



**CALIFORNIA STATE SCIENCE FAIR
2017 PROJECT SUMMARY**

Name(s) John Kim	Project Number S1515
Project Title EmNet: Emotion Recognition from Human Voice Using Machine Learning for Affective Computing	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The goal of this research is to develop an algorithm that can analyze and accurately recognize the speaker's emotion from human speech using machine learning. This is not only a challenging research problem but also has great potential for real-world applications. Adding human emotions as context information can help transform the emerging human-machine interactions to the next level, especially when video capture of human facial gestures is not a preferable option for the user due to privacy.</p> <p>Methods/Materials Conventional approaches to recognize human emotion use pattern recognition on the static features, obtained by taking the mean and variance of time-varying characteristics of feature vectors. However, the procedure to obtain the static features eliminates important information contained in the temporal information of feature vectors. "EmNet," the proposed method in this project, allows the neural network to learn this unknown temporal trajectory. The system consists of feature extraction, followed by Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks.</p> <p>The data used for this project comes from the Berlin Emotional Speech Database (EMO-DB). The database consists of 535 utterances from 10 talkers, with each utterance representing one of seven different emotional states: anger, boredom, disgust, fear, happiness, and sadness.</p> <p>Results The recognition rate achieved by the proposed method reaches above 86%, which is much higher than the 77.3% obtained from the conventional method using Support Vector Machine (SVM). This is a significant error rate reduction by about 40% over the conventional approach.</p> <p>Conclusions/Discussion A new method is proposed to recognize emotion from human speech. The method consists of 1) acoustic speech analysis to maximize the efficiency of well-known feature extraction and 2) machine learning to learn an unknown mechanism of temporal information processing. Validation on other databases and languages remains as further work.</p>	
Summary Statement EmNet is the proposed method for emotion recognition from human speech, comprised of acoustic feature extraction and machine learning, and demonstrates an error rate reduction of about 40% compared to the static approach.	
Help Received Received help to get appropriate packages for feature extraction.	