

音声認識ビジネスへの取り組み

ラインタイトル

NECでは、自社開発の音声認識エンジンを継続的に強化して、さまざまな分野での適用製品を提供してきています。特に近年は、話し言葉音声の認識と、コンパクトな音声認識の研究開発に注力し、特に業務用途への適用を中心にビジネスへの展開を進めています。

本稿では、音声認識技術の事業化の進展について音声認識を活用したソリューション、製品開発の取り組み状況、今後のクラウドサービス時代に向けた技術開発、実験的な取り組みを紹介し、本特集の論文の位置付けをご説明します。

市場開発推進本部
音声認識事業化推進センター長
池田 勇吉

1 製品化とビジネスへの取り組み

日本における音声認識技術の研究開発は、1960年に京都大学様とNECの共同研究によりスタートしたのが最初です。以来、NECでは、自社開発の音声認識エンジンを継続的に強化して、さまざまな分野での適用製品を提供してきています。その間、機械に対して単語を認識させたり、文章を読み上げてテキスト入力する書き言葉音声の認識ソフトウェアを開発し、次々と製品化してきました。車載機器向けや携帯電話向けに組み込むための半導体・ミドルウェア、機械が人との対話を支援するためのロボットプラットフォーム、音声による文字入力や通訳機能を提供するパーソナルソフトウェア製品、そして業務用途では音声応答システムにおける単語音声認識など、その適用領域を広げてきました（図1）。

そして近年は、話し言葉音声の認識と、コンパクトな音声認識の研究開発に注力し、特に業務用途への適用を中心にビジネスへの展開を進めています。前者は、コンタクトセンターや会議の議事録作成の重要な業務に、後者はハンズフリーが求められる現場業務の支援や自動通訳によるコミュニケーション支援の

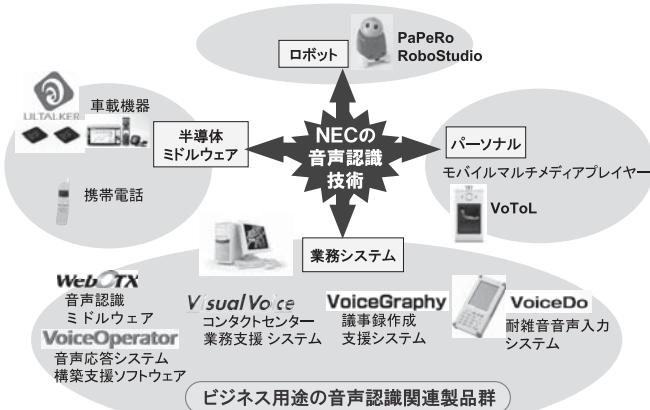


図1 これまでの音声認識関連製品

業務に適した製品・技術の開発を進めており、これらの活用事例・実績も増えつつあります。

1.1 話し言葉認識を活用したビジネス領域

2005年にコンタクトセンターのオペレータが話す言葉を認識できる製品CSVIEW/VisualVoiceの提供を開始しました。これ

は、オペレーター一人ひとりがお客様の応対を短時間で的確に行ない、業務効率を高めるために利用いただく製品です。本製品の導入により、通話音声の内容をリアルタイムにテキスト化して、発話内容を目で確認するほか、会話に出てきたキーワードを使ってFAQ情報を検索したり、応対後の業務データ入力を音声入力を使って時間短縮するなど、複雑化する業務システムの操作を軽減しながら、運用コストの削減を実現できます。

その後、2007年にオペレータの音声に加えて顧客側の音声も認識して対話内容をフルテキストで把握できる強化版を国内で初めて提供、更に2009年に発表した最新版では、音声認識を利用する専門辞書をお客様自身がメンテナンスできるようになり、運用にかかるシステム投資コストを低減でき、導入しやすい製品へと強化が図られています。

また、同エンジンを利用して、会議で話した通りの会話をテキスト化して議事録作成に役立てるためのソフトウェア製品VoiceGraphyを2007年から提供しています。

これは、全国の自治体様の議会において、議事録を作成・公開するために導入されており、作成から公開までの所要時間が、音を何度も聞きながら書き起こす従来の作業に比べて、大幅に短縮できるなどの効果が出ています。

その後も、議会だけでなく、役員会議などで法的に議事録を残す必要がある一般企業様にも、導入が広がってきています。また最近では、本ソフトウェアをSaaS型サービスの形で提供を開始しました。これは、企業様が社外に向けて発信する情報をより早く充実した形で提供するため、広報・IRの会見、決算説明会の議事録作成を中心に利用いただいている（図2）。

最新の事例としては、2009年5月から始まった裁判員制度でも弊社の話し言葉認識を利用いただいている。全国の裁判員

裁判の法廷に導入されており、評議の場において証言などの中から確認したい映像を検索するために利用されています。このシステムには、不特定の話者が複数人で同時に発話することがあっても高い認識精度を実現する技術が導入されています。

1.2 コンパクトな音声認識を活用したビジネス領域

音声認識の実用化開発で、もう1つの大事なポイントは、小型情報端末のようなマシンリソースに限りがある環境でも動作する、コンパクトな音声認識ソフトウェアの実現です。

工場などの現場においては、キーボードから情報を入力することが難しい業務シーンがあり、小型情報端末などを使ったハンズフリーでの音声入力が求められています。この場合、騒音のある環境下での用途が多いことも特徴的なニーズとして挙げられます。PDAタイプの業務端末Pocket@iEXに組み込んで利用できる耐騒音音声認識装置VoiceDoは、2つのマイクを使った独自の耐騒音処理方式により、地下鉄の車両内に相当する騒音下でも高い認識性能を実現する製品になっています。

これまでVoiceDoは、自動車製造などを含む製造業全般（製造工程、検査工程）、流通業、電力、鉄道、船舶、住宅、医療などの幅広い分野で応用されています。最近では、鮮魚市場、酒類卸売販売業に導入されるなど適用領域が拡大しています。

また、音声認識に翻訳機能などを組み合わせて実現する自動通訳も、携帯型の端末への組込み利用が求められる技術です。

NECでは、1983年から自動通訳技術の研究開発を続けており、旅行会話の日英自動通訳技術について、2002年にPDAタイプの試作品を発表、2006年に発売したモバイルマルチメディアプレーヤー製品に機能搭載、2009年には、携帯電話端末でも動作可能な自動通訳ソフトウェア技術のコンパクト化を実現してい

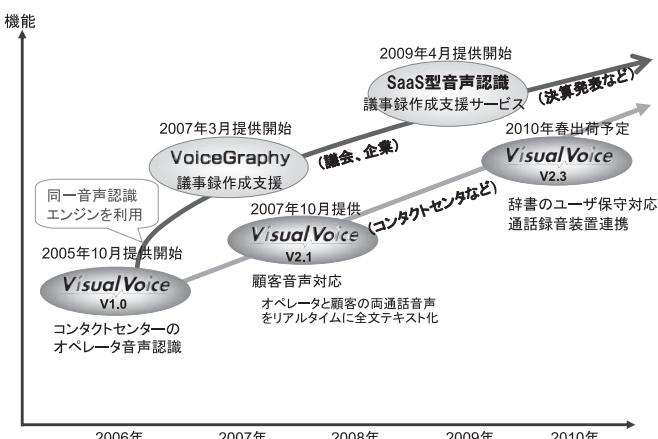


图2 話し言葉認識を活用した製品のビジネス展開

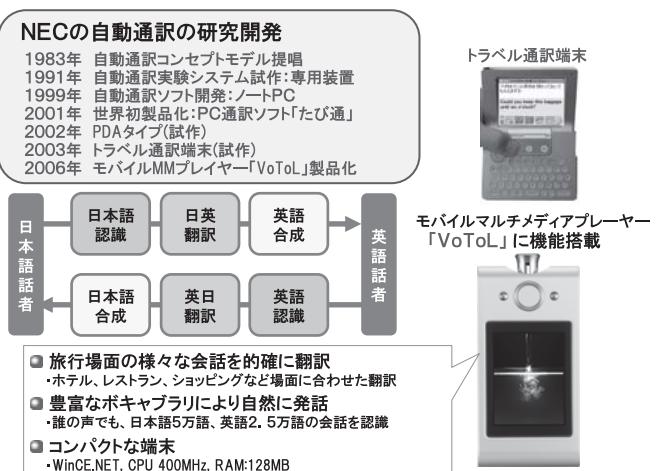


图3 自动通訳技术的端末組込みの取り組み

ます（図3）。

2 クラウドサービスに向けた取り組み

IT・ネットワーク技術の飛躍的な発展により、クラウドコンピューティングのシステムを、企業をはじめとする利用者が安全・安心に利用できる時代が到来しつつあります。

今後は、コンパクトな音声認識が組み込まれた多様な端末と、会議や電話などの会話音声を認識できるサービス基盤の両方を組み合わせることで、利用者がいつでもどこでも柔軟に音声をユーザインターフェースとして利用できるITサービス環境が求められてくると想定しています。

弊社では、クラウドサービスに向けた研究開発の取り組みとして、遠隔医療相談や観光地域での自動通訳、営業情報入力の業務シーンを例にした実証実験などを通じて、音声認識技術の実用化研究を進めています。

奥多摩町での遠隔医療相談の実証実験では、住民と遠隔の医師・医療スタッフとをブロードバンドネットワークを利用した遠隔相談システムで結び、未病状態の改善や成人病予防の相談・指導が行われました。医療スタッフが利用する端末では、健康相談内容の記録作業効率化のために音声認識を利用いただきました。今後は、遠隔相談を行っているときに、その会話音声をリアルタイムにテキスト化する仕組みも求められてきます。このような遠隔相談のサービスは医療をはじめ、多様な分野で広がっていく方向にあり、ネットワークを介して利用できる音声認識サービス基盤は本格的に提供されることが望れます。

観光地域での自動通訳実証実験は、旅行会話のような会話領域が限定された自動通訳技術が実用レベルに近づきつつあるため、この技術を地域の観光振興への貢献に役立てるために実施しています。モバイル型やウェアラブル型の端末を利用した自動通訳システムは、端末上に自動通訳機能を搭載して利用するだけでなく、多言語化への対応を効率良く行うためにサービス基盤上での自動通訳も利用して、全体として観光施設における訪日外国人の接客シーンを適切にサポートする仕組みとなっています。

営業情報入力業務における実証実験については、企業内の業務システムからWebサービス連携で利用できるリアルタイム音声認識サービス基盤を開発し、弊社内で運用しているSFAアプリケーションに組み込んで、サービスの有効性を検証したものです。クラウドサービスがWebサービスを基本として展開されていく中、音声認識をレスポンス良く、多様なWebアプリケーションから共通的に利用できるサービス基盤を提供することは、

企業向けシステムがサービス化されていく中で多様な適用シーンが出てくるものと考えています。

これらの実証実験を進めながら、クラウドサービスに対応するための音声認識サービス基盤の実用化課題を早期に発見し、それを解決できるよう、研究開発を進めていきます。

3 更なる精度向上へ向けた取り組み

更なる音声認識精度の向上に向けて、認識処理するための音声を的確に捉える技術も重要になります。その取り組みとして、雑音下音声認識技術、話者クラスタリング技術の研究開発を進めています。

雑音下の音声認識技術として、音声らしさをモデル化した知識を用いることで、さまざまな雑音環境に対し効果のある音声認識技術を開発しています。この技術では、非常に軽い処理でマイクの数に関わらず従来よりも高い音声認識性能が得られますので、車載機器や携帯型端末の組込み用途へ音声認識を展開することが可能となっています。

話者クラスタリング技術は、1つのマイクに話す人が途中で交代してもそれをリアルタイムに自動検知し、話者ごとに声の性質を自動的に学習することで認識性能を向上させる技術です。例えば会議室において1人に1つずつのマイクを設置していない環境においても高い音声認識性能を達成することができるため、今後、TV会議や社内一般の会議などにおいて音声認識を活用することが可能となります。

4 本特集の構成

本特集では、1章、2章、3章でご説明したそれぞれの取り組みの具体的な内容を以降の論文でご紹介します。

まず、話し言葉認識の特長を活かした最新の「ソリューション・サービス」、続いて、現在お客様に提供中の中心的な「音声認識製品」、そして、「研究開発」として実証実験のフィールドにおける技術実用化への取り組み、最後に、「導入事例」に関する論文を集めました。

NECが進めている音声認識ビジネスの全体像を本稿ではご紹介しました。それぞれの取り組みについて、その位置づけを理解の上、各論文をご一読いただけますと幸いです。