

JEITA 一般社団法人 電子情報技術産業協会

平成 29(2017)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 —

2018 年 7 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（平成 29 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

平成 29（2017）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](mailto: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，平成 29（2017）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。  
当製品調査は、調査対象の製品分類を大きく

- 「ミドルウェア組み用」、
- 「ミドルウェア PC 用」、
- 「ハードウェア製品」、
- 「その他」

の4カテゴリ（第1分類）に分けた上で、さらに細分類（第2分類）を設けて製品情報をまとめている。

調査結果を表 A-1～9 に示す。

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とその推移について述べる。

表 1 に、調査対象とした 9 カテゴリについて製品件数と推移を示す。

表 1 分類毎の音声合成製品の件数と推移

分類		表番号*	H29 (2017)	H28 (2016)	H27 (2015)	H26 (2014)	H25 (2013)
第1分類	第2分類		件数(増減**)	件数	件数	件数	件数
ミドルウェア 組み用	規則音声合成	表 A-1	33 (+1)	32	39	39	38
	その他	表 A-2	1 (±0)	1	1	1	1
ミドルウェア PC 用	規則音声合成	表 A-3	24 (-3)	27	31	29	29
	その他	表 A-4	2 (±0)	2	2	1	1
ハードウェア	規則音声合成 (規則音声合成 LSI)	表 A-5	1 (±0)	1	1	2	2
	その他	表 A-6	2 (-1)	3	3	2	1
その他	サーバ製品	表 A-7	12 (-2)	14	14	12	10
	PC ソフト製品	表 A-8	23 (+1)	22	22	22	24
	・ツール製品						
	サービス	表 A-9	15 (±0)	15	13	12	9
合計			113 (-4)	117	126	120	115

\*表番号：合成製品一覧の対応表番号， \*\*増減：前年度からの増減

今年度（平成 29 年度）の回答製品数は総計 113 件となり、平成 28 年度調査と比較して製品件数の総数で 4 件の減少となった。

以下、平成 28 年度と比較した今年度の製品件数の推移をまとめる。

第 2 分類で見ると、製品件数が増加した項目は

- 「ミドルウェア組み込み用」の「規則音声合成」製品（表 A-1、1 件増）

- 「その他」の「PC ソフト・ツール」製品（表 A-8, 1 件増）  
の 2 項目,

逆に, 製品件数が減少したのは,

- 「ミドルウェア PC 用」の「規則音声合成」製品（表 A-3, 3 件減）
- 「ハードウェア」の「その他」製品（表 A-6, 1 件減）
- 「その他」の「サーバ」製品（表 A-7, 2 件減）

の 3 項目で製品件数の合計が減少した。それ以外の 4 つの項目では製品件数の合計は同じであった。

但し, 製品件数はベンダ各社ごとに何を 1 製品として扱うかの基準が異なり, 上記の項目で製品件数が減少している製品カテゴリでも, 複数の製品をひとつの製品へと統合したケースも存在している。また, 製品カテゴリごとに集計すると各社の製品の増加と減少が相殺されるケースも存在するため, 製品件数の増減値は参考程度に捉えるのが妥当である。

表 2 では, 該当製品が存在した 9 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とその推移を示す。

表 2 分類毎の音声合成製品の提供ベンダ数とその推移

分類		表番号*	H29 (2017)	H28 (2016)	H27 (2015)	H26 (2014)	H25 (2013)
第 1 分類	第 2 分類		社数 (増減**)	社数	社数	社数	社数
ミドルウェア 組み込み用	規則音声合成	表 A-1	12 (±0)	12	11	11	11
	その他	表 A-2	1 (±0)	1	1	1	1
ミドルウェア PC 用	規則音声合成	表 A-3	13 (-1)	14	13	13	13
	その他	表 A-4	1 (±0)	1	1	1	1
ハードウェア	規則音声合成 (規則音声合成 LSI)	表 A-5	1 (±0)	1	1	2	2
	その他	表 A-6	2 (-1)	3	3	2	1
その他	サーバ製品	表 A-7	10 (±0)	10	10	9	8
	PC ソフト製品 ・ツール製品	表 A-8	11 (+1)	10	10	11	11
	サービス	表 A-9	9 (±0)	9	8	9	7
合計(延べベンダ数)			60 (-1)	61	58	59	55

\*表番号：合成製品一覧の対応表番号, \*\*増減：前年度からの増減

音声合成製品の提供ベンダ数を, 製品カテゴリ（第 2 分類）ごとの延べベンダ数で見ると,

製品を提供しているベンダ数が増加した製品カテゴリは、

- 「その他」の「PC ソフト・ツール」製品（表 A-8, 1 ベンダ純増）

の 1 項目であった。

逆に今年度の調査では、提供ベンダ数が減少した製品カテゴリは

- 「ミドルウェア PC」の「規則音声合成」製品（表 A-3, 1 ベンダ純減）
- 「ハードウェア」の「その他」製品（表 A-6, 1 ベンダ純減）

の 2 項目であった。

それ以外の 6 項目では製品カテゴリのベンダ総数での増減はなかった。

今回の音声合成の製品動向調査に製品情報を提供したベンダ数は、全カテゴリを合わせた延べ数で 60 社であり、前年度の調査からは 1 ベンダ減少した結果となった。複数のカテゴリに製品を提供しているベンダの重複を考慮すると、調査に回答したベンダ（異なり数）は 24 社で、前年度調査から 1 社が減った。

### ○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 26 年度～29 年度の製品件数とその推移を表 3 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 4 に示す。

表 3 分類毎の音声認識製品の件数と推移

分類		表番号*	H29		H28	H27	H26
			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	13	0	13	13	13
	PC 向け	表 B-2	11	-1	12	11	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	13	18	14
	Web サービス	表 B-4	11	+2	9	7	7
	その他	表 B-5	7	+2	5	4	6
③ エンドユーザ向け 応用製品	応用製品	表 B-6	6	-2	8	6	2
	応用ソフトウェア製品	表 B-7	17	0	17	15	14
合計		-	79	+2	77	74	67

\*表番号：認識製品一覧の対応表番号， \*\*増減：平成 28 年度からの増減

表 4 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*	H29		H28	H27	H26
			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	7	0	7	7	7
	PC 向け	表 B-2	6	0	6	6	7
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	9	10	10
	Web サービス	表 B-4	8	+2	6	5	6
	その他	表 B-5	6	+2	4	4	4
③ エンドユーザ向け 応用製品	応用製品	表 B-6	3	0	3	3	1
	応用ソフトウェア製品	表 B-7	5	+2	3	4	4
合計		-	45	+7	38	39	39

\*表番号：認識製品一覧の対応表番号， \*\*増減：平成 28 年度からの増減

回答して頂いた 19 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、該当製品情

報を掲載していない。

従って、本調査結果のみから、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。

- ・ 昨年度からベンダ数が増えたことで、1 社あたりの製品数は減ったが、製品数は増加した。
- ・ 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 向け製品ともに昨年度と大きな変化はなかった。
- ・ ソリューション利用者向け製品については、全体で 5 製品増加した。その中で、Web サービスとその他が共に 2 製品増加した。
- ・ エンドユーザ向け製品については、全体で 2 製品減少した。応用ソフトウェア製品は変わらなかったが、応用製品が 2 製品減少した。
- ・ ベンダごとの内訳をまとめると、19 製品がバージョンアップされた。また、新規に 10 製品が登録された一方で、昨年度に登録されていた 8 製品が今回は登録されなかった。既存製品の見直しが進むとともに、全体の製品数が増えたことから新製品開発が依然として活発であると考えられる。
- ・ ベンダ数については、昨年度から 4 社増加した。この結果、分類のカバー範囲を減らしたり、一つの分類の中で製品を統廃合により減らしたりするベンダが 3 社あったが、ベンダの延べ数は 7 社増加した。

その他、本調査に含まれない委員会の独自調査からも、スマートフォンやロボットや音声アシスタント端末からアクセスするサービス、また、2020 年の東京オリンピックに向けた音声翻訳としての実証実験や製品化応用サービスが引き続き増加しており、今後の動向が注目される。

なお、今後も、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増加することが予想され、製造企業、販売企業、サービス提供企業などが複雑化するとともに、販売製品という形態ではなく、有償・無償とにかかわらずエンドユーザに対する情報検索や音声翻訳や介護福祉支援などのサービスとして提供される応用システムの増加も考えられる。次年度以降も、なるべく多くの情報を収集し、当該分野の動向を把握するとともに、関連製品開発者、導入検討者、エンドユーザに有益な情報を提供できるよう、調査方法について検討を続ける予定である。

## ○平成 29 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)

平成 29 (2017) 年度の製品動向調査の結果の集計には含まれない、全般的な動向のひとつとして、深層学習 (ディープラーニング) 技術を実際の音声合成製品に搭載する事例が出始めたことが挙げられる。Google は、2016 年には研究段階だった WaveNet<sup>1</sup> を利用した音声合成を 2017 年 10 月からは AI 音声アシスタント Google Assistant に採用したと発表した<sup>2</sup>。また、NTT テクノクロス社も 12 月にバージョンアップした製品 FutureVoice Crayon で、深層学習を適用した製品の発売を開始した<sup>3</sup>。2018 年 1 月には HOYA も VoiceText の新コンセプトサイト (音声合成の声優事務所) にてディープラーニング版のデモを公開した<sup>4</sup>。これらの動向により、音声合成製品の品質はますます向上していくことが期待される。

また、もうひとつの動向として、2017 年は遂に日本国内市場でも、音声アシスタント端末 (スマートスピーカ・AI スピーカ) の発売が本格化した<sup>5</sup>。

音声アシスタント端末は製品の性格上、本体に GUI を備えておらず、音声 UI (音声認識 & 音声合成) がユーザインタフェースの主体となっているため、音声アシスタント端末の普及を通じて、家庭での音声合成技術の重要度も、さらに高まっていくと言える。

こうした音声アシスタント端末や、ロボット機器、2020 年の東京オリンピックも見据えた音声翻訳システムの普及に伴い、通常の読み上げスタイルの音声合成だけでなく、「使用している本人と同一の声で音声合成を行いたい」「多言語での音声合成が必要」「災害時に使うため、緊急な緊迫感のある音声合成を行いたい」「ロボットのキャラクターごとに声色

---

<sup>1</sup> WaveNet 深層学習を用いて音声波形を直接モデル化する音声合成の新方式

<sup>2</sup> Google WaveNet <https://deepmind.com/blog/wavenet-launches-google-assistant/>

<sup>3</sup> corevo® を活用した音声合成ソフトウェア「FutureVoice Crayon」に国内で初めて深層学習を適用し商用販売を開始。  
<https://www.ntt-tx.co.jp/whatsnew/2017/171205.html>

<sup>4</sup> ディープラーニング搭載の VoiceText が「音声合成の声優事務所」に登場！  
<http://voicetext.jp/news/product/180112/>

<sup>5</sup> “世界初「スマートスピーカー」のカオスマップを公開！” (ロボットスタート, 2017 年 11 月 8 日) <https://robotstart.info/2017/11/08/smart-speaker-map2017.html>  
“日本で購入できるスマートスピーカー 15 製品まとめ & 全世界の AI スピーカー一覧” (SmartHacks Magazine, 2017 年 11 月 15 日)  
<https://smarthacks.jp/mag/18832>



を変えたい」など、音声合成技術に対するニーズが多様化し、具体化してきていることが2017年の傾向と言える。

### (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が進むとともに、新製品開発が依然として活発であると考えられる。また、製品動向調査の範囲外の応用として、スマートフォンや専用端末からアクセスするクラウド型音声対話サービスの普及が進んでいる。例えば、このようなサービスを活用して家庭からの物品購入や情報検索などを容易にする音声アシスタント端末として、米国で普及している Amazon 社のスマートスピーカ Amazon Echo の国内販売が始まった。また、同様に、Google 社の Google Home、LINE 社の WAVE、ONKYO 社の G3 (Google Assistant 搭載) および P3 (Amazon Alexa 搭載)、ソニー社のスマートスピーカ LF-S50G、コミュニケーションロボット Xperia Hello!、NTT ドコモ社が開発した petoco などの国内販売が始まった。ここで、1 台の音声アシスタント端末を家族などの複数ユーザで共有する際のパーソナライズ手段として、話者認識技術が活用され始めている。例えば、「今日の予定は？」と話しかけた話者を認識し、その人の予定を示すといった使い方である。また、米国では、会議システムやスケジュールシステムなどとの連携を行う Alexa for Business といった、法人向けサービスも始まっている。

他には、オリンピックに向け、インバウンド対応を目的とした音声翻訳の製品化が活発になってきている。また、自動車関連では、安全性と便利さの実現のために、トヨタ社や NTT 社などが感情認識の研究を進めている。さらに、音声認識を支える背景技術として深層学習が普及した。