WISE: Web-based Interactive Speech Emotion Classification

Sefik Emre Eskimez*, Melissa Sturge-Appleº, Zhiyao Duan* and Wendi Heinzelman*

*Dept. of Electrical and Computer Engineering
ºDept. of Clinical and Social Sciences in Psychology
University of Rochester, Rochester, NY

Motivation

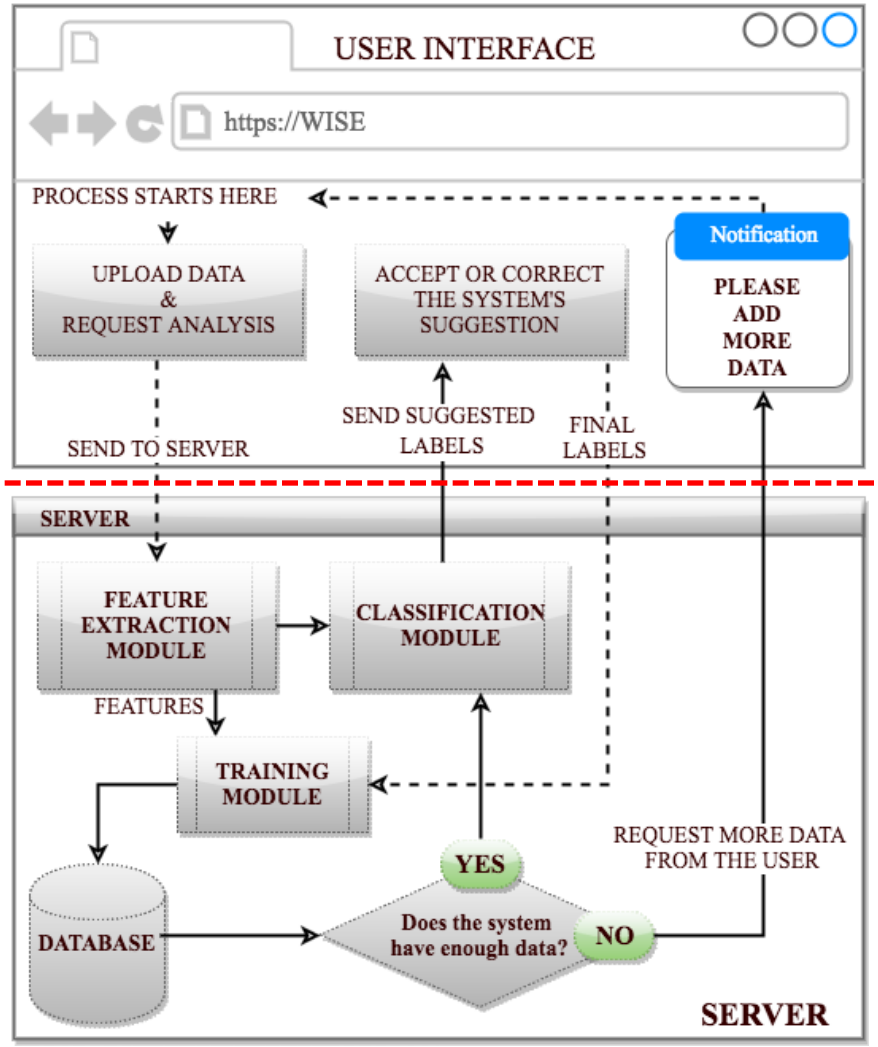
- Fully automatic speech emotion classification systems may not reflect the user's perceived emotions
- Manual speech emotion classification is costly and not efficient

Introduction

- WISE: a web based interactive speech emotion classification system
  - WISE has an automatic speech emotion recognition module, which is trained by a user's choices over time
  - WISE gives suggestions to users, which can be accepted or corrected by the user

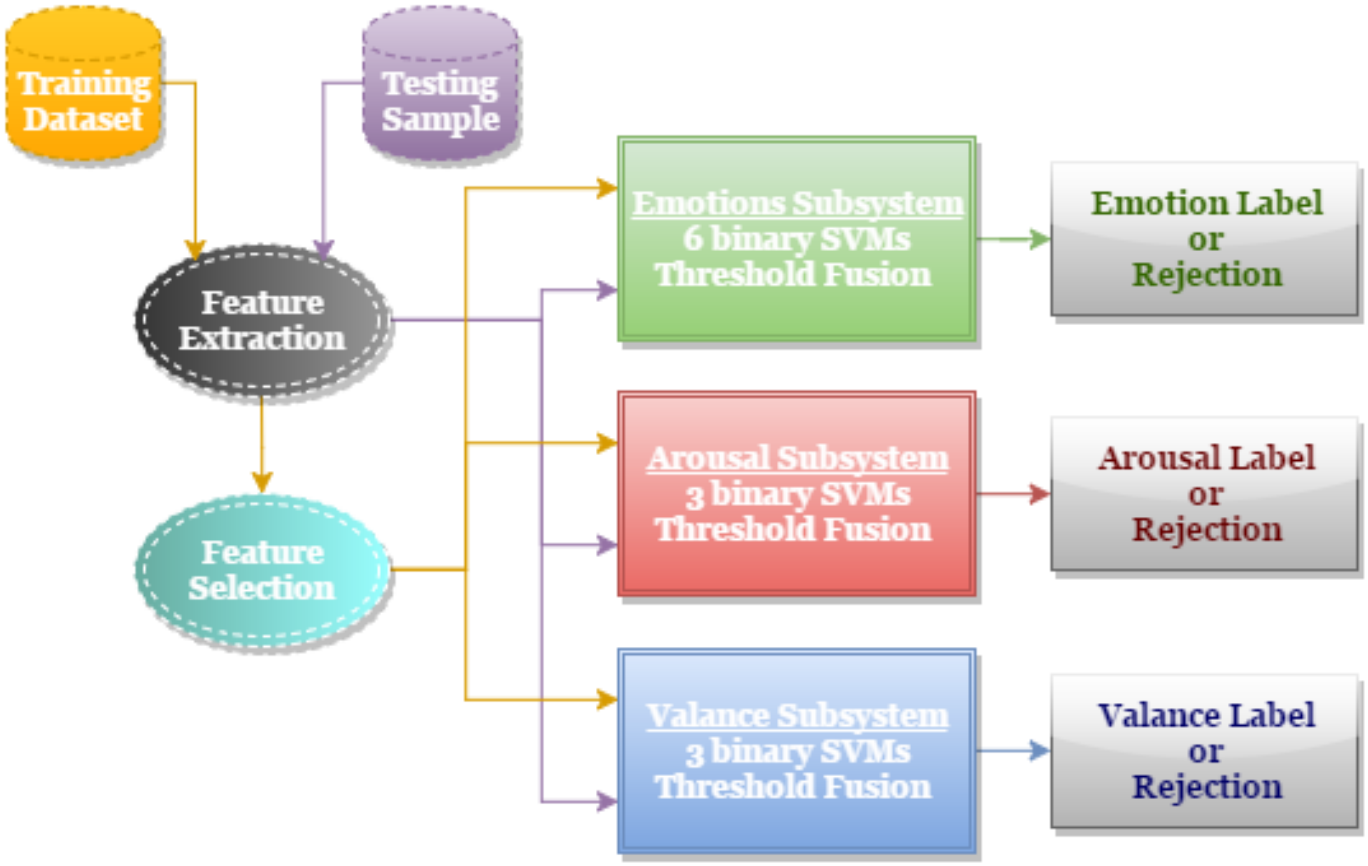
The WISE System Overview

Web-based Interface



Server Processes

Automatic Speech Emotion Classification Module



Features

- 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)
  - Support Vector Machine (SVM) Recursive Feature Elimination

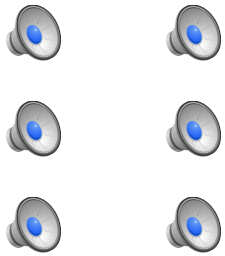
Feature name	#	Feature name	#
Fundamental Frequency (f0)	10	Spread	10
Energy	10	Skewness	10
Frequency and bandwidth for the first four Formants	80	Kurtosis	10
12 Mel-frequency Cepstral Coefficients (MFCCs)	120	Flatness	10
Zero-cross rate	10	Entropy	10
Roll-off	10	Roughness	10
Brightness	10	Irregularity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

Automatic Emotion Classifiers

- System uses binary SVM classifiers with RBF kernel for each emotion
  - 6 binary SVMs for first sub-system:
    - **anger, disgust, panic, happy, neutral, sadness**
  - 3 binary SVMs for second and third sub-systems:
    - Arousal Categories: **active, passive and neutral**
    - Valence Categories: **positive, negative and neutral**
  - Total of 12 binary SVMs

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



Experiments

- Simulating user interactions:
  - Divide dataset into training, validation and testing subsets
  - Steps of simulation:
    1. Classification module is trained with initial training subset
    2. Models are evaluated on testing subset
    3. A single sample from validation subset is added to training subset
    4. Models are evaluated on testing subset
    5. Repeat until validation subset is empty
- In high level, user uploads a new sample in each iteration and models are evaluated to see if they are adapting to new data or not

Experiments

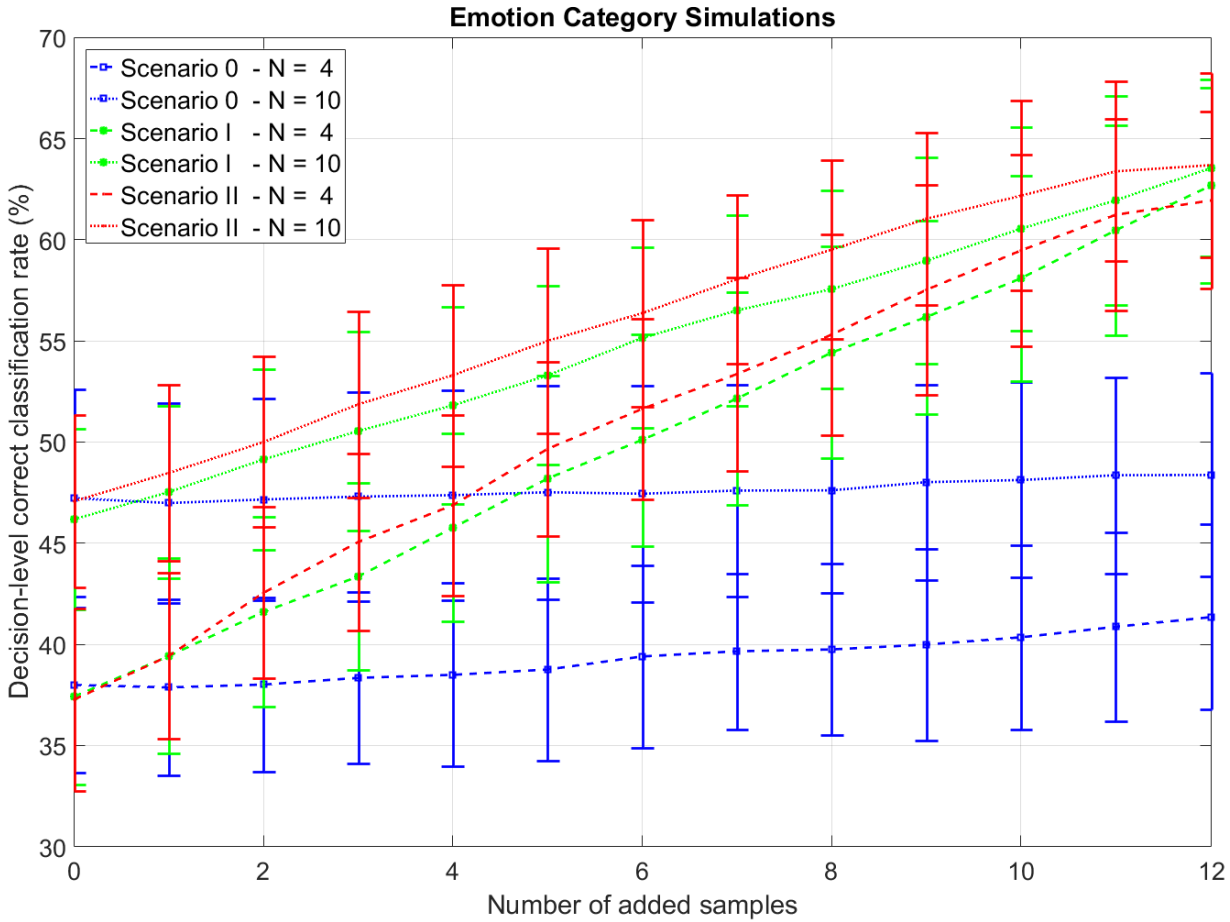
- Scenario 0 – *baseline – no adaptation*:
  - 6 out of 7 speakers' data are used for training and validation data
  - Testing data is chosen from the remaining speaker
  - This is repeated for all speakers and results are averaged over 7 speakers and 200 trials
  - Validation data is known to the system
- Scenario I – *simulation of "system adapting to user upload data"*:
  - Same setting as scenario 0, except validation data is chosen from the remaining speaker
  - Validation data is unknown to the system

Experiments

- Scenario II – *simulation of "system requesting ground-truth from user"*:
  - Same setting as scenario I, except the validation data is ordered in according to their classification confidence level in the system, and the least confident sample is added to the system in each round:
    - The system chooses a sample which it has least information on from the validation subset
    - Adds it to training subset
    - The models are evaluated on testing subset
    - Repeat until validation subset is empty
  - This scenario is beneficial when adding more data is costly

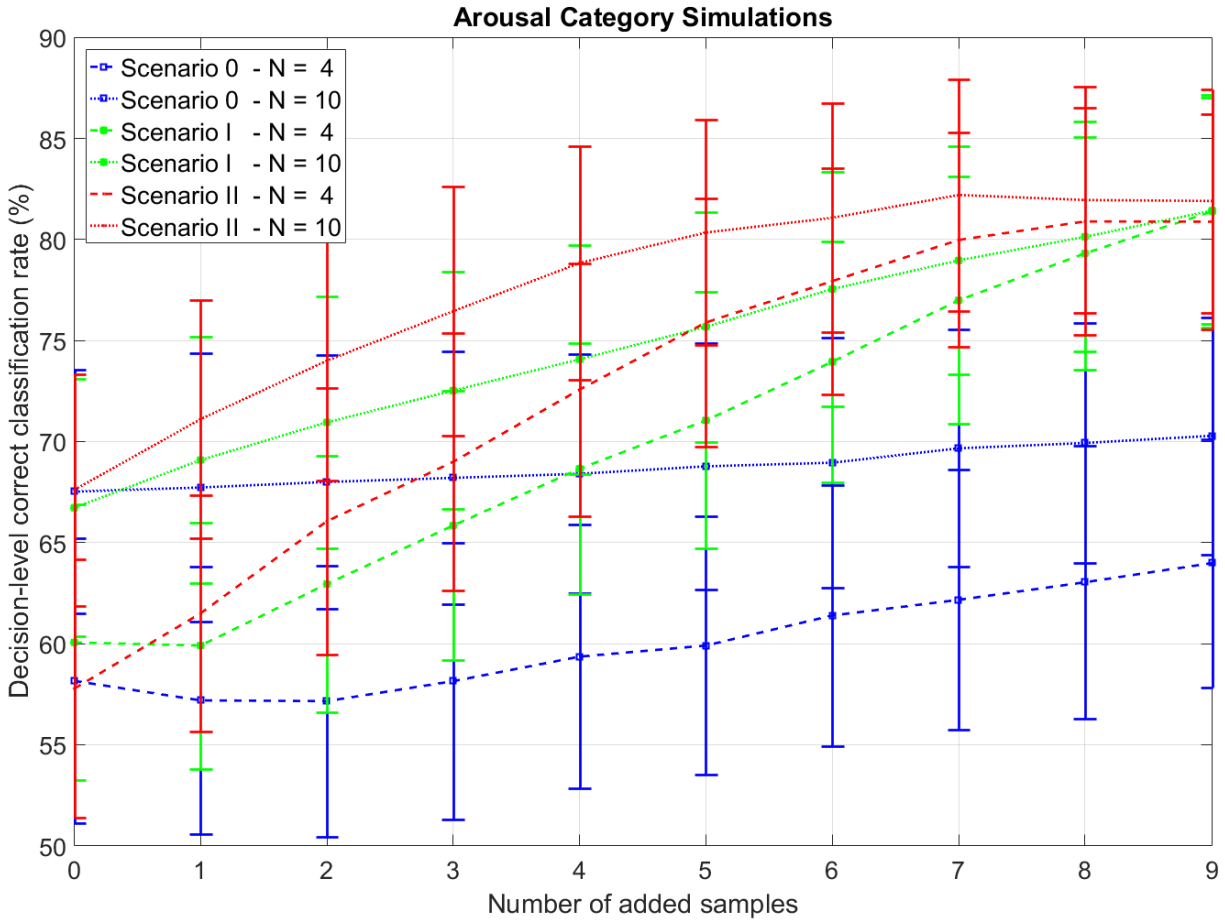
Experiments – Emotion Category

- N is the number of samples for each class in training
- Validation data has 2 samples from each class, total of 12



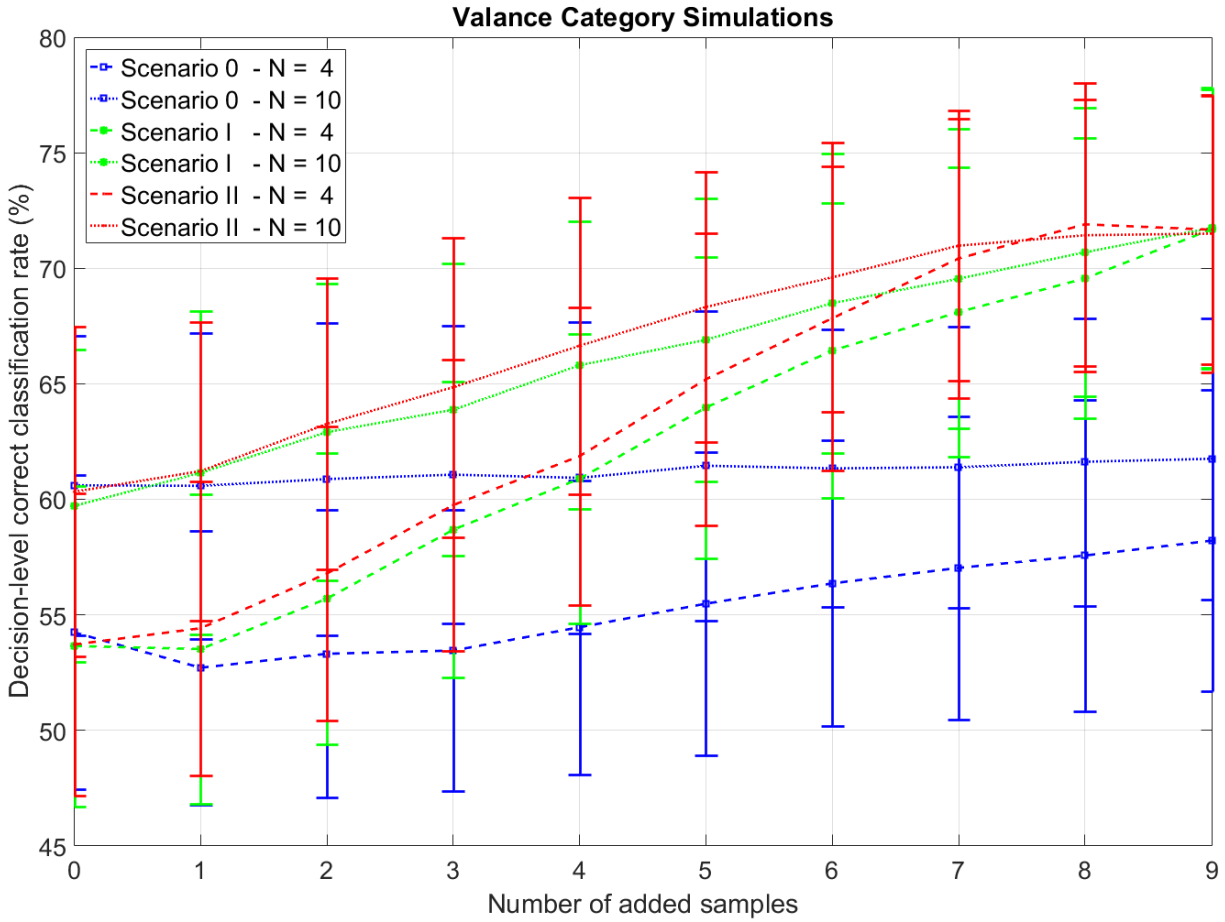
Experiments – Arousal Category

- N is the number of samples for each class in training
- Validation data has 3 samples from each class, total of 9



Experiments – Valance Category

- N is the number of samples for each class in training
- Validation data has 3 samples from each class, total of 9



WISE WEB ACCESS

http://system.wise.audio

Conclusion

- In this study, The WISE system is introduced and evaluated
- The WISE system is available for the community to use
- Evaluation results show that the system can adapt to a user's emotional choices over time
  - Future work: user study

The End…

Thank you!