

多言語音声認識プラットフォーム

NTTサーブिसイノベーション総合研究所



多言語音声認識プラットフォームについて

「多言語音声認識プラットフォーム」は、世界最高精度を実現する最先端ディープニューラルネットワーク(DNN)技術と、アジアを中心とした10言語に対応した最新の音声認識エンジンを備えた、音声認識プラットフォームです。

本技術は、雑音下でも高精度な音声認識を実現、5秒程度の短い発話をもとに音声の特徴のみを利用して話している言語の判別を可能とするほか、研究所で開発された従来バージョンと比較して、少ない学習データでの音声認識が可能となることから、音声認識システムの運用稼働の多くを占めるモデルチューニング作業を大幅に削減することができます。

本技術の導入で、音声認識システムの運用稼働が削減され、人の稼働やICT機器利用による環境負荷の削減が見込まれます。



多言語音声認識プラットフォーム 概要 出典：NTT R&Dポータルサイト



環境貢献度評価

● 評価条件

コールセンタで利用される音声認識システムのモデルチューニング作業(※1)について、開発技術(開発バージョン)を利用した音声認識システムの場合と、従来技術(従来バージョン)を利用した音声認識システムの場合のCO₂排出量を比較することにより、開発技術の環境貢献度を定量化しました。

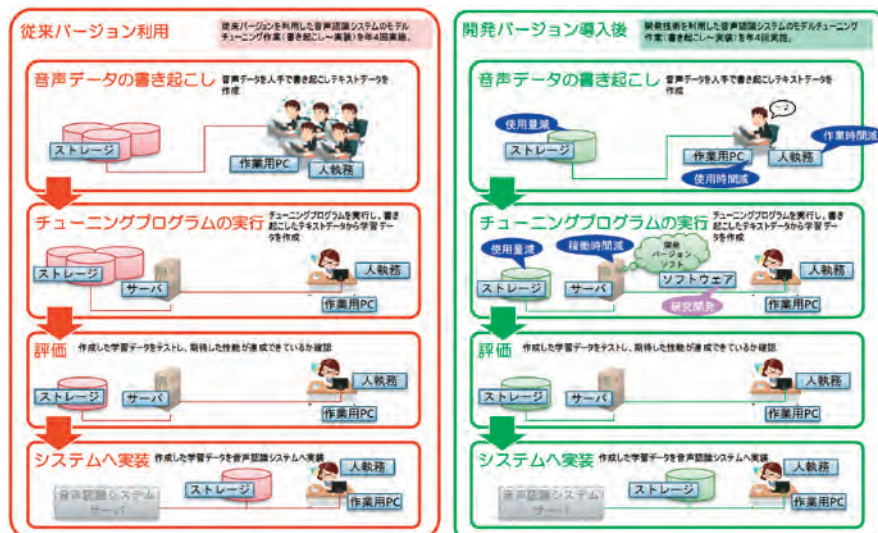
なお、評価値は、コールセンタ(※2)で利用される音声認識システムのモデルチューニング作業を年4回実施すると仮定しました。

● 評価結果

コールセンタ等で利用される音声認識システムに、開発バージョンを導入することにより、従来バージョンを利用した場合と比べて1,089kg-CO₂(削減率54%)の削減効果が見られました。

主な削減要因は、開発バージョンを導入することにより、音声データの書き起こしに関わるICT機器や人の稼働が削減されたことによるものです。

(※1) モデルチューニング作業：音声データ書き起こし・チューニングプログラム実施・評価・システムへの実装
(※2) 席数100程度のコールセンタを想定



評価モデル図

