

20

あなたの声に即座にあわせて音声認識

～音声認識精度を高めるニューラルネットの高速適応技術～

どんな研究

近年、様々な場面で音声認識が利用されるようになってきましたが、現在の技術でも、**誰の声でも同じように正確に認識できるわけではありません**。本展示では、システムが話者の声の特徴に即座に適応することで、常に高い音声認識精度を実現するモデル適応技術を紹介します。

どこが凄い

音声認識のための**ニューラルネット**を、**数秒間の音声データ**だけを用いて、その話者の声を最も認識しやすいように**即座に適応させる**ことができます。ニューラルネットワークの膨大なモデルパラメータを話者の特徴に合わせるには、従来は大量の音声データでの学習が必要でした。

めざす未来

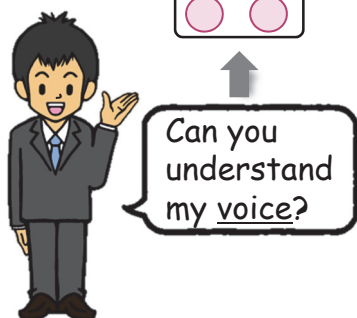
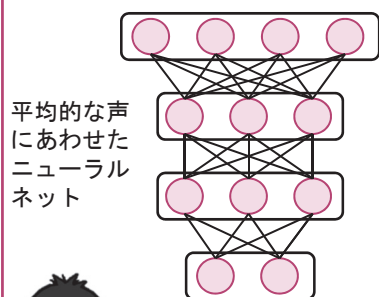
話者の声の特徴だけではなく、収録している部屋の音響特性や、雑音の性質に応じて、**常に高い精度の音声認識**が実現できるようになります。さらに、音声認識だけではなく**入力データの特徴に合わせて振る舞いを即座に変更・最適化**する人工知能技術の実現を目指します。

【従来の音声認識】

平均的な声にあわせたモデルを使用するため、話者の声の特徴によっては、高い精度は得られない。

Can you understand my choice? 

↑ 単語誤り率: 5.6%※

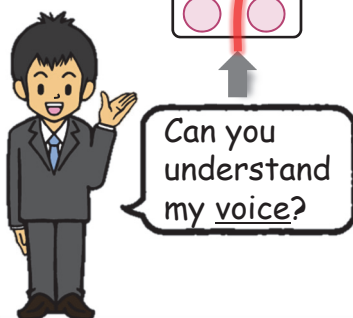
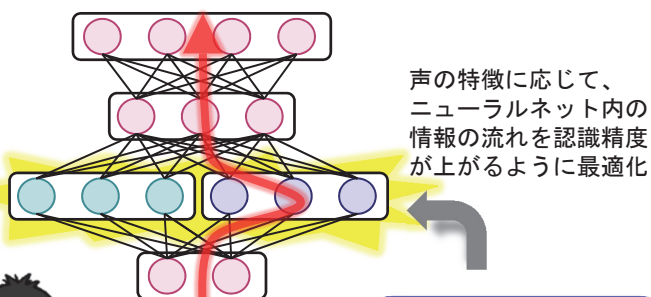


【話者の特徴に即座に適応する音声認識】

話者の声の特徴を抽出し、その特徴に応じて認識しやすいように、ニューラルネットワークの情報の流れを制御することで、各話者にあわせた音声認識を実現。

Can you understand my voice? 

↑ 単語誤り率: 4.8%※
※英語新聞読み上げ音声認識での評価



発話からその話者の音響的特徴を抽出

数秒程度の短い音声データからでもOK!

関連文献

- [1] M. Delcroix, K. Kinoshita, T. Hori, T. Nakatani, "Context adaptive deep neural networks for fast acoustic model adaptation," in *Proc. ICASSP*, pp. 5270-5274, 2015.
- [2] M. Delcroix, 木下慶介, Yu Chengzhu, 小川厚徳, 吉岡拓也, 中谷智広, "音響コンテキスト適応型DNNに基づく高速音響モデル適応," *日本音響学会春季研究発表*, 2016.
- [3] M. Delcroix, K. Kinoshita, A. Ogawa, T. Yoshioka, D. Tran, T. Nakatani, "Context adaptive neural network for rapid adaptation of deep CNN based acoustic models," in *Proc. Interspeech*, 2016.

連絡先

デルクロア マーク (Marc Delcroix) メディア情報研究部 信号処理研究グループ
E-mail: marc.delcroix(at)lab.ntt.co.jp