# The Human Speechome Project

Brandon Roy
MIT Media Lab

ASRU 2007

# The Human Speechome Project

## MIT Media Lab

*Cognitive Machines Group*

Prof. Deb Roy

Philip Decamp

Brandon Roy

Jethran Guinness

Rony Kubat

## Northeastern University

*Communication Analysis and Design Laboratory*

Prof. Rupal Patel

Alexia Salata

# HUMAN SPEECHOME PROJECT

# HUMAN SPEECHOME PROJECT

Goal: Advance our understanding of how children acquire language in natural contexts

# HUMAN SPEECHOME PROJECT

Goal: Advance our understanding of how children acquire language in natural contexts

Approach: Longitudinal, ultra-dense, in vivo recordings + data mining and behavioral modeling

# HUMAN SPEECHOME PROJECT

Goal: Advance our understanding of how children acquire language in natural contexts

Approach: Longitudinal, ultra-dense, in vivo recordings + data mining and behavioral modeling

Differentiators: Two orders of magnitude more behavioral data than previous studies, far fewer observer effects, new analysis tools

# HUMAN SPEECHOME PROJECT

Goal: Advance our understanding of how children acquire language in natural contexts

Approach: Longitudinal, ultra-dense, in vivo recordings + data mining and behavioral modeling

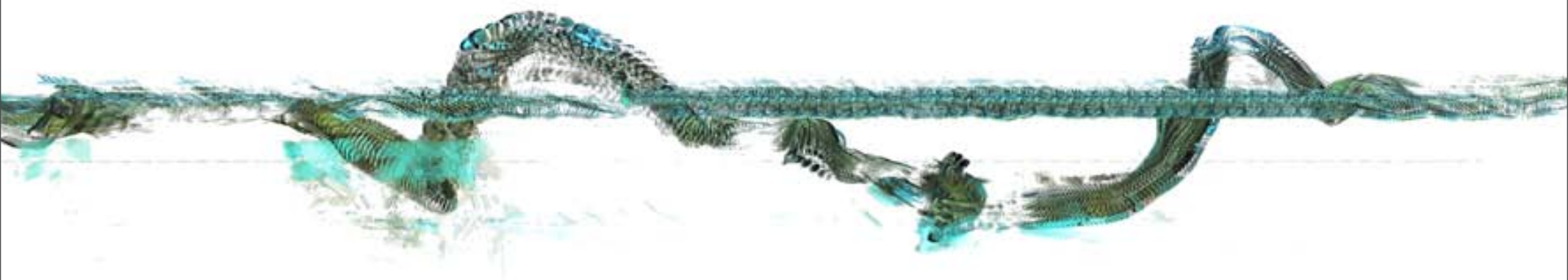
Differentiators: Two orders of magnitude more behavioral data than previous studies, far fewer observer effects, new analysis tools

Impact: Illuminate language acquisition, behavioral phenotyping, video search, smart homes, parenting aids, security, retail,...

# Speech in the home

# Speechome

# DATA COLLECTION

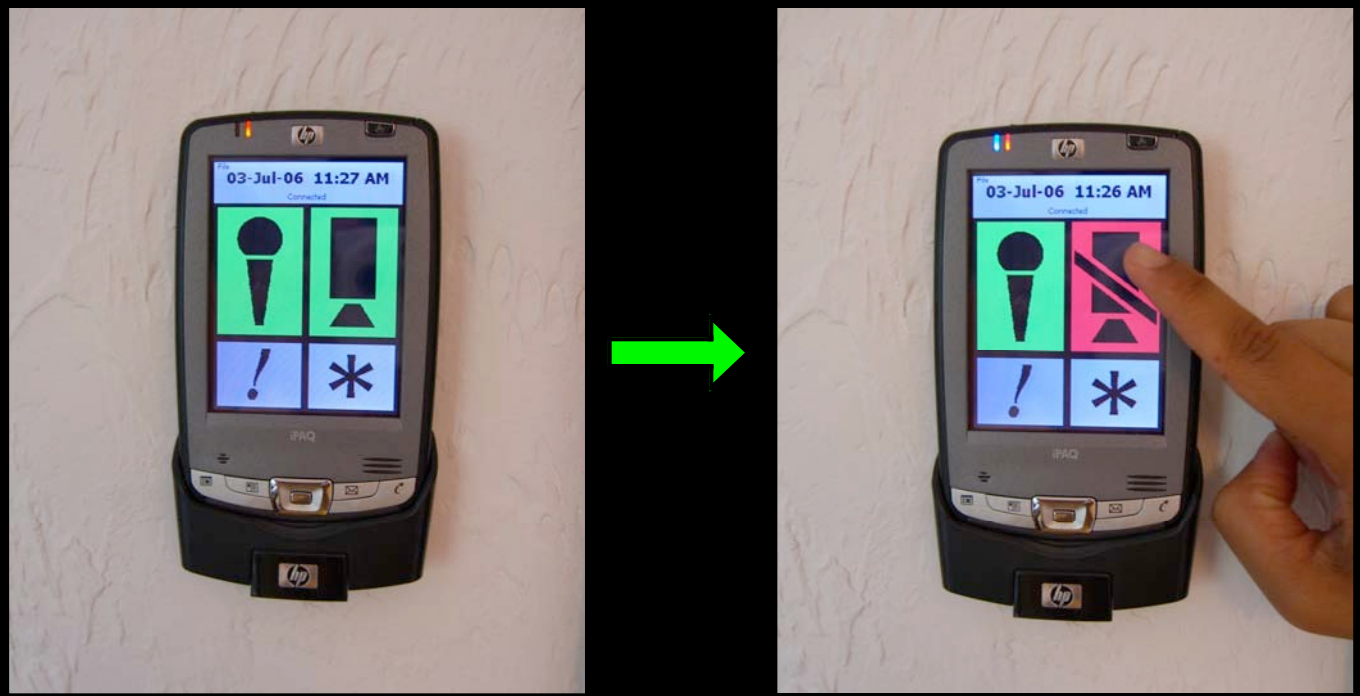


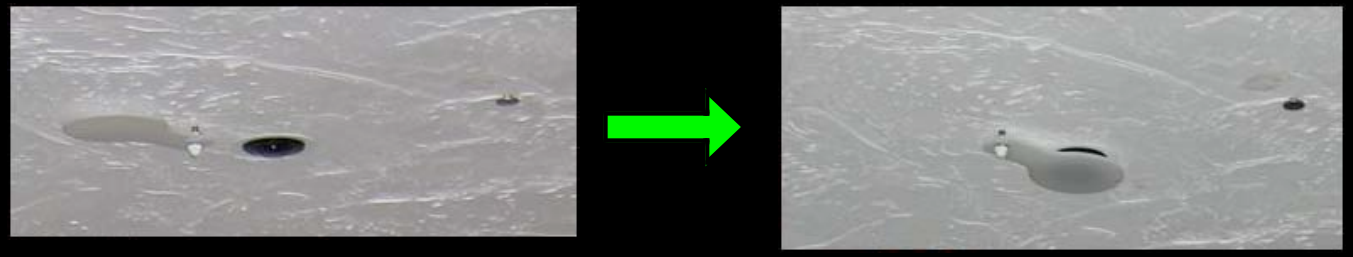
A house in the Boston area

# Looking up at the ceiling

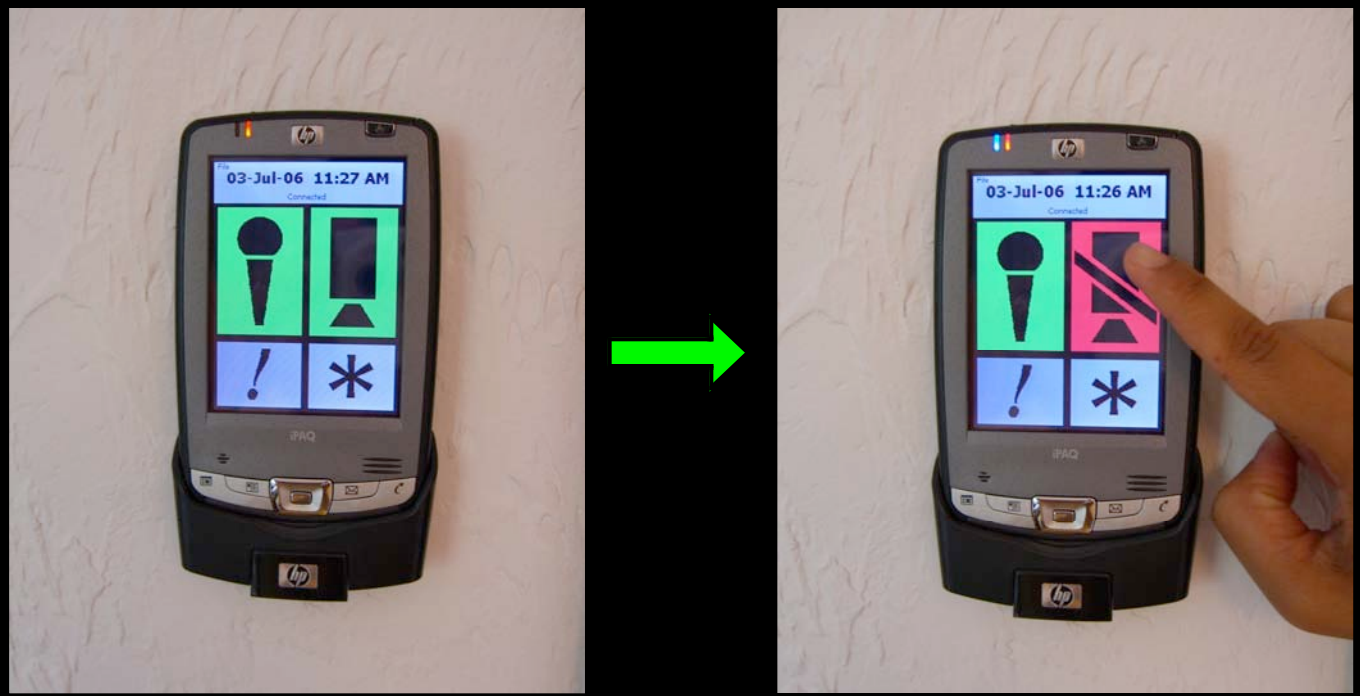
# Looking up at the ceiling

# Recording control system

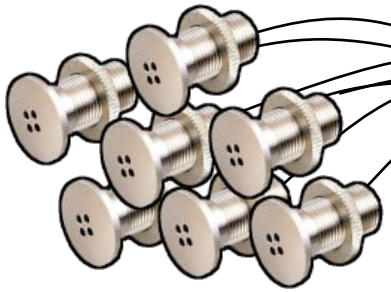




# Recording control system

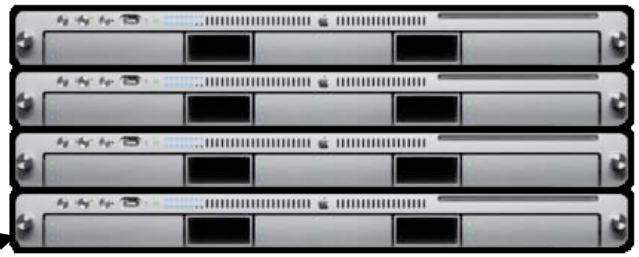


# Recording Infrastructure



14 Microphones

Preamps



Audio/Video Processing

Ethernet

11 Cameras

GbS Ethernet

Control Server

11 PDAs

Wireless Hubs

5TB Raid

# Storage at the Media Lab

250,000 GB capacity

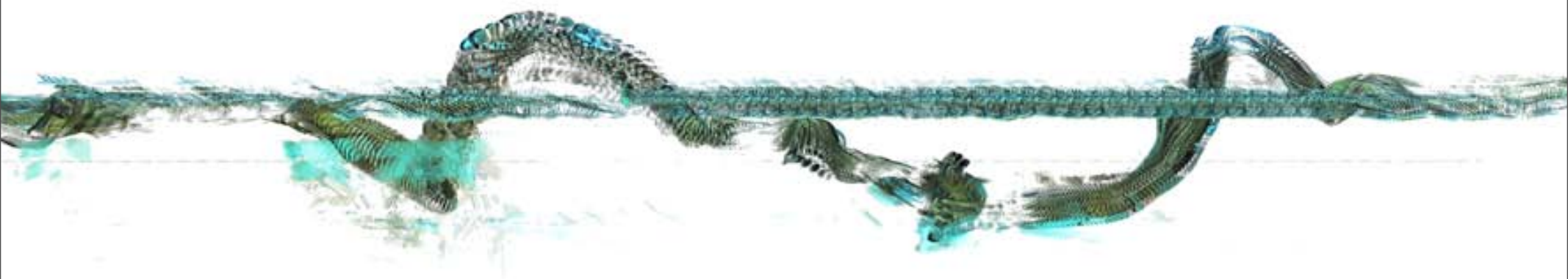
80,000 hrs video

120,000 hrs audio

~2 yrs of recording

48 KHz audio,
1 MP video,
~15 fps

# DATA ANALYSIS



# Tracing the Birth of a Word

# Tracing the Birth of a Word

Transcribe all speech heard and produced by child

# Tracing the Birth of a Word

Transcribe all speech heard and produced by child

Annotate video surrounding all uses of target word

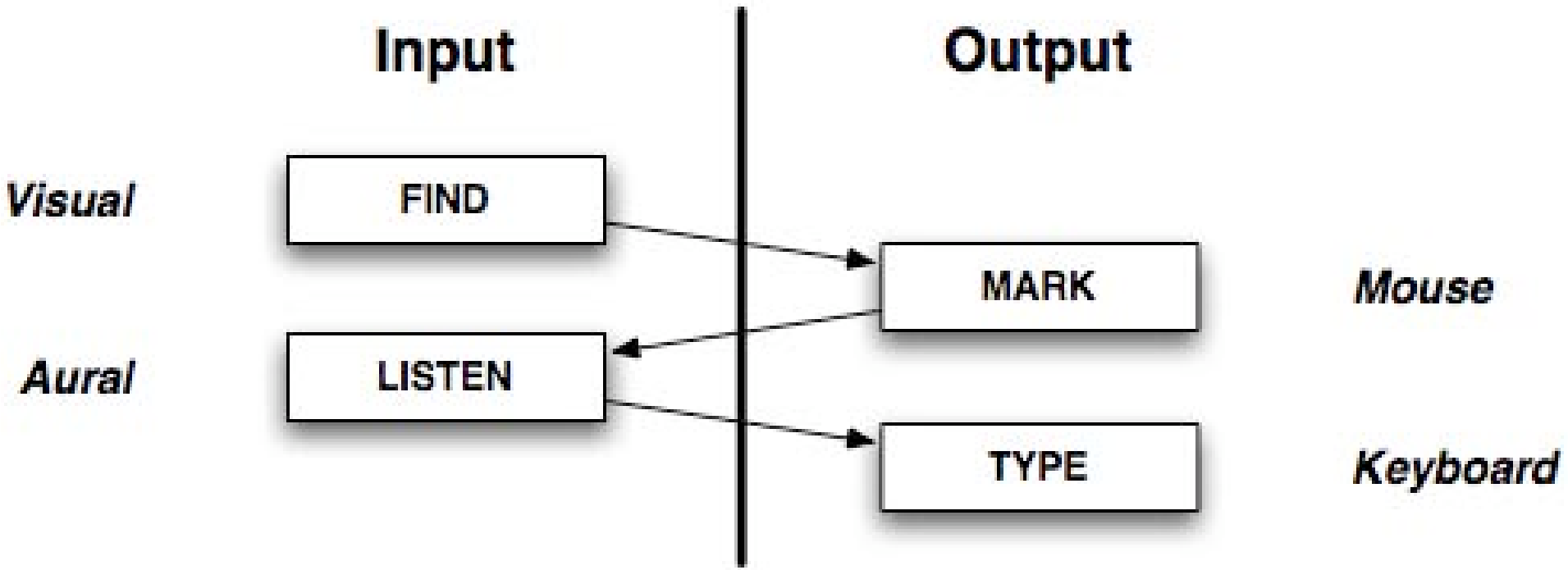
# Tracing the Birth of a Word

Transcribe all speech heard and produced by child

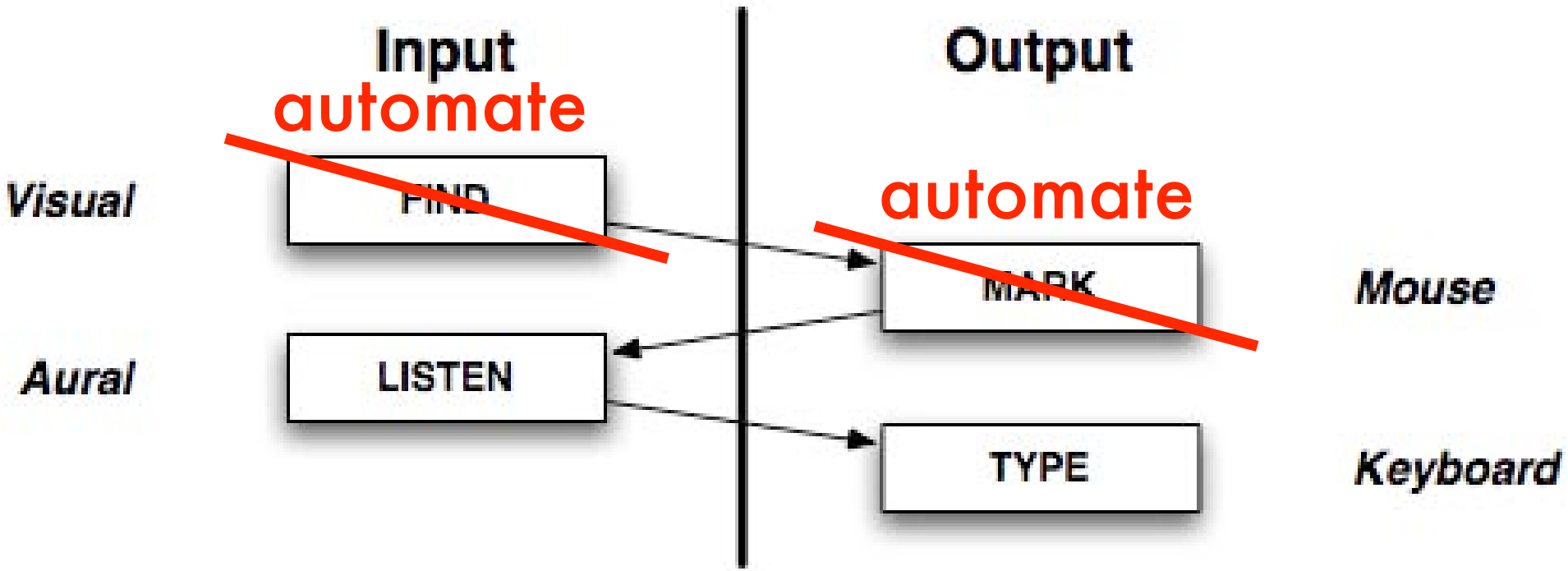
Annotate video surrounding all uses of target word

Analyze role of contextual factors in word learning

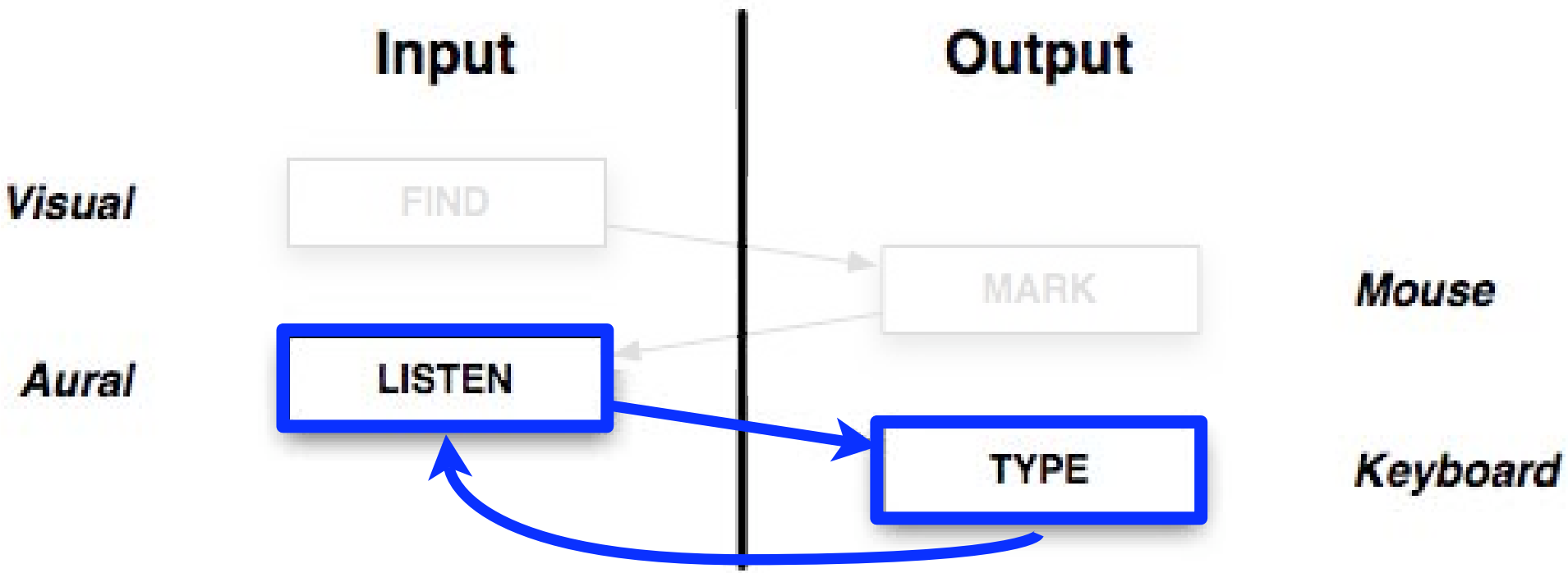
# Speech Transcription

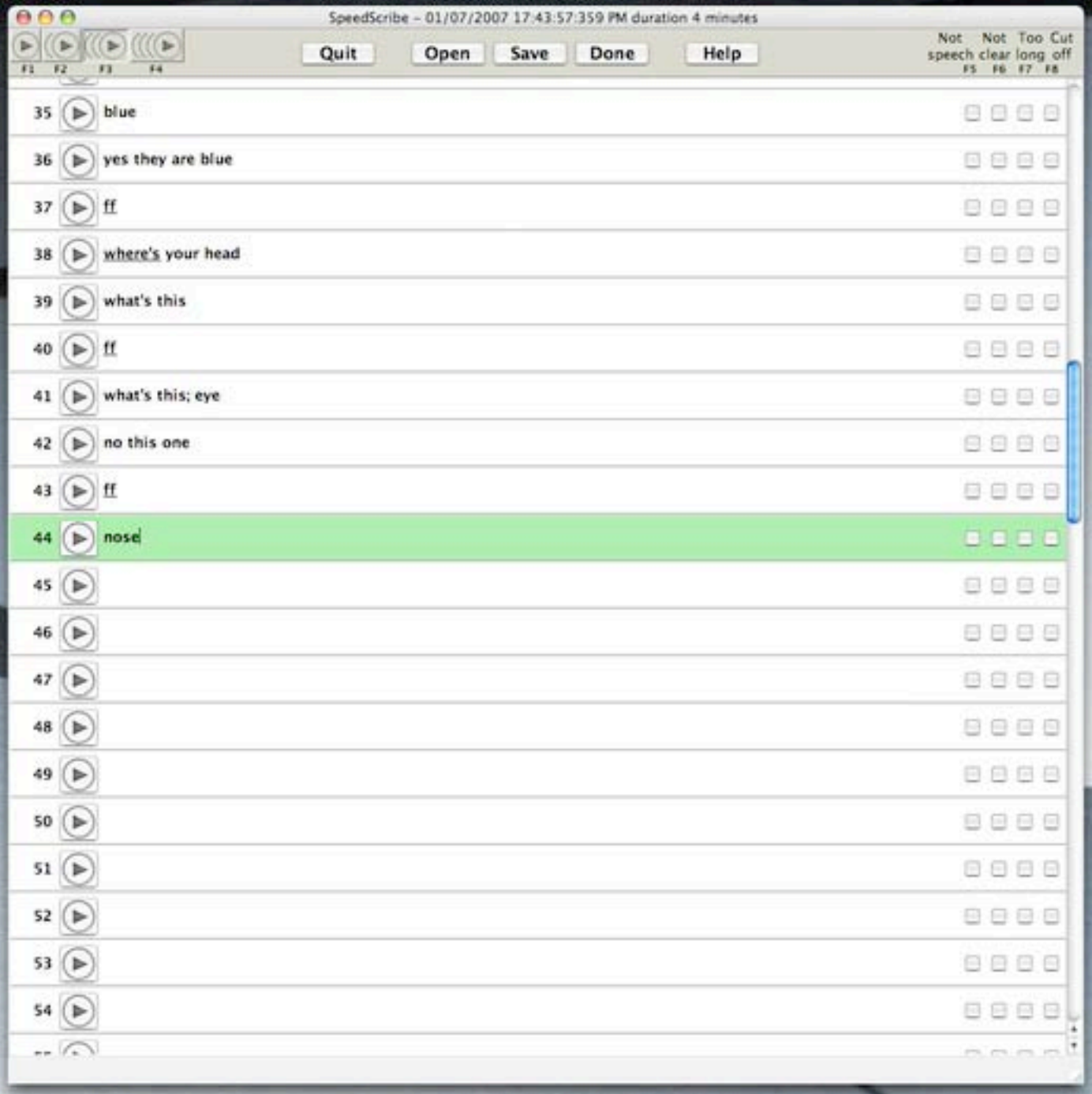


# Semi-automatic speech transcription



# Semi-automatic speech transcription




SpeedScribe – 01/07/2007 17:43:57:359 PM duration 4 minutes
F1
F2
F3
F4
Quit
Open
Save
Done
Help
Not speech F5
Not clear F6
Too long F7
Cut off F8
35 blue
36 yes they are blue
37 ff
38 where's your head
39 what's this
40 ff
41 what's this; eye
42 no this one
43 ff
44 nose
45
46
47
48
49
50
51
52
53
54

Evolution of “water”

# Video Annotation

# Video Annotation

Tracking algorithms to follow people throughout the house

# Video Annotation

Tracking algorithms to follow people throughout the house

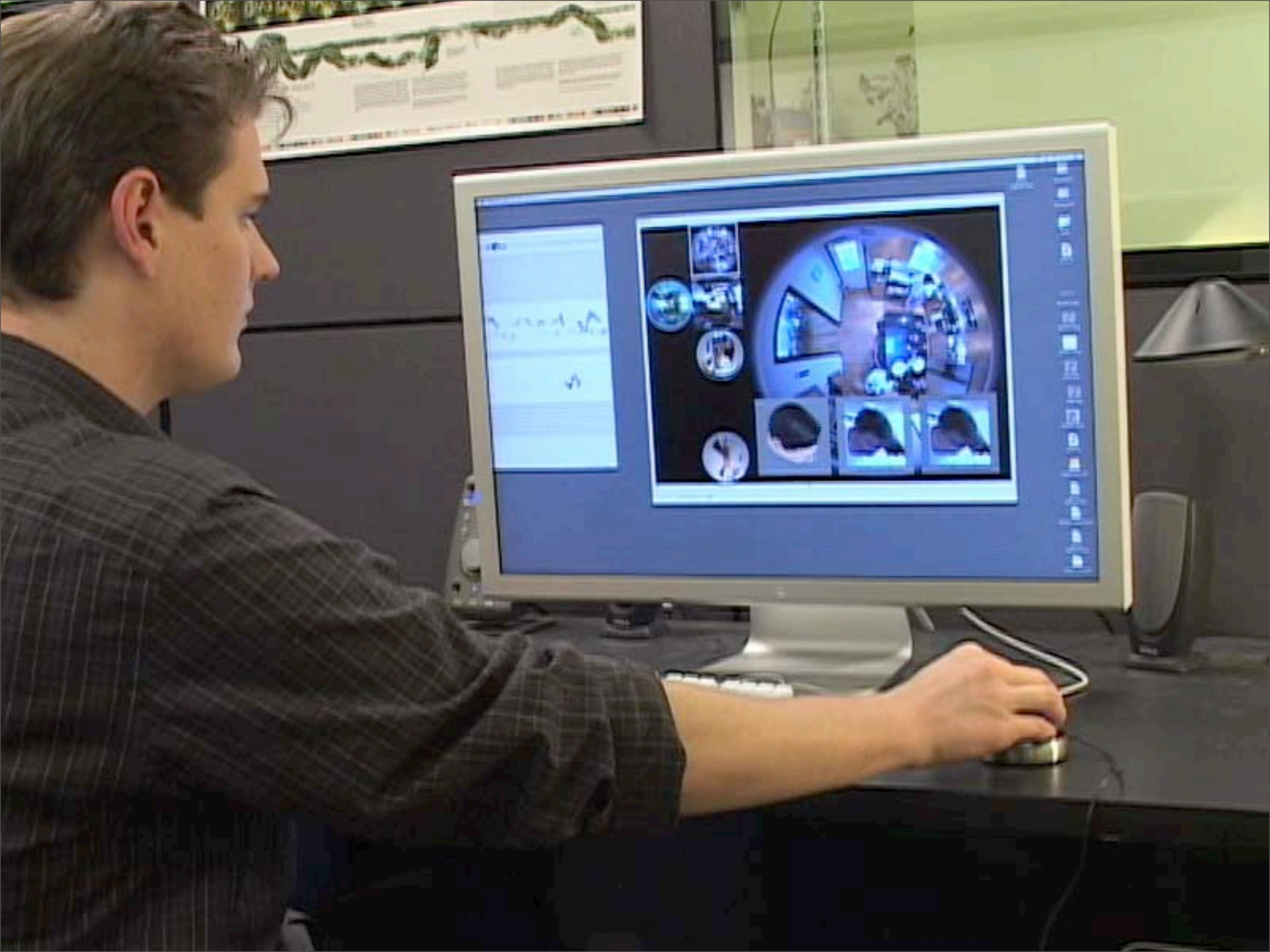
Estimate head pose to help determine focus of attention

# Video Annotation

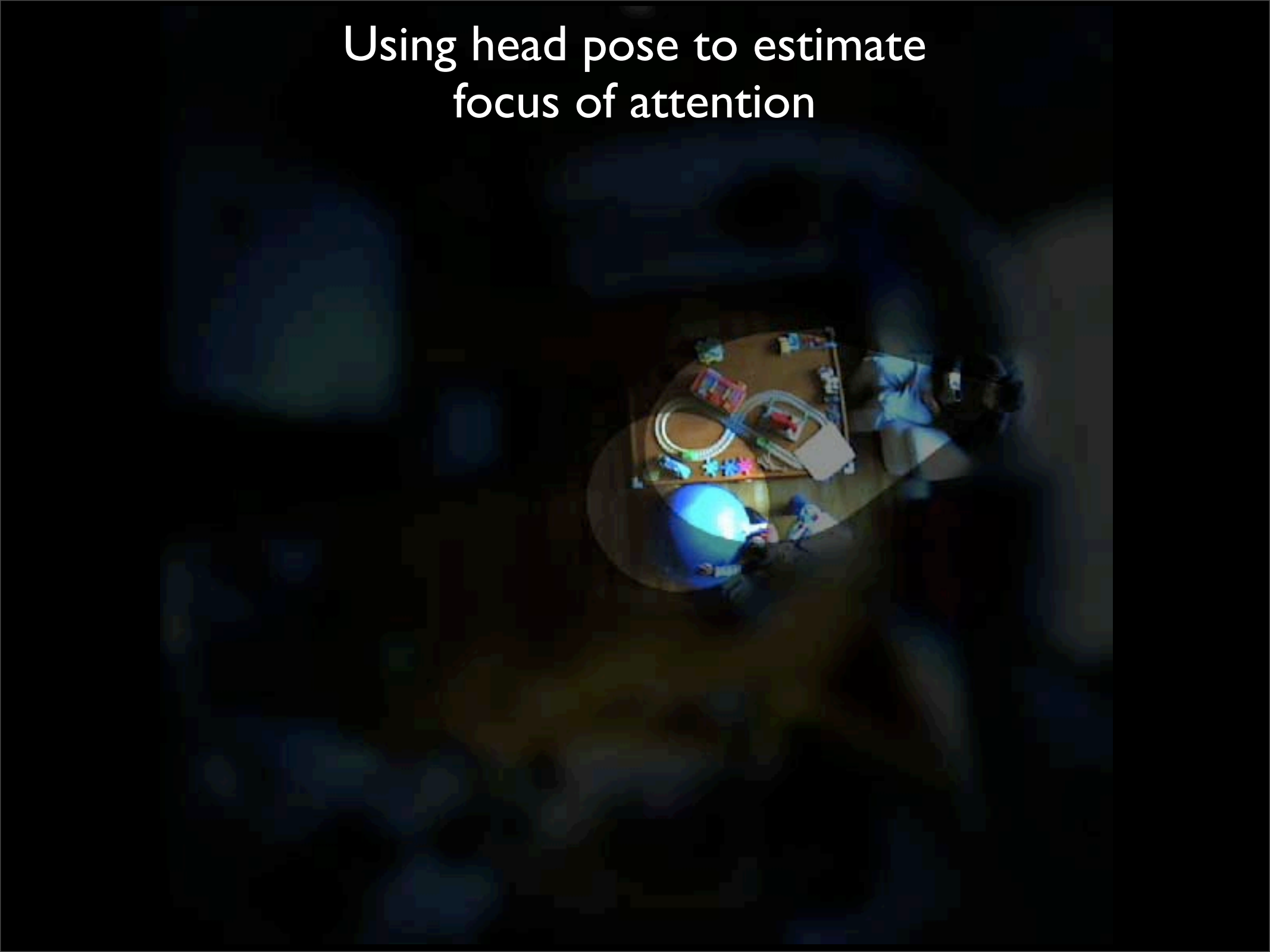
Tracking algorithms to follow people throughout the house

Estimate head pose to help determine focus of attention

Combine automatic and manual methods for tracking and head pose annotation



# Using head pose to estimate focus of attention



# Summary

# Summary

Build a system to make it easy to collect lots of multimodal data

# Summary

Build a system to make it easy to collect lots of multimodal data

Interesting speech and language modeling challenges

# Summary

Build a system to make it easy to collect lots of multimodal data

Interesting speech and language modeling challenges

Issues of privacy management, making the system cheaper, smarter recording strategies...

# Summary

Build a system to make it easy to collect lots of multimodal data

Interesting speech and language modeling challenges

Issues of privacy management, making the system cheaper, smarter recording strategies...

How to scale to N > 1?

# Summary

Build a system to make it easy to collect lots of multimodal data

Interesting speech and language modeling challenges

Issues of privacy management, making the system cheaper, smarter recording strategies...

How to scale to N > 1?

...

# Summary

Build a system to make it easy to collect lots of multimodal data

Interesting speech and language modeling challenges

Issues of privacy management, making the system cheaper, smarter recording strategies...

How to scale to N > 1?

...

www.media.mit.edu/cogmac