音声言語情報処理 45－9
（2003．2．7）

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

中村 哲[1]，武田一哉[2]，黒岩眞吾[3]，山田武志[4]，
北岡教英[5]，山本一公[6]，西浦敬信[7]，藤本雅清[8]，水町光徳[1]

[1] ATR音声言語コミュニケーション研究所，[2] 名古屋大学，
[3] 徳島大学，[4] 筑波大学，[5] 豊橋技術科学大学，[6] 信州大学，[7] 和歌山大学，[8] 龍谷大学

あらまし 本稿では，2001年10月に音声言語情報処理研究会内に設立した雑音下音声認識の評価に関するワーキンググループの活動状況の報告を行う．このワーキンググループでは，雑音下音声認識の評価のための共通のコーパスの策定、および収録、その評価法の検討を進めている．現在までに行ったデータベース収集、評価系の構築について経過報告を行う．

キーワード 雑音下音声認識,AURORA

Progress Report of SLP Working Group for Noisy Speech Recognition

—— Data Collection ——

Satoshi Nakamura[1], Kazuya Takeda[2], Shingo Kuroiwa[3], Takeshi Yamada[4],
Norihide Kitaoka[5], Kazumasa Yamamoto[6], Takanobu Nishiura[7],
Masakiyo Fujimoto[8], Mitsunori Mizumachi[1]

[1] ATR Spoken Language Translation Research Labs., [2] Nagoya University,
[3]Tokushima University, [4]Tsukuba University, [5]Toyohashi University of Technology, [6]Shinshu University, [7]Wakayama University, [8]Ryukoku University

Abstract This paper reports current status of the SLP working group establised in October 2001 on the noisy speech recognition. The working group aims to develop standards, common corpus, and noisy speech recognition system in conjunction with Europian ETSI AURORA evaluation projects. In this paper, we report current status of data collection, standard backend recognition system, and other activities of the working group.

key words Noisy speech recognition, AURORA

1 はじめに

現在の音声認識は、雑音のない環境での協力的な音声に対してはかなりの性能改善が達成されている．しかし、実際の環境では雑音、残響の混入が避けれらず、認識性能の劣化が避けられない．このため、多くの研究がなされ、種々の改善法が提案されてきている．しかしながら、これらの研究は、異なるデータに対して、異なる基準で評価されたものが殆どで比較が非常に困難であった．この理由は、一つには音声認識のタスク自体の多様性、さらには実音響環境の多様性である．本報告では、2001年10月に音声言語情報処理研究会内に設立した雑音下音声認識の評

価に関するワーキンググループの活動状況の報告を行う．本プロジェクトでは、上記の問題を鑑み、評価のための共通の雑音データベース、雑音込み音声のデータベース、それらの評価法を検討することを目的としている．また、このワーキンググループは，特にEUROSPEECH2001における欧州のAURORAセッションに刺激を受けて発足した．この欧州のAURORAプロジェクト[1]というのは，携帯電話などのネットワークを利用した音声認識サービスのために，音声認識の前処理部を標準化しようという試みである．このネットワークで利用する音声認識は分散型音声認識(DSR:Distributed Speech Recognition)と呼ばれている．標準化においては，十分利用環境において高い性能を期待できる音声認識前処理が必要となる．この標準化は実際には，ETSI[2]のDSR標準化に加わっているグループ（主として企業）が具体的な要求基準，技術開発，評価を進めているが，それと平行して評価データをELRA(European Language Resources Association)から一般の研究者に配布し研究を啓蒙して，主にISCAのEUROSPEECH, ICSLPにおいて評価結果を発表するAURORAスペシャルセッションが開催されている．本ワーキンググループでは，このAURORAスペシャルセッションに連動して，日本において，並列に同様の雑音下音声認識の評価のためのデータ収集，評価法の検討，性能評価のための仕組みの検討，AURORAスペシャルセッションへのコミットメント，日本からの参加者のプロモーションなどを活動の趣旨としている．

2 WGの活動内容

本ワーキンググループ(WG)が発足してから、既に8回の会合を通して議論を進めてきた[7]．議論は大きく2つに分けられる．1つめは、騒音下の音声認識を本来どのように評価すべきかという課題、もう一つは欧州で進んでいるAURORAプロジェクトとの関係である．WGとしては、一つ目の課題を十分時間をかけ調査などをしながら進めつつ、内容がすでに確定し比較の容易なAURORA と同様のデータを収録していくこととした．AURORAでは、比較的容易なTI-DIGITデータに種々の雑音を人工的に加えて作成した評価データを用いて性能を比較するAURORA2[3]、自動車内の種々の騒音環境で実際に発話して収録した評価データを用いて性能を評価するAURORA3[4]があり、さらに、連続音声に雑音を加算して作成した評価データを用いて性能を比較するAURORA4[5]も計画されている．この評価プロジェクトでは、学習データ、評価データに加え、標準認識系（標準バックエンドと呼ぶ）を作成し評価を行うためのHTKスクリプトが配布される．また、さらに標準バックエンドにより得られる性能をベースにして、相対改善率、種々の条件の改善を平均して総合的な改善率を求めるExcel spread sheetが配布される．さらに詳細なAURORAのタスクのデータに関する解説は前回報告済みであるのでそちらを参照されたい[6]．本WGにおいても、AURORAと同様の方式をとり、学習、評価用データベース、学習用、評価用HTKスクリプト、改善率評価用Excel spread sheetを作成、配布することを目指している．以下、配布を計画しているデータベースについて述べる．

2.1 データベース

当面のデータとして、AURORAと同じタスクを日本語で収録し配布することとした．まず、AURORA2に対応するものとしてAURORA2J、次にAURORA2Jのサブセットの話者に同一の騒音を聴取させながら同一内容を発話させるAURORA2.5Jを収録する．これは、人工的に雑音加算することと実際の環境内の発話の差異を調べることを目的としている．さらに自動車内における種々の騒音環境での発話であるAURORA3Jを収録する．AURORA4Jについては、AURORA4と同様な読み上げ連続音声に雑音を加算したデータとするかどうかについては、現在議論を進めている．

2.1.1 AURORA2J

AURORA2Jの発話内容は、AURORA2で用いられている各連続数字をそのまま日本語に翻訳したものとした．話者数もAURORA2と同数の220話者を収録する予定であり、現在、名古屋大学で収録を進めている．学習8440発話、評価1001発話である．評価用のHTKスクリプト、集計用spread sheetも現在作成中である．音声データに重畳する雑音については、欧州のAURORAで用いられているものを入手し、全く同じデータを重畳する．

2.1.2 AURORA2.5J

AURORA2の内の少数の話者にAURORA2と同一の発話内容をヘッドホンで雑音を受聴させながら、発話してもらう．現在、収録方法について実験を進めている．テストセットのみを作成する予定であり、収録はAURORA2収録後を予定している．再生する雑音はAURORA2Jと同一の雑音である．

2.1.3 AURORA3J

AURORA3と同様に自動車内で発話した単語を収録するが、具体的な発話内容はAURORA3とは若干異なる．最終的な発話者は80名とする予定．現在、名古屋大学で収録、切り出しを進めている．発話内容は、適応用連続数字、評価用10桁連続数字、評価用単語、学習用音素バランス文である．騒音環境は、走行状態がアイドリング、市街地走行、高速走行の3種類、車両状態がエアコン(ON/OFF)、オーディオ、ハザード、窓開、通常の6種類で収録されている．

2.1.4 評価用バックエンド

評価用バックエンドについては、今年度、Microsoft Research から提供された改良版の20混合のHMMを基本に改良を加えて構築するが、今年度評価を行いながら確定する予定である．

2.1.5 データベースの配布

上記、AURORA2データベースについては、2003年4月頃に配布を開始する予定であり、逐次、AURORA3, AURORA2.5を配布していく．

2.2 共通ツール

現在のところSNR測定ツールの検討を進めている他、標準的音声切り出しモジュール、雑音下音声認識において標準的になりつつある雑音減算処理、その他の標準的な処理モジュールを作成し、公開していくことを計画している．また、上記評価データの集計なども検討していく．

2.3 耐騒音アルゴリズムの分類について

上記のデータを用い、評価用バックエンドをスクリプトを変更しながら、研究を進めていくことになるが、性能の比較をする際にどこまで変更しても良いのかが不明確であり、公平な比較が困難な場合が生じてくる．実際に、今年度のICSLP2002のspecial sessionでは、配布されたもの以外の学習データを用いたものや、バックエンドの混合数を増やしたもの、単語unitでないHMMを用いたものなど、多様なシステムが発表された．しかし、これらを同一の基準で評価するのは公平でない．そこで、WGでは開発されたシステムに応じて、下記の分類を提案することとした．

[クラス 0] 基本的に与えられているスクリプトを改変していないレベル．

[クラス 1] 学習されるHMMの構造がオリジナルと同じならば，学習過程では何をしても構わないレベル．例えば識別学習等を導入したければしても良い．(ただ、認識時の計算コストはオリジナルと全く同じ)

[クラス 2] HMMの構造が変わらなければ，認識過程において適応手法を導入しても良いレベル．話者適応や，雑音HMMが1状態1混合のHMM合成などが含まれる．(認識時のコスト増は適応化に対してのみ)

[クラス 3] Microsoft complex baselineのように混合数を増やしたり，状態数を変えたりする程度の変更(チューニング)なら許されるレベル．ただし，モデルはwhole wordモデルのままとする．HMMの構造を変えるようなHMM合成はここに含まれる．

[クラス 4] バックエンドにおける計算コストの指針があって，それに見合った計算量になるなら，何をしても良いレベル．例えば，モデルが複雑になっても，パラメータの次元数を減らして，トータルの計算量を制御できていれば良い等．

[クラス 5] 計算量は度外視し、何をやってもよく、最高の認識率を上げた者勝ちのレベル．

[規定外学習モード] 評価セットが同じでありさえすれば，違うデータベースのデータを追加してモデルを学習したとしても構わないレベル．

上記の内、規定外学習モードは、クラス1－5で混在し得る．性能評価データ記載時に上記のクラスを明記する必要があると考える．(当WGとしては欧州のAURORAにも同様の記載を求めるべく要求する予定である．)

3 音声認識実用化に関する調査(アンケート)

3.1 目的

従来，音声データベースの収集は，シーズベースで行っており，それらのデータを用いた評価がニーズに合致しておらず，商用の音声認識システムのカタログ等にそれらのデータを用いた認識率等が掲載される例は希であった．本WGでは，そのような事実に対する反省のもと，音声認識を利用した商品・サービス提供者および利用者が，本来の意味で参考にできる認識性能を提供するための標

準評価環境構築を目指している．このような環境を構築するためにはタスク設定や環境設定が現実に則しており，かつフェアであるという条件を満足させる必要がある．そこで，音声認識が利用されているサービスの現状，および将来期待されるタスクや利用環境を把握するためにアンケート調査を進めている．

3.2 調査項目

調査項目は大きく分けて以下の4点である．

- 回答者と音声認識の関わり (音声認識開発者，ベンダー，利用者等)
- 音声認識を用いたサービスに関して (現状，数年以内，将来)
- 本WGに期待するデータベースに関して (タスク・雑音・環境等)
- 本WGへの要望

ほとんどの部分を自由回答としており，アンケート作成者のバイアスがかからないよう配慮した．

3.3 現在までの結果

2003年1月現在，音声認識システムを実用化している7機関から回答を得ている．

現在実用化されているサービスの問題点 問題となる使用環境に関しては，実用化の場面ごとに異なりまとまりがなかったが，“利用者以外の音声に起因する雑音 (人ごみを含む)” や “突発性雑音” が問題となるケースが多いようである．話者に関しては，“高齢者・子供” を5機関があげていた．また，“小声・ぼそぼそ声” も3機関が指摘している．

数年以内／将来実用化したいサービス 家庭内での利用を数年以内と見るか将来と見るか等ばらつきがあったので，実用化の具体的時期は無視して集計した．その結果，“家庭内 (家電・情報検索)”，“自動車内”，“人ごみ” での利用を各々4機関以上があげていた．また，“自動字幕作成 (講演会場等での表示装置を含む)” を3機関があげていた．

データベース (発声環境・雑音の種類) “屋外 (5機関)”，“家庭内 (4)”，“自動車内 (3)”，“列車内 (2)”，“オフィス (2)” が複数回答あった環境である．

データベース (タスク) 「複数の認識装置を比べる時に，カタログにどのような内容が記述されていると良いか?」との観点から，どのようなタスクの音声データベースを収集すれば良いかを尋ねた．複数回答が得られたのは “姓名”，“数字” (各2機関) のみであった．注目すべき回答としては，どの程度の騒音レベル (家電のカタログには書いてある) で動くかを記述，という主旨の提案があった．一般の利用者にはSNRよりも騒音レベルが馴染みが深い．また，今後ロンバート効果等を視野に入れていく上では，同じSNRでも騒音レベルにより認識率は大きく異なることが予想される．十分検討したい提案である．

データベース (認識語彙数) 回答は，全体で 5,000 ～ 100,000 単語，同時認識単語数 200 ～ 50,000 単語とばらついた．バイフォンやトライフォンをカバーする語彙セットという意見もあった．

データベース (発声様式) 本項目は “離散単語，5単語程度の連続，連続，音節区切り，その他” からの選択とした．結果は “連続” 7，“5単語程度の連続” 4，“離散単語” 3であった．アンケート製作者は，誤認識語の言い直しや未知語入力に音節区切り発声が今後必須になると思っていたが，1機関が△をつけたのみであった．やはり，これからは “連続” のようである．

3.4 今後の予定

現在までの回答が7機関と，少ないにもかかわらず今後の方針を考えていく上で有用な情報・意見を収集できたように思う．今後，ベンダー系の企業や音声認識利用企業からも意見を頂く予定である．

4 Aurora Special Session in ICSLP2002

図1にICSLP2002において主催者のD.Pearce氏がまとめたグラフを示す。A2クリーンでもエラー率8ようにシステムの条件が曖昧であることに起因して、比較が困難な状況も生じ始めている。

5 今後の計画とまとめ

雑音下音声認識評価に関するSLP-WGの活動の内容と現状について報告を行った．音声認識の雑音環境に於ける頑健性の問題は，今日では非常に重要な課題である．WG設立以来，AURORAの評価プロジェクトと平行し

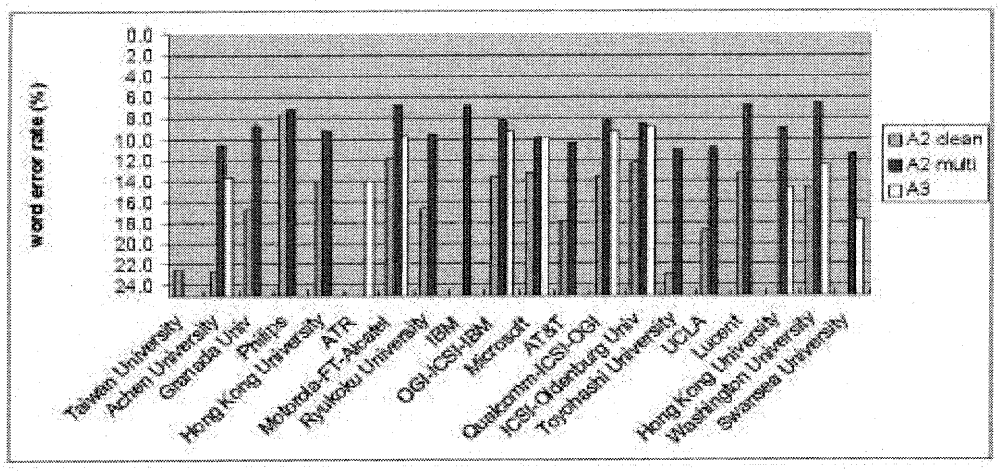

図 1: Performance comparison of Aurora2 evaluation by D.Pearce at ICSLP2002

て活動を進めてきた．今後，AURORA については，ICSLP,EUROSPEECH のセッションに参加し情報交換をするとともに，日本語データの収録が終了した際には，AURORA の評価 DB に日本語 DB を公開することなどを計画している．また，上述の日本語のデータ収集を継続するとともに，評価手法などの検討を進めていく．WG としては，本活動が雑音下音声認識に於ける評価の枠組みの確立の一助になればと考えている．

[謝辞] ご多忙中アンケートにご回答頂きました皆様に感謝いたします．また、本研究の一部は，通信・放送機構の研究委託により実施いたしました．

参考文献

[1] http://eurospeech2001.org/ese/ NoiseRobust/index.html,
http://www.elda.fr/proj/aurora1.html,
http://www.elda.fr/proj/aurora2.html

[2] ETSI standard document, "Speech processing, transmission and quality aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithm", ETSI ES 201 108 v1.1.2 (2000-04), 2000

[3] H.G.Hirsh, D.Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", ISCA ITRW ASR2000, september, 2000

[4] D.Pearce, "Developing the ETSI AURORA advanced distributed speech recognition front-end & What next", Proc. EUROSPEECH2001, 2001

[5] Aurora document no. AU/345/01, "Large vocabulary evaluation of front-ends- baseline recognition system description", Mississipppi State University, Aug 2001

[6] 中村、武田、黒岩、山田、北岡、山本、西浦、藤本、水町, "SLP 雑音下音声認識評価ワーキンググループ活動報告", 音声言語情報処理, SLP42-11, pp.65-69, 2002.

[7] http://www2.slt.atr.co.jp/ñakamura/ SLPWG2001/ Noise-WG.html