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**
  - Speech-to-text
    - Speech recognition
    - Captioning
  - Sensing of comprehension & interest level
    - Assist comprehension
    - Presentation upon interest
    - Annotations

### Approach

- **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 → Interaction analysis → Content-based indexing
Audio analysis → Interaction analysis → Interactive presentation
Video analysis

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

Targets

large formal few
audience style interaction small informal a lot

one

Seminar talk
Poster conversation

# main speakers

Parliament meeting

many

Real-time assistance
Captioning Transcription
Explaining technical terms
Switching slides
Annotations for browsing & search
Offline archiving

2012/4/23
Multi-modal Sensing & Analysis

- Video
- Motion
- Audio

<table>
<thead>
<tr>
<th></th>
<th>Pointing</th>
<th>Gaze (head)</th>
<th>Nodding</th>
<th>Backchannel</th>
<th>Laughter</th>
<th>Utterance</th>
</tr>
</thead>
</table>

- (mental state)
- attention
- comprehension
- interest
- courtesy