

コンピュータが音を読み解く

～深層学習を活かした実環境下音声認識・音響処理～

どんな研究

多様な音環境での**会話音声認識**や**会話シーン分析**の研究を進めています。これに不可欠な基盤技術の一つとして、深層学習を用いた雑音や残響除去技術や、自由な話し言葉音声を高精度に認識する技術を提案しています。

どこが凄い

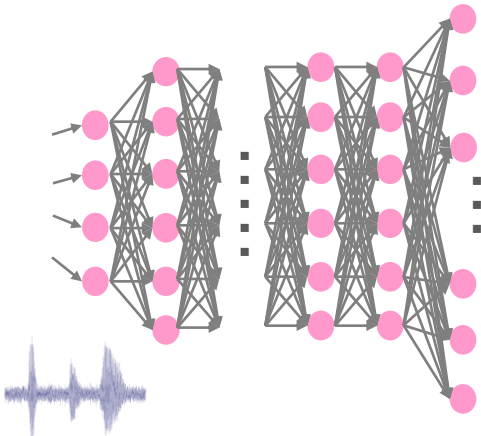
深層学習が広く検討されている音声認識だけでなく、雑音・残響除去や環境音認識も含めた幅広い音声処理技術について、**深層学習に基づく効果的な方法**を提案しました。これにより、実環境で収録した音から「何が起きているか」「何を話しているか」などを、より正確に推定することができるようになりました。

目指す未来

オフィス打合せスペースでの議論や、カフェやリビングでの会話など、騒がしい所で多くの人が話す場面でも**快速に動作する会話音声認識が可能**になります。例えば騒がしい場所での打合せの議事録自動作成や、ウェアラブル端末での音声ライフログ記録などが可能となり、**音声インタフェースがより便利**になります。

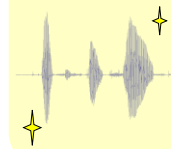
深層学習

- 深い層構造を持つニューラルネットワーク (deep neural network: DNN) で、複雑な現象も表現
- 観測した音声の品質向上や音声認識に有効な特徴量を、自動的に学習
- 近年、さまざまな音声認識タスクで深層学習が高性能を達成している



騒がしい場所で収録した音声
(会話・雑音など)

音響処理



音声を取り出すフィルタ計算に最適なパラメタを自動選択して、録音音声に含まれる雑音や残響を低減！

音声認識

「アレだと思うんですけど...」

話し言葉の長い文脈や収録環境の違いなども区別することで、認識精度を格段に向上！

環境音認識



環境音(扉の開閉音・拍手など)の識別に最適な特徴量を自動学習して、高精度な認識を実現！

実用化



DNN音声認識をリアルタイム処理可能な音声認識エンジンとして実用化※！

【応用例】リアルタイム会議音声認識システム：RexMeet

※メディアインテリジェンス研究所にてエンジン化しました

関連文献

- [1] 久保陽太郎, “音声認識における深層学習の活用とその進展,” 電子情報通信学会技術研究報告, Vol. 114, No. 52, pp. 39-44, 2014.
- [2] S. Araki, T. Hayashi, M. Delcroix, M. Fujimoto, K. Takeda, T. Nakatani, “Exploring multi-channel features for denoising-autoencoder-based speech enhancement,” in *Proc. ICASSP*, 2015.
- [3] D.Q. Truong, S. Nakamura, M. Delcroix, T. Hori, “WFST-based structural classification integrating DNN acoustic features and RNN language features for speech recognition,” in *Proc. ICASSP*, 2015.

連絡先

荒木章子 (Shoko Araki) メディア情報研究部 信号処理研究グループ
E-mail : araki.shoko(at)lab.ntt.co.jp

