Understanding Nonverbal Behaviors

The Role of Context in Recognition

Louis-Philippe Morency | 09/24/2008

The projects or efforts depicted were or are sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM),and/or the US Army Research Institute. The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

USC

ICT

INSTITUTE FOR CREATIVE TECHNOLOGIES

Research Agenda

Nonverbal Behaviors for interactive interfaces

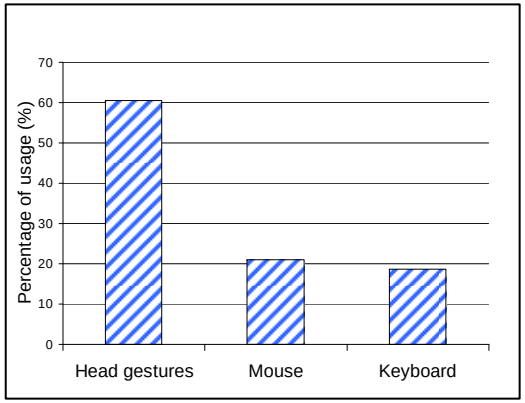
- Natural and useful feedback
- Recognition algorithms
- Contextual information

User Studies: Natural Behaviors

Mel, robotic penguin [HRI 2006]

GUI interface [IUI 2006]

Peter, virtual human [ICMI 2006]



USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Research Agenda

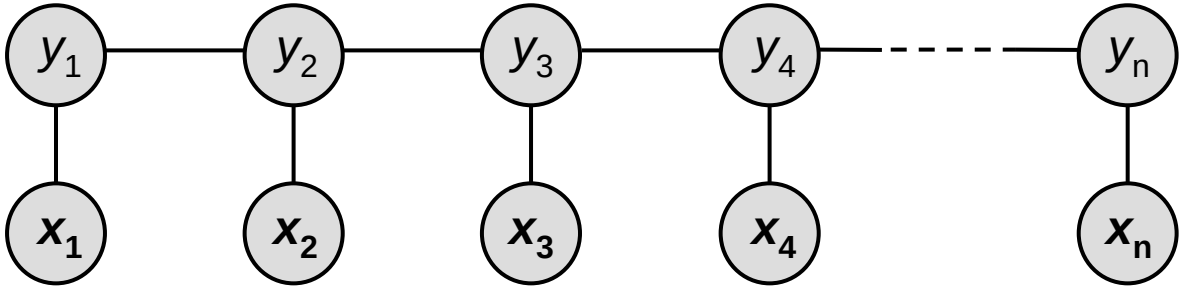
Nonverbal Behaviors for interactive interfaces

- Natural and useful feedback
- Recognition algorithms
- Contextual information

User Study Observations

- Natural gestures:
  - Head gestures: head nods and head shakes
  - Eye gestures: gaze aversion
- Internal sub-structure
- Dynamics between gestures

Conditional Random Field (CRF) [McCallum 2001]



$$P(\mathbf{y} \mid \mathbf{x}, \theta) = \frac{1}{Z(x, \theta)} \exp\left(\sum_k \theta_k \mathbf{F}_k(\mathbf{y}, \mathbf{x}) \right)$$

CRF outperforms HMM for activity recognition

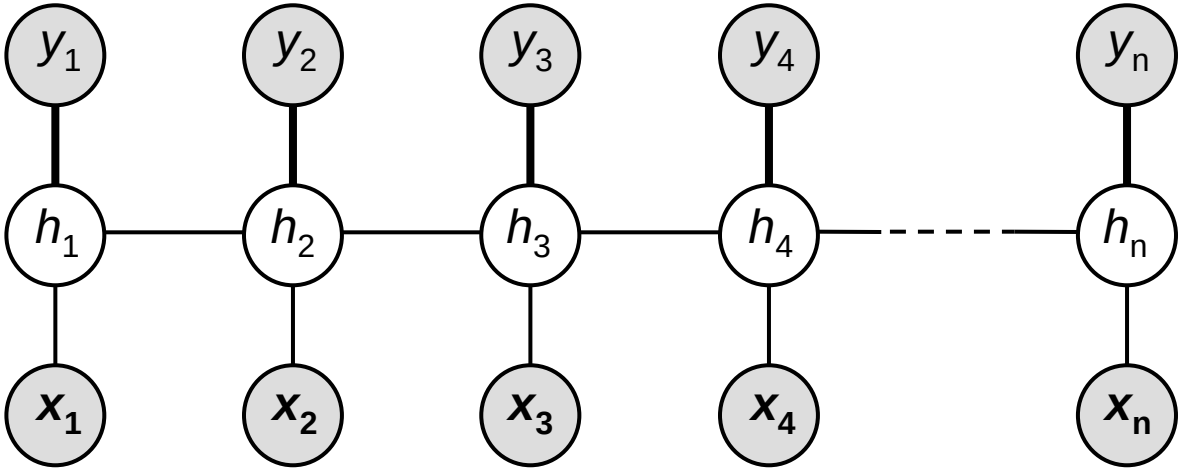
[Sminchisescu05]

Problem: CRF does not model internal sub-structure

USC

ICT INSTITUTE FOR CREATIVE TECHNOLOGIES

Latent-Dynamic CRF [CVPR 2007]



y_j : label

h_j : hidden state

$\boldsymbol{x}_j$: observations

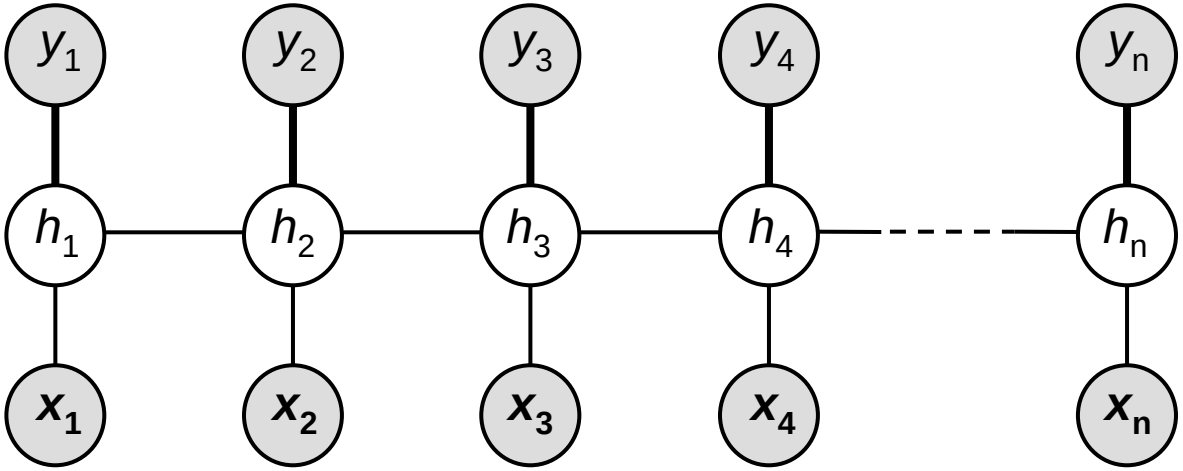
$$P(\boldsymbol{y} \mid \boldsymbol{x}, \theta) = \sum_{\boldsymbol{h} \in \mathcal{H}^m} P(\boldsymbol{y} \mid \boldsymbol{h}, \theta)\, P(\boldsymbol{h} \mid \boldsymbol{x}, \theta)$$

where $\boldsymbol{h} = [h_1, h_2, h_3, \ldots, h_n]$

USC

ICT INSTITUTE FOR CREATIVE TECHNOLOGIES

Latent-Dynamic CRF [CVPR 2007]



y_j : label

h_j : hidden state

$\boldsymbol{x}_j$: observations

$$P(\boldsymbol{y} \mid \boldsymbol{x}, \theta) = \sum_{\boldsymbol{h}:\forall h_j \in \mathcal{H}_{y_j}} P(\boldsymbol{h} \mid \boldsymbol{x}, \theta) = \sum_{\boldsymbol{h}:\forall h_j \in \mathcal{H}_{y_j}} \frac{1}{Z(x,\theta)} \exp\left(\sum_k \theta_k \mathbf{F}_k(\boldsymbol{h}, \boldsymbol{x}) \right)$$

Same as CRF

USC

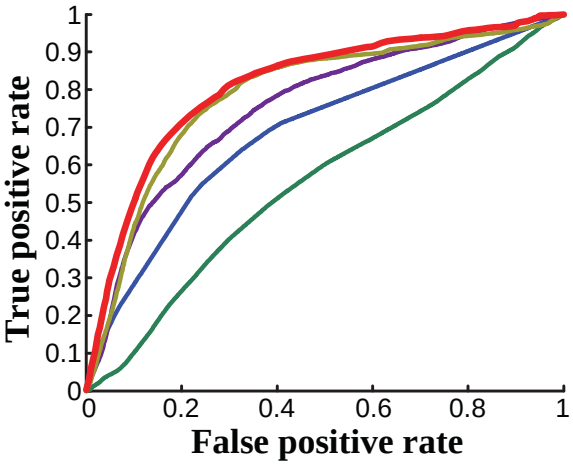
ICT INSTITUTE FOR CREATIVE TECHNOLOGIES

Head and Eye Gaze Estimation

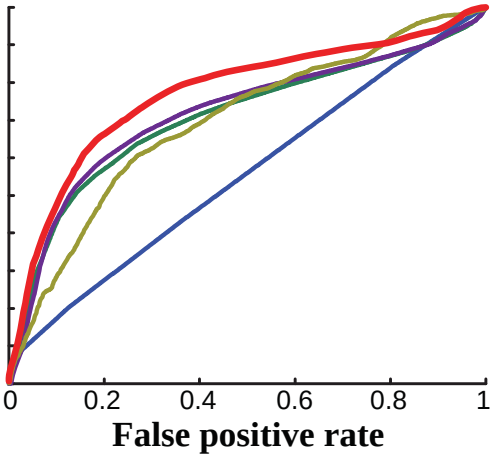
- Head gaze tracking
  [CVPR2003]
  - Adaptive View-based Appearance Model
- Eye gaze estimation
  [ICMI 2006]
  - View-based eigenspaces from 18 participants looking at 35 targets

Vision-based Recognition using LDCRF

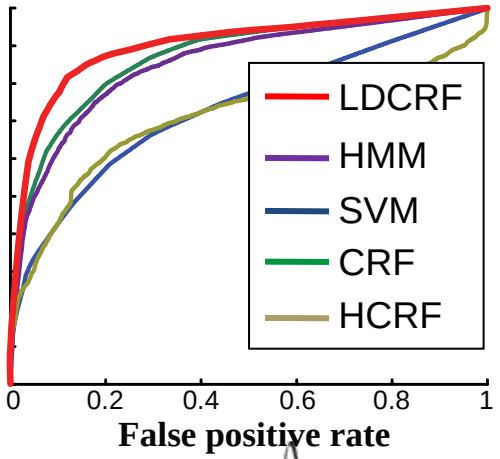
Head nods



Head nods



Gaze aversion



USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Research Agenda

Nonverbal Behaviors for interactive interfaces

- Natural and useful feedback
- Recognition algorithms
- Contextual information

Vision-based Recognition

Input Images

Head nod
or look down

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Visually Ambiguous Gestures

Head nod

Look down

Context-based Recognition

Input Images

Context

Head nod or look down

Head nod

Head nod or head shake?

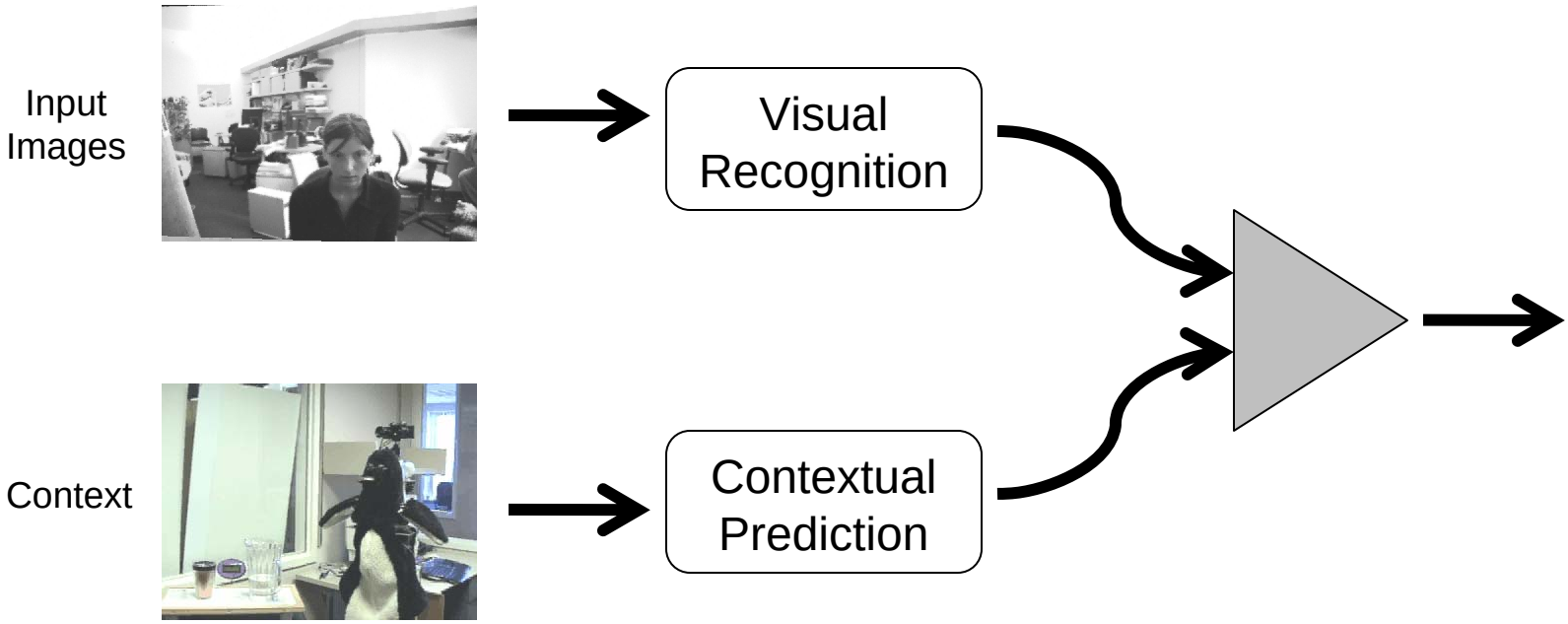
"By the way, do you know Paul?"

USC

ICT INSTITUTE FOR CREATIVE TECHNOLOGIES

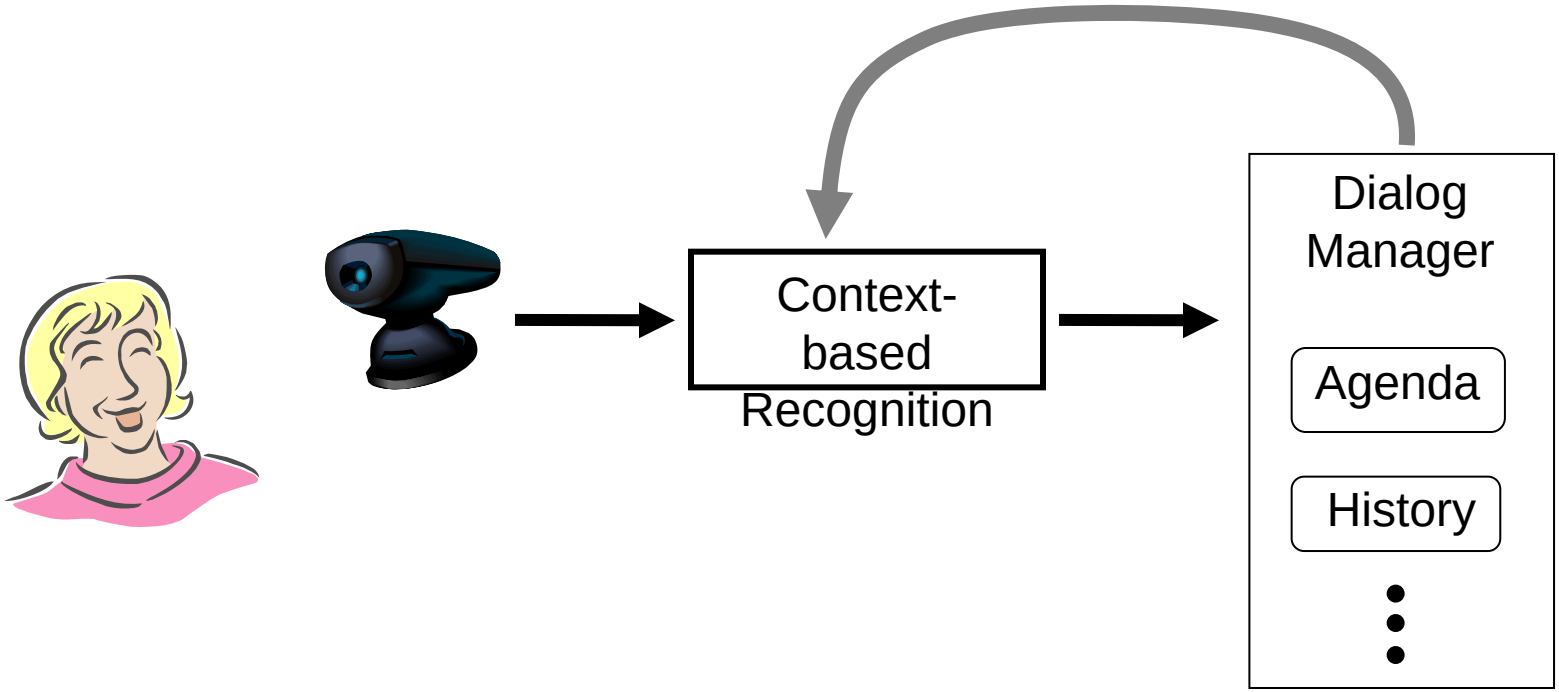
Understanding Nonverbal Behaviors

- **Use contextual knowledge from the interactive system to improve recognition of nonverbal behaviors**



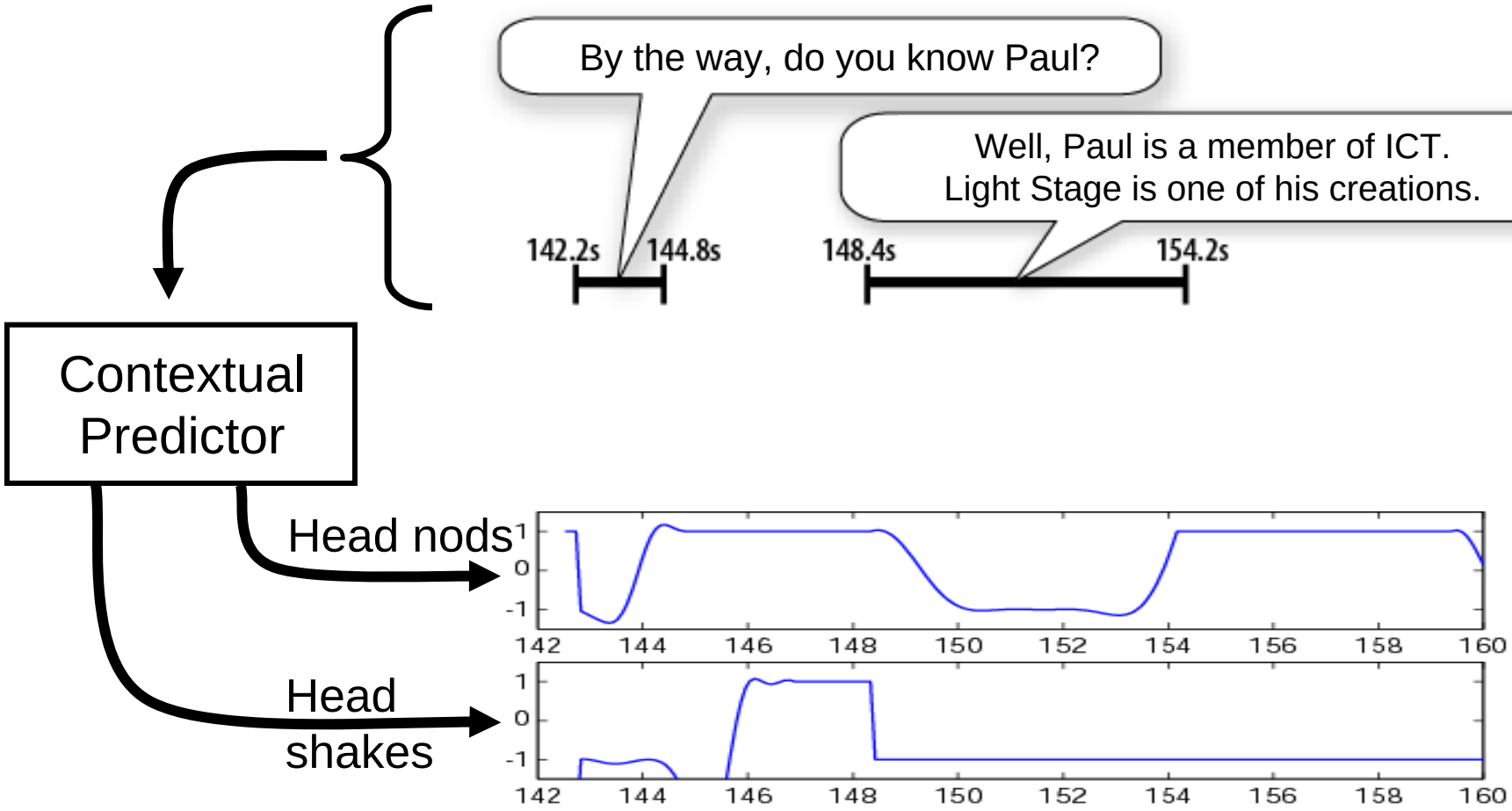
Dialogue Context

- Next spoken utterance
- Prosody/punctuation
- Avatar gestures



USC
ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Learning Context

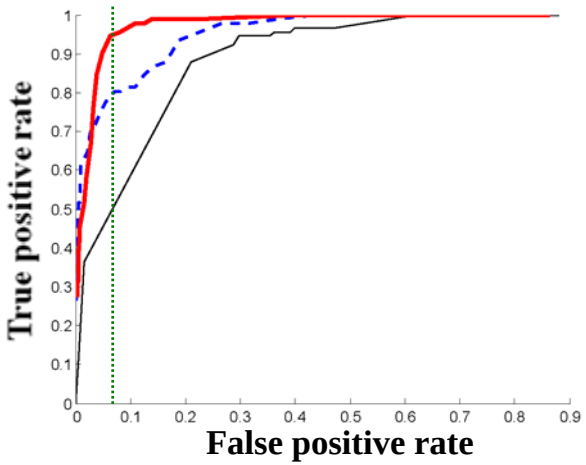


USC

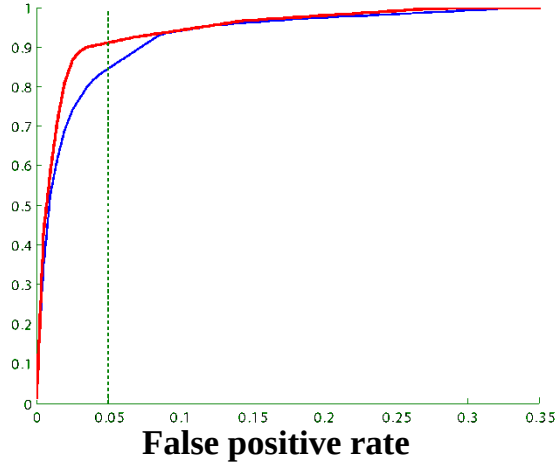
ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Context-based Recognition

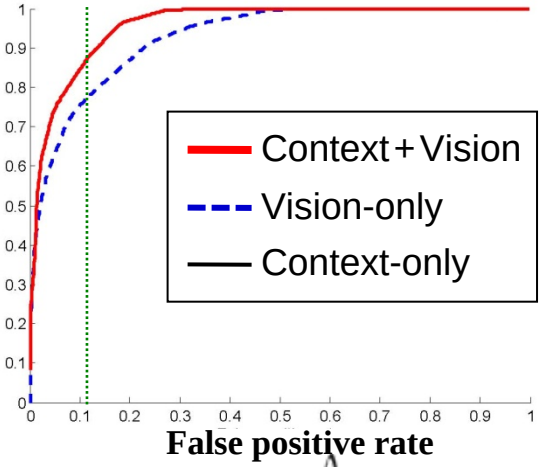
Head nods



Head nods



Gaze aversion



USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Contributions

Nonverbal Behaviors for interactive interfaces

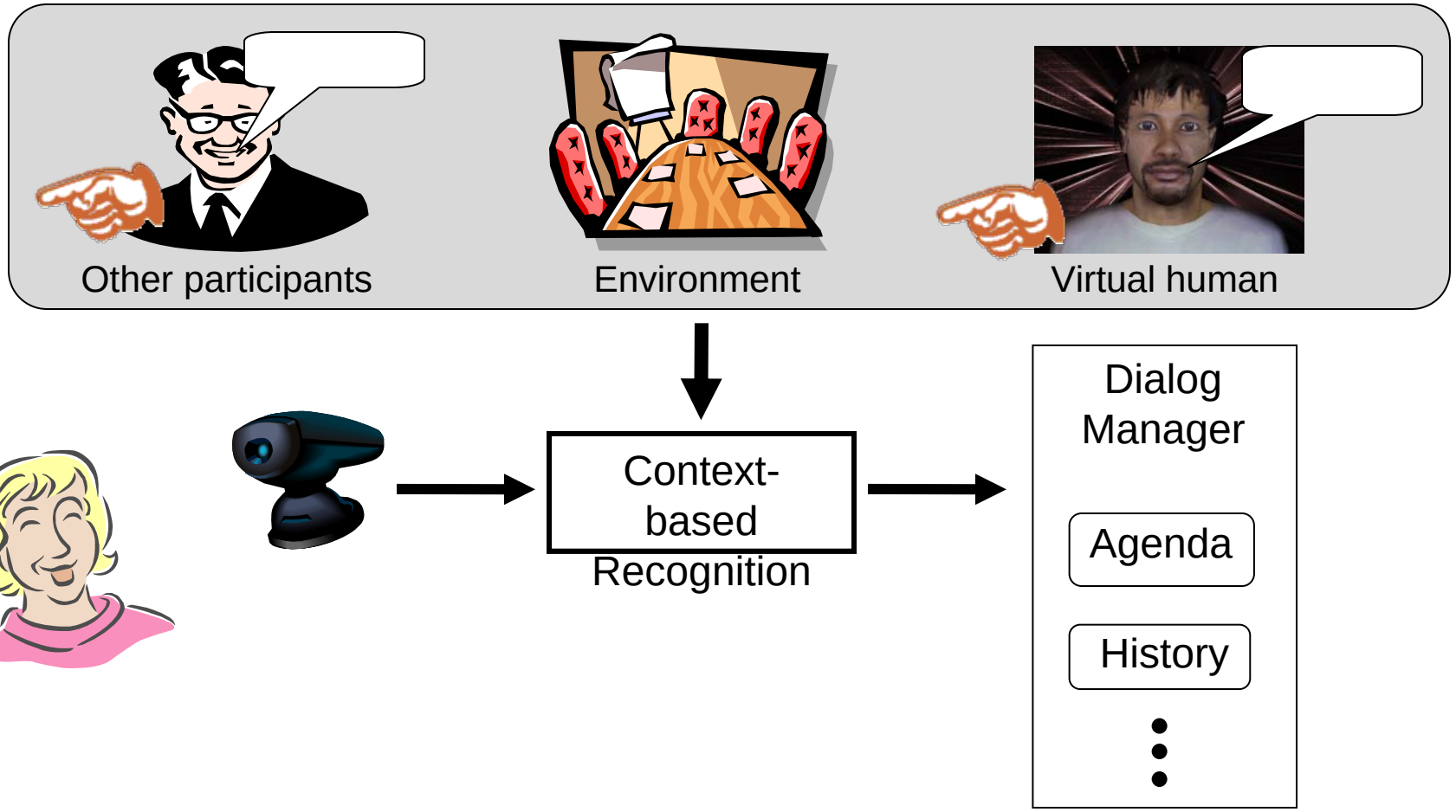
- Natural and useful feedback
  - User studies with head and eye gestures
- Recognition algorithms
  - New model for visual gesture recognition
- Contextual information
  - Context-based Visual Feedback Recognition

Future Directions

Nonverbal Behaviors for interactive interfaces

- Natural and useful feedback
  - Facial expressions, arm gestures and body posture
- Recognition algorithms
  - Shared latent-dynamic: multiple gestures and modalities
- Contextual information
  - Automatic of contextual features

Contextual Information



Collaborations

- **Watson, real-time nonverbal behavior recognition**
  - Head tracking, eye gaze estimation and gesture recognition
  - Downloaded by more than a hundred research groups
  - http://projects.ict.usc.edu/vision/watson/
- **hCRF library, discriminative dynamic models**
  - Matlab and C++ implementations of LDCRF, CRF and HCRF
  - More than 15 downloads every weeks
  - http://hcrf.sourceforge.net/

MIT Media Lab

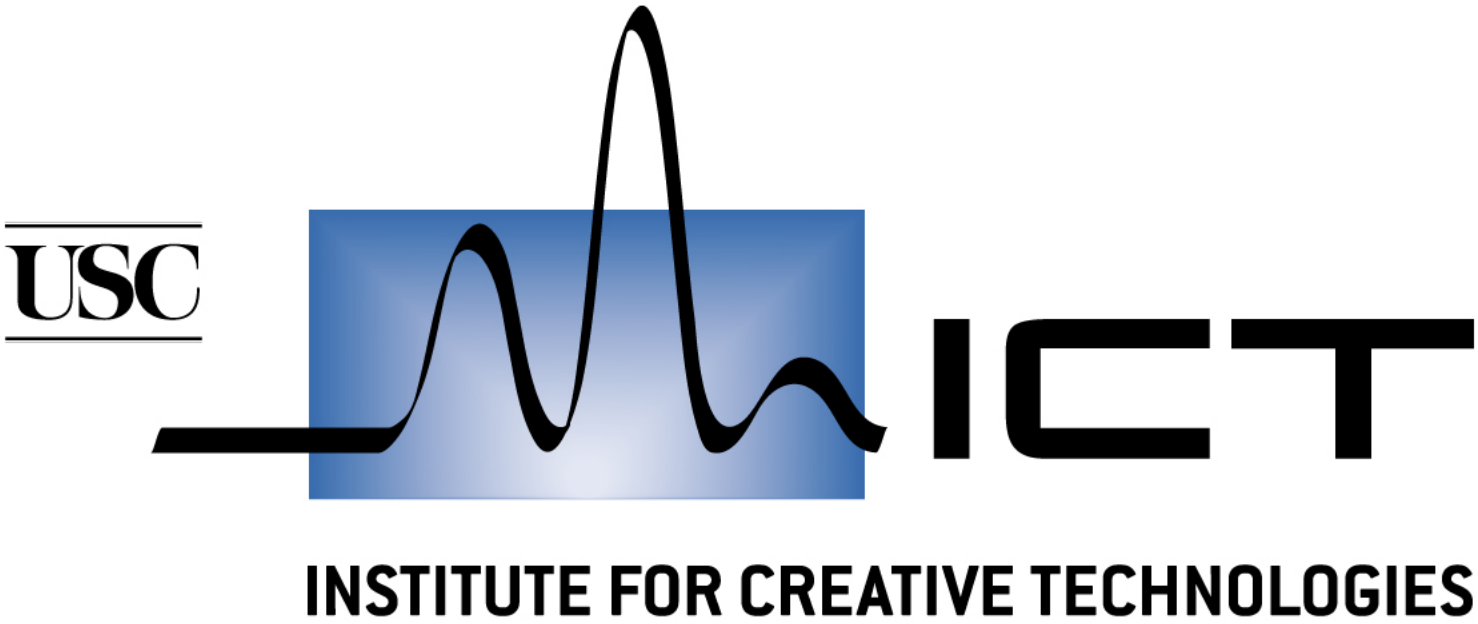
NTT

NU-MACK

ICT/USC

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES


USC
ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES