13 Anytime, anywhere, we can speak like a native!

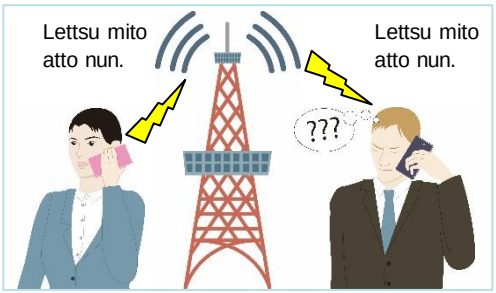
- Speech rhythm conversion by mobile application -

Abstract

Most native Japanese speakers have difficulty speaking English and therefore cannot communicate well in English with native English speakers. We have proposed a method 'Speak like a native' that can convert the speaking rhythm of English sentences spoken by native Japanese speakers into stress-timed rhythm by a native English speaker. However, our previous speaking rhythm conversion technique needed the same sentences to be spoken by a native speaker. In this study, we devised rules for converting the speaking rhythm of native Japanese speakers into that of native English speakers using English speech corpus and showed a speaking rhythm conversion of arbitrary speech sentences automatically. Moreover, we developed a conversion system using a mobile device. We hope that this technique will eventually alleviate the burden involved in communication using non-native languages.

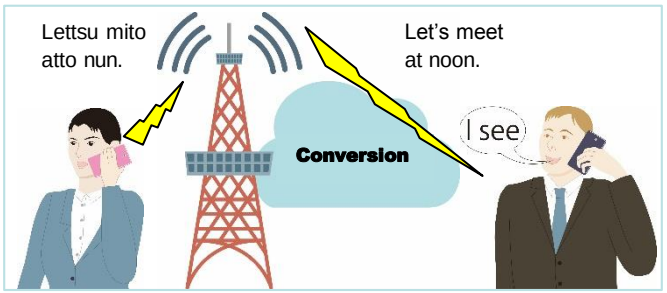
Difficult to communicate in Japanese English with native speakers

Japanese native speakers tend to speak English using mora-timed rhythm.



Speaking rhythm conversion technique can help it

Convert the speech rhythm spoken by native Japanese speakers into stress-timed rhythm of native English speakers



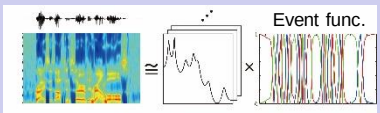
Three base techniques of rhythm conversion

Estimation of vocal-tract spectrum

PEAR (Phase equalization-based AR) [1]

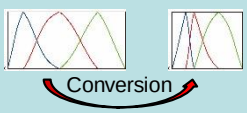
Extract event functions from vocal-tract spectra

NTD (Non-negative temporal decomposition) [2]



Training of event functions

Using an English speech corpus (350 native Japanese and 700 native English speakers), we train rules for converting the speaking rhythm of native Japanese speakers into that of native English speakers. [3]



Speaking rhythm conversion using a mobile device

PC server receives an English speech spoken by a native Japanese speaker and converts it into like a native.

Anytime, anywhere, we can convert the speaking rhythm by a mobile device.

References

[1] S. Hiroya, T. Mochida, "Phase equalization-based autoregressive model of speech signals," in *Proc. Interspeech*, pp. 42-45, 2010.

[2] S. Hiroya, "Non-negative temporal decomposition of speech parameters by multiplicative update rules," *IEEE Trans. Audio, Speech, and Lang. Process.*, Vol. 21, No. 10, pp. 2108-2117, 2013.

[3] S. Hiroya, "Speaking rhythm conversion of arbitrary speech using a non-negative temporal decomposition and machine learning," in *Proc. Acoustical Society of Japan, Autumn meeting*, pp. 423-424, 2017.

Contact

Sadao Hiroya Sensory and Motor Research Group, Human Information Science Laboratory