Workshop on Speech, Music and Mind 2018
01 September 2018, Hyderabad, India

Discriminating between High-Arousal and Low-Arousal Emotional States of Mind using Acoustic Analysis

Esther Ramdinmawii[1] *and V. K. Mittal* [2]

[1]Indian Institute of Information Technology Chittoor, Sri City, A. P. , India
[2]CEO, Ritwik Software Technologies Pvt. Ltd., Hyderabad, Telangana, India
[1]esther.rm@iiits.in, [2]DrVinayKrMittal@gmail.com

Abstract

Identification of emotions from human speech can be attempted by focusing upon three aspects of emotional speech: valence, arousal and dominance. In this paper, changes in the production characteristics of emotional speech are examined to discriminate between the high-arousal and low-arousal emotions, and amongst emotions within each of these categories. Basic emotions *anger, happy* and *fear* are examined in high-arousal, and *neutral* speech and *sad* emotion in low-arousal emotional speech. Discriminating changes are examined first in the excitation source characteristics, i.e., instantaneous fundamental frequency (F0) derived using the zero-frequency filtering (ZFF) method. Differences observed in the spectrograms are then validated by examining changes in the combined characteristics of the source and the vocal tract filter, i.e., strength of excitation (SoE), derived using ZFF method, and signal energy features. Emotions within each category are distinguished by examining changes in two scarcely explored discriminating features, namely, zero-crossing rate and the ratios amongst the spectral sub-band energies computed using short-time Fourier transform. Effectiveness of these features in discriminating emotions is validated using two emotion databases, Berlin EMO-DB (German) and IIT-KGP-SESC (Telugu). Proposed features exhibit highly encouraging results in discriminating these emotions. This study can be helpful towards automatic classification of emotions from speech.

Index Terms: high/low arousal emotions classification, ratios of signal sub-band energies, zero crossing rate, F0, zero-frequency filtering, strength of excitation

1. Introduction

Basic emotions such as *anger, happy, fear, neutral,* and *sad*, are explored in this study. A 3-dimensional Valence-Arousal Dominance (VAD) dimensional space of emotions can be used to distinguish these basic emotions [1]. *Valence* represents pleasantness or unpleasantness of the emotion, *arousal* represents the degree of excitement (active/calm), and *dominance* represents the degree of power over the emotion. *Anger* and *fear* emotions are in the high-arousal, negative-valence plane, whereas *happy* emotion is in the high-arousal, positive-valence plane. *Sad* emotion lies on the low-arousal, negative-valence plane. *Neutral* speech lies at the centre of this 3-dimensional space [1, 2]. In this paper, the differences between *high-arousal* and *low-arousal emotions* are examined using acoustic analysis.

Analysis and classification of emotional speech can be carried out by observing the spectrograms [3, 4, 5]. Spectrograms prove to be a good method for analysis of speech signal since these focus upon specific regions of interests, instead of the entire speech signal [6, 7]. Emotion classification between German and Thai database was carried out in [6]. Analysis of emotional speech analysis was carried out using TIMIT database [7]. Classification into high-arousal (*anger, happy*) and low-arousal (*neutral, sad*) emotions was attempted using subsegmental features in [8, 9, 10]. Identification and recognition of emotions were also explored in [8, 9, 10, 12, 13], where neutral emotional speech was used as reference. Paralinguistic sounds such as shout (that may correspond to *anger*) and laughter (that may correspond to *happy*), and few basic emotions were extensively studied in [12, 13, 14, 15, 16, 17, 18]. Nonverbal speech sounds and expressive voices (e.g., *Noh* voice) have also been studied [19, 20, 21]. The effect of source-system interaction that occurs significantly during production of emotional speech and paralinguistic sounds was studied for few consonants in [22].

Selection of appropriate features is very important for classifying emotions. Several studies on emotions have focused upon the excitation source characteristics by examining changes in the vocal fold vibrations measured through F0 [23, 24]. The F0 was computed using the ensemble method in [23], and using the ZFF method in [24]. The F0 estimation using ZFF method involves efficient detection of epochs, i.e., instants of significant excitation [25, 26]. A modified version of the ZFF, called *modZFF*, is used recently for examining the source characteristics of emotional speech [15, 16, 17], paralinguistic sounds [18], and expressive voices [19, 20, 21]. Combined effects of the source and vocal tract filter (i.e., system) have been examined in [10, 14, 16], through the feature strength of excitation (SoE) derived as the slope of the ZFF signal [25, 26]. Signal energy, Mel Frequency Cepstral Coefficients (MFCCs) and F0 were used for emotion classification in [27]. Zero-crossing rate was used for discriminating between the voiced and unvoiced regions of a speech signal and for emotion recognition [27, 28]. Sub-band spectral energy computed for Berlin Emo-DB gave better recognition accuracy than features MFCCs and Log Frequency Power Coefficients (LFPC) in [29, 30].

Emotion classification has been carried out using several features along with machine learning techniques. Different classifiers such as K-Nearest Neighbor, Support Vector Machines, Artificial Neural Network, Hidden Markov Model have been used [29, 30, 31, 32]. Nine different emotional speech database performances were analyzed with the help of HMMs and suprasegmental models in [33].

Production characteristics of emotional speech vary significantly for different emotions, was established in [9, 10, 11, 14] and [15, 16, 18, 20]. In our previous work [34], production features F0 and formants were examined as source and system characteristics, respectively. However, detailed analysis of high-arousal and low-arousal emotions by studying the production characteristics such as ZCR and subband spectral energies is seldom focused in the studies carried out so far.

This study proposes a new approach for discriminating first between high-arousal and low-arousal emotional speech, and

10.21437/SMM.2018-1

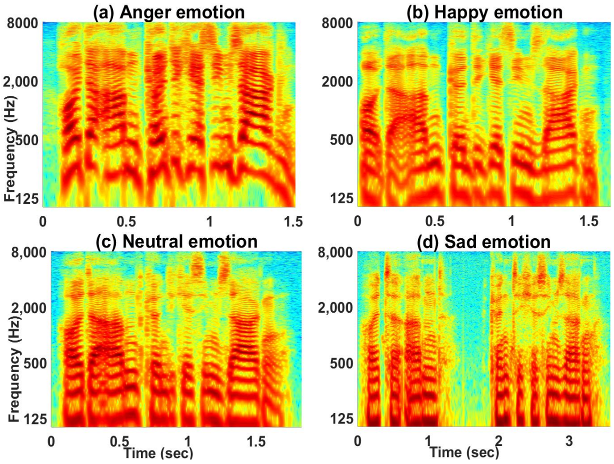

Figure 1: *Spectrogram of German emotional speech for S6 (M) where (a) anger, (b) happy, (c) neutral, and (d) sad.*

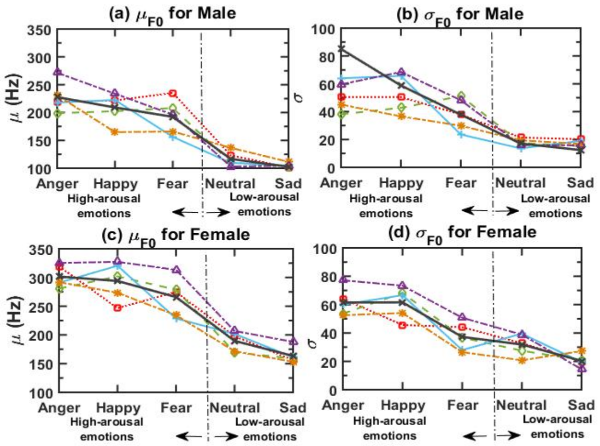

Figure 2: *F0 values for German emotional speech, (a) μ_{F0} and (b) σ_{F0} are for male, and (c) μ_{F0} and (d) σ_{F0} are for female.*

Table 1: *F0 values for Berlin Emo-DB using μ_{F0} and σ_{F0}, where A: anger, H: happy, F: fear, N: neutral, and S: sad.*

	μ_{F0} (Hz)					σ_{F0} (Hz)				
(M/F)	**A**	**H**	**F**	**N**	**S**	**A**	**H**	**F**	**N**	**S**
S1 (M)	219	222	235	123	101	50	50	38	21	20
S4 (M)	199	202	208	111	103	38	43	51	17	15
S5 (M)	218	223	156	110	105	64	66	24	14	19
S6 (M)	231	165	166	137	111	45	37	30	19	17
S9 (M)	272	234	196	103	104	60	68	48	16	16
μ_{F0_M}	*228*	*209*	*192*	*117*	*103*	*85*	*59*	*38*	*17*	*12*
S2 (F)	320	247	275	198	156	64	46	44	33	20
S3 (F)	281	302	279	170	159	54	68	37	27	21
S7 (F)	291	320	229	202	161	60	67	28	40	19
S8 (F)	292	273	234	172	153	53	54	26	21	27
S10(F)	325	327	313	207	188	77	73	51	39	14
μ_{F0_F}	*302*	*294*	*266*	*190*	*163*	*61*	*62*	*37*	*32*	*20*

Table 2: *F0 values for IITKGP-SESC using μ_{F0} and σ_{F0}, where A: anger, H: happy, F: fear, N: neutral, and S: sad.*

	μ_{F0} (Hz)					σ_{F0} (Hz)				
(M/F)	**A**	**H**	**F**	**N**	**S**	**A**	**H**	**F**	**N**	**S**
S4 (M)	234	244	198	199	183	128	73	99	80	53
S6 (M)	207	222	201	216	161	71	35	52	48	50
S8 (M)	213	161	200	189	196	68	85	35	70	39
S9 (M)	174	145	182	152	144	35	28	28	29	31
S10(M)	207	193	195	189	171	30	20	17	32	23
μ_{F0_M}	*215*	*195*	*202*	*187*	*178*	*66*	*48*	*46*	*52*	*39*
S1 (F)	403	347	305	319	331	122	91	50	73	72
S2 (F)	292	302	217	260	283	35	35	35	31	34
S3 (F)	377	319	397	314	308	69	70	30	61	46
S5 (F)	298	237	322	247	282	40	56	35	32	31
S7 (F)	286	294	230	285	237	48	38	27	42	25
μ_{F0_F}	*331*	*300*	*294*	*285*	*288*	*63*	*58*	*36*	*48*	*41*

then discriminating amongst emotions within each category. The 5 basic emotions: *anger, happy, fear, neutral,* and *sad*, are studied. First, changes in the source characteristics are examined through F0 derived using the ZFF method. Then, combined effect of the source and system characteristics are analysed using the strength of excitation (derived using the ZFF method) and signal energy features. After discriminating between the high-arousal and low-arousal emotional speech, the emotions within each category are distinguished using the proposed features ZCR and sub-band spectral energies. Spectrograms of the emotional speech signals are used for validation. Effectiveness of the features is validated on two databases of emotional speech, German Emo-DB [35] and Telugu IITKGP-SESC [36].

This paper is organized as follows. Databases used in this study are discussed in Section 2. In Section 3, the features and signal processing methods used are discussed. Analysis by examining the changes in the production characteristics is presented in Section 4. Analysis by examining ZCR and sub-band spectral energies features is carried out in Section 5. Section 6 consists of the summary, along with the scope of future work.

2. Databases

The Berlin EMO-DB database of emotional speech in German language [35] and the IITKGP-SESC database of emotional speech in Telugu (Indian) language [36] are used in this study. The Berlin EMO-DB consists of approximately 800 sentences for 7 emotions: *anger, happiness, fear, neutral, disgust, boredom, sadness*. Actors (5 male, 5 female) had recorded simulated emotional speech for 10 sentences by each speaker. For this study, 4 sentences of each of 10 speakers for the 5 basic emotions: *anger, happy, fear, neutral,* and *sad* are considered.

The IITKGP-SESC database consists of 12000 utterances of emotional speech, recorded by professional All India Radio (AIR) artists. Utterances of 10 artists (5 male, 5 female) were recorded for 15 sentences, for 8 basic emotions in each of 10 sessions. Thus, the corpus consists of total 12000 utterances (10 artists $\times$ 8 emotions $\times$ 15 sentences $\times$ 10 sessions), for 8 emotions: *anger, happiness, fear, compassion, disgust, neutral, sarcastic* and *surprise*. For this study 4 utterances by each Telugu speaker, for 5 emotions are considered.

Thus, total 400 sentences (5 emotions $\times$ 4 utterances $\times$ 10 speakers $\times$ 2 databases) are examined.

3. Features and Signal Processing Methods

Production features used in this study are as follows:

(a) Fundamental Frequency (F0): Since the excitation source characteristics are expected to change significantly during production of emotional speech [15, 16, 18, 19], changes

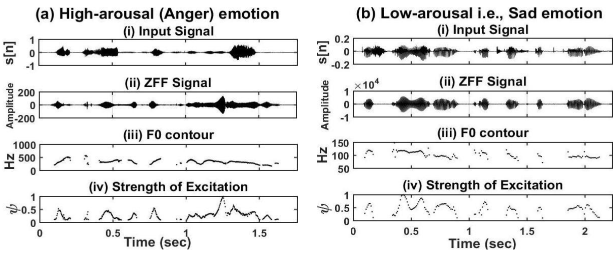

Figure 3: *Illustration of SoE contours for (a) high-arousal (anger) emotion for S2(F), and (b) low-arousal (sad) emotion for S9(M) in German emotional speech database.*

Table 3: *Average SoE (ψ_{ave}) and signal energy (E_{ave} × 1000) for German emotional speech, where A: anger, H: happy, F: fear, N: neutral, and S: sad.*

	(I) ψ_{ave}					*(II)* E_{ave} *× 1000*				
(M/F)	**A**	**H**	**F**	**N**	**S**	**A**	**H**	**F**	**N**	**S**
S1(M)	.32	.36	.44	.50	.53	6.0	9.6	5.8	0.6	6.2
S4(M)	.34	.44	.40	.49	.46	4.1	2.3	2.6	1.2	2.6
S5(M)	.30	.32	.43	.50	.54	7.9	6.4	0.7	1.0	0.3
S6(M)	0.2	.35	.28	.44	.49	.57	7.9	5.5	0.9	1.2
S9(M)	.24	.38	.42	.47	.47	10.6	3.2	5.6	0.9	1.1
Ave(M)	*.31*	*.36*	*.42*	*.49*	*.51*	*7.3*	*5.4*	*3.1*	*2.4*	*0.7*
S2(F)	.27	.43	.39	.44	.46	6.8	2.2	3.3	2.8	0.3
S3(F)	.29	.23	.45	.49	.46	5.7	5.1	3.4	0.8	0.1
S7(F)	.28	.37	.45	.43	.41	2.6	9.0	1.7	0.7	0.3
S8(F)	.28	.40	.44	.52	.43	6.5	3.0	1.9	0.9	0.5
S10(F)	.24	.22	.37	.44	.44	2.4	3.9	4.1	1.2	0.4
Ave(F)	*.27*	*.33*	*.42*	*.46*	*.44*	*4.8*	*4.6*	*2.9*	*1.3*	*0.3*
Ave	*.29*	*.34*	*.42*	*.48*	*.48*	*6.0*	*5.0*	*3.0*	*1.8*	*0.5*

in the frequency of vibration of vocal folds, i.e., F0 are examined for each emotion. The F0 is derived using the ZFF method [25, 26]. Illustration of changes in the F0 contours for *anger* and *sad* emotions is shown in subplots (iii) in Fig. 3 (a) and 3 (b), respectively. Parameters average F0 (μ_F) and standard deviation (σ_{F0}) are computed for this study.

(b) Strength of Excitation: SoE is obtained from the slope of the ZFF signal at each epoch [25, 26]. It represents the strength of impulse-like excitations at the GCIs. Subplots (iv) in Fig. 3 (a) and Fig. 3 (b) show the SoE contour for *anger* and *sad* emotion, respectively, for German emotional speech.

(c) Signal energy: The energy of a signal x[n] is obtained as $E_x = \sum_{n=-w}^{n=+w} x^2[n]$, for signal samples taken in a window of fixed size (from -w to +w).

(d) Zero Crossing Rate: Zero crossing rate is the number of times the audio waveform crosses the zero axis. The voiced region in a speech signal has low ZCR as opposed to unvoiced region where the ZCR signal is always higher [27].

(e) Sub-band Energies: Subband spectral energies are computed using short-time Fourier transform (STFT), with sampling frequency 16 KHz, and Hanning window with N=1024 and overlap of $\frac{N}{2}$ samples. The sub-band spectral energies are then computed for the specific ranges of frequencies, details of which are mentioned in Section 5. The sub-band frequency ranges are chosen empirically, by observing the spectrograms.

4. Analysis using Production Features

Spectrogram analysis of the emotional speech was carried out using Hanning window of size N (4096 samples), with a shift

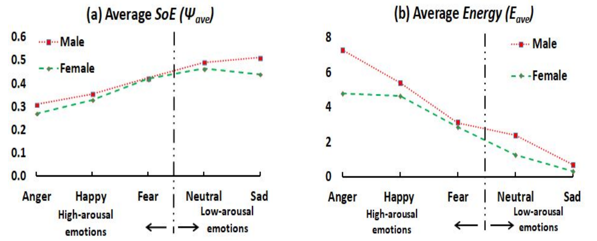

Figure 4: *Average (a) SoE (ψ_{ave}) and (b) energy (E_{ave}) for male and female speakers, for German emotional speech database.*

Table 4: *Mean values for Zero-Crossing Rate (z_c) values for German emotional speech, where z_c(M) and z_c(F) are the average ZCR values for male and female, respectively [μ_{zc} × 1000].*

Speaker	**anger**	**happy**	**fear**	**neutral**	**sad**
z_c(M)	162.4	128.8	144.6	93.7	109.9
z_c(F)	157.0	133.7	149.1	91.1	113.4

of a sample. Spectrograms are denser in the high-arousal emotions (*anger* and *happy*), in frequencies below 2000 Hz. Spectrograms also reveal significant differences in the source characteristics for emotions as can be observed in the low frequency regions around 100 - 500 Hz, in Fig. 1 (a), (b), (c), and (d).

Changes in the excitation source characteristics for each emotion are examined by observing changes in the F0 contour. Patterns of decreasing average F0 from *anger* to *sad* emotions can be observed in Table 1 and Fig. 2, for German emotional speech. Fluctuations i.e., changes in F0 also show similar trend, in Fig. 2. Validation of the changes in the excitation source characteristics is carried out using Telugu emotional speech (Table 2), which also shows similar patterns. Significant differences are observed between high-arousal emotions and low-arousal emotions in Table 1, Table 2 and Fig. 2.

Differences observed from the spectrograms between high-arousal and low-arousal emotional speech are also examined by observing changes in the combined source-system characteristics using SoE and signal energy. The average SoE (ψ_{ave}) for each speaker in column section I in Table 3, shows the trend opposite to that of signal energy. The average SoE is lower for high-arousal emotional speech and gradually increases across emotions for low-arousal emotional speech, as shown in Fig. 4 (a). The SoE also shows clear distinction between the high-arousal and low-arousal emotions.

The discrimination between high-arousal and low-arousal emotions categories given by SoE is validated by signal energy feature. Signal energy and SoE show complementary trend. It can be observed from column section II of Table 3 that the average signal energy values (multiplied by 1000) for German database are higher for high-arousal emotions, whereas SoE values are lower for the corresponding emotions in column section I of Table 3. It can be observed from Fig. 4 (b) that the average signal energy of *anger* emotion is distinctly higher than that of *sad* emotion. It also shows clear discrimination between high-arousal and low-arousal emotions.

5. Analysis by ZCR and Subband Energies

Analysis of the combined effect of source-system characteristics for distinguishing between high-arousal emotional speech and low-arousal emotional speech, and discrimination amongst

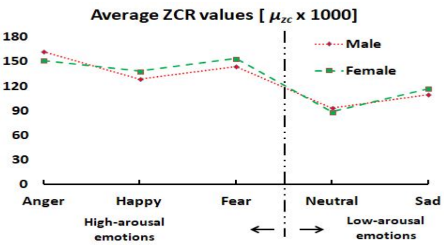

Figure 5: *Average ZCR values for German emotional speech.*

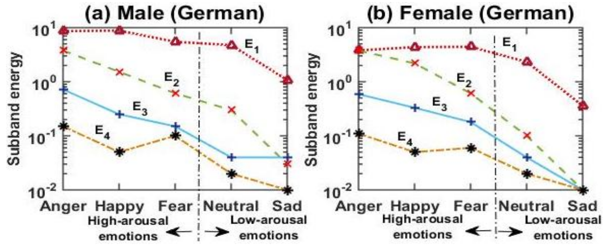

Figure 6: *Subband energies for German emotional speech.*

Table 5: *Average subband spectral energy values of 10 speakers for Berlin EMO-DB, where β_1, β_2, and β_3 are the ratios of the subband spectral energies, i.e., SERs (E_2, E_3, E_4 w.r.t. E_1).*

M/F	SER	anger	happy	fear	neutral	sad
M	β_1	436.79	173.15	111.76	62.93	27.02
	β_2	83.90	28.69	26.93	9.09	34.90
	β_3	17.72	5.69	17.47	3.83	10.40
F	β_1	992.50	501.08	138.28	45.11	19.27
	β_2	156.06	74.82	41.63	15.43	30.26
	β_3	28.91	11.41	13.13	7.02	22.92

Table 6: *Average subband spectral energy values of 10 speakers for IITKG-SESC Database, where β_1, β_2, and β_3 are the ratios of the subband energies, i.e., SERs (E_2, E_3, E_4 w.r.t. E_1).*

M/F	SER	anger	happy	fear	neutral	sad
M	β_1	52.51	32.08	21.73	32.08	19.14
	β_2	12.82	8.51	6.57	8.51	6.76
	β_3	2.07	2.72	2.10	2.72	1.86
F	β_1	135.64	63.27	25.83	63.27	27.41
	β_2	40.65	17.80	5.05	17.80	6.44
	β_3	5.68	1.98	2.12	1.98	1.86

emotions within each of these two categories is further carried out using the features ZCR and subband spectral energies. Table 4 and Fig. 5 show the average ZCR values of emotional speech in German language for five basic emotions. All values are multiplied by 1000 for computational convenience and ease of comparision. *Anger* emotion shows the highest average ZCR (z_c). The z_c is significantly higher for high-arousal emotional speech w.r.t. *neutral* speech. However, in the case of *sad* emotion, longer speech duration and slower speaking rate perhaps cause higher zero-crossing rate as compared to *neutral* speech.

The observations made from spectrogram analysis led authors towards the selection of four energy subbands for better classification of emotions within each category: high-arousal emotional speech (*anger*, *happy* and *fear*) and low-arousal emotional speech (*neutral* and *sad*). Frequency bands chosen for computing the subband spectral energies are: 1 – 1000 Hz (for E_1), 1001 – 2750 Hz (for E_2), 2751 – 4500 Hz (for E_3), and 4501 – 7000 Hz (for E_4). Average subband spectral energies are computed for each speaker. Then ratios of the subband spectral energies (β) are computed, which better reflect the changes in the subband spectral energies for different emotions.

The subband spectral energy ratios (β) are computed as: $\beta_1 = \frac{E_2}{E_1} \times 1000$, $\beta_2 = \frac{E_3}{E_1} \times 1000$, and $\beta_3 = \frac{E_4}{E_1} \times 1000$. Table 5 and Table 6 give the subband spectral energy ratios (β) for German emotional speech and Telugu emotional speech, respectively. Average subband spectral energy values are also illustrated in Fig. 6 (a) for male and in Fig. 6 (b) for female German speakers.

Subband spectral energies can help in discriminating between *anger* and *happy* emotional speech, since the subband spectral energy ratio β_1 is higher for *happy* emotion for both male and female German speakers. The average subband spectral energy values in Table 5 and Fig. 6 show clear distinction between high-arousal and low-arousal emotions. These observations can be validated from the corresponding spectrograms for *anger* and *happy* emotions in Fig. 1, in which the energy components in lower frequency bands (approx. < 2500 Hz) show significant differences in energy and intensity. The ratios of subband spectral energies i.e., β_1, β_2, and β_3, discriminate well amongst emotions within each category. The β values are larger for *anger* than *happy* emotions, and for *neutral* than *sad* emotions, for both male as well as female speakers.

In Telugu emotional speech, the valence, arousal, dominance, and thus expressivity of emotions appear to be relatively less than for German emotional speech. That is why the values of production features (μ_{F0}, σ_{F0}, E, SoE, z_c) and the parameters (β_1, β_2, and β_3) are larger for German data. Inferences can be drawn that the features F0, SoE, E, and z_c are helpful in discriminating between high-arousal and low-arousal emotions. Discriminating amongst the emotions in the same category, e.g., *anger* and *happy* (both high-arousal emotions) which is usually very challenging, can be carried out using the features z_c (ZCR) and β_1, β_2, β_3 (i.e., ratios of subband spectral energies).

6. Summary

In this study, changes in the excitation source feature F0 derived using the ZFF method is analyzed, along with source-system combined features the strength of excitation, signal energy, zero crossing rate and ratios of subband spectral energies (β_1, β_2, β_3). Spectrograms are used as ground truth. Discrimination of high-arousal emotions (*anger, happy* and *fear*) from low-arousal emotions (*neutral* and *sad*) can be achieved better using the F0, since production characteristics of emotional speech differ mainly in the excitation source features. Source-system combined features (SoE and E) prove useful in discriminating between these two categories of emotions. The signal energy and SoE exhibit complementary trends. The average values of SoE, E and ZCR for male speakers are higher than that for female speakers. F0 values are higher for female speakers.

The average subband energies show decreasing trend across emotions from *anger* emotion to *sad* emotion. The observations from spectrograms indicate that the energy and intensity are higher in case of *anger* and *happy* emotions, and can further be discriminated from each other by using ZCR and subband energies features. *Fear* emotion lies between the high-arousal and low-arousal emotions, as can be observed from the explored features. Further analysis and classification will be performed for better results and comparison to state-of-the-art systems as an extension to this study.

7. References

[1] L. C. Yu, L. H. Lee, S. H. Hao, J. Wang, Y. He, J. Hu, K. R. Lai, and X. Zhang, "Building Chinese affective resources in Valence-Arousal Dimension," in *Proc. North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, pp. 540–545, Jun. 12–17, 2016.

[2] S. Parthasarathy, C. Zhang, J. H. L. Hansen, C. Busso, "A study on speaker verification performance with expressive speech," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 5540–5545, Mar. 5, 2017.

[3] A. V. Oppenheim, "Speech spectrograms using the fast Fourier transform," *Spectrum*, IEEE, vol. 7, no. 8, pp. 57–62, 1970.

[4] K. C. Wang, "Time-frequency feature representation using multi-resolution texture analysis and acoustic activity detector for real-life speech emotion recognition," *Sensors*, Multidisciplinary Digital Publishing Institute, vol. 15, no. 1, pp. 1458–1478, 2015.

[5] K. C. Wang, "Speech emotional classification using texture image information features," *International Journal of Signal Processing Systems*, vol. 3, no. 1, pp. 1–7, Jun., 2015.

[6] S. Prasomphan, "Detecting human emotion via speech recognition by using speech spectrogram," in *International Conference on Data Science and Advanced Analytics*, IEEE, pp. 1–10, 2015.

[7] J. Pribil, A. Pribilova, "An experiment with evaluation of emotional speech conversion by spectrograms," *Measurement Science Review*, vol. 10, no. 3, May, 2017.

[8] C. Busso, S. Lee, and S. S. Narayanan, "Using neutral speech models for emotional speech analysis," in *INTERSPEECH – 8^{th} Annual Conference of the International Speech Communication Association*, Belgium, pp. 2225–2228, Aug. 27–31, 2007.

[9] P. Gangamohan, S. R. Kadiri, and B. Yegnanarayana, "Analysis of emotional speech - A review," in *Toward Robotic Socially Believable Behaving Systems*, Springer, vol. 1, pp. 205–238, 2016.

[10] P. Gangamohan, S. R. Kadiri, and B. Yegnanarayana, "Analysis of emotional speech at subsegmental level," in *INTERSPEECH – 14^{th} Annual Conference of the International Speech Communication Association*, France, pp. 1916–1920, Aug. 25–29, 2013.

[11] P. Gangamohan, S. R. Kadiri, S. V. Gangashetty, and B. Yegnanarayana, "Excitation source features for discrimination of anger and happy emotions," in *INTERSPEECH – 15^{th} Annual Conference of the International Speech Communication Association*, Singapore, pp. 1253–1257, Sep. 14–18, 2014.

[12] G. Reddy and K. S. Rao, "Neutral to happy emotion conversion by blending prosody and laughter," in *8^{th} International Conference on Contemporary Computing (IC3)*, IEEE, pp. 342–347, 2015.

[13] S. Prasomphan, "Improvement of speech emotion recognition with neural network classifier by using speech spectrogram," *International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, pp. 73–76, 2015.

[14] S. H. Dumpala, K. V. Sridaran, S. V. Gangashetty, and B. Yegnanarayana, "Analysis of laughter and speech-laugh signals using excitation source information," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 975–979, 2014.

[15] V. K. Mittal and B. Yegnanarayana, "Production features for detection of shouted speech," in *Consumer Communications and Networking Conference (CCNC)*, IEEE, USA, pp. 106–111, 2013.

[16] V. K. Mittal and B. Yegnanarayana, "Effect of glottal dynamics in the production of shouted speech," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 3050–3061, 2013.

[17] V. K. Mittal and B. Yegnanarayana, "An automatic shout detection system using speech production features," in *International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*, Springer, pp. 88–98, 2014.

[18] V. K. Mittal and B. Yegnanarayana, "Analysis of production characteristics of laughter," *Computer Science & Language*, Elsevier, vol. 30, no. 1, pp. 99–155, 2015.

[19] V. K. Mittal, B. Yegnanarayana, "An impulse sequence representation of the excitation source characteristics of nonverbal speech sounds," in *Proc. SLPAT – Workshop on Speech and Language Processing for Assistive Technologies*, San Francisco, USA, pp. 69–74, Sep. 13, 2016.

[20] V. K. Mittal and B. Yegnanarayana, "Study of characteristics of aperiodicity in Noh voices," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. 3411–21, Jun., 2015.

[21] V. K. Mittal and B. Yegnanarayana, "Significance of aperiodicity in the pitch perception of expressive voices," in *INTERSPEECH – 15^{th} Annual Conference of the International Speech Communication Association*, Singapore, pp. 504–5048, Sep. 14–18, 2014.

[22] V. K. Mittal, B. Yegnanarayna, and P. Bhaskararao, "Study of the effects of vocal tract constriction on glottal vibration," *The Journal of the Acoustical Society of America*, vol. 136, no. 4, pp. 1932–1941, 2014.

[23] D. Morrison, R. Wang, and L. C. D. Silva, "Ensemble methods for spoken emotion recognition in call-centres," in *Speech Communication*, Elsevier, vol. 49, no. 2, pp. 98–112, 2007.

[24] P. Gangamohan, V. K. Mittal, and B. Yegnanarayana, "A Flexible Analysis Synthesis Tool (FAST) for studying the characteristic features of emotion in speech," in *Consumer Communications and Networking Conference (CCNC)*, IEEE, pp. 250–254, 2012.

[25] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[26] B. Yegnanarayana and K. S. R. Murty, "Event-based instantaneous fundamental frequency estimation from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 614–624, 2009.

[27] M. Jalil, F. A. Butt, and A. Malik, "Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals," in *International Conf. on Technological Advances in Electrical, Electronics and Computer Engineering*, IEEE, pp. 208–212, 2013.

[28] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," in *Speech Communication*, vol. 48, no. 9, pp. 1162–1181, 2006.

[29] S. Amarakeerthi, C. Morikawa, T. L. Nwe, L. C. D. Silva, and M. Cohen, "Cascaded subband energy-based emotion classification," *IEEJ Transactions on Electronics, Information and Systems*, vol. 133, no. 1, pp. 200–210, 2013.

[30] S. Amarekeerthi, T. L. Nwe, L. C. D. Silva, and M. Cohen, "Emotion classification using inter-and intra-subband energy variation," *INTERSPEECH – 12^{th} Annual Conference of the International Speech Communication Association*, pp. 1569–1572, 2011.

[31] V. Sethu, E. Ambikairajah, and J. Epps, "Speaker dependency of features and speech production cues for automatic emotion classification," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Taiwan, pp. 4693–4796, Apr. 19–24, 2009.

[32] M. E. Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern recognition*, Elsevier, vol. 44, no. 3, pp. 572–587, 2011.

[33] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll, and A. Wendemuth, "Acoustic emotion recognition: A benchmark comparison of performances," *IEEE Workshop on Automatic Speech Recognition & Understanding*, pp. 552–557, 2009.

[34] E. Ramdinmawii, A. Mohanta, and V. K. Mittal, "Emotion recognition from speech," *Region 10 Conference (TENCON)*, IEEE, Penang, Malaysia, pp. 1562–1567, Nov. 5–8, 2017.

[35] F. Burkhardt, A. Paeshke, M. Rolfes, W. Sendlmeier, and Benjamin Weiss, "A database of German emotional speech," in *INTERSPEECH – 6^{th} Annual Conference of the International Speech Communication Association*, Portugal, Sep. 4–8, 2005.

[36] S. G. Koolagudi, S. Maity, A. K. Vuppala, S. Chakrabarti, and K. S. Rao, "IITKGP-SESC: Speech database for emotion analysis," in *International Conference on Contemporary Computing (IC3)*, Springer, vol. 40, pp. 485–492, 2009.