

音声認識技術

音声を変換する音声認識技術は、スマートスピーカの普及など近年身近な技術になってきました。その裏には、End-to-End型音声認識モデルによる目覚ましい精度改善があります。しかし、話し方や周囲の環境（雑音や音声の回り込み）、処理を行うコンピュータの能力など、さまざまな状況・制約下で高い認識精度を維持することは容易ではありません。NTTでは基盤的な音声認識技術の深化に加え、限られた学習データで高い認識精度を実現するための自己教師あり学習技術や、人間でも聞き取りづらい音声を聞き取りやすく強調する音声強調技術の研究開発にも取り組んでいます。

NTTの音声認識は、演算能力の低いコンシューマデバイスでも高い認識精度を維持しつつ、リアルタイムに動作させることが出来るため、従来のクラウド型の音声認識サービスに限らず、手元のデバイスだけでネットワークを使わずに音声認識を利用することが可能になります。これにより、クラウド処理が受け入れ難い、機微な情報を含む会話の自動文字起こしなど、サービス領域のさらなる拡大が期待されます。

● End-to-End型音声認識

一つのニューラルネットワークモデルで、音声から文字に直接変換



● 処理負荷が小さく、NW不要でOn-Deviceでも音声認識処理が可能

端末に閉じた音声認識処理



クラウドでの音声認識処理



適応シーン例：

議事録・通話対応ログ・作業ログの自動作成、AIエージェント・自動応答サービス、音声メモなど