

JEITA 一般社団法人 電子情報技術産業協会

令和 3(2021)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 —

2022 年 8 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 3 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 3（2021）年 12 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.jp）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 3（2021）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

提供いただいた情報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については，委員会の判断にて変更をしている場合がある．

記載されている内容については委員会にて確認をしておらず，内容はすべて情報提供企業にその責任がある．

○音声合成のシステム開発者向け製品の動向調査結果

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。

製品カテゴリの分類

今回の調査（2021 年度調査）で採用している製品分類の 7 カテゴリを 表 1 に示す。

表 1 音声合成製品の製品分類

ID	分類	
A-1	組込み用 ミドルウェア・SDK	組込み 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 に調査結果示す。

今年度の回答製品数は総計 97 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, 25 件)

の製品数が最多であった。

続いて

- 「ツール製品」(表 A-5, 20 件)
- 「組込み用ミドルウェア・SDK」製品 (表 A-1, 17 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	R3 (2021) 製品件数	R3 (2021) ベンダ数
表 A-1	組込み用 ミドルウェア・SDK	17	9
表 A-2	汎用 OS 用 ミドルウェア・SDK	25	12
表 A-3	サーバ用 ミドルウェア	10	8
表 A-4	Web サービス	13	9
表 A-5	ツール製品	20	11
表 A-6	ハードウェア製品	4	2
表 A-7	その他	8	5
合計		97	56

の製品件数が多かった。

- 「web サービス」製品（表 A-4, 13 件）
- 「サーバ用ミドルウェア」製品（表 A-3, 10 件）

も製品件数は 10 件以上であった。

また、今年度の提供ベンダ数は総計 56 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, 12 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, 11 社）
- 「組込み用ミドルウェア・SDK」製品（表 A-1, 9 社）
- 「web サービス」製品（表 A-4, 9 社）
- 「サーバ用ミドルウェア」製品（表 A-3, 8 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 24 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、2021 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

新 ID	新しい分類	R3	R2	R1	旧 ID	旧分類 (第 2 分類)	H30	H29
		(2021)	(2020)	(2019)			(2018)	(2017)
		件数(±)	件数	件数			件数	件数
A-1	組込み用 MW SDK	17 (±0)	17	23	A-1	MW 組込み用 音声合成	30	33
A-2	汎用 OS 用 MW SDK	25 (±0)	25	30	A-3	MW PC 用 音声合成	26	24
A-3	サーバ用 MW	10 (±0)	10	10	A-7	その他 サーバ製品	11	12
A-4	Web サービス	13 (+2)	11	10	A-9	その他 サービス	11	15
A-5	ツール製品	20 (+3)	17	19	A-8	その他 PC ソフト・ツール	21	23
A-6	ハードウェア製品	4 (+1)	3	1	A-5	ハードウェア 音声合成	1	1
A-7	その他	8 (-1)	9	8	A-2	MW 組込み用 その他	1	1
					A-4	MW PC 用 その他	0	2
					A-6	ハードウェア その他	2	2
合計		97 (+5)	92	101	合計		103	113

表 4 分類毎の音声合成製品の 提供ベンダ数 とその推移

新 ID	新しい分類	R3	R2	R1	旧 ID	旧分類 (第 2 分類)	H30	H29
		(2021)	(2020)	(2019)			(2018)	(2017)
		件数(±)	社数	社数			社数	社数
A-1	組込み用 MW・SDK	9 (±0)	9	10	A-1	MW 組込み用 音声合成	11	12
A-2	汎用 OS 用 MW・SDK	12 (±0)	12	13	A-3	MW PC 用 音声合成	13	13
A-3	サーバ用 MW	8 (±0)	8	8	A-7	その他 サーバ製品	9	10
A-4	Web サービス	9 (+1)	8	6	A-9	その他 サービス	6	9
A-5	ツール製品	11 (+1)	10	9	A-8	その他 PC ソフト・ツール	10	11
A-6	ハードウェア製品	2 (±0)	2	1	A-5	ハードウェア 音声合成	1	1
A-7	その他	5 (-1)	6	5	A-2	MW 組込み用 その他	1	1
					A-4	MW PC 用 その他	0	1
					A-6	ハードウェア その他	2	2
合計		56 (+1)	55	52	合計		53	60

今年度の回答製品数は総計 97 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、5 件ほど増えている（表 3）。「ツール製品」（表 A-5）で 3 件、「web サービス」製品（表 A-4）で 2 件¹、「ハードウェア」製品（表 A-6）で 1 件増え、「その他」製品（表 A-7）で 1 件減った。

今年度の提供ベンダ数は総計 56 社であった。提供ベンダ数を前年度の調査と比較すると、1 社ほど増えている（表 4）。「web サービス」製品（表 A-4）と「ツール製品」（表 A-5）でそれぞれ 1 社増え²、「その他」製品（表 A-7）で 1 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 24 社で、前年度の調査から変わらなかった。

【参考】2019 年までの分類カテゴリとの対応

昨今の技術の進展と動向に合わせて、2019 年の調査から分類のカテゴリを見直した。2019 年度調査までの分類カテゴリと現在の分類カテゴリは必ずしも 1:1 に対応する訳ではないが、参考までに旧分類カテゴリと現在の分類カテゴリの対応を表 5 に示す。

表 5 音声合成製品の新旧分類カテゴリの対応

新 ID	新しい分類	旧 ID	前年度調査までの分類	
			第 1 分類	第 2 分類
A-1	組み込み ミドルウェア・SDK	A-1	ミドルウェア・組み込み用	規則音声合成
A-2	汎用 OS 用 ミドルウェア・SDK	A-3	ミドルウェア・PC 用	規則音声合成
A-3	サーバ用 ミドルウェア	A-7	その他	サーバ製品
A-4	Web サービス	A-9	その他	サービス
A-5	ツール製品	A-8	その他	PC ソフト製品・ツール製品
A-6	ハードウェア製品	A-5	ハードウェア	規則音声合成 (規則音声合成 LSI)
A-7	その他	A-2	ミドルウェア・組み込み用	その他
		A-4	ミドルウェア・PC 用	その他
		A-6	ハードウェア	その他

¹ 前年度(2020 年度)の報告書では「web サービス」製品の件数を 10 件としていたが、11 件の誤りであったため、前年度の調査からの比較で 2 件の増加となる。

² 同じく「web サービス」提供ベンダ数を 7 社としていたが、8 社の誤りであったため、前年度の調査からの比較では 1 社の増加となる。

○音声認識の製品動向調査

本年度調査にあたって、その製品分類・記載項目について検討し、製品分類を一部見直すこととした。具体的には、製品分類は音声認識製品を使用する立場から、①システム開発者向けとして、組込み向け(ボード等含む)と PC 向け、②ソリューション利用者向けとして、CTI ソリューション、Web API サービスソリューション、その他ソリューション、そして③エンドユーザ向けとして、応用製品とソフトウェア製品の、計 7 つに分類している。近年、Web 上でのサービス提供が主流となっていることから、従来の Web サービスソリューションという分類は有用性が低くなっていると考え、Web API サービスに特化した分類に変更した。Web API に該当しない製品は、各ベンダがそれぞれの判断で他の適切な分類に移行した。また記載項目についても、分類に応じて、利用シーン、処理単位、通信プロトコルなどを追加した。

各分類における平成 30 年度～令和 3 年度の製品件数とその推移を表 6 に示す。また、製品を提供しているベンダ数とその推移について表 7 に示す。今回、一部 分類の見直しを行ったため、昨年度までの Web サービスと本年度の Web API サービスの製品数を比較することは控えた。

表 6 分類毎の音声認識製品の件数と推移

分類		表番号*1	R3		R2	R1	H30
			件数	増減*2	件数	件数	件数
① システム開発者向け 音声認識エンジン	組込み向け(ボード等含む)	表 B-1	9	0	9	11	11
	PC 向け	表 B-2	4	0	4	6	6
② ソリューション利用者 向け製品	IVR/CTI/CRM 分野	表 B-3	14	0	14	15	13
	Web サービス(~R2)	表 B-4			13	14	10
	Web API サービス(R3)		7				
	その他	表 B-5	8	0	8	7	9
③ エンドユーザ向け 応用製品	応用製品	表 B-6	9	-6	15	17	13
	応用ソフトウェア製品	表 B-7	33	+11	22	19	14
合計		-	84		85	89	76

*1 表番号：認識製品一覧の対応表番号。 *2 増減：令和 2 年度からの増減。

表 7 分類毎の音声認識製品のベンダ数とその推移

分類		表番号*1	R3		R2	R1	H30
			件数	増減*2	件数	件数	件数
① システム開発者向け 音声認識エンジン	組込み向け(ボード等含む)	表 B-1	4	0	4	5	5
	PC 向け	表 B-2	4	0	4	5	5
② ソリューション利用者 向け製品	IVR/CTI/CRM 分野	表 B-3	10	0	10	11	10
	Web サービス(~R2)	表 B-4			9	8	6
	Web API サービス(R3)		7				
	その他	表 B-5	6	-1	7	6	8
③ エンドユーザ向け 応用製品	応用製品	表 B-6	7	-1	8	6	5
	応用ソフトウェア製品	表 B-7	7	+3	4	3	3
合計		-	45		46	44	42

*1 表番号：認識製品一覧の対応表番号。 *2 増減：令和 2 年度からの増減。

回答していただいた 25 社³ に関する調査結果の詳細を表 B-1～B-7 に掲載する。

調査結果は、上記方法による本委員会からの依頼に対して回答のあった情報のみを掲載しているため、網羅性の観点からは完全な調査ではない。よって、周知であるにも関わらず調査結果に記載されていない製品が存在する。

また、昨年度は回答があったが今年度は回答が得られなかったベンダについて、製品自体が継続して販売されていることが企業ホームページから確認できる場合でも、該当する製品情報を掲載していない。

したがって、本調査結果のみから、詳細な動向を判断することは困難であるが、表 6 と表 7 からわかるここ数年の傾向を以下に示す。

- ・ 昨年度に対し、ベンダ数、製品数とも 1 ずつ減少と、大きな変動はなかった。
- ・ システム開発者向けの音声認識エンジンについては、組込み向け（ボード等含む）製品、PC 向け製品とも変動がなかった。
- ・ また、ソリューション利用者向け製品についても、既存の分類である IVR/CTI/CRM 向けソリューションとその他製品は変動がなかった。
- ・ エンドユーザ向け製品については、全体で 5 製品増加した。個別にみると、昨年度、減少へと転じた応用製品はさらに 6 製品減少した。ただ、特定の用途の製品が大きく減少するような傾向はみられなかった。なお、減少した 6 製品のうち 3 製品については、本年度より担当者の変更があったベンダから、コンシューマ向け製品を除外する旨のコメントがあった。この減少は、製品自体の変化によるものではなく、本調査への回答方針の精査によるものといえる。一方で、応用ソフトウェア製品は 11 製品増加し、3 年連続の増加となった。新規製品としては、議事録作成に関するものが最も多く、その他に、オンライン会議、オンライン商談、物流センターピッキングなどに関する製品があった。また増加した中には、昨年度の Web サービスの分類から移動されたものが 4 製品あり、分類の変更の影響もあると考えられる。

³ 回答の集計単位を本年度より変更した。昨年度までは、調査依頼先の窓口の数を集計していたが、本年度より、各製品のメーカーの数を集計することとした。たとえば、昨年度・本年度とも、NEC 社に調査協力を依頼し、結果 NEC 社と NEC ソリューションイノベーション社の製品を回答いただいたが、これを昨年度は 1 社、本年度は 2 社と数えている。

○令和3年(2021年)度の音声認識・合成技術による応用製品の全般的な動向

(音声合成)

製品動向調査の結果だけからは読み取りにくい、2021年の音声合成製品の1つの動向として、無料や安価での利用が可能な音声合成のサービスの発表が相次いだことが挙げられる。

2021年4月には、大学発のベンチャーであるCoeFont社⁴から商用利用可能なサービスである「CoeFont STUDIO」が無償公開⁵されて話題となり、7月にはカスタムボイス作成サービスを含むプラットフォーム「CoeFont CLOUD」がリリースされた⁶。

2021年8月には、商用利用も可能なオープンソースのAI音声合成システム「VOICEVOX」がリリースされ⁷、2021年12月には、商用利用可能な無料のAIトークソフト「COEIROINK(コエイロインク)」も公開された⁸。同じく8月には、東芝デジタルソリューションズからクリエイター向けの音声合成ツール「Voice Track Maker」も提供された⁹。

このような無料もしくは安価なサービスが広く提供されることによって、音声合成を使いたい個人利用者による利用のハードルが下がり、音声合成利用者の裾野が大きく広がることが期待される。

その他の動向では、一般の人々が音声合成による自動音声を書く機会が増えたことが挙げられる。NHKでは、2020年秋に「第2世代音声合成システム」¹⁰が発表され、一部の放送での使用が開始された。2021年には、朝の定時ニュース（おはよう日本）でも、一部のニュースが「AI自動音声」で読み上げられ始めている模様だ。また、2021年7月から開催された東京オリンピック・パラリンピックでも、NHKによるバスケットボールやバレーボール、卓球、水泳などの一部の試合のオンラインライブ配信では、音声合成を用いた「ロボット実況」が話題になった¹¹。

⁴ サービス公開時は Yellston 社。2021年12月に CoeFont に社名変更。

<https://www.yellston.com/company/>

⁵ CoeFont STUDIO. すべてのクリエイターに声を届ける Web 音声合成サービス。

<https://prtmes.jp/main/html/rd/p/000000001.000078329.html>

⁶ CoeFont CLOUD. 500円・15分の収録で AI 音声合成化。

<https://prtmes.jp/main/html/rd/p/000000005.000078329.html>

⁷ VOICEVOX.

<https://pc.watch.impress.co.jp/docs/news/1341708.html>

⁸ COEIROINK(コエイロインク).

<https://gigazine.net/news/20211229-coeiroink/>

⁹ Voice Track Maker.

<https://www.global.toshiba.jp/company/digitalsolution/news/2021/0817.html>

¹⁰ 新開発「第2世代音声合成システム」“初”の放送利用

<https://www.nhk.or.jp/info/pr/toptalk/assets/pdf/soukyoku/2020/10/002.pdf>

¹¹ 聞きやすい！NHKの五輪中継“ロボット実況”好評 合成音声が入りタイ実況 落ち着く声はどう作られた？

<https://news.yahoo.co.jp/byline/okadayuka/20210805-00251641>

(音声認識)

昨年度に続き、新型コロナウイルス感染症（COVID-19）が猛威をふるう中で、人々のライフスタイルや業務の様々な変化に対して、音声認識技術の利用がさらに進んだ一年と言える。

特に、非対面のコミュニケーション手段として定着がみられるオンライン会議においては、リアルタイムなコミュニケーション支援や議事録作成支援などに音声認識が重宝されている。製品数が大きく増加し、オンライン商談向けのように特定の会議形態に特化した製品も登場している。

また、新型コロナウイルス感染症に関する問い合わせへの対応や自宅療養者への健康確認の業務により業務負荷が高まるコールセンターでは、対応の自動化やオペレータ業務の効率化のために音声認識が貢献している。

その他、オンラインでの消費者の購買行動が増えたことで需要が増す製造・物流業界向けに、製品組立・製品検査の作業や、物流センターのピッキング作業などを支援する製品も増加した。