MODELING SOCIAL CUES:
EFFECTIVE FEATURES FOR PREDICTING LISTENER NODS

FAISAL KHAN, BILGE MUTLU, XIAOJIN ZHU | DEPARTMENT OF COMPUTER SCIENCES | UNIVERSITY OF WISCONSIN–MADISON

RESEARCH GOALS

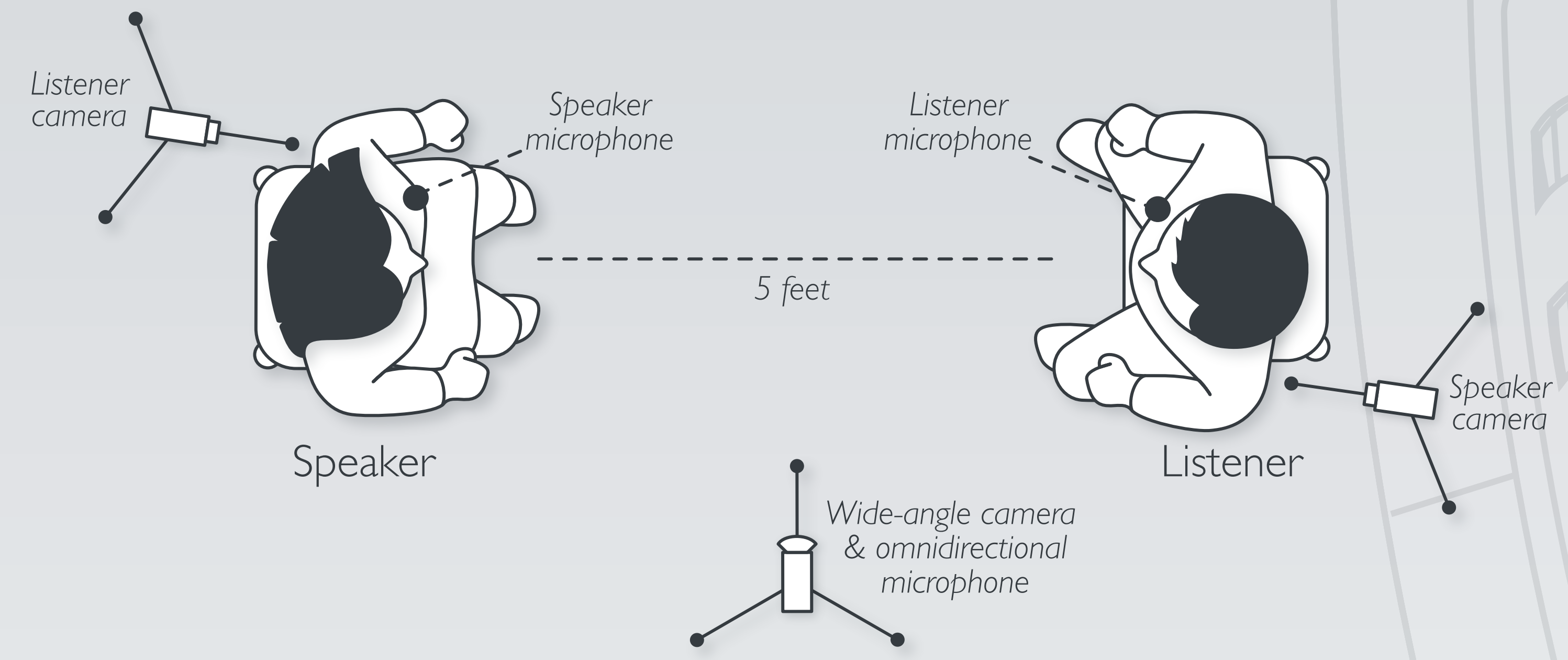
Gaining a computational understanding of human social behavior

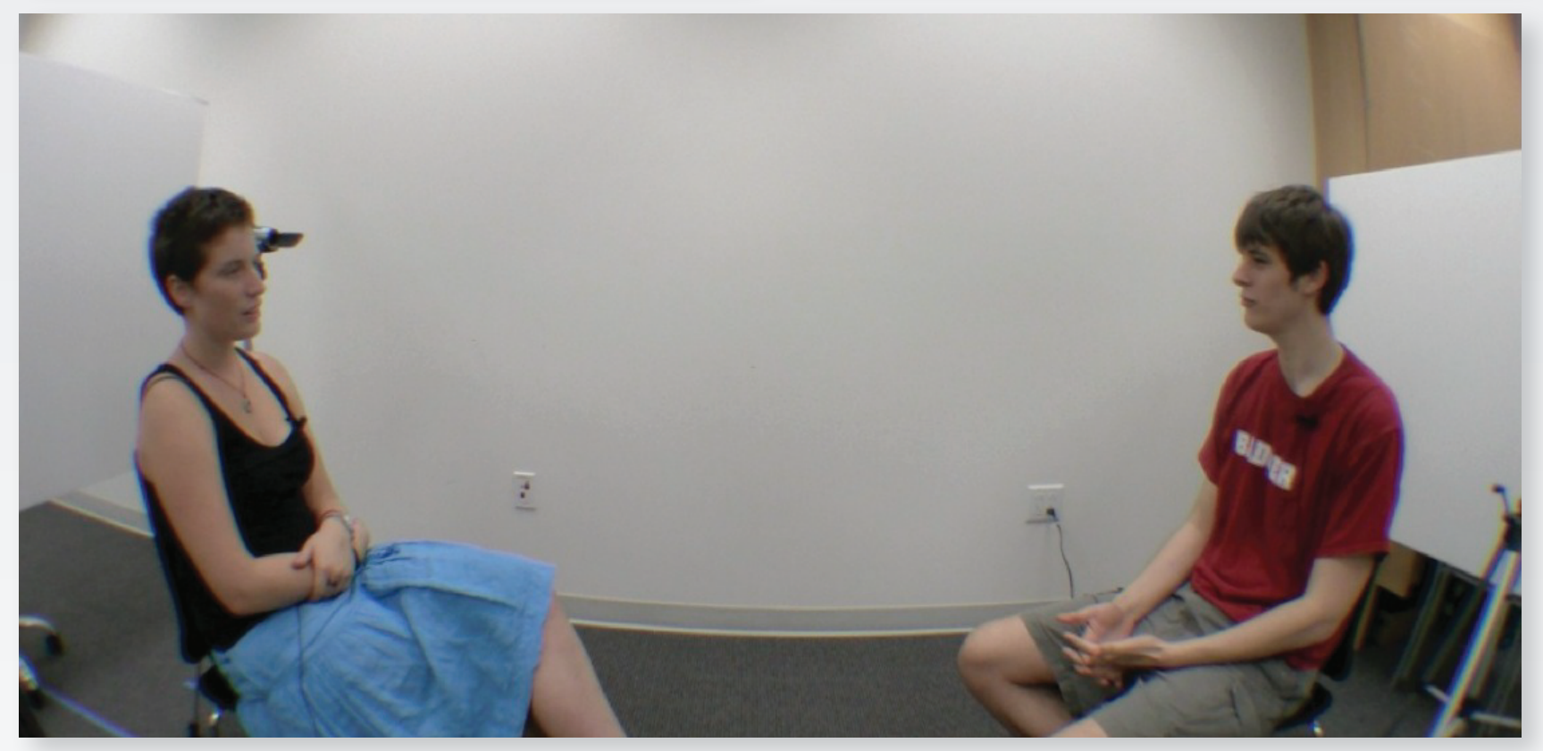
Building socially interactive systems such as agents and robots

Current study explores:

Using of a small set of real-time features to predict listener nods

DATA COLLECTION SETUP





Data collection with 24 dyads

Equal number of MM, FM, MF, and FF gender combinations

Perform a "storytelling task"

Participants were paid $10

SPONSORS

The National Science Foundation Award IIS-1017952 and an equipment loan from Mitsubishi Heavy Industries supported this research.

RAW FEATURES

A "raw" set of features extracted automatically from multimodal data

- Speech segmentation (speech/pause)
- Speaker classification (speaker/listener)
- Pitch values and slopes (rising/falling intonation)
- Speaker head movements including *nodding*

DERIVED FEATURES

Temporal dependencies between raw of features and listener nods captured by a derived set of features based on multiple windows of averages of raw features and differences across window averages

Final feature vector for a given frame:

$$\mathbf{f}_i = [\ \mathbf{r}_i \quad \mathbf{g}_i^1 \quad . \quad . \quad . \quad \mathbf{g}_i^7 \quad \mathbf{h}_i^1 \quad . \quad . \quad . \quad \mathbf{h}_i^7\ ]'$$

Where

$$\mathbf{r}_i = [\ speech \ \ speaker \ \ head_x \ \ head_y \ \ nodding \ \ pitch \ \ s_1 \ . \ . \ . \ s_9\ ]'$$

$$\mathbf{g}_i^m = \frac{1}{2^m} \sum_{k=0}^{2^m-1} \mathbf{r}_{i-k} \ \text{ and } \ \mathbf{h}_i^m = \mathbf{g}_i^m - \mathbf{g}_{i-2^m}^m$$

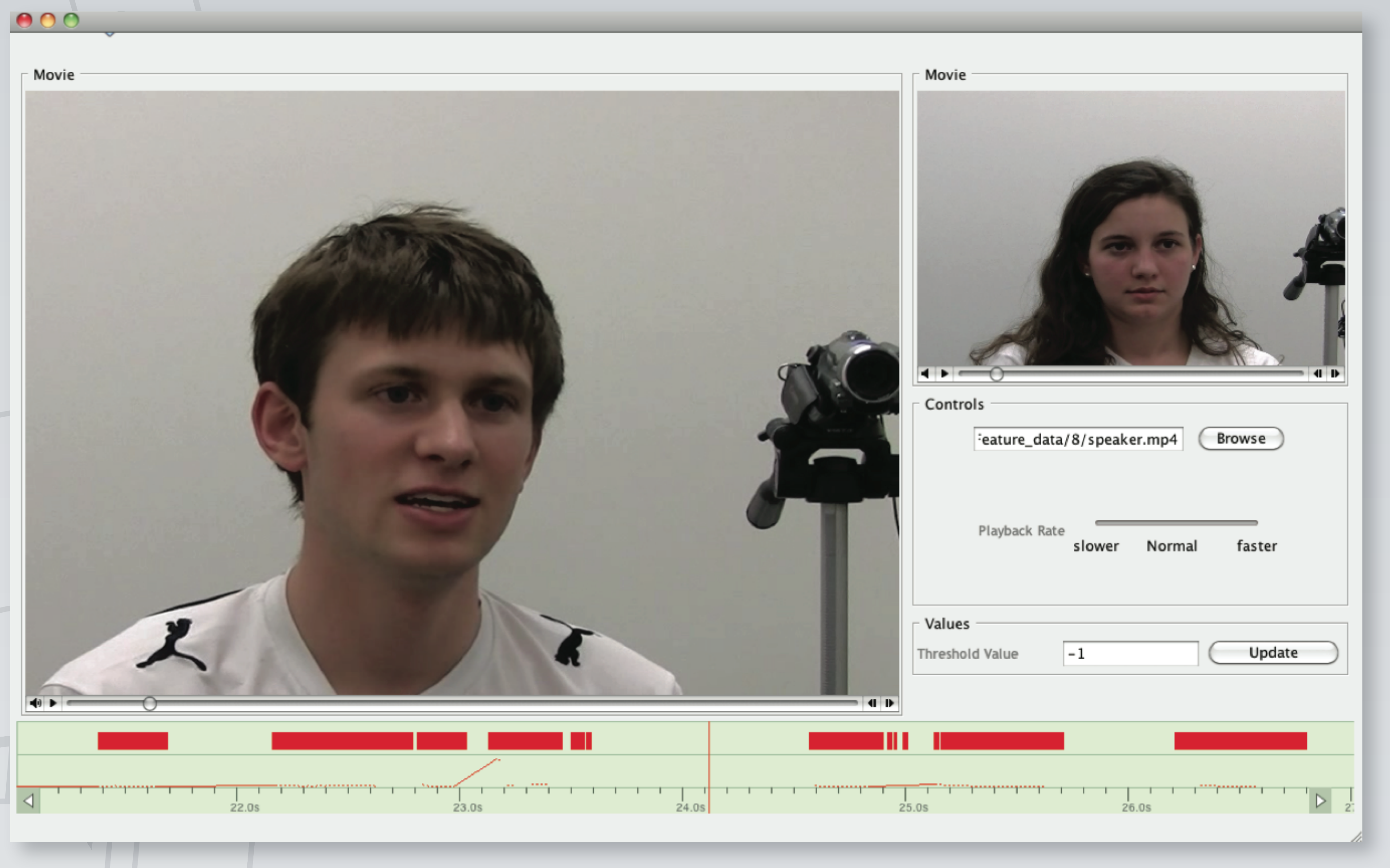
CLASSIFICATION RESULTS

Predictions using a Support Vector Machine (SVN) classifier

Four-fold cross validation

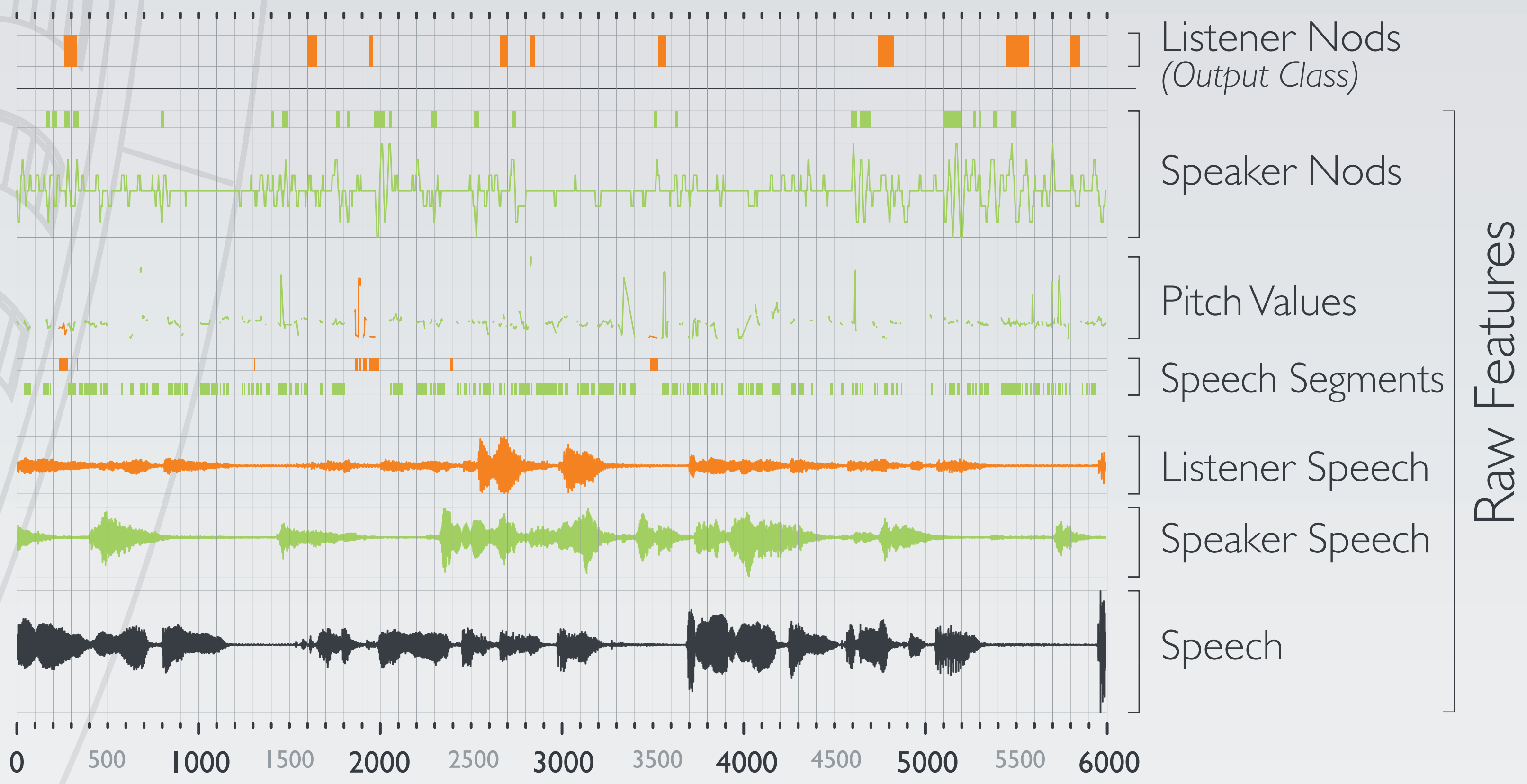
Precision	=	**0.1083**
Recall	=	**0.3165**
F-measure	=	**0.1605**

DATA ANNOTATION & VISUALIZATION



eature_data/8/speaker.mp4

Threshold Value -1



NEXT STEPS

Improving the modeling of temporal dependencies using:

- Encoding templates (Morency et al., 2010)
- Sequential models (e.g., CRF, HMM)

Using model predictions to control a robot's nods

Conducting human-robot interaction studies to test effectiveness