2005－SLP－58（6）－1
2005／10／21

カーナビ音声認識の商品開発

赤堀　一郎
(株)デンソー　基礎研究所
E-mail: akahori@jo1.denso.co.jp

カーナビ向けの音声認識の開発とその商品化に初期段階から携わってきた。その間に体験してきた研究と商品化のギャップについて紹介し、音声認識実用化のための課題の共有化を図りたい。

Development of Speech Recognition for Car Navigation System

The author has been involved in the development of speech recognition for car navigation system since its early stage till commercialization. The gap between research and commercialization experienced through the development is described to share the problems in putting speech recognition to practical use.

Ichiro AKAHORI
Research Laboratories, DENSO CORPORATION
E-mail: akahori@jo1.denso.co.jp

1. はじめに

カーナビに音声認識機能が搭載され始めたのは 1995 年頃であった。当初は少数のコマンドが認識できるだけであったが、その後急速に機能や性能が向上していき、現在では数百個のコマンド、全国の住所(2900 万ヶ所)、施設名(10 数万ヶ所)および電話番号などが認識できるようになってきている。性能についても、時速 100km の走行騒音下でほとんど性能の低下なく認識できるレベルに到達している。

筆者はこれまでカーナビ音声認識の商品化にその初期段階から一貫して携わってきたが、この間、様々な面で研究と商品化とのギャップを痛感してきた。このギャップの中には解決できたものもあるが、依然として未解決なものもある。本稿では性能評価に的を絞って紹介し、音声認識実用化のための課題の共有化を図りたい。

2. 性能評価におけるギャップ

1) 認識率評価

一般に論文等では認識率として認識語彙や話者についての平均値が用いられることが多い。平均値によって提案方式の優位性を客観的に示すことは可能である。

一方、商品として認識率を評価する場合は平均値だけでは不十分である。ナビに 100 種類のコマンドがあったとしよう。そのうち 99 個のコマンドの認識率が 100%であっても残り 1 個の認識率が 0%であれば、平均認識率は 99%という高い値ではあるが商品としては受け入れがたい。話者についても同様で、平均すれば高い認識率であっても、認識率が低い特定の話者にとっては、その商品

は満足できるものとはならない。

つまり、商品では平均値ではなく最悪値が問題となる。

コマンドに関しては全てのコマンドの認識率を評価することが可能であり、低認識率コマンドがなくなるように性能を向上(あるいは認識語彙を再設計)してから商品を発売することができる。一方、話者に関してはその商品の使用者すべての認識率を事前に評価することはできない。話者別認識率の分布を推測する手法[1]などを検討しているが、まだ課題が多い。別のアプローチとして、難認識音声のデータベースを充実する方向も考えられる。

住所などは、その語彙が非常に多く、全てに対して評価データを用意することが困難である。このような場合、事前に認識率が低い単語を特定する手法の開発が望まれる。このような方向の研究として[2]がある。

2) 耐ノイズ性評価

音声認識カーナビでは走行騒音に対する耐ノイズ性評価が重要となる。研究ベースではスタジオで収録した音声にホワイトノイズ等を重畳することで耐ノイズ性の評価を行うことが多い。

音声認識カーナビの耐ノイズ性評価も、スタジオで収録した音声に車室内の伝達特性を加味し走行騒音を重畳することで評価することが考えられる。 このような試みは多くなされているが、少なくとも我々の経験では、実際の性能との一致は不十分であった。

そのため非効率ではあるが、実際の車両で走行しながら収録した音声を使って評価している。認識エンジンの性能向上を確認する目的であれば、一度車両で音声データを収録しておけば、認識実験はオフラインで繰り返し行うことができる。しかし、車両、マイク、語彙(コマンド)などの諸条件が変わった場合は、実際の車両で収録し直す必要がある。

このような非効率な評価方法を取らざるを得ないのは、スタジオあるいは停止した自動車内で収録した音声には、ロンバート効果が含まれないためである。 ロンバート効果は認識率にかなり大きな影響を与え、これを無視しては正確な性能評価ができない。これまでロンバート効果を模擬することなどを試みてきたが満足する結果は得られておらず、たびたび車両による音声データ収録を繰り返しているのが現状である。

最近では CIAIR や CENREC-3 のように実走行車内で音声データを収録する例が増えてきており、実際の使用環境に即した研究の進展が期待できるようになってきた。

3. おわりに

性能評価での研究と商品化のギャップについて述べてきた。このギャップを埋める研究が進むことを期待している。

しかし認識率、耐ノイズ性などがどれだけ向上しても商品としては不十分である。 性能が向上することで「使える」ようにはなる。 だが「使いたくなる」レベルまで到達するにはまだ多くの課題が残されている。 標準的ユーザは、マニュアルは読まず、コマンドは覚えず、ちょっと試してうまくいかないと二度と使ってくれない。このようなユーザでも「使いたくなる」ためにはどうしたらいいであろうか。この点についても議論したい。

[1] 一ツ松孝文, 赤堀一郎: ″話者別認識率の分布推定法″, 日本音響学会講演論文集, Vol.I, 1-8-16, pp.37-38, 2004

[2] R.Terashima,H.Hoshino,T.Wakita: ″Prediction of Low Recognition Rate Words for Isolated Word Recognition System″, Proc. of Eurospeech 2001,pp. 2095-2098, 2001

2005－SLP－58（6）－2
2005／10／21

音声認識の実用化の現状と課題

NEC　渡辺隆夫

あらまし　話し言葉認識を対象とした大語彙連続音声認識の実用化への取り組みについて紹介するとともに今後の課題について述べる。

Toward Practical Speech Recognition

Takao Watanabe , NEC Corporation

Abstract　Application of continuously spoken large vocabulary spoken language recognition is presented.　Problems to be discussed are also proposed.

1. 実用化への取り組み

音声認識は、図１に示すように、コンシューマ領域からビジネス領域に至るまで、デジタルデバイド解消、バリアフリー実現やコスト低減などに貢献するものとしてさまざまな分野でその応用が期待されている。ＮＥＣでは、サーバから PDA までさまざまな環境で動作する、話し言葉認識向け大語彙連続音声認識システムを開発し、実用化に向け種々の応用システムの開発を進めている[1]。

- **旅行会話向け自動通訳 PDA**　日英双方向の旅行会話を自動通訳. システムは音声認識, 翻訳, 音声合成を統合して PDA 上で動作.
- **耐騒音音声入力ハンディターミナル**　製造、流通、物流、電力、建設、鉄道など現場でのデータ入力用。システムは２入力ノイズキャンセル機能つき音声認識・合成エンジンを統合して PDA 上で動作。
- **携帯電話マニュアルの音声検索システム**　外から操作マニュアルを簡単に検索・参照。電話音声認識サーバを持つシステムに電話をかけて声で携帯電話の使い方に関する質問を行う. システムは質問文の音声認識結果テキストを用いて操作マニュアルを検索し, 得られた検索結果候補を携帯電話画面に表示する。
- **AV コンテンツの検索システム**　蓄積された AV コンテンツのアーカイブを音声認識（不特定話者）して、認識結果テキストと対応する時間情報からなるアノテーション情報を付与する. キーワードを入力して認識結果テキストに対し検索を行う.
- **コンタクトセンタ向け音声認識ソリューション**　コンタクトセンタにおけるオペレータ通話音声を認識する. 図２に示すように、通話音声をテキスト化することによってオペレータ業務（ナレッジ検索キーワード入力, 応対記録作成など）や, スーパーバイザ業務（特定単語検出によるリアルタイム状況検知, モニタリング業務での通話内容確認など）を支援する. なお、本応用は、経営におけるコンタクトセンター運営の重要性に見られるようにニーズが明確であること、また、業務用途であるため使用条件を限定しやすいなどの点で、実用化する上で、コンシューマ応用と比較して有利といえる。

2. 実用化の課題

上述の応用システムの実用での評価はこれからであるが、これらのシステムを含めこれまでの種々の実用化の試みの経験をもとに、音声認識の実用化における課題をまとめる。

① **高い付加価値（ニーズ）があること**　モバイル・ユビキタス環境、入力する情報量が多いケース（地名、人名、複雑な操作のガイド、検索要求、通訳など）、会話から取り出した情報を有効に活用できる場合（議事録など）などは付加価値が高いケースと考えられる。付加価値を考える上で、GUI など他の手段との比較は重要である。特に、情報家電など GUI を持つ機器では、コマンド機能の単純な音声化ではなく操作ガイド・ヘルプなどの付加機能との統合が重要である。

② **想定した範囲での認識機能・性能**　類似語や類似文の識別や、話者や環境によらない性能の確保などである。これまでの研究開発においてすでに意識されている課題であるが、認識率の低い話者・単語の存在は、製品としての品質保証の観点からは重要な問題である。誤認識が多いと作業中断によるユーザの心理的負荷が増える点も考慮が必要である。

③ **システムのコストおよびシステムを開発する際のコスト**　開発コストには音響モデル・言語モデルの設計コスト、アプリケーション、ユーザイ

ンタフェース設計のコストがある。システムのコスト低減は研究開発の主要テーマであるが開発コストの重要性は必ずしも認識されていない。

④ **システムの想定外の問題**　認識できる範囲を表現することが容易でないこと、ユーザに使用条件をガイドすることが容易でないことに起因して、カジュアルな発話、対象外の音・声、想定外の音環境（マイク位置・条件、雑音など）などシステムの想定外の入力（システムと発話のギャップ）が問題となる。図３に示すように音声認識システムは応用にあわせてシステム自身を制御（適応）する機能をもつが、ギャップそのものは解消されない。この問題は、これまで、あまり技術的課題としてとらえられてこなかったが、こうしたギャップを解消するためには、たとえば、応用に依存しない汎用の知識を最大限利用するなど技術的なアプローチも必要と思われる（想定外のことが起きていることを汎用の音響モデル、言語モデル、対話モデルなどにより早期発見しユーザへ知らせる、安定して音を収める機構をつくるなど）。

参考文献

[1]　磯谷、畑崎、服部、奥村、渡辺：話し言葉認識に向けた基本技術と応用、情処研報、2005-NL(2005-9)（予定）

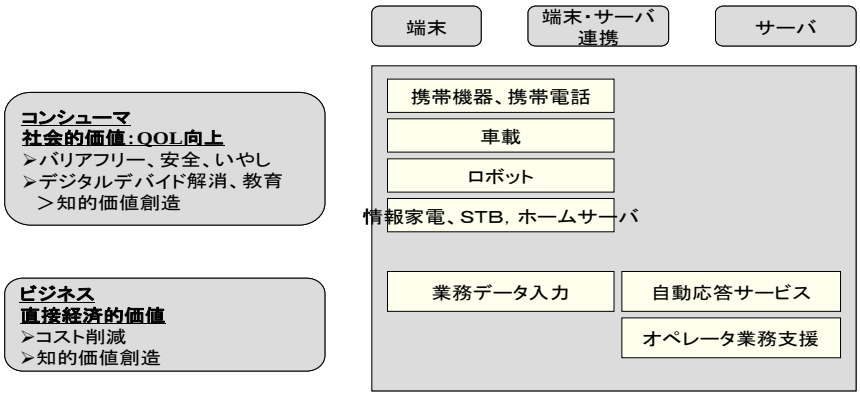

図１　音声認識の応用

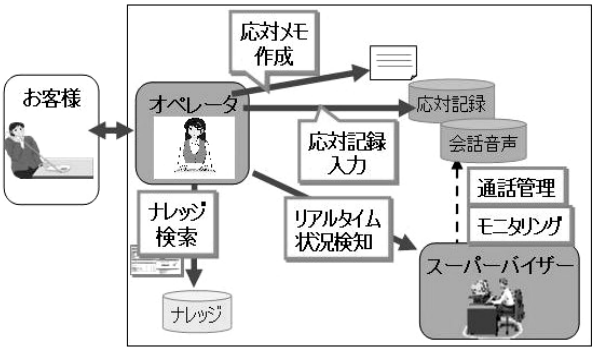

図２　コンタクトセンターにおける音声認識の応用

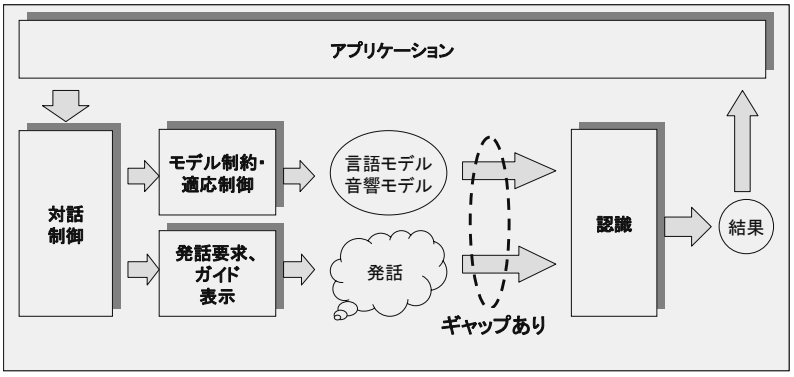

図３　音声認識システムの課題

2005－SLP－58（6）－3
2005／10／21

音声認識を利用した携帯電話サービスの開発

河井　恒†

†KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15
E-mail:　Hisashi.Kawai@kddilabs.jp

あらまし　携帯電話音声認識の技術上の主な問題点は、符号化歪み、背景雑音、伝送エラーであるが、サービスの有用性の観点からは、インターネットサービスとの連携も重要である。本稿では、この問題に対する取り組みとして、KDDI 研究所で開発した音声認識アプリケーション事例を紹介する。

Development of Speech Recognition Applications for Mobile Telephones

Hisashi KAWAI†

†KDDI R&D Laboratories Inc.　2-1-15 Ohara, Fujimino, Saitama, 356-8502 Japan
E-mail:　†Hisashi.Kawai @ kddilabs.jp

Abstract　Although major technical problems in automatic speech recognition (ASR) for mobile telephones are coding distortions, background noises, and transmission errors, integration of ASR and internet services is also important in view of practical usefulness. This paper describes some approaches for this problem developed at KDDI R&D labs.

1. 固定から携帯へ

KDDI 研究所では、長年にわたり電話網を対象とした音声認識技術の開発、音声認識応用サービスの開発を行ってきた。西暦 2000(平成 12 年)前後までは、サービスの対象は主に固定電話であり、携帯電話は付加的な扱いにすぎなかった。この間、内線番号案内システム、オペレーターアシストシステム、悪戯呼自動排除システム、ボイスダイヤリングなど様々な音声認識応用システムを開発してフィールド試験・試行サービスを行い、それらの中には実際に商用化され、現在も使い続けられているものもある。

固定電話の加入者数は、1997 年をピークに漸減し続けているのに対して、携帯電話の加入者数は 1995 年頃から急速に立ち上がり、2000 年には固定電話を逆転した(**図 1**)。この状況をふまえ、KDDI 研究所では、2000 年以降音声認識アプリケーションの主な対象を固定電話から携帯電話に移している。

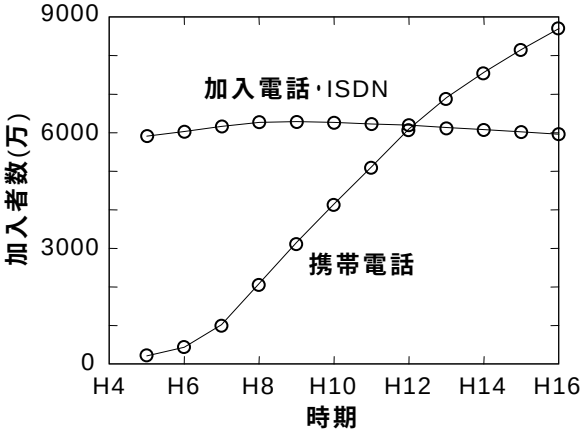

図 1. 固定電話/携帯電話加入者の推移。(総務省発表)

一方、携帯電話で電子メール、Web などインターネットサービスを利用するための IP 接続契約の比率は年々増加し、現在は 85%を超えている。現時点では、データ通信の ARPU(月間電気通信事業収入)は音声通信には及ばないが、携帯電話の重要な使用目的であることは間違いなく、携帯電話の音声認識サービスを開発する上でインターネットサービスは無視できない。

2. 携帯音声認識の問題点

携帯電話による音声認識では、固定電話と比較して次のような性能劣化要因がある。

・低ビットレート音声符号化による非線形歪
・背景雑音
・電波状態の変動による伝送エラー

携帯電話音声の SNR の分布は、固定電話と比較して低い側と高い側の両方に広がっているのが特徴である。低い側は、携帯電話が屋外で使われる機会が多いことによるもので、雑踏、人声、自動車などの雑音が主な原因である。筆者らの調査によると、約 3 割に発声で何らかの非定常雑音が混入している。一方、SNR の高い側は、

CODEC に前置される雑音抑圧機能の効果であるが、消し残った雑音は符号化歪みを受け、認識性能の劣化要因となる。伝送エラーは、約 2 割の通話で生じており、復号時には雑音となって認識率を低下させる。

一方、利用状況での特徴としては、

・人前で使うことが多い。

・歩きながら等キー操作困難な場合が多い。

などがあり、人前で機械に語りかけることへの羞恥心から使用がためらわれる可能性が高い反面、真の利便性向上につながる潜在的可能性もある。

さらに、利便性の観点からは、音声認識技術のみでサービスが完結する事例はまれであり、インターネットサービスとの連携が重要である。

3. 事例１：音声認識/Web の連携

KDDI 研究所では、第 1 世代の携帯電話音声認識システムとして、音声呼のみを使用して利用者の発声内容を認識し、その結果に応じた情報を音声として応答するシステムを開発し、フィールド試験を行った。

次に、第 2 世代のシステムとして、携帯電話 Web と音声認識を連携し、文字入力の代替手段として音声認識を利用可能にするシステム(**図 2**)を開発した。このシステムでは、利用者が Web ページ中のフィールドに対して音声入力を選択すると、ブラウザ機能を利用して音声認識装置へ発呼する(音声呼)。音声認識終了後、呼を切断し、端末から再度インターネットに接続し、利用者の操作により認識結果をコンテンツ Web に転送し、情報を表示する。このシステムを道案内のためのスポット・住所の認識、インターネット検索のためのキーワード認識、楽曲の演奏者名認識など様々な携帯 Web コンテンツに応用し、トライアルまたは商用提供を行った。

4. 事例２：PDA による分散型音声認証

第 2 世代のシステムでは、音声呼とデータ呼を切り替えるのに時間がかかるため、利用者の利便性向上に必ずしもつながらない面があった。そこで、音声信号の音響分析を端末で行い、結果をデータ接続によってサーバに転送する分散処理方式にもとづくシステムを試作した(**図 3**)。ただし、端末は携帯電話ではなく、汎用 OS を搭載した PDA、無線インターフェースは無線 LAN であり、タスクは話者認証である。

このシステムは、符号化による非線形歪み・伝送エラーの影響を受けないため、認識性能に関しては原理的に有利であるが、伝送エラーの少ない条件下では、音声呼方式に対して認識性能の決定的な差はない。むしろ、この方式の優位性は、音声呼/データ呼の切り替えが不要であるためインターネットサービスとの相性がよい点にある。

5. 展望：サーバ vs. 端末

KDDI 研究所では、これまでサーバ型に絞って音声認識システムの研究開発を行ってきた。サーバ型には、ヒット曲名のように認識内容が頻繁に変化するタスクで辞書の更新が容易なこと、強力な計算能力を利用できること、音声認識ソフトウェアの更新が容易であること、という利点がある一方で、通信が発生するという問題点もあるため、今後は携帯電話端末の CPU 能力の向上にともない、端末型との用途応じた棲み分けが進むものと思われる。

文　献

[1] 内藤他：携帯電話 Web コンテンツ向け音声入力システム、音講論、2002-10.

[2] 加藤他、“統合 PDA 端末の開発(6)～分散型音声認証システムの実装”、信学総大、B-15-16, p.717, 2005-1.

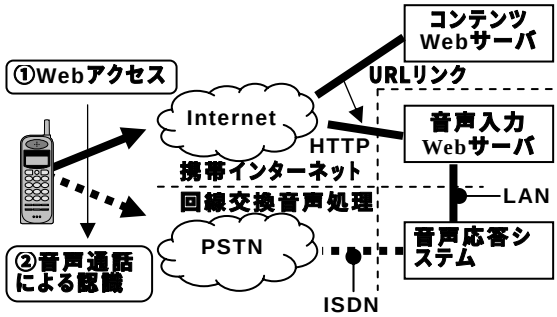

図 2. 音声認識/Web 連携システム

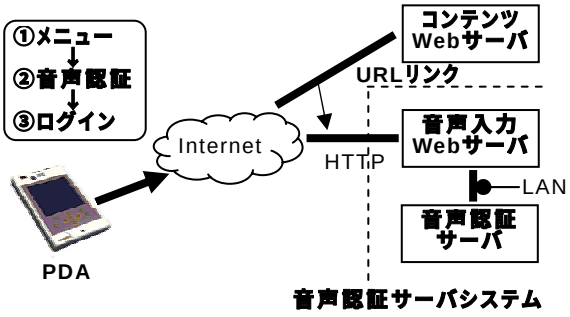

図 3 分散型音声認証システム

2005－SLP－58（6）－4
2005／10／21

コーパス，モデリング，ベンチマークのあるべき姿

旭化成株式会社　情報技術研究所
庄境　誠

アブストラクト

音声認識の実用化，音声認識利用の普及を促進する上で，解決すべき重要な技術課題は，今や，コーパス，モデリング，そして，ベンチマークである．本論では，組込型音声認識ベンダーの立場から，それらの重要性について言及するとともに，音響空間俯瞰技術がそれらの解決にどのように貢献するかについて論じる．

Ideal Methodologies on Corpus, Modeling and Benchmarking

Asahi Kasei Corporation, Information Technology Laboratory
Makoto Shozakai

Abstract

The technical problems to be solved at present are corpus, modeling and benchmark to accelerate commercialization and popularization of speech recognition. The importance of these problems is described from a point of view of embedded-type speech recognition vendor. Furthermore, how the technique to overlook multidimensional signal space contributes to solve those problems is discussed.

１．背景

現在の音声認識のアプリケーション領域の一つは，カーナビなどの情報家電のハンズフリー操作である．情報家電は，低コスト化，高信頼性，リアルタイム性などの要求から，いわゆる組み込み機器の形態で商品化されている．従って，音声認識に対しても，高認識率，低演算量，低メモリ量を要求する．

高認識率を提供する音声認識を実現するためには，入力信号の特徴を十二分に把握することが必要である．実環境での情報家電の音声による利用における入力信号の多様性をもたらす要因として，マイクロフォンの周波数歪みなどの入力・伝送系要因，利用者の声質や発話様式，方言などの利用者要因，雑音や残響などの環境要因，システムの設計語彙や利用者語彙などの語彙要因などがある．それぞれが独立に発生し，入力信号の多様性を非常に複雑にしている．

音声認識製品を提供する立場として，入力信号の多様性を知ることなく，高認識率の音声認識製品を提供することはできない．実環境の実製品の実利用の入力信号の多様性を，果たして我々は把握したと言えるのだろうか？多次元で暗闇の中で，見えざる敵（入力信号の多様性）と戦っているに過ぎないのではないだろうか？

２．音響空間俯瞰技術

入力信号の多様性を把握する技術として，筆者らは，音響空間俯瞰技術 COSMOS 法[1]-[5]の研究を進めている．この COSMOS 法は，様々な要因の組み合わせ毎にHMMなどの統計モデルを作成し，それらを二次元上平面に可視化して，音響空間地図を作成する統計的多次元尺度法と位置付けられる．ビジュアルデータマイニング技術とも呼ぶべきこの手法は，多次元空間上での入力信号の多様性の解析を大幅に簡素化することが分かってきた．多次元信号の二次元可視化には，副産物としてミクロな写像歪みが伴う．しかし，マクロに見れば，写像歪みの悪影響は十分に小さいため，COSMOS 法は有効であると判断している．

３．コーパス

既に収集された様々な音声コーパスから一枚の音響空間地図を作ることにより，それぞれの音声コーパスの網羅性，粗密分布，重複性を把握することが出来る．例えば，カーナビの音声認識機能は，自動車のドライバーが自動車空間の中でハンズフ

リー操作の目的で使用するわけだから，対象となる入力信号空間は，音響空間地図のある部分領域を占めるはずである．

1）部分領域を決定することが出来れば，その補空間に位置する音声コーパスは，音響モデルの学習に含める必要はないし，含めるとかえって性能が劣化するおそれもある．

2）部分領域のある一角が過度に密であれば，そこに密集している音声コーパスを全て使わなくても良いので，音響モデルの学習コストが削減される．

3）部分領域のある一角が過度に疎であれば，その一角に存在すべき音声コーパスをさらに収集する必要があると判断できる．疎な領域に既に位置する音声コーパスの特徴を調査し，それと同様の特徴を持つ音声コーパスの収集を行えばよい．

このように，音声コーパスの音響空間を俯瞰することは，音声コーパスの評価に極めて有用である．もし，ターゲットの部分領域に対して，過多の音声コーパスが不足する場合，音声コーパスの収集コストは無視できない．さらに，カーナビなどの情報家電は，世界に輸出される．従って，音声認識ベンダーに対する多言語対応の要求は極めて強い．音声認識ベンダーにとって，音声コーパス収集コストの削減は重要な課題であり，経営判断の中で大きなウエイトを占めるようになるだろう．

４．モデリング

情報家電に搭載されるプロセッサやメモリは，コストの観点からＰＣに比べて貧弱である．キラー技術として市場から認知されていない音声認識技術への割り当てはさらに些少である．従って，入力信号の多様性を網羅する大容量の音響モデルを情報家電に搭載することは現状では不可能である．一方で，情報家電は，不特定多数の利用者に使用されるというよりは，特定少数（一人または数人）の利用者に使用される．この点に着目すれば，利用者に適合した小容量の音響モデルを情報家電に搭載すれば良い．すなわち，情報家電では，入力・伝送系要因，利用者要因，環境要因，語彙要因の組み合わせにマッチした，音響モデルの要求が強くなる．各利用者が位置する音響空間地図上の位置を同定できれば，その位置に相当する音響モデルを提供すればよい．いずれ，衣服や眼鏡のように，利用者に合わせて，音響モデルを選んで購入し，日々使用する時代が到来すると筆者は信じる．カスタム音響モデルの安価なモデリング技術を有するベンダーが市場を獲得する日も近いだろう．

５．ベンチマーク

音声認識ベンダーは，音声認識技術のベンチマークに一体どのくらい時間とお金をかけてきたであろうか？従来は，手持ちの限定的な音声コーパスを利用して，ベンチマークを行い，認識率が，例えば９０％を越えることが確認できたので，音声認識技術を製品化してきた歴史がある．その結果，予期せぬ利用環境の利用者から，「認識しない」とそっぽをむかれることが多かった．市場に対し，信頼される音声認識製品を提供する責務を負う音声認識ベンダーは，今まで以上にベンチマークを優先することが求められる．そのことが，音声認識市場の拡大のために何より必要ではないだろうか？

一方で，ベンチマークにあまりに多くの時間とお金をかけすぎると，市場投入が遅れるし，採算性も危うくなる．今こそ，効率的なベンチマーク方法論を確立する必要がある．毎度毎度，新たに評価用音声データを収集しなければならないとしたら，それは音声認識市場拡大上の大きなボトルネックになる．ある情報家電に音声認識製品を提供する場合，その入力・伝送系要因，利用者要因，環境要因，語彙要因を考慮した，ベンチマークをどのように設計すればよいのだろうか？評価話者のセットはどのように選べば良いのだろうか？加法性雑音や周波数歪みをどう考慮すればよいのだろうか？それらの疑問に対する答は，既に収集された多数の音声コーパスの分析から得られるはずである．その分析を効率的に行う上で，音響空間俯瞰技術は，威力を発揮するに違いない．

今後は，既に収集された多数の音声コーパスからの音響空間地図の作成の研究を進める．そして，音響空間地図上の位置に依存した音響モデルライブラリから，利用者に適合した音響モデルを提供するモデリング技術の研究を継続する．その上で，ターゲットの要因を考慮した音響空間地図を利用した，効率的なベンチマーク方法論について研究する予定である．

６．提言

音声認識技術が，市場からの認知を得るために，コーパス，モデリング，ベンチマークの研究開発は今後ますます重要になる．この分野の研究を深耕し，産業界を支援していただければ幸いである．

７．参考文献

[1] Shozakai et al., ICSLP, 717-720, 2004.
[2] Nagino et al., ICSLP, 2965-2968, 2004.
[3] Nagino et al., ICASSP, 449-452, 2005.
[4] Shozakai et al., NSIP, 430-435, 2005.
[5] Shozakai et al., Eurospeech, 921-924, 2005.

2005－SLP－58（6）－5
2005／10／21

国家プロジェクト:音声認識技術の実用化

畑岡　信夫
（株）日立製作所　中央研究所
〒185-8601 東京都国分寺市東恋ヶ窪 1-280
e-mail: hataoka@crl.hitachi.co.jp

本稿では、真の音声技術の実用化を図ることを目標に、実用化に関する課題と、現在と今後の取り組みに関して纏めた。実用化への課題としては、ビジネスの観点からは、市場分野と応用製品投入、およびビジネスの仕方等、いわゆるビジネスモデルの明確化が必要であることが分かった。さらに、技術開発の観点からは、①音声・非音声の分離を主とする音響認識、②HMI(Human Machine Interface)の観点からの使い勝手の向上、③発話された単語、文を精度良く認識する音声認識、が重要となっている。現在、NEDO技術開発機構委託「音声認識技術実用化に向けた先導研究」を受けて、産学官連携に基づく、技術開発を主導する開発機関の設立を前提に、早稲田大学内で具体的な活動が開始されている。本稿では、本先導研究の事業概要と活動状況の報告、および今後の進め方に関して整理した。

Key Words:　音声認識、音声合成、音響処理/認識、音声対話処理/理解、HMI (Human Machine Interface)

Activities for Real Use of Speech Recognition Technologies

Nobuo Hataoka
Central Research Laboratory, Hitachi Ltd.
1-280 Higashi-koigakubo, Kokubunji, Tokyo 185-8601, JAPAN
e-mail: hataoka@crl.hitachi.co.jp

In this paper, the current problems of speech recognition and the future necessary R&D activities are summarized in order to pursue real use of speech recognition technologies. For the business aspect, the business models such as the market areas and how to put the real products into these areas are important and open questions. For the technical issues, the problems for the real use are, first, the discrimination between speech and non-speech clearly, second, the speech dialog understanding from HMI(Human Machine Interface) viewpoints, and third, speech recognition itself to recognize word and sentence utterances. We propose a neutral R&D organization which pursues R&D activities to overcome the technical problems for the real use of speech technologies by the collaborative consortium among companies, universities and governmental R&D institutes. In this paper, we summarize the activities on “pre-research activities for the real use of speech recognition technologies” supported by New Energy and Industrial Technology Development Organization (NEDO) in Japan.

Key Words: ASR (Automatic Speech Recognition), TTS (Text-to-Speech), Acoustic Processing/Recognition, Speech Dialog Processing/Understanding, HMI (Human Machine Interface)

1. はじめに

音声認識や音声合成の研究の歴史は長い。その結果、タスクや使用環境を限定すれば、現実に使用可能なレベルでの装置、システム、ソフトウェアが、製品化されている。音声認識等の処理も、現在では、マイクロプロセッサ(マイコン)でも実現でき、カーナビ端末や携帯電話、携帯端末機による新しいサービスが期待されている。

このように、音声処理技術、特に、音声認識技術の応用展開の夢は大きく、かつ音声処理技術は、HMI(Human Machine Interface)を実現する重要な技術となっているが、事業、ビジネスの観点からは、まだ大きな展開へとなっていない。すなわち、まだ「実用化」一歩手前という状況である。

2 音声認識技術実用化に向けた先導研究

2.1 事業概要

本事業では、音声認識技術を用いたHMIの性能を飛躍的に向上させ、情報機器の操作性を改善することを目的とする。このため、音声認識技術を多種多様な情報機器に幅広く使用できるように、組込み型実装の取り組みを目指して、音声認識技術に関する市場分析、音声認識技術の抱える課題、およびそれらの解決手段を見出し、今後の音声認識技術の研究開発の方向性を明確にするための先導研究を行う。

2.2 実施内容の抜粋(NEDO 技術開発機構 HP)

①市場分析:

音声認識技術の利用可能な市場の種類と市場規模を調査し、今後の成長性とその可能性を検討し、有望な市場の絞込を行う。

②技術課題の整理、順位付け:

音声認識技術を用いた各アプリケーションにおいて、それらの利用促進に必要な技術課題と、その課題解決の優先順位付けを行う。

③技術課題を解決するための方向性の提言:

優先順位の高い技術課題に対して、解決するために必要な研究開発項目と開発ステップを具体的に検討して提言にまとめ、今後の研究開発の方向性を明確にする。

④研究開発体制:

優先順位の高い技術課題を解決するための最適な研究開発体制案を提言する

⑤事業化に至る方向性の提言:

音声認識技術を用いた各アプリケーションを事業化するに当って、技術開発終了後の計画、方向性を具体的に検討して提言する。

⑥報告:

先導研究で取り組んだ検討内容を整理し、成果報告会を実施し、当該分野の有識者の見識やパブリックコメントを反映し、最終的に本先導研究の報告書をまとめる。

2.3 具体的な体制と活動

早稲田大学内に、「音声技術実用化研究所」を設置し、NEDO 技術開発機構の先導研究事業を実施している。プロジェクトリーダは、東工大古井教授、サブリーダは早大小林教授、主要研究員は、石川・藤井氏(三菱)、赤羽氏(ソニーCE)、森戸氏(沖電気)、渡辺氏(NEC)、金澤氏(東芝)、庄境氏(旭化成)、畑岡・大淵(日立)等である。

先導研究事業を実施するにあたり、①ビジネスモデル・研究開発戦略策定分科会(BS 分科会)、②音声認識技術予備評価分科会(技術分科会)、③標準化予備検討分科会(標準化分科会)を設置し、国内の音声関連研究者と有識者へも参加を依頼し、具体的な戦略策定にあたっている。

3 今後の進め方

国家プロジェクト提案による産業界コンソーシアム設立を行い、実用化技術の開発を推進する。さらに、その後に、各社、あるいはベンチャ会社が、開発した成果を基に、音声技術の事業化が図れるような体制の実現を狙う。

4 まとめ

NEDO 技術開発機構「音声認識技術の実用化に向けた先導研究」を紹介し、その事業内容と具体的な活動内容に関して纏めた。さらに、今後の進め方と展開に関して整理した。

謝辞: 本稿を纏めるにあたり、東工大古井教授、早大小林教授、三菱石川氏、ソニーCE 赤羽氏をはじめ、関係者の皆様に感謝します。本研究は、(独)新エネルギー・産業技術総合開発機構(NEDO 技術開発機構)の委託を受けて実施している。

参考文献

[1] 畑岡、他:音講論、1-8-10、2004 年 3 月
[2] 畑岡:情処理 SLP05-55、2005 年 2 月