Goal

- Enhance existing speaker recognition and phrase verification systems by creating the ability to detect distress in a subject’s voice
- The system will compare features extracted from the user’s speech by using existing “machine learning” training and classification techniques

Motivation

- Verification of employee hours in tandem with existing login systems that can confirm the identity of individuals entering facilities
- Confirmation that routine tasks are carried out
- Maintaining a safe working environment by detecting emergency situations
- Personnel and visitor tracking to identify the whereabouts of individuals on the premises
- Detect distress of personnel to determine crisis and help identify emergency scenarios
- Passively monitor mental health of personnel by tracking mood, i.e., depression

Project Challenges

- Mitigating the effects of ambient noise that exists in voice samples
- Choosing best voice characterization method to extract relevant features to improve classification of subjects
- Choosing appropriate functions to better cluster data for classification
- Difficulty implementing system in live demo aspect

Methodology

1. Raw Training Data
   - (Normalized & extracted features)
2. Raw Test Data
   - (Normalized & extracted features)
3. Feature Storage
4. Training
5. Accept/Reject
6. Similarity measure

Analysis and Results

The figure above depicts the decisions made on the voice sample per frame. As seen, there is a high concentration of classification on speaker 4, which is the speaker the system correctly classified.

The figure below portrays the confidence level of the system per frame when making a decision.

The closer the confidence value is to 1, the more certain the system is while identifying the speaker.

Acknowledgements

- Dr. Dario Pompili for his support throughout the R&D of our capstone design
- Guidance from PhD student Çağdaş Karatas
- Brendan Jenkins, Executive Director of recap video
- Machine learning utility creates a model from training data to be compared against test data for classification purposes
- Raw Data (Training or Testing): Recorded Audio Samples
- Normalization: Necessary for accurate feature extraction and classification
- Feature Extraction: Mel Frequency Cepstrum Coefficients extracted from voice sample
- Feature Storage: Raw audio files are discarded; extracted features are saved
- Training: Using feature data, Support Vector Machine (SVM) algorithm builds a training structure
- Similarity Measure: SVM prediction function measures similarity between training & test data
- Accept/Reject: Graphical User Interface returns classification per frame and consensus decision

References