

JEITA 一般社団法人 電子情報技術産業協会

平成 30(2018)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 —

2019 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（平成 30 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 30（2018）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，平成 30（2018）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

- 「ミドルウェア組込み用」,
- 「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*	H30 (2018)	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	件数
ミドルウェア 組込み用	規則音声合成	表 A-1	30 (-3)	33	32	39	39	38
	その他	表 A-2	1 (±0)	1	1	1	1	1
ミドルウェア PC 用	規則音声合成	表 A-3	26 (+2)	24	27	31	29	29
	その他	表 A-4	0 (-2)	2	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	1	2	2
	その他	表 A-6	2 (±0)	2	3	3	2	1
その他	サーバ製品	表 A-7	11 (-1)	12	14	14	12	10
	PCソフト製品 ・ツール製品	表 A-8	21 (-2)	23	22	22	22	24
	サービス	表 A-9	11 (-4)	15	15	13	12	9
合計			103 (-10)	113	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 30 年度）の回答製品数は総計 103 件となり、平成 29 年度調査と比較して製品件数の総数で 10 件の減少となった。

以下、平成 29 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 PC 用」の「規則音声合成」製品（表 A-3、2 件増）

の 1 項目のみ、

逆に、製品件数が減少した項目は、

- 「ミドルウェア組込み用」の「規則音声合成」製品（表 A-1, 3 件減）
- 「ミドルウェア PC 用」の「その他」製品（表 A-4, 2 件減）
- 「その他」の「サーバ」製品（表 A-7, 1 件減）
- 「その他」の「PC ソフト・ツール」製品（表 A-8, 2 件減）
- 「その他」の「サービス」製品（表 A-9, 4 件減）

の 5 項目で製品件数の合計が減少した。それ以外の 3 つの項目では製品件数の合計は同じであった。

但し、製品件数は当該のカテゴリでの提供を取りやめたために製品数が減じているカテゴリもある一方で、複数の製品を新製品へと統合したため製品数が相殺されて減じている分類カテゴリも存在している。また、製品カテゴリごとに集計した結果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*	H30 (2018)	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	社数
ミドルウェア 組込み用	規則音声合成	表 A-1	11 (-1)	12	12	11	11	11
	その他	表 A-2	1 (±0)	1	1	1	1	1
ミドルウェア PC 用	規則音声合成	表 A-3	13 (±0)	13	14	13	13	13
	その他	表 A-4	0 (-1)	1	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	1	2	2
	その他	表 A-6	2 (±0)	2	3	3	2	1
その他	サーバ製品	表 A-7	9 (-1)	10	10	10	9	8
	PC ソフト製品 ・ツール製品	表 A-8	10 (-1)	11	10	10	11	11
	サービス	表 A-9	6 (-3)	9	9	8	9	7
合計(延べベンダ数)			54 (-6)	60	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、今回の調査では存在しなかった。

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

- 「ミドルウェア組込み用」の「規則音声合成」製品（表 A-1, 1 ベンダ減）
- 「ミドルウェア PC 用」の「その他」製品（表 A-4, 1 ベンダ純減）
- 「その他」の「サーバ」製品（表 A-7, 1 ベンダ減）
- 「その他」の「PC ソフト・ツール」製品（表 A-8, 1 ベンダ減）
- 「その他」の「サービス」製品（表 A-9, 3 ベンダ減）

の 5 項目であった。

それ以外の 4 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 53 社であり、前年度の調査からは 7 ベンダ減少した結果となった。但し、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 25 社で、前年度調査から 1 社が増えている。

○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、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 27 年度～30 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*	H30		H29	H28	H27
			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	-2	13	13	13
	PC 向け	表 B-2	6	-5	11	12	11
② ソリューション利用者 向け製品	IVR/CTI/CRM 分野	表 B-3	13	-1	14	13	18
	Web サービス	表 B-4	10	-1	11	9	7
	その他	表 B-5	9	+2	7	5	4
③ エンドユーザ向け 応用製品	応用製品	表 B-6	13	+7	6	8	6
	応用ソフトウェア製品	表 B-7	14	-3	17	17	15
合計		-	76	-3	79	77	74

*表番号：認識製品一覧の対応表番号， **増減：平成 29 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*	H30		H29	H28	H27
			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	-2	7	7	7
	PC 向け	表 B-2	5	-1	6	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	10	0	10	9	10
	Web サービス	表 B-4	6	-2	8	6	5
	その他	表 B-5	8	+2	6	4	4
③ エンドユーザ向け 応用製品	応用製品	表 B-6	5	+2	3	3	3
	応用ソフトウェア製品	表 B-7	3	-2	5	3	4
合計		-	42	-3	45	38	39

*表番号：認識製品一覧の対応表番号， **増減：平成 29 年度からの増減

回答していただいた 20 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、製品数ともに 3 つ減少した。
- ・ 特に、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）製品は 2 製品、PC 向け製品は 5 製品、ともに減少した。
- ・ 一方で、エンドユーザ向け製品については、全体で 4 製品増加した。応用ソフトウェア製品は 3 製品減少したが、応用製品が 7 製品増加した。
- ・ また、ソリューション利用者向け製品については、全体の製品数は変化がなかった。内訳としては、IVR/CTI/CRM 向けソリューションと Web サービスがともに 1 製品減少したが、その他製品が 2 製品増加した。
- ・ 認識エンジン製品が減少し、応用製品、Web サービス製品の製品数が増加している。また、PC 向け認識エンジンおよび応用ソフトウェア製品が他分野製品に比較して大きく減少している。これらのことから、ユーザが、音声操作 HMI を構築するための製品よりも、音声の特徴を生かして作業・操作の効率化を図るために、あらかじめ音声と製品の機能が、ハードウェア・ソフトウェア込みでパッケージングされたサービス・製品が主流となってきたことが推測される。

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。

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。

○平成 30 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 30 (2018) 年度の製品動向調査の結果の集計には含まれない、全般的な音声合成製品の技術と市場の動向のひとつとして、個別の「声色」に着目した活動が挙げられる。すわなち、単に「合成音声」を使うというだけにとどまらず、「誰の音を使うか」に着目した動向が 2018 年のトレンドであった。

たとえば、東芝デジタルソリューションズ社は「声」を収集・蓄積・提供するプラットフォーム「コエステーション™¹」サービスを 2018 年 4 月に正式なアプリとしてリリースした（その後 個人向けアプリに加えて、11 月には法人向けサービスも開始）。このサービスでは一般ユーザが自ら発声収録することで自分の”コエ”による音声合成をアプリ上にて無償で作成し、他者も利用することができる。また著名人などの声の音声合成をあらかじめ作成しておき、利用したい顧客企業を仲介する事業形態も提供している。一般ユーザ、著名人を含め、「声」を流通させるプラットフォームという活動は、個別の「声色」の利用を促進する活動として注目される。

声を流通させるには、声の持ち主と作成された声の利用者の権利化が課題となってくる。この課題に対して、2018 年 6 月には、一般社団法人 デジタルボイスパレット²が設立された。デジタルボイスパレットは、音声合成技術の普及、利用者と権利者のためのルール作り、利活用の推進を目的とした団体で、「発声権」という権利概念の普及を目指すとしている。これによれば「発声権」という概念により、収録した音声の権利（知的所有権）は発話者に帰属する。近年急速に発展している機械学習と人工知能の分野では、データとデータから作成されたモデルの所有権と利用許諾の範囲は、まさに議論がなされている分野であり、今後も動向が注目される。

前年（2017 年）は深層学習技術を実際の音声合成製品に搭載する事例が出始めた年であった³が、2018 年は同様の動向が歌声合成技術で開花した年であった⁴。また、音声合成の研究分野では、声質変換技術に深層学習を適用するなどの技術的な進展があり⁵、音声合成技術と声質変換技術を組み合わせることで、作成される合成音声の多様性を拡大することに寄与している。たとえば、言語を超えた声質変換（収録した発声者のネイティブ言語以外

¹ 東芝コエステーション <https://coestation.jp/>

² 一般社団法人 デジタルボイスパレット <http://www.digitalvoice.or.jp/>

³ JEITA 「2017 年度音声認識・合成技術 応用製品の動向について」 https://www.jeita-speech.org/survey/JEITA_productsurvey2017.pdf

⁴ 名古屋工業大学「超高品質な歌声を再現する AI 歌声合成システム～名工大と大学発ベンチャーが共同開発に成功～」. 2018 年 12 月。
<https://www.nitech.ac.jp/news/press/2018/7162.html>

⁵ J. Lorenzo-Trueba, J. Yamagishi, T. Toda, D. Saito, F. Villavicencio, T. Kinnunen, Z. Lin, "The Voice Conversion Challenge 2018: Promoting Development of Parallel and Nonparallel Methods", arXiv:1804.04262v1, 2018.

の音声合成を可能にする) の試みも行われている。

(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、活発に製品投入されているスマートスピーカをはじめ、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も継続して進んでいる。これらは、既存の家電（照明、エアコン、調理器具など）と連携して動作・操作することで、高機能化した複数の家電製品を統合操作・制御し、より快適・高度なサービスの実現を図るものである。特に、車載機器に関しては、機器の高機能化が進むなかで、手軽かつ安全な操作の要求が非常に高まっている。したがって従来のコマンド操作の置き換えではなく、自然な発話の中からユーザの意図を推定し、機器操作を行う HMI の実用化が進んでいる。例えば、BMW, Daimler, Audi などにおいては、そのようなシステムの製品化がされており、この傾向は自動車全体に進むのではないかとと思われる。

他には、昨年度までと同様、オリンピック向け、インバウンド対応を目的とした音声翻訳の製品導入が引き続き活発である。

(音声認識・音声合成 共通)

近年、音声認識・音声合成の技術的發展、市場への普及を物語る事例として、2018年には、特許庁による「特許出願技術動向調査⁶」の12のテーマのうちのひとつとして「マンマシンインターフェースとしての音声入出力⁷」が公開されている。この調査報告では、技術分野の俯瞰から始めて、用途、要素技術、課題などの観点から、出願人・出願国・出願内容(分野)・出願件数を詳細に分析し、日本の強みや弱みを分析しており、参考になる。

⁶ 特許庁。「特許出願技術動向調査」 <https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/index.html>

⁷ 平成 29 年度「電気・電子」。「マンマシンインターフェースとしての音声入出力」（発行平成 30(2018)年 2 月）
https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/document/index/29_11.pdf