社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－63 (9-1)
2006／10／20

音声認識の実用化の阻害要因と課題
－音声インタフェースのユーザビリティ評価－

三菱電機株式会社 情報技術総合研究所
石川 泰

アブストラクト

音声インタフェースは、長年の音声認識研究の成果により実用化がすすみつつある。しかし、その普及、応用分野の拡大は期待に対して十分なものとは言えない。本稿では、実用化を阻害している要因とその課題について、外観し、特にユーザインタフェースの設計、ユーザビリティ評価の視点での問題点と今後の課題を述べる。

Hindrance and Problem of Practical Application of Speech Recognition
- Assessment on Usability of Speech Interface -

Mitsubishi Electric Corporation, Information Technology R&D Center
Yasushi Ishikawa

Abstract

Speech recognition has been put to practical use, through the results of research activities. However, application areas are still limited and diffusion is not sufficient comparing its expectations. In this paper, hindrance and problems of practical application are surveyed, and the problem on user interface design and its assessment are discussed.

1. 背景

音声認識は、統計的手法などによる認識技術の進歩とマイクロプロセッサの高性能化により、1990 年代後半から、本格的な実用段階を迎えた。しかしながら、高い期待を受け続けているにもかかわらず、電話系のシステムは代替手段の発展もあり、応用システムは広まっておらず、また携帯電話のボイスダイアリングなど。普及せずにいるアプリケーションも存在する。この数年では、カーナビゲーション向けの音声インタフェースが、ハンズフリー、アイフリーの観点から注目され、音声認識応用製品の代表例となっているが、同様に、期待されるほどには、普及は進んでいない。音声認識技術の実用化を進め、その普及を図るには、その阻害要因を分析し、課題に対する適切な対策をとることが急務である。

2. 音声認識の実用化阻害要因

音声認識を用いたインタフェースの実用化がすすまない原因については、過去にも議論され、すぐれた分析もなされている。これを再度、簡単にまとめると、以下のような要因が考えられる。

1) 音声認識の基本性能に関する要因

実使用の音響環境における認識性能が不十分、性能の話者依存性により、極端に認識率が悪い話者がある、語彙外や、騒音に対するリジェクト性能が低いなどが問題となる。音響モデルと実環境音声のミスマッチによる問題である。

2) インタフェース設計に関する要因

「なんと言っていいかわからない」「予期せぬ動作をしたが、なにがいけないのかわからない」など、音声認識の利用方法、音声インタフェース設計技術が確立しておらず、十分にユーザビリティの高いインタフェースが実現できていない。インタフェース設計と、ユーザの期待とのミスマッチによる問題である。また、それをどう評価し、どのように改善すべきかについての知見、方法論が不十分である。

3) 開発コスト、体制にかかわる問題

音声のみのインタフェースとなる電話系システム以外では、基本的にマルチモーダルインタフェースとなるが、その実現には、現状では多大な開発コストが必要であり、応用製品の拡大や、設計から改良の効果的開発サイクルが実現できない。また、インタフェース設計者とエンジン開発者の間での問題の共有、解決方策の検討などの連携が必ずしも円滑にできない場合もある。

4) その他

ユーザ視点では、誤認識が起こると、さらに利用方法が理解できなくなるなどの悪循環が生じ、また、開発現場では、コストの問題などから、製品の投入とその利益による改良開発などへの投資のサイクルが形成できず、さらには、各社の独

自性の追及などにより、共通化(標準化)がなされないために、さらにユーザ視点での有効性が阻害されるなどの問題が生じている。

3. インタフェース設計とユーザビリティ評価

上述の問題はいずれも重要な問題であるが、現実のシステムの評価によれば、音声インタフェース設計上の問題、それにより生じる語彙外発話は、極めて深刻な状況を引き起こし、「音声認識は使えない」という意識をユーザに与えかねない問題である。以下に、インタフェース設計とユーザビリティ評価の観点での問題を詳細化し、今後の課題を検討する。

3.1 インタフェース設計上の問題

音声インタフェースの設計や評価の問題は、広く議論されている問題ではあるが、利用上に生じるユーザの期待とシステム設計のミスマッチをさらに分類すると、以下のようになる。

1) 音声コマンド

設計にあたっては、起こりえるユーザの発話を受理できるよう、類義語が定義されることが多いが、実際の利用場面では、定義されない発声が起こる場合が見られる。

さらには、音声コマンドから推測される機能がユーザと設計者の間で異なる場合もある。たとえば、「戻る」という音声コマンドが、どの状態まで戻るのかに誤解が生じると、「システムの状態」に対する誤解が生じ、以下の問題を引き起こす要因ともなる。

2) システムの状態と受理可能な音声コマンド

インタフェースの設計の原則として、すべての音声コマンドが基本的に常に受理可能で、1 つの音声コマンドにより、1 つのシステム動作が起こるような設計と、受理可能な音声コマンドがアプリケーションや音声対話の状態に依存し、複数ターンによる機能の実現を含むような設計とに大別できる。後者は、音声認識の対象語彙を限定し、高い認識性能が出せる可能性があること、さらには、適切なヘルプ画面や、GUI との一貫性などにより、語彙外発話が防げる可能性があることなどが期待できるが、一方では、現在、受理可能な音声コマンドがなにであるかの誤解による語彙外発話が生じる場合がある。特に、誤認識があった場合に顕著な問題となる。

3) システムの機能自体に対する誤解

GUI が基本的には、システムの機能をユーザに提示し、選択させるのに対し、音声認識では「ユーザの要求」を直接入力させるため、ユーザの想定するシステムの機能自体に誤解が生じる場合がある。特に複雑な機能を有するシステムでは、「なんと言えばいいのか」以前に「なにができるのか」が問題になる場合がある。

3.2 ユーザビリティ評価の問題

インタフェースデザインでは、ISO13407 などが規定され、開発のためのユーザビリティ評価と改良のサイクルが企業でも実践されつつある。たとえば、GUI や H/W による I/F 設計では、インスペクション評価や 5 名程度の被験者の発話法やプロトコル解析の方法により、インタフェースデザインの問題点の 70〜80%が抽出できるとされているが、音声インタフェースでは、確率的な現象である誤認識に対する問題点を考えれば、同様の手法を適用することは、十分な評価結果をもたらさない。さらには、評価では、前述の 3)の問題が見落とされ、タスクを与えての評価、すなわち、被験者に機能を既知とした評価だけが行われる場合も多い。また、インタフェース評価では、可学習性の観点から、被験者やインストラクションに大きく結果は依存する。

さらには、ユーザビリティ視点での評価結果からのみ音声インタフェースの改良案を作成すると、音響的な類似語による音声認識の困難性の視点が欠如する場合もある。

3.2 課題

上述の問題点に対して一般的な解を導くことは困難であるが、音声認識エンジン開発者と、インタフェース設計者、評価者の間で、上述のような問題点の整理と共有が行われ、設計指針の明確化と、とそれを考慮したユーザビリティ評価の手法の確立と普及、結果に対するエンジン開発者とインタフェース設計者のそれぞれの視点での分析と、改善案作成での協調体制が重要である。さらには、実使用環境での問題を抽出把握するためのフレームワークも必要である。技術課題としては、これらの実現を支援するために、以下が必要である。

1) 複雑なアプリケーションに対するインタフェース設計と評価を低コストで実現するための、ラピッドプロトタイプ構築環境
2) 比較的大規模な評価、あるいは、実使用時の問題点抽出を実現し、エンジン開発者やインタフェース開発者へ有効な情報を提供する、環境

これらを踏まえ、インタフェース設計の指針や、ガイドラインが内包された音声インタフェースの構築ツールの開発を進めるべきであろう。

4. まとめ

期待されながら、普及が期待ほど進んでいない音声インタフェースの実用化のため、問題点と課題をインタフェース設計、ユーザビリティ評価の観点からまとめた。これらの解決のためには、技術課題のみではなく、産業界における組織を超えた検討の場や、技術開発の場、さらには、協調による適切な情報の提供と標準化も重要である。

社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－63 (9-2)
2006／10／20

自動車用音声インタフェースへの期待

日産自動車株式会社 総合研究研究所 モビリティ研究所
神沼 充伸

アブストラクト

音声インタフェースは操作に対する視認時間を小さくできることから、自動車用のインタフェースにおいて、キラー技術の一つと言われている。然るに、音声インタフェースは商品的な視点、技術的な視点、安全面において、未だ問題を抱えている。本稿では、音声IFの現状と将来への期待について議論する。

Expectation to speech interface for the automobile

Mobility Laboratory, Nissan Research Center,
NISSAN MOTOR CO., LTD.
Atsunobu Kaminuma

Abstract

The speech interface is called one of the killer technologies in the interface for the automobile since total glance time can be reduced. However, the speech interface for the automotive still has some problems on a commodity aspect, a technical aspect, and the safety side. In this paper, present status and future expectations of the speech interface for the automobile are discussed.

1．はじめに

メーカーオプションによるカーナビゲーションシステム（以下カーナビ）の普及と共に、その多くに装着されている音声インタフェース（音声認識システム、音声合成によるガイダンスシステム、広義には携帯電話のハンドフリーシステムも含む。以下音声IF）の装着率も向上しつつある。しかしながら、音声IFそのものを購入する目的でメーカーオプションによるカーナビを選択するケースは多いとはいえず、また、ユーザーが音声IFにおける音声入力機能を率先して使用する頻度は装着率に比して低いとの指摘もある。本稿では、商品化の観点から、音声IFの車載化の効果を改めて考察し、更に、車載化を行う際に生じる問題点について議論する。

2．音声IFの車載化の効果

家電等のインタフェース設計と比較して、車載音声IF設計では、安全性に最大の注意が向けられる。例えば、運転中に操作可能な機能は、その機能を遂成するために必要な注意量、操作時間等を考慮した上で決定され、運転操作（安全性）に影響があると判断された機能については、運転中に操作ができないように設計されている。一方、顧客の視点では、すべての機能について必要なときに使用できることが理想と考えられる。これらの点から、「すべての機能を、安全に、運転中でも操作できる」ことは音声IF設計の目標の一つとして挙げられる。

運転中に操作ができない機能（例えばカーナビの住所入力）について、運転中に操作できるようにするためには、当該操作が、運転操作に対し影響を与えないことを客観的に示す必要がある。この評価指標として捜査情報の表示（例えばカーナビ画面）に対する視認時間や、手操作の運動時間等が用いられている。当該機能操作に関する視認時間には一定の目標値が存在するため、当該機能を実現するためには少なくとも視認時間や運動時間が目標値を満たしていることが要件となる。音声IFでは、従来、操作情報が表示された画面を視認し、スイッチを押すことによって成立していた操作が、画面を注視せず（アイズフリー）、手をステアリングから離さずに操作できる（ハンドフリー）点にある。この効果は、専門家以外（例えば経営層）に対しても容易に説明が可能であることから、音声IFの商品化に対する強力な後押しとなっている。

3．自動車用音声 IF の問題点

音声 IF が十分な商品価値を有することを示すためには、商品広告レベルで効果を語れる必要がある。現実には、使用者が画面を注視することなく、音声ガイダンスによる情報だけを頼りにタスクを達成することは人間の STM（短期記憶）のチャンク数や音声情報の時系列性からも困難であるし、音声区間切出しの困難さから、ユーザーに PTT スイッチを押させることなく音声認識を開始する構成の実現も容易ではない。本節では商品化のために必要なシステム性能と評価法について紹介する。

（1）システム性能

音声 IF（携帯電話を除く）は入力における誤り（誤認識）が問題となることは周知である。この要因は、車室内雑音、発話変形、入力語彙の莫大さ、発話誤りの多さ等に起因する。

車室内雑音）ベンダーが提供する音声認識システムの内部処理（SS、マルチパス音響モデル等）によって対策されているが、車種の違いから下限性能は必ずしも保証されず、開発工数浪費の要因となっている。筆者らは、音声認識システムおよび携帯電話のハンドフリーシステムの前段に共通の線形性を有する雑音除去システムを設置することで、認識性能の安定化、電話音質の保証および開発工数の低減を試みている[1, 2]。

発話変形）　走行時の発話では、特定のホルマントの変形が指摘されている[2]。筆者らの調査によれば、発話変形の効果が運転スキル、男女等の個人差によっても変動する現象も確認されており、発話変形の原因が走行雑音のみならず、走行時の緊張も一因である可能性が示唆されている。

入力語彙）　カーナビの目的地設定に対するタスク語彙の膨大さ（住所、施設名称等まで含めると 100 万語を下らないとも言われている）が問題となる。現状は、孤立単語認識と動的な辞書の切替えとの組み合わせや、ネットワーク階層を用いた辞書（連続単語認識）による認識率の向上が図られている。然るに、孤立単語認識を用いて階層構造を順番に（あるいは複数階層をランダムに）辿らせる手法では、操作時間の点で基準を超える場合があり走行中には使用できない。また、連続単語認識を用いた場合は、１発話の長さゆえに、使用者の発話誤りが増加し、使用者にとって心理的な負荷をかける恐れがある（「神奈川県横須賀市夏島町一番地」を初めて発話するとき、流暢に発話できる使用者は何%いるだろうか？）。顧客が必要としているのは、タクシーの運転手が理解できる程度の発話で入力できる音声 IF ではないだろうか？

発話誤り）　長い発話以外にも、ユーザーにとってインタフェースの操作法（音声コマンド）が分からなければその機能が存在しないことと等価である。近年では音声対話によって解決を試みている例を目にするが、音声対話は安全性の問題から多用できるとは限らない[5]。また、開始時に何を言えばよいのか分からないなどの問題もある。本課題に対しては、トークバック等を用い、自動的に操作法の音声ガイダンスを発行する手法が特許公開されているが、情報提示タイミング、デザイン、安全性の観点も含め、更なる改善が望まれる。

（2）評価法

音声インタフェースの商品化において音声認識率向上の必要性を語る場合、音声認識率が向上することによって得られる音声 IF に対する顧客満足が何かを事前に語る必要がある。音声認識率が十分ではないと思われている理由を説明した研究[4]や、音声認識率が使用者の感性に与える影響を論じた研究[5]はあまりにも少なく、応用メーカーの開発部署は手探りで仕様検討をしている状態である。また、タスク達成率、タスク達成時間に代表される認知的な要素に関する評価、操作中の注意量、ストレス度等を測定する生理評価、SD 法などを用いシステムの好み等を評価する感性評価等、一般的なインタフェース評価手法についての評価に関する研究も少ない。今後は、音声 IF に特化した評価法を創出した上で、評価法の正当性の評価（クリーン環境で収録した音声に雑音を重ねて作成した音声の認識率を求めることに意味があるのか？視認時間よりも注意時間ではないのか？）も必要であるし、物理、認知、感性等の各評価階層における評価結果が構造的に結び付けられるような評価構造を検討する必要があると考える。

4．将来への期待

Norman による行為遂行の７段階理論[6]では、インタフェースは外界と人間のメンタル領域をつなぐ架け橋となる役割を果たすと説明される。商品化されて間もない音声 IF に関し、「ユーザー」と「システム」をつなぐ架け橋となるような研究が、数多く創出されることに期待している。

5．参考文献

[1] 神沼、他, 音講論集, 1-1-12, pp819-822, 2005（秋）.

[2] Saitoh, et al., Eurospeech2005, 4CP-5, (2005)

[3] 竹山、他, 音講論集, 1-Q-18, pp369-370, 2006（春）

[4] 池田、他., HIS2006, 2314, pp613-616, (2006).

[5] 李、他, HIS2005, 2311, pp333-338, (2005)

[6] D.A.ノーマン, (株)新曜社,誰のためのデザイン？,1990

社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－63 (9-3)
2006／10／20

音声認識の実用化に対する大学の役割

中川聖一
豊橋技術科学大学

1．大学の研究

なんと言っても、大学に求められている研究は営利目的の企業ではやれない基礎研究であろう。特に工学部の研究は企業の開発研究・実用化研究と結びやすく、大学の役割を意識しておく必要がある。一方、大学の教員が研究をする必要があるのは、大学院生の教育のためでもある。大学院生は、研究を通して、新しい問題の発掘、未知の問題に対する解決、研究の調査・計画・実行・考察・まとめ・発表の訓練を受ける。この指導を担う大学教員は、自ら研究に邁進する必要がある。

2．音声認識研究の特質

音声認識の研究には、科学の側面と工学の側面がある。これは、音声や言語が人間に密接に関わっているためである。これは、電気や機械の工学の学問は自然原理を応用し、人間に役に立つことを目指しているが、その技術内容には人間とは直接関係がなく、科学的側面を持っていないのと対照的である。ここでは、音声認識の工学的側面について、大学の役割を考えることにする。

音声言語の認識処理には、信号処理、パターン認識、記号処理・知識処理のレベルがあり、これらの技術が有機的に統合化される必要がある。1970年代は音声理解という枠組みで記号処理・知識処理の重要性が指摘され、1980代はパターン認識レベルの技術が進歩し、1990年代になってディクテーションシステムとして結実し、一部実用化を達成した。また、この年代で、音声対話システム研究が隆盛になった。20000年代に入ってからは、信号処理の技術が重要であるという方向になっている。たとえば、今年度の秋の日本音響学会では、電気音響のセッションで音声認識をターゲットとした研究発表が多くあった。音声言語の認識・理解は、上述の3つのレベルの研究がいずれも脳機能の解明と同じく困難な課題のため、解決の糸口を見つけるために、他のレベルの未熟さを追求する方向に流れが移るのかもしれない。

3．最近の進展技術

・信号処理技術
マイクロフォンアレー処理による雑音・残響処理、移動音源追跡処理、音源分離の技術が進展した。その他、変調スペクトルやミッシングフィーチャ理論などが提案された。

・パターン認識技術
識別学習としての最小識別誤り学習（MCE）や最大相互情報量学習（MMI）に続き、音素認識誤り最小学習（MPE）やマージン最大化学習が認識率の向上に貢献している。また、ベイズ理論に基づく学習理論が発展し、点推定よりも分布推定の優位性の理論的整理がなされた。特徴ベクトルを非線形変換して高次元空間に写像して、効率よく（カーネルトリック）識別関数を作成したり次元圧縮する技術も開発された。また、複雑怪奇な音声現象を単一モデルでモデル化するには限界があり、Rover 法や MoE 法 (Mixture of Experts)が試みられている。これは複数識別器を順次構成し統合するバギングやブースティング法に通じる。

・時系列パターン認識技術
HMMを包含するモデルとしてグラフィカルモデル、ダイナミックベイジアンモデルが提案された。これらは、HMMを上回る能力を発揮できるかまだ結論付けられないが、回帰係数などの動的特徴を組み込んだ HMM で十分という意見もある。最大エントロピー法の時系列版である条件付確率場（CRF）は今後の利用が見込める。HMMのフレーム相関の記述能力に欠けるのを補う様々なセグメントモデルが提案されてきた。

4．データベースと基本ツール

音声認識のための大規模共通データベースの開発は技術の進展に寄与した。わが国でも、ASJ 連続音声認識データベース、JNAS 新聞読み上げ音声データベース (SLP/WG)、CSJ 話し言葉音声データベース、雑音環境下音声データベース (SLP/WG) などが開発された。音声検索評価用データベース作成等の目的に音声ドキュメント処理（SLP・WG）の活動も始まった。

HTK の HMM ツールキットや CMU 言語モデルツールキットも音声認識の発展に寄与した。わが国では、SLP/WG から発展した IPA 認識ツール（Julius）が多大な貢献をしている。このようなデータベースや基本ツールの開発は大学の貢献が大きい。

5．実用化のための基礎研究

・実環境下での音声認識やロボットとの対話処理には、遠隔発話（ハンズフリー）の音声認識技術が必要である。これは、雑音・残響の除去や音声区間検出などの古くからの問題を扱う。

・音響モデルや言語モデルの適応が有効であることは明らかになっており、これをオンライン・教師なしで行う技術を開発する必要であるが、

教師あり適応技術の開発と表裏一体である。
・対話システムでは、ユーザ発話が対システム（機械）発声であるか対人間発話であるかの識別、ドメイン外発話であるかの識別、タイミングなどの応答技術の開発が必要である。
・HMMによる音響モデルとトライグラムによる言語モデルの可能性と限界（人間と比べて、他のモデルと比べて）を明らかにすることは技術の健全な発展に必要である。
・共通データベースや基本ツールの他に、システムの評価技術（認識率やパープレキシティ。最近では、音声歪尺度と認識率の関係、音声要約の評価尺度やゴースト話者に対する最低保証基準）の開発が技術の進展に必要である。

6．大学の特長

大学の実用化研究への役割を考える上で、大学の研究上の特長を考えるとよい。
・大学の研究室では、学部の卒業研究、大学院生の修士論文や博士論文の作成と指導、と種々のレベルの研究と多くのテーマが設定可能である。教員は多種多様な研究テーマを設定できる特権と指導する能力が求められる。
・目先の実用化にとらわれないで、教員自身が興味あるテーマを設定することが可能であり、この自由度が大学教員の魅力であった。独立行政法人化後、この魅力がだんだん薄れ、実用化・外部資金獲得の風潮が強まっている。
・大学の教員はライフワークとして、何十年と継続的に研究を続け、長期的・経験的な視野から研究の動向や本質を捉えられる。しかし、この経験が固定観念となって自由な発想を妨げることがある。また、10年研究を続ければ、一流になれなければ駄目という話もあり、長期にわたる研究はアイデアの枯渇になることもありうる。

7．大学の役割と貢献

大学の研究者と企業の研究者に能力の差はないという前提に立てば、大学教員には、大学の特長を活かした役割と貢献があるはずである。
・新しい認識モデル・パラダイムの提案
音声認識の個々の問題はいずれも困難であり未解決のものが多いのも事実であるが、それを打開する手段として、新しい発想で認識モデルや研究パラダイムの提案が求められる。本質的でない研究や従来の改良・改善という細かい研究が多く見られる。最近の例をあげれば、無音声認識（NAM）やHMM合成の提案は斬新であり、新しいパラダイムを提供した好例である。
・種々の手法を体系化し・個々の手法を位置づけることは、研究の発展に必要であり、幅広い研究をおこなっている大学教員に求められている役割であろう。
・共通データベース・基本ツールの構築、評価技術の提案。
最近20年を振り返っても、欧米型の競争的研究の貢献は大きいが、その貢献を支えたのは、データベース、基本ツール、共通評価尺度の共有である。
・人材の養成
大学の最も基本的な役割・貢献は、研究を通した問題解発見・決・計画・実行・プレゼンの能力を備えた人材の養成である。大学で行った研究分野を就職してからも続ける学生は、博士課程の学生を除いてはむしろ稀である。その意味では、東海地区の音声関連研究生修士2年生の発表件数40件はあながち多すぎるとは言えない。

8．大学と企業の共同研究および国の政策

特に国立大学の教員は、国民の税金で給料や研究費の大部分が賄われていることを考えれば、特定の企業だけに利益をもたらす共同研究等は慎重でなければならないが、独立行政法人化後は、大学が戦略として企業等からの外部資金導入に力を入れるようになっている。以前と比べて、大学側も特許戦略を重要視するようになり、企業との共同研究の契約に支障をきたしている。たとえば、教員と相手企業が共同研究に合意しても、双方の知財関係部の調整がうまくいかず共同研究を断念した例がある。一方、企業は大学との共同研究には積極的だが、大学への奨学寄附金は以前より少なくなっているようである。

国は、国立大学の運営交付金を削減し、定常的な研究経費を減らし、競争的研究資金で配分し、富めるものが益々富める欧米型の政策をとっているが、これが日本人の特性と合致するか疑問であり（藤原正彦氏の「国家の品格」参照）、いずれボディブローのように我国の研究力を弱めることになると思う。筆者は、国立大学の独立行政法人化に反対し、新聞の意見広告に賛意を表明した経緯がある（朝日新聞平成15年4月23日朝刊）

9．まとめ

以上をまとめると、大学の音声認識の実用化への役割、実用化への貢献のためには、
・音声認識モデルの限界・上限を見極めた上での改善・改良研究、新しい認識手法・パラダイムの提案（上下限モデル例として人間を意識した研究）
・共通データベース・基本ツール・評価手法の整備、各研究の体系化・位置づけ
・大学のシーズと企業のニーズの相互刺激・マッチング型共同研究
・組織的開発研究と個別的基礎研究のバランス・融合を図り、競争的・集中型（アメリカ型）研究に対して、企業も含めた大学相互の競争的・協調的（日本型）研究
が望ましいと考える。

社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－63 (9-4)
2006／10／20

音声認識実用化における課題

磯健一†

† 株式会社アドバンスト・メディア
〒170-6048 東京都豊島区東池袋3-1-1 サンシャイン60 48階
E-mail: †iso@advanced-media.co.jp

あらまし 本稿では弊社ですすめている音声認識実用化の取り組みをいくつか紹介し，それらにおける課題を整理する．とくに顧客の投資に見合う価値を提供するためには，顧客の利用環境に合わせてシステムをチューニングするコストの低減（一部の自動化）や，経時変化によって新たに生じるミスマッチによる性能劣化を自動検出・通知するパフォーマンスモニタリングが重要であることを指摘する．
キーワード チューニング，ミスマッチ，パフォーマンスモニター，ロギング

Some issues on speech recognition system deployment to markets

Ken-ichi ISO†

† Advanced Media, Inc.
48F Sunshine 60 Building, 3-1-1 Higashi-Ikebukuro, Toshima-ku, Tokyo, 170-6048 Japan
E-mail: †iso@advanced-media.co.jp

Abstract This paper introduces our speech recognition system deployment to markets and summarizes some issues to be overcome to increse the customers' ROI (return of investment). First, the cost for system adjustment to user environments must be minimized. Secondly, accuracy degradation caused by environment change after deployment should be detected and notified to administrators.
Key words system adjustment, mismatch, performance monitoring, logging

1. まえがき

本稿では弊社ですすめている音声認識実用化の取り組みをいくつか紹介し，それらにおける課題を整理する [1].

2. 実用化の事例

2.1 医療文書作成支援

医療現場における文書作成支援システムとして，ディクテーションパッケージを製品化している．辞書，言語モデルを個別分野ごとに用意することにより認識精度を向上させている．

医療文書中には病名，症状名，部位名，薬品名などの専門用語が多用されており，それらの用語をキーボードと仮名漢字変換で正確に入力するのは容易ではなく，音声入力の有用性が認められている．

また放射線画像診断では画像から，病理診断では顕微鏡から目を離さずに診断レポートを入力できることも音声入力が選択される理由の一つになっている．調剤薬局では電子薬歴システムの導入がすすめられており，服薬指導結果の記録などの需要も増えてきている（保険点数の加算なども動機の一つである）．

一方で近年の個人情報保護の流れを受けて，辞書・言語モデル作成用に過去のカルテ情報の参照が難しくなりつつある点が，今後の懸念材料の一つである．また国内では医療関係者に口述筆記の習慣がない点も普及を阻む要因になっている．習熟したユーザは発話スタイルも協力的で高い認識精度で文書ドラフトを作成できるが，不慣れなユーザは初期バリアが高く，日常的な利用に至らないケースが生じている．

2.2 議会議事録作成支援

国や地方自治体には議会議事録の迅速な作成・公開と，その経費効率化が求められている．そこで議会録音音声を音声認識によりテキスト化し，その認識誤りだけを手作業で修正して議事録を作成することによって，テープの聞き起こしに比べて効率的に作業が行えるように支援するシステムを製品化している [2].

過去の議事録の電子テキストの入手は比較的容易であり，辞書・言語モデル作成の基礎データとすることができる．本会議などのフォーマルな議会では，発話も比較的ていねいで原稿を読み上げるのに近い発話スタイルになる．一方で委員会などでは話し言葉スタイルの発話が増える．書き言葉として記録された議事録テキストと，話し言葉スタイルのあいだのミスマッチを埋めることが認識精度改善に重要になる．

また社会情勢の変化に応じて議会で議論される話題が変化するため，辞書・言語モデルの迅速な更新が必要になる．維持管理コストを考慮すると，利用者サイドで辞書・言語モデルを適時更新できることが望まれている．

誤認識の修正ツールは，発話単位での繰り返し再生や話速変換再生，単語単位での候補選択などにより作業効率を向上させている．ユーザが日常的に使用している文書作成環境との連携，統合などがさらなる利便性向上には重要である．

2.3 コールセンター会話音声認識

コールセンターにおけるエージェントと顧客のあいだの会話音声を音声認識するシステムを製品化している．音声認識によって会話全体をテキスト化して検索やマイニングに利用するアプリケーションや，会話を常時モニタリングして関連する情報を随時提示したりアラーム警告するアプリケーションなどに用いられている．

これらの音声認識用の辞書・言語モデルは，同じコールセンターの録音音声を書き起こしたテキストコーパスから作成するのが最も効果的と考えられる．しかし書き起こしの人的コストの問題や，個人情報保護のためにコールセンター外部への録音データの持ち出しを制限する場合もあり，十分なコーパスを得るのが容易でないことがある．

一方，コールセンター会話の特徴として，エージェントは話者が既知であり，発話スタイルも比較的ていねいである，などのエージェントと顧客の非対称性を利用することも認識性能向上には有効である．

2.4 モバイル

PDA や携帯電話などの情報端末を用いた音声入力システムを製品化している．水産市場におけるセリ情報の入力や工場などの作業現場でのデータ入力などに利用されている [3]．また携帯電話を用いた情報検索（乗換案内，地図検索，など）や，営業レポート入力などのアプリケーションも提供している．携帯電話を入力デバイスとして使用して，サーバー側で音声認識する場合には，電波強度変化などによる音質の変化や劣化が問題になることが多い．携帯電話側で音声を特徴量に変換して，データ通信でサーバー側へ送信して音声認識を行う分散音声認識（Distributed Speech Recognition，DSR）は，電波強度変化の影響を受けにくく，サーバーとのネゴシエーション時間も短いため応答を速くすることができる，などのメリットがある．

3. 課　　題

周知のように特定の音響的，言語的環境において十分な実運用音声データを収集できれば，近年提案されているさまざまな技術を用いてシステムの認識精度を改善することができる．

しかし実際のビジネス現場においては，実運用音声データを顧客サイトの外へ持ち出すことが難しい場合があり（個人情報保護，機密保持，など），またデータの処理コストも問題になる．顧客サイトで辞書・言語モデルなどを随時自動で作成，更新できることなどが望まれている．あるいは安全にデータを持ち出すために，個人情報，機密事項などは判読不能であるが，チューニングに必要十分な情報は保持しているような形式の研究も興味深いテーマである．

またシステム導入の初期段階においてこれらの問題を克服してチューニングを行うことができた場合にも，顧客サイトでの運用環境は日々変化していく．そのため導入後しばらくするとミスマッチが増大して初期チューニングの効果が薄れて性能劣化してしまうことがある．そこでシステムの運用状況を常時モニタリングして性能劣化を検知することが重要になる．モニタリングすべき項目としては，音声入力レベル，背景雑音レベル，などの音響的な情報から，認識結果のスコアやリジェクション回数，システムの使用頻度，などのアプリケーションレベルの情報までさまざまである．電話音声認識サーバーのように，サーバーにデータが集中的に蓄積される場合にはモニタリングは比較的容易であるが，スタンドアロンシステムの場合にはログ情報の収集方法も課題になる．

4. おわりに

顧客が投資に見合った価値を得られるように音声認識システムを提供するためには，パフォーマンスモニタリングを徹底し，利用環境の変化による性能劣化を迅速に検出して通知する技術の開発が重要である．もちろん検出した変化に自動的に対応できる技術の開発も望まれている．

またそこで収集した実運用データを用いたチューニングの（半）自動化やコスト低減も重要な課題である．

文　　献

[1] http://www.advanced-media.co.jp/.

[2] 山崎恵喜，音声認識システムを活用した会議録作成：一北海道議会における実例一，情報管理，Vol.49, No.4, 2006, pp.165-173.

[3] http://jad.fujitsu.com/adver/produce/report/case_06/.

社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－63 (9-5)
2006／10／20

音声認識実用化における標準化協同作業の重要性

豊橋技術科学大学 大学院工学研究科
新田 恒雄

アブストラクト

利用に制約の多い音声認識という商品の市場を拡大する上で，今何が必要かを考察している。本文では，音声認識がコトバを相手にしていることを再認識すると共に，技術の高度化，エンジンと応用という両ベンダー間の距離の拡大，およびインターネットに向けた市場のシフトという観点から，共通仕様に向けた協同作業の欠如の現状を指摘し提言をまとめている。

Importance of Collaborative Works in Standardization for Developing Speech Recognition Systems

Graduate School of Engineering, Toyohashi University of Technology
Tsuneo Nitta

Abstract

To expand the market of voice-input products that do not satisfy easy-to-use interface yet, the emergent actions now are described. After discussing (a) ambiguity of spoken language, (b) complex technologies, (c) expanded distance between engine vendors and application developers, and (d) market shift toward Internet, the importance of collaborative works for standardization is emphasized.

1. 背景

音声認識の市場が小規模に留まっている理由に，技術自体が未成熟，キラーアプリがない，認識エンジンと応用システムの関係が疎であるなどが挙げられることが多い。しかし，「制約の多い商品であることを知らない人たちが話すコトバ」を，「コンピュータというどこの国とも分からないヒト？」に伝えるという「制約の多い商品」を扱っていることを認識すれば，自ずと商品開発に必要な他の側面が見えてくる。

音声認識の関連技術開発には，エンジン機能・性能はベンダー間の競争原理で進めるとして(差別化技術=勝てる要素を持たなければ敗者)，エンジンと応用システムに係わる基本仕様は，エンジン屋と応用システム屋が協同で共通仕様作成に取り組む必要がある。「制約の多い商品」が制約を徐々に外すことで，消費者に価値を認められる例は少なくない。ここで述べていることは，どちらも" 当たり前のことを当たり前に実行することが大切" という基本に立てば分かることである。しかし，勝つ要素もないのにエンジン開発を進める，あるいは利用者のことを考慮せずに異なる仕様の商品を出すといった転倒がまだまだ多く観られる。

こうしたことが起きる背景の一つには，情報システム開発が長年デファクトスタンダードで来たことがあると考えられる。しかしインターネット時代には，開発者から見えない利用者が，多様な端末で多様なサービスにアクセスする。同じサービスなのにアクセス方法が異なる，利用できるコトバが違っているではマーケットに受け入れられない（ここはカーナビも同じ)。

以下では，音声認識市場を形成するために何故標準化作業が大切なのかを四つの視点から説明した後，情報処理学会の情報規格調査会試行標準化委員会 WG4 の活動を中心に，これまでの作業と今後必要な作業を具体的に述べる。

2. だから標準化が大切

(A) 音声認識はコトバを相手にする： 辞書を開くと，そこには豊富な「語彙」が展開され，各項目にはこれまた豊富な「語意」が並んでいる。自語の持つ多義性を考えるなら，「語彙の制約問題」に協同で対処しなければならない。同時に，「異なる自語表現を共通の意味で括る方法」も協同で提供する時期にきている。

(B) エンジンも応用システムも技術内容が難しくなっている：

技術は確実に高度化する。音声もこれまでのように，一社で全ての技術問題に解を与える行きかたはいずれ失敗する。課題の対処を一つでも誤ると（二流以下の技術が一つでも入ると)，

商品価値は95%に落ちるのではなく0%になるからである。差別化のためのキー技術開発は必要であるが，その他の技術については標準仕様に沿って開発すればよい。ネット上の分散処理が前提となる音声処理では尚更である。

(C) エンジン屋と応用システム屋の距離が遠くなっている：

以前はエンジン開発者たちが同時に応用事業部・企業と組んで開発を行っていた。エンジンのモジュール性が高まるにつれ，この距離が次第に疎になって来ている。これは仕方のないことである。そこで，インタフェース仕様・利用方法や制約に関する仕様を標準化する重要性が大きくなる（べきと思うのだが，そうなっていますか？）。

(D) 主要な市場はインターネットに係わる：　ネット上でエンジンは仮想的なデバイスとなる。利用者はどこのエンジンを利用しているか見えないのである（本来は何語で話されるかも分からないと覚悟しなければ…）。少なくとも，エンジン間で同じサービスに利用される語彙は共通化する必要がある。一方，アプリケーションから見ると，(A)に述べた「異なる言語表現を共通の意味で括る方法」，すなわち共通タグセットの共通化と提供がなければ，煩雑な開発作業をしてまで，能力の劣るデバイスを利用したいとは思わないであろう。

3. 情報処理学会試行標準委員会WG4がやったこと[1]

音声言語インタフェース委員会(WG4)では，標準化目的と対象テーマを関連企業の協力を得て討議した後，以下の項目を標準化が急がれるテーマとして討議した。(1) ユーザが新語（未知語）を登録する際に使用する「読み」表記の標準化，(2) 音声関連製品の取扱い説明書などで使用する「用語」の標準化，(3) アプリケーションに特化した標準化，(4) 対話記述言語の標準化。

(1)と(2)は，これまで各社で独自に対応してきたが，委員会で調査し6社の比較表を作成したところ，全ての企業に共通する表記・表現が少ないことが明らかになった。特に，(1)はエンドユーザに混乱を招くため，早急な対応が必要とされ，最初の試行標準成果となった[2]。(2)については，討議結果をもとに産官学における，この分野の有識者へアンケート調査した後，試行標準案をまとめる予定である。次に(3)は，CTI，ディクテーション，カーナビ，PDAなどを対象に，操作コマンドや対象の呼称に対する「読み」の統一と，評価方法のガイドラインが検討された。このうち「読み」に関する標準化は，多くのアプリケーションで共通化が困難とされ，当面，標準化を見送ることとなった。理由は，コマンドの場合，同じアプリケーションでも製品により機能やその意味が異なること，あるいは既に長年使用されていることなどであった。また，対象の呼称では，検索したい目的地の読みや，楽曲の読みなどで問題が大きいことが共通に認識されたが，音声インタフェースを手掛ける企業とコンテンツを提供する企業が異なることが問題とされた。この課題は作業量が膨大となるが，音声対話技術の発展のためにも，第三者機関が本格的に取り組むことが望ましい。一方，ディクテーションは音声入力の基本的な応用であり，標準化の対象範囲を「ディクテーション中に入力する（キーボード上の）記号に対する読み」に絞って作業した結果，二番目の試行標準としてまとめることができた[2]。

次に，評価のガイドラインでは，文を入力するため評価方法が容易なディテーションや，古くから評価方法が検討されてきたCTIと比較し，カーナビの音声入力評価方法が問題とされた。カーナビでは使用状況が複雑多岐にわたる上，操作方法，コマンド名称も各社まちまちのため，統一的な性能評価が困難である。また，騒音下のハンズフリー大語彙連続音声認識といった非常に難しいタスクが対象となるため，環境の影響を大変受けやすく，正しい手順に従って評価することが特に重要である。カーナビ向け音声入力装置を開発する関係者の参加により，ガイドラインは三番目の試行標準として発行された[2]。

4. 標準化next？ [3]

2節で理由を説明したが，音声認識応用拡大向けて以下の標準化作業が急務である。

(a) 同じサービスに利用される語彙の共通化
(b) 共通タグセットの共通化
(c) モジュール化仕様とそれらの標準技術仕様
(d) エンジンと応用システム間のインタフェース仕様，および利用方法と制約内容に関する標準化

5. 提言

情報処理学会試行標準委員会は，上に述べたように標準化の必要な案件を迅速に処理するのに適した組織である。国内もしくは国際標準の場へ提案する前段階の案を討議することも推奨されている。今後，益々多くの技術者の参加を望んでいる。

参考文献

[1] 新田，松浦，西本，西村：音声言語インタフェースのための情報処理学会試行標準，Vol. 47, No. 7, pp. 762-767 (2006).
[2] 情報処理学会試行標準ホームページ
http://www.itscj.ipsj.or.jp/ipsj-ts/index.html
[3] 新田，甘粕，芦村，荒木，西本，桂田，石川：音声・マルチモーダル対話記述とその標準化，FIT2006 シンポジウム資料 (2006).