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, United States

Motivation

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

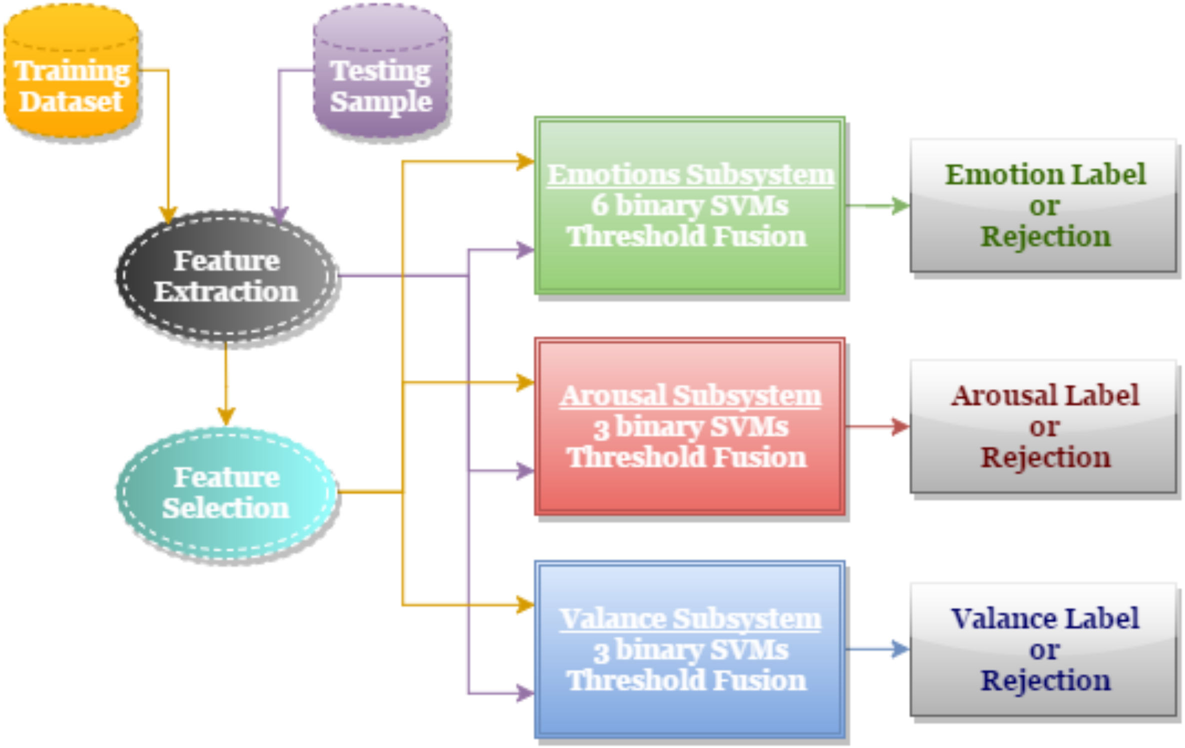
Introduction

- In this study naïve human coders and an automated system are evaluated in terms of speech emotion classification performance.

  - The 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 effect their classification accuracy, while automated system has increased accuracy when it is confident in classification.

Automatic Speech Emotion Classification System Overview



Feature Extraction

- All features and their 1st order derivative 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

- 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 and get top N best features.
    - In our experiments we use N = 80 (out of 331);

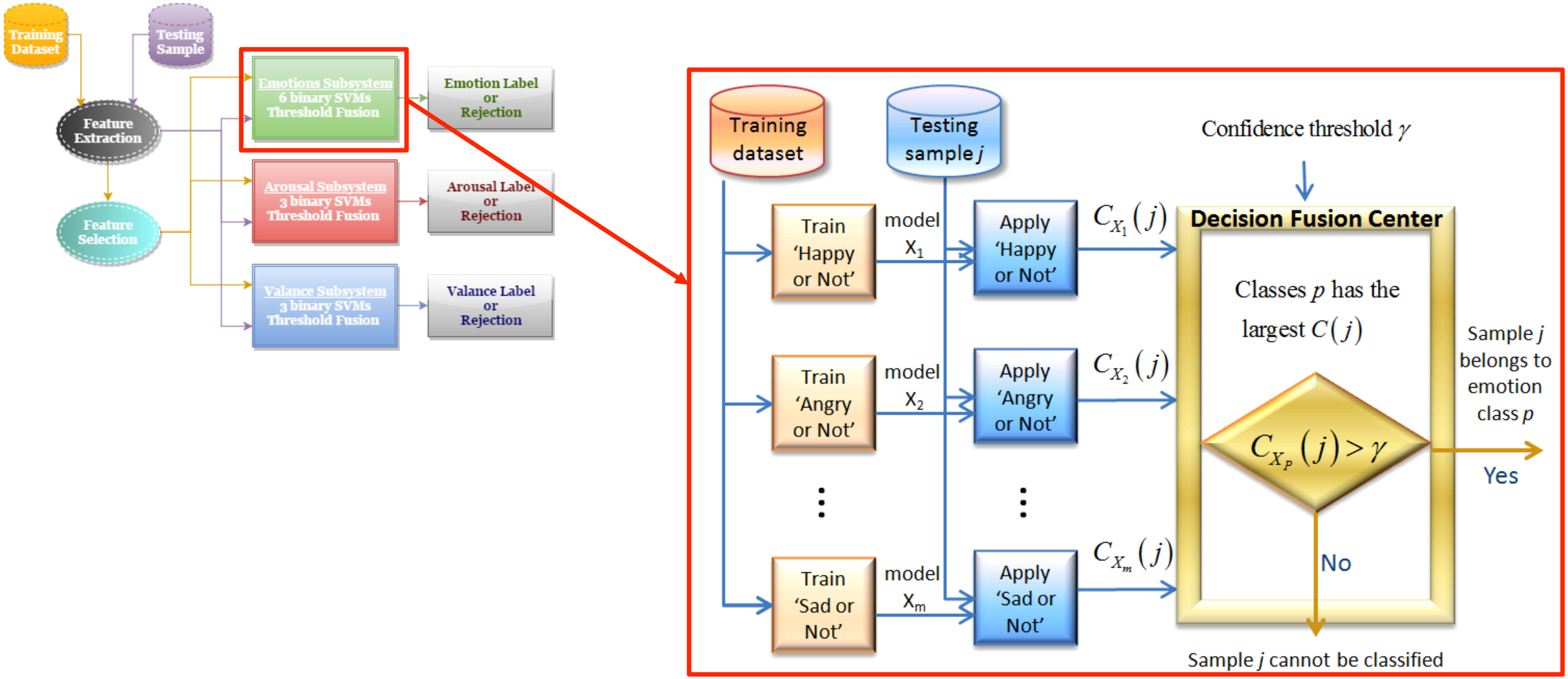
Automatic Emotion Classification

- 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 Support Vector Machine (SVM) classifiers with RBF kernel for each emotion, resulting:
    - 6 binary SVMs for first sub-system.
    - 3 binary SVMs for second and third-sub system.
    - Total of 12 binary SVMs.

Automatic Emotion Classification Threshold Fusion



LDC Dataset

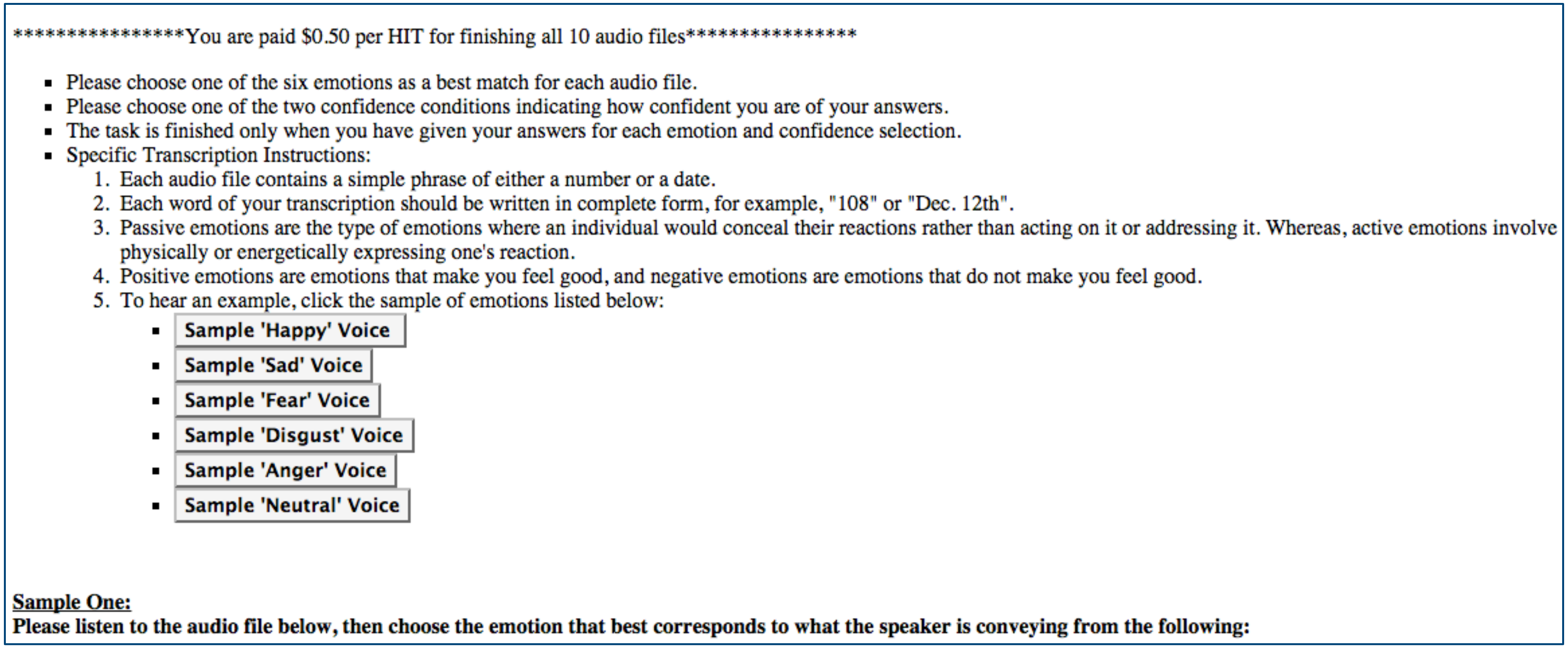
- 15 Emotions,
- Speakers: 4 actress and 4 actors.
- Total of 2433 utterances.
- Acted.
- In our experiments:
  - 6 Emotions: Anger, disgust, panic, happy, neutral and sadness.
  - Speakers: 4 actress and 3 actors.
  - 727 utterances.

Experimental Setup: Automatic Emotion Classification System

- 7-fold cross validation
  - 6/7 of the data is used for training, 1/7 of the data is used for testing.
  - In each fold, training and testing data has been randomly chosen.
  - Data has 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.



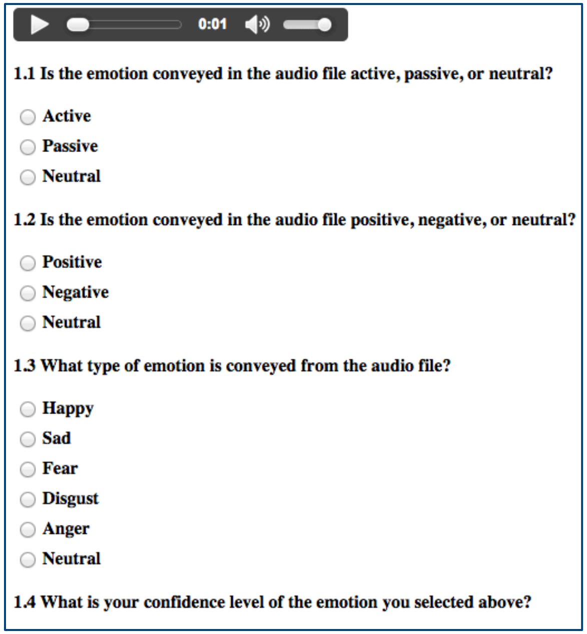
****************You are paid $0.50 per HIT for finishing all 10 audio files****************

- 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 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.
- On the right figure, Turkers are asked for demographic information.



1.1 Is the emotion conveyed in the audio file active, passive, or neutral?

○ Active
○ Passive
○ Neutral

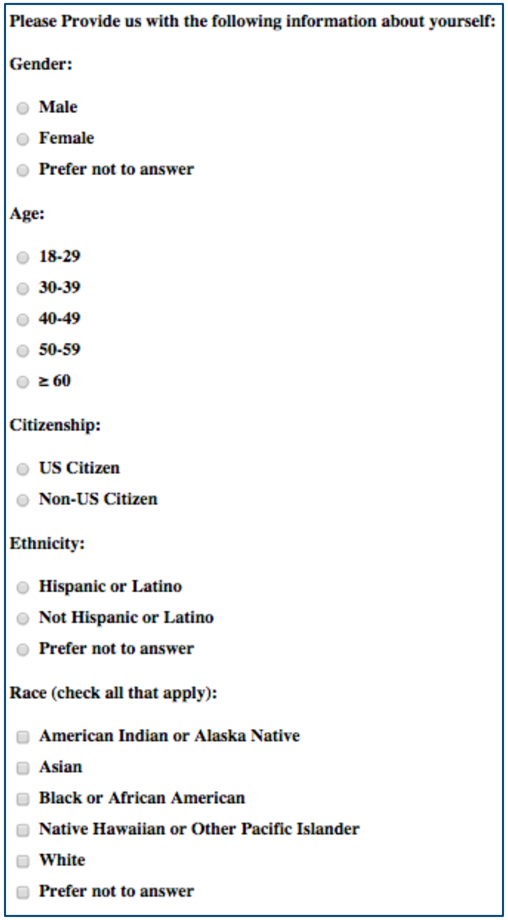
1.2 Is the emotion conveyed in the audio file positive, negative, or neutral?

○ Positive
○ Negative
○ Neutral

1.3 What type of emotion is conveyed from the audio file?

○ Happy
○ Sad
○ Fear
○ Disgust
○ Anger
○ Neutral

1.4 What is your confidence level of the emotion you selected above?



Please Provide us with the following information about yourself:

Gender:

○ Male
○ Female
○ Prefer not to answer

Age:

○ 18-29
○ 30-39
○ 40-49
○ 50-59
○ ≥ 60

Citizenship:

○ US Citizen
○ Non-US Citizen

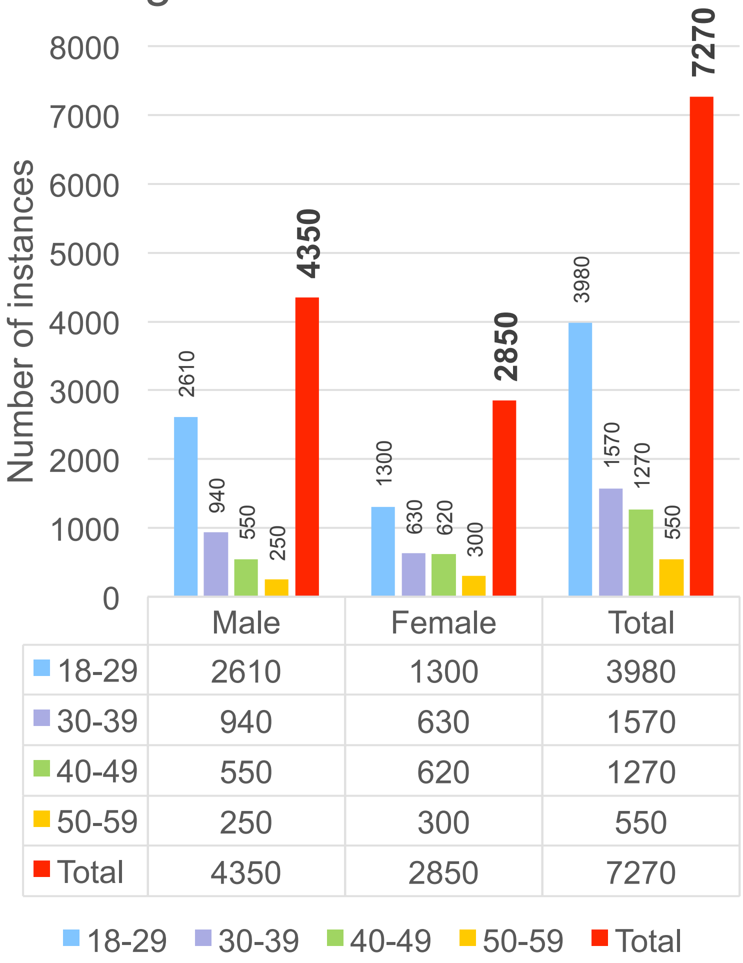
Ethnicity:

○ Hispanic or Latino
○ Not Hispanic or Latino
○ Prefer not to answer

Race (check all that apply):

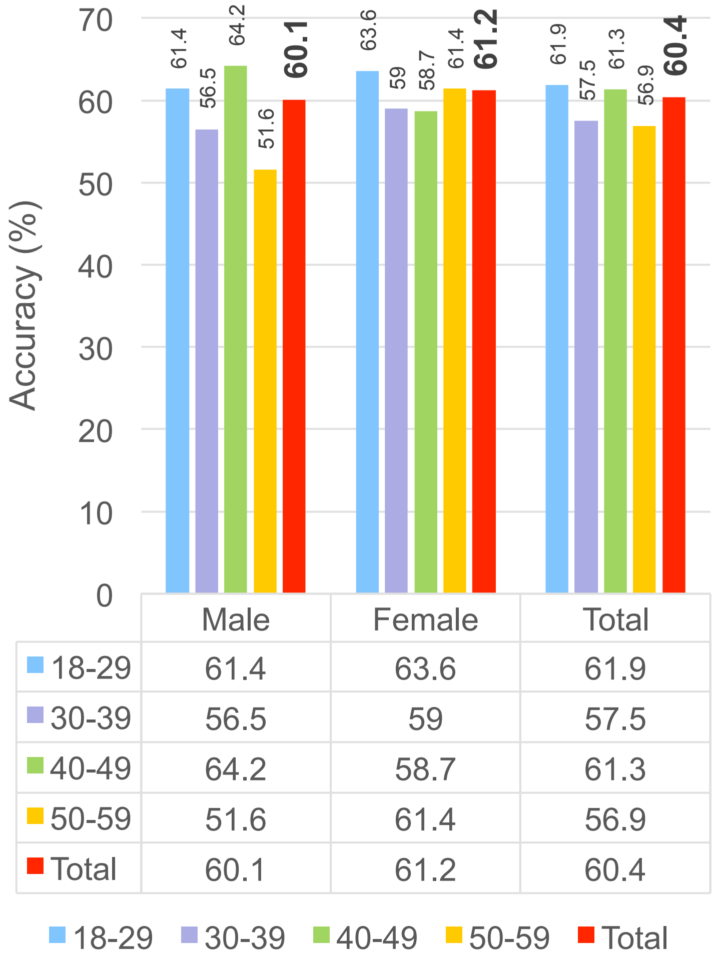
□ American Indian or Alaska Native
□ Asian
□ Black or African American
□ Native Hawaiian or Other Pacific Islander
□ White
□ Prefer not to answer

Number of labeled instances according to Turker's age and gender information:



	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

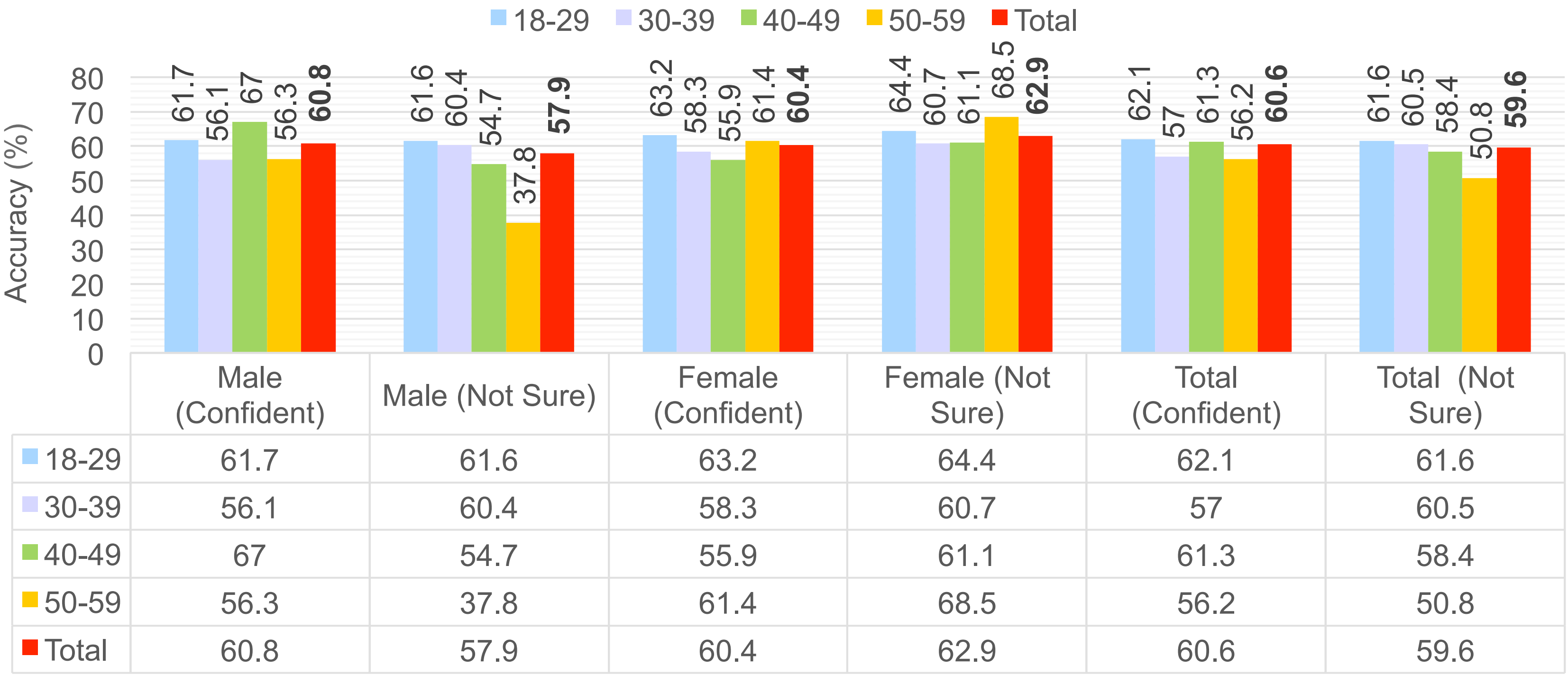
Accuracy percentage according to Turker's age and gender information:



	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

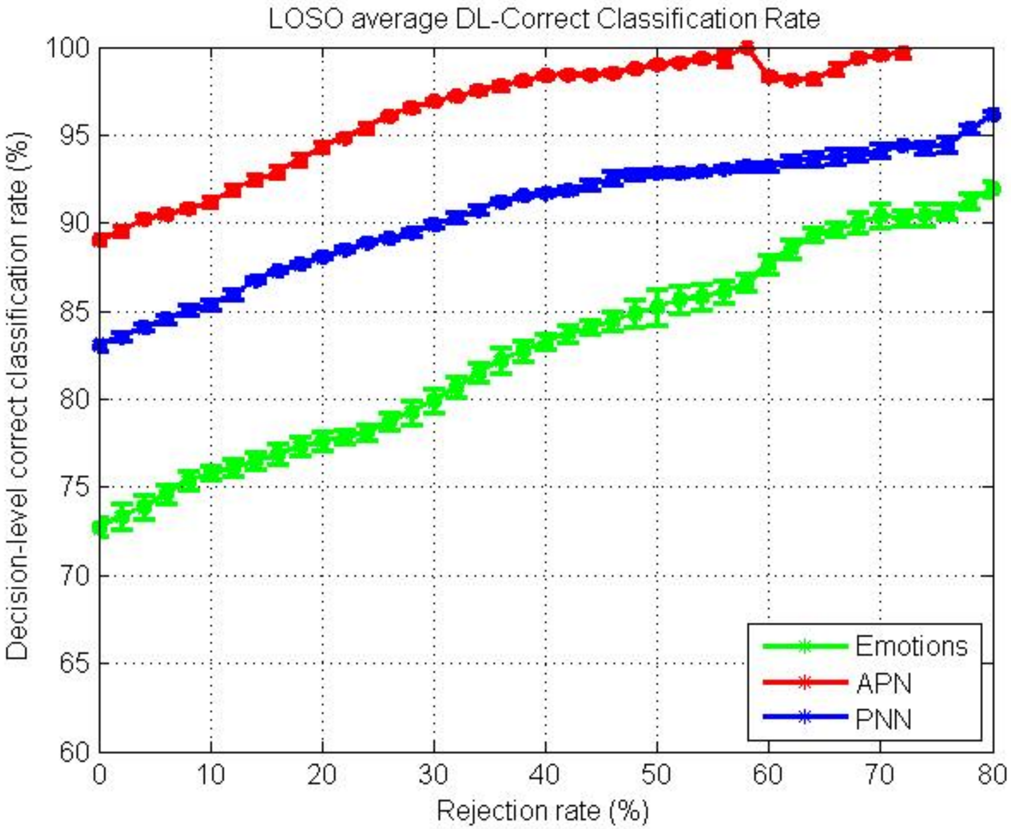
Results: Turkers

Accuracy percentage according to Turker's confidence level

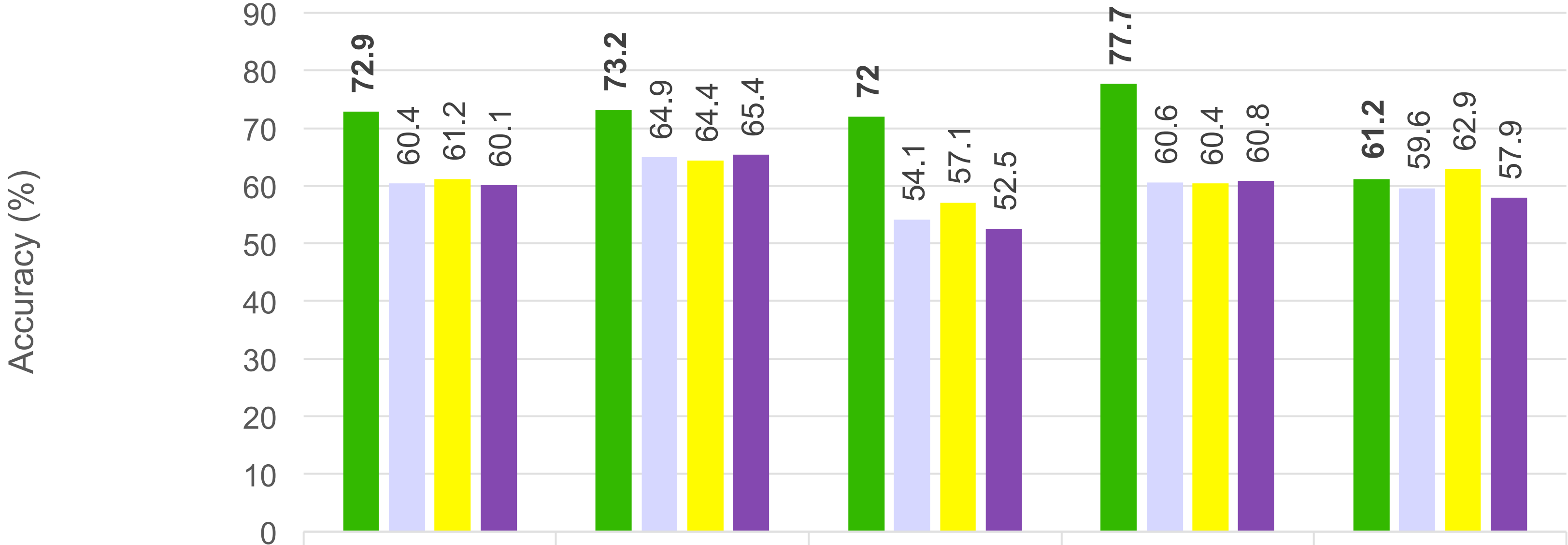


	Male (Confident)	Male (Not Sure)	Female (Confident)	Female (Not Sure)	Total (Confident)	Total (Not Sure)
🟦 18-29	61.7	61.6	63.2	64.4	62.1	61.6
🟪 30-39	56.1	60.4	58.3	60.7	57	60.5
🟩 40-49	67	54.7	55.9	61.1	61.3	58.4
🟨 50-59	56.3	37.8	61.4	68.5	56.2	50.8
🟥 Total	60.8	57.9	60.4	62.9	60.6	59.6

Results: Computer System



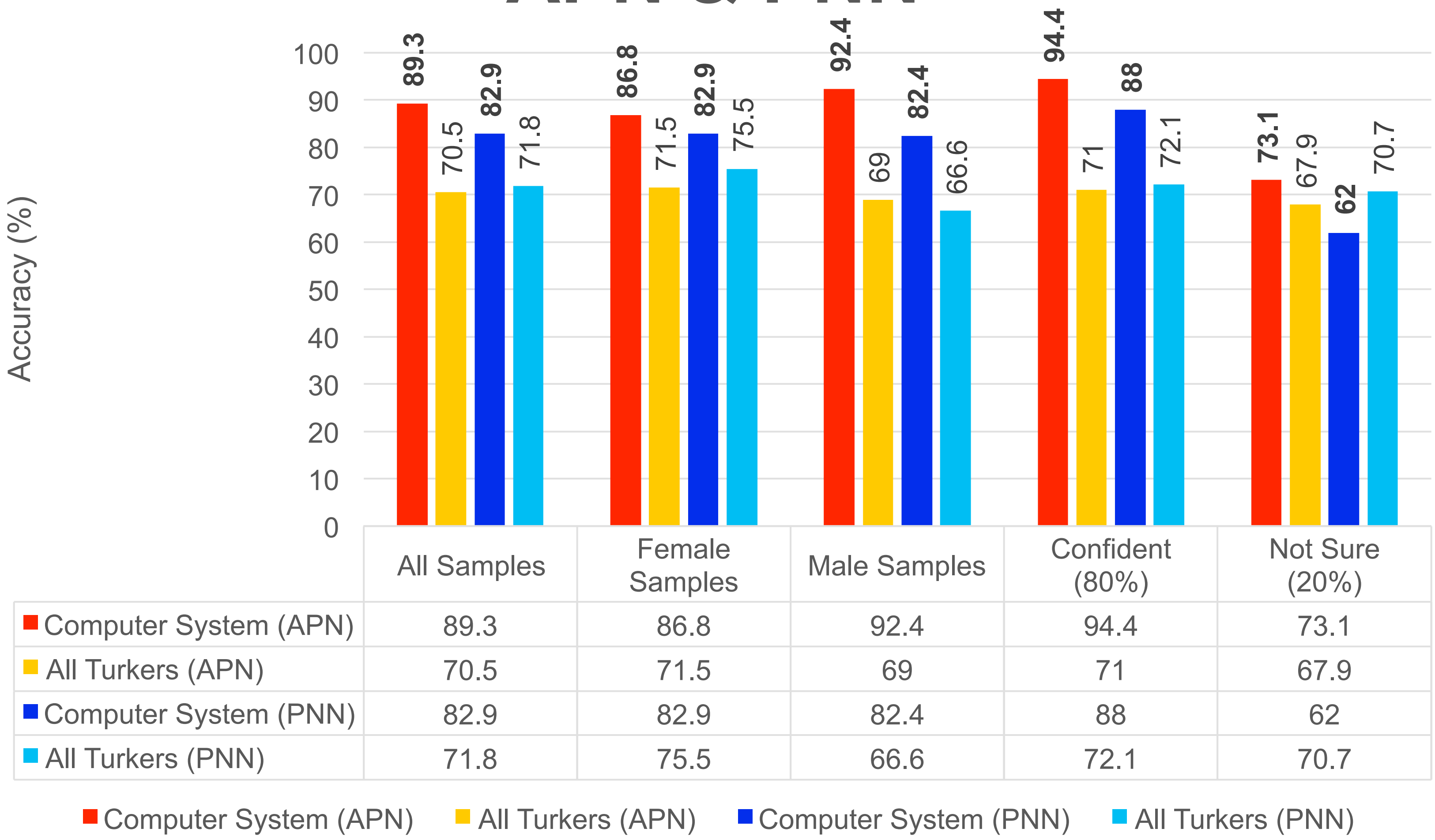
Turkers vs. Computer System: Emotions



	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

🟩 Computer System ▫ All Turkers 🟨 Female Turkers 🟪 Male Turkers

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

- This study compares naïve human coders with the automatic emotion classification system.
- The automatic system achieves much better accuracy in almost all cases.
- The automatic system can improve the 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
- The results show that it is feasible to replace naïve human coders with automatic emotion classification systems.

The End…

Thank you!