Vol.2009-SLP-78 No.6
2009/10/23

音声認識 実用化事例の紹介とその課題

藤田泰彦†

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取り上げる．特に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などの初期コストの低減，初期段階での認知，使うことの動機付け，慣れやわかりやすさなど，普及への初期のハードルが存在することを指摘し，弊社の取り組みを紹介する．

Practical Applications and Issues of a Speech Recognition System

Yasuhiko Fujita†

This paper introduces practical applications and issues of speech recognition system (ASR). Numerous factors exist that prevent the wide spread use of the ASR system, such as initial tuning cost and understanding of the system including awareness, motivation, and ease of use. This paper discusses our approach to these obstacles.

1. はじめに

弊社は音声認識・認証に特化した専門企業として，1997 年より音声ソリューションの実用化・普及に努めてきた．音声ソリューションは新しい市場が多く，実用化に際しては個々の課題が存在する．なかでもコストだけでなく，動機付け，認知，わかりやすさなども含め，導入時あるいは初期の課題が特に重要である．本稿では弊社で進めている音声認識実用化の一端を紹介し，それらにおける課題を整理する．

2. 実用化事例

2.1 医療文書作成支援

電子カルテの入力や調剤薬局での服薬指導の記録など，医療現場における音声入力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（***AmiVoice® Ex***）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．

弊社では既に 2,900 施設を越える医療関連機関への導入実績がある（2009 年 3 月末現在）．中でも放射線科 660 施設，電子カルテ向け 880 施設などが多く，医師の負担を軽減するためや，診療にあてる時間をより確保するために利用されている．また調剤薬局でも 1,050 施設で使われている．電子薬歴システムの導入がすすめられており，保険点数の加算などの動機付けもあることから，服薬指導の記録入力の需要が増えてきている．最近では眼科電子カルテ用や内視鏡レポート用[6]などにも展開している．

効果については，キーボード卓越者で 34%，キーボード初心者では 73%も入力時間が短縮されたケース[3]など，医療系学会や医療関係誌で多数報告されている．[4][5]

国内では医療関係者に口述筆記の習慣がない点が普及を阻む要因になっていたが，普及が進むにつれ，かなり認知されてきた感はある．習熟したユーザーは発話の仕方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，不慣れなユーザーは初期のハードルがあり，日常的な利用に至らない場合も生じている． 音声認識以外の部分も含め，わかりやすいアプリケーションにするとともに，マニュアルやヘルプデスクでのサポートなども継続的使用には重要である．

† 株式会社アドバンスト・メディア 取締役 技術部長
〒170--8630 東京都豊島区東池袋 3-1-4 サンシャインシティ文化会館 6 階
Advanced Media, Inc.

ⓒ2009 Information Processing Society of Japan

2.2 東京都議会議事録作成支援システム

弊社が開発した「議事録作成支援システム（***AmiVoice® Recorder／AmiVoice® Rewriter***）」は 2006 年 9 月より東京都議会様で総務委員会において運用開始され，本年５月より，全ての常任委員会を含む全 12 委員会で本格導入となった．これは、この約２年半の実績と経済性，及び弊社のノウハウ，セキュリティを高く評価頂いた結果であろうと考えている．

当システムは、現在、全国自治体 34 自治体で導入済みであり、民間企業への導入を含めると 55 ユーザーと国内でも最大の導入実績（当社調べ）を誇っているが、今回の東京都議会様 12 委員会での運用は、当社における議会向け「議事録作成支援システム」として過去最大規模となる．

本システムは，ネットワーク接続された音声録音用端末で録音された音声を，音声認識用サーバで認識処理を行い，最終的に作成された認識データを自動的にＦＴＰサーバへ送信する．上記端末およびサーバは，ネットワーク管理サーバにより一元管理され，そのサーバにより各端末の遠隔操作および音声認識ソフトウェアの動作状態を確認できる．ＦＴＰサーバは速記会社からアクセス可能で，速記会社は委員会開催後 30 分単位で認識データを取得して，修正・確認の作業が開始できるため，音声認識によりコストの短縮が図られたと同時に，納期短縮が実現された．

間違いのないテキストを要求されるケースでは，人的リソースとの融合が必要であり，速記会社と連携したビジネスモデルが構築できたことは，導入に向けて重要なポイントであった．

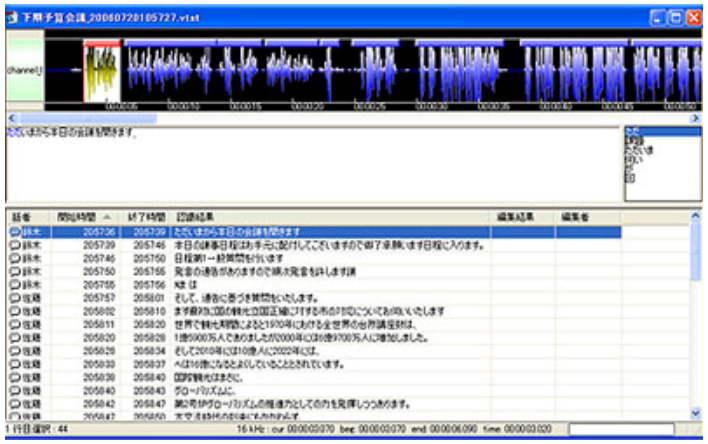

図１　確認修正用ソフト ***AmiVoice® Rewriter*** の画面イメージ

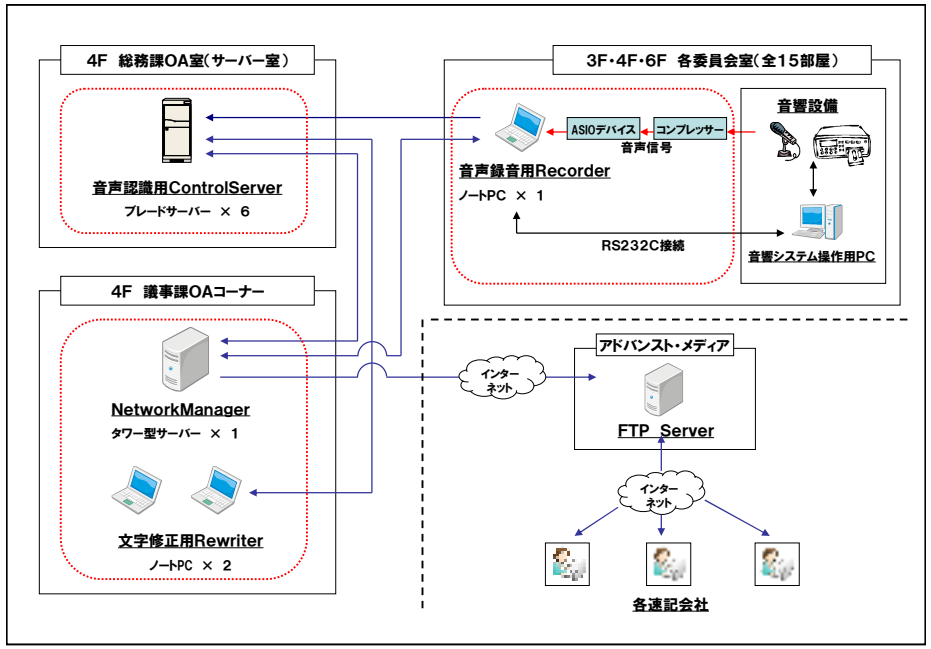

図２　東京都議会議事録作成支援システム全体構成

2.3 コールセンタでの通話認識の利用

コールセンタにおけるコミュニケータと顧客のあいだの会話音声を音声認識するシステム（***AmiVoice® Assist／AmiVoice® CallScriber***）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通話の分類，マイニングに利用する．または会話を常時モニタリングして，モニタリング効率の向上，関連する情報を随時提示，アラーム警告するなどを目的としたアプリケーションに用いられている．特に銀行・保険といった金融系コールセンタにおいて実績が多い．

これらの音声認識用の辞書・言語モデルは，同じコールセンタの録音音声を書き起

ⓒ2009 Information Processing Society of Japan

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．先の金融系コールセンタでは，協力会社とも協力し，ベースとなる言語モデルの整備が進み，初期開発の効率化を進めている．

一方，コールセンタ会話の特徴として，コミュニケータ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，などのコミュニケータ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 iPhone での音声認識メールサービス

モバイル分野においては，分散型音声認識（DSR: Distributed Speech Recognition）の技術を使い，（株）富士通様のらくらくホンⅣにおいて，乗り換え案内やナビ機能の音声入力として 2007 年 8 月に採用されたのにはじまり，2008 年 4 月には（株）ＮＴＴドコモ様の世界初の「音声入力メール」サービスに採用．iPhone においては音声ミニゲームや音声認識を使用した英語学習コンテンツをリリースする一方,「音声認識メール」については，2009 年 4 月からサービスを開始した．試行期間として無償でのダウンロードを開始し，7 月（約 3 ヶ月）には 13 万ダウンロードを突破し，有償化に移行した．

図３　iPhone 音声入力メールの画面イメージ

iPhone においては，マルチタッチボードにより，ある程度簡単に文字入力することができるものの，使いづらい点も多く，音声入力を活用することで，声で素早くより簡単にメールを作成できるようになった．

機能上の多くの工夫を行っているが，特に単語，ユーザー言語モデルの自動学習機能に力を入れた．たとえば，自動単語追加機能として，ユーザーが明示的な操作で単語を追加できるだけでなく，音声認識メールアプリ中で，入力した単語をサーバへ送信，既にシステム辞書に存在する場合は出現頻度を上げ，存在しない場合は新たに登録を行なう．また，ユーザー言語モデルの自動学習も実装した．

音声認識メールは，その利便性・音声認識性能の高さについて多くのユーザーから「実用的」･「認識率もかなり良い」と好評をいただいている．カスタマー評価では 72%の方から実用レベルに達していると支持いただいた．一方で,「使えない」･「認識しない」という否定的な声も少なくない．最初の印象がいいと，継続して使っていただくことにより，上記学習機能の効果も効いてくることや，ユーザー自身も使い方に慣れていただけることにより，結果として認識率が向上する　⇒　さらに便利に使っていただける，といういい循環になってくれるのだが，初めの印象が悪い，もしくは発話することに対する心理的抵抗が大きいと，即使えないと判断されてしまう傾向があるように思われる．

もちろんユーザーの好みやマルチタッチボードの習熟度による選択はあると思うが，本製品は使える便利なツールという域には達していると考えている．課題としては，価格や更なる精度向上はもちろんであるが，初めて使っていただくユーザーに，便利と思っていただけるまで，どうしたら使い続けてもらえるか？という点も大きいと考えている．

2.5 e-learning 導入事例（進研ゼミ中学講座＋ｉ）

音声認識の利用分野として，エンターテインメントや教育分野においても力を入れてきた．特に英語教育においてはＣＡＬＬ（Computer Assisted Language Learning）として近年コンピュータの利用が進んでいる．従来英語教育においては，米国 Native 用の音声認識エンジンをベースに利用される場合が多いが，中上級者ならともかく，初心者/初級者は Native 用エンジンでは認識しづらい．弊社は日本人英語用に特化した認識エンジンと，日本人が誤りやすい発音間違いを判定する２つのエンジンを開発することにより，さまざまな英語教材を開発している．中でも発音評定を核とした学校向けの英語教育ソフト ***AmiVoice® CALL –pronunciation-*** は OEM 提供分を含めて，既に約 200 施設で利用されている．

（株）ベネッセコーポレーション様の進研ゼミ中学講座＋ｉ（以下プラスアイ）は 2008 年 4 月より中学１年生向けにスタートし，今年は中学２年生向けにもサービスされている．プラスアイでは，クイズ形式のコンテンツやビデオドラマの中に出てくる

ⓒ2009 Information Processing Society of Japan

フレーズを，登場人物になりきってアフレコするなど，音声認識を利用することにより，従来なかった新しいコンテンツを実現している．英語を話すコンテンツが作れ，かつ発音評定できることで，「英語に自信がついた」，「発音が良くなった」，「楽しく勉強できる」などの評価をいただいている．

音響的には中学生英語に重点を置いた音響モデルを構築したが，親が試す場合なども想定されるので，完全な中学生用とはしていない．

それ以上に労力がかかった問題として，次の２点がある．ひとつはコンピュータリテラシーの高くないユーザーにも使ってもらえる工夫．もうひとつはコンピュータがユーザー所有の物なので，機種/ＯＳバージョン/インストール済みの他ソフトなどの違いに対応する必要があることである．幸いマイクは統一したものを配布できた．一般的な市販ソフトの場合も同様の問題はあるので，紙面の都合上対応策ひとつひとつを挙げることは省略する．特に音声認識に関係する注意点として，インストール時の対応と認識失敗時のコメントをあげたい．インストール時の対応としては，マニュアルだけでなく，インストールガイドＣＤを配布し，インストールがわかりやすくなっているほか，既にインストールされているソフトのバージョンやＯＳの設定も自動でチェックしている．またマイクの使い方などもていねいに説明している．一方認識失敗時は，その失敗理由がある程度予測できるものについては，できるだけ失敗理由を都度表示するようにしている．たとえば「もう少し大きい声で話してください」とか，「間違えないように最後まで話してください」とかである．

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のように，サーバ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．

コンシューマソリューションにおいては，声を出すという心理的抵抗を和らげることや，正しい使い方の認知，サポートなど使い続けてもらえる環境を作ることも重要である．

4. おわりに

音声認識は確実に実用化の段階に来ている．少なくとも「あれば便利」，「安ければ（タダなら）使う」という段階を越え，ビジネスとして成立するソリューションも増えてきている．しかし初期コストのハードルをいかに低くするかは，普及への大きな要素であると考える．さらに認識率を含めた使い勝手の面でも初期のハードルを低くする努力が重要である．そのためにはビジネスモデルがいかに作れるか，という課題も避けては通れない．

参考文献

1) （株）アドバンスト・メディア http://www.advanced-media.co.jp/.
2) 磯健一，音声認識実用化における課題，情報処理学会 研究報告，2006-SLP-63(9-4)，pp51-52.
3) 高原太郎，中嶋美佳，似鳥俊明，蜂谷順一，連続音声認識による MRI レポートの試み，日本磁気共鳴医学会雑誌 21(supp.)；155
4) 放射線科の LifeHack，Rad Fan 2009 年 8 月号（メディカルアイ），pp.41-.
5) 電子カルテの入力は簡単になったか，新医療　2008 年 7 月号（産業科学株式会社），pp.62-.
6) 内視鏡データベース活用の時代，消化器内視鏡　2009 年 7 月号（東京医学社），pp.1089-.
7) 山崎恵喜，音声認識システムを活用した会議録作成 ：―北海道議会における実例―，情報管理，Vol.49, No.4, 2006, pp.165-173.
8) （株）ＮＴＴドコモ 音声入力メール http://www.nttdocomo.co.jp/service/safety/voiceinput_mail/

ⓒ2009 Information Processing Society of Japan