特集

音声認識技術の実用化への取り組み

編集にあたって

三木 清一
NEC 情報・メディアプロセッシング研究所

IT 技術の進展，情報機器の浸透に合わせ，音声認識技術は実用化の時代が来ている．カーナビや携帯電話には音声認識技術を利用したインタフェースやサービスが搭載され，コールセンタの自動応答も昔ながらのプッシュボタンから，音声で応答できるものが増えてきた．また，一般にはあまり目につかないかもしれないが，議会における会議録の作成支援，裁判員裁判における公判映像の検索，物流現場や市場でのハンズフリーでのデータ入力，コールセンタの中での情報分析といった，目に見えないところで社会を支える用途にも活用されている．このように，効率良く快適な，また安全なユーザインタフェースを提供したり，業務向けにさまざまな用途に活用されたりしている音声認識技術であるが，目指すべき人間の音声を聞き取る能力と比べると，まだその差は大きい．ある意味，永遠に課題の尽きない技術領域であるともいえる．本特集では，音声認識技術の適用範囲をさらに拡大させる上での，さまざまな課題への取り組みについてまとめて解説する．

音声認識技術とは，音声信号からテキストを取り出す技術である．現在主流となっている音声認識方式は，あらかじめ用意された音響モデルと言語モデルとを組み合わせ，入力音声に対して最も可能性が高い文を認識結果として出力するものである．概略，音響モデルとは「あ」「い」「う」といった単位シンボル(音素)とその音声信号との対応を示すモデルであり，言語モデルとはアプリケーションの用途等に応じ，その場で発話されやすい単語や文を表現するモデルである．すなわち，システムとして，音声的にも言語的にもさまざまな現象をどれくらいうまく事前の想定に組み込むことができるかが，高い認識精度を得るためのポイントとなる．実場面での応用において対応が必要な現象としては，種々の雑音や，利用者の多様な言語表現があり，システムとしてもさまざまな現象を表現できる，高い自由度を持つモデル表現の枠組みが求められる．また，技術開発により音声認識精度を高めていくことに加え，現在の音声認識技術の特性を踏まえ，実場面での使われ方を考慮して音声認識エンジンの実力を最大限発揮できるように，音声アプリケーションを設計・開発することも求められる．

本特集は大きく分けて 3 部から構成される．最初に，音声認識技術の実用化の現状について，技術解説や事例紹介を行う．「**1. 音声認識技術の実用化への取り組み**（古井）」では，音声認識技術の基本原理を説明し，現在公開されているツールやデータベー

ス，国内外の応用事例について幅広く紹介する．また，適用範囲の拡大に向けて解決すべき技術課題についても概観する．「**2. 携帯電話における分散型音声認識システムの実用化**（加藤）」，「**3. 使い勝手の良い音声インタフェースの実現**（庄境）」では，音声認識技術がインタフェースとして活用される代表的な場面として，携帯電話と車載情報機器（カーナビ）についての取り組みを紹介する．それぞれ，利用者の行動を分析し，どのような音声インタフェースが好ましいかについて説明する．

次に，実環境下で生じるさまざまな問題に対する技術的な取り組みについて，最新のトピックを紹介する．実環境下で生じる問題の代表的なものは雑音・騒音である．携帯電話や家電のリモコン等で利用可能な，コンパクトで高性能なマイクロホンアレイの開発について，「**4. 正方形マイクロホンアレイによる音源分離技術**（矢頭他）」で紹介する．雑談等の人の音声も雑音として入ってくるような，マイクを常にオンにして音声認識を動作させるチャレンジングな取り組みについて，「**5. ボタンレス音声インタフェースのための音声コマンド検知技術**（大淵）」で紹介する．記事の中で述べられているように，「リモコンで一番困ることはリモコンが見つからないこと」であり，マイクオンボタンを探さずに，機器を操作したいときにはただ発声するだけでシステムが「聞き分けて」くれることは実用化のさらなる拡大に向けて意義のある取り組みである．

利用者が，どんなしゃべり方をするかシステムが分からない，あるいは，利用者も，どんなしゃべり方をすればシステムに受け付けられるか分からないというのも実環境下で生じる大きな問題である．「**6. 音声認識実用化に向けた高次言語モデルの検討**（花沢）」では，利用者の多様な発話から，音声による操作で意味のあるキーワードを精度良く抽出する技術を紹介する．「**7. 音声インタフェースの現状とイノベーションの可能性**（西村他）」では，やはり利用者の多様な発話から，それら細かい表現の差異に影響されない利用者の「意図」を抽出する技術を説明し，その有用性を示す実験結果を紹介する．

現在の音声認識技術は機械学習に基づく部分が多く，大規模なデータを用いて複雑なモデルを学習することで音声認識精度を向上させることができる．しかしながらモデルが複雑になればなるほど，データが大量になればなるほど，その扱いは難しくなる．「**8.WFST に基づく T^3 音声認識デコーダ**（大西他）」では，こういった複雑なモデルを簡便に扱うことができる，非常にフレキシビリティの高い，スケーラビリティを持つ音声認識デコーダについて紹介する．「**9. 音声認識の多言語化技術**（河村）」では音声認識技術のグローバル展開に向けて，日本語にはない音韻的特徴を持つ言語，たとえば中国語やタイ語の音声認識技術や，多言語展開する際にそのコストを大幅に低減する技術について説明する．

最後に，音声認識技術（エンジン）を実際にアプリケーションに組み込む際に，その実力を発揮させるにはどうすればよいかについて，現状の課題を踏まえて解説する．「**10. サーバ連携に基づく継続的な音声認識応用システム開発**（小林他）」，「**11. 組込み機器向け音声インタフェース技術の開発プロセス**（平沢他）」では，音声アプリケーションを開発する上で，利用者，アプリケーション開発者，音声認識エンジン開発者が密に連携すべきであるという立場のもと，前者では運用時のログ収集や周辺機能のサービス提供を行う継続的な音声認識応用システム開発基盤を提案する．後者では，主に音声認識エンジン開発者の観点から，これまでの成功体験，失敗体験に基づいて，よい音声アプリケーションを開発するためにはどのようなことを押さえなければならないかについて紹介する．

本特集により，音声認識技術の研究者にとどまらず，各種アプリケーション開発者や利用者も含めた議論が活発化し，今までにない製品分野における音声インタフェースの実用化や，音声認識技術のさまざまな応用への展開が促進され，現在よりさらに利用者や社会に貢献できる音声アプリケーションが次々登場してくることを期待する．

（平成 22 年 10 月 13 日）