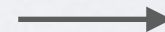
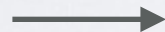


# SOUNDPOUND

**Input:**  
video



**Output:**  
video + sound



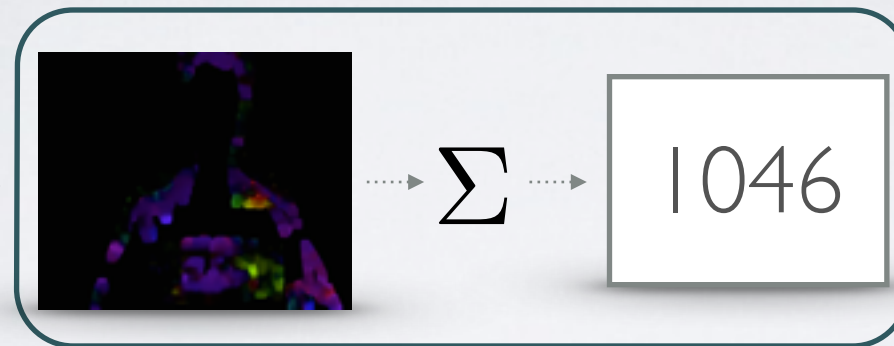
David Abel  
CS295I-B  
12/17/14

# PIPELINE OVERVIEW

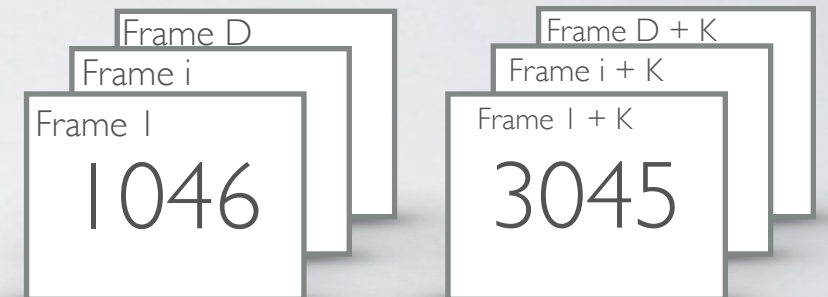
**Input:**  
video



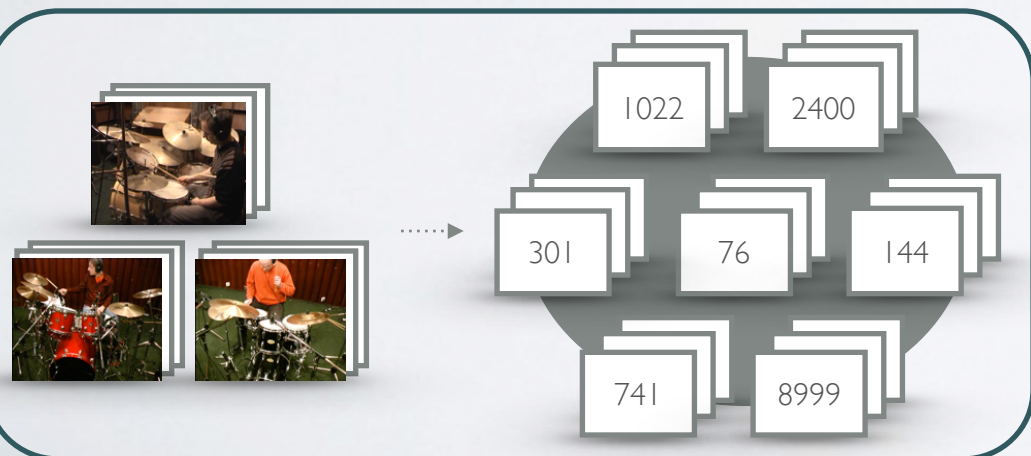
Featurize  
(optical flow + pooling)



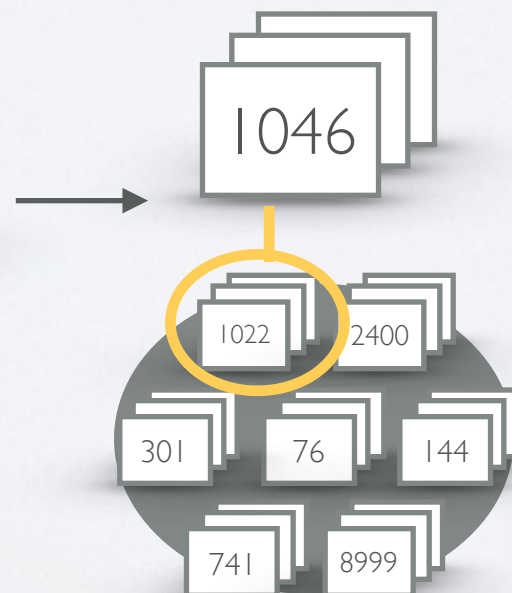
Group frames into  
Segments



Turn dataset into segments



Match each input segment  
with dataset (NN)



Add sound from each  
matched segment



**Output:**  
video + sound



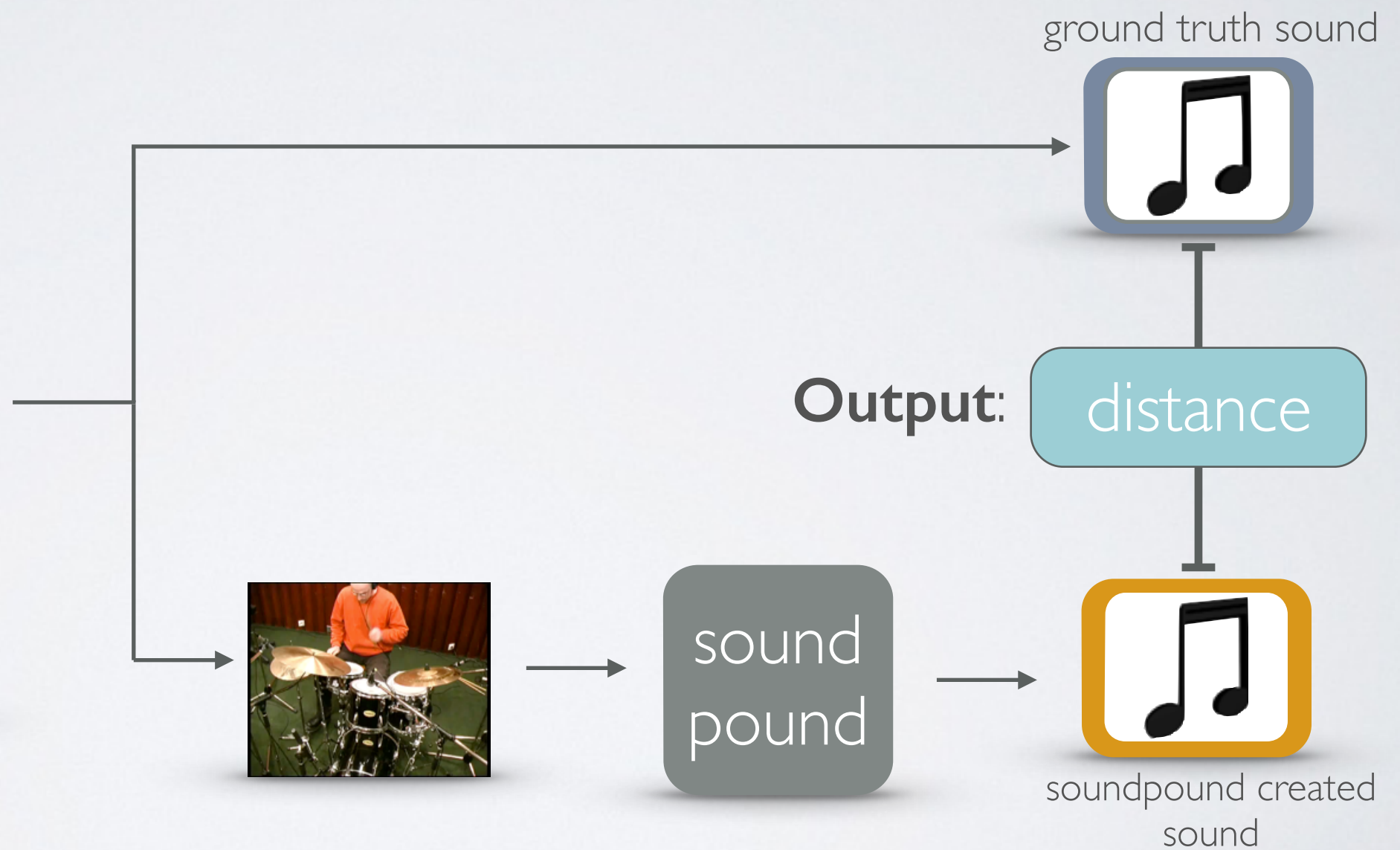
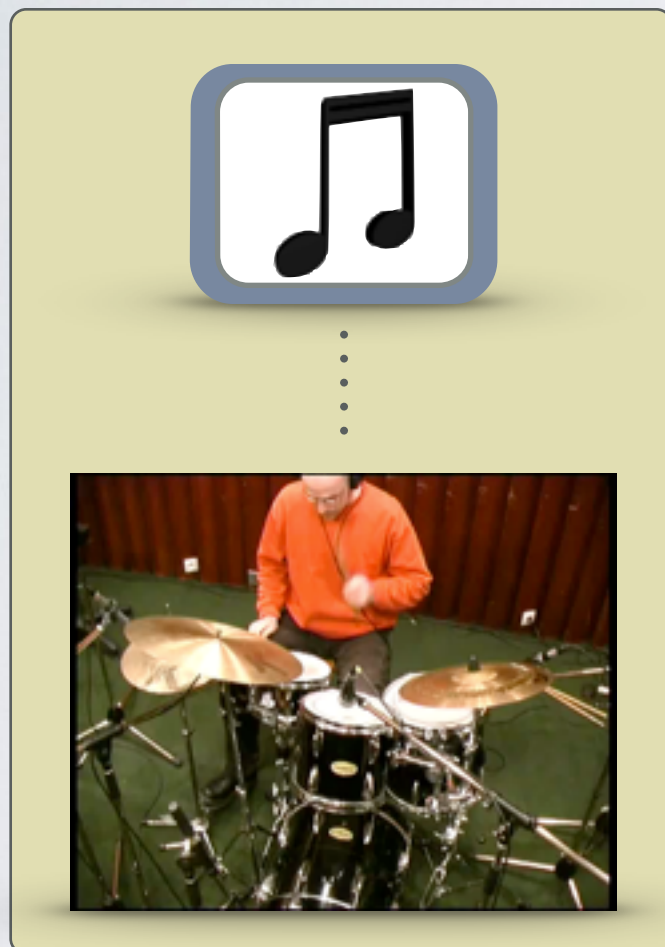
# CHANGES

1. Quantitative Evaluation Framework
2. Feature Selection
3. Hyperparameter Optimization
4. Dataset Modifications



# EVALUATION

**Input:** video +  
ground truth sound



# DISTANCE

1. **MSE:** Mean Squared Error in audio **per quanta**
2. **MTE:** Mean Temporal Error - accounts for playing a *nearly* correct sound a bit too soon/too late.

# FEATURE SELECTION

- Average Magnitude
- Max Magnitude
- Sum of Magnitudes
- Max Magnitude Vector's Angle
- Weighted Average of Optical Flow Vectors

# PERFORMANCE: MTE

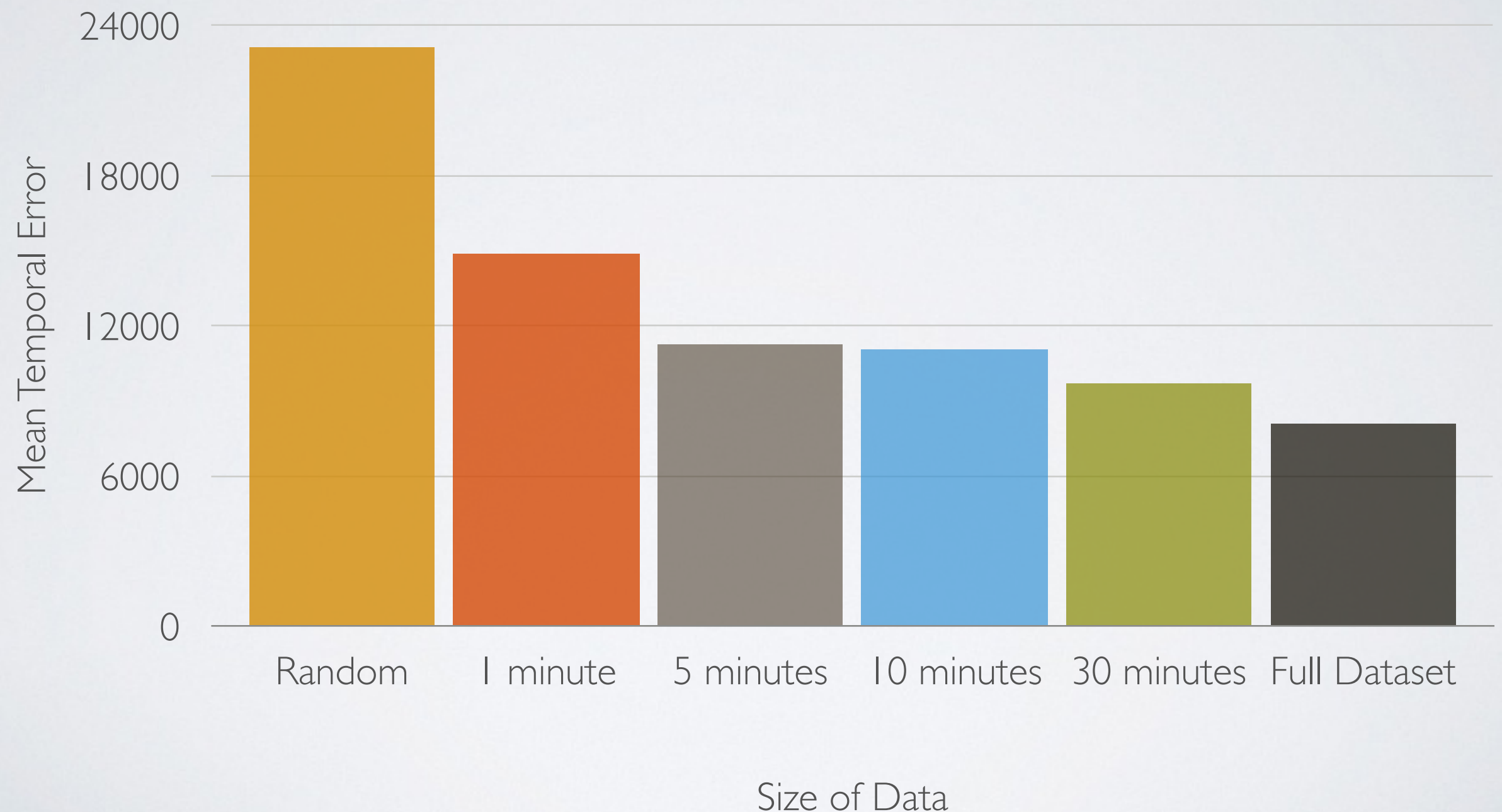
Comparison of Feature Types





# PERFORMANCE: DATA SIZE

Average Pooling with Varied Dataset Size





# PRELIMINARY RESULTS



# NEW RESULTS I





# NEW RESULTS II



# CONTRIBUTIONS

- Developed SOUNDPOUND. Works in around 20 seconds for a 10 second video.
- Designed and implemented quantitative evaluation framework, including new distance metric for sound.
- Performed feature selection across possible representations.

