音声処理技術

Speech Processing Technologies

巻頭言

音声インタフェース技術の将来

Prospects for Speech Interface Technologies

古井　貞熙
FURUI Sadaoki

音声を用いて人とコンピュータシステムが情報のやり取りをするために必要な音声インタフェース技術には，何を話しているかを判定する“音声認識”や，誰が話しているかを判定する“話者認識”，コンピュータが音声で話しかける“音声合成”，音声を情報圧縮して伝送する“音声符号化”などの技術があります。更に，雑音などが重畳した実際の音声に対してこれらを実現するためには，“エコーキャンセラ”や，“音源分離”（マイクロホンアレー），“雑音抑圧”，“残響抑圧”などの技術が必要となります。

音声認識や音声合成の研究は1950年代から始まっており，人が音声を発声したり聞き取ったりする方法を，アナログ回路などで模擬していました。コンピュータが使えるようになった1970年代から，パターン認識などの数学的なアルゴリズムに基づいた研究が行われるようになり，1980年代から，統計的な処理が中心的に使われるようになって，技術が大きく進歩しました。近年では，多数の人が発声した多様な音声を録音して書き起こした膨大な音声データ（ビッグデータ）に基づいて，機械学習アルゴリズムを利用した音声のモデル化が行われるようになり，コンピュータの高度化に支えられて大きな技術的進歩をもたらしています。

筆者が主に従事してきた音声認識技術について言えば，研究を始めた1970年頃には実用化されると思っていた人はいませんでしたが，現在では音声対話システムや自動書き起こしなど多様な応用が実現しています。音声を文字に変換するだけでなく，他の言語に翻訳したり，話し手の意図を理解して応答したり（対話処理），音声でコンピュータに質問して回答を得たり（質問応答），録音された音声から必要な部分だけを検索したりするシステムも開発されています。

人がどうやって音声を発声し聞き取っているかについては，まだわかっていないことが多いですが，音声の音としての性質（音響モデル）だけでなく，語彙や，文法的な規則などの言語的な性質（言語モデル）を，巧みに組み合わせていることは確かです。人と同様の，あるいは人を超える能力を持った音声インタフェース技術の実現には，まだまだ基礎的な研究と，それに基づく大きな技術飛躍が必要です。

豊田工業大学シカゴ校　学長，工学博士　　President, Toyota Technological Institute at Chicago