

# JEITA

電子情報技術産業協会規格

Standard of Japan Electronics and Information Technology Industries Association

*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

**The Guidelines for TTS Speaking Style Classification**

2018年7月制定

作 成

音声入出力方式標準化専門委員会

Speech Input/Output Systems Standardization Subcommittee

発 行

一般社団法人 電子情報技術産業協会

Japan Electronics and Information Technology Industries Association

## 目 次

ページ

まえがき	
1 適用範囲	1
2 引用規格並びに文書	1
3 用語及び定義	1
4 要求事項	2
5 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	2
5.1 感情や意図を表現する音声合成技術の現状	2
5.2 JEITA「音声合成 話し方種別ガイドライン」作成の目的	3
5.3 「音声合成・話し方種別」ガイドラインとは	3
5.3.1 本ガイドラインにおける「話し方」とは	3
5.3.2 本ガイドラインの特徴	4
5.4 本ガイドラインが対象としていない内容	4
5.5 本ガイドラインの位置付け	5
6 話し方種別	5
6.1 話し方種別にグループとカテゴリを割り当てた理由	6
6.2 話し方種別のグループとカテゴリ	6
6.3 話し方種別の一覧	7
7 話し方種別の実現フェーズ	9
7.1 実現フェーズを設定する意義	9
7.2 実現フェーズの定義	9
7.3 実現フェーズごとの話し方種別の一覧	10
7.3.1 実現フェーズ 1	10
7.3.2 実現フェーズ 2	10
7.3.3 実現フェーズ 3	11
7.3.4 実現フェーズ 4	12
7.4 ユースケース	13
7.4.1 実現フェーズ 1 までの話し方種別で対応が期待できるユースケース	13
7.4.2 実現フェーズ 2 までの話し方種別で対応が期待できるユースケース	13
7.4.3 実現フェーズ 3 までの話し方種別で対応が期待できるユースケース	13
7.4.4 実現フェーズ 4 までの話し方種別で対応が期待できるユースケース	13
8 例文	14
8.1 平静グループ	14
8.2 感情グループ	14
8.3 発話意図グループ	18
8.4 その他グループ	19
9 解説	20
9.1 音声合成における「話し方」の技術開発の歴史	20
9.1.1 音声合成を利用する側（コンテンツ作成サイド）の事情	20
9.1.2 音声合成を技術開発・提供する側（エンジンベンダサイド）の事情	20

9.2	本ガイドラインの特徴	21
9.3	本ガイドラインで選定した話し方種別 55 種の選定過程	22
9.4	本ガイドラインにおける「話し方」とは	22
9.5	個人性, キャラクタ, 役割	23
9.6	本ガイドラインの範囲から除外した内容	23
9.7	本ガイドラインの活用法	23
9.7.1	音声合成を利用する側 (コンテンツ作成サイド)	23
9.7.2	音声合成を技術開発・提供する側 (エンジンベンダサイド)	24
10	今後の課題	25
10.1	話し方種別同士の関係	25
10.2	話し方の強度や程度	25
10.3	入力テキストでの記法	25
10.4	出力される音声自体の特徴に関する規定	25
10.5	出力される合成音声の解釈, 評価に関する記述	26
11	まとめ	26
	付録	27
A.1	SSML	27
A.2	各社音声合成における「話し方種別」の現状	27
A.2.1	ANIMO Fine Speech (株式会社アニモ)	27
A.2.2	AITalk (株式会社イーアイ)	27
A.2.3	FutureVoice (NTT テクノクロス株式会社)	27
A.2.4	RECAIUS (東芝デジタルソリューションズ株式会社)	28
A.2.5	Ruby Talk (株式会社 日立超 LSI システムズ)	28
A.2.6	VoiceText (HOYA 株式会社)	28
A.3	感情に関する研究	28
A.3.1	基本感情	28
A.3.1.1	ポール・エクマンの基本感情	28
A.3.1.2	プルチックの感情モデル	29
A.3.2	次元モデル	29
A.3.2.1	ヴントのモデル	29
A.3.2.2	ラッセルの円環モデル	30
A.4	発話意図タグに関する研究	30
A.5	感情や意図ラベルのセットに関する提案例	32
A.5.1	OGVC	32
A.5.2	UADB	33
A.5.3	感情表現辞典	33
A.5.4	テキストマイニングにおける感情・意志分析	34
A.6	その他の文字コード・規格との関係	37
A.6.1	Emotion ML	37
A.6.2	UNICODE 6.0 の顔文字 (emoticon ブロック)	37
A.6.3	Amazon Alexa Speechcon	38
A.7	IT-4012 審議委員会の構成表	39

## まえがき

本ガイドラインは、音声合成技術の一層の利用拡大を図ることを目的として、昨今、ニーズの高い「感情や意図を表現する話し方」について、話し方の種別の一覧をガイドラインとして示すものである。

本ガイドラインは、一般社団法人 電子情報技術産業協会 音声入出力標準化専門委員会 音声合成グループにより作成された。

## 電子情報技術産業協会規格

# 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

## The Guidelines for TTS Speaking Style Classification

### 1 適用範囲

本ガイドラインは、音声合成技術において、感情や意図を表現する「話し方」について、「話し方種別」に関する提案を示すものである。

これまでの音声合成技術では、人間が日常的なコミュニケーションにおいて活用しているような、多彩な「話し方」を実現させることができていなかっただけでなく、そもそも、どのような「話し方」が要望されているのか、どのような「話し方」から開発を進めて行けばよいのか、を利用者と開発者の間で議論するための共通の基盤すら確立できていなかった。

そこで、このような現状を鑑みて、本ガイドラインでは、

- 1) 「話し方」種別の一覧、及びグループやカテゴリへの分け方
- 2) 「話し方」種別の実現性（実現フェーズ）
- 3) それぞれの「話し方」種別ごとの例文のサンプル

を示している。

### 2 引用規格並びに文書

なし。

### 3 用語及び定義

以下、本ガイドラインで用いる主な用語とその説明を示す。（50音順）

#### 1) 韻律

発声される音声の音調（イントネーション）、強弱、話速、リズム（音長、ポーズを含む）などの音声的な性質。

#### 2) 音声合成（Text To Speech, TTS）

音声を機械的に生成する処理。テキストを入力として、入力されたテキストに対応する音声を生成するものを「テキスト音声合成」と呼ぶ。以降、本ガイドラインでは「テキスト音声合成」を「音声合成」と称している。

#### 3) 音声合成エンジン

音声合成技術により、音声を生成するソフトウェア・プログラム。

#### 4) 音声対話システム

入力された音声を認識結果テキストとして出力する音声認識技術、その認識結果テキストに応じて次に応答すべき内容を定める対話処理技術、出力すべきテキストを音声として出力する音声合成技術などから構成されるシステム。アプリケーション。人間（ユーザ）との間で音声言語によるコミュニケーションを可能とする。