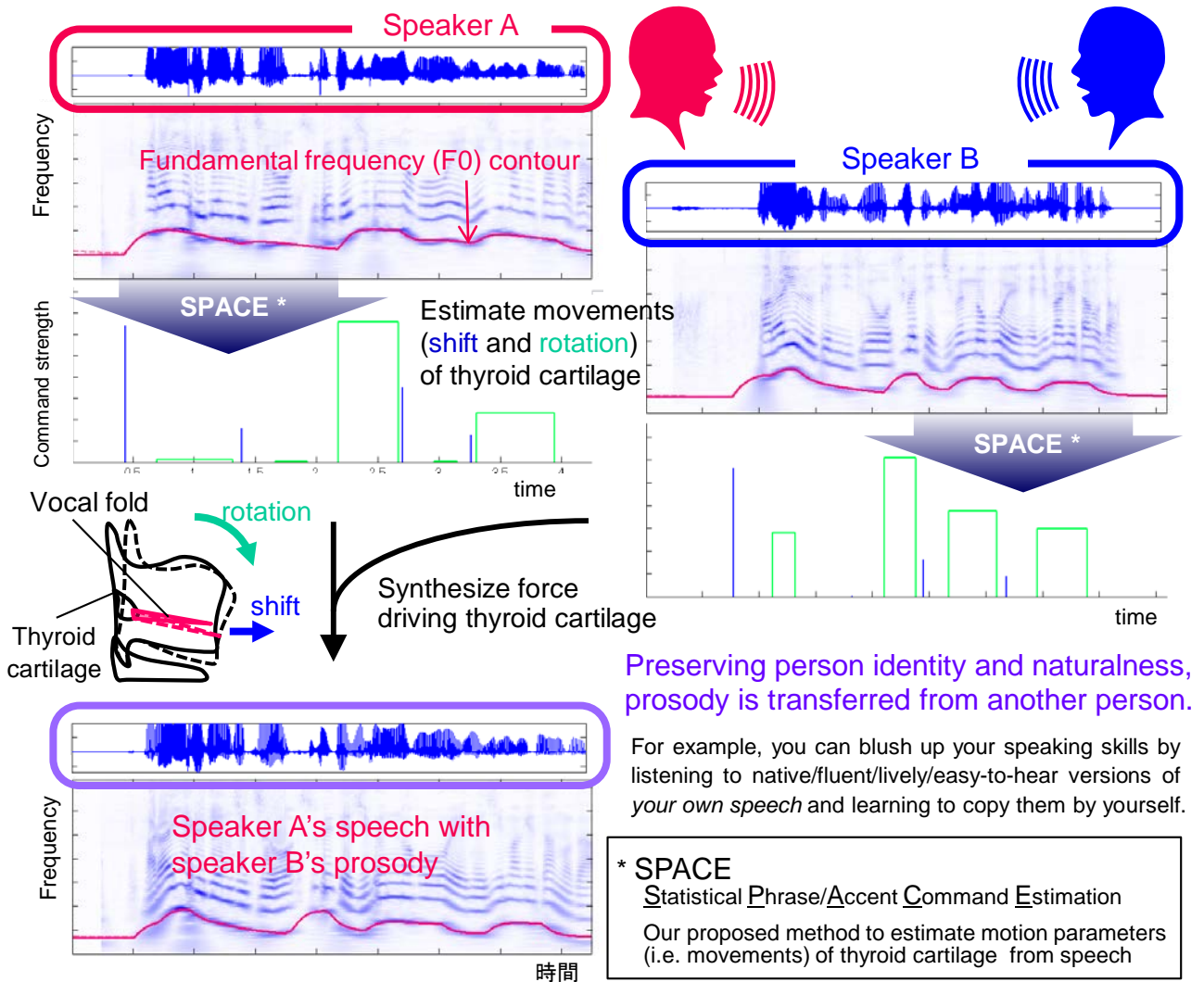


Abstract

A voice F0 contour, one of the essential components of human speech prosody, is generated by the vocal fold, whose tension is controlled by the thyroid cartilage movement. This mechanism is known to be well described by the Fujisaki model, a physical model consisting of a set of intonation (shift) and accent (rotation) parameters, but accurate estimation of those parameters has been considered difficult for a long time. Our proposed method "SPACE" makes it possible to accurately estimate the movements of the thyroid cartilage from speech, taking advantages of our newly established statistical framework with a specially designed HMM. Here we demonstrate intonation morphing, where you can speak with your own voice, but borrowing another person's intonation and accent. The potential applications of the system range from naturalness-guaranteed text-to-speech systems or voice conversion systems, to self-training systems for improving presentation/language skills.



【Reference】

- [1] H. Kameoka, K. Yoshizato, T. Ishihara, K. Kadowaki, Y. Ohishi, K. Kashino, "Generative modeling of voice fundamental frequency contours," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 6, pp. 1042-1053, Jun. 2015.
- [2] H. Fujisaki, O. Fujimura, Ed., "A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour," in *Vocal Physiology: Voice Production, Mechanisms and Functions*. New York, NY, USA: Raven, 1988.

【Contact】

Hirokazu Kameoka Recognition Research Group, Media Information Laboratory
E-mail : kameoka.hirokazu(at)lab.ntt.co.jp