# Analyzing Emotions in the Fearless Steps Corpus

*Ashish Rawat, Daniel Lambert*

EESC 6366 - Speech and Speaker Recognition
5 May 2018

Ashish.Rawat@utdallas.edu
Daniel.Lambert@utdallas.edu

## Abstract

This project deals with analyzing emotional states of the speaker in the Fearless Steps corpus. We consider all the speech segments to fall into the broad emotional categories of happy, sad, loud (anger and disgust) or soft (boredom, fear and anxiety). Three phases of the Apollo 11 mission (lift off, lunar landing, and lunar walking) are studied for all speakers and for specific speaker sets who contributed towards distinct diversity in results. This project uses a four mixture Gaussian mixture model (GMM) for each of four emotions, trained over the Berlin database of emotional speech. Acoustic features taken into consideration for training the GMM are pitch, zero crossing rate, energy density, and delta energy density. We analyze a two-dimensional representation of the four emotions contained within tested speech data, and address challenges related to emotion detection for the specific speech data provided. Challenges include the lack of emotional intensity in speakers, noise in the provided testing data, and silence regions in audio segments.

**Index Terms**: Emotions, Gaussian mixture models, Fearless Steps corpus, noise reduction, pitch, energy density, delta energy density, zero crossing rate, speech classification

## 1. Introduction

Emotion is an integral aspect of human interaction. It is an abstract concept which has been thought upon for years. Speech is one of the many ways by which humans communicate emotions, so it is natural to measure the various emotions present in acoustic speech data. Recognising emotions from speech is subjective to the listeners discretion, as it is also not easy to quantify combinations of emotions. This project is an attempt to analyze the emotional states of the speakers from the Fearless Steps corpus [1], which contains recorded audio from the first successful moon landing mission: Apollo 11.

Generally, emotion recognition from the speech is performed on data which has emotion tags for truth. These tags on the testing data help us understand and quantify the accuracy of the designed emotion recognition system. One of the challenges of using speech from the Fearless Steps corpus is that we do not have pre-tagged emotions for any segments of speech. Therefore, our project functions to reveal the emotional content of the speech data given to us after validating our analysis system with known truth data.

Also considering the environment, situation, and the subjects of the given speech data, we expect to detect a much smaller amount of strong emotional content verses that from television broadcasts or personal conversational speech.

## 2. Methods

We proceed to construct an emotion classification system by reducing the noise of the test speech data, and recognizing and classifying emotions.

### 2.1. Noise Reduction

The given data from Apollo 11 has background noise, primarily from the channel. This noise has a humming effect which is additive in nature. We attempted using the noisy data as-is and verified that the results for emotion analysis are not satisfactory. Erroneous results were observed where the emotion classifiers group speech primarily by the pitch of the background noise hum. As expected, the noise hindered the performance of the system, and therefore we use Audacity software [2] to perform noise reduction.

We perform noise reduction using Audacity chain processing functions on the given 100 hours of data. A chain is created using the default noise reduction option. Default parameters are noise reduction of -12dB, sensitivity of 6.00 and frequency smoothing of level 3. We apply this chain to the folders containing speech data for each channel in the three phases of the Apollo 11 mission, and save the reduced noise waveforms as .WAV files. While stronger reduction settings can attenuate the noise further, they also degrade the quality of speech. Therefore we do not increase the reduction gain any further to avoid corrupting the speech which can produce erroneous outcomes.

### 2.2. Emotion Recognition and Classification

The data with reduced noise is now used for emotion recognition using a four-mixture GMM in an algorithm inspired by [3]. The features calculated are energy density, delta energy density, pitch, and zero crossing rate. The pitch is calculated for each frame using the subharmonic-to-harmonic algorithm provided by [4]. These features are acoustic features which are independent of the linguistic properties of the given test and training data. Taking this fact into consideration, we have chosen the Berlin Database of Emotional Speech [5] which has training emotional speech data in German language. The resulting models are valid for the English speakers in the Apollo mission because modulations in pitch, intensity and other acoustic parameters are similar for the two languages [6].

We choose elementary, opposing emotions to represent the extremes of both axes: happy and sad on one axis, and loud and soft emotions on the other axis. We combine anger and disgust as *loud* emotions, and boredom and fear/anxiety as *soft* emotions. Next, we generate scores for detected emotions by calculating the GMM posterior probability, and plot the opposing emotions on two axes. Here we assume that detected emotions are a combination of multiple emotions. We consider the origin of the axes as neutral emotion, and expect our results to be

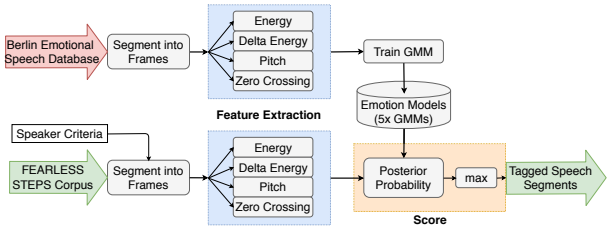


Figure 1: *Block diagram of the Matlab emotion analysis system.*

mostly clustered around the origin since we expect the speakers to have a neutral emotion overall.

### 2.3. Matlab Implementation

We use Mathworks Matlab to create a series of scripts (Figure 1) to process the dataset files, train the emotion models, and score the emotions of each segment of speech. First, the training script loads the German emotional speech files and calculates the feature vector for each 20ms frame of speech to train a GMM for each emotion. Next, we score each segment of speech in the Fearless Steps corpus as categorized by the speaker identification metadata included in the corpus. Last, we plot the log likelihood values for each of the four emotion models in a two-dimensional plot by taking the difference of opposing emotions as outlined in Section 2.2. For example, the happy-sad axis plots the log likelihood score of the happiness model (in the positive direction) with the subtracted score of the sadness model (to place it in the negative direction). Every segment of speech in the corpus placed at a two-dimensional position to illustrate the type and degree of emotions observed.

## 3. Results

We begin by validating the emotion analysis system using tagged English testing data, then proceed to examine the results from the Apollo 11 speech.

### 3.1. Validation

To validate the performance of our analysis system, we test the German-trained models on an English emotion corpus (RAVDESS) [7] to view the results for various known truth emotions in the database. Figure 2 shows the happiness emotion correctly classified by the analysis system, plotted in the upper quadrants of the two-dimensional plot, while Figure 3 shows the fear and calm emotions (both designated as *soft*) are correctly classified. Table 1 lists the accuracy results considering only one dimension of the plot, where we count the number of correctly-classified segments of speech corresponding to happy-vs-sad and loud-vs-soft emotions, respectively. The *soft* and happy emotions perform well, whereas the *loud* and sad emotions results are quite poor. We attribute this discrepancy to differences in enacting the specified emotions in a prompted setting. In the Apollo data, we expect the *loud* emotions to be less frequent in the formal mission control setting, and therefore the emotion classification system performance appears adequate for our application.

Figure 4 represents the distribution of the emotions of all the speakers in the Apollo 11 for each of the three phases. This result very well meets with our expectation of having major clusters around the origin to signify neutral speech among the mission personnel. Therefore, we conclude that the majority of

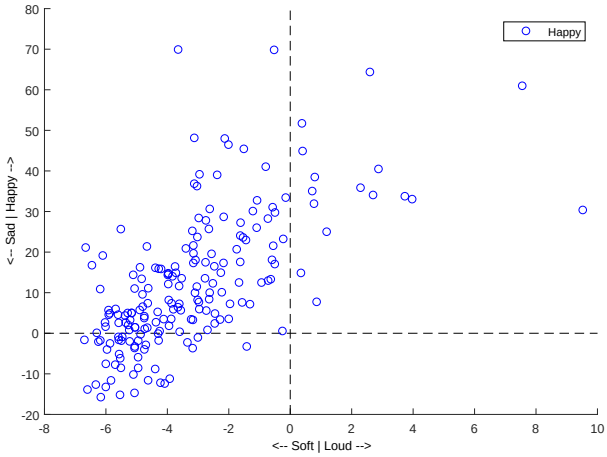


Figure 2: *Two-dimensional emotion space plot for RAVDESS happy emotion speech segments. 77.6% of the points are correctly classified in the upper two quadrants, revealing the success of the emotion analysis engine for recognizing happiness.*

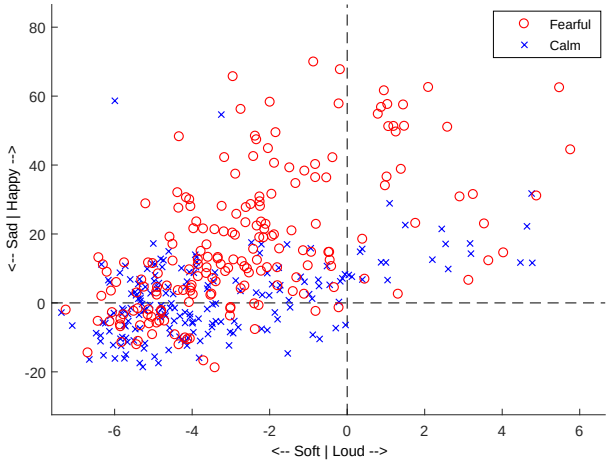


Figure 3: *Two-dimensional emotion space plot for RAVDESS fearful (red circle) and calm (blue cross) emotion speech segments, with 85.4% and 89.6% classification success rates, respectively, as possessing soft emotions in the left two quadrants.*

utterances during the mission revealed a relatively-neutral emotional state from their speech.

One more distinct observation is that the speakers had a quality of softness in their voice. Boredom, fear and anxiety are the key emotions that are being represented by soft emotions. There are very few speakers with loud states. Also, out of the three phases, lift off has the most speakers with loud emotions. A probable rationale could be that lift off is the most tense phase of the three. From Figure 4 we verify, by manually playing specific audio segments, that the segments satisfactorily justify their location on the plots.

Figure 5 represents the emotional states of the Capsule Communicator (CAPCOM) for the three phases. We observe that the speakers have been soft for all the three phases. We have the least data for CAPCOM speakers for the lift off phase and significantly more for the other two phases. Results are mostly soft emotions with equal distribution for happy and sad emotions.

Table 1: *Accuracy scores for the English RAVDESS corpus tested using the German-trained GMM emotion models, calculated by the single axis relevant for the truth emotion. Calm, fearful, and happy emotions display favorable results, while angry, disgust, and sad emotions show poor results.*

| Emotion | Correct | Total | Accuracy |
|---|---|---|---|
| Calm | 172 | 192 | 89.6% |
| Fearful | 164 | 192 | 85.4% |
| Angry | 43 | 192 | 22.4% |
| Disgust | 16 | 192 | 8.3% |
| Happy | 149 | 192 | 77.6% |
| Sad | 71 | 192 | 37.0% |

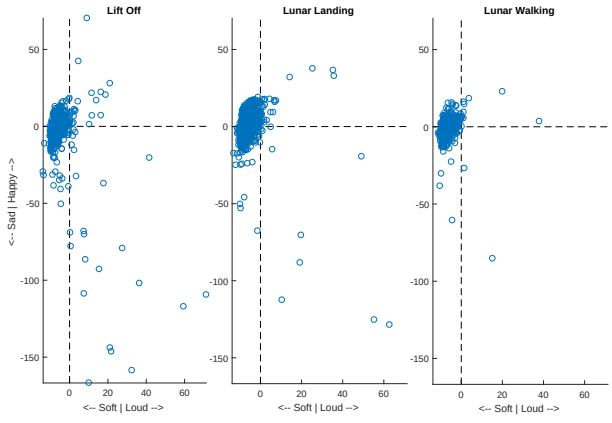


Figure 4: *Two-dimensional distribution of emotions for all speakers, divided into the three phases of the Apollo 11 mission.*

According to the observations in Figure 6, we understand that the Guidance speakers have been typically soft and mostly sad as compared to other channels. This observation holds true for all the three phases.

We distinctly observe that most of the speakers in the loud region in Figure 4 are from BKS. By plotting the only the BKS speaker in Figure 7, we see a clearer picture of the same locations. BKS has significant number of speakers having a higher score for loud emotions. This is verified by listening to some selected audio segments, where the results here are satisfactorily convincing.

## 4. Conclusion

We successfully accomplish tagging for the Apollo 11 Fearless Steps audio corpus by using an automated emotion classification system we designed in Matlab. This system was verified for its accuracy using the English RAVDESS data set as our testing data.

Tagging was performed uniquely by combining related emotions into the elementary categories of soft and loud. This tagging helped us analyze the Fearless Steps audio corpus by plotting four emotions on a two-dimensional plot representing the emotional state of the speakers.

For very few corrupted audio segments with high noise or extended periods of silence, the system erroneously tags one of the available emotions. Such erroneous results make up the majority of the extreme values visible in Figure 4.

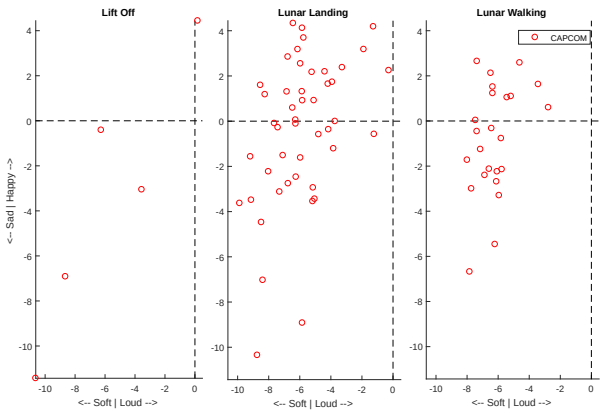


Figure 5: *Speech emotional states of the Capsule Communicator (CAPCOM) speakers for three phases.*

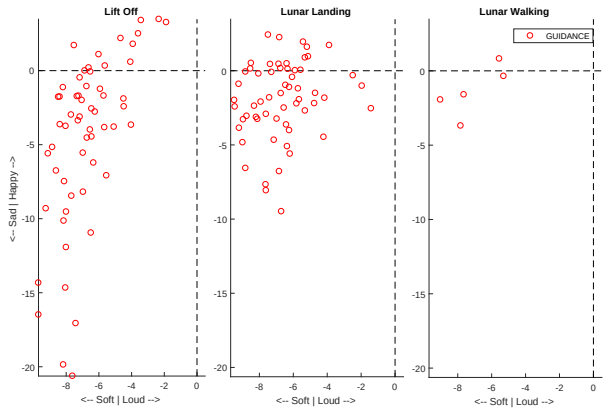


Figure 6: *Speech emotional states of GUIDANCE speakers for three phases.*

For better performance of emotion detection on the Fearless Steps corpus, we can utilize a more robust voice activity detection system instead of relying on the provided speaker ID metadata. This would help us score the audio segments with higher accuracy. Another task that could improve the performance of emotion detection would be adaptive noise reduction. Complete noise removal without tampering with the speech features would be an ideal way to proceed with emotion recognition and analysis in future works.

## 5. References

[1] J. H. Hansen, A. Sangwan, L. Kaushik, and C. Yu, "Fearless steps: Advancing speech and language processing for naturalistic audio streams from earth to the moon with apollo," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. 1868–1868, 2018.

[2] A. Team, "Audacity (version 2.2.2)," *Free audio editor and recorder*, 2018.

[3] V. Sethu, E. Ambikairajah, and J. Epps, "Speaker normalisation for speech-based emotion detection," in *Digital Signal Processing, 2007 15th International Conference on*. IEEE, 2007, pp. 611–614.

[4] Xuejing Sun. Pitch detection algorithm. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/1230-pitch-determination-algorithm

[5] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and

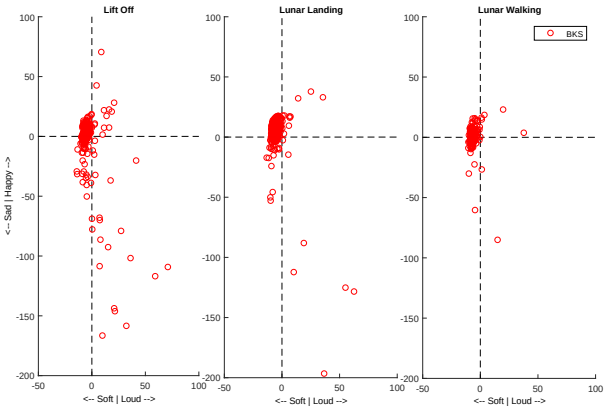


Figure 7: *Speech emotional states of BKS speakers for three phases.*

B. Weiss, "A database of german emotional speech," in *Ninth European Conference on Speech Communication and Technology*, 2005.

[6] W. Strange, O.-S. Bohn, S. A. Trent, and K. Nishi, "Acoustic and perceptual similarity of north german and american english vowels," *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1791–1807, 2004.

[7] S. R. Livingstone, K. Peck, and F. A. Russo, "Ravdess: The ryerson audio-visual database of emotional speech and song," in *Annual meeting of the canadian society for brain, behaviour and cognitive science*, 2012, pp. 205–211.