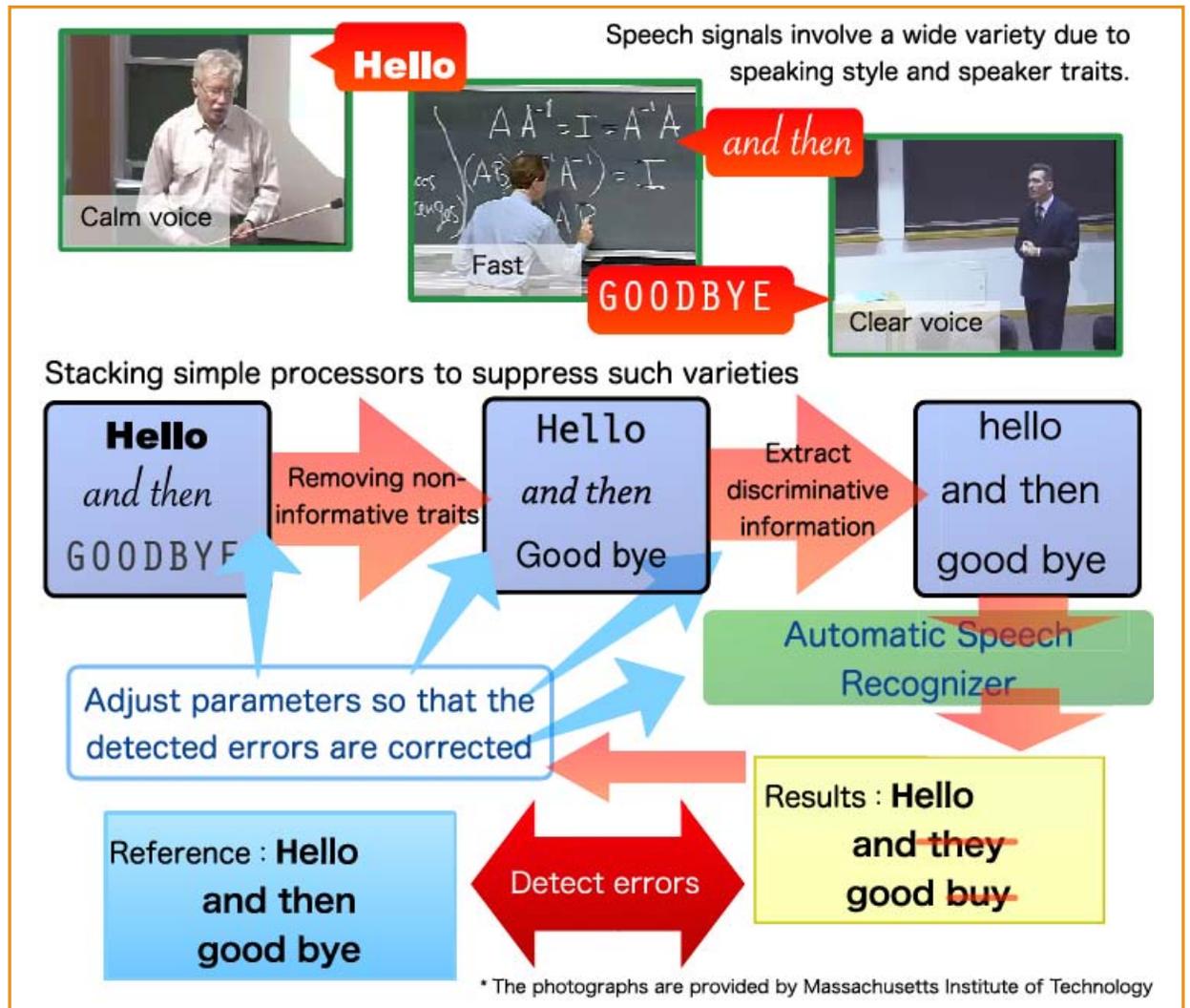




# Transcribing every word whoever speaks

## Speech recognizers robust to casual speech variations

**Abstract**— Automatic speech recognition is considered to be a mature technology that can accurately recognize sentences that are correctly and formally uttered. However, the recognition of casual speech is still a challenging task since speakers and speaking styles inevitably vary greatly in such situations. To tackle this, we developed a unified statistical model for both acoustic and language representations, and enhanced it by using a multi-layered algorithm that extracts recognizer-friendly representations from the observed speech by combining simple processing layers. Using these technologies, we are investigating a robust speech recognizer that can transcribe every word that anyone speaks. The recognizer will help to promote a natural interaction between humans and machines.



### Related works

- [1] Y. Kubo, T. Hori, A. Nakamura, "Integrating deep neural networks into structured classification approach based on weighted finite-state transducers," in *Proc. INTERSPEECH*, 2012.
- [2] Y. Kubo, T. Hori, A. Nakamura, "Large vocabulary continuous speech recognition based on WFST structured classifiers and deep bottleneck features," in *Proc. ICASSP*, 2013.

### Contact

**Yotaro Kubo** Signal Processing Research Group, Media Information Laboratory  
E-mail : kubo.yotaro{at}lab.ntt.co.jp (Please replace {at} with @)