

## 雑音下音声認識評価ワーキンググループ活動報告： 認識に影響する要因の個別評価環境(2)

北岡 教英<sup>1</sup> 山田 武志<sup>2</sup> 滝口 哲也<sup>3</sup> 柘植 寛<sup>4</sup> 山本 一公<sup>5</sup>  
宮島千代美<sup>1</sup> 西浦 敬信<sup>6</sup> 中山 雅人<sup>6</sup> 傳田 遊亀<sup>6</sup> 藤本 雅清<sup>7</sup>  
田村 哲嗣<sup>8</sup> 松田 繁樹<sup>9</sup> 小川 哲司<sup>a</sup> 黒岩 眞吾<sup>b</sup> 武田 一哉<sup>1</sup>  
中村 哲<sup>9</sup>

<sup>1</sup>名古屋大学 <sup>2</sup>筑波大学 <sup>3</sup>神戸大学 <sup>4</sup>徳島大学 <sup>5</sup>豊橋技術科学大学  
<sup>6</sup>立命館大学 <sup>7</sup>NTT CS 基礎研 <sup>8</sup>岐阜大学 <sup>a</sup>早稲田大学 <sup>b</sup>千葉大学 <sup>9</sup>NiCT/ATR  
E-mail: kitaoka@nagoya-u.jp

**あらまし** 音声認識実用化において、雑音下の音声認識の性能向上が叫ばれている。現在も多くの研究が行われているが、これらの手法を客観的に比較評価する標準評価基盤が必要と考えられる。我々は2001年10月から情報処理学会音声言語情報処理研究会の下で雑音下音声認識評価ワーキンググループとして活動し、標準評価基盤 CENSREC シリーズを構築・配布している。これまでの CENSREC シリーズを概観し、さらに今年度新たに配付する残響下音声認識評価基盤 CENSREC-4 の概要を述べる。そして、ワーキンググループ最終年度に向けて、今後どのような方針で評価基盤を設計・構築・配付していくのかを述べる。

**キーワード** 音声認識, 雑音, 評価基盤, CENSREC

## Progress Report of SLP Noisy Speech Recognition Evaluation WG: Individual evaluation framework for each factor affecting recognition performance (2)

Norihide KITAOKA<sup>1</sup>, Takeshi YAMADA<sup>2</sup>, Tetsuya TAKIGUCHI<sup>3</sup>,  
Satoru TSUGE<sup>4</sup>, Kazumasa YAMAMOTO<sup>5</sup>, Chiyomi MIYAJIMA<sup>1</sup>,  
Takanobu NISHIURA<sup>6</sup>, Masato NAKAYAMA<sup>6</sup>, Yuki DENDA<sup>6</sup>,  
Masakiyo FUJIMOTO<sup>7</sup>, Satoshi TAMURA<sup>8</sup>, Shigeki MATSUDA<sup>9</sup>, Tetsuji OGAWA<sup>a</sup>,  
Shingo KUROIWA<sup>b</sup>, Kazuya TAKEDA<sup>1</sup>, and Satoshi NAKAMURA<sup>9</sup>

<sup>1</sup> Nagoya Univ. <sup>2</sup> Univ. of Tsukuba <sup>3</sup> Kobe Univ. <sup>4</sup> Univ. of Tokushima <sup>5</sup> Toyohashi Univ. of Tech.

<sup>6</sup> Ritsumeikan Univ. <sup>7</sup> NTT CS research Lab. <sup>8</sup> Gifu Univ. <sup>a</sup> Waseda Univ. <sup>b</sup> Chiba Univ. <sup>9</sup> NiCT/ATR

E-mail: kitaoka@nagoya-u.jp

**Abstract** Performance improvement of noisy speech recognition is urgent for practical use of speech recognition and methods for this purpose should be compared on common evaluation frameworks. We organized a working group under Special Interest Group of Spoken Language Processing in Information Processing Society of Japan, to develop evaluation frameworks of noisy speech recognition to compare many methods for processing of noisy speech. In this paper, we review the series of CENSREC series and then introduce the reverberant speech recognition evaluation framework CENSREC-4, the newest CENSREC. Finally we describe the road-maps of future CENSRECs.

**Key words** Speech recognition, noise, evaluation frameworks, CENSREC.

## 1. はじめに

音声認識の実用化研究が進められている。その中で常に中心的な研究課題となるのが、実環境のさまざまな音声認識劣化要因への対処法である。これらを広義で雑音と呼ぶ。かねてより多数の雑音下音声認識手法の研究があるが、個々の実験条件の下で性能が評価されてきた、これらの手法を客観的に比較評価できる枠組を提供することが望まれている。これにより、たとえば研究成果の実用システムへの利用が促進されるものと考えられる。

我々はこうした問題意識のもと、2001年10月に雑音下音声認識評価ワーキンググループを組織し、これらの評価基盤の構築を行ってきた。これらは、雑音下音声認識手法をさまざまな側面から評価できる評価基盤群をなしてきている[3]~[7]。

この活動は、欧州におけるAURORA研究プロジェクトに準ずる形で開始されたものである。雑音抑圧手法の標準評価コーパスとHTKを利用した標準認識スクリプト、および標準評価用Excelシートをセットとした評価基盤を配布している。このAURORAから提案されてきたものとして、TI digitに雑音を付与した連続数字音声認識タスクのAURORA-2、自動車内で実際に発声した連続数字認識タスクのAURORA-3、Wall Street Journalタスクをベースとした雑音下大語彙連続音声認識タスクのAURORA-4がある[1], [2]。我々の提供する評価基盤がAURORA-2の日本語版として提供したAURORA-2J(CENSREC-1)から始まったように、AURORAプロジェクトの動向を注視しながら活動を進めてきた。そして、AURORAの進展が複雑な環境での評価に移りつつあるのに対して、実環境における認識性能低下要因の個別評価環境の構築という方針のもと、現在では独立した活動を行っている。

本稿では、まずこれまでの基盤群を簡単に振り返り、これらの間の関係を整理する。これに基づき、現在構築中の残響化音声認識評価基盤CENSREC-4の位置づけとその概要を述べる。さらに、今後の展開について述べる。

## 2. CENSREC シリーズ

まず、本ワーキンググループで雑音下音声認識の標準評価基盤として作成・配布している、CENSREC(Corpus and Environment for Noisy Speech REcognition) シリーズを概観する。

### 2.1 CENSREC-1/AURORA-2J [8]

欧州AURORAプロジェクトの作成した雑音環境下連続英語数字音声認識タスクの共通評価フレームワークであるAURORA-2の日本語版である。2003年7月の配布開始から、100部を超える部数を配布し、各種研究発表で利用されている。

学習/テストデータとともに学習認識スクリプトとベースライン性能を提供し手軽に実験を開始できる。タスクが連続数字認識とシンプルで、加算性雑音の影響とその対策にほぼ絞られている。学習環境は、clean training (クリーン音声によるモデル学習)、multicondition training (雑音重畳音声による学習)があり、Multicondition trainingの場合にのみ4種類の雑

音(Subway, Babble, Car, Exhibition)を5種類のSNRレベル(clean, 20dB, 15dB, 10dB, 5dB)で重畳した音声を用いて学習を行う。テストデータは8種類の雑音を7種類のSNRレベル(clean, 20dB, 15dB, 10dB, 5dB, 0dB, -5dB)で重畳しており、半分は既知雑音、残りは未知雑音環境下での評価となる。

### 2.2 CENSREC-2 [9]

CENSREC-1と同様の連続数字データを自動車内で発話したデータに基づいており、CENSREC-1の雑音加算シミュレーションに対し実収録データへの効果を確認できる。本評価環境は、2005年12月より配布を開始している。

自動車内で接話マイクロホンと天井の1本(遠隔マイクロホンと呼ぶ)で収録された音声を用いた[12]。収録条件は、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, エアコン On, オーディオ On, 窓あけ)を組み合わせた11種類の環境である。これらの音声データを組み合わせ、そのときの学習データと評価データの収録条件(マイク種別, 収録環境)の一致度合により、以下の4種類の音声認識評価条件が設定されている。

- Condition 1: マイク種別, 収録環境がともに一致
- Condition 2: マイク種別が一致, 収録環境が相違
- Condition 3: マイク種別が相違, 収録環境が一致
- Condition 4: マイク種別, 収録環境がともに相違

### 2.3 CENSREC-3 [10]

CENSREC-3も実環境での発話を対象とする。実走行車内での孤立単語音声認識の評価環境で、2005年2月より配布を行っている。

収録環境はCENSREC-2と同じで、接話マイクロホンと遠隔マイクロホンの2種類を用いて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、6種類の車内環境(通常走行, ハザード On, エアコン(Low), エアコン(High), オーディオ On, 窓開)を組み合わせた16種類の環境[14]である。学習データと評価データの環境の組み合わせにより、欧州AURORAプロジェクトのAURORA-3のWell-matched condition(WM), Moderate-mismatched condition(MM), High-mismatched condition(HM)に準じた6種類の評価環境(Condition 1~6)を設定している。

- Condition 1, 2, 3 マイク種別, 収録環境がともに一致(WM)
- Condition 4 マイク種別が一致, 収録環境が相違(MM)
- Condition 5, 6 マイク種別, 収録環境(の一部)が相違(HM)

### 2.4 雑音下音声区間検出評価基盤 CENSREC-1-C [11]

音声区間検出(Voice activity detection; VAD)は雑音下の音声認識において大きな役割を果たすとして、最近多くの研究例がある。我々は2007年9月から、雑音下VAD評価基盤であるCENSREC-1-Cの配布を開始した。これは、雑音下の連続数字発話が雑音区間を挟んでさらに接続されたデータと、VADの結果の評価ツールからなる。

シミュレーションデータは1話者による9ないし10発話を接続して作成された。接続されるデータは全てCENSREC-1のものである。CENSREC-1のデータ中の1秒の無音を間に

挟んで接続されている。また、実環境におけるデータも収録している。このデータは、二つの実際の雑音環境（レストランおよび高速道路脇）で、雑音レベルの高い時間帯と低い時間帯（それぞれ低 SNR 条件と高 SNR 条件に対応）を選んで、人間が 50cm 離れたマイクロフォンに向かって発声したものを収録した。10 人の話者が、1 環境 1SNR につき 4 回の収録を行い、各回につき連続数字 8~10 発話を 2 秒程度の無音を挟んで断続的に発している（従って、各人が 1 条件につき、38~39 発話している）。一人の話者は発声に問題があったため除外したので、9 名により、トータルで 1380 発話収録された。また、2 種類の評価基準も定義している。ひとつはフレーム単位の検出性能であり、もうひとつは音声認識を指向した発話単位の検出性能である。音声パワーに基づくベースラインの VAD による結果をベースラインとして公開し、それとの比較が行える。

### 2.5 CENSREC 基盤間の関係

昨年度の報告 [7] において、(広義での) 雑音下音声認識において評価すべき項目を、「環境」と「対処のアプローチ」の面で検討した。

この観点では、CENSREC-1~3 は環境に着目した基盤であり、その中でも最大の要因である加算性雑音の範囲内で詳細に検討できることを意図したものである。CENSREC-1 は連続数字発声に種々の雑音を付加したデータによる実験環境である。これに対し、自動車内の実環境下で同様の内容を実際に発声して収録したデータを用いて、実環境とシミュレーションの差が比較できる。さらに、タスクによる影響も考えられるため、その検討のために、CENSREC-2 を孤立単語認識タスクに変更したものが CENSREC-3 である。すなわち、雑音下音声認識の困難さを「雑音の違い」「シミュレーション/実環境の違い」および「タスクの違い」の軸上で比較できる環境群となっている。一方、CENSREC-1-C はアプローチに着目した基盤であり、いわゆる（フロントエンドやバックエンドを含む）認識手法とは異なる戦略での雑音下音声認識への取り組みを考慮したものである。我々はこれらの面を考慮しながら基盤整備を進めていると述べ [7]、実際に構築を進めている。特に環境に関しては、大きな要因として加算性雑音と残響があることも言及した。発声された音声信号が受信されるまでの過程が

$$y = h \otimes x + n$$

としばしば書かれるように、加算性雑音と残響に代表される乗算性歪みが 2 大要因と考えてよいであろう。次節では、まもなく公開予定である、残響下音声認識評価基盤 CENSREC-4 について述べる。

### 3. 残響下音声認識評価基盤 CENSREC-4

CENSREC-4 は、様々な残響環境における遠隔発話の音声認識の評価を目的とする評価環境であり、19 年度中に配布が開始される予定である。CENSREC-4 は、これまでの CENSREC-1 と同様に連続数字発話で構成されており、大きく二つのサブセットに分けることができる。ひとつは「基本データ (Basic data)」であり、もうひとつが「追加データ (Extra data)」で

表 1 残響収録機材と条件

Table 1 Recording equipment and conditions

Microphone	SONY, ECM-88B
Microphone amplifier	Thinknet, MA-2016C
A/D board	TOKYO ELECTRON DEVICE, TD-BD-8CSUSB-2.0
Loudspeaker	B&K, Mouth simulator Type 4128
Speaker amplifier	YAMAHA, P4050
Sampling frequency	16 kHz (Recording is 48 kHz)
Quantization	16 bits

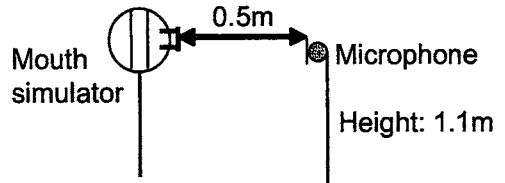


図 1 インパルス応答収録時の設定

Fig. 1 Recording setup for impulse responses

あり、基本データに対してのみ、HTK による HMM 学習/認識スクリプトが提供される。

#### 3.1 基本データ

##### 3.1.1 インパルス応答データ

種々の残響環境をシミュレーションするために、いくつかのインパルス応答を収録した。インパルス応答の収録には Time stretched pulse (TSP) 法 [13] を用いた。TSP 信号の長さは 131,072 点で、16 回の同期加算により高 SNR を確保している。

CENSREC-4 では 8 種類（オフィス、エレベータホール、車内、リビングルーム、ラウンジ、和室、会議室、浴室）の環境で収録したインパルス応答を用いている。収録条件を表 1 に示す。

車内と浴室を除く多くの環境では図 1 に示す設定で収録されている。多くの場合において、マイクロフォンは部屋の中央付近に設置した。車環境としては中型セダン車を用い、マウスシミュレータを運転席に、マイクロフォンをサンバイザ位置に設置した。マウスシミュレータとマイクロフォン間の距離は 0.4m である。また、ラウンジ環境ではマイクロフォンをテーブルの上に設置した。浴室環境ではマイクロフォンを水で満たされているバスタブの上方に、マイクロフォンは 0.3m 離れた壁に設置した。

各環境の収録時における残響時間 ( $T_{60}$ )、暗騒音レベル、湿度および気温を表 2 に示す。残響時間および暗騒音レベルの解像度はそれぞれ 0.05 秒および 0.5dB である。

##### 3.1.2 シミュレーションデータ (セット A, セット B)

残響環境下の音声を、クリーン音声にインパルス応答を畳み込むことによってシミュレーションしたデータセットを提供する。クリーン音声は CENSREC-1 のものと同一であり、11 種類 (“ichi,” “ni,” “san,” “yon,” “go,” “roku,” “nana,” “hachi,” “kyu,” “zero,” および “maru”) の日本語数字の連結された連続数字である。これらは CENSREC-1 作成時に収録されたものを用いている。防音室内でヘッドセットマイク

表 2 インパルス応答収録環境の残響時間, 暗騒音レベル, 湿度および気温

Table 2 Reverberation time, ambient noise level, humidity, and temperature of recording

Room	Test set	Room size [m]	Reverberation time [ $T_{60}$ ]	Temperature	Humidity	Ambient noise level [dBA]
Office	A	9.0 x 6.0	0.25 sec	30°C	40%	36.5 dB
In front of an elevator	A	11.5 x 6.5	0.75 sec	30°C	50%	39.0 dB
In-car	A	Middle-sized sedan	0.05 sec	29°C	44%	32.0 dB
Living room	A	7.0 x 3.0	0.65 sec	30°C	54%	34.0 dB
Lounge	B	11.5 x 27.0	0.50 sec	27°C	50%	52.5 dB
Japanese tatami-floored room	B	3.5 x 2.5	0.40 sec	30°C	54%	30.0 dB
Meeting room	B	7.0 x 8.5	0.65 sec	27°C	52%	48.5 dB
Bath	B	1.5 x 1.0	0.60 sec	31°C	62%	29.5 dB

ロフォン (Sennheiser MHD25) を用いて収録された。ただし、CENSREC-1(8kHz) とは異なり 16kHz サンプルングである。各残響環境につき話者数 104 名 (男性 52 名, 女性 52 名) である。基本データのテストセットは表 2 に示すようにセット A とセット B の二つのサブセットからなる。CENSREC-1 におけるセット A (既知雑音) とセット B (未知雑音) と同様に、CENSREC-4 では、セット A で用いられたインパルス応答は提供される学習データの作成にも用いられ、セット B で用いられるものは学習データには用いない。

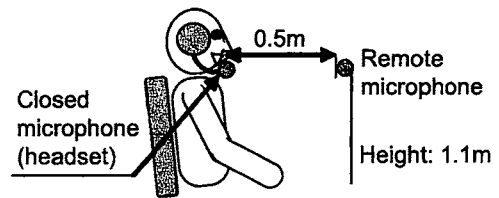
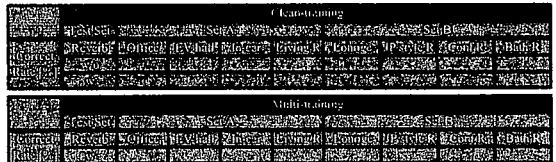


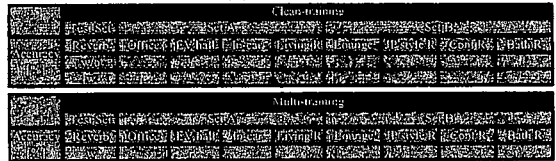
図 2 Recording setup for real data

表 3 Evaluation sheet for basic sets

(a) Correct



(b) Accuracy



### 3.2 追加データ

#### 3.2.1 加算性雑音も重畳されたシミュレーションデータ (セット C)

追加データには、残響のインパルス応答の畳み込みと加算性雑音の重畳の両方が施されたデータが収録されている。このシミュレーションデータは、まず室内のインパルス応答を畳み込んだ後、そのインパルス応答を収録した場所で収録された雑音を、SNR を調整して加算することで作成されている。これらはセット C と呼ばれ、これに含まれるテストセットはセット A 中の 2 環境とセット B 中の 2 環境が含まれている。重畳雑音は、SNR がクリーン、20dB、10dB、および 5dB となるように調整されている。雑音は各環境で 120 秒録音されており、最初の 60 秒をテストセットの作成に、後半 60 秒を学習セットの作成に用いた。

#### 3.2.2 実環境収録データ

実際に残響および加算性雑音が存在する環境において発声されたデータも収録されている。

実環境データは、マウスシミュレータを用いずに人間により発声された音声を、表 1 の条件で近接マイクおよび遠隔マイクの二つのマイクロフォンで同時に録音されている。このデータセットをセット D と呼ぶ。収録環境はセット C のものと同一であり、各環境 10 名 (男性 5 名, 女性 5 名) が連続数字を発話している。マイクと発話者の位置関係などの設定もセット A、B と同一である。図 2 にその設定を図示する。発声されたデータはテストデータ (49 ないし 50 発話) と学習データ (11 データ, 適応用) からなり、全部で 2536 発話 (2536 ファイル) となっている。

### 3.3 評価尺度

CENSREC-4 では利用者が任意の認識手法の性能を評価し、

ベースラインと比較できるような評価ツールを提供する。性能は Excel の表にまとめられるようになっている。原稿執筆時点でその表のフォーマットは正式に決定していないが、基本データセットに対するフォーマットは、おおよそ図 3 のようになる予定である。

CENSREC シリーズの特徴として、実環境における音声認識への影響要因の個別評価環境の提供がある。加算性雑音や残響など、様々な要因が重なりあって初めて実環境と呼び得ることはもちろんであるが、このような複雑な環境では、その評価環境を用いることにより、なにを解こうとするのかが不明確になってしまう恐れがある。また現時点での統合された環境に対処できる技術の開発および評価は時機尚早であり、利用者も多くは望めない。従って、セット C とセット D に対しては学習/認識/評価ツールは現時点で提供しない。時期が来たら検討するものとする。

## 4. 今後の展開

ここでは、これまで環境による影響と呼んできたものを音声(言語)信号への影響要因とし、再度、今後どのような評価基盤を構築するかを考える。

### 4.1 音声信号への影響要因

これまで、外的な要因による音声信号への影響は加算性雑音と乗算性歪みに集約されると考えてきた。しかし実際にはそのほかに、

- 加算性や乗算性とは異なる非線型な信号の変形(デジタル伝送におけるパケットロスなど)
- 外的な影響を受けた信号のフィードバックによる影響(ロンバード効果など)<sup>(注1)</sup>

が存在する。これまでも、例えば中村 [15] は、実環境における各種要因を、「音源」(ロンバード効果などに対応)「話者および雑音源からの音響パス」(残響などの伝達特性に対応)「雑音」(加算性雑音に対応)「回線特性」(線形/非線型型の伝達特性に対応)に分類しているように要因を分類している。これは、おおよそ

$$y = c(h \otimes l_n(x) + n)$$

と書ける。ここで  $c$ 、 $l_n$  はそれぞれ(デジタル)伝送路における非線型変換、特に加算性雑音に影響を受けた発話者による非線型な発声変換を表す。このうち、外的要因によりその変換が変化するため収録が容易でなく、個別の影響評価が難しいのがロンバード効果に代表される  $l_n$  である。

#### 4.1.1 ロンバード効果音声

ロンバード効果は、特に大きな加算性雑音下において発声するものであり、声が大きくなるほか、スペクトルにも影響を及ぼしているといわれている。しかし、大きな雑音下であるがために、その影響を独立して受けた音声の収録は困難で、実施された例は少ない。しかしそのいくつかの例によると、ロンバード効果は音声認識性能に影響を及ぼしていることが分かる。

いうまでもなく、収録には工夫が必要である。例えば小川ら [16] は、ダミーヘッドでバイノーラル録音された雑音をオープンエアヘッドホンを通して聞きながら発声することにより雑音下での発声をシミュレートし、ロンバード音声のみを収録することを試みている。その中でヘッドホンからの雑音の洩れや、自身の声の耳からのフィードバックなどを慎重に考慮している。また、その音声を用いた音声認識実験を通して、加算性雑音や残響の付加によるシミュレーションの妥当性を評価して、加算性雑音ではロンバード効果を考慮する場合としない場合では差が生じることも言及している。

この研究を含め、データのバリエーションも少なく、検討すべき点も数多く残されている。本 WG では、より充実したロン

バード音声 DB の作成に着手していこうと考えている。データが多くないことから多くは研究されていない分野でもあり、研究促進のためのデータ提供となろう。

#### 4.1.2 今後の収集に向けて

対象とする雑音が、加算性から乗算性、ロンバードとバリエーションが増してきた。これまで、特に加算性の雑音について、実在の雑音のカバレッジを向上させるという目的でデータの分類・可視化ができればよいと述べてきたが、より広くとらえることができれば、他の要因についても手当たり次第の収集を必要とせず、さらによい。

コーパスを可視的に表現して見る方法の提案があるが [17], [18], 加算性雑音の SNR やスペクトル、非正常性、残響の特性を表現する量と、それに基づく雑音環境の姿の記述、およびそれに基づく分類・可視化に取り組む。

### 4.2 雑音下音声認識のためのアプローチ

昨年度は、音声認識手法そのものとは異なるアプローチの評価基盤として、音声区間検出(VAD)の評価基盤である CENSREC-1-C の配布を開始した。その他に、フロントエンドの処理としてマルチマイクロフォンを用いた遅延和アレイをはじめとするマイクロフォンアレイ処理も多く研究されていることを述べたが、マイクロフォンの配置にはそれぞれの手法や研究者の意図によってさまざまな工夫が施されており、共通に利用可能な収録条件が設定できるかどうか難しい面があることも指摘した。これに関しては現在も議論しているが、このほど収録した CENSREC-4 では、マイクロフォンアレイによる収録も同時に行っている。この際には、比較的オーソドックスなリニアアレイを用いて多くの手法がそのデータによって実現可能であることを期待した設計とした。今後の議論により、なんらかの形で提供できるものと考えている。

さらに、映像情報と音声のハイブリッドによるバイモーダル手法は数多く提案されてきている。従って直接比較可能な評価基盤の整備は急務である。これに対しては、昨年度から CENSREC-1-AV, CENSREC-2-AV の作成を継続しており、近く公開できるものと考えている。

## 5. まとめ

本稿では、雑音下音声認識評価ワーキンググループのこれまでの標準評価基盤 CENSREC シリーズのそれぞれの位置付けを説明した。また、今年度新たに構築した残響下音声認識評価基盤 CENSREC-4 の概要について述べた。さらに、音声信号への影響要因と雑音下音声認識のためのアプローチの観点から今後の展開を述べ、それに基づく形でロンバード効果音声データベースとバイモーダル音声認識評価基盤が作成中であることを説明した。

これまで、雑音下音声認識への影響要因を分類して個別評価環境の提供をするという立場でいくつかの基盤を配布/作成してきているが、この方針が必ずしも最善であるとは考えていない。例えば残響に効果のある手法が加算性雑音の存在する環境では効果を発揮しないことは十分にありうる。また、その両者を統合的に扱う手法では、個々には加算性雑音、残響ともより

(注1): さらに、雑音下/残響下で明瞭性を得るためや、あるいは話し相手の反応などに応じて、用いる語彙や、言い回しを変えるなどの影響も考えられる。いずれの場合にも高度なユーザモデルを要するなど、一般的な信号処理・音声認識処理の範疇を越える処理が必要である。昨年度の報告において、対話処理やマルチモダリティによるアプリケーションレベルでの高精度化を対象外としたのと同様に、これらの影響は対象外とする。

すぐれた手法があったとしても、両者が存在する場合にはそれらを合わせるよりも効果的であるかもしれない。このようなことも考慮すると、統合的な評価環境も存在すべきであろうと考えている。

来年度は SLP 傘下のワーキンググループとしての活動最終年度を迎える。これまでリリースした、あるいは近日リリースする評価基盤を全体として捉えた場合に、一体何が明らかになるのか、また何が必要か、これまでのユーザ利用例などを参照しながら、包括的に捉え直していきたい。

## 文 献

- [1] D. Pearce, "Developing the ETSI AURORA advanced distributed speech recognition front-end & What next," Proc. Eurospeech 2001, 2001.
- [2] Aurora document no. AU/345/01, "Large vocabulary evaluation of front-ends- baseline recognition system description," Mississippi State University, Aug. 2001.
- [3] 中村 哲, 武田一哉, 黒岩眞吾, 山田武志, 北岡教英, 山本一公, 西浦敬信, 藤本雅清, 水町光徳, "SLP 雑音下音声認識評価ワーキンググループ活動報告," 情報処理学会研究報告, 2002-SLP-42-11, pp.65-70, July 2002.
- [4] 中村 哲, 武田一哉, 黒岩眞吾, 山田武志, 北岡教英, 山本一公, 西浦敬信, 藤本雅清, 水町光徳, "SLP 雑音下音声認識評価のための WG: 評価データ収集について," 情報処理学会研究報告, 2002-SLP-45-9, pp.51-56, Feb. 2003.
- [5] 中村 哲, 武田一哉, 黒岩眞吾, 北岡教英, 山田武志, 山本一公, 西浦敬信, 佐宗晃, 水町光徳, 宮島千代美, 藤本雅清, 遠藤俊樹. "実環境下音声認識の評価の標準化とその動向," 音声言語シンポジウム, 2004-SLP-54-20, pp. 139-144, Dec. 2004.
- [6] 中村哲, 武田一哉, 黒岩眞吾, 北岡教英, 山田武志, 山本一公, 西浦敬信, 佐宗晃, 水町光徳, 宮島千代美, 藤本雅清, 遠藤俊樹, 滝口哲也. "SLP 雑音下音声認識評価 WG 活動報告—評価用データと評価手法について—," 音声言語シンポジウム, 2005-SLP-54-26, Dec. 2005.
- [7] 北岡教英, 山田武志, 滝口哲也, 柘植寛, 山本一公, 宮島千代美, 西浦敬信, 中山雅人, 傳田遊亀, 藤本雅清, 田村哲嗣, 黒岩眞吾, 武田一哉, 中村哲. "雑音下音声認識評価ワーキンググループ活動報告: 認識に影響する要因の個別評価環境," 音声言語シンポジウム, 2006-SLP-65-1, Dec. 2006.
- [8] 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, 2005.
- [9] 藤本 雅清, 武田一哉, 中村 哲, "CENSREC-2: 実走行車内における連続数字音声データベースと評価環境の構築," 情報処理学会研究報告, SLP-60-3, pp. 13-18, 2006.
- [10] M. Fujimoto, K. Takeda, and S. Nakamura, "CENSREC-3: An Evaluation Framework for Japanese Speech Recognition in Real Driving-Car Environments," *IEICE Transactions on Information and Systems*, (accepted).
- [11] 北岡教英ら "CENSREC-1-C: 雑音下音声区間検出手法評価基盤の構築," 情報処理学会研究報告, 2006-SLP-63-1, pp. 1-6, 2006.
- [12] K.Takeda et al., "Construction and Evaluation a Large In-Car Speech Corpus," *IEICE Transactions on Information and Systems*, Vol.E88-D, No.3, Mar. 2005.
- [13] Y. Suzuki et al., "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, Vol. 97, No. 2, pp. 1119-1123, 1995.
- [14] 武田一哉 他 "走行状況別車内音声データベースとその予備評価," "音講論集, 3-P-10, pp. 185-186, Mar. 2002.
- [15] 中村 哲, "実環境に頑健な音声認識を目指して," 信学技報, SP2002-12, pp. 31-36, 2002.
- [16] 小川哲司, 倉持公社, 小林哲則, "シミュレーションに基づく騒音下音声認識システム評価におけるロンバード効果の影響の検証—複数の認識タスク, 騒音レベルに対する評価," 日本音響学会 2007 年秋季研究発表会, 1-P-29, Sep. 2007.

- [17] 庄境 誠. "複数音声コーパスの俯瞰的分析," 音声言語シンポジウム, 情報処理学会研究報告, 2005-SLP-59-23, 2005.
- [18] 山川仁子, 松井知子, 板橋秀一, "多次元尺度化構成法を用いた複数音声コーパスの可視化," 音講論集, 1-P-20, pp. 447-448, Sep. 2007.