

JEITA 一般社団法人 電子情報技術産業協会

令和 5(2023)年度 音声認識・合成技術 応用製品の動向について

— 音声入出力方式標準化専門委員会 実施の製品動向調査結果より —

2024 年 10 月

音声入出力方式標準化専門委員会

○概要

音声認識・合成技術の適用分野の拡大、応用製品の普及を目的として、本委員会では継続して製品動向調査を実施している。本調査は応用製品の開発者、製品の導入・購入を考える企業・個人に有益な情報を提供するとともに、継続実施することにより、技術の進歩、利用領域の変遷を理解する貴重な資料となっている。

本年度（令和 5 年度）の調査は平成 22 年度以降の調査方法を基本的に踏襲し、以下の方法により実施した。

(1) 情報提供の依頼

令和 6（2024）年 1 月に JEITA 本委員会ホームページに情報提供の依頼を掲示するとともに、音声研究・開発者向けメーリングリスト（onsei-mail@aist.jpn.org）に同一内容のメールを配信。なお、前年に情報を提供いただいた企業には直接メールを発信し、依頼した。

(2) 掲載製品

情報提供企業が製造あるいは販売する音声認識・合成関連製品であり、令和 5（2023）年 12 月末現在で日本国内で購入可能な製品とした。

(3) 提供内容

委員会で作成した、製品分類ごとの製品特徴・仕様を記載するフォームを配布、記載いただき電子ファイルとして受領した。

(4) 掲載内容

情報提供企業からいただいた内容は基本的に加工を行わず掲載している。ただし、内容の変更を伴わない書式の統一、情報提供者の明らかな誤解による分類誤りの修正は、委員会の判断にて行っている場合がある。

○本年度(令和 5 年・2023 年)調査からの改訂のポイント

本年度の調査から、以下のポイントについて調査方式の見直しを行った。

・ 掲載内容（回答項目）

調査結果の読者が必要とする情報の詳細度、回答ベンダによる記入の負荷、回答ベンダにとって公開がメリットとなる詳細度のレベルなどを考慮し、今回の調査から掲載内容（回答項目）の一部を見直した。

・ 掲載対象（「委員会による調査」の追加実施）

前年までの調査（ベンダからの回答による調査）では、回答ベンダにより記入された回答項目を元に JEITA にて回収・集計し、調査結果として公開していた。これに対して本年からは、前年までの「ベンダからの回答による調査」に加えて、JEITA 音声入出力方式標準化専門委員会（以下 本委員会）による調査（委員会による調査）の結果も調査報告書に加えた。これは、本委員会から直接的に回答協力を依頼・回収することが困難なベンダも多く、調査結果の網羅性が低くなってしまう事態に対応したためである。

○音声合成のシステム開発者向け製品の動向調査結果(ベンダからの回答による調査)

日本国内で販売されている音声合成製品を調査した結果について以下に概要をまとめる。

製品カテゴリーの分類

今回の調査（2023 年度調査）で採用している製品分類の 7 カテゴリーを 表 1 に示す。

表 1 音声合成製品の製品分類

ID	分類	
A-1	組み込み用 ミドルウェア・SDK	組み込み OS にて動作するライブラリ提供の音声合成ミドルウェア製品。
A-2	汎用 OS 用 ミドルウェア・SDK	汎用 OS 上で動作するライブラリとして提供される音声合成ミドルウェア製品。アプリケーションに組み込むことで動作する製品。 対象 OS: windowsOS, MacOS, iOS, AndroidOS を始め, Linux OS も含む。
A-3	サーバ用 ミドルウェア	クライアントサーバ構成を実現できるなど, アプリ単体に組み込んでランタイムのみで動作するというような制限がなく, 外部からの呼び出しで動作する音声合成エンジンソフトウェア。 マルチスレッドで動作するなど, 複数リクエストを受けることのできる製品。 対象 OS: windows Server OS, Linux OS など。
A-4	Web サービス	利用者が音声合成エンジンの動作するソフトウェアを入手することなく利用可能な提供形態の音声合成サービス・製品。webAPI にて提供されるクラウドサービスなど。
A-5	ツール製品	利用者が音声ファイルを生成したり, 辞書記述を編集するなど, ツールとして利用可能な製品・サービス。
A-6	ハードウェア製品	ハードウェアとして提供される音声合成製品。LSI 提供, ボード提供。
A-7	その他	上記 6 カテゴリーに当てはまらない音声合成関連製品・サービス。

音声合成製品の製品件数とベンダ数

調査結果を、「合成関連製品一覧表（別 pdf ファイル）」に示す（別冊 1 に表 A-1～表 A-3, 別冊 2 に表 A-4～表 A-7 を所収）。

以下では, 今年度の調査結果をもとに, 調査対象とした 7 カテゴリーの分類ごとの製品件数と製品を提供しているベンダ数について述べる。表 2 に調査結果を示す。

今年度の回答製品数は総計 79 件であった。製品分類カテゴリごとに製品件数をみると,

- 「汎用 OS 用ミドルウェア・SDK」製品（表 A-2, 17 件）
- 「ツール製品」（表 A-5, 17 件）

の製品数が最も多かった。

続いて

- 「その他」製品（表 A-7, 13 件）

- 「組込み用ミドルウェア・SDK」製品（表 A-1, 11 件）

表 2 分類毎の音声合成製品の件数と製品を提供しているベンダ数

カテゴリID	分類カテゴリ	R5 (2023) 製品件数	R5 (2023) ベンダ数
表 A-1	組込み用 ミドルウェア・SDK	11	6
表 A-2	汎用 OS 用 ミドルウェア・SDK	17	11
表 A-3	サーバ用 ミドルウェア	5	4
表 A-4	Web サービス	9	5
表 A-5	ツール製品	17	8
表 A-6	ハードウェア製品	7	2
表 A-7	その他	13	7
合計		79	43

の製品件数が 10 件以上であった。

また、今年度の提供ベンダ数は総計 43 社であった。ここで「提供ベンダ数」とは「のべベンダ数」であり、複数の製品カテゴリに回答したベンダは重複してカウントされている。

製品分類カテゴリごとにベンダ数をみると、7 つの製品カテゴリの中では、

- 「汎用 OS 用ミドルウェア・SDK」製品（表 A-2, 11 社）

の提供ベンダ数が最多であった。

以下は

- 「ツール製品」（表 A-5, 8 社）
- 「その他」製品（表 A-7, 7 社）
- 「組込み用ミドルウェア・SDK」製品（表 A-1, 6 社）
- 「Web サービス」製品（表 A-4, 5 社）
- 「サーバ用ミドルウェア」製品（表 A-3, 4 社）

のように続いた。

- 「ハードウェア」製品（表 A-6, 2 社）

は 2 社の回答が寄せられた。

なお、複数のカテゴリに製品を提供しているベンダを重複してカウントせずに集計すると、調査に回答したベンダ（異なり数）の総数は 18 社であった。

音声合成製品の製品件数と提供ベンダ数の変化

以下では、2023年度調査までの5年間の製品数（表3）とベンダ数（表4）を示す。

表3 分類毎の音声合成製品の製品件数と推移

ID	新しい分類	R5	R4	R3	R2	R1
		(2023)	(2022)	(2021)	(2020)	(2019)
		件数(±)	件数	件数	件数	件数
A-1	組込み用 MW SDK	11 (-4)	15	17	17	23
A-2	汎用 OS 用 MW SDK	17 (-4)	21	25	25	30
A-3	サーバ用 MW	5 (-2)	7	10	10	10
A-4	Web サービス	9 (-3)	12	13	11	10
A-5	ツール製品	17 (-3)	20	20	17	19
A-6	ハードウェア製品	7 (+2)	5	4	3	1
A-7	その他	11 (-3)	16	8	9	8
合計		79 (-17)	96	97	92	101

表4 分類毎の音声合成製品の提供ベンダ数とその推移

ID	新しい分類	R5	R4	R3	R2	R1
		(2023)	(2022)	(2021)	(2020)	(2019)
		ベンダ数 (±)	ベンダ数	ベンダ数	ベンダ数	ベンダ数
A-1	組込み用 MW・SDK	6 (-2)	8	9	9	10
A-2	汎用 OS 用 MW・SDK	11 (-1)	12	12	12	13
A-3	サーバ用 MW	4 (-2)	6	8	8	8
A-4	Web サービス	5 (-4)	9	9	8	6
A-5	ツール製品	8 (-4)	12	11	10	9
A-6	ハードウェア製品	2 (±0)	2	2	2	1
A-7	その他	7 (-3)	10	5	6	5
合計		43 (-16)	59	56	55	52

今年度の回答製品数は総計 79 件であった。製品件数の総数を前年度の調査と比較すると、17 件ほど減っている（表 3）。「組込み用 MW SDK」製品（表 A-1）で 4 件が減、「汎用 OS 用 MW SDK」製品（表 A-2）で 4 件が減、「サーバ用 MW」製品（表 A-3）で 2 件が減、「Web サービス」製品（表 A-4）で 3 件が減、「ツール製品」（表 A-5）で 3 件が減、「その他」製品（表 A-7）で 3 件が減った一方、「ハードウェア製品」（表 A-6）で 2 件が増えた。

最も減少幅が大きい 4 件の減少があった「組込み用 MW SDK」製品（表 A-1）では、製品の統合

による減少が 1 件，製品カテゴリの変更による減少が 1 件，今年度回答のなかった 1 社による減少が 1 件で，実質的な減少は製品の削除による 1 件であった。同じく 4 件減った「汎用 OS 用 MW SDK」製品（表 A-2）については，今年度回答のなかった 2 社による減少が 2 件，製品の統合による減少が 2 件であり，実質的な減少であるかの判断は難しい。他のカテゴリでの減少の大部分は，今年度回答のなかったベンダによる減少であった。

今年度の提供ベンダ数は総計 43 社であった。提供ベンダ数を前年度の調査と比較すると，16 社ほど減っている（表 4）。「組込み用 MW SDK」製品（表 A-1）で 2 社，「汎用 OS 用 MW SDK」製品（表 A-2）で 1 社，「サーバ用 MW」製品（表 A-3）で 2 社，「Web サービス」製品（表 A-4）で 4 社，「ツール製品」（表 A-5）で 4 社，「その他」製品（表 A-7）で 3 社減った。

一方，複数のカテゴリに製品を提供しているベンダを重複してカウントせずに集計したときの，調査に回答したベンダ（異なり数）の総数は 18 社で，前年度の調査より 8 社減った。今年度回答がなかった 6 社による減少と，M&A による 2 社の減少により，トータルで 8 社の減少となった。

○音声合成製品:JEITAによる調査(委員会調査)の結果

前節では、回答ベンダにより記入・回収された回答に基づいて、音声合成製品を調査した結果（ベンダからの回答による調査）について述べた。本節では、JEITA 音声入出力方式標準化専門委員会（本委員会）による調査（委員会調査）の結果について述べる。ここでは、上記調査（ベンダからの回答による調査）では収集しきれなかった音声合成製品について、インターネット上の web サイトで収集可能な範囲で情報を集め、集約した。なお「製品数」の数え方については当該のベンダによる確認を経ることができないため、情報は「提供者単位」とした。

表 5 JEITA 委員会調査による音声合成製品の提供者数

カテゴリID	分類カテゴリ	R5 (2023) 委員会調査・提供者数	R5 (2023) ベンダ調査・ベンダ数
表 A-1	組込み用 ミドルウェア・SDK	—	6
表 A-2	汎用 OS 用 ミドルウェア・SDK	—	11
表 A-3	サーバ用 ミドルウェア	—	4
表 A-4	Web サービス	12	5
表 A-5	ツール製品	16	8
表 A-6	ハードウェア製品	—	2
表 A-7	その他	2	7
合計		30	43

本年から開始した「委員会による調査」の結果、提供者(ベンダ)数にして延べ 30 提供者について、新たに調査結果に加えることとなった。しかし、このうち 12 提供者(異なり数にして 6 提供者)は、前年まで「ベンダ回答による調査」にて掲載していたが、本年から回答回収の協力が得られず、「委員会による調査」での掲載へと移行した提供者である。従って、本年からの「委員会による調査」で新たに調査結果に加えることとなった純増の提供者は 18 提供者となる。

言い換えると、従来からの「ベンダ回答調査」での総ベンダ数が前々年の延べ 59 社から、本年は延べ 43 社と 16 社の減となっているものの、減っている 16 社のうち 12 提供者は「ベンダ回答による調査」での掲載から、「委員会による調査」での掲載へと移行しているケースである。

従来からの「ベンダ回答による調査」と本年からの「委員会による調査」の双方の結果を総合して概観すると、やはり「A-4 Web サービス」「A-5 ツール製品」の充実ぶりが見てとれる。

○令和 5 年(2023 年)度の音声合成製品の全般的な動向

2023 年の音声合成関連製品市場では、以下のような動向が見られた。

(業界構造の変化)

2023 年は、2022 年までに進行していた業界構造の変化がさらに本格化した年といえる。従来から音声合成製品・サービスを提供していた企業に加えて、新興企業やオープンソースを活用した製品・サービス提供者が台頭した¹。この変化に伴い、カスタムボイスのバリエーションが従来に比べ大幅に増加していることは象徴的と言える。業界構造のオープン化²という変化は、音声合成エンジンの製品・サービス提供だけでなく、素材となるボイス提供者にも企業やプロ声優以外からの参入というオープン化を起こしている³。また、感情を表現できる音声合成技術の普及も進んだ⁴。さらに、生成 AI の技術進化により、音声対話システムを開発する際に、音声合成の応答テキストの作成が格段に容易になっている。

(実社会への影響)

2023 年には、ディープフェイク技術による合成音声が悪用されるなど、犯罪性のあるニュースが現実になっている⁵。これにより、技術の利用に伴う倫理的な問題や社会への影響についての議論が活発化しており、もはや未来の話としてではなく、目の前の現実への対応が迫られている時代となっている⁶。

¹ 本報告書 (2023 年 JEITA 音声合成関連製品動向調査結果) など参照。

² 【2023 年 11 月時点】日本語 TTS モデルで高品質な OSS まとめ. 2023 年 11 月 5 日。
<https://zenn.dev/k99758/articles/f52ec6382c6db0>

³ MYCOEIROINK (MYCOE).
<https://coeiroink.com/mycoeiroink>

⁴ 音声合成 AI 「Style-Bert-VITS2」 のインストール方法・使い方！感情豊かな日本語ボイスを生成しよう。くろくまそふと(2024 年 2 月 14 日)
<https://kurokumasoft.com/2024/02/14/style-bert-vits2/>

⁵ AI を悪用した音声詐欺が世界で増加中！10%が遭遇する音声詐欺から身を守る方法は？
Audiostart News. (2023 年 5 月 26 日)
<https://audiostart.info/2023/05/26/ai-voice-imposter/>

⁶ 進化するディープフェイクボイスチェンジ攻撃とその見抜き方 4 つ. 安野貴博 note (2023 年 5 月 7 日)
<https://note.com/takahiroanno/n/naf00564d384b>

○音声認識の製品動向調査(ベンダからの回答による調査)

本年度の調査にあたって、その製品分類・記載項目について、近年の製品形態の傾向を鑑み、昨年度の分類を見直した。

まず、製品分類に関し、以下の2つの変更を行った。

- a) システム開発者向けの PC 用のみに特化した製品が無いことからその分類を削除した。
- b) エンドユーザ向け応用製品について、それを昨年度までのように「応用製品」と「応用ソフトウェア製品」にさらに分類することは、その定義の曖昧さから、回答者にとって難しい例も少なからず出ていたため、それらを統合し「応用製品」という1つの項目とした。

これらの結果、今年度の製品分類はシステム開発者を対象とした①組込み向け（ボード等含む）、ソリューション利用者を対象とした、②CTI ソリューション、③Web API サービスソリューション、④その他ソリューション、そしてエンドユーザを対象とした、⑤応用製品の、計 5 つに分類となった。

次に、記載項目に関しても検討し、昨年度までの項目は、この報告書の対象読者にとって必要以上に詳細であり、製品を比較するときに、それらの仕様詳細が重要となる場合は極めて稀であると判断した。また、回答する各ベンダの立場としては、各社の守秘義務規定に抵触する恐れのある仕様詳細項目も含まれていること、あるいは定められた仕様詳細項目のみを基準にして競合他社と比較されることは、場合によっては回答者側に不利益をもたらす懸念もあり、回答しづらい項目も含まれていた。

そこで、昨年度までの仕様詳細項目を、改めて本報告書の読者およびベンダの両観点から見直した結果、製品の技術的特徴については、各ベンダに自由に記述して頂く形をとることにより、昨年度までの仕様詳細項目を大幅に簡素化した。具体的には、以下に示す項目とした。

- カテゴリ（次から選択）
 - 組込み向け（ボード等含む）音声認識エンジン
 - ソリューション (IVR CTI CRM)
 - ソリューション (Web API サービス)
 - ソリューション (その他)
 - 応用製品
- 製品名/型番
- 製品情報の URL
- 製品概要
- 音声認識機能・特徴
- 対応言語
- 動作環境
- 想定用途・利用シーン（カテゴリが「エンジン」「ソリューション(その他)」「応用商品」の場合記入）
- アプリケーション開発ツール（カテゴリが「ソリューション(IVR CTI CRM)」の場合記入）
- 通信プロトコル（カテゴリが「ソリューション(Web API サービス)」の場合記入）

- 連絡先
- 備考

ここで、この項目の変更により、昨年度まで行っていた製品件数の推移観測は実施できなくなる。ただ、近年は製品件数の推移に大きな変化は見られていなかったこと、そして、近年の製品形態の変化により、何をもちて 1 件の製品と見做すかの定義が困難になってきていること、などの理由により、製品件数の推移観測が実施できなくなることについて、特段問題にはならないと判断した。

○音声認識製品:JEITAによる調査(委員会調査)

また、前節で述べた、JEITAによる調査に関しては、報告書の網羅性を高めることを目的とし、その製品・サービスが一般公開されている場合はできる限り掲載する方針を採って、本委員会が自主的に調査を行った。JEITAによる調査の調査対象は、昨年度の調査まで種々の理由により複数年に亘って回答が得られていないベンダおよび今年度の調査で締切日までに回答が得られなかったベンダとした。JEITAによる調査の項目を以下に示す。

- カテゴリ(次から選択)
 - 組込み向け(ボード等含む)音声認識エンジン
 - ソリューション (IVR CTI CRM)
 - ソリューション (Web API サービス)
 - ソリューション (その他)
 - 応用製品
- 製品名/型番
- 製品情報の URL
- 製品概要
- 備考

ここで、JEITAによる調査のベンダ名について、外国に本社があるものについては、それぞれ一般に広く知られている商標の企業名を採用した。

以上のように、ベンダからの回答による調査と、JEITAによる調査の2種類の調査を行った。そして、その2種類の調査結果は統合せず、ベンダからの回答による調査、JEITAによる調査、の順に、それぞれ報告書に掲載した。

ベンダからの回答による調査では、17社から延べ59製品の回答を得た。また、JEITAによる調査では、13社の27製品について調査した。

調査結果をカテゴリ別に見る。「組込み向け(ボード等含む)音声認識エンジン」については、省リソースのための軽量設計という特徴に加え、車載カーナビを想定する製品は、ノイズ耐性が重要となっている。また、端末側のハードウェア性能の向上に伴い、ネットワーク通信が可能な環境ではサーバ側で処理をする特徴も見られ、別カテゴリの「ソリューション(Web API サービス)」との境界が曖昧になってきている。つまり、オフライン時に単体で処理できることも、このカテゴリの重要な特徴のうちの一つとなっている。「ソリューション(IVR CTI CRM)」「ソリューション(Web API サービス)」「ソリューション(その他)」について、これらのカテゴリの製品は、他システムとの連携の柔軟性、電話以外のチャネルも含めた管理情報の一元化、書き起こしや要約、感情分析などの情報の可視化が特徴となっている。また、例えばIVR用のWeb APIを提供しているなど、これら3つのうちの複数のカテゴリにまたがる製品も多く、近年の製品の実状に沿えずカテゴリ分けが曖昧になってしまったことは来年度以降の課題である。「応用製品」については、様々な製品が出されているが、中でも、オンライン以外の会議音声の書き起こしという課題に対しては複数のベンダから製品が出されており、オンライン会議の書き起こしも含め、継続した需要があることが分かる。

○令和 5 年(2023 年)度の音声認識製品の全般的な動向

生成 AI 技術開発の盛り上がりが昨年度以降も続いており、それを応用した様々なアプリケーションが市場に登場してきている。一方、本調査は、音声を変換する音声認識技術に焦点を当てているため、音声認識技術の日本国内の適用分野は従来どおり、コールセンタ(IVR)、カーナビ、家庭用スマートスピーカ、モバイル端末での入力インタフェース、そして、会議における音声の書き起こしなどが中心となっている。ただし、この生成 AI 技術の普及によって、音声認識結果を生成 AI エンジンに入力して要約文章や翻訳文章を出力させるなど、音声認識を生成 AI の入力インタフェースとして利用する事例が増えてきている。これらの事例が示しているように、生成 AI アプリケーションが、今後、より多くのユーザにとって、音声認識技術を使うきっかけとなっていく可能性は大いにあると思われる。

また、音声認識技術の利用環境については、技術開発による近年の著しい精度向上によって、業界全体の音声認識技術の品質も向上しており、また、米国のプラットフォーム企業を中心に、SaaS 形態での音声認識サービスや、オンライン会議ツールの機能の一部として書き起こし・概要作成機能などを提供しているため、個人・法人に関わらず、誰でも安価に品質の良い音声認識技術が利用できる環境の普及は急速に進んでいる。言い換えると、音声認識技術のコモディティ化が急速に進んでいると言える。

このように、生成 AI という新たな技術の出現や、音声認識技術のコモディティ化が進むことによって、音声認識技術の普及が加速することは、社会にとって歓迎すべきことである。一方、急速に進む人口の高齢化に伴う労働人口の減少や介護リソースの不足など、音声技術の貢献できる可能性がある課題はまだ山積している。今後のさらなる技術の発展が望まれる。