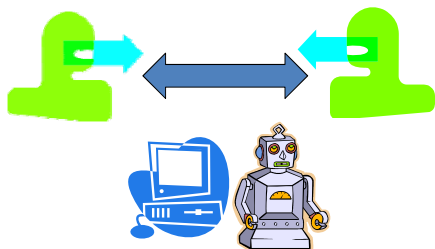# Smart Seminar Room based on Multi-Modal Recognition of Verbal and Non-Verbal Information

Project Core Members

- PI : **Prof. Tatsuya Kawahara (Kyoto University)**

- Kyoto University
  - Prof. Yuichi Nakamura (Video Processing)
  - Prof. Takashi Matsuyama (Computer Vision)
  - Prof. Sadao Kurohashi (Natural Language Processing)

- Nara Institute of Science & Technology
  - Prof. Kiyohiro Shikano (Speech Processing)
  - Assoc. Prof. Hiroshi Saruwatari (Acoustic Processing)
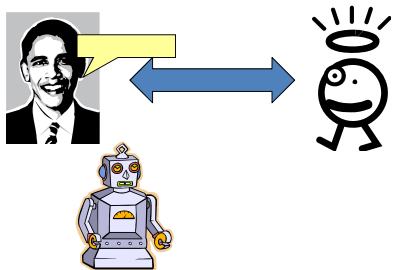
Problems

Speech
Communication

Meetings & Conversations

- Speech-to-text
  - Speech recognition
  - Captioning

- Sensing of comprehension
  & interest level
  - Assist comprehension
  - Presentation upon interest
  - Annotations

Approach

Speech
Communication

**Multi-modal
sensing & analysis**

- Listening to main speakers
  (Content-based)
  - Speech recognition
  - Natural language analysis
  - Keyword-based indexing (tf-idf)

- Sensing audience reaction
  (Interaction-based)
  - Non-verbal information
    Gaze, nodding, backchannel
  - Attention, interest-level

Process Overview

Speech recognition → Content analysis → Content-based indexing

Audio analysis →

Video analysis → Interaction analysis → Interactive presentation

Reaction-based indexing

[multi-modal sensing]   [archiving & presentation support]

Targets

large / formal / few — **audience style interaction** — small / informal / a lot

 # main speakers — one ... many

lecture → **Seminar talk**

Parliament meeting

Poster conversation

Real-time assistance — Captioning — Explaining technical terms   Switching slides

Offline archiving — Transcription — Annotations for browsing & search

Multi-modal Sensing & Analysis

Video
Motion
Audio

| Pointing |
| Gaze (head) |
| Nodding |
| Backchannel |
| Laughter |
| Utterance |

(mental state)

attention

compre-hension

interest

courtesy