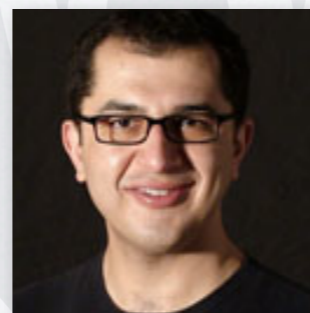# MODELING SOCIAL CUES

EFFECTIVE FEATURES FOR PREDICTING LISTENER NODS
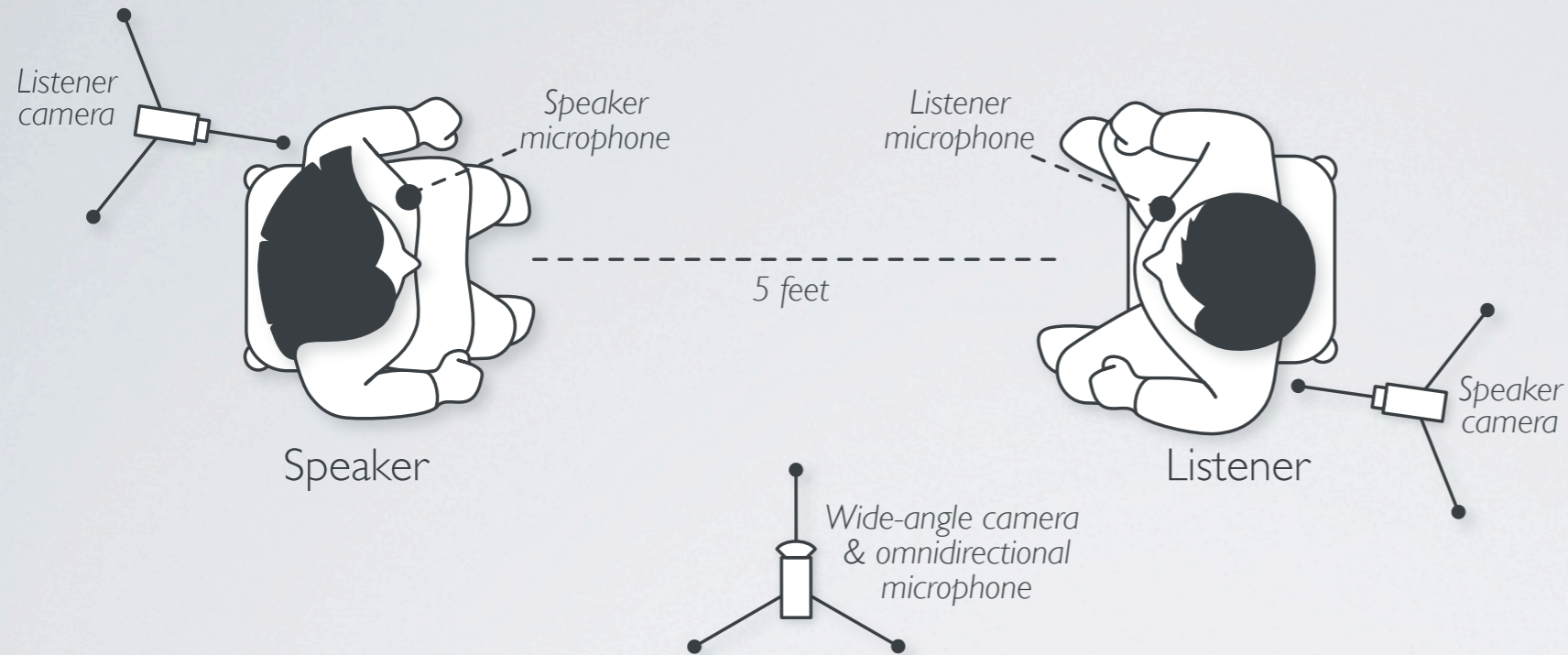
FAISAL KHAN   BILGE MUTLU   JERRY ZHU

Department of Computer Sciences, University of Wisconsin–Madison

NIPS 2010 Workshop on Modeling Human Communication Dynamics

December 10, 2010

DATA COLLECTION SETUP

DATASET

Data collection with 24 dyads

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

Perform a "storytelling task" for 3 minutes

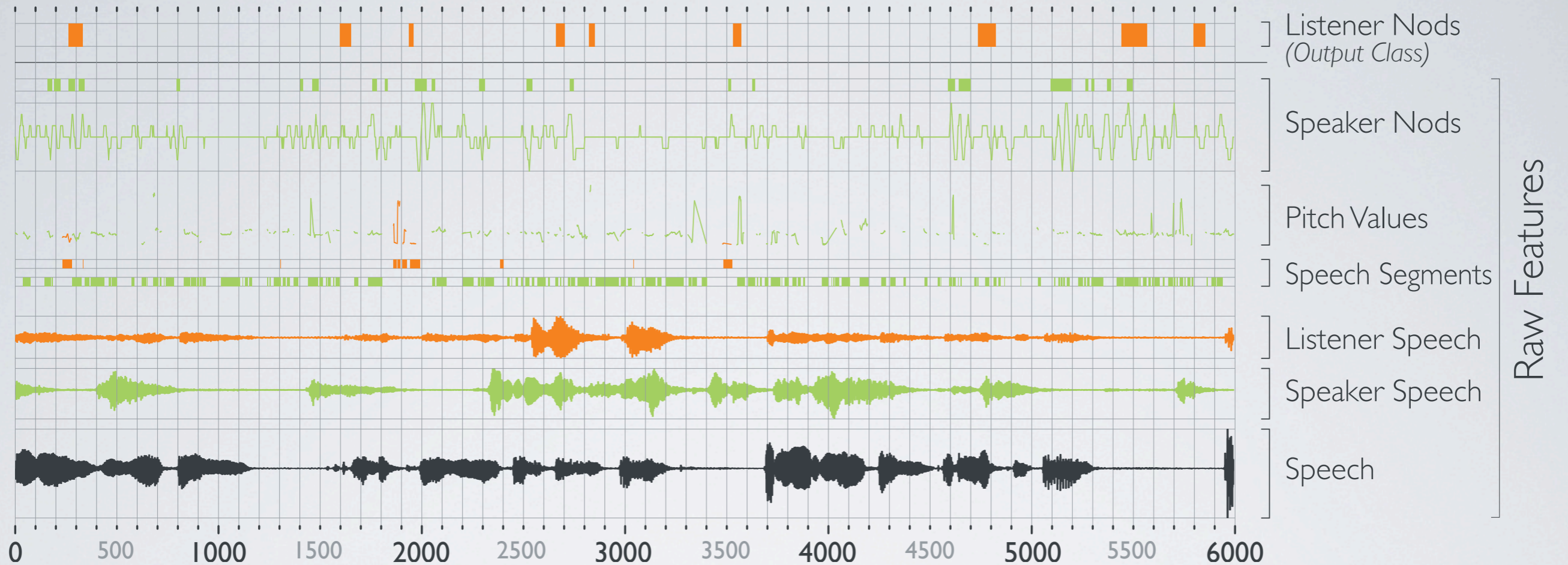Four hours and 52 minutes of multimodal data (audio/video)

THE UNIVERSITY of WISCONSIN MADISON

RAW FEATURES

- Listener Nods *(Output Class)*
- Speaker Nods
- Pitch Values
- Speech Segments
- Listener Speech
- Speaker Speech
- Speech

Raw Features

DERIVED FEATURES

$$\mathbf{r}_i = \begin{bmatrix} speech & speaker & head_x & head_y & nodding & pitch & s_1 & \ldots & s_9 \end{bmatrix}'$$

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

$$\mathbf{f}_i = \begin{bmatrix} \mathbf{r}_i & \mathbf{g}_i^1 & \ldots & \mathbf{g}_i^7 & \mathbf{h}_i^1 & \ldots & \mathbf{h}_i^7 \end{bmatrix}'$$

PREDICTION RESULTS

Support Vector Machine (SVN) predictions

Four-fold cross validation

Precision   =   **0.1083**

Recall   =   **0.3165**

F-measure   =   **0.1605**