解 説

JEITA 話し方種別ガイドライン
――感情や意図を表現する音声合成――

The Guidelines for Text-to-speech Speaking Style Classification

平沢純一　中嶌信弥

Abstract

人間と機械が音声を通じてコミュニケーションできるロボットや AI の普及により，音声合成技術には，感情や意図も表現できる多彩な対話調の話し方が求められてきている．一般社団法人電子情報技術産業協会（JEITA）音声入出力方式標準化専門委員会では，技術的な難易度を踏まえながら，合成音声の話し方に対する，一般の利用者からの要件を明確に定義しやすくすることを目的として，「話し方種別のガイドライン」を策定した．これにより非専門家（利用側）と専門家（提供側）が協同して要件を検討できるツールとして活用できる．

キーワード：音声合成，話し方，感情，ガイドライン

1. は　じ　め　に

1.1　音声合成技術の進展

音声合成（Text-to-speech．以下 TTS）技術の進展が目覚ましい．かつて“ロボットボイス”と揶揄された音声合成技術も，聞き取りやすさに困難を感じることが少なくなり，声色の個人性も再現できるようになってきている．その結果，音声合成技術に求められる要件は，単に聞き取ることができればよいという「読み上げ調」では不十分となっている．すなわち，人間同士の日常会話で見られるような，感情や意図を表現できる，多彩な「対話調」の話し方を実現することが急務となっている．

感情や意図を表現できる対話調の話し方は，誰もが日常的になじみのある対象であるために，技術的に実現することも難しくないと思えてしまうかもしれない．しかしながら，音声合成技術を開発する各企業は，多彩な対話調の音声合成の実現に困難を感じてきた．

1.2　多彩な話し方の音声合成を開発することの難しさ

開発を難しくしてきたと考えられる理由の一つが，そもそも「感情や意図を表現する“話し方”」とはどんなものであるか？が自明でなかったことにある．例えば，感情とはどう分類されて，何種類あるのかという問い[(1)~(3)]に対してもいまだ明確な記述や定義は合意されていない．また現在，W3C による SSML（Speech Synthesis Markup Language[(4)]）を用いて，音声合成の読み上げテキストをマークアップすることが可能だが，話し方について制御できるのは，音量，ピッチ，速度などの物理レベルの制御にとどまり，話し方を直感的に指定できるものではない．

このような現状の中で，TTS 利用者は感情研究を専門としているわけでもなく，日常生活の中でなじみのある“喜怒哀楽”のような大まかな把握にとどまりがちである．そのため，具体的にどんな感情の音声合成が欲しいのか，何種類あれば足りるのか，など必要な音声の特徴を改めて TTS 開発者に“技術的な要件”として伝えることが難しかった．一方で，技術の提供側（開発者）にとっても，感情や意図を表現する対話調の話し方は対象として余りに広大で，利用者に対して技術の全体像や難易度を示すことができずにきた．

音声合成に多彩な話し方をさせようとするとき，おなじみの“喜怒哀楽”レベルでは粗過ぎる一方で，具体的な感情を改めて体系的に網羅しようとしたら際限がない．つまり，日常的な対象であるがゆえに，利用者と提供者の間で，ニーズを具体化し，技術的な要件として擦り合わせていくことに難しさがあり，技術開発のゴールを設定しにくい状況にあった．

平沢純一　一般社団法人固有名詞協会
E-mail　hirasawa@jeita-speech.org
中嶌信弥　国士舘大学理工学部電子情報学系
E-mail　nakajima@kokushikan.ac.jp
Jun-ichi HIRASAWA, Nonmember (Association for Named Entity, Tokyo, 101-0054 Japan) and Shinya NAKAJIMA, Nonmember (Faculty of Engineering, Kokushikan University, Yokohama-shi, 227-0047 Japan).
電子情報通信学会誌　Vol.104 No.1 pp.55-59 2021 年 1 月

2. 話し方種別ガイドラインの策定

2.1 利用者と開発者が共有できるコミュニケーションツール

対話調の音声合成技術を開発するにあたり，利用者と開発者が共通のゴールとしての要件を設定することが難しい状況に対して，一般社団法人電子情報技術産業協会（JEITA）音声入出力方式標準化専門委員会は，両者が共通に議論できるコミュニケーションツールとして活用できるように「話し方種別ガイドライン（IT-4012）(注1)」を2018年7月に策定した[5]．このガイドラインでは，対話調音声の性質と役割を整理し，感情だけに限定せず，意図や態度の表現も含めて，55種の話し方種別として提示した．併せてそれぞれの話し方種別の技術的な実現の難易度（実現フェーズ）と，対応するユースケースを示している．更に個々の話し方種別に対して2〜3文の例文を挙げている．

2.2 ガイドラインが解決している課題

音声合成の利用者，すなわち，アプリケーションやサービスの提供者にとっては，従来「どのような感情や意図の話し方の合成音声が欲しいのか？」を技術要件として指定することが難しかった．“話し方”という対象が余りに多彩であるがゆえに，非専門家である利用者からの要望はとかく曖昧になりがちであることが課題であった．利用者は本ガイドラインを参照すれば，55種の【話し方種別】の中から必要とする話し方を選択することにより，色見本で指定するかのように開発者に明確な要求仕様を伝えることができる．もちろん55種だけであらゆる話し方のリクエストをカバーし切れるものではないが，希望する話し方が「ガイドライン内の55種類には含まれない」ことや，「近い話し方を挙げるとすれば」のように伝えることで，音声合成の技術者との意思疎通の一助となる．

一方，音声合成技術の開発者（提供者）の立場からも，“話し方”という対象が余りに多彩であるがゆえに，技術の全体像やロードマップを顧客に十分に示し切れないことが課題であった．しかし「感情とは何か」はいまだ容易に回答を期待できる段階にない[3]．そこで，本ガイドラインを活用することで，感情の何たるかまで自ら立ち返らずとも，利用者に技術開発の実情を把握してもらうことから始められる．利用者からの期待と技術の現実がすれ違うことのないよう，リクエストされる“話し方”が提供可能か，開発中であるか，提供予定がないかなど，自社の開発ロードマップを示すのに本ガイドラインを活用することにより，利用者との間で共通のツールとすることができる．

3. 話し方種別55種

3.1 【グループ】と【カテゴリ】

本ガイドラインは，出力される音声を通じて表現される，感情，口調，態度，意図，印象などの話し方の種別を一覧にして提示している．音声上の特徴が重要な違いになりそうな話し方について，音声合成の実際の利用場面で必要となりそうな話し方種別へのニーズと，技術的な実現の可能性をバランスさせて，55種の話し方種別としている．従来研究での分類[6],[7]を参考にしつつ，55種が漏れや重複を起こさないように【グループ】に分け，似ている話し方種別を把握しやすくするため一部を【カテゴリ】にまとめた（表1）．

55種の話し方種別は，まず〈平静（な話し方）〉〈感情（を表現する話し方）〉〈発話意図（を伝える話し方）〉〈その他〉の四つのグループに分類されている．〈感情グループ〉は38種と多様な話し方を含むため，更に〈ポジティブ（な感情の話し方）〉〈ネガティブ〉〈興奮・緊張〉〈沈静・弛緩〉の4軸でサブグループを設定している．

また，似た性格を持つ複数の話し方種別を【カテゴリ】としてまとめることで把握しやすくしている．例えば〈ポジティブ感情〉サブグループの《喜び》カテゴリでは，〔からかった感じ〕〔冗談っぽく〕〔感謝して〕などの，5種の異なる《喜び》の話し方が共通の一つのカテゴリにまとめられている．

【グループ】や【カテゴリ】はあくまで把握の便宜のために設定しているにすぎず，本ガイドラインは“人間の感情”を網羅的，体系的に分析分類したり，新たに学術的な定義を提案することを意図しているものではない．

3.2 話し方種別55種

以下に，本ガイドラインで与えている55種類の話し方種別を示す．表2（左列）には〈平静〉グループ，及び，《喜び》，《好き》，などの〈ポジティブ感情〉サブグ

表1 話し方種別のグループとカテゴリ

グループ	サブグループ	カテゴリ数	話し方数
〈平静〉		1カテゴリ	1種
〈感情〉	〈ポジティブ〉	2カテゴリ	6種
	〈ネガティブ〉	6カテゴリ	17種
	〈興奮・緊張〉	2カテゴリ	10種
	〈沈静・弛緩〉	1カテゴリ	5種
〈発話意図〉		6カテゴリ	7種
〈その他〉		6カテゴリ	9種
	計	24カテゴリ	55種

(注1) 正式なタイトルは「音声合成技術で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」.

表2　話し方種別（1）　〈平静グループ〉・〈感情グループ〉

カテゴリ	話し方種別名	ID
〈平静グループ〉（1カテゴリ・1種）		
《平静》（1種）	〔平穏な様子で〕	#1
〈感情グループ（ポジティブ）〉（2カテゴリ・6種）		
《喜び》（5種）	〔喜んでいる感じで〕	#2
	〔笑いながら〕	#3
	〔からかった感じで〕	#4
	〔冗談っぽく〕	#5
	〔感謝して〕	#6
《好き》（1種）	〔好意的な感じで〕	#7
〈感情グループ（ネガティブ）〉（6カテゴリ・17種）		
《嫌悪》（7種）	〔不満げに嫌がった様子で〕	#8
	〔おろおろと困惑した感じで〕	#9
	〔疲れた様子で〕	#10
	〔眠そうに〕	#11
	〔いやみっぽく〕	#12
	〔呆れた様子で〕	#13
	〔なげやりな感じで〕	#14
《怒り》（2種）	〔怒って腹を立てた様子で〕	#15
	〔敵意をもった感じで〕	#16
《恐れ》（2種）	〔怖れた様子で〕	#17
	〔悲鳴を上げるように〕	#18
《悲しみ》（4種）	〔悲しんで〕	#19
	〔沈痛に嘆く様子で〕	#20
	〔くよくよと落ち込んで〕	#21
	〔憐れんで同情して〕	#22
《苦笑い》（1種）	〔苦笑いしながら〕	#23
《後悔》（1種）	〔残念そうに後悔して〕	#24

カテゴリ	話し方種別名	ID
〈感情グループ（興奮・緊張）〉（2カテゴリ・10種）		
《昂り》（7種）	〔ドキドキし，昂った様子で〕	#25
	〔緊張した様子で〕	#26
	〔焦って気が急いだ様子で〕	#27
	〔元気にハツラツとして〕	#28
	〔自慢げに〕	#29
	〔高圧的な感じで〕	#30
	〔威嚇して脅迫するように〕	#31
《驚き》（3種）	〔思いもよらず驚いた様子で〕	#32
	〔緊急で注意喚起するように〕	#33
	〔うろたえた様子で〕	#34
〈感情グループ（沈静・弛緩）〉（1カテゴリ・5種）		
《安らぎ》（5種）	〔落ち着いて，やさしく〕	#35
	〔子供に話し掛けるように〕	#36
	〔お年寄りに話し掛けるように〕	#37
	〔気楽な感じで〕	#38
	〔のんびりとした様子で〕	#39

表3　話し方種別（2）　〈発話意図グループ〉・〈その他グループ〉

カテゴリ	話し方種別名	ID
〈発話意図グループ〉（6カテゴリ・7種）		
《質問》（1種）	〔相手に質問するように〕	#40
《要求》（2種）	〔要求するように〕	#41
	〔命令するように〕	#42
《希望》（1種）	〔希望するように〕	#43
《勧誘》（1種）	〔誘うように〕	#44
《意見》（1種）	〔主張するように．決意を語るように〕	#45
《謝罪》（1種）	〔謝罪して申し訳なさそうに〕	#46

カテゴリ	話し方種別名	ID
〈その他グループ〉（6カテゴリ・9種）		
《甘え》（2種）	〔相手に甘えた感じで〕	#47
	〔色っぽく〕	#48
《励まし》（1種）	〔励ますように〕	#49
《慰め》（1種）	〔慰める感じで〕	#50
《褒め》（1種）	〔相手を褒めるように〕	#51
《迷い》（2種）	〔迷いながら〕	#52
	〔疑っているように〕	#53
《恥》（2種）	〔恥ずかしそうに照れて〕	#54
	〔自信なさげに〕	#55

ループ，《嫌悪》，《怒り》をはじめとした6カテゴリから成る〈ネガティブ感情〉サブグループの話し方種別を示している．

表2（右列）には，《昂り》，《驚き》のような〈興奮・緊張〉を表現するサブグループの感情と，《安らぎ》のような〈沈静・弛緩〉感情のサブグループの話し方種別を示す．

本ガイドラインは「音声合成において必要となる話し方」を挙げているため，必ずしも“感情”とは分類しがたい話し方であっても，必要になると考えられる口調や

表 4　例文（抜粋）

グループ	カテゴリ	話し方種別（ID）	例文
〈感情〉（ネガティブ）	《苦笑い》	〔苦笑いしながら〕（#23）	・一人で買い物に行かせるのはまだ早かったかな． ・昨日までは上手くできていたのになぁ． ・こんな段差でつまずいちゃったよ，もう年だねえ．
〈感情〉（興奮・緊張）	《驚き》	〔緊急で注意喚起するように〕（#33）	・火災発生の恐れがあります．ただちに避難してください． ・近所で熊が出没したらしい．急いで家の中に入って！
〈発話意図〉	《勧誘》	〔誘うように〕（#44）	・一緒に映画を観に行きませんか？ ・同じサークルに入ろうよ．
〈その他〉	《慰め》	〔慰める感じで〕（#50）	・イベントに行けなくて本当に残念だったね． ・誰だって失敗することはあるよ．

態度であれば積極的に話し方種別に含めている．表 3 には，話し手の意図を伝える〈発話意図〉グループの話し方種別，及び，感情にも発話意図にも分類しがたい話し方として〈その他〉グループを示した．

3.3　実現フェーズとユースケース

一般に利用者が技術開発の難易度を目利きとして持つことは難しい．本ガイドラインは，利用者と技術開発者の間で要件を擦り合わせていく際のコミュニケーションツールとなることを目指しているので，参考として【実現フェーズ】の情報を示した．技術の難易度を定量的な指標として提示することは難しいため，「既に製品化されている」「一部のベンダでは開発を終えている」「研究開発を進めている」「まだ実現の見通しが立っていないに等しい」などの実現フェーズを用いて，4 段階で示した．これによりガイドラインの利用者は，自らの希望する話し方がどのような開発段階にあるのかを把握することができる．

また実現フェーズを捉えやすくする目的で，それぞれの実現フェーズに割り当てられた話し方を使うことで，具体的にどんなサービスを実現できるのか？をユースケースとして示している．

3.4　例文

それぞれの話し方がどのようなものか，感覚的に把握することの一助となるよう，個々の話し方種別に対して二〜三つの例文も示した．55 種類の話し方種別に対して，全 123 文の例文が提示されているが，ここでは一部を抜粋として表 4 に示す．例文が提示されていることにより，利用者の側から希望する話し方のイメージを伝える場合に，あるいは，各社の音声合成による音声サンプルを比較する場合など，検討を進めやすくなると考えられる．

また“話し方”である以上，なかなか文字だけではイメージしにくいため，JEITA の Web サイト[5]ではプロのナレータ録音による音声サンプルも参考として試聴できる．

4. 今 後 の 課 題

2020 年現在，TTS ベンダ各社は“感情音声”などへの対応バリエーション数を増やしつつある．本ガイドラインは発行された 2018 年での技術動向などを考慮して，55 種の話し方種別を提示した．今後，更なる技術の進展と，市場からのニーズの変化に合わせて，話し方種別も見直していく必要がある．以下では主な今後の課題を述べる．

（1）　音声の特徴記述と評価方法

大きな課題の一つが「音声の特徴記述」であることは間違いない．本ガイドラインは仕様要件を詰めていく際のコミュニケーションツールを目指したため，まずは話し方種別のセットを提示することを第一義とした．そのため，例えば〔いやみっぽい〕話し方（#12）の音声がどのような音声の特徴を持っているか？については規定していない．実際，学術的にも十分に解明できているとは言いがたく，大きな課題の一つである．

音声の特徴が十分に解明されていないことは，そのまま評価方法の課題ともなっている．得られた音声がどれくらい〔いやみっぽい〕話し方に合致しているのか，どのように評価を実施すればよいのかを定めていくことは今後の課題である．

（2）　個人性・キャラクタ音声・役割声

本ガイドラインの話し方種別は特定の個人によらず，一般的に一人の声が備えておきたい話し方のセットを示したものである．他方で“あの人らしい話し方”のような個人性やキャラクタの話し方をどう扱うか？も課題となってくる．また，音声合成技術の用途を考えると，“駅員さんらしい話し方”など職業や役割に由来する話し方への指標も検討の対象となろう．

（3）　話し方種別．更なる再考と精査

現状のガイドラインは，話し方の「強度」を考慮に入れられていない．例えば〔笑いながら〕（#3）という一

つの話し方種別を考えても，「軽く微笑みながら」話すのと「爆笑しながら」話すのは決して同じではない．話し方種別の強度をどう定めればよいのかも今後の検討課題である．また，本ガイドラインで示した55種の話し方種別同士がどのような関係にあるのか，すなわち，混在，共存，排他のような個々の話し方の組合せについても考察を深める必要があるだろう．例えば，〔怒って腹を立て〕（#15）ながらかつ〔命令するように〕（#42）話すことは十分想像できる一方で，〔なげやり〕（#14）なのに〔甘えた〕（#47）話し方などというものが存在するのかは自明ではない．

5. おわりに

JEITA音声入出力方式標準化専門委員会では，昨今ニーズの高まっている，感情や意図を表現する対話調の話し方についてガイドラインを策定した．本ガイドラインでは55種の話し方種別を提示し，併せて技術の難易度の参考となる実現フェーズや対応するユースケースも示している．これにより，非専門家であることが多い技術の利用者と開発を担う提供者の間で，共通の情報を参照することが可能となり，本ガイドラインは両者が協同して技術的な要件を検討できるツールとして活用できる．

文　献

（1） P. Ekman, “Basic emotions,” in Handbook of Cognition and Emotion, T. Dalgleish and M.J. Power, eds., pp. 45-60, John Wiley & Sons Ltd., New York, 1999.

（2） J.A. Russell, “A circumplex model of affect,” Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161-1178, 1980.

（3） プタシンスキ・ミハウ，“感情処理：感情を理解するコンピュータ，”心を交わす人工知能―言語・感情・倫理・ユーモア・常識―，荒木健治，ジェプカ・ラファウ，プタシンスキ・ミハウ，ディバワ・パヴェウ，第3章，森北出版，2016.

（4） SSML Speech Synthesis Markup Language, http://www.w3.org/TR/2010/REC-speech-synthesis11-20100907

（5） 一般社団法人電子情報技術産業協会規格 JEITA IT-4012，“音声合成技術で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，”July 2018, https://www.jeita-speech.org/standard/standard_4012.html

（6） Y. Arimoto, H. Kawatsu, S. Ohno, and H. Iida, “Naturalistic emotional speech collection paradigm with online game and its psychological and acoustical assessment,” Acoust. Sci. Tehnol., vol. 33, no. 6, pp. 359-369, 2012.

（7） “宇都宮大学　パラ言語情報研究向け　音声対話データベース UUDB（Utunomiya University Spoken Dialogue Database for Paralinguistic Information Studies），”http://uudb.speech-lab.org/about.html

（2020年7月31日受付　2020年8月25日最終受付）

平沢（ひらさわ）　純一（じゅんいち）

1995奈良先端科学技術大学院了．同年日本電信電話株式会社入社．音声対話システムの研究開発に従事．ニュアンスコミュニケーションズ株式会社，(株)フュートレックを経て，2018一般社団法人固有名詞協会．2014からJEITA音声入出力方式標準化専門委員会音声合成グループ主査．

中嶌（なかじま）　信弥（しんや）

1982慶大大学院工学研究科修士課程了．同年日本電信電話公社（現NTT）入社．横須賀通研勤務．1990～1991米国ロチェスター大計算機科学部客員研究員．現在国士舘大・理工・教授．音声画像処理技術，対話処理技術，ヒューマンインタフェース技術の研究に従事．工博．2010からJEITA音声入出力方式標準化専門委員会委員長．