

JEITA 一般社団法人 電子情報技術産業協会

令和元(2019)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 —

2020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（令和元年度）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

令和元（2019）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.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，令和元（2019）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、これまで経年の変化を追うことも目的としていたため、調査開始当初に設定した製品分類のカテゴリ（第1分類4カテゴリ、第2分類で全9カテゴリ¹⁾）を採用してきた。しかし、今回の調査（2019年度調査）から製品分類を見直し、7カテゴリとした。

今回、製品分類カテゴリを見直した理由のひとつは、昨今の技術の進展に伴い、旧分類が最近の技術動向に見合わなくなってきたため、現在の技術動向に合わせた製品分類へと見直した。とりわけ、サーバ環境やクラウドサービスでの製品動向が反映されるように見直した。また、ともするとこれまでの旧分類が提供側の視点による分類であったため、新分類では音声合成製品のユーザ（本調査の読者）に分かりやすくなることを重視して、製品分類を見直した。

結果として、新しい製品カテゴリは、利用用途や利用場面から追うことができる分類体系に改編されたと言える。新しい製品分類のカテゴリを表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。

¹⁾ 2018 年度調査以前の製品カテゴリの分類については、表 4 を参照。

音声合成製品の製品件数

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，分類ごとの製品件数について述べる。

表 2 に，調査対象とした 7 カテゴリについて製品件数を示す。

表 2 分類毎の音声合成製品の件数

カテゴリ ID	新しい分類	R1 (2019) 製品件数
表 A-1	組込み用 ミドルウェア・SDK	23
表 A-2	汎用 OS 用 ミドルウェア・SDK	30
表 A-3	サーバ用 ミドルウェア	10
表 A-4	Web サービス	10
表 A-5	ツール製品	19
表 A-6	ハードウェア製品	1
表 A-7	その他	8
合計		101

今年度の回答製品数は総計 101 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，2 件ほど減っている。

製品分類ごとに製品件数で見ると，

- 「汎用 OS 用ミドルウェア・SDK」製品（表 A-2, 30 件）

は 7 つの製品カテゴリの中では，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, 23 件）

- 「ツール製品」（表 A-5, 19 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）

はともに製品件数では 10 件であった。

- 「ハードウェア」製品（表 A-6, 1 件）

は今年度の調査では 1 件の回答が寄せられるのみであった。

音声合成製品の提供ベンダ数

表 3 では、調査回答が得られた 7 カテゴリについて、製品件数ではなく、製品を提供しているベンダ数を示す。

表 3 分類毎の音声合成製品の提供ベンダ数

カテゴリID	新しい分類	R1 (2019) ベンダ数
表 A-1	組込み用 ミドルウェア・SDK	10
表 A-2	汎用 OS 用 ミドルウェア・SDK	13
表 A-3	サーバ用 ミドルウェア	8
表 A-4	Web サービス	6
表 A-5	ツール製品	9
表 A-6	ハードウェア製品	1
表 A-7	その他	5
合計		52

今年度の回答ベンダ数は総計 52 社であった。製品の提供ベンダ数の総数を前年度の調査と比較すると、1 社ほど減っている。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。

製品分類ごとにベンダ数でみると、

- 「汎用 OS 用ミドルウェア・SDK」製品（表 A-2, 13 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

- 「組込み用ミドルウェア・SDK」製品（表 A-1, 10 社）
- 「ツール製品」（表 A-5, 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, 6 社）

のように続いた。

- 「ハードウェア」製品（表 A-6, 1 社）

は 1 社のみ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、前年度調査から 1 社が減っている。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を相殺した結果、異なりベンダ数で 1 社が減っている。

【参考】従来の分類カテゴリとの対応

上述の通り、今年度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。前年度調査までの分類カテゴリとは必ずしも 1:1 に対応する訳ではないが、参考までに、新しい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 4 に示した。

表 4 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	表 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	表 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	表 A-7	その他	サーバ製品
A-4	Web サービス	表 A-9	その他	サービス
A-5	ツール製品	表 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	表 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	表 A-2	ミドルウェア・組み込み用	その他
		表 A-4	ミドルウェア・PC 用	その他
		表 A-6	ハードウェア	その他

併せて、前年度調査までの製品数（表 5）とベンダ数（表 6）を示す。今年度調査ではそれぞれの分類カテゴリに対して、定義を明確化したため、新しい分類の定義に従って、前年度調査とは異なるカテゴリに製品を移動させた例もあるため、製品数とベンダ数の増減は参考としての目安である。

表 5 分類毎の音声合成製品の件数と推移

新 ID	新しい分類	R1 (2019)	旧 ID	旧分類 (第 2 分類)	H30 (2018)	H29 (2017)	H28 (2016)	H27 (2015)
		件数(±)			件数	件数	件数	件数
A-1	組込み用 MW SDK	23 (-7)	A-1	MW 組込み用 音声合成	30	33	32	39
A-2	汎用 OS 用 MW SDK	30 (+4)	A-3	MW PC 用 音声合成	26	24	27	31
A-3	サーバ用 MW	10 (-1)	A-7	その他 サーバ製品	11	12	14	14
A-4	Web サービス	10 (-1)	A-9	その他 サービス	11	15	15	13
A-5	ツール製品	19 (-2)	A-8	その他 PC ソフト・ツール	21	23	22	22
A-6	ハードウェア製品	1 (±0)	A-5	ハードウェア 音声合成	1	1	1	1
A-7	その他	8 (+5)	A-2	MW 組込み用 その他	1	1	1	1
			A-4	MW PC 用 その他	0	2	2	2
			A-6	ハードウェア その他	2	2	3	3
合計		101 (-2)	合計		103	113	117	126

表 6 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	R1 (2019)	旧 ID	旧分類 (第 2 分類)	H30 (2018)	H29 (2017)	H28 (2016)	H27 (2015)
		社数(±)			社数	社数	社数	社数
A-1	組込み用 MW・SDK	10 (-1)	A-1	MW 組込み用 音声合成	11	12	12	11
A-2	汎用 OS 用 MW・SDK	13 (±0)	A-3	MW PC 用 音声合成	13	13	14	13
A-3	サーバ用 MW	8 (-1)	A-7	その他 サーバ製品	9	10	10	10
A-4	Web サービス	6 (±0)	A-9	その他 サービス	6	9	9	8
A-5	ツール製品	9 (-1)	A-8	その他 PC ソフト・ツール	10	11	10	10
A-6	ハードウェア製品	1 (±0)	A-5	ハードウェア 音声合成	1	1	1	1
A-7	その他	5 (+2)	A-2	MW 組込み用 その他	1	1	1	1
			A-4	MW PC 用 その他	0	1	1	1
			A-6	ハードウェア その他	2	2	3	3
合計		52 (-1)	合計		53	60	61	58

○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 28 年度～令和元年度の製品件数とその推移を表 7 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 8 に示す。

表 7 分類毎の音声認識製品の件数と推移

分類		表番号*	R1		H30	H29	H28
			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	11	0	11	13	13
	PC 向け	表 B-2	6	0	6	11	12
② ソリューション利用者 向け製品	IVR/CTI/CRM 分野	表 B-3	15	+2	13	14	13
	Web サービス	表 B-4	14	+4	10	11	9
	その他	表 B-5	7	-2	9	7	5
③ エンドユーザ向け 応用製品	応用製品	表 B-6	17	+4	13	6	8
	応用ソフトウェア製品	表 B-7	19	+5	14	17	17
合計		-	89	+13	76	79	77

*表番号：認識製品一覧の対応表番号， **増減：平成 30 年度からの増減

表 8 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*	R1		H30	H29	H28
			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	5	0	5	7	7
	PC 向け	表 B-2	5	0	5	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	11	+1	10	10	9
	Web サービス	表 B-4	8	+2	6	8	6
	その他	表 B-5	6	-2	8	6	4
③ エンドユーザ向け 応用製品	応用製品	表 B-6	6	+1	5	3	3
	応用ソフトウェア製品	表 B-7	3	0	3	5	3
合計		-	44	+2	42	45	38

*表番号：認識製品一覧の対応表番号， **増減：平成 30 年度からの増減

回答していただいた 22 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、上記手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機関について、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 3 と表 4 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 つ増加し、製品数は 13 個増加した。
- ・ 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 向け製品ともに増減がなかった。
- ・ 一方で、ソリューション利用者向け製品については、全体で 4 製品増加した。内訳としては、その他製品が 2 製品減少したが、IVR/CTI/CRM 向けソリューションが 2 製品増加し、Web サービスが 4 製品増加した。
- ・ また、エンドユーザ向け製品については、全体で 9 製品増加した。応用製品が 4 製品増加し、応用ソフトウェア製品が 5 製品増加した。
- ・ 昨年度と比較して、認識エンジン製品の製品数は変わらない一方、Web サービス製品、応用製品、応用ソフトウェア製品の製品数が顕著に増加している。音声認識の利用用途が広がるにつれて、様々なプラットフォームで利用可能な Web サービスの製品数が増えていると推測される。また、新しい用途に音声認識が用いられることで、応用製品や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。

昨年度までの調査と同様に、クラウド音声認識を活用する携帯端末上のアプリケーションや、サーバ上での音声認識利用サービスが継続して提供されていることを確認した。クラウドの活用により、ハードウェア・ソフトウェアの開発負荷を小さくすることができるため、今後もこの方向性は継続していくと考えられる。また、サービスの利便性を向上させるためには、多くの情報を短時間で処理することが必要である。したがって、音声の特徴を生かして、作業の効率化・利便性・機器の操作性を向上したサービスは、今後も提案されていくことが予想される。そして、音声認識の利用用途の拡大に伴い、それぞれの利用用途に合わせたカスタマイズが行われ、応用製品の製品数はさらに増加すると考えられる。

これらの傾向は既存の製品、サービスの範疇に収まらない、新たな分野で提案される可能性もあり、継続して広く製品・技術動向を調査していく必要があると考えられる。

○令和元年度の音声認識・合成技術による応用製品の全般的な動向

(音声合成)

令和元（2019）年度の製品動向調査の結果の集計だけからでは把握しにくい、音声合成技術の特徴的な動向として、2018年に引き続き「音声の個人性（カスタムボイスの音声合成技術）」に注目が集まったことが挙げられる。2018年の本報告書では、「声色」に関わる動向が注目されることを述べた。さらに2019年は単に「個人(の声色)」にとどまらず、「故人(の声色)」にまで注目が集まった²。象徴的なのは、ヤマハとNHKにより開発された「AI美空ひばり³」であろう。AI美空ひばりは2019年大晦日のNHK紅白歌合戦にも登場し、大きな議論を巻き起こした。音声(歌声)データを用いて再現性の高い合成音声(歌声合成)を行う技術自体に目新しさがある訳ではないが、具体的に実在した人物での再現であった点、さらにそれが故人であった点、そして、生成された作品(楽曲)が著作権のある“新作”であったことなど、社会がこの技術をどう受け止めればよいのか、様々な議論が生じた。

音声合成技術に関する、もうひとつの特徴的な動向は、感情を中心とした多彩な話し方への対応が、2019年に一層顕著になったことが挙げられる。既に2018年頃より、音声合成エンジンの各ベンダは感情音声への対応を進めていたが、2019年にはLINE社がLINE BRAINとして音声合成への対応も発表した⁴。LINE Clovaにも使われたLINE社の音声合成技術は、3Dキャラクターを伴う”キャラクター召喚装置”として発売された GateBoxにも採用され、2019年10月に発売された⁵。多彩な話し方への対応については、JEITA本委員会でも、「IT-4012 話し方種別ガイドライン」として、2018年7月に発行されている⁶。このガイドラインに関しては、2019年10月のCEATECでも講演セッションが開催され⁷、聴衆の関心を集めた。

(音声認識)

音声認識については、引き続き既存製品の見直しとともに、音声を利用した応用製品・サービスに開発の主軸が移っていることが調査からうかがえる。近年、音声合成や自動対話といった技術と音声認識を組み合わせ、機械との音声対話を実現し、電話の自動応答を行うソ

² 音声合成はAIの力で“棒読み”を脱した 飛躍の年、2019年を振り返る。ITメディア 2019年12月27日 <https://www.itmedia.co.jp/news/articles/1912/27/news069.html>

³ AI美空ひばり <https://www.itmedia.co.jp/news/articles/1909/03/news121.html>

⁴ LINE BRAIN. <https://www.linebrain.ai/> (実は音声合成自体はまだ coming soon)

⁵ Gatebox 量産モデル、発売延期から1年の成果 「音声合成は3回やり直し」、武地 CEOに聞く舞台裏。IT Media. 2019年10月11日。

<https://www.itmedia.co.jp/news/articles/1910/11/news031.html>

⁶ 感情や意図を表現するための話し方種別のガイドライン (IT-4012). https://www.jeita-speech.org/standard/standard_4012.html

⁷ 2019年10月 CEATEC講演。JEITA 話し方種別ガイドライン。

https://www.jeita.or.jp/japanese/assets/pdf/about/2019/ceatec_2019_guide.pdf

ソリューションが出現している。働き方改革の要請が高まる中、従業員の負荷を緩和しつつもサービス利用時間を拡大できるため、期待が高まっている。

また、2020年に入ってから猛威をふるっている新型コロナウイルス感染症（COVID-19）への感染防止策として、従来行われてきた様々な対面でのやり取りが、電話やWeb会議システムといった音声入出力デバイスを介したやり取りに急速に置き換わっている。この変化は一時的な変化に留まらず、ニューノーマルとして一定程度定着すると考えられ、今後はこれらの遠隔でのやり取りに音声認識を活用したソリューションが多数現れると予測される。