ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102971, October 2016*

Recognition of Anger, Irritation and Disgust Emotional States based on Similarity Measures

Hemanta Kumar Palo[1]*, Jyoti Mohanty[1], Mihir Narayan Mohanty[1], Mahesh Chandra[2]

[1]Department of Electronics and Communication Engineering, Siksha 'O' Anusandhan University, Bhubaneswar - 751030, Odisha, India; hemantapalo@soauniversity.ac.in, Jyotimohanty@soauniversity.ac.in, mihir.n.mohanty@gmail.com
[2]Department of Electronics and Communication Engineering, Birla Institute Technology, Ranchi - 835215, Jharkhand, India; shrotriya69@rediffmail.com

Abstract

Background/Objectives: The objective of this paper is to distinguish similar and overlapped emotional states like disgust and irritation from the primary angry state based on similarity measures. **Methods/Statistical Analysis:** The similarity test among these emotions have been carried out using the signal waveform, frequency coherence, power spectral density, log-likelihood score and Dynamic Time Wrapping (DTW) technique. **Findings**: Disgust state has more number of frequency coherence peaks which are leaning more towards unity like angry state. Further the phase angle between irritation and anger state is larger than between anger and disgust state. The minimum cost path using dynamic time wrapping technique founds to be 147.5917 between disgust and angry state as compared to 184.2386 between angry and irritation samples. From these analyses, it is concluded that, the primary anger state is more closure to disgust state than irritation state. The results are promising and we are able to put a boundary among these emotional states with our chosen techniques. **Application/Improvements:** Anger detection can lead to an improvement in human relationship and social system refinement. Demarcation of subcategory emotions that leads to anger state prevents confusion and can improve its recogntion.

Keywords: Emotional States, Dynamic Type Wrapping, Short Time Fourier Transform, Similarity Measure, Spectral Coherence

1. Introduction

As an intense state of emotional expression, anger is often encounterd in human behaviour due to preceived threat or provation. It is a strong emotional outburst when a person thinks his/her personal boundaries has been violated. The state can be stratified into three modalities: apprisals (cognitive), agitations and tension (affective) and antagonism and withdrawl (behavioral)[1]. Disgust and irritation are two subsidary less intense emotional states of emotion and are often overlapped with anger state hence create confusion for the recognizer. These state affects both body and mind that may reslult in confilict and break in human relationship. These human states are surfaced as protective mechanism due to percevied fear, sadness or humilating conditions. Detection of these related emotion by human machine interface is crucial in the study of behavioral mangement. It affects negatively to the social well-being and may cause emotional/Psychological trauma to the sorrounding people including the affected person. A dangerous feature of these expressive anger related states is an exponential rise in anger of others.

*Author for correspondence

A survey on literature of emotional speech recogntion reveals recogntion and comparison of different emotional classes such as boredome, sadness, fear, anger, surprise etc. in most of the cases[2–4]. However, very little research has been done to scale down a single class of emotion or similar emotions as a comparing platform. Lee *et. al.,* 2008 compared high and low anger states based on arousal level of the speech signal in call center application[5]. – Classification of anger emotion using linguistic and acoustic cues of speech signal using realistic emotional utterances of IVR and WoZ database has been explored and compared[6]. With fusion of these features the authors have reported an enhanced accuracy for IVR data set. Autoregressive features along with linear predictive spectrum and Mel-frequency cepstral coefficients have been utilized as a comparing standard for recognition of anger using far-end noisy telephone speech signal and Berlin (EMO-DB) database[7]. Similarly, an attempt has been made to put a boundary between two similar states as anger and disgust using pitch, formants, energy and auto-correlation based speech features in a neural network environment by the authors in [8]. However, these authors have approached the detection of anger state using different features or classification based models. However, to the best of our knowledge no effort has been made by the speech community in distinguishing similar and overlapped emotional states based on similarity measure. This has provided the research motivation to the authors in this direction.

The objective of this work is to demarcate sub-categorical emotional states as irritation and disgust as compared to the primary anger state. The work is directed to find the similarity among these states and observe the level of overlapping among these. There have been number of similarity measures applied in the field of medical imageing, for colabrative, document comarison, time series data mining, spatial clustering etc. [9–11] filtering techniques, Similarities are investigated using magnitude of the signal, spectral coherence, cross-correlation coefficients and dynamic time wrapping (DTW) technique. Application of DTW has been performed using the STFT (Short time fourier transorm) features of emotional utterances. Standard Berlin emotional speech database (EMO-DB) has been used for collection of anger and disgust utterances[12]. The irritation utterances are collected from different sources due to absence of any standard accessible dataset for our purpose.

The rest of the paper is organized as follows. A description on similarity measures applied to the chosen set of emotions are explained in section 2. Results on our findings are shown both graphically in section 3 along with detailed discussion. Section 4 concludes the work.

2. Similarity Measure

Three states of speech emotion as angry, irritation and disgust are compared here. Irritation and disgust can convey complementary angry states hence are secondary to angry state of mind, hence can be treated as similar. These states are often overlapped or displayed during hot conversations. The object is to correlate these similar states of emotions with angry state and single out the differences. In the process, thiry utterances of angry and disgust samples of German emotional database (EMO-DB) are taken. Since irritation emotional speech samples are either unaccessible or unavailable, seventy utterances of this emotional state has been collected from different sources. Out of which thirty utterances based on a listening test from five subjects of both gender has been chosen. The collected irritation signals are resampled to 16KHz sampling rate as that of EMO-DB database.

Following similarity measure methods has been used to distinguish the primary angry emotion against its secondary irritation and disgust state.

2.1 Similarity Measure using Signal Waveform

Signal amplitude plays a major role in establishing the type of emotion displayed. The magnitude shows the arousal level of the emotions hence provide energy information. People speak with higher magnitude when angry as compared to low arousal emotions such as bore and sad. However, irritation and disgust signals are allies of angry emotions. How far is these secondary emotions are closure to angry state? It has been least analysed for speech emotions. Hence an attempt is taken to compare these emotions to establish a relation among them based on the amplitude of the signal. We have taken the absolute value of the signal for measuring the similarity among the emotional states.

2.2 Similarity Measure using Cross-correlation Coefficients

Cross-correlation is an efficient tool to measure the similarity between two signals as a function of delay m between these signals. The cross-correlation relationship between the angry and disgust signal can be represented as

$$C_{angry,disgust}(m) = E[Angry^{*}(n+m)\ disgust(n+m)] \quad (1)$$

where, $Angry^{*}$ is the complex conjugate value. The correlation measure between angry and irritation samples has been performed similarly.

2.3 Similarity Measure using Frequency Coherence

A power spectrum displays the power present in each frequency. Let the power spectral density (PSD) for an angry sample is represent as

$$PSD_{angry}(\omega) = \lim_{T\to\infty} E\left[\left|\widehat{angry}_T(\omega)\right|^2\right] \quad (2)$$

where, w is the signal frequency, $\widehat{angry}_T(\omega) = \frac{1}{T}\int_0^T angry(t)e^{-j\omega t}dt$ is the Fourier transform of the signal and $E[\cdot]$ $E[\cdot]$ is the expected value. Subsequently, for disgust and irritation signals the PSDs can be indicated as $PSD_{disgust}(\omega)$ and $PSD_{irritation}(\omega)$ respectively. The frequency domain correlation among these signals can be identified using spectral coherence. Coherence of two signals has a value between zero and unity that establishes a fitting relationship between these two signals at each frequency. Ideally, coherence value inclined more towards unity indicates correlated frequency components. Correspondingly, uncorrelated signals have coherence values leaning more towards zero. The establishment of coherence between two signals is explained in following steps.

1. The signal is framed and windowed initially. A frame size of 25ms with 50% overlapping between frames has been chosen.
2. Extract the PSD of the windowed emotional samples.
3. A magnitude squared coherence (MSC) estimate between angry, disgust and angry, irritation using STFT feature vector is done. Welch's average periodogram method has been used for this in this work. The MSC between angry and disgust signal is given by

$$MSC_{Angry,disgust} = \left(\frac{abs\left(P_{Angry,disgust}\right)^2}{\left(P_{Angry,Angry}\right)\left(P_{disgust,disgust}\right)}\right) \quad (3)$$

where, $P_{Angry,Angry}$ and $P_{disgust,disgust}$ are obtained by averaging the MSC of the N-*DFTs* of angry and disgust states.

2.4 Similarity Measure using DTW

DTW has been applied for speech and emotional speech recognition using sequences of MFCC feature vectors[13–15]. The algorithm aims for optimal match between the feature vectors of two sequences by warping the time axis iteratively. The time-series algorithm is tested in this piece of work for angry vs. disgust and angry vs. irritation states using STFT features as a similarity measure.

Let's consider the angry and disgust emotional states. The feature vectors of these two states are arranged along the *x*-axis and *y*-axis of a grid starting from the top left corner of the grid as shown in result section 3. The corresponding STFT feature values of the two emotional states are compared based on a distance measure in each cell. Euclidean distance measure has been used with DTW algorithm earlier[13,14]. We have used the cosine distance between the absolute STFT feature values to find the optimized match between the emotional states as in [16]. The aim is to find and maximize the local match between the frames of the designated emotional samples. A similarity cost is found using the algorithm that indicates the extent to which these signals match. The algorithm is explained.

Initially, the STFT magnitude is extracted from the Hamming windowed signal. A frame size of 30ms with 25% overlapping between frames has been used. Let STFT features of angry denoted as $A_{STFT_m} = \{A_{STFT_1}, A_{STFT_2}, \dots, A_{STFT_j}\}$. Here, $j = 1, 2, \dots, J$ is the number of features per frame of an utterance. The STFT features of an angry utterance consisting of $q = 1, 2, \dots, Q$ number of frames is thus represented by

$$A_{STFT_{j,q}} = \{A_{STFT_{1,1}}, A_{STFT_{2,1}}, \dots, A_{STFT_{j,1}}\}, \{A_{STFT_{1,2}}, A_{STFT_{2,2}}, \dots, A_{STFT_{j,2}}\}, \dots, \{A_{STFT_{1,Q}}, A_{STFT_{2,Q}}, \dots, A_{STFT_{j,Q}}\} \tag{4}$$

The features for all the utterances $r = 1, 2, \dots, R$ of angry state is given by

$$A_{STFT_{j,q}}(r) = \{A_{STFT_{1,1}}(1), A_{STFT_{2,1}}(1), \dots, A_{STFT_{j,1}}(1)\}, \{A_{STFT_{1,2}}(2), A_{STFT_{2,2}}(2), \dots, A_{STFT_{j,2}}(2)\}, \dots, \{A_{STFT_{1,Q}}(R), A_{STFT_{2,Q}}(R), \dots, A_{STFT_{j,Q}}(R)\} \tag{5}$$

Similarly, for disgust state the STFT features of all utterances are given by

$$D_{STFT_{j,q}}(r) = \{D_{STFT_{1,1}}(1), D_{STFT_{2,1}}(1), \dots, D_{STFT_{j,1}}(1)\}, \{D_{STFT_{1,2}}(2), D_{STFT_{2,2}}(2), \dots, D_{STFT_{j,2}}(2)\}, \dots, \{D_{STFT_{1,Q}}(R), D_{STFT_{2,Q}}(R), \dots, D_{STFT_{j,Q}}(R)\} \tag{6}$$

Next to it, the cosine distance between the absolute STFT feature values is computed to construct a local match score matrix. The cosine angle between these STFT vectors or the normalized inner product is estimated as

$$E_{Angry} = \sqrt{\sum_{r=1}^{R} \left[A_{STFT_{j,q}}(r)\right]^2} \tag{7}$$

$$E_{Disgust} = \sqrt{\sum_{r=1}^{R} \left[D_{STFT_{j,q}}(r)\right]^2} \tag{8}$$

$$Local\ match = \frac{\left(A_{STFT_{j,q}}(r)\right)' \left(E_{Disgust}\right)}{\left(E_{Angry}\right)' \left(E_{Disgust}\right)} \tag{9}$$

This way a full local match matrix is constructed by estimating the distance between features of each pair of frames of disgust state from that of angry state. A path is established through the cells of the grid that will minimize the overall distanced between the feature values. This will provide the highly similar feature values shown as dark line along the leading diagonal in the result section. Now the dynamic programming algorithm is used to find the minimum cost path from the opposite corners of the local match matrix or cost matrix. Similarly, the minimum cost values of angry and irritation state are computed by constructing a local match matrix of their absolute STFT feature values as explained earlier.

2.5 Similarity Measure using Likelihood Ratio

Likelihood measure is a statistical test used to find the goodness of fit between similar signals. For the reference angry emotion tested against irritation and disgust states, the measure will provide the desired optimized relationship among them. STFT features of these emotions are extracted initially. The optimized values of STFTs are used as a basis for similarity measure using likelihood ratio. Both irritation and disgust states are tested individually against the angry state in this case. The method is explained.

Let, ρ is a notional parameter used to model angry, disgust and irritation states based on the null (M_0) and alternative (M_1) hypothesis which can specify these classes optimally. Assuming a fixed value for ρ, we can specify these hypotheses as

$$\begin{aligned} M_0 &: \rho_0 \\ M_1 &: \rho_1 \end{aligned} \tag{10}$$

The likelihood ratio Λ can be expressed as [17,18].

$$\Lambda(r) = \frac{S\left(\rho_0/_d\right)}{sup\{S(d):\ \rho \in \{\rho_0, \rho_1\}\}} \quad (11)$$

Here d is chosen based on the similarity among features. The terms $S\left(\rho_0/_d\right)$ and "sup" signify the likelihood and supremum function respectively. The governing rule for the decision of likelihood function is generalized.

If $\Lambda > y,\quad 0 \le y \le 1,\quad \rho_0$ is selected, else if $\Lambda < y,\quad 0 \le y \le 1,$ then ρ_0 is discarded. ρ_0, is also discarded with probability 'p' when $\Lambda = y$.

Selection of y and d is made owing to some fixed threshold β as per the equation given by

$$p \cdot P\,(\Lambda = y|\rho_0) + P(\Lambda < y|\rho_0) = \beta \quad (12)$$

This way for negligibly small likelihood ratio statistic, a null hypothesis is discarded based on β. Here, β signifies the rejection of a true null hypothesis i.e. the extent to which the type I error probability can be tolerable.

3. Results and Discussion

In Figure 1, irritation and disgust signals are shown in the first and second subplots. The objective is to find how the angry signal shown in the third subplot differs from the other two signals. It is observed from the time series plot of Figure 1 that angry (subplot 3) is more closure to disgust (subplot 2) than irritation (subplot 1). Disgust and angry signals have higher amplitudes than irritation signal. Thus a person speaks at lower energy when irritated as compared to angry or disgust state of mind. Further, disgust signal is more similar to angry compared to irritation signals as the shape of the waveforms reveal.

In Figure 2, the presence of correlation coefficients of angry state in irritation and disgust utterances is shown. Higher peak of the cross-correlation coefficient in subplot implies that angrier information is present in disgust utterances than irritation samples. For angry signal, maximum absolute cross-correlation coefficient amplitude of 41.7336 is reported when compared with disgust signal. The value is 9.9329 between angry and irritation states. The angry state is present in subplot 2 starting after -84.75ms using -1.5s to 1.5s time axis range as shown in this figure. Thus, the disgust signal lags the angry signal. From subplot 1 it is however found that the angry signal is present in irritation signal with a lead of 11.66ms.

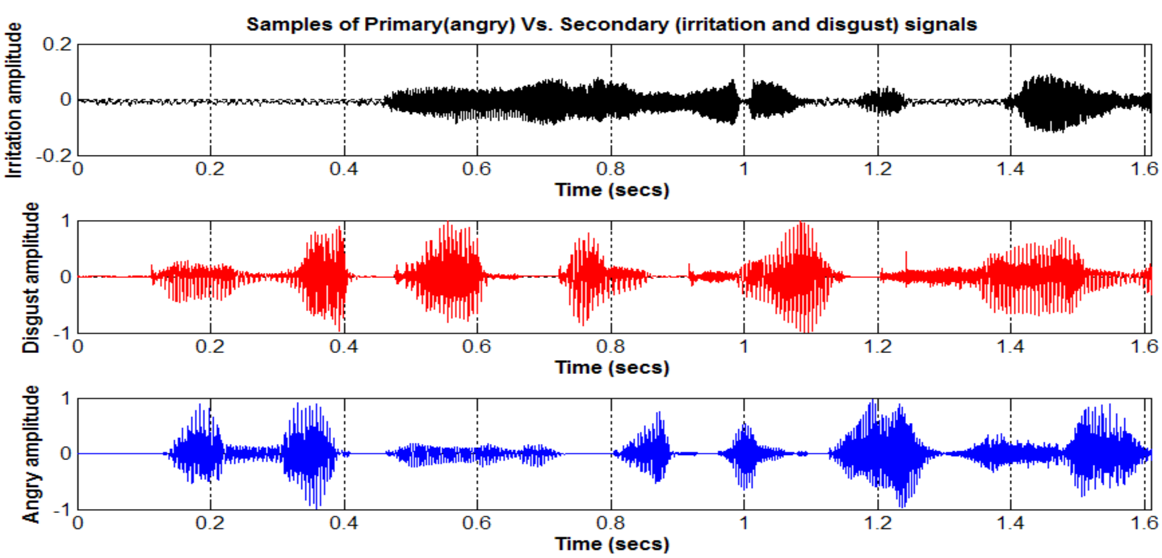

Figure 1. Similarity measure between angry, disgust and irritation samples based on signal waveform.

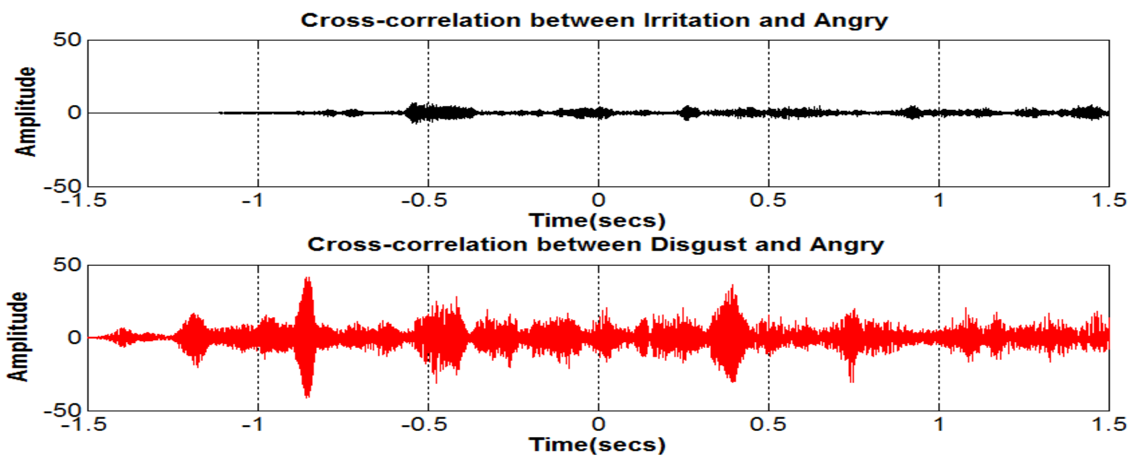

Figure 2. Similarity measure of angry, disgust and irritation samples based on cross-correlation.

Therefore, it can be concluded that irritation is less correlated to angry than disgust state.

The plots of power spectrum for angry, disgust and irritation signals have been shown in the Figure 3. It is concluded that the irritation and disgust signals are more correlated and similar type of emotions. The magnitude/HZ indicates that these signals are correlated to angry emotions in a large extent.

From PSD plot it is seen that, most of the signal components are lying in the range between 0 to 1000Hz. Hence this range is used as our analysis band for coherence estimation. The signals are tested for possible coherence peaks between 0-200Hz, 200-600Hz, and 400 to 850Hz range. Most of the coherence peaks among signals have appeared between 400 to 850Hz frequency range and is plotted in Figure 4. It has been observed that disgust sig-

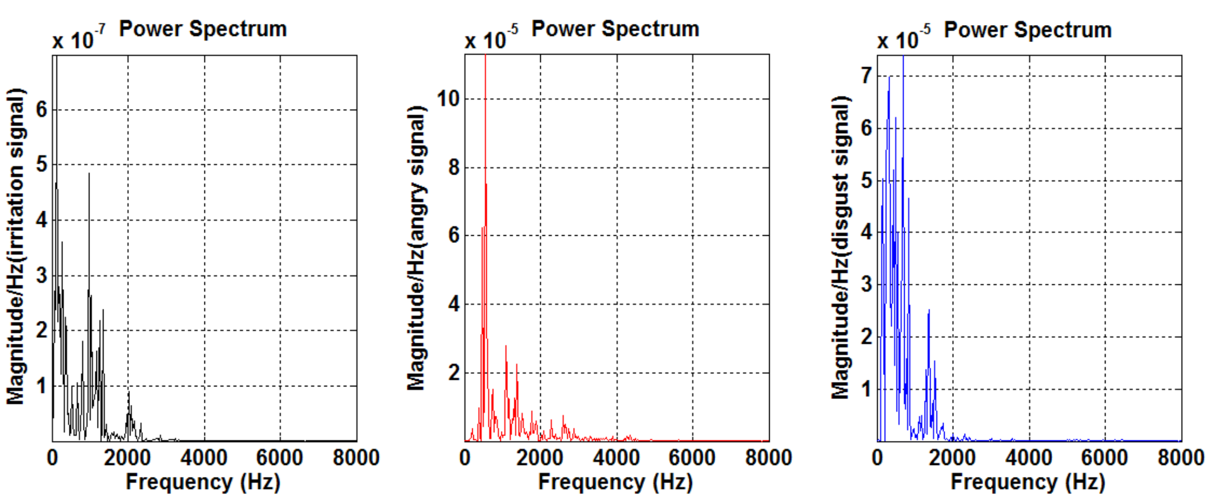

Figure 3. Similarity measure of angry, disgust and irritation samples based on power spectrum.

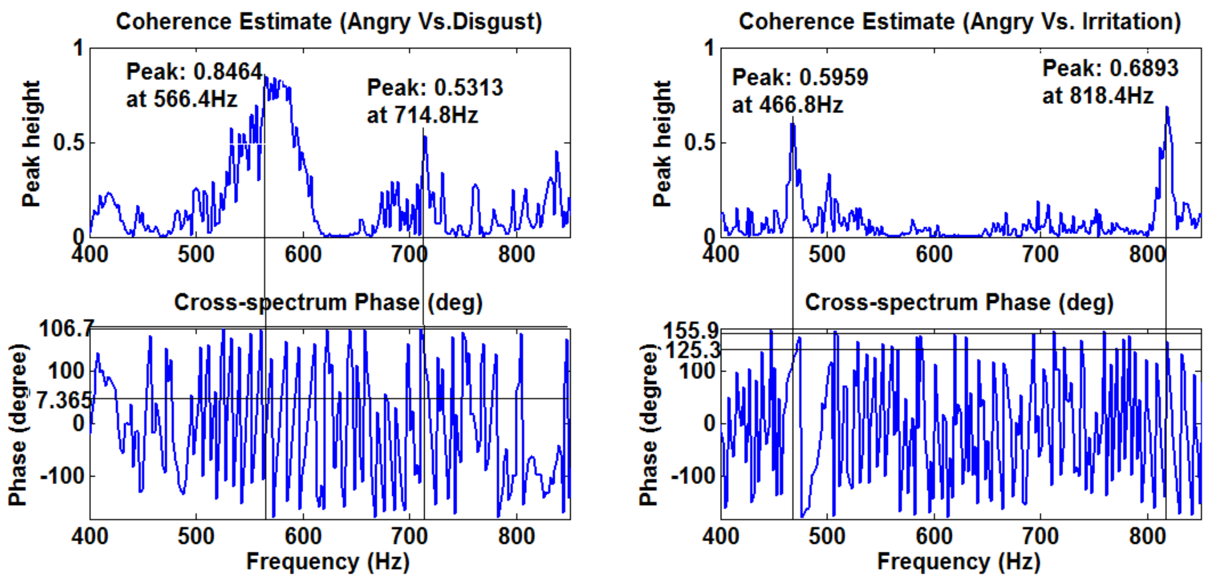

Figure 4. Similarity measure of angry, disgust and irritation samples based on frequency coherence.

nal is more correlated to angry state. These signals have two correlated peaks of value above 0.5. These peaks are 0.8464 and 0.5313 occurring at the frequency of 566.4Hz and 714.8Hz respectively. These peaks are leaning more towards unity indicates the correlation characteristic of angry with disgust. As compared to this the irritation and angry signals have these correlated peaks of 0.5959 and 0.6893 at 466.8Hz and 818.4 Hz respectively. Since the peaks are inclined less towards unity as compared to angry and disgust, hence are more uncorrelated. The phase difference between angry and disgust is 7.365 degree and 106.7 degree around frequency 566.4Hz and 714.8Hz respectively for the signals under analysis. Uncorrelated angry and irritation signals have larger phase angles of 125.3 and 155.9 degree around 466.8Hz and 818.4 Hz respectively.

The total `similarity cost' using DTW algorithm provides the similarity between the sample and template as an indication of best match. The dark line in the diagonal indicates the lowest cost path as a measure of suitable

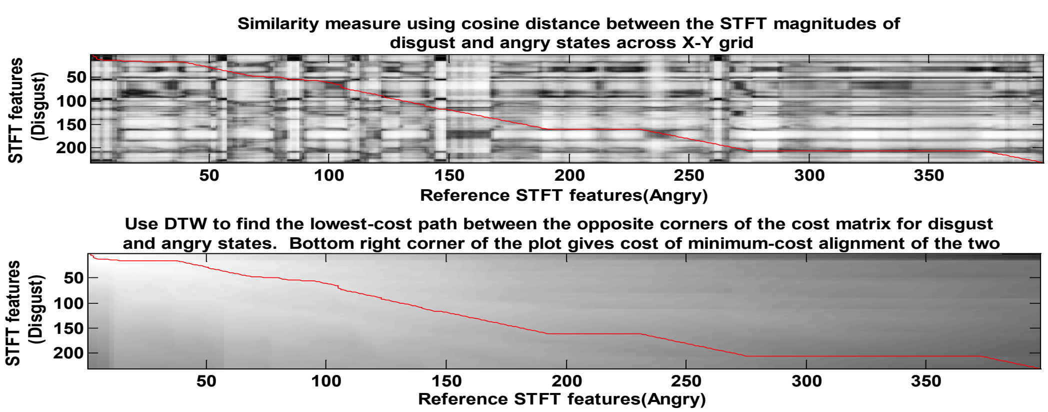

Figure 5. Similarity measure of angry and disgust states using DTW.

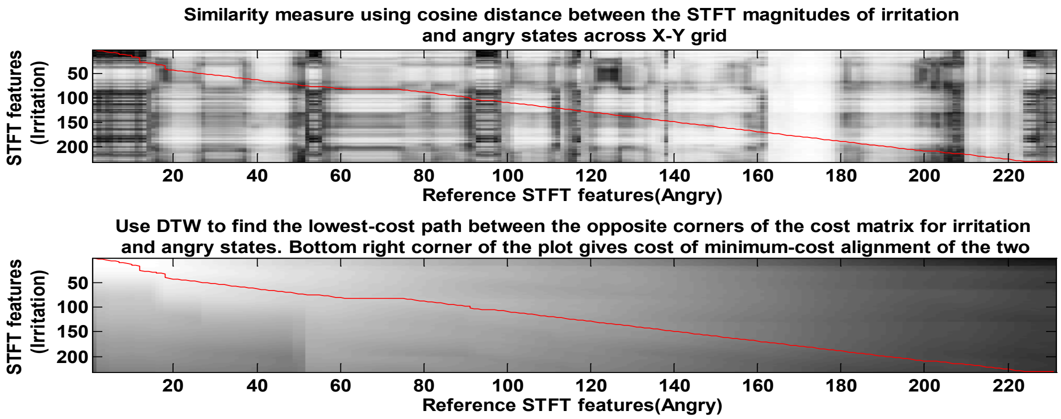

Figure 6. Similarity measure of angry and irritation states using DTW.

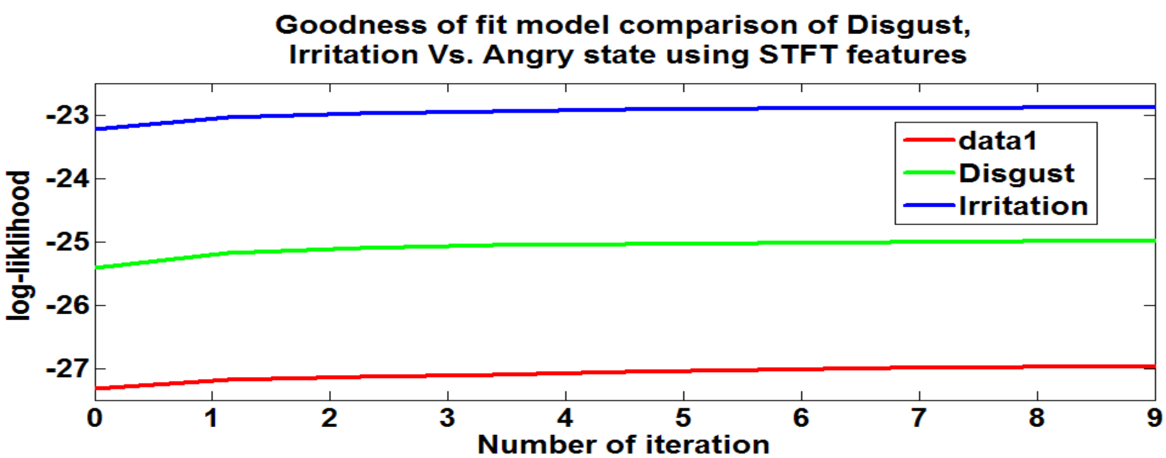

Figure 7. Similarity measure of angry, disgust and irritation states based on log-likelihood score.

match as shown in Figure 5 between disgust and angry state. Similar plot has been shown for irritation and angry state in Figure 6. The minimum cost path founds to be 147.5917 between disgust and angry state as compared to 184.2386 between angry and irritation samples. It indicates the more closeness of disgust signal to angry state than irritation state to angry state

Figure 7 shows the likelihood model of disgust and irritation state against the angry states graphically. A likelihood model is generated for each of the emotional states using the optimized STFT feature values. Based on the probability distribution of likelihood values, a lower probability of occurrence of the values under the null hypothesis as compared to the values under the alternative hypothesis is an indication of similarity. A better fitting model results for Lower value. It can be hypothesized that, a likelihood values more negative or leaning more towards zero is an indication of better fitting model. Since the disgust states are more closure having more negative values to angry state values, hence it is more similar to angry state than irritation state.

4. Conclusion

A number of sub-categories of emotional states associated with primary angry emotion often lead to confusion in its detection. This piece of work has taken an attempt to separate two sub-category emotion such as irritation and disgust from the angry emotional state based on different similarity measuring techniques. A boundary is drawn between the sub-categories emotional states against the primary state of emotion as reflected in result section. Also, a comparison on irritation and disgust states are made based on closeness of these states with the angry state.

5. References

1. Raymond N. Anger as a clinical and social problem. Advances in the study of aggression. Academic Press: London. 1986:2.
2. Ayadi EM, Kamal MS, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition. 2011 Sep; 44(3):572–87.
3. Subhashree R, Rathna GN. Speech emotion recognition: Performance Analysis Based On Fused Algorithms and GMM modelling. Indian Journal of Science and Technology. 2016 Mar; 9(11):1–8.
4. Vaid S, Singh P, Kaur C. Classification of human emotions using multiwavelet transform based features and random forest technique. Indian Journal of Science and Technology. 2015 Oct; 8(28):1–7.
5. Lee FM, Li LH, Huang RY. Recognizing low/high anger in speech for call centers. Proceedings of 7th International Conference on Signal Processing, Robotics and Automation. World Scientific and Engineering Academy and Society (WSEAS), University of Cambridge, UK; 2008. p. 171–6.
6. Polzehl T, Schmitt A, Metze F, Wagner M. Anger recognition in speech using acoustic and linguistic cues. Speech Communication. 2011 Oct; 53(9–10); 1198–209.
7. Pohjalainen J, Alku P. Automatic detection of anger in telephone speech with robust autoregressive modulation filtering. Proceedings of ICASSP, Vancouver, Canada; 2013. p. 7537–41.
8. Palo HK, Mohanty MN. Classification of emotions of angry and disgust. Smart Computing Review. 2015 Jun; 5 (3):151–8.
9. Karhikeyan C, Ramadoss B. Comparative analysis of similarity measure performance for multimodality emage fusion using DTCWT and SOFM with various medical image fusion techniques. Indian Journal of Science and Technology. 2016 Jun; 9(22):1–6.
10. Saranya KG, Sadasivam GS, Chandralekha M. Performance comparison of different similarity measures for collaborative filtering technique. Indian Journal of Science and Technology. 2016 Aug; 9(29):1–8.
11. Kalpana S, Vigneshwari S. Selecting multiview point similarity from different methods of similarity measure to perform document comparison. Indian Journal of Science and Technology. 2016 Mar; 9(10):1–6.
12. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W, Weiss B. A database of German emotional speech. Proceedings of the Interspeech, Lissabon, Portugal; 2005. p. 517–1520.
13. Muda L, Begam M, Elamvazuthi I. Voice recognition algorithms using MEL Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. Journal of Computing. 2010 Mar; 2(3):138–43.
14. Krishna NM, Lakshmi PV, Srinivas Y, Devi JS. Emotion recognition using dynamic time warping technique for isolated words. International Journal of Computer Science. 2011 Sep; 8(5):306–9.
15. Pawar VKR, Patel N. Emotion recognition from hindi speech using MFCC and sparse DTW. International Journal of Engineering Research and Technology. 2015 Jun; 4(6):1–5.
16. Turetsky R, Ellis D. Ground-truth transcriptions of real music from force-aligned midi syntheses. 4th International Symposium on Music Information Retrieval ISMIR-03, Baltimore; 2003. p. 135–41.
17. Casella G, Berger RL. Statistical Inference, 2nd edn. Duxbury Press; 2001.
18. Mood AM, Graybill FA. Introduction to the theory of statistics, 2nd edn. McGraw-Hill; 1963.