

# Speech Interface

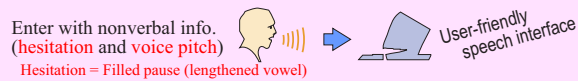
M. Goto (AIST) K. Itou (Nagoya Univ.) T. Kobayashi (Waseda Univ.)

## Exploiting Intentionally-Controlled Nonverbal Speech Information

### New Direction of Speech Interface

- ❑ Exploit **nonverbal** speech information
  - Current speech-input interfaces have **not** fully exploited the potential of speech
  - Most speech recognizers utilize only **verbal** information (phoneme / word)

➔ Make use of **nonverbal** speech information intentionally controlled by a user



- ❑ Achieve **four interface functions** that can exploit the potential of speech
  - Ensure both **verbal** and **nonverbal** info. are recognized and used in well-balanced ways
  - Convey **new types of information** from the user side to the computer side
  - Experimental results showed that four functions are robust & effective for Japanese speech

### Conventional Speech Input

- ❑ Did not use **nonverbal** info. or used only **nonverbal** info. **unconsciously** uttered in natural speech input
  - **Hesitation** caused erroneous recognition
  - **Pitch** was only used to improve recognition rates
  - Exceptions: [Schmandt 1988] [Igarashi, et al. 2001] [Olwal, et al. 2005]

### Four Functions

Please enjoy video and live demonstrations!

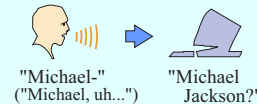
#### Speech Completion

Filled pause

- ❑ Help a user enter an **uncertain** word / phrase by **completing the missing part** of a partially uttered fragment
- ❑ What is a good **completion trigger** for speech?



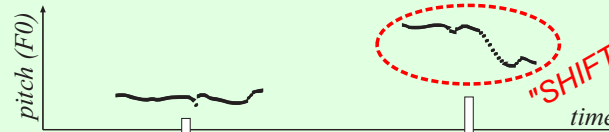
You can input uncertain phrases using **filled pauses**!



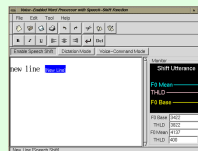
#### Speech Shift

Filled pause + Voice pitch

- ❑ Enable a user to switch speech-input modes by **intentionally changing the pitch**



You can **directly** enter a word in the intended speech-input mode!

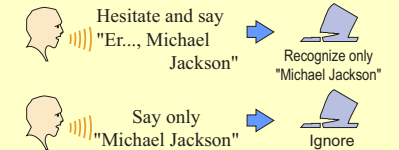
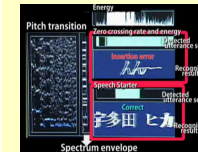


#### Speech Starter

Filled pause

- ❑ Enable a user to specify the **beginning** of each utterance with a filled pause
- ❑ **Noise-robust endpoint-detection** without using other input devices (e.g., button) for **non-stationary noisy environments**

You can enter a **voice command** in noisy environment!



#### Speech Spotter

Filled pause + Voice pitch

- ❑ Enable a user to enter **voice commands** into a speech recognizer in the midst of **natural human-human conversation**
- ❑ Monitor human conversation and accept only a **speech-spotter utterance** that is intentionally unnatural

You can enter a **voice command** in the midst of **human-human conversation**!

