

JEITA 一般社団法人 電子情報技術産業協会

令和 2(2020)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 —

2021 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 2 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 2（2020）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（[onseir-mail@aist.jp](mailto:onseir-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 2（2020）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。

### 製品カテゴリの新しい分類

音声合成製品の動向調査では、前々回（2018 年度調査）まで音声合成製品を 9 つの製品カテゴリに分類して経年の変化を追ってきた。しかし、昨今の技術の進展と動向に合わせて、前回の調査（2019 年度調査）から製品分類を見直し、7 カテゴリとしている。

今回の調査（2020 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られ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。

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	R2 (2020) 製品件数	R2 (2020) 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	10	7
表 A-5	ツール製品	17	10
表 A-6	ハードウェア製品	3	2
表 A-7	その他	9	6
合計		91	54

今年度の回答製品数は総計 91 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

- 「組込み用ミドルウェア・SDK」製品 (表 A-1, 17 件)
- 「ツール製品」(表 A-5, 17 件)

の製品件数が多かった。

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

もともに製品件数は 10 件であった。

また、今年度の回答ベンダ数は総計 54 社であった。ここで「提供ベンダ数」とは「のベンダ数」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、

- 「汎用 OS 用ミドルウェア・SDK」製品 (表 A-2, 12 社)

が 7 つの製品カテゴリの中では、回答ベンダ数が最多であった。

以下は

- 「ツール製品」(表 A-5, 10 社)
- 「組込み用ミドルウェア・SDK」製品 (表 A-1, 9 社)
- 「サーバ用ミドルウェア」製品 (表 A-3, 8 社)
- 「web サービス」製品 (表 A-4, 7 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。

### 【参考】従来の分類カテゴリとの対応

上述の通り、前年度(2019 年)の調査から分類のカテゴリを見直し、各カテゴリの定義も明確にした。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、現在の分類カテゴリに対応すると考えられる、旧分類カテゴリを表 3 に示す。

表 3 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	組み込み用 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

### 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、2019 年度調査までの製品数（表 4）とベンダ数（表 5）を示す。今年度の回答製品数は総計 91 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、10 件ほど減っている（表 4）。

今年度の回答ベンダ数は総計 54 社であった。製品の提供ベンダ数の総数（のべ数）を前年度の調査と比較すると、2 社ほど増えている（表 5）。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、前年度調査から変わっていない。これは、「新たに調査に登録した新規のベンダ」、「分社により新たな社名で登録したベンダ」のように異なりベンダ数の増加となった社数と、「製品提供を中止したり、製品調査への回答を取りやめたベンダ」の社数を相殺した結果、異なりベンダ数で見ると変化していない結果となった。

表 4 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	R2	R1	旧 ID	旧分類 (第 2 分類)	H30	H29	H28
		(2020)	(2019)			(2018)	(2017)	(2016)
		件数(±)	件数			件数	件数	件数
A-1	組込み用 MW SDK	17 (-6)	23	A-1	MW 組込み用 音声合成	30	33	32
A-2	汎用 OS 用 MW SDK	25 (-5)	30	A-3	MW PC 用 音声合成	26	24	27
A-3	サーバ用 MW	10 (±0)	10	A-7	その他 サーバ製品	11	12	14
A-4	Web サービス	10 (±0)	10	A-9	その他 サービス	11	15	15
A-5	ツール製品	17 (-2)	19	A-8	その他 PC ソフト・ツール	21	23	22
A-6	ハードウェア製品	3 (+2)	1	A-5	ハードウェア 音声合成	1	1	1
A-7	その他	9 (+1)	8	A-2	MW 組込み用 その他	1	1	1
				A-4	MW PC 用 その他	0	2	2
				A-6	ハードウェア その他	2	2	3
合計		91 (-10)	101	合計		103	113	117

表 5 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	R2	R1	旧 ID	旧分類 (第 2 分類)	H30	H29	H28
		(2020)	(2019)			(2018)	(2017)	(2016)
		社数(±)	社数			社数	社数	社数
A-1	組込み用 MW・SDK	9 (-1)	10	A-1	MW 組込み用 音声合成	11	12	12
A-2	汎用 OS 用 MW・SDK	12 (-1)	13	A-3	MW PC 用 音声合成	13	13	14
A-3	サーバ用 MW	8 (±0)	8	A-7	その他 サーバ製品	9	10	10
A-4	Web サービス	7 (+1)	6	A-9	その他 サービス	6	9	9
A-5	ツール製品	10 (+1)	9	A-8	その他 PC ソフト・ツール	10	11	10
A-6	ハードウェア製品	2 (+1)	1	A-5	ハードウェア 音声合成	1	1	1
A-7	その他	6 (+1)	5	A-2	MW 組込み用 その他	1	1	1
				A-4	MW PC 用 その他	0	1	1
				A-6	ハードウェア その他	2	2	3
合計		54 (+2)	52	合計		53	60	61

## ○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、昨年同様の分類ならびに記載項目で調査を実施することとした。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 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。各分類における平成 29 年度～令和 2 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。

表 6 分類毎の音声認識製品の件数と推移

分類		表番号*	R2		R1	H30	H29
			件数	増減**	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	-2	11	11	13
	PC 向け	表 B-2	4	-2	6	6	11
② ソリューション利用者 向け製品	IVR/CTI/CRM 分野	表 B-3	14	-1	15	13	14
	Web サービス	表 B-4	13	-1	14	10	11
	その他	表 B-5	8	+1	7	9	7
③ エンドユーザ向け 応用製品	応用製品	表 B-6	15	-2	17	13	6
	応用ソフトウェア製品	表 B-7	22	+3	19	14	17
合計		-	85	-4	89	76	79

\*表番号：認識製品一覧の対応表番号、 \*\*増減：令和元年度からの増減

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*	R2		R1	H30	H29
			件数	増減**	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	-1	5	5	7
	PC 向け	表 B-2	4	-1	5	5	6
② ソリューション利用 者向け製品	IVR/CTI/CRM 分野	表 B-3	10	-1	11	10	10
	Web サービス	表 B-4	9	+1	8	6	8
	その他	表 B-5	7	+1	6	8	6
③ エンドユーザ向け 応用製品	応用製品	表 B-6	8	+2	6	5	3
	応用ソフトウェア製品	表 B-7	4	+1	3	3	5
合計		-	46	+2	44	42	45

\*表番号：認識製品一覧の対応表番号、 \*\*増減：令和元年度からの増減

回答していただいた 24 社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、網羅性の視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、ベンダ数は 2 社増加した。しかし、製品数は 4 製品減少し、製品の見直しが行われている。
- ・ 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 向け製品ともに 2 製品ずつ減少した。
- ・ また、ソリューション利用者向け製品については、全体で 1 製品減少した。内訳としては、その他製品が 1 製品増加したが、IVR/CTI/CRM 向けソリューションと Web サービスが 1 製品ずつ減少した。
- ・ 一方で、エンドユーザ向け製品については、全体で 1 製品増加した。応用製品が 2 製品減少したが、応用ソフトウェア製品が 3 製品増加した。
- ・ 大きく製品数が増加した昨年度と比較して、今年度は製品数が減少へと転じた。例えば、認識エンジン製品の製品数は近年の減少傾向を引き継ぎ今年度も減少した。一方、応用ソフトウェア製品では、製品数は昨年度からの増加傾向を引き継ぎ今年度も増加している。様々な応用分野に向けて認識精度面とユーザーインタフェースの両面でカスタマイズされた音声認識製品が拡充されており、それによって応用ソフトウェア製品の製品数が伸びていると考えられる。

## ○令和 2 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)

令和 2 年(2020 年)の製品動向調査の結果を見るだけでは読み取りにくい、音声合成技術の特徴的な動向を振り返ると、これまでも増して「個人」が主役となった 1 年と言える。

2018 年から、個人の声や音声合成製品のマーケットで流通させる「コエステ」アプリがリリースされて来たが、2020 年 2 月には、新たに「コエステ社」が設立され<sup>1</sup>、コエステーション事業が東芝デジタルソリューションズ社からコエステ社に承継<sup>2</sup>された。

さらに 2020 年には、開発企業が「音声の個人性（声色）」を対象として取り扱うだけでなく、ユーザやクリエイターなどの「個人」が出資するクラウドファンディングにより、音声合成技術が進化した。たとえば、2020 年 7 月には「東北イタコ」の音声合成データベース制作プロジェクトがクラウドファンディングにより成立した<sup>3</sup>。このプロジェクトでは歌唱データベースを開発するのに加えて、パブリックドメインとして利用可能<sup>4</sup>な「ITA コーパス<sup>5</sup>」の作成も行われ、今後のテキスト音声合成の普及への貢献が期待される。また、同じく 2020 年 7 月には「小春六花」の制作資金のクラウドファンディングも成立し<sup>6</sup>、テキスト音声合成ソフト VOICEROID、及び、歌唱ソフト Synthesizer V に向けた開発が実現した。

同時に AI 対応の歌唱ソフトの世界では、個人ユーザ・個人クリエイターにも利用可能な AI 対応の歌唱ソフトのリリースが相次いだ<sup>7</sup>。

このような「個人化」の潮流は強まりこそすれ、今後も弱まることはないと考えられ、従来型の音声合成技術の開発、ビジネス展開と共に、注目される。

1 コエステ株式会社。 <https://coestation.jp/>

2 コエステーション事業のコエステ株式会社への譲渡のお知らせ。  
<https://coestation.jp/news/detail.php?id=1082015>

3 東北イタコ音声合成データベース制作プロジェクト。  
<https://greenfunding.jp/pub/projects/3891>

4 高い、使いにくい、読みにくいー音声合成研究者を悩ませるハードルを解決する”台本”，明治大学らが発表。2021 年 6 月 18 日。ITmedia NEWS。  
<https://www.itmedia.co.jp/news/articles/2106/18/news099.html>

5 ITA コーパス。  
<https://github.com/mmorise/ita-corpus>

6 小春六花 新ボイスロイド制作プロジェクト。  
<https://greenfunding.jp/pub/projects/3660>  
<https://prtimes.jp/main/html/rd/p/000000005.000058500.html>

7 「今の歌声合成ソフト市場は 30 年前のシンセ市場のよう」ーAHS が見る業界の現在地。2020 年 11 月 2 日。ITmedia NEWS。  
<https://www.itmedia.co.jp/news/articles/2011/02/news059.html>

### (音声認識)

音声認識製品に関する技術的な動向として、言語モデルや対話モデルのオーサリングツールがクラウドサービスで提供されるようになり、IVR やチャットボットなどを顧客が使用環境に合わせて独自に開発することが可能となっており、今後さらに幅広い分野での音声認識の活用が広がっていくと考えられる。

また、今年度は特に、新型コロナウイルス感染症（COVID-19）が猛威をふるう中、それに伴う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進んだのが大きな特徴と言える。特にコールセンターでは、新型コロナ感染症に関する問い合わせへの応対や自宅療養者などへの健康確認の業務が急増したことから、サービス拡充やオペレータ業務の効率化のために、自動受付や自動問い合わせなどの実現に利用されている。加えて、窓口においても感染防止のため非接触応対の実現に利用されている。さらに非対面のコミュニケーション手段として利用が増加しているオンライン会議では、発言の確認のために、議事録作成の支援などに利用されている。また、医療分野では、医療従事者の業務の効率化や感染リスク抑制のために、音声入力による効率的かつ非接触な、電子カルテなどの医療記録の作成や自動問診の実現に利用されている。