Nonverbal Behaviors

- Head gestures
- Head gaze
- Eye gaze
- Eye gestures
- Facial expressions
- Arm gestures
- Body posture
Interactive Interfaces

Mel, robotic penguin
- Conversational
- Physical embodiment

GUI interface
- Non-conversational
- Non-embodied

Peter, on-screen avatar
- Conversational
- Virtual embodiment
Nonverbal Behaviors for Interactive Interfaces

Head nod
- Grounding
- Answering

Head shake
- Answering

Gaze aversion
- Turn-taking
- Thinking
Mel, Interactive Robot [HRI 2006]

- 2 sets of interactions (iGlassware and self-demo)
- 16 participants
- ~3 mins / interaction
Mel, Interactive Robot  [HRI 2006]

- Early user study without recognition [Sidner04]
  - Participants head nodded (Average: 20.8%)

- Recent user study with recognition [Sidner06]
  - Average: 29.3%

- Robot nods back
  - Average: 42.3%
19 participants

Two tasks, each one performed 3 times:
  – Reading a short text
  – Emailing answers

Dialog box appears randomly

~10mins interactions

End with short questionnaire
User Study [IUI 2006]

- Three interaction phases:
  1. Conventional inputs
  2. Head gestures
  3. Free choice
- Short tutorial before 2nd phase
- ~3 minutes per phase
Free Choice Results [IUI 2006]

Percentage of usage (%)

- Dialog box answering
- Document browsing

- Head gestures
- Mouse
- Keyboard
Peter, On-screen Avatar [ICMI 2006]

- 6 participants
- 100 questions
  - Is your age an even number?
  - Does Canada have a president?
  - If five a prime number?
- 10-12 minutes / interaction
Gaze Aversion [ICMI 2006]

- **Definition:** Eye movements to empty or uninformative regions of space, reflecting “look-away” of “thinking”.

- 3 coders labeled 114, 72 and 125 gestures
- Intersection of all 3 coders
- 72/600 answers
- Average: 12%
User Studies

Mel, robotic penguin
Head nods
• Grounding/Answering

GUI interface
Head nods
• Command

Peter, on-screen avatar
Gaze aversions
• Thinking
Recognition of Nonverbal Behaviors

Head shake
Contextual Prediction

Look left

“Right there is the iGlassware cup.”

Head shake
Contextual Prediction

Look left
Head shake
"By the way, do you know Paul?"
the iGlassware cup.”
Contextual Prediction

Look left

“Right there is the iGlassware cup.”

Head shake

“By the way, do you know Paul?”
Key Idea

- Use contextual knowledge from the interactive system to improve recognition of nonverbal behaviors
  - Embodied interfaces (robot, on-screen character, …)
  - Non-embodied interfaces (dialog boxes, …)
Related Work

- Phone-based interactive systems [SLS Jupiter]
- Interactive information kiosks
Related Work

- BEAT: Behavior Expression Animation Toolkit [Cassell01]
Related Work

- Eye gaze for pointing/scrolling [Toyama98, Kjeldson01]
- Body/arm tracking [Demirdjian02]
Related Work

• Engagement with an interactive robot [Sidner03]
• Sociable robot [Breazeal00]
Related Work

- Audio-visual speech recognition [Dupont00]
Related Work

- Prosody and head gestures [Fugie04]
Related Work

- Visual Recognition
- Speech Recognition
- Dialog Manager
  - Agenda
  - History
- Visual Synthesis
- Speech Synthesis

• Dynamic grammars [Lemon04]
Our Approach

- Context-based visual feedback recognition
Context-based Recognition Framework

Context-based recognizer

- Gaze estimator
- Gesture recognizer

Contextual predictor

Multimodal integrator

Dialogue context

Visual feedback
Head Pose Tracking

View-based Model

Pitch

Yaw

View-based Model

Pose-change Measurements
Adaptive View-based Model [CVPR 2003]

Online Model Acquisition

Pitch

Pose-change Measurements

Yaw

Adaptive View-based Model
Adaptive View-based Model [CVPR 2003]
Head and Eye Gaze Estimation
Context-based Recognition Framework

- Gaze estimator
- Contextual predictor
- Multimodal integrator
- Gesture recognizer

Dialogue context
Visual feedback
ROC Curves

Head nods

Gaze aversion
Context-based Recognition Framework

Context-based recognizer

Gaze estimator

Gesture recognizer

Contextual predictor

Multimodal integrator

Dialog Manager

Agenda

History
Contextual Features

- Generic to most architecture
- Easily available
- Fast to compute
- Helps recognition

Diagram:

- Dialog Manager
  - Agenda
  - History

Contextual features
Contextual Features

- Lexical features
- Prosody/punctuation
- Timing
- Gesture display
Contextual Features

MEL: Do you see the copper in the glass?

MEL: Well, it transmits to the readout display by inductance with the surface of the table.

\[ f_1 : \text{timing} \]

\[ f_2 : \text{question mark} \]

\[ f_{\text{do you}} : \text{bigram "do you"} \]
Contextual Predictor

- **Input:**
  - Contextual features

- **Output:**
  - A likelihood measurement for each visual gesture

- Learn from examples
Training Contextual Predictor

- **Interactions with Mel, robot-penguin**
  - Conversation manager: Collagen [Rich *et al.*, 2001]
  - Developed at MERL

- **2 sets of interactions**
  - iGlassware: 9 participants
  - Self-demo: 7 participants

- **Training set: iGlassware**
Contextual Predictor

\[
[f_t, f_?, f_{"do you"}]
\]

\[f_t: timing\]

\[f_?: question mark\]

\[f_{"do you"}: bigram "do you"\]

Head nods

Head shakes
Multimodal Integration

- **Integrates:**
  - Contextual prediction
  - Visual observation
- **Determine if a head gesture is happening**
Multimodal Integration

Vision-based recognizer

Contextual predictor
Multimodal Integration

Vision-based recognizer

Contextual predictor
Multimodal Integration

Vision-based recognizer

Contextual predictor

Multimodal Integration

Head nods

Ground truth

Head nod

No gesture
# 2006 Conference Program

**Mozilla Firefox - Automatic update**

A new version of Firefox is available. Do you want to install it?

- Yes
- No

- People wanting to buy extra tickets for the Dinner Cruise need to contact Shigeki Amitani when registering. They will be informed of the cost of extra tickets.
- **Dinner Cruise:** we should already indicate that we meet at the wharf: King Street Wharf No. 4 (People who want to buy extra tickets need to tell their spouse where to meet them!)

<table>
<thead>
<tr>
<th>Sunday, Jan 29</th>
<th>Monday, Jan 30</th>
<th>Tuesday, Jan 31</th>
<th>Wednesday, Feb 01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorials and Workshops</td>
<td>Welcome</td>
<td>Papers Recommendation 1</td>
<td>Invited Talk Jeffrey Shaw</td>
</tr>
<tr>
<td>Invited Talk Hiroshi Ishiguro</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papers Gestural Input</td>
<td>Papers Multimedia and</td>
<td>Papers Personal Assistants 2</td>
<td></td>
</tr>
</tbody>
</table>
Non-conversational Interfaces

- Context-based recognizer
  - Gaze estimator
  - Gesture recognizer
  - Contextual predictor
  - Multimodal integrator

- Window Manager
  - Input Events
  - Output Events
Contextual Features

- Conventional input devices
  - Mouse
  - Keyboard
- Current display
  - Dialog boxes
User Studies and Context-based Recognition

Mel, robotic penguin

Head nods
- Grounding/Answering

GUI interface

Head nods
- Command

Peter, on-screen avatar

Gaze aversions
- Thinking
Context-based Recognition

Head nods

Gaze aversion

![Graphs showing performance metrics for head nods and gaze aversion with different recognition methods.](Image)
Conclusion

- **Context-based Visual Feedback Recognition**
  - Embodied interfaces (robot, on-screen character, …)
  - Non-embodied interfaces (dialog boxes, …)
- **Adaptive View Appearance Model**
  - Algorithm for head gaze tracking
- **User studies**
  - Mel: Head nods with interactive robot
  - GUI: Gesture-based widgets
  - Peter: Gaze aversion with on-screen avatar