

自動車内における連続数字音声コーパス CENSREC-2 の設計と評価

藤本 雅清[†] 武田 一哉^{††} 中村 哲[†]

[†] ATR 音声言語コミュニケーション研究所
E-mail: {masakiyo.fujimoto, satoshi.nakamura}@atr.jp
^{††} 名古屋大学大学院 情報科学研究科
E-mail: kazuya.takeda@nagoya-u.jp

あらまし 本稿では、SLP 雑音下音声認識評価ワーキンググループの活動成果として、自動車内における連続数字音声認識の評価用データベース CENSREC-2 と、標準評価スクリプトによるベースライン評価結果について述べる。音声データの収録は、接話マイクと遠隔マイクの 2 種類を用いて、3 種類の走行速度と 4 種類の車内環境を組み合わせた 11 種類の環境下で行っており、これらの音声データを用いた 4 種類の評価環境を提供する。

キーワード 雑音下音声認識、共通評価フレームワーク、自動車内音声データベース、CENSREC-2

CENSREC2: Corpus and Evaluation Environments for In Car Continuous Digit Speech Recognition

Masakiyo FUJIMOTO[†], Kazuya TAKEDA^{††}, and Satoshi NAKAMURA[†]

[†] ATR Spoken Language Communication Research Laboratories
E-mail: {masakiyo.fujimoto, satoshi.nakamura}@atr.jp
^{††} Graduate School of Information Science, Nagoya University
E-mail: kazuya.takeda@nagoya-u.jp

Abstract This paper introduces a common database and an evaluation framework for connected digit speech recognition in real driving car environments, CENSREC-2, as an outcome of IPSJ-SIG SLP Noisy Speech Recognition Evaluation Working Group. Speech data of CENSREC-2 was collected using two microphones, a close-talking microphone and a hands-free microphone, under carefully controlled 11 different driving conditions, i.e., combinations of three car speeds and four car conditions. CENSREC-2 provides four evaluation environments which are designed using speech data collected in these car conditions.

Key words noisy speech recognition, common evaluation framework, in-car speech database, CENSREC-2

1. はじめに

近年の音声認識技術の進歩は、統計モデルの導入と大規模コーパスの収集によりもたらされた。しかし、現在の音声認識を実際に利用されるような雑音のある環境で利用すると、未だ著しい性能劣化が避けられない。このような問題に対し、筆者らは、2001 年 10 月に情報処理学会 音声言語情報処理研究会内に、有志によるワーキンググループを作り、雑音下日本語音声認識の評価のための議論を進めてきた。ワーキンググループの活動成果として筆者らは、雑音下音声認識の標準環境である AURORA-2J [1]、及び CENSREC-3 (Corpus and Environments for Noisy Speech REcognition) [2] を作成し、配布を行った。本稿では、新たな雑音下音声認識の標準環境である

CENSREC-2 の概要と、ベースライン性能について述べる。

2. 自動車内音声データの収録

2.1 収録語彙

CENSREC-2 は、連続数字音声の認識を対象としており、語彙は数字 11 種類 (1~9, 0 (まる), Z (ぜろ)), 無音 (sil), ショートポーズ (sp) の 13 種類である。発話内容、発音スタイル等は、AURORA-2J [1] と同様である。

2.2 音声データの収録環境

自動車内音声の収録は、特別に設計された実験車両を用いて行う。CENSREC-3 では、接話マイクと運転手席のバイザー部分に設置された遠隔マイクを用いて音声を収録する [2]。それぞれのマイクには、SONY ECM77B を使用し、標準化周波数 16kHz、語長 16bit で A/D 変換を行った。

データの収録は、表 1 に示す、3 種類の走行速度（アイドリング、低速（市街地）走行、高速走行）と、4 種類の車内環境（通常走行、エアコン On、オーディオ On、窓開）を組み合わせた 11 種類の環境で行う。

表 1 音声データの収録環境

走行速度	車内環境
アイドリング	通常走行、エアコン On、オーディオ On、窓開
低速走行	通常走行、エアコン On、オーディオ On、窓開
高速走行	通常走行、エアコン On、オーディオ On

評価データの発話者数は 104 名（男性 52 名、女性 52 名）であり、収録音声の総数は 17,651 発話である。これらの収録音声のうち、73 名（男性 33 名、女性 40 名）の話者のデータを学習データとし、31 名（男性 19 名、女性 12 名）の話者のデータを評価データとした。学習データの総数は 14,687 発話（接話マイク: 7,492、遠隔マイク: 7,195）であり、評価データの総数は 2,964 発話（遠隔マイクのデータのみ）である。

3. 評価環境の設計

CENSREC-2 では、様々な環境で収録された音声データを用いて、4 種類の音声認識評価環境（Cond. 1~4）を構成する。

表 2、3 の ○ 印は、各々の評価環境で用いるデータ、表 4、5 は、使用するデータ量を示しており、それぞれの環境は、
Cond. 1: 学習と評価でマイク種別、収録環境が共に一致
Cond. 2: 学習と評価でマイク種別が一致、収録環境が相違
Cond. 3: 学習と評価でマイク種別が相違、収録環境が一致
Cond. 4: 学習と評価でマイク種別、収録環境が共に相違
という条件を設定している。

4. ベースライン認識性能

4.1 評価用スクリプト

評価用ベースラインスクリプトは、HTK [3] を用いて HMM の学習、評価実験が容易に行えるように作成し、以下のようなベースライン評価の仕様を設計した。

- 単語音声の認識は単語単位 HMM により行い、数字 (0~9, Z) は 16 状態 20 混合分布、無音 (sil) は 3 状態 36 混合分布、ショートポーズ (sp) は 1 状態 36 混合分布（無音モデルの第 2 状態と共有）である。

- 特徴量は、HTK の HCopy により抽出された、12 次 MFCC + log-power + Δ MFCC + Δ log-power + $\Delta\Delta$ MFCC + $\Delta\Delta$ log-power の 39 次元とする。分析条件は、 $1 - 0.97z^{-1}$ のプリエンファシス、ハミング窓、24 次元のメルフィルタバンク、20ms の分析フレーム長、10ms のフレームシフトとする。また、自動車雑音特有の低周波成分に対処するため、メルフィルタバンク分析時に 250Hz 以下の低周波成分を取り除く。

4.2 ベースライン認識結果と認識性能の比較

表 6 は Microsoft Excel にて作成されたスプレッドシートである。表の上段は、ベースライン性能を示しており、中段に自身の手法による認識結果を入力する。中段に認識結果を入力すると、下段にベースライン性能との相対的な改善性能（誤り改善率）が自動で出力される。また、CENSREC-2 においても、AURORA-2J、CENSREC-3 と同様に、バックエンドの変更に対して、その度合に応じたカテゴリーを設定する [1], [2]。

表 2 各評価環境で用いられる学習データ条件

評価環境	Cond. 1	Cond. 2	Cond. 3	Cond. 4
マイクロホン	接話	遠隔	接話	遠隔
アイドリング	—	○	—	○
低速走行	—	○	—	○
高速走行	—	○	—	○

表 3 各評価環境で用いられる評価データ条件

評価環境	Cond. 1	Cond. 2	Cond. 3	Cond. 4
アイドリング	○	—	—	—
低速走行	○	○	○	○
高速走行	○	○	○	○

表 4 各評価環境における学習データ数

走行速度	マイク	Cond. 1	Cond. 2	Cond. 3	Cond. 4
アイドリング	接話	—	—	2,737	2,737
	遠隔	2,397	2,397	—	—
低速走行	接話	—	—	2,728	—
	遠隔	2,757	—	—	—
高速走行	接話	—	—	2,027	—
	遠隔	2,041	—	—	—
合計		7,195	2,397	7,492	2,737

表 5 各評価環境における評価データ数

走行速度	Cond. 1	Cond. 2	Cond. 3	Cond. 4
アイドリング	906	—	—	—
低速走行	1,180	1,180	1,180	1,180
高速走行	878	878	878	878
合計	2,964	2,058	2,058	2,058

表 6 CENSREC-2 スプレッドシートとベースライン性能

CENSREC-2 Evaluation Results				
CENSREC-2 Baseline Results (%)				
Condition 1	Condition 2	Condition 3	Condition 4	Average
80.58	74.49	61.46	48.87	66.35
CENSREC-2 Word Accuracy (%)				
Condition 1	Condition 2	Condition 3	Condition 4	Average
CENSREC-2 Relative Improvement				
Condition 1	Condition 2	Condition 3	Condition 4	Average

5. む す び

本稿では、自動車内音声認識の評価用データベース CENSREC-2 と、ベースライン評価結果について報告した。今後、非定常雑音下及び、残響下での認識など、環境を徐々に難しくした評価環境を設計する予定である。

AURORA-J/CENSREC Web site:

<http://sp.shinshu-u.ac.jp/CENSREC/>

謝辞 本研究の一部は、情報通信研究機構の研究委託により実施したものである。

文 献

- [1] S. Nakamura et al., "AURORA-2J: An Evaluation Framework for Japanese Noisy Speech Recognition," *IEICE Transactions on Information and Systems*, Vol. E88-D, No. 3, pp. 535-544, Mar. 2005.
- [2] 藤本 雅清 他, "実走行車内単語音声データベース CENSREC-3 と共通評価環境の構築," 情報処理学会研究報告, SLP-55-8, pp.41-46, Feb. 2005.
- [3] HTK Web site, <http://htk.eng.cam.ac.uk/>