

JEITA

電子情報技術産業協会規格

Standard of Japan Electronics and Information Technology Industries Association

JEITA IT-4005

音声認識エンジン性能評価方法のガイドライン The Guidelines for Performance Evaluation of Speech Recognition Engine

2008年2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

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	3
4.1 評価の目的	3
4.1.1 エンドユーザ	3
4.1.2 音声認識応用システム開発者	3
4.1.3 音声認識エンジン開発者	4
4.2 評価の対象	4
4.3 評価への要求	4
5 音声認識性能に影響を及ぼす要因	5
5.1 認識対象に起因する要因	5
5.1.1 単語の数	5
5.1.2 文法	5
5.1.3 音の類似性	5
5.1.4 対象外単語	5
5.2 環境に起因する要因	6
5.2.1 マイクロフォンを用いる音声認識	6
5.2.2 電話を用いる音声認識	8
5.2.3 雑音	10
5.2.4 音声収録機器・再生機器	11
5.3 話者に起因する要因	11
5.3.1 特定の話者を対象とした音声認識	11
5.3.2 不特定の話者を対象とした音声認識	12
5.3.3 話者に係わるその他要因	12
5.4 評価の目的と要因	13
6 評価手順	14
6.1 評価語彙・文法の決定	14
6.1.1 対象語彙・文法が既知の場合	14
6.1.2 対象語彙・文法が未知の場合	14
6.1.3 語彙・文法の作成の留意点	14
6.2 評価用発声内容の決定	15
6.2.1 全単語・単語列による評価	16
6.2.2 対象の一部の選択	16

6.3	被験者の決定	17
6.3.1	人数	17
6.3.2	話者の選択	17
6.3.3	その他の留意点	18
6.4	評価方法の決定	18
6.5	使用機器と環境の決定	18
6.5.1	マイクロフォン	19
6.5.2	電話	19
6.5.3	実験環境	19
6.5.4	環境の模擬	20
6.6	被験者へのインストラクションと発声	22
6.6.1	被験者へのインストラクション	22
6.6.2	話者、発話情報の作成	22
6.6.3	発声	22
6.7	音声認識エンジンへの入力と結果の収集	23
6.8	認識結果の集計と評価尺度	25
6.8.1	正解の作成	25
6.8.2	正解認識・誤認識の判定	25
6.8.3	評価尺度	26
6.9	認識結果の評価と検定	27
6.9.1	平均認識率	27
6.9.2	最低認識率	27
6.9.3	認識率に対する話者分布	27
6.9.4	統計検定	28
6.10	認識結果の公開	28
	解説	29

まえがき

本ガイドラインは、有効で正確な音声認識エンジンの評価方法の普及による音声認識の利用者と利用分野の拡大を目的として、音声認識の適用用途、利用分野を考慮した音声認識エンジンの標準的な評価方法と評価時の留意点を示すものである。本ガイドラインは、社団法人 電子情報技術産業協会 音声入出力方式標準化専門委員会、音声認識グループにより作成された。

電子情報技術産業協会規格

音声認識エンジン性能評価方法のガイドライン

The Guidelines for Performance Evaluation of Speech Recognition Engine

1 適用範囲

本ガイドラインは、音声を入力として、その内容を表す記号や記号列を出力する音声認識エンジンの性能評価方法について示すものである。

本ガイドラインにおいて、性能とは、入力音声に対して、いかに正しく記号や記号列を出力できるかを示し、性能評価とは、得られた記号や記号列の正しさを評価することである。したがって、本ガイドラインでは、音声認識エンジンにおける応答速度やコストの評価、音声認識エンジンを用いる応用システムの評価は含めない。

2 引用規格並びに文書

社団法人 電子情報技術産業協会「日本語音声認識用読み記号 (JEITA IT-4003)」

3 用語及び定義

音声認識応用システムの一般的な構成を図に示す。

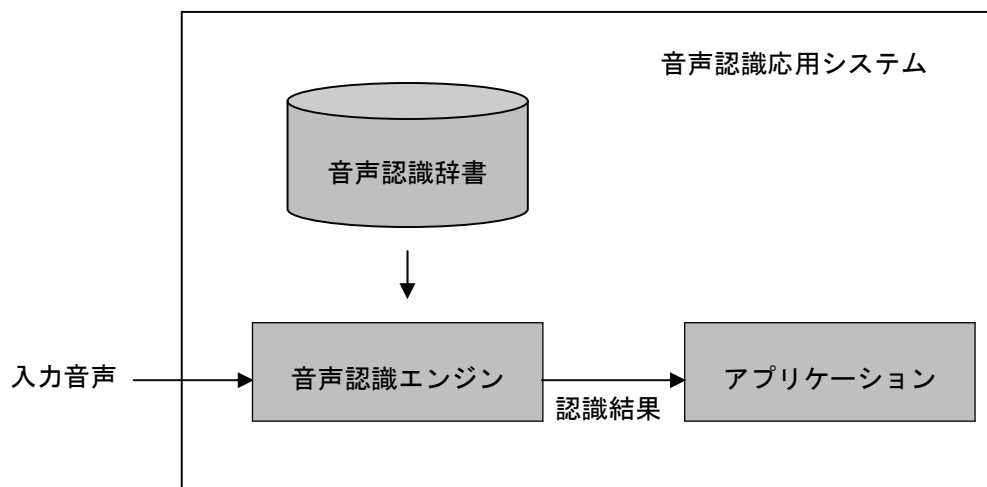


図1—音声認識応用システムの構成

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。

(1) 音声認識

入力された音声を処理し、その音声に対応する記号あるいは記号列を認識結果として出力すること。ここで、記号とは単語や意味記号、記号列とは単語列や意味記号列を指す。

(2) 単語

開発者あるいはユーザにより定義される音声認識の単位。言語学上の形態素（単語）である必要はない。