクトセンターの自動化は、企業の競争力を左右する最重要のユースケースとなっている。

6.2. アンビエント・スマートホームにおけるプロアクティブな生活支援

CES 2026 で多くの企業が提示した「アンビエント AI」の概念は、スマートホームのあり方を根本から変革する。従来のスマート家電は、「ユーザーが明示的なコマンドを発する（例：『Alexa、リビングの電気を消して』）」ことによるのみ受動的に動作していた。しかし次世代のシステムは、視覚と聴覚のコンテキストを常時モニタリングし、ユーザーの行動意図を予測して自律的に支援を行う¹。

例えば、Samsung の「AI Companion Vision」や Amazon の「Alexa Plus」が目指す世界観では、AI は擬人化された独立したロボットとしてではなく、家全体の環境に溶け込んだ存在として機能する⁴。ユーザーがキッチンで両手を塞いだ状態で夕食の準備をしている際、環境音と映像から「食材に悩んでいる」という状況を察知した AI が、画面を介さずに音声のみで適切なレシピを提案する。ユーザーは AI の提案が終わるのを待たずに、「いや、冷蔵庫にあるトマトを使いたいんだけど」と被せるように割り込んで（Full-duplex）対話を修正することができる。AI は即座に提案を修正し、さらに自律的な Agentic 機能を用いて、オーブンの予熱を開始し、キッチンの換気扇を連動させる。このような「存在感を消しながらも常に寄り添う（Screenless & Ambient）」知能の提供こそが、次世代の究極の顧客体験（CX）となる⁴。

6.3. 高齢者ケアとヘルスケア・モニタリングの高度化

先進国を中心とした人口動態の急速な高齢化に伴い、高齢者の見守りやケアはテクノロジーの最も重要な適用領域の一つとなっている。前述の「DESAMO」に代表されるような、Audio LLM を搭載した端末側で動作するスマートホームシステムは、この分野で革新的な価値をもたらす¹⁶。

カメラを用いた監視システムは高齢者のプライバシーを著しく侵害するため、居住空間での普及には高いハードルがある。しかし、音声と環境音に特化したエッジ処理システムであれば、プライバシーを保護しながら精緻な見守りが可能になる。Oura Ring の第 4 世代のような Arm ベースの生体センサーからのバイタルデータと、空間内のマイクが捉える音響データを組み合わせることで²⁴、インテリジェントな状況判断が行われる。例えば、AI が「ドスン」という重い衝撃音（転倒音）と「うめき声」をローカルデバイス内で即座に解析し、同時に「〇〇さん、大丈夫ですか？」と音声で呼びかける。数秒間応答がない、あるいは苦痛を含んだ音声が続いてきた場合、AI は自律的に緊急連絡先や医療機関へ通報を行い、玄関のスマートロックを解除して救急隊員の進入をサポートする。これらの全プロセスにおいて、日常のプライベートな会話データは一切外部に送信されず、エッジデバイス内で安全に破棄される仕組みが担保されているため、導入の心理的障壁を大きく引き下げることができる¹⁴。

6.4. 車載エンタテインメントにおけるマルチユーザー空間の制御

自動車産業もまた、車両を巨大なコンシューマー・コネクテッドデバイスとして再定義しつつある。車内は特有の環境であり、ロードノイズや風切り音が絶えず発生する過酷な音響空間であると同時に、複数の乗員が近い距離で同時に発話する複雑なマルチユーザー空間である¹⁹。

従来の車載アシスタントは、運転席からの単一のコマンドを処理することしかできなかった。しかし、次世代の音声システムは、Audio LLM の音源分離技術と空間認識能力を用いて、運転席、助手席、後部座席のそれぞれの声を物理的に分離し、それぞれの意図を同時に並行して理解する。ドライバーが「ナビの目的地を駅に変更して」と言いかけたところに、後部座席の子供が「アニメの音楽かけて！」と割り込んでも、システムは Full-duplex 通信と高度なエージェント機能によりパニックを起こすことはない¹⁹。AI はドライバーの文脈を優先して維持しつつ（修復優先ポリシーの適用）、バックグラウンドの自律タスクとして後方スピーカーから子供向けの音楽を流すといったマルチタスクを完璧に処理する。これにより、運転中の視線移動を完全に排除した安全なハンズフリー環境と、同乗者への快適なエンターテインメントの提供を両立させる⁸。

6.5. 音声コマース（Voice Commerce）の爆発的成長とエージェント機能の融合

音声を用いたショッピング行動は、すでに米国市場を中心に爆発的な成長を見せている。統計によれば、米国消費者の約 50% が音声検索を利用したショッピングを経験しており、世界の音声ショッピング支出額は 2025 年までに約 820 億米ドルに達し、さらに 2030 年までには E コマース全体の売上の 30% を音声主導のトランザクションが占めると予測されている¹⁰。

これまでの音声ショッピングは、「いつもの洗剤を再注文して」といった、過去の購入履歴に基づく単純で反復的なトランザクションに限られていた。しかし、VUI に Agentic AI（推論と自律行動能力）が統合されることにより、音声コマースの質は根本的に変化する⁶。ユーザーが「今度の週末の家族 4 人でのキャンプ向けに、予算 3 万円以内で設営が簡単なテントを見繕って」と曖昧な条件で発話した場合、AI はパーソナルコンシェルジュとして機能する。複数の EC サイトを巡回して製品を比較し、在庫状況と配送日数を自律的に確認した上で、「A 社のワンタッチテントはいかがですか？ レビューも高く、予算内で収まります。防水スプレーもセットにしておきましょうか？」と音声で提案を行い、ユーザーの「それで頼む」の一言で決済から配送手配までを完結させる。このような UCP（Unified Commerce Platform）プロトコルと Agentic UI の連携は、Amazon という単一の目的地に依存していたこれまでのコマースの物理学を塗り替え、消費者の購買行動をより直感的で摩擦のないものと進化させる⁶。

7. 政策的視座と企業の技術戦略：実装の死の谷を越えるための最適解

技術的な要素がどれほど出揃ったとしても、これらの高度な AI システムを実際のビジネスプロセスや商用プロダクトに安定的に組み込み、消費者に真の価値を持続的に提供するための道のりは決して平坦ではない。単に最新の LLM の API キーを取得して呼び出すだけ、あるいはオープンソースの強力な音声認識モデル（Whisper 等）を既存のアプリ画面に「付け足す」だけでは、エンドツーエンドの優れた UX は到底実現できない²。ストリーミングオーディオのパイプライン構築、ウェイクワード検出器の実装、低遅延ネットワークキングの最適化など、水面下での泥臭いエンジニアリングの統合が不可欠である²。

Convergence Lab. 株式会社が数多くの受託研究開発を通じて蓄積してきた実践的な知見が示すように、企業の AI プロジェクトの多くは「概念実証（PoC：Proof of Concept）」の段階で高い精度を達成しながらも、実環境のノイズ、システム統合の複雑さ、あるいは運用コストの壁に直面し、本番環境への実装に至らない「PoC 死（死の谷）」に陥っている²⁵。企業が次世代の音声対話技術を自社製品に実装し、グローバル市場における競争優位性を確立するためには、以下の三つの戦略的アプローチを組織全体で徹底することが強く求められる。

7.1. 「後付け開発」からの脱却と「音声ファースト設計」の徹底

多くの企業が犯す最大の過ちは、音声インターフェースを既存のタッチ中心の UI に対する「新機能」や「代替の入力ボタン」として扱ってしまうことである²。マイクのアイコンを押

して話し、結果を画面で確認して再びタップ操作に戻るような中途半端なフローは、音声対話を持つ本来のポテンシャルを殺してしまう。

音声対話は、それ自体が独自の設計パラダイムを要求する根本的に異なるインタラクションのモードである。製品企画の初期段階から、音声こそが「主要なインターフェース (Primary Interface)」であるという前提に立ち、システムをゼロから設計し直さなければならない²⁾。画面を一切見なくてもタスクが完結するアンビエントな操作体系を構築し、ユーザーの自然な「言い淀み」「言い間違い」、そして「割り込み」を処理できる Full-duplex 対応のアーキテクチャを標準仕様として採用すべきである。

7.2. 研究開発から運用保守までを貫く一貫したパートナーシップの構築

音声 AI システムの実装は、特定のアルゴリズムを納品して終了するものではない。AI による技術的差別化を図るためには、現状の業務プロセスや製品特性を深く分析し、数ある技術 (画像認識、音声処理、自然言語処理など) の中から最も適合するものを専門的知見に基づいて選定・カスタマイズする「戦略立案とコンサルティング」のフェーズが不可欠である²⁵⁾。

Convergence Lab. が提案する受託研究開発の理想的なフローが示すように、企業は単なる下請けの開発ベンダーではなく、「要件定義」から「実装」、そして本番環境での「運用保守」に至るまで、エンドツーエンドで一貫して伴走できる技術パートナーを持つべきである²⁵⁾。特に、クラウドと NVIDIA Jetson 等のエッジハードウェアを組み合わせたハイブリッド・アーキテクチャの構築においては、ハードウェアの制約とソフトウェアのモデル軽量化 (LoRA 等) の両方を深く理解した高度な摺り合わせの技術が求められる¹⁷⁾。

7.3. Human-in-the-Loop (HITL) を前提とした堅牢なシステム運用

最後に、経営層および開発現場が認識しなければならない冷徹な事実、「AI の推論精度は本質的に 100% には到達しない」ということである²⁵⁾。どれほど高度な Audio LLM を用いたとしても、未知のノイズ環境や想定外のコンテキストにおいては、誤認識やハルシネーションが発生する。特に医療モニタリング、自動車の制御、あるいは重要インフラの予知保全など、実業務におけるエラーが重大なインシデントに直結する領域においては、AI の完全自動化を盲信することは極めて危険である。

したがって、システムのアーキテクチャの中に、あらかじめ人間の介入を想定した設計を組み込む必要がある。AI の判断に対して、必要に応じて人間のオペレーターがレビューを行い、あるいはユーザー自身が音声で即座に修正を指示できる「Human-in-the-Loop (HITL)」の仕組みをシステム運用フローの中核に据えることが極めて有効である²⁵⁾。AI が 100% 完璧ではないことをシステムの前提条件として受け入れつつ、人間と AI が相互に補完し合う協調的なワークフローを通じて、業務への適合性とモデルの精度を継続的に向上させていく漸進的なアプローチこそが、最も確実に ROI の高い実装戦略となる。

8. 結論：環境に溶け込む知能が導く未来

コンシューマーエレクトロニクスにおける音声対話技術は、長らく続いた停滞期を脱し、今まさに技術的特異点を迎つつある。テキスト変換というボトルネックを排除し、生の音響データから豊かな感情と空間コンテキストを直接理解する Audio LLM。人間同士の会話に特有の被せ合いや割り込みを許容し、遅延のない自然なコミュニケーションを可能にする全二重 (Full-duplex) 通信技術。そして、これら膨大な処理をクラウドに依存することなく、プライバシーを厳格に保護しながら極小の遅延で実行可能にするセルフホスト AI ハードウェアの革新。これら要素技術の高度な融合により、デバイスは単なる「操作される機械」から「環境に溶け込み、自律的にユーザーを支援するプロアクティブなパートナー」へと完全に変貌を遂げる。

企業にとって、この VUI を中心としたパラダイムシフトは、自社製品に新たな機能を追加するだけのマイナーチェンジの機会ではない。それは、ユーザーと製品との関わり方、ひいて

はブランドが提供する顧客体験そのものを根本から再構築するための、不可逆的かつ戦略的な転換点である。

次世代の市場において圧倒的な競争優位性を確立するためには、最新の学術的知見を絶えず追究するとともに、エッジハードウェアからクラウドインフラに至るエンドツーエンドの実装力を有し、ユーザーのプライバシー保護を絶対的な前提としたアーキテクチャを設計し抜く覚悟が求められる。この技術的進化の波を的確に捉え、自社のビジネスモデルに深く統合できた企業こそが、来るべき「アンビエント AI」時代のエレクトロニクス市場を牽引していくことになるだろう。

引用文献

1. 8 Emerging Trends Shaping Consumer Electronics. - Kadence, 3月9, 2026 にアクセス、
<https://kadence.com/knowledge/8-emerging-trends-shaping-the-consumer-electronics-segment/>
2. Voice is the Interface - Synervoz, 3月9, 2026 にアクセス、 <https://synervoz.com/blog/voice-is-the-interface/>
3. The impact of voice assistants on consumer behavior - PwC, 3月9, 2026 にアクセス、 <https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html>
4. The Top CES 2026 Highlights – and the CX Shift Hiding in Plain Sight - CMSWire, 3月9, 2026 にアクセス、 <https://www.cmswire.com/digital-experience/the-most-buzz-worthy-tech-from-ces/>
5. Top 5 trends you can expect from CES 2026 in Las Vegas - Arm Newsroom, 3月9, 2026 にアクセス、 <https://newsroom.arm.com/blog/top-trends-for-ces-2026>
6. Voice User Interfaces in 2026: The Future of Ambient AI Now - Zignuts, 3月9, 2026 にアクセス、 <https://www.zignuts.com/blog/voice-user-interfaces>
7. Japan Smart Speakers Market Outlook [2025 and 2032] Key Drivers | Challenges | Opportunities - Apiary, 3月9, 2026 にアクセス、 <https://japansmartspeakersmarket.docs.apiary.io/>
8. 日本の会話 AI 市場は 2033 年までに 30 億 9,200 万米ドルに達する見込み - IMARC Group, 3月9, 2026 にアクセス、 <https://www.imarcgroup.com/pressrelease/ja/japan-conversational-ai-market-statistics>
9. Demand for Connected TVs in Japan | Global Market Analysis Report - 2035, 3月9, 2026 にアクセス、 <https://www.futuremarketinsights.com/reports/japan-connected-tvs-market>
10. Voice Search Trends 2025: Statistics, Industry Insights, and SEO Strategies - SevenAtoms, 3月9, 2026 にアクセス、 <https://www.sevenatoms.com/blog/voice-search-trends>
11. In the gen AI economy, consumers want innovation they can trust - Deloitte, 3月9, 2026 にアクセス、 <https://www.deloitte.com/us/en/insights/industry/telecommunications/connectivity-mobile-trends-survey.html>
12. PERSONAPLEX: VOICE AND ROLE CONTROL FOR FULL DUPLEX CONVERSATIONAL SPEECH MODELS - Research at NVIDIA, 3月9, 2026 にアクセス、 [https://research.nvidia.com/labs/adlr/files/personaplex/personaplex_preprint.pdf](https://research.nvidia.com/labs/adlr/files/personaplex/personaplex_preprint.pdf)

13. Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone, 3月9, 2026 にアクセス、 <https://research.google/blog/google-duplex-an-ai-system-for-accomplishing-real-world-tasks-over-the-phone/>
14. Teaching Physical Awareness to LLMs through Sounds - arXiv, 3月9, 2026 にアクセス、 <https://arxiv.org/html/2506.08524v1>
15. AudioLLMs/Awesome-Audio-LLM: Audio Large Language Models - GitHub, 3月9, 2026 にアクセス、 <https://github.com/AudioLLMs/Awesome-Audio-LLM>
16. DESAMO: Smart Homes for Elders with Audio LLM - Kukarella, 3月9, 2026 にアクセス、 <https://www.kukarella.com/news/desamo-smart-homes-for-elders-with-audio-llm-p1756296002>
17. 音声解析 AI の開発事例 - Convergence Lab., 3月9, 2026 にアクセス、 <https://www.convergence-lab.com/blog/2022-04-15-2022-04-15/>
18. Full-Duplex-Bench v1.5: Evaluating Overlap Handling for Full-Duplex Speech Models - arXiv, 3月9, 2026 にアクセス、 <https://arxiv.org/html/2507.23159v1>
19. Unlocking the Future of Voice AI: Introducing the Duplex Conversation Datasets on MagicHub.com, 3月9, 2026 にアクセス、 <https://magichub.com/unlocking-the-future-of-voice-ai-introducing-the-duplex-conversation-datasets-on-magichub-com/>
20. Solving Voice AI Latency Could Herald an Entirely New Human-Computer Interaction, 3月9, 2026 にアクセス、 <https://www.speechtechmag.com/Articles/Editorial/Industry-Voices/Solving-Voice-AI-Latency-Could-Herald-an-Entirely-New-Human-Computer-Interaction-166799.aspx>
21. The Smart Home Privacy Gap: Why Cloud Voice Fails Appliances | Sensory, 3月9, 2026 にアクセス、 <https://sensory.com/blog-smart-home-privacy/>
22. Smart Speakers Market Outlook 2025-2032 - Intel Market Research, 3月9, 2026 にアクセス、 <https://www.intelmarketresearch.com/smart-speakers-2025-2032-605-948>
23. Technical challenges of voice privacy | Skyted, 3月9, 2026 にアクセス、 <https://www.skyted.io/blog/technical-challenges-voice-privacy/>
24. CES 2026 Ushers In AI Revolution Across Devices - Evrim Ağacı, 3月9, 2026 にアクセス、 <https://evrimagaci.org/gpt/ces-2026-ushers-in-ai-revolution-across-devices-521918>
25. 事業紹介 - Convergence Lab.株式会社, 3月9, 2026 にアクセス、 <https://www.convergence-lab.com/business/>
26. CES 2026 Decoded: Physical AI, Voice Interfaces & Grid Chaos - Mobisoft Infotech, 3月9, 2026 にアクセス、 <https://mobisoftinfotech.com/resources/newsletters/mobisoft-in-the-loop-ai-and-tech/ces-2026-physical-ai-voice-interfaces-energy>
27. 2025年最新動向：セルフホスト AI デバイス普及とプライバシー保護の最前線、オフライン処理能力向上とデータセキュリティの技術的解決策 | 海辺の部屋 - note, 3月9, 2026 にアクセス、 <https://note.com/umibenoheya/n/neb8cd65454c7>

28. Voice Recognition Market Size and YoY Growth Rate, 2025-2032, 3 月 9, 2026 にアクセス、
<https://www.coherentmarketinsights.com/industry-reports/voice-recognition-market>
29. 2025 年に知っておくべき AI 搭載のトップ音声アシスタント - PageOn AI, 3 月 9, 2026 にアクセス、
<https://www.pageon.ai/jp/blog/ai-powered-voice-assistants>