



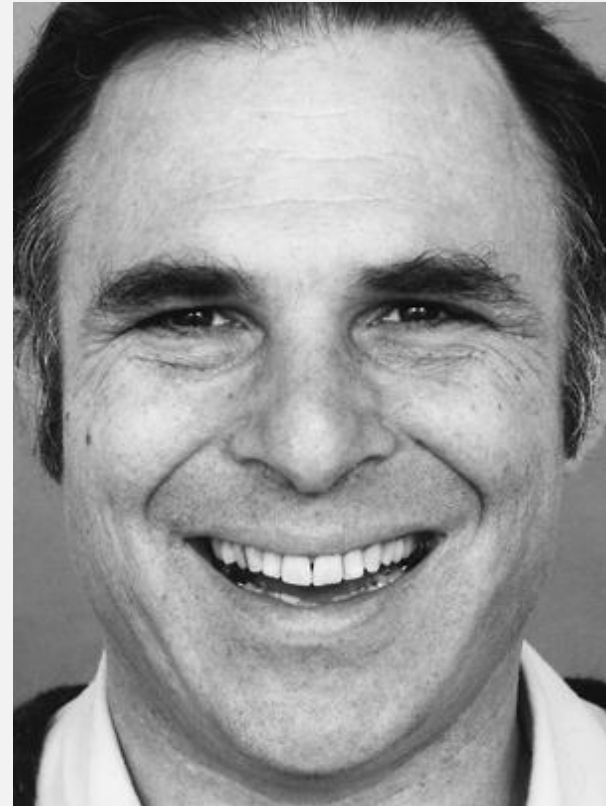
Criteria and metrics for thresholded AU detection

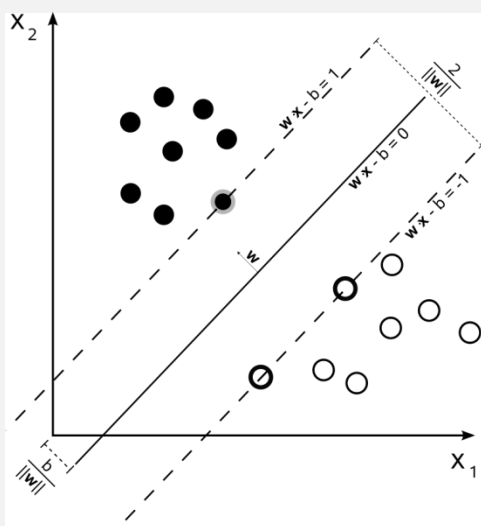
Jeff Girard and Jeff Cohn
University of Pittsburgh

BeFIT Workshop, ICCV 2011



- Facial action informs:
 - Emotion Displays
 - Pain Displays
 - Social Signaling
- FACS Coding System:
 - Anatomically-based
 - Common vocabulary
 - Objective and reliable
 - Action Unit 6+12+25+26

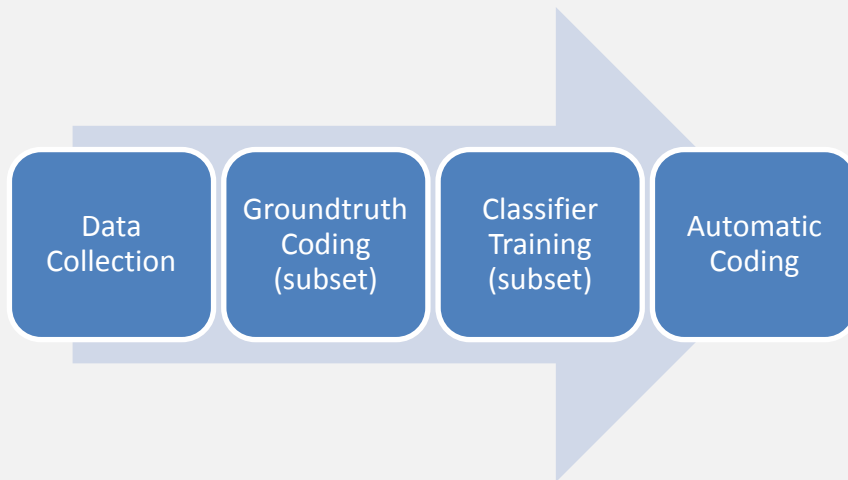




- FACS training, reliability, and manual coding are time consuming
- Automatic AU detection would be faster and enable “online” coding
- However, automatic coding requires a trained classifier
- Classifier training requires time, expertise, and ground truth coding
- Classifiers are considered corpus-specific; thus a new classifier is usually trained for each corpus



Novel Classifier Training



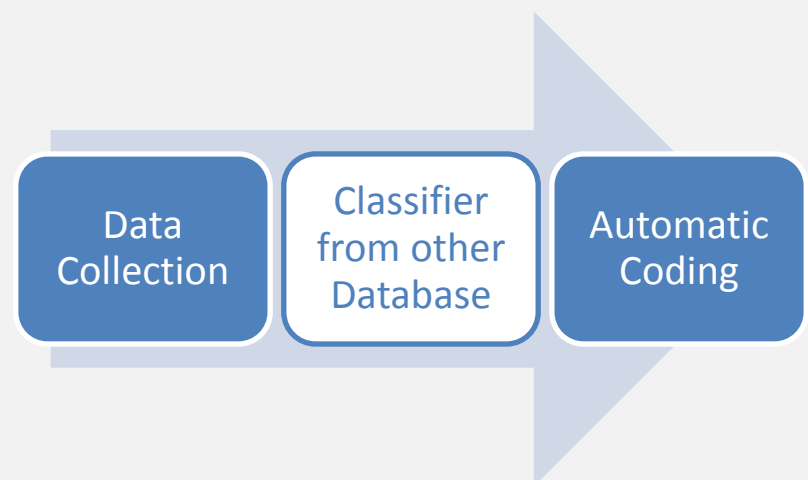
Strengths:

- +Classifier trained on same database

Limitations:

- Requires ground truth coding
- Requires classifier training

Naïve Classifier Implementation



Strengths:

- +Requires no ground truth coding
- +Requires no classifier training

Limitations:

- Classifier not trained on same database



SVM Classifier

- SIFT Feature Data
- Radial Basis Kernel
- 3-fold cross-validation

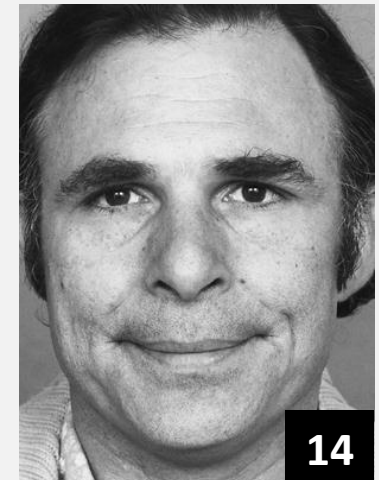
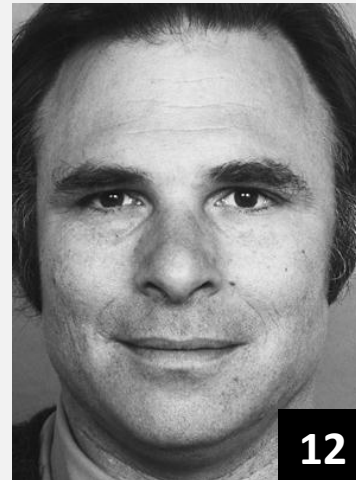
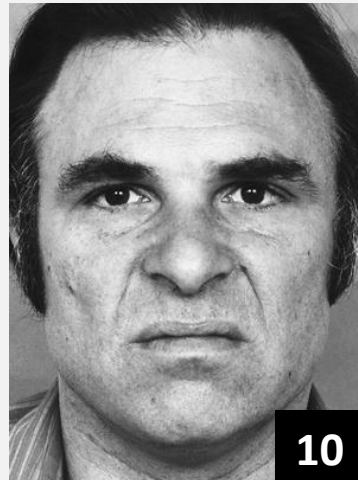
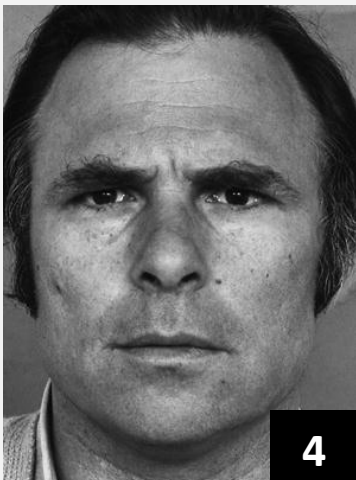
RU-FACS-1 Database

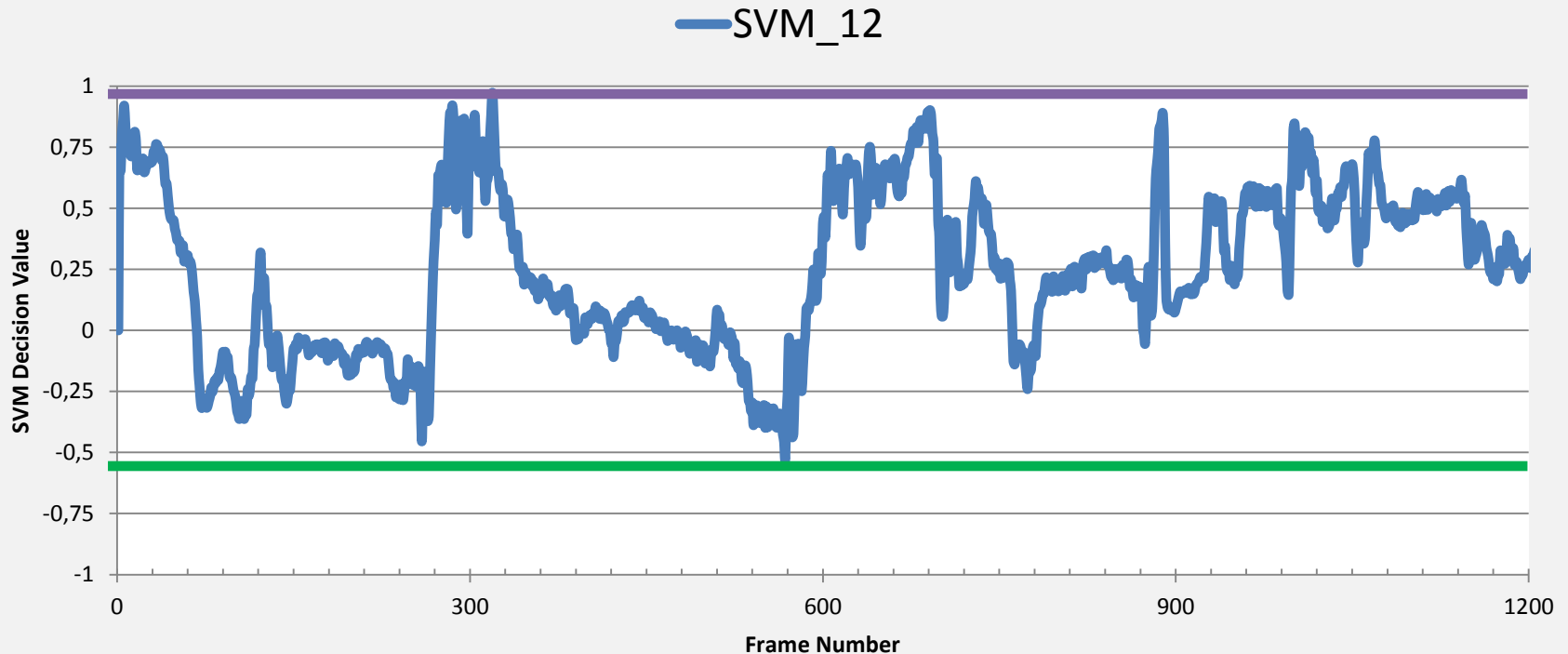
- Classifier Training Set
 - 17 subjects (97000 frames)
- Classifier Testing Set
 - 11 subjects (67000 frames)
- False Opinion Paradigm



Spectrum Database

- Threshold Training Set
 - 23 subjects (88000 frames)
- Threshold Testing Set
 - 13 subjects (37000 frames)
- HRSD-17 Depression Interview

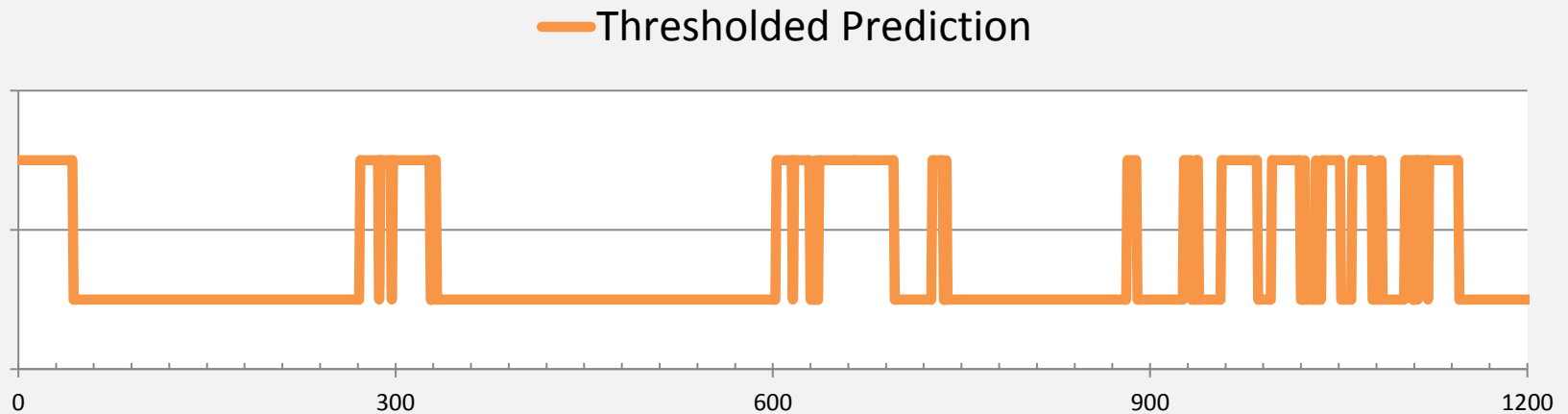
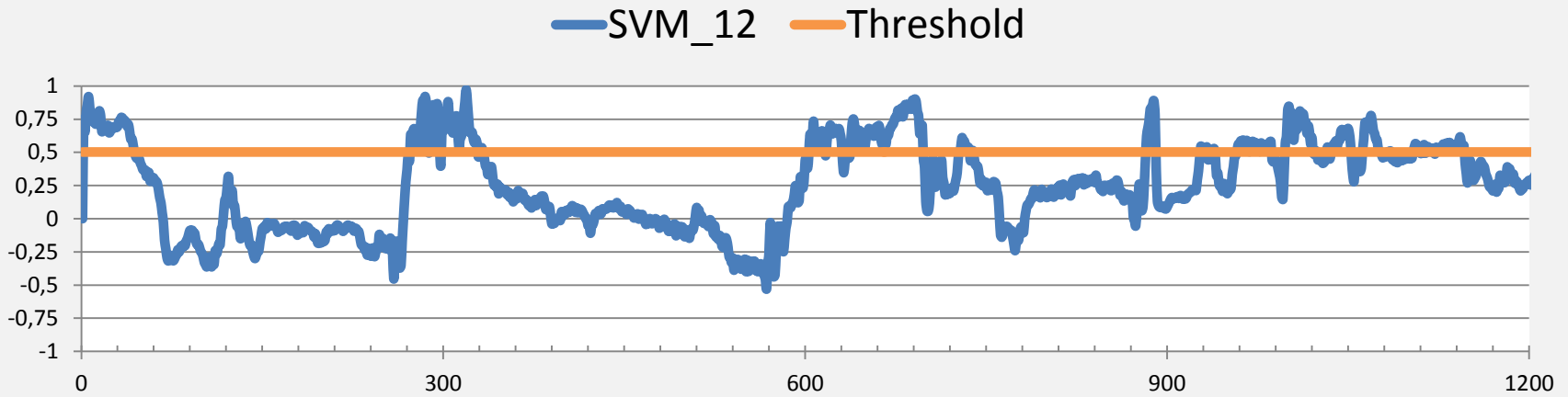




- Find **minimum** and **maximum** SVM decision values for each AU
- Separate this range into equal steps to identify potential thresholds
- This study compared 250 thresholds for each of the action units



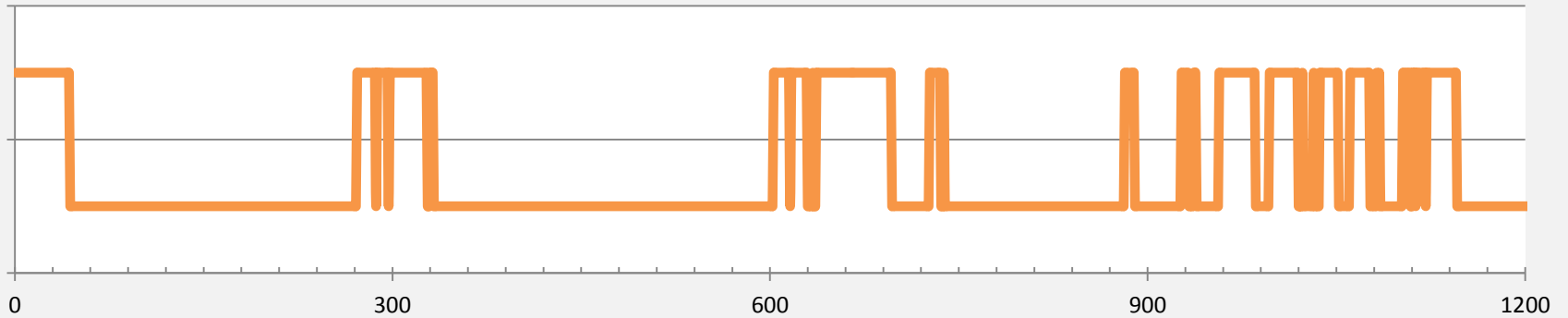
Step 4 – Generate Predictions



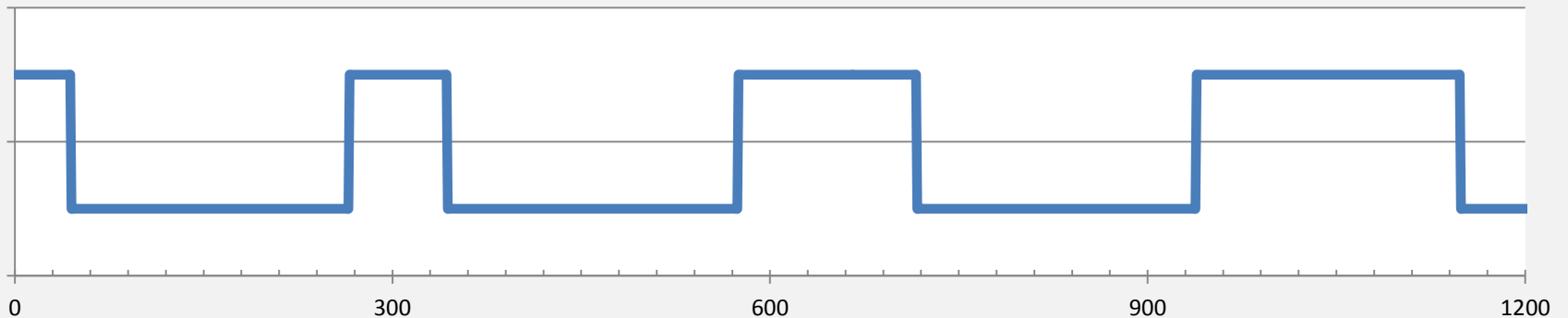


Step 5 – Compare to Groundtruth

— Thresholded Prediction



— Groundtruth Labels



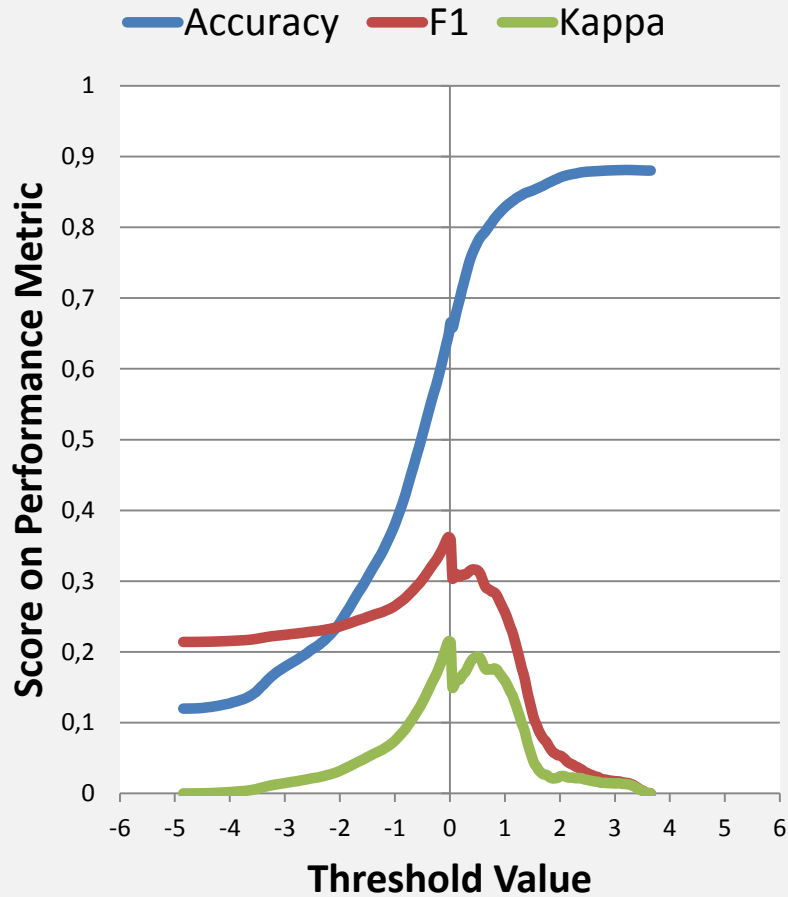
Accuracy = **0.855**

F1 = **0.756**

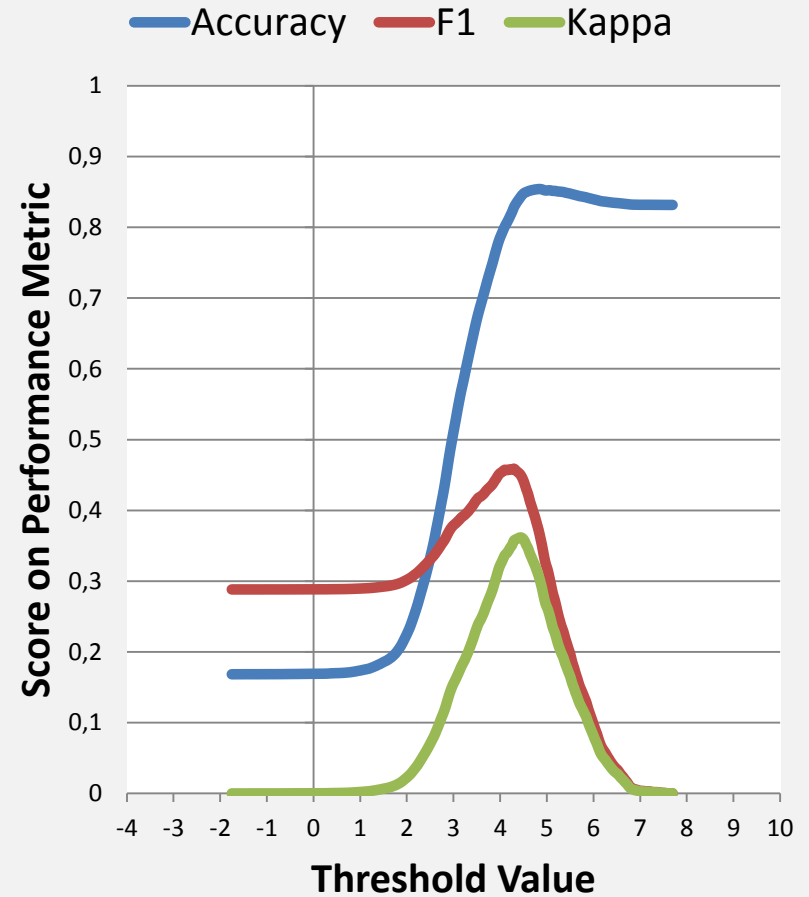
Kappa = **0.656**



AU_4 Threshold Training

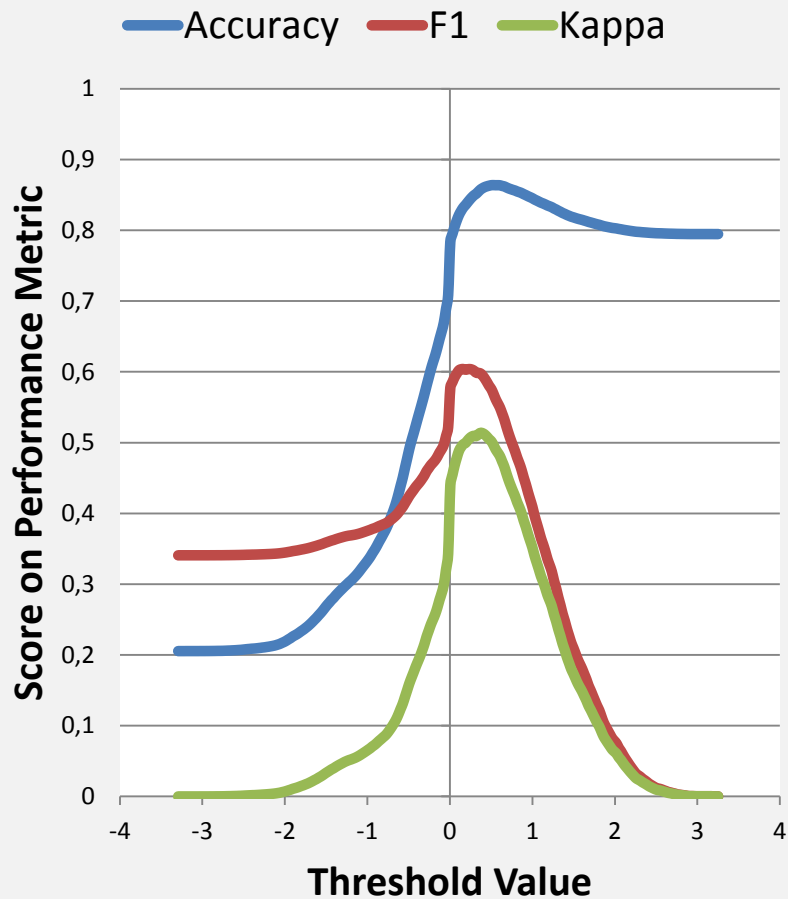


AU_10 Threshold Training

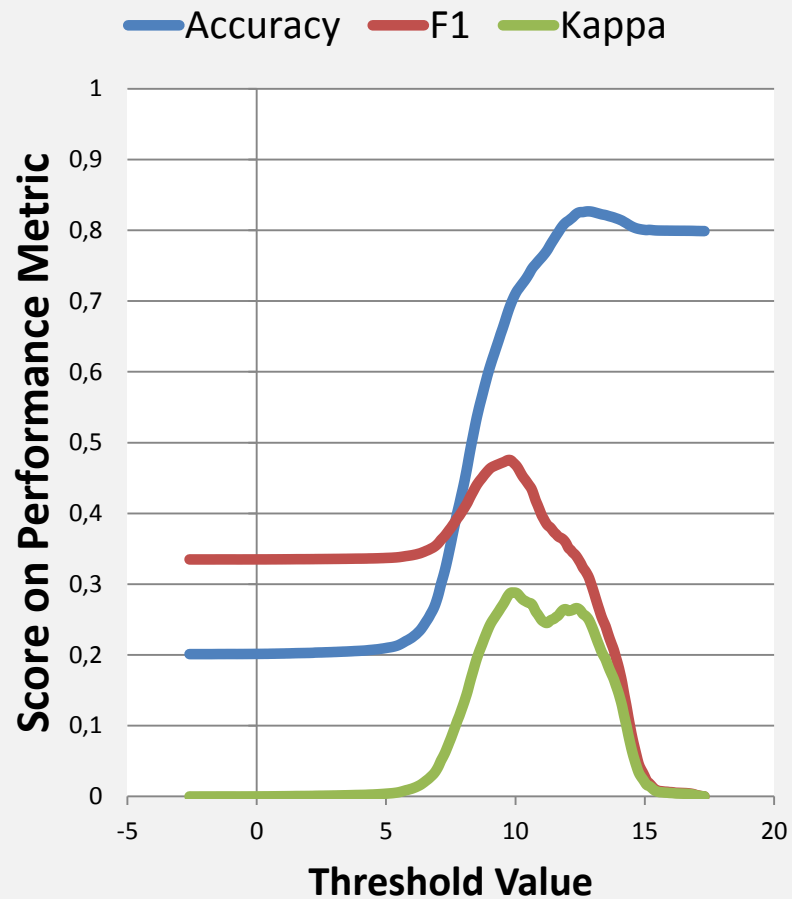




AU_12 Threshold Training

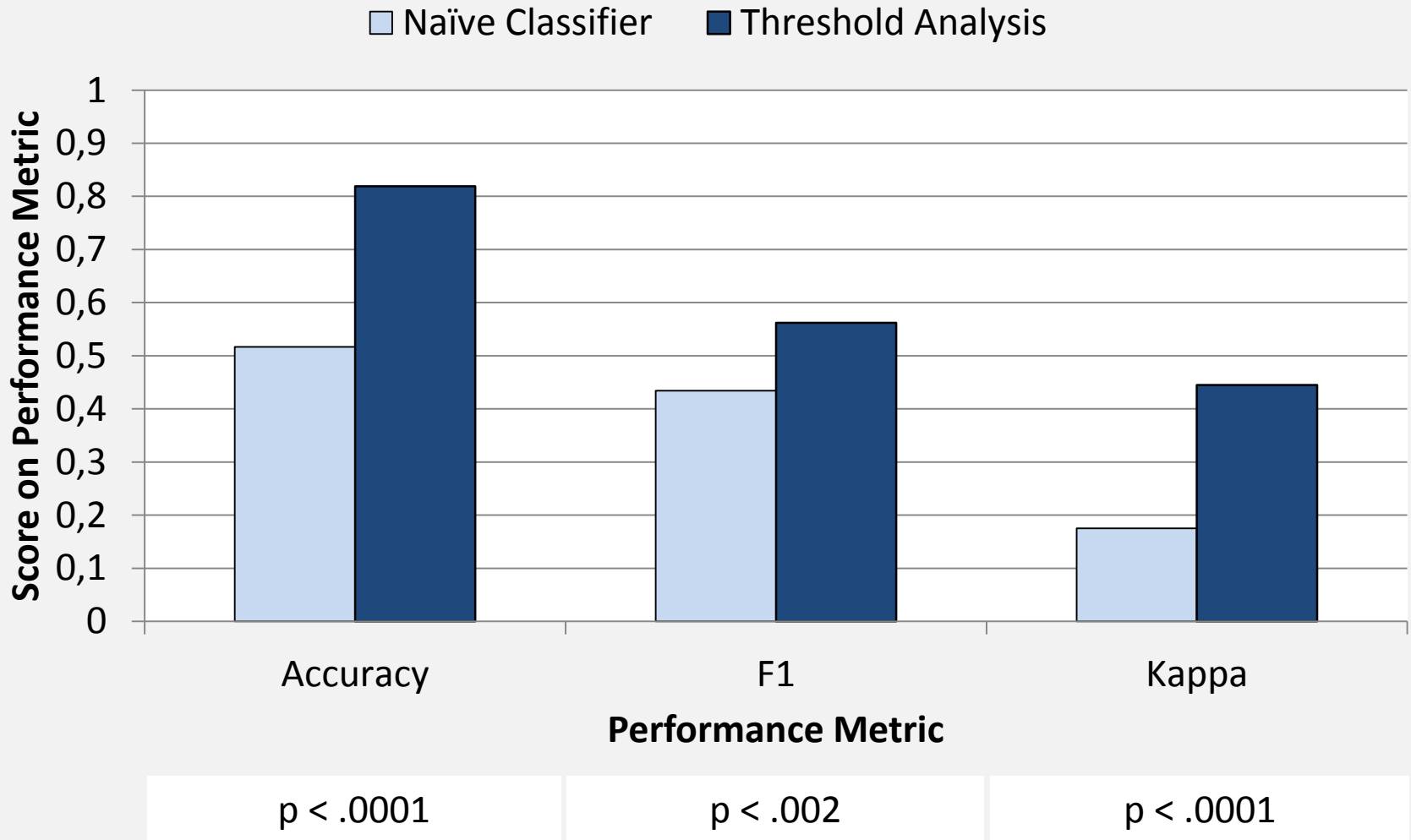


AU_14 Threshold Training



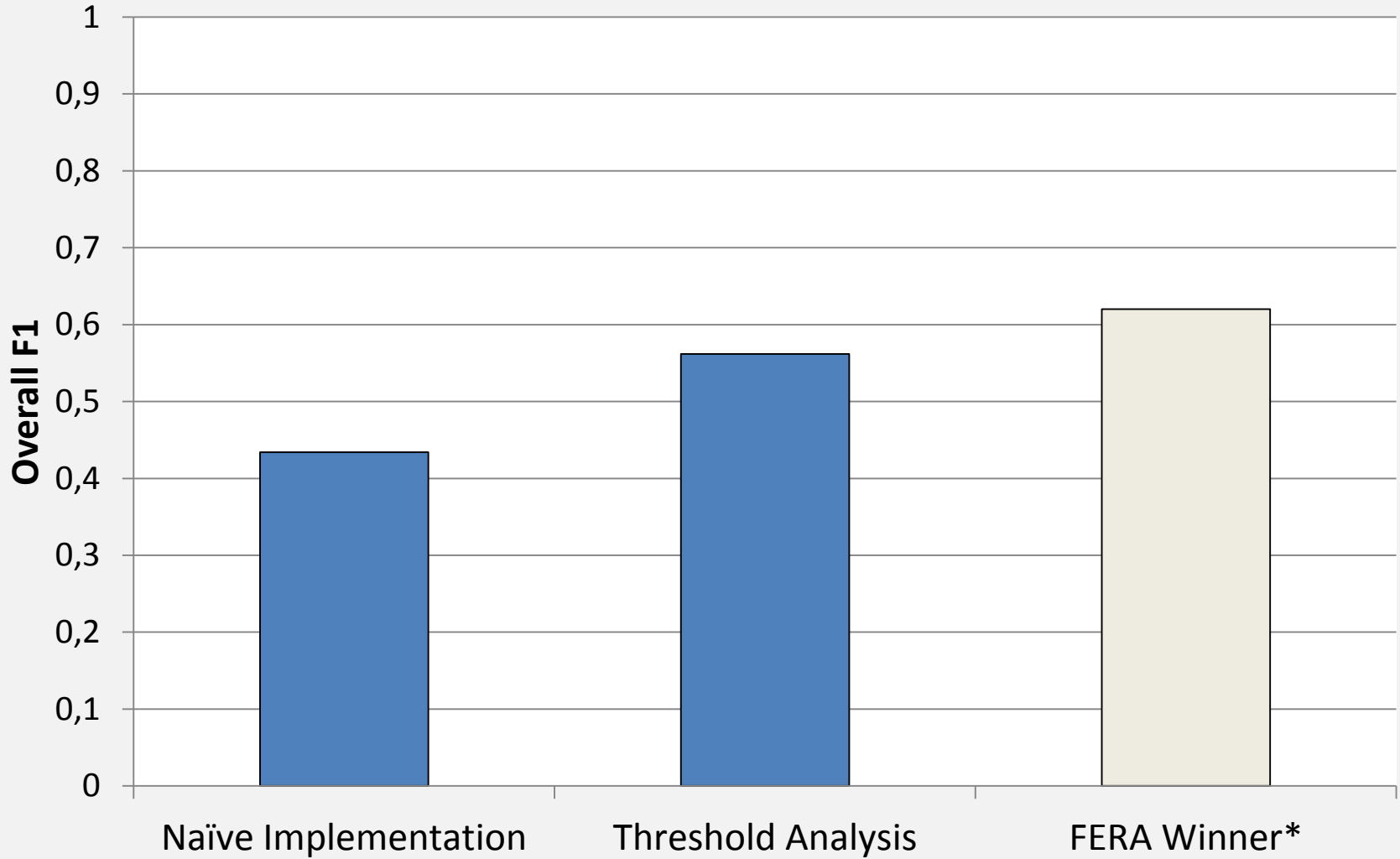


Overall Results in Testing Subset



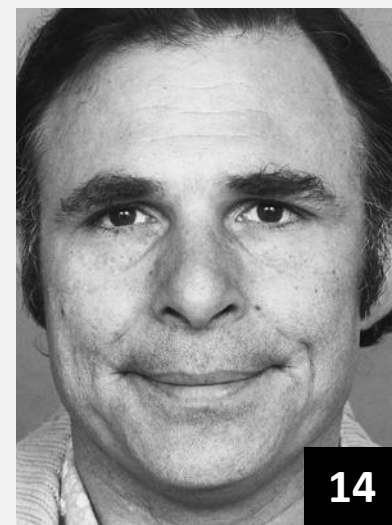
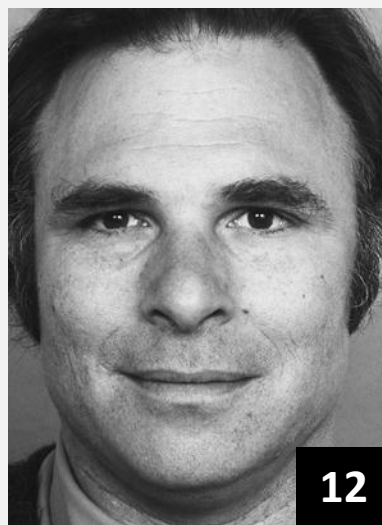
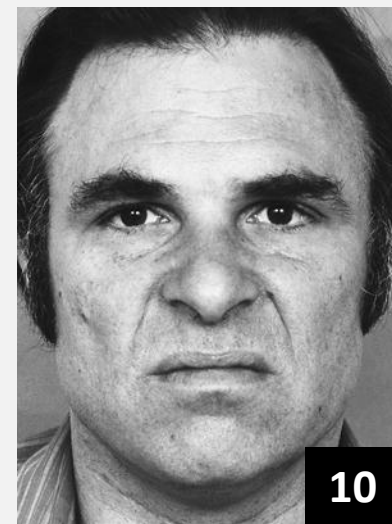
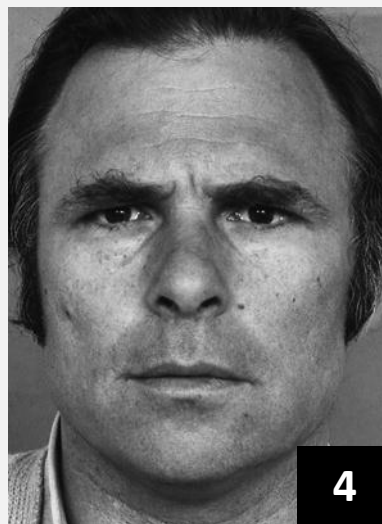
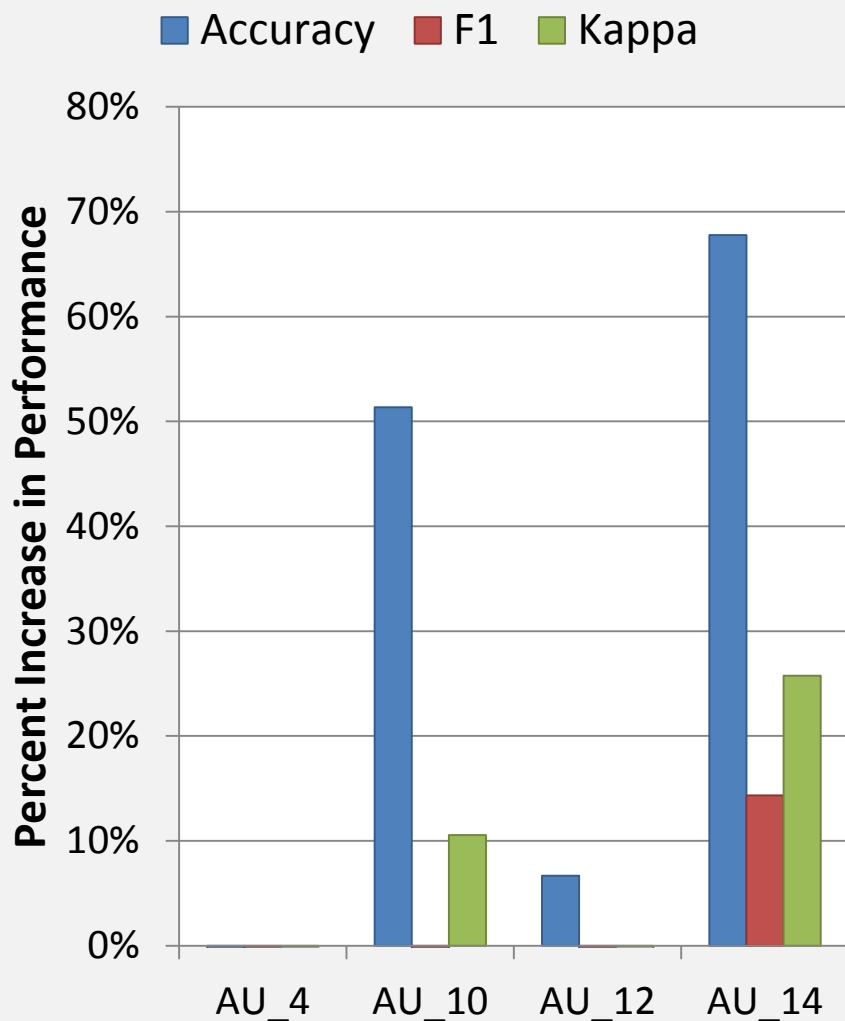


Comparison to FERA Winner





Performance Gains by Action Unit



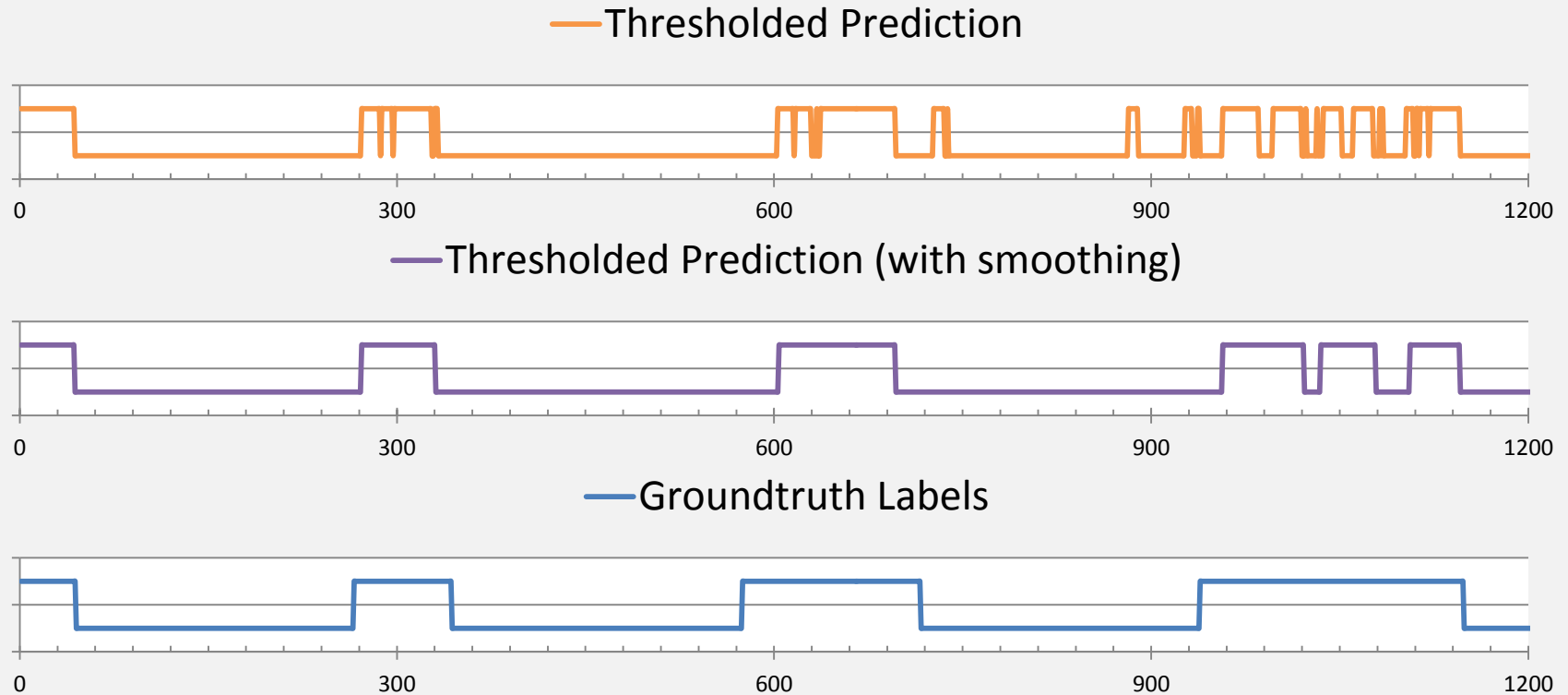


- Additional Databases and Action Units
- Additional Feature and Classifier types
- Determine required size of training set
- Smoothing to remove noise in predictions
- Compare directly to Novel Classifier Training

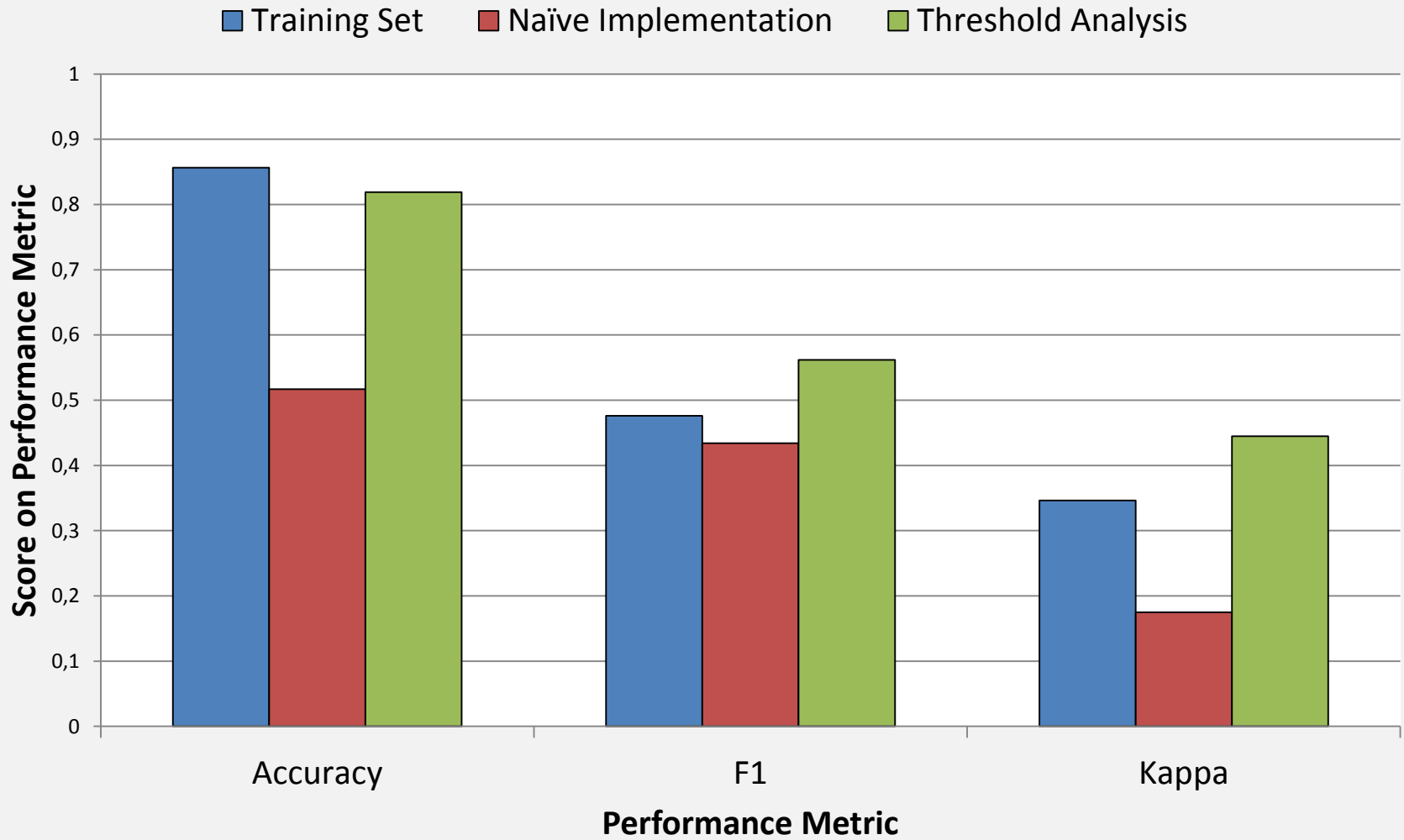
jmg174@pitt.edu

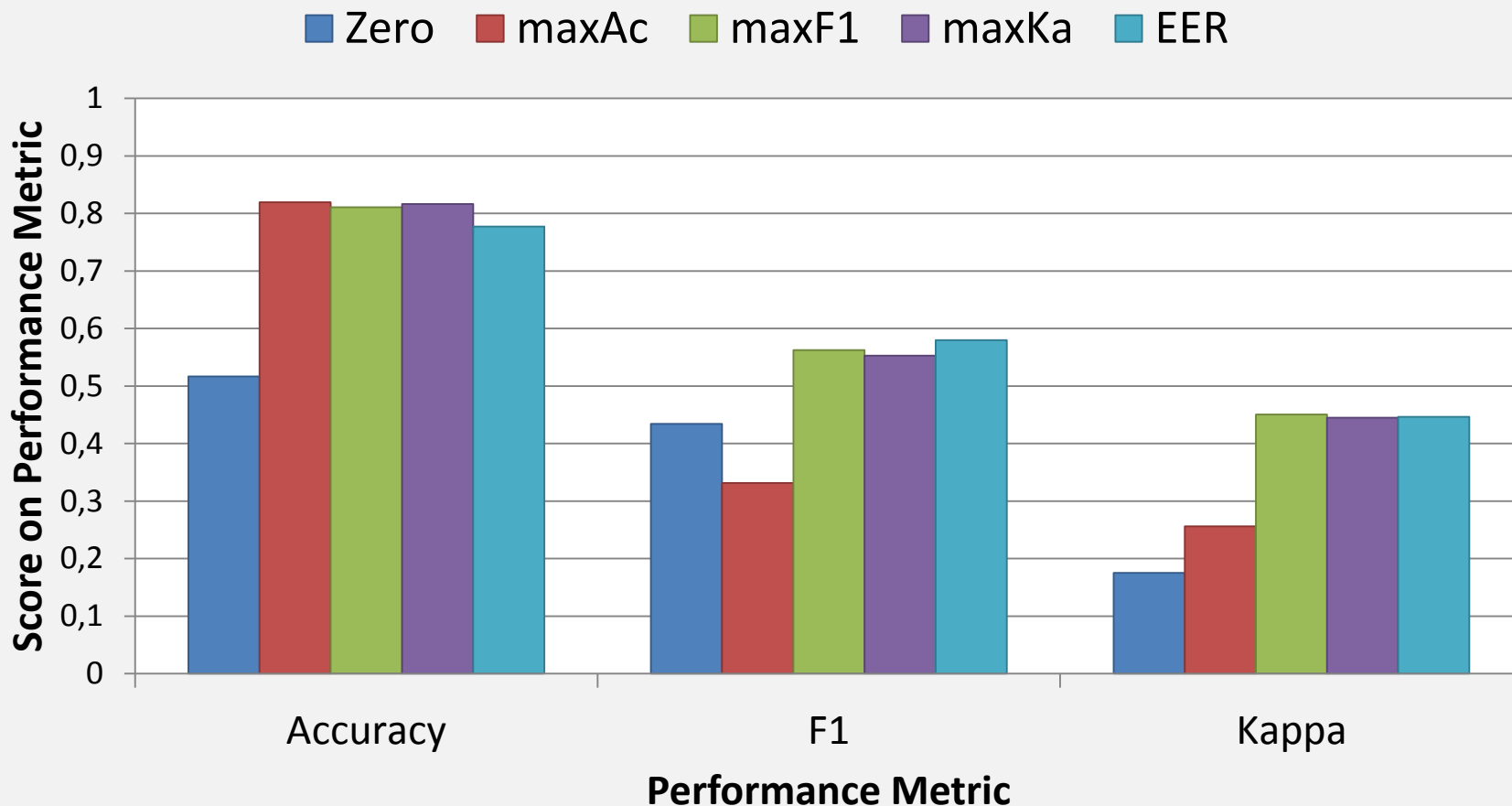
AFFECT ANALYSIS GROUP
Research Lab at the University of Pittsburgh





Accuracy = 0.855	F1 = 0.756	Kappa = 0.656
Accuracy = 0.896	F1 = 0.826	Kappa = 0.754





The threshold that maximized Accuracy performed poorly on F1 and Kappa. Thresholds that maximized F1, Kappa, and EER performed best on all metrics.