音声認識向上へ深層学習

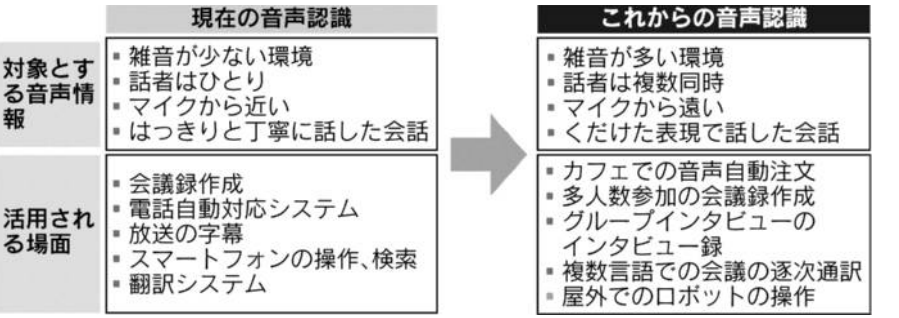
ＮＴＴデータ経営研究所
ＬＶＣユニットシニアコンサルタント
岸本 純子氏

音声認識技術は話した言葉をテキスト化する技術である。ＮＴＴドコモの「しゃべってコンシェル」や米アップルの「Ｓｉｒｉ」など、スマートフォン（スマホ）の音声インターフェースとして身近にある。最近はテレビなどの家電製品も音声で操作できるようになってきた。

音声認識技術はカーナビゲーションシステム、スマホやパソコンの音声インターフェース、多言語翻訳などのシステムに使われている。ビジネス分野では放送の字幕付与、会議録作成やコールセンターの電話自動対応システムなどに活用されている。登場した当初の音声認識率は低い印象を受けた。現在では話し言葉や、一定の雑音があっても認識できるほど精度が向上した。

雑音が少ない環境での音声認識技術の認識率は８～９割と高い。大容量のデータベース（ビッグデータ）の利用が可能になったこと、深層学習（ディープラーニング）の手法の１つであるＤＮＮが利用され始めたことが認識精度の飛躍的な向上に貢献している。音声認識技術はディープラーニングを実用レベルで活用している代表例と言える。

課題も残っている。友人と交わすくだけた話し言葉、雑音の多い環境での会話、入力マイクから離れた位置の発言、多人数の議論などについての認識率は極端に低下する。現在の技術は雑音の少ない環境で、１人がマイクに向かい、はっきりと丁寧に話すことが前提になっている。



機械学習させるために大量のデータは必要だが、音声データから雑音を分離したり、話者を識別したりする処理にもディープラーニングを活用できる可能性は十分ある。課題が解決されると、カフェやレストランでの音声による自動注文、会議での逐次通訳や議事録作成、屋外でのドローンやロボットの音声操作などができる。音声認識技術の利用場面はさらに広がるだろう。

きしもと・じゅんこ　グローバル、ヘルスケア、防災の分野で先端科学技術をベースとしたコンサルティングや新規事業開発に取り組む。博士（工学）。