EMOTION CLASSIFICATION: HOW DOES AN AUTOMATED SYSTEM COMPARE TO NAÏVE HUMAN CODERS?

Sefik Emre Eskimez, Kenneth Imade, Na Yang, Melissa Sturge-Apple, Zhiyao Duan, Wendi Heinzelman

University of Rochester

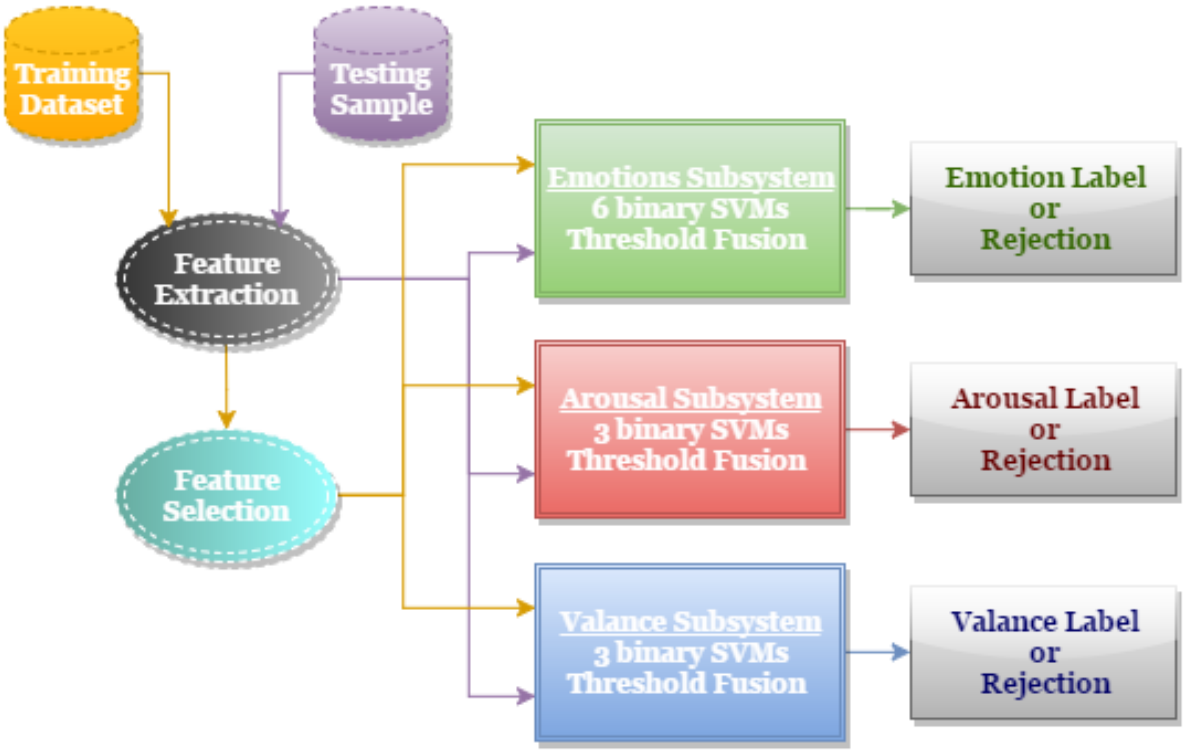
Motivation

- Emotions play a vital role in social interactions
  - Realistic human-computer interactions require accurately determining affective state of the user
- How does an automated system compare to naïve human coders?
  - Can automated systems replace naïve human coders in speech-based emotion classification applications?

Introduction

- In this study, an automated system is compared with naïve human coders in terms of speech emotion classification performance
  - Results show that it is feasible to replace naïve human coders with automatic emotion classification systems
  - Naïve human coders' confidence level in classification does not affect their classification accuracy, while automated system can increase accuracy by only classifying samples in which it is confident

Automatic Speech Emotion Classification System Overview



Feature Extraction

- All features and their 1st order derivatives (except speaking rate) are calculated in overlapping frames
- Statistical values are calculated using all frames
  - min, max, mean, standard deviation and range (max-min).

Feature name	#	Feature name	#
Fundamental Frequency (f0)	10	Spread	10
Energy	10	Skewness	10
Frequency and bandwidth for the first four Formants	80	Kurtosis	10
12 Mel-frequency Cepstral Coefficients (MFCCs)	120	Flatness	10
Zero-cross rate	10	Entropy	10
Roll-off	10	Roughness	10
Brightness	10	Irregularity	10
Centroid	10	Speaking Rate	1
		Size of Feature Vector:	**331**

Feature Selection

- Support Vector Machine (SVM) Recursive Feature Elimination
  - Train the SVMs to obtain weights
  - Eliminate the feature that has the lowest weight value
  - Continue until there is no feature left
  - Rank the features according to reverse of the elimination order to get top N best features
    - In our experiments, we use N = 80 (out of 331)

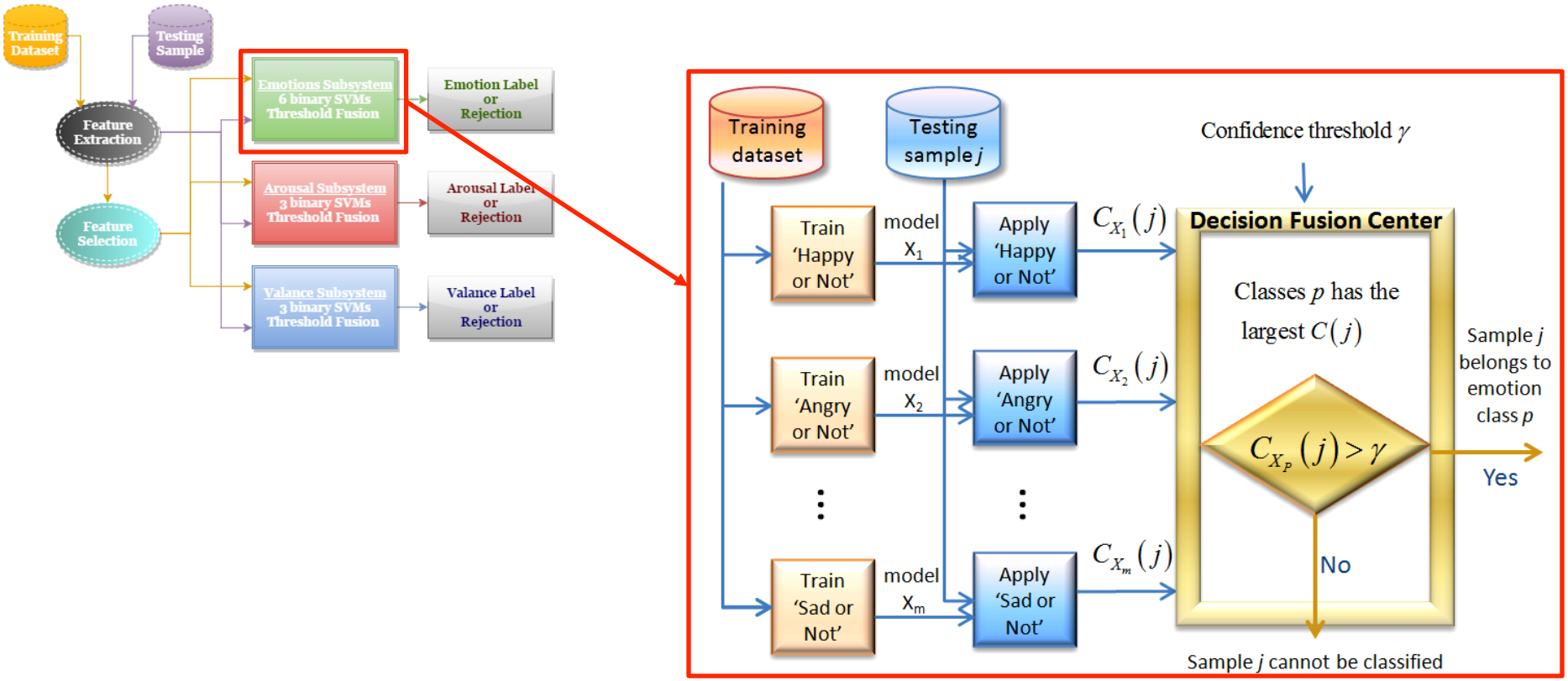
Automatic Emotion Classification

- The system labels each sample with three different labels from the following sub-systems:
  - 6 Emotion Categories: anger, disgust, panic, happy, neutral, sadness.
  - Arousal Categories: active, passive and neutral (APN).
  - Valence Categories: positive, negative and neutral (PNN).

Automatic Emotion Classifiers

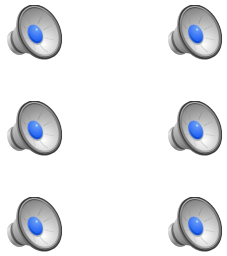
- System uses binary SVM classifiers with RBF kernel for each emotion
  - 6 binary SVMs for first sub-system
  - 3 binary SVMs for second and third-sub systems
  - Total of 12 binary SVMs

Automatic Emotion Classification Threshold Fusion



LDC Dataset

- 15 Emotions
- Speakers: 4 actresses and 4 actors
- Total of 2433 utterances
- Acted dataset
- In our experiments
  - 6 Emotions: anger, disgust, panic, happy, neutral and sadness.
  - Speakers: 4 actresses and 3 actors
  - 727 utterances



Experimental Setup: Automatic Emotion Classification System

- 7-fold cross validation
  - 6/7 of the data used for training, 1/7 of the data used for testing
  - In each fold, training and testing data have been randomly chosen
  - Data have been up-sampled to even out all classes
- Leave-One-Subject-Out (LOSO) test

Experimental Setup: Amazon's Mechanical Turk

- 138 unique workers participated
- 10-100 random samples per worker
- Only one sample per emotion category is presented beforehand

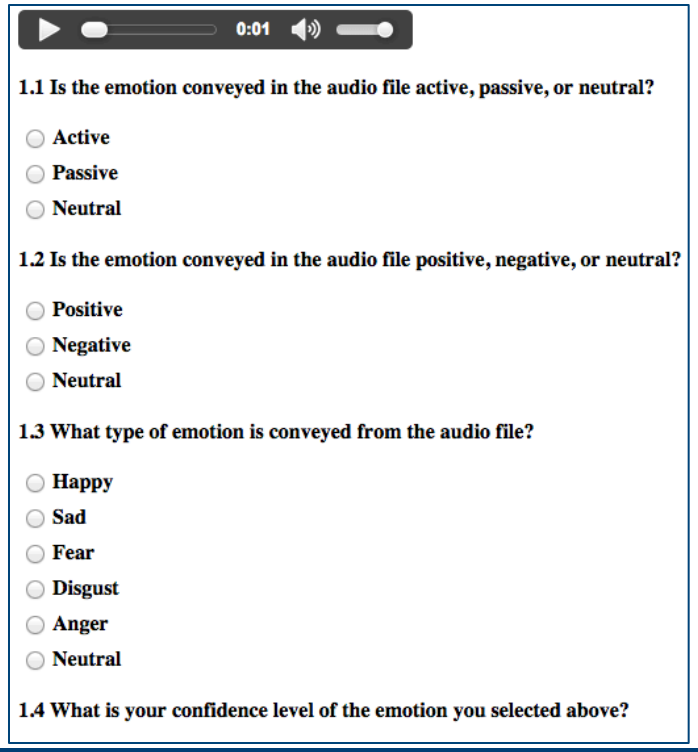
**********You are paid $0.50 per HIT for finishing all 10 audio files**********

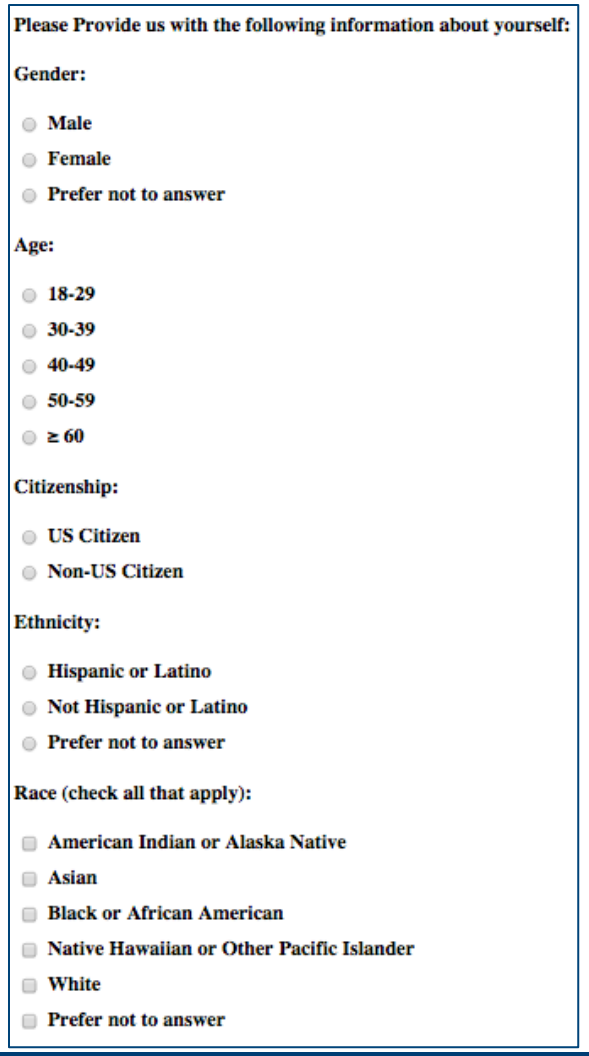
- Please choose one of the six emotions as a best match for each audio file.
- Please choose one of the two confidence conditions indicating how confident you are of your answers.
- The task is finished only when you have given your answers for each emotion and confidence selection.
- Specific Transcription Instructions:
  1. Each audio file contains a simple phrase of either a number or a date.
  2. Each word of your transcription should be written in complete form, for example, "108" or "Dec. 12th".
  3. Passive emotions are the type of emotions where an individual would conceal their reactions rather than acting on it or addressing it. Whereas, active emotions involve physically or energetically expressing one's reaction.
  4. Positive emotions are emotions that make you feel good, and negative emotions are emotions that do not make you feel good.
  5. To hear an example, click the sample of emotions listed below:
     - Sample 'Happy' Voice
     - Sample 'Sad' Voice
     - Sample 'Fear' Voice
     - Sample 'Disgust' Voice
     - Sample 'Anger' Voice
     - Sample 'Neutral' Voice

Sample One:
Please listen to the audio file below, then choose the emotion that best corresponds to what the speaker is conveying from the following:

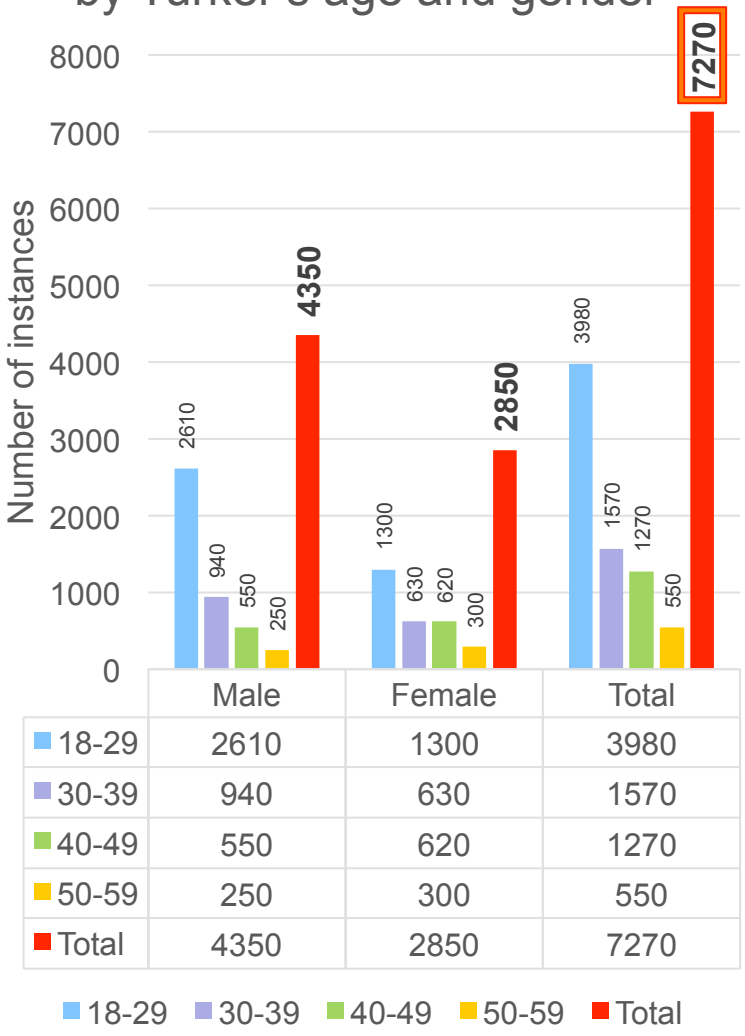
Experimental Setup: Amazon's Mechanical Turk

- Turkers are asked to listen, label and transcribe the audio sample
- Turkers are asked for demographic information after they are done labeling





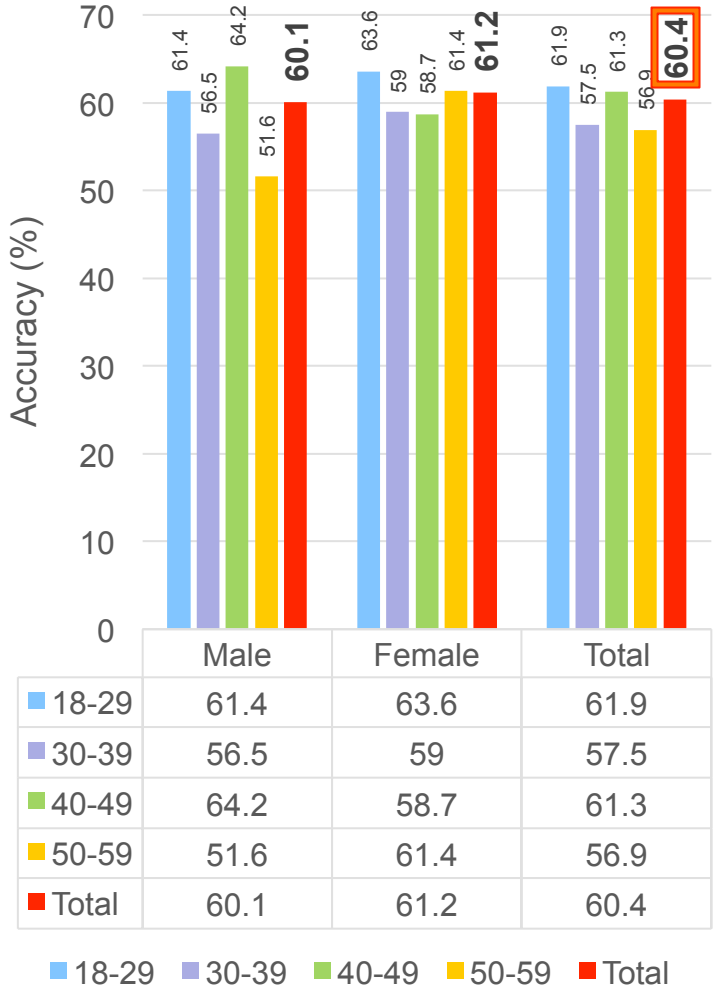
Number of labeled instances by Turker's age and gender



	Male	Female	Total
18-29	2610	1300	3980
30-39	940	630	1570
40-49	550	620	1270
50-59	250	300	550
Total	4350	2850	7270

18-29 30-39 40-49 50-59 Total

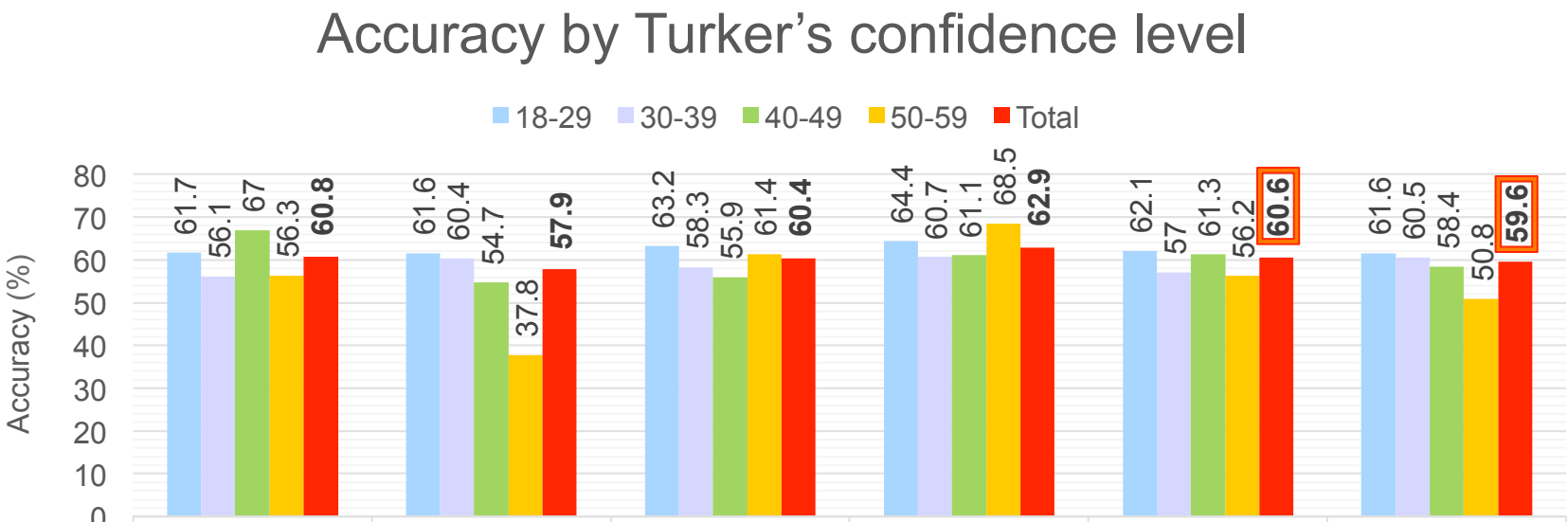
Accuracy by Turker's age and gender



	Male	Female	Total
18-29	61.4	63.6	61.9
30-39	56.5	59	57.5
40-49	64.2	58.7	61.3
50-59	51.6	61.4	56.9
Total	60.1	61.2	60.4

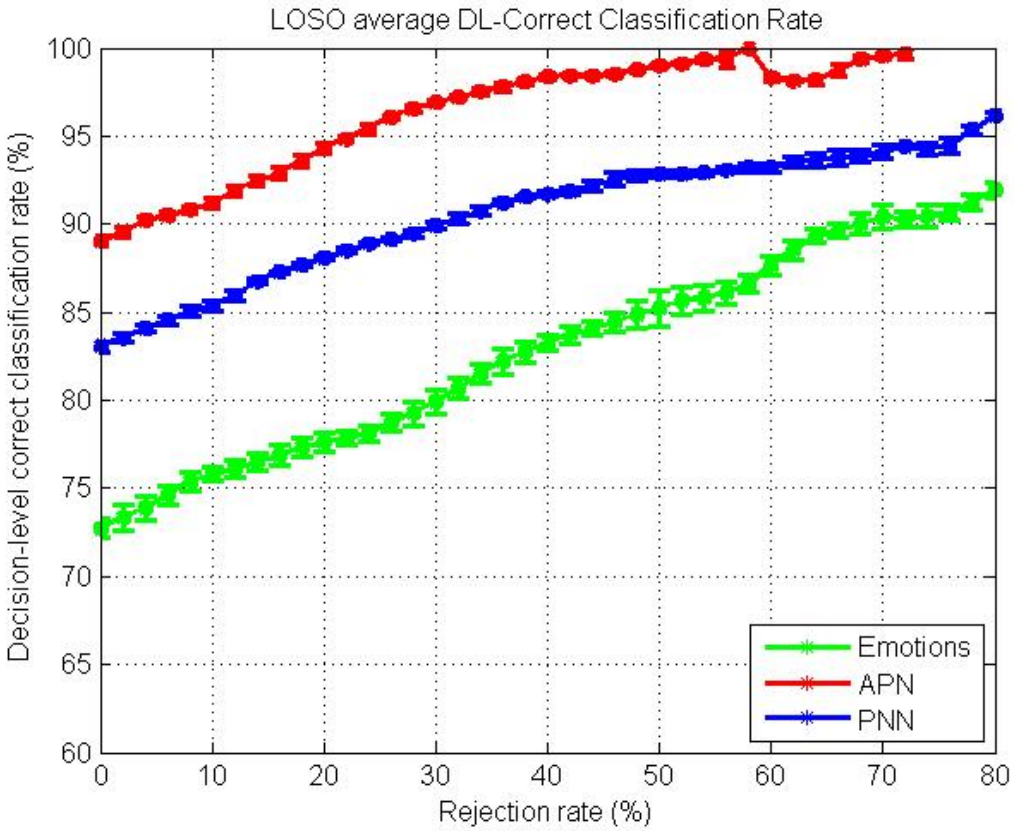
18-29 30-39 40-49 50-59 Total

Results: Turkers

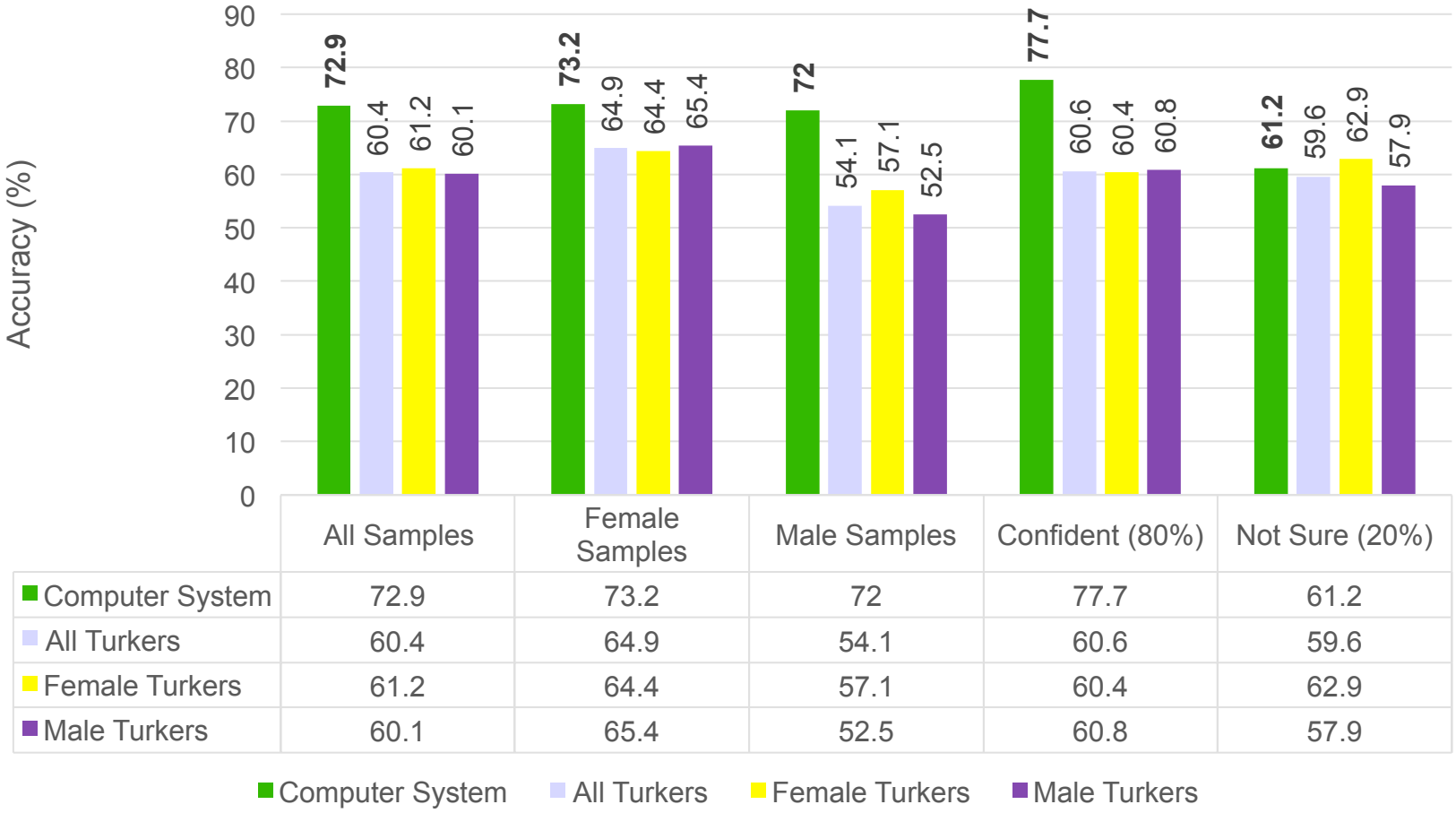


	Male (Confident)	Male (Not Sure)	Female (Confident)	Female (Not Sure)	Total (Confident)	Total (Not Sure)
18-29	61.7	61.6	63.2	64.4	62.1	61.6
30-39	56.1	60.4	58.3	60.7	57	60.5
40-49	67	54.7	55.9	61.1	61.3	58.4
50-59	56.3	37.8	61.4	68.5	56.2	50.8
Total	60.8	57.9	60.4	62.9	60.6	59.6

Results: Computer System

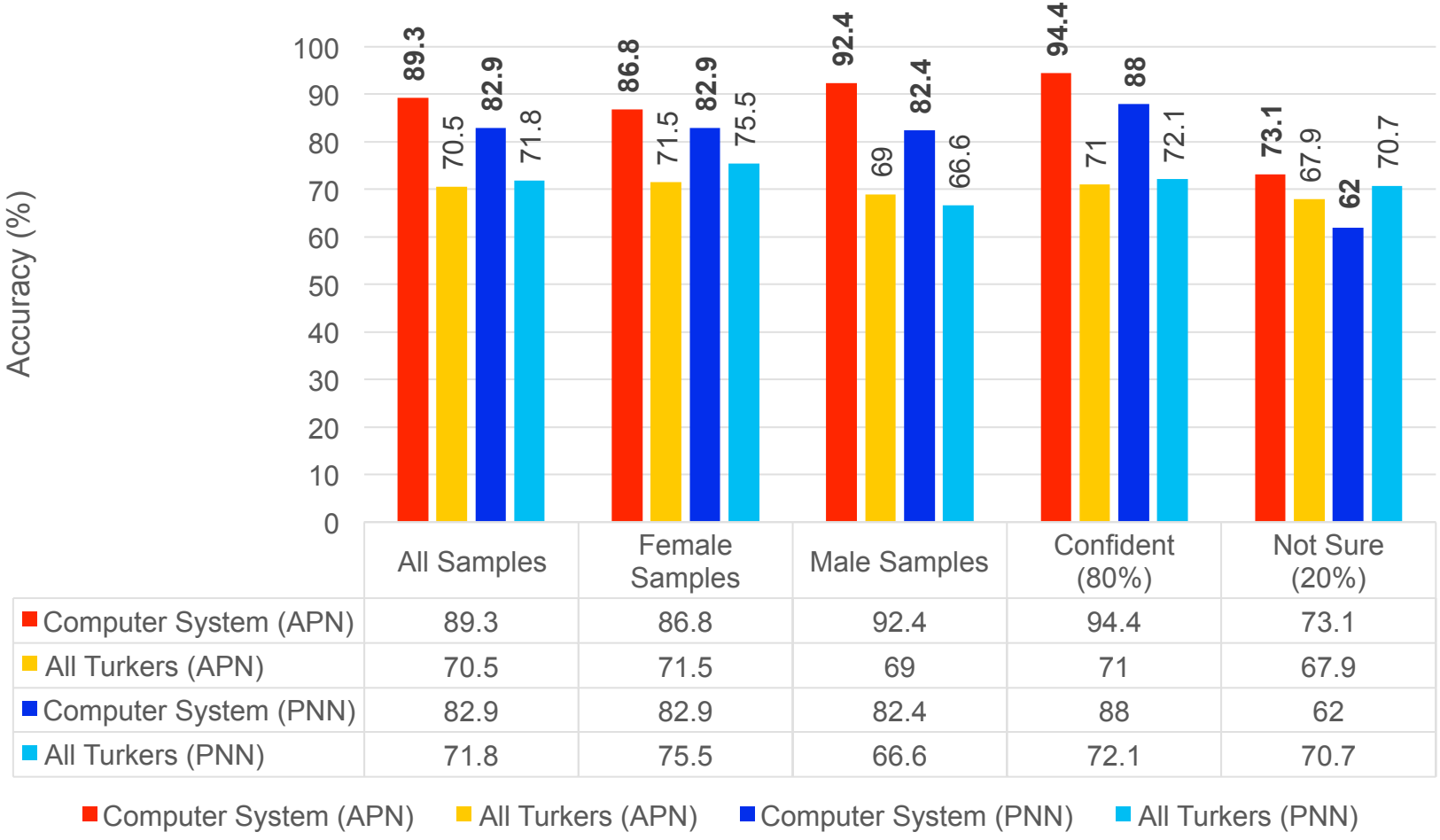


Turkers vs. Computer System: Emotions



	All Samples	Female Samples	Male Samples	Confident (80%)	Not Sure (20%)
Computer System	72.9	73.2	72	77.7	61.2
All Turkers	60.4	64.9	54.1	60.6	59.6
Female Turkers	61.2	64.4	57.1	60.4	62.9
Male Turkers	60.1	65.4	52.5	60.8	57.9

Turkers vs. Computer System: APN & PNN



	All Samples	Female Samples	Male Samples	Confident (80%)	Not Sure (20%)
Computer System (APN)	89.3	86.8	92.4	94.4	73.1
All Turkers (APN)	70.5	71.5	69	71	67.9
Computer System (PNN)	82.9	82.9	82.4	88	62
All Turkers (PNN)	71.8	75.5	66.6	72.1	70.7

Conclusion/Discussion

- This study compares naïve human coders with a computer emotion classification system
- Expressed vs. perceived emotions!
- The computer system achieves better accuracy in almost all cases
- The computer system can improve classification accuracy by rejecting samples with low confidence
- Naïve human coders were not able to improve their accuracy through specifying their confidence in their classification
- Results show that it is feasible to replace naïve human coders with automatic emotion classification systems

The End…

Thank you!