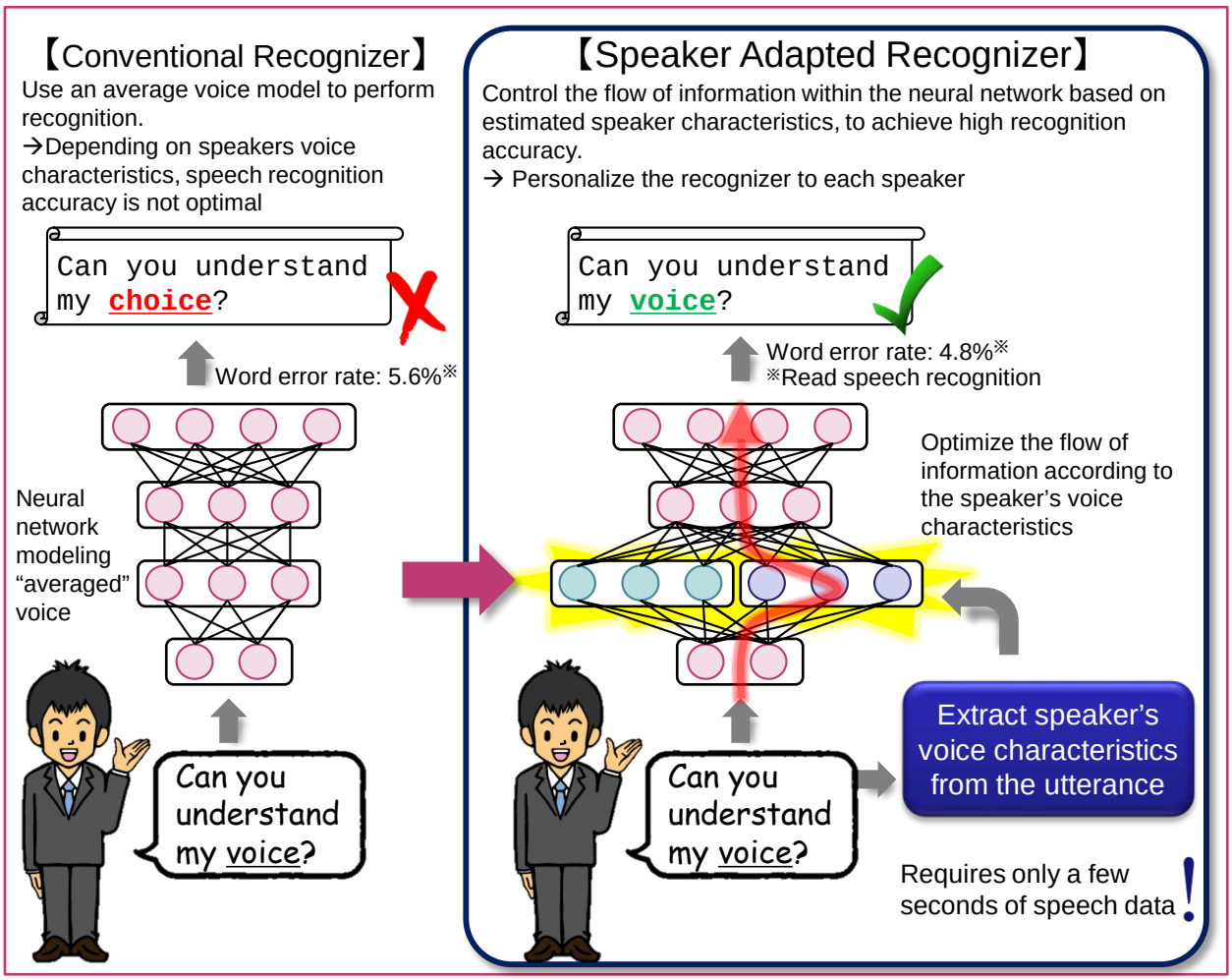
20 Personalizing your speech recognizer

- Neural network adaptation for automatic speech recognition -

Abstract

Automatic speech recognition is used more and more often in our everyday life. However, the accuracy of speech recognizers largely varies depending on the speakers. In this exhibit, we present a system that can adapt to the speaker's characteristics to maintain high recognition accuracy for all speakers. The proposed system first extracts the speaker's voice characteristics and then uses them to adjust neural network parameters for optimal speech recognition accuracy. Since only a few seconds of speech data are sufficient for estimating the voice characteristics, the proposed system can adapt a large number of network parameters using very little speech data. In addition, this approach can potentially be extended to other types of acoustic variations, such as noise, to realize noise-robust speech recognition. Moreover, the same ideas could be applied to other AI problems, where we want to control the behavior of a network depending on the input characteristics.



Reference

[1] Delcroix, M., Kinoshita, K., Hori, T. and Nakatani, T., "Context adaptive deep neural networks for fast acoustic model adaptation," in Proc. *ICASSP, 2015*

[2] Delcroix, M., Kinoshita, K., Ogawa, A., Yoshioka, T., Tran D., and Nakatani, T., "Context adaptive neural network for rapid adaptation of deep CNN based acoustic models," in Proc. *Interspeech, 2016*

Contact

Marc Delcroix Signal Processing Research Group, Media Information Laboratory
Email : marc.delcroix(at)lab.ntt.co.jp