Development and Analysis of Speech Emotion Corpus Using Prosodic Features for Cross Linguistics

Syed Abbas Ali, Sitwat Zehra, Mohsin Khan and Faisal Wahab

Abstract— In daily life communication, spoken language not only carries linguistic information but also conveys nonlinguistic information such as the speaker's emotions, gender, social status and age, etc. This paper introduces the Emotion-Pak corpus, a multilingual emotional speech database consisting of emotional sentences elicited in provincial languages of Pakistan: Urdu, Sindhi, Balochi, Punjabi and Pashto for observing the speech emotions present in acoustic signals.The proposed database is recorded from people (naive speakers)having different linguistics, ages, gender, education level and cultural backgrounds from different regions of Pakistan and cross linguistics compared with the Berlin database of emotional speech (EMO-DB)(actor speakers)in order to study whether emotions are gender and language dependent or independent using prosodic features. The statistical analysis of Emotion-Pak corpus shows that emotions with short duration and strong intensity (Anger and Happiness) or longer duration and weak intensity (Sadness and Comfort) have similar acoustic features. The subjective listening tests used for evaluating the quality of speech emotions in proposed emotions corpus. The subjective listening test results found quite similar with the results obtained from prosodic analysis of Emotion-Pak speech corpus.

Index Terms— Emotion, Multilingual Emotional Speech Corpus, Prosodic Features.

1 INTRODUCTION

EMOTION is a physio-psycological process initiated by unconscious and / or conscious perception of situation or an object and is related with temperament, mood, personality and motivation. Emotions in human communication play an important role and can be conveyed either by expressing nonverbal clues or verbally through vocabulary of emotions such as facial expressions and gestures. Emotions carried in speech reveal the speaker's psychological and physiological states. In daily life conversation, people may often understand the word spoken by speaker but unfortunately misinterpret the emotion in his / her conversation and vice versa. Searching among the sixty four speech resources reviewed in [1] for a multilingual emotional speech database, it is surprising to find out that such resources are very rare, especially when dealing with European languages. According to it, there are very few resources containing speech material in two languages (English and German in [2, 3], English and Slovenian [4], English and Spanish in [5]), and no database containing material in languages of Pakistan. Due to this lack of consideration in multilingual emotional speech resources, Emotion-Pak is realized. The Berlin database of emotional speech (EMO-DB) [6] has been used to perform a cross linguistic analysis of German language with languages of Pakistan and to observe the difference in emotional speech uttered by naive and actor speakers. The key objective of this paper is to introduce the Emotion-Pak corpus and make use of proposed database to analyse the mean values of prosodic features and subjective listening test to study the speech emotion dependency on gender and linguistics. Paper details are as follow. The consequent section briefly reviews the concept of human speech production model and cognitive process of emotions. In section 3, we discuss the expression of feeling using prosodic elements of speech. Proposed multilingual Emotion-Pak speech corpus is presented in section 4. We present our experimental results and subjective testing evaluation in section 5. The conclusion is drawn in section 6.

2 HUMAN SPEECH AND EMOTION

The human speech production starts with the downward movement of the diaphragm, the muscle which lies between the chest cavity and the abdomen, letting air flow up to the lungs. The vocal tract is generally considered as a set of speech-producing organs above the vocal folds (see figure 4.1), which consists of the following: larynx, oral cavity, nasalpharynx (above the velum, rear end of nasal cavity), and nasal cavity (above the palate and extending from the pharynx to the nostrils) [7]. The production of sound of the human voice, after the movement of the diaphragm [8] shown in figure-1.

- *Syed Abbas Ali, M.Engg in Communication Systems working as Astt. Professor at the Department of Computer and Information Systems Engineering, NED University of Engineering & Technology, Karachi, Pakistan.*
  E-Mail: saaj@neduet.edu.pk
- *Sitwat Zehra, Ph.D. in Physiology working as Astt. Professor at the Karachi Institute of Biotechnology and Genetic Engineering, University of Karachi, Pakistan.*
  E-mail: sitwat.zehra@kibge.edu.pk
- *Mohsin Khan, B.E. in Telecomunication Engineering working as Consultant Engineering, Karachi, Pakistan.*
  E-mail: khan.mohsin@engineer.com
- *Faisal Wahab, B.E. in Telecomunication Engineering working as Consultant Engineering, Karachi, Pakistan.*
  E-mail: faisalwahab@engineer.com

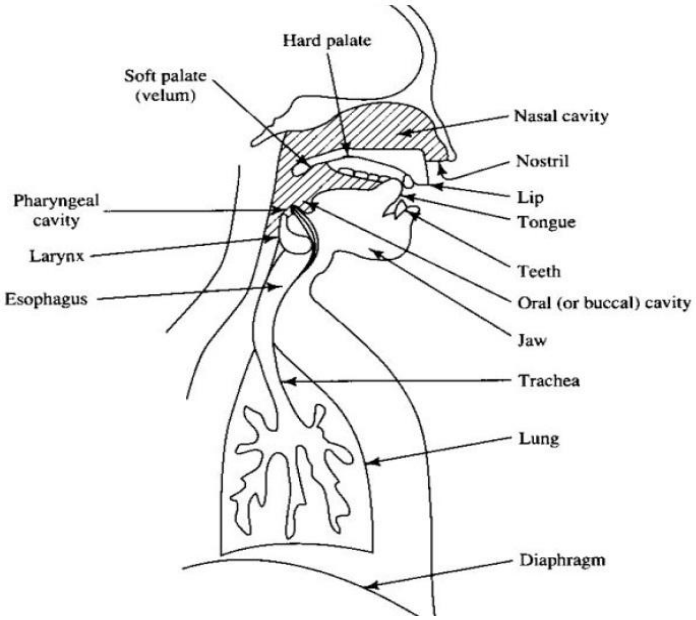
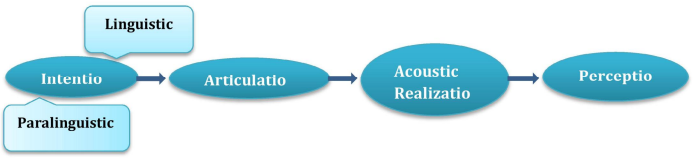


Figure-1: The human speech production [9]

- Air pressure from the lungs creates a steady flow of air through the trachea (windpipe), larynx (voice box) and pharynx (back of the throat).
- The vocal folds in the larynx vibrate, creating fluctuations in air pressure that are known as sound waves.
- Resonances in the vocal tract modify these waves according to the position and shape of the lips, jaw, tongue, soft palate, and other speech organs, creating formant regions and thus different qualities of sonorant (voiced) sound.
- Mouth and nose openings radiate the sound waves into the environment.

According to the classic definition, prosody has to do with speech features whose domain is not a single phonetic segment, but larger units of more than one segment, possibly whole sentences or even longer utterances. Consequently, prosodic phenomena are often called supra-segmental; it uses to structure the speech flow and is supposed as stress, or number of modification of intonation, rhythm and loudness.

Figure-2: Stages of oral communication



Human emotions are associated with cognition; it stimulates cognitive processes which help in building strategies. Emotions don't need any especial language it can easily be universally spoken and understood.

2.1 Cultural

The qualities of emotional expression can be determined through cultural characteristic. Emotional qualities and expressions cannot be determine through hormones, neurotransmitters and autonomic relations these biological them[10], basically emotions are interdependent and interpenetrating with other cultural phenomena according to activity theory.

2.1 Behavioral

Behaviorism includes acting, thinking and feeling of an organism [11]. In the view of Behavioral theorist regard emotion is a natural phenomenon of reinforcing stimuli and classical conditioning.

3 Prosody Features

In linguistics, prosody is the stress, rhythm and intonation of speech. Different prosody features of the utterance or speaker are as follow: the emotional state of the speaker; the type of the utterance (command, statement, or question); the presence of sarcasm or cynicism; dissimilarity, stress and motivation; or other features of language that may not be determined by choice of vocabulary or grammar. Prosody has been studied as vital source of knowledge in emotional speech research community for speech understanding and analyzing the emotional expressions present in acoustic speech [12]. Linguistically relevant prosodic measures correspond to structure of sentence which highlights linguistic components by suggesting their function and marking their boundaries. Different types of prosodic units have been proposed such as sentences, paragraphs, intermediate groups, intonation groups, stress groups etc. The family of prosodic occurrences includes the features of stress, intonation, speech rate and rhythm which reflect the function of the different prosodic components: a decreasing intonation contour mark the conclusion of a sentence, stress mark the prominent syllable in the word, minor intonation and a earlier speech rate characterize a parenthetical phrase. All of these prosodic features are physically apprehended in the speech in terms of set of acoustic parameters variation.

Prosody is the combination of duration of speech segment, energy variation during speech and pitch of speech. Prosody features deliver an added sense to the spoken words which has immense significance in natural speech to provide speaker emotion such as sadness, anger, happiness and comfort. Prosody is not only helpful in finding physical state of the speaker such as speaker is an energetic person or a lazy one, but also has an importance in speech synthesis. Prosody is essential in order to attain an acceptable naturalness. Four acoustic signal prosodic features such as duration of speech segment, intensity, pitch and formants are used in this experiment to parameterize the speech signal.

- Duration: One of main prosody parameter is the duration of speech segments. The timing structure of a sentence has immense significance to provide naturalness to speech. The duration of Phoneme influ-

enced by following parameters such as surrounding phones, phonetic identity and level of stress or position in the word or in the sentence. Word duration also have massive importance in the sentence.

- Intensity: Intensity refers to the energy of the speech signal. While intensity is generally assumed to encode prosodic information, few studies on the use of intensity in speech prosody exist [13]. Physiologically, intensity variations partially correlate with fundamental frequency variations in case of in to national prominences [14, 15, 16]. Nevertheless, recent studies pointed out the role of intensity variations into prosodic organization [17], and information focus [18] independently of fundamental frequency variations. Several measurements of the speech intensity exist, from the conventional short-term intensity measure, to the long-term integration over prosodic units.
- Pitch: Pitch is an important prosody parameter having perceptual property that makes use of frequency-related scale to ordering of sounds. Pitch may be measured as a frequency, but pitch is not a purely objective physical property. Pitch is a subjective psycho acoustical attribute of sound [19].
- Formants: Formant refers as an acoustic resonance of the human vocal tract. Formants are the unique or significant frequency components of human singing and of speech. Formants are quantified by the frequency content of the vowel sounds which contain information that humans require to discriminate between vowels. Most of these formants are produced by chamber resonance and tube but few whistle tones come from periodic collapse of Venturi effect low-pressure zones [20].

4 EMOTION-PAK: A MULTILINGUAL EMOTIONAL SPEECH CORPUS

Emotion-Pak is a first attempt to experimentally collect an emotional speech database in provincial languages of Pakistan (Urdu, Sindhi, Balochi, Punjabi and Pashto). The purpose of this speech corpus development is to study the vocal emotions in a cross-linguistic and cross-cultural study. The objective of the proposed Emotion-Pak is to analyze people's ability to decode vocal emotions in provincial languages of Pakistan. The main idea behind the development of Emotion-Pak speech corpus is to identify elements which are common between nearer cultures than in cultures being poles apart (e.g. American vs. Japanese or Chinese).

4.1. General observations

Research on vocal emotions has followed research on facial expression in [21], underlines "Contrary to research on the face, researchers interested in the voice cannot present a valid 'snapshot' which represents the vocal attributes of an emotion." This difficulty has been challenged in different ways causing emotional speech resources to differ for their characteristics. These include primarily choice and number of encoders (with reference to gender and type: actor vs. naive speakers); recorded speech units (such as vowels or syllables, words, non-sense utterances, interjections, and so on); kind of speech (spontaneous vs. laboratory speech, read or acted); emotion elicitation procedures (photographs, scenarios, verbal labels, etc.); last but not least, the number of emotions considered differs greatly, too. Because of the great differences between emotional speech databases, a comparison among them is often not possible. This issue suggested the development of the Emotion-Pak database.

4.2. Emotion-Pak Corpus Description

The Emotion-Pak database has been recorded bearing in mind to administer a perceptual test to listeners to verify their ability to decode emotions as well as to verify how much of the differences in the recognition accuracy rates were due to language differences. For this reason the Emotion-Pak Corpus is recorded for five languages of Pakistan: Urdu, Sindhi, Balochi, Punjabi and Pashto, considering the emotions defined in [22] as basic emotions (anger, happiness, sadness, comfort) using 10 (5 female and 5 male) naive speakers from different regions of Pakistan. The male speakers are selected from the age group of 23-50 years; two of them were engineering students while rests were professionals. Similarly female speakers are selected from the age group of 24-32 years; three were arts students while rests were professionals.

4.2.1. Corpus Recording Specifications

For the recoding of the emotional speech data, standardized procedure for speech corpora development based on the ITU recommendations has been adopted. The recording has been done in standard recording environment having SNR>=45dB. We use built-in sound recorder of Microsoft Windows 7to record the entire corpus. The recording format is 16 bit, PCM, Mono and sampling rate is 48 KHz, using microphone with impendence of 2.2W, sensitivity of 54dB±2dB, pulp type of 3.5mm stereo and cable length of 1.8m.

4.2.2. Choosing a Carrier Sentence

The choice of a carrier sentence was highly sought-after following some a-priori established desiderata, according to which the sentence had to be:

- Semantically neutral if taken out of a specific context: every sentence spoken during a speech act carries an amount of information that cannot be excluded a-priori. Since the interpretation and emotional value is in the mind of the listener and not in the sentence itself, these mantically neutral refers here to the possibility of inserting the sentence in any given scenario of a specific emotion.
- Consistent with the general construct of any of the situations presented.
- Correct according to the general rules of each language examined avoiding confusion or distraction firstly in the encoders, and subsequently in the listeners.

- Be easy to analyze, meaning that the sentence on set should be easily identified for analysis: sentence on set with unvoiced consonants was for this reason avoided. Following previous studies (such as [23, 24]) the resulting sentence choice was:

"In Seven Hours It Will Happen"

- For German: In siebenStundenwirdessoweitsein.
- For Urdu: یہ سات گھنٹےمینہو گا.
- For Sindhi: ڪلا ڪن ۾ ٿيندو. ستهي
- For Balochi: بئيتہاڪلاڪہفتهي.
- For Punjabi. ہو ئگاوِچگھنٹےستایہ .
- For Pashto:. دابهيه اوو ساعتو کيږي

4.2.3. Database Validation and Stimuli Selection

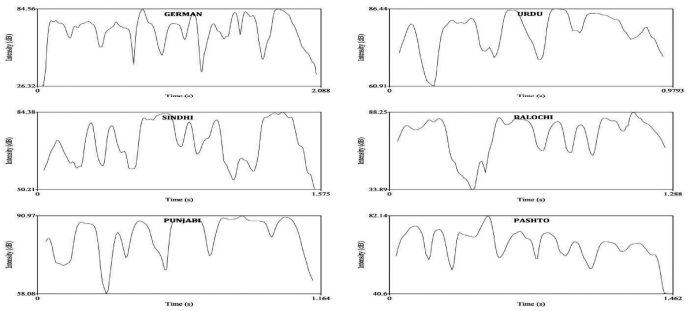
The validation procedure was carried out by means of a subjective listening test with four options available (anger, happiness, sadness, and comfort) using the PRAAT software [25]. Judges were asked to carefully listen to the randomly presented files and to indicate which of the choices available they recognized in the presented files. Judges were not allowed to go back to previously presented stimuli.

Figure-3: Working with PRAAT

5 Experimental Results and Subjective Evaluation

Experiments were done for analyzing the emotions in speech signal using PRAAT software. The pictorial view of the PRAAT is shown in Figure-3. Computer software package (PRAAT) was developed with the aim to analyze speech in phonetics. PRAAT offers features ranging from simple tasks (i.e. recording sounds) to apply statistical computation on sound related data. It offers user guidebook which makes it handy for all kind of users [26]. The proposed Emotion-Pak speech corpus used in the experiments here consists of 40 speech sample taken from the recording of male and female speakers in provincial languages of Pakistan: Urdu, Sindhi, Balochi, Punjabi, Pashto and 8 speech samples of male and female in German language taken from Berlin database of Emotional speech (EMO-DB) having four different emotions: Anger, Happiness, Sadness and Comfort. In this work, we investigate four different emotions in the provincial languages of Pakistan using prosodic features: Duration of speech segment, Intensity, Pitch and Formant frequency to observe the dependency of emotions on gender and cross linguistic. The following observations has been made using PRAAT Software for different prosodic features in four emotions of provincial language (Urdu, Sindhi, Balochi, Punjabi and Pashto) speakers.

In figure 4 and 5, graphical results represent the intensity of



the male and female speakers for Anger and Sadness respectively. The overall observation of intensity analysis for four different emotions can be summarized as:

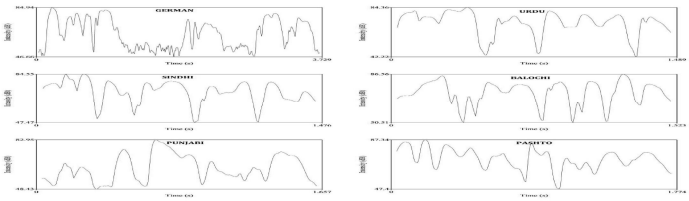


Figure-4: Intensity of male speakers' anger.

Figure-5: Intensity of female speakers' sadness.

- Anger: The mean intensity is high, the pace is fast, a lot of syllables are pronounced, the end word is not emphasized and the intensity outlines of all syllables are decreasing.
- Happiness: The mean intensity is high but not more than anger, the pace is slightly fast, few syllables are pronounced, the last word is emphasized and the intensity outlines of all syllables are increasing.
- Sadness: The mean intensity is low, the pace is slow, very limited syllables are pronounced, the last word is not emphasized and the intensity outlines of all syllables are decreasing.
- Comfort: The value of the mean intensity is higher but not more than happiness, the overall pace is slow and very few syllables are pronounced whereas, the end syllable is emphasized and the intensity outlines of syllables are increasing.

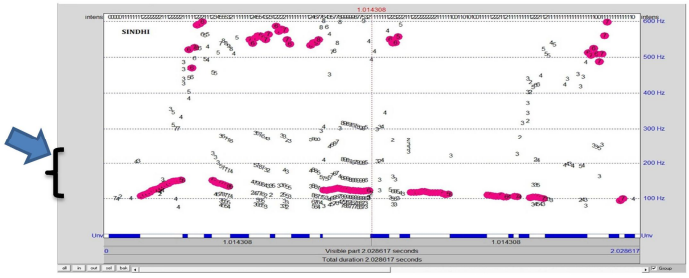


Figure 6 and 7, represent the variation in pitch of Sindhi male and Pashto female speakers for happiness and comfort respectively. The observation can be summarized as:

Figure-6: Pitch of Sindhi male speaker's happiness.

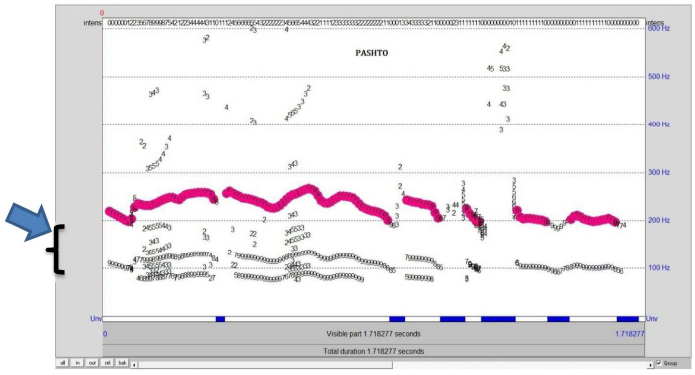


Figure-7: Pitch of Pashto female speaker's comfort.

- Anger: The mean pitch is high, greater number of voiced samples represented by pink circles found in the frequency range of 200-300 Hz for male, 300-400 Hz for female and the pitch outlines of entire syllables are decreasing.
- Happiness: The mean pitch is high, greater number of voiced samples represented by pink circles found in the frequency range of 100-200 Hz for male, 200-300 Hz for female and the pitch outlines of entire syllables are increasing.
- Sadness: The mean pitch is low, greater number of voiced samples represented by pink circles found in the frequency range of 100-200 Hz for male, 200-300 Hz for female and the pitch outlines of entire syllables are decreasing.
- Comfort: The mean pitch is high but not more than happiness, greater number of voiced samples represented by pink circles found in the frequency range of 100-200 Hz for male, 200-300 Hz for female and the pitch outlines of all syllables are rising.

Figure-8 and Figure-9, show the formant frequencies of male and female speakers for Anger and Comfort respectively. It is observed that the formant frequencies in:

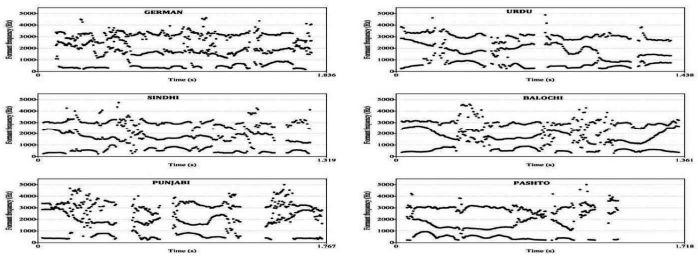


Figure-8: Formants of male speakers' anger.

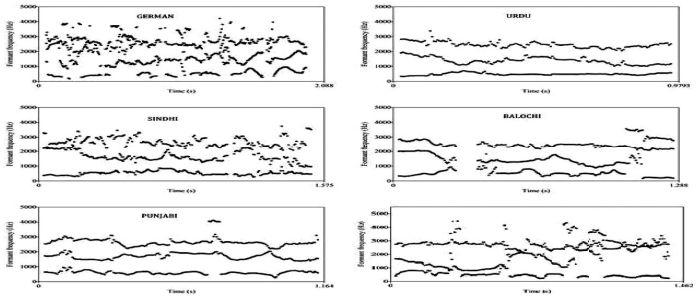


Figure-9: Formants of female speakers' comfort.

- Anger: The mean formant frequency is high with high variance but less than sadness and having slight variance of phoneme durations.
- Happiness: The mean formant frequency is high but less than anger, and has a high variance.
- Sadness: The mean formant frequency is high with low variance and having high variance of phoneme durations.
- Comfort: The mean formant frequency is low and with a high variance of phoneme durations.

Figure-10 to Figure-17, provide the comparative analysis of prosodic features in provincial languages of Pakistan having four different emotions: Anger, Happiness, Sadness and Comfort with German database of emotional speech (EMO-DB). To develop this comparison of speech emotions, we used our proposed Emotion-Pak speech corpus consists of male and female speakers in provincial languages of Pakistan. The figures describe the mean values of intensity, pitch and Formant frequency in term of speech emotion samples.

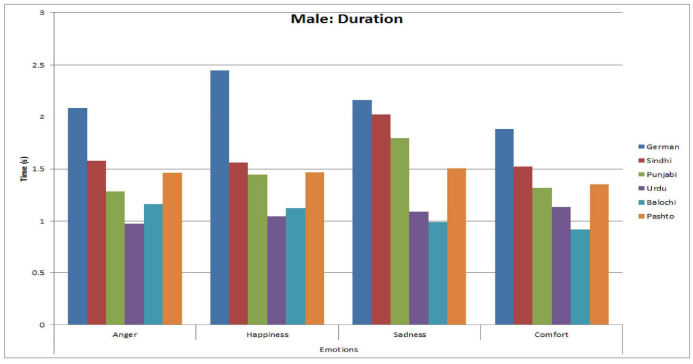


Figure-10: Duration of male speech utterances.

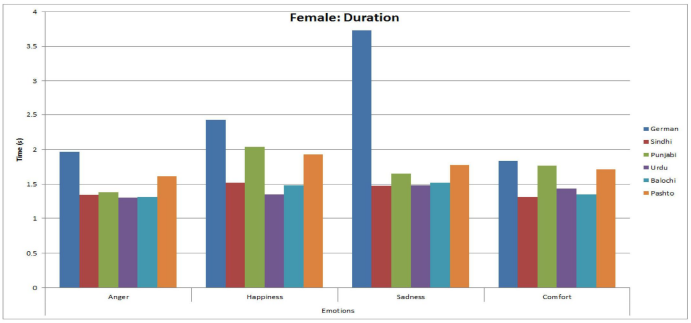


Figure-11: Duration of female speech utterances.

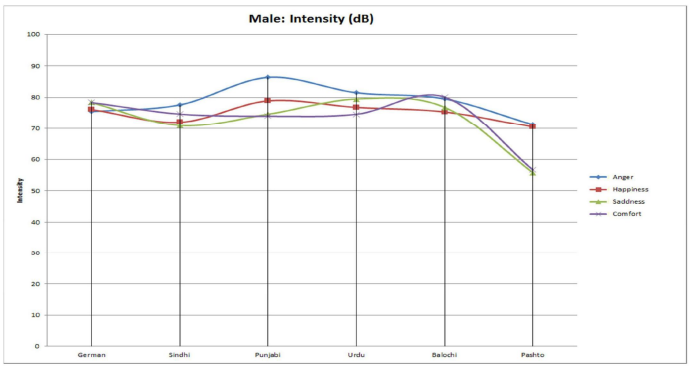


Figure-12: Male mean intensities.

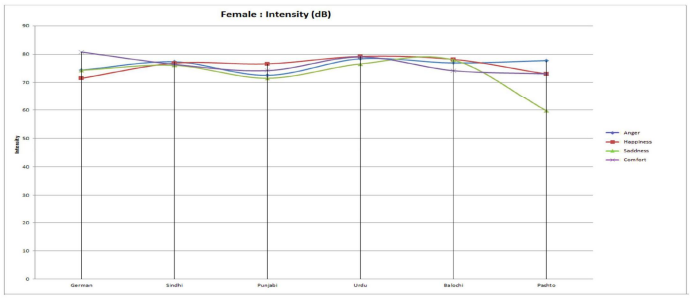


Figure-13: Female mean intensities.

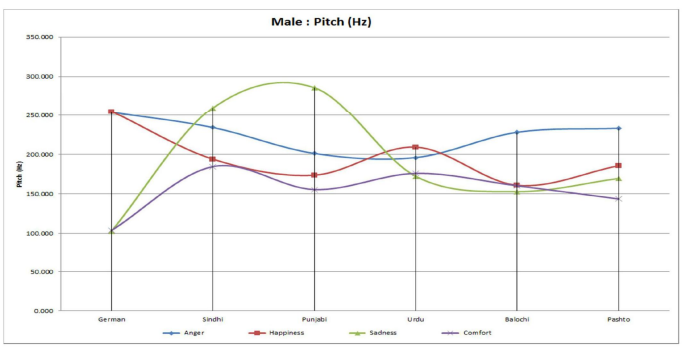


Figure-14: Male mean pitches

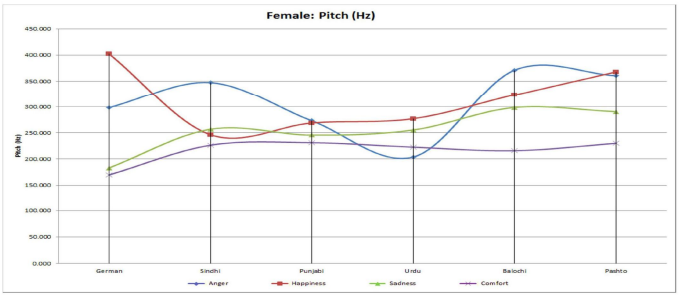


Figure-15: Female mean pitches.

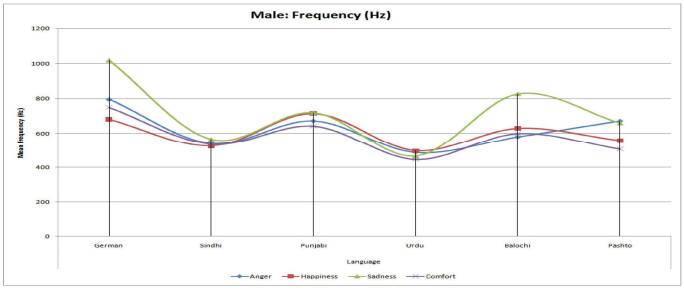


Figure-16: Male mean formants frequencies.

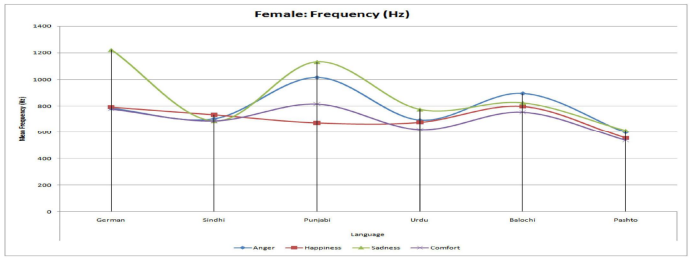


Figure-17: Female mean formants frequencies.

The comparison has been made on the basis of the mean values of speech emotion samples. Our experimental results show that the following observations:

- A significant variation is observed in German sadness. Whereas, slight variations in happiness are observed in German, Punjabi and Pashto.
- Intensity lowers sharply transiting from Balochi to Pashto. There is a normative distribution over the course of length from German to Urdu in comfort. Anger and happiness have similar representations throughout the languages. In female mean intensities only major drop in intensity occurs from Balochi and Pashto.
- Comfort and happiness show similarity from Urdu to Balochi, whereas sadness and happiness show similarity from Balochi to Pashto. The other emotions deviate rapidly from one another throughout the languages. In female mean pitch Sadness and comfort have no significant variations till Urdu. Whereas, happiness and anger have curves exactly opposite in terms of deviations.

- There is a steep increase in formant frequency midway from Urdu to Pashto which shows that data values are directly proportional to the frequencies. Sadness shows highest frequencies throughout the language curve, however it over rides from Sindhi to Punjabi. In female formant frequency analysis, emotion happiness is streamlined throughout the languages. The means of all the frequencies match from Sindhi to Pashto. However, happiness only matches from Urdu to Pashto. There is a clear overlapping of anger and comfort observed in German.

The subjective listening test is used to evaluate the quality of speech emotions expressed in the proposed speech emotion corpus. The quality characterizes how well the speakers pretend the emotions from the spoken sentence. The human subjects are used to assess the realness of the emotions entrenched in speech utterances. Subjective evaluation is carried out by 20 faculty members of the university belong to different discipline of engineering. This subjective evaluation is also useful to identify the discriminative and confusing emotions among four speech emotion classes. The randomly ordered speech emotion from speech corpuses are played to the listeners. For every single emotional speech, the listener has to mark the emotion category from the set of four emotion classes. The overall performance of subjective listening test for selected female and male naïve speakers' using berlin database for emotional speech (EMO-DB) and proposed Emotion –Pak speech corpus shown in Figure-18 and Figure-19. The following observations have been made during subjective listening tests: (1) The average speech emotion recognition rate is about 74%and 76% for female and male speech utterances respectively. (2) Happiness and anger are well recognized whereas; sadness and comfort are relatively less accurately recognized. The prosodic features and performance of subjective listening tests found quite similar, which specify that emotions are independent of gender and cross linguistics.

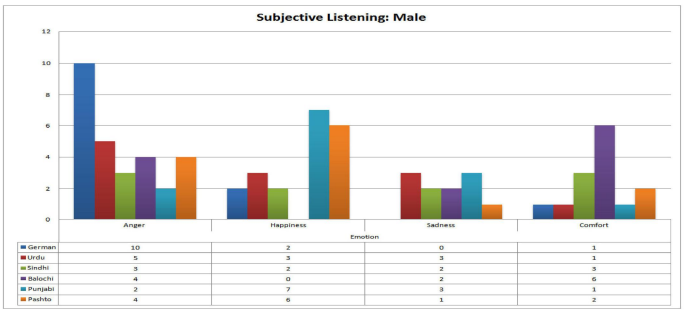


Figure-18: Result of Male Speech Utterances

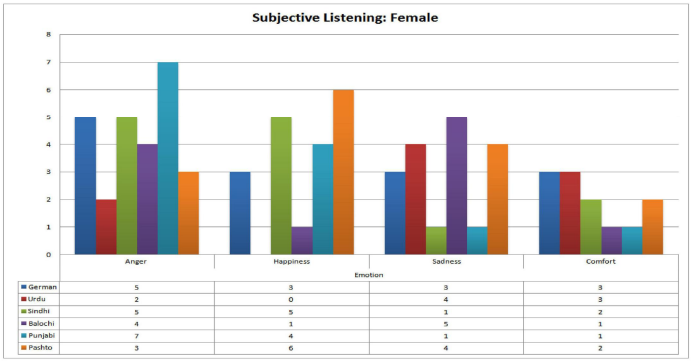


Figure-19: Results of female speech utterances

CONCLUSION

In this paper, we have introduced speech emotion corpus (Emotion-Pak) recorded in provincial languages of Pakistan: Urdu, Sindhi, Balochi, Punjabi and Pushto. The emotions considered for developing Emotion-Pak corpus are anger, happiness, sadness and comfort using prosodic features (Duration, Intensity, Pitch and Formant Frequencies). This paper presented initial study for observing the emotion present in speech signal to study the whether emotions are dependent on gender and cross linguistic. The proposed speech emotion corpus has wide variety of features in terms of languages, ages, gender, educational level and cultural backgrounds of speakers. The experimental results and subjective listening test showed that females had profound emotions than males. Whereas, females are more responsive to the affective prosody, on the other hand, men are better in the linguistic utterances. The statistical analysis of prosodic features and subjective listening tests found quite similar, which indicates that emotions are independent of gender, languages and cultural backgrounds. Our future research focus is to enhance the developed speech emotion corpus, investigate the difference between patient's emotions, the variations in emotional speech utterances of normal and abnormal people and gender recognition using proposed Emotion- Pak speech corpus.

REFERENCES

[1] D. Ververidis and C. Kotropulos, "Emotional speech recognition: Resources, features, and methods," Speech Communication Elsevier, Vol. 48, pp. 1162-1181, 2006.

[2] A. Batliner, C. Hacker, S. Steidl, E. Nöth, S. D'Arcy, M. Russel and M. Wong, "'you stupid tin box' - children interacting with the AIBO robot: a cross-linguistic emotional speech corpus," in Proceedings of the 4th International Conference of Language Resources and Evaluation (LREC '04), pp. 171-174, 2004.

[3] K.R. Scherer, D. Grandjean, L.T. Johnstone, G. Klasmeyer, " Acoustic correlates of task load and stress," in Proceedings of the International Conference on Spoken Language Processing (ICSLP '02), pp.2017-2020, 2002.

[4] D.C. Ambrus, "Collecting and recording of an emotional speech database," Technical Report, Faculty of Electrical Engineering, Institute of Electronics, University of Maribor., 2000.

[5] G.M. Gonzalez, "Bilingual computer-assisted psychological assessment: an innovative approach for screening depression in Chicanos/Latinos," Technical Report TR-0039, University of Michigan., 1999.

[6] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of German emotional speech," in Proceedings of Interspeech '05, Lisbon, Portugal, pp. 1517-1520, 2005.

[7] Rosistem bar code, biometric education. http://www.barcode.ro/tutorials/biometrics/voice.html.

[8] Voice organ-Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Voice organ.

[9] Speech production. http://kom.aau.dk/group/04gr742/pdf/speech_production.pdf

[10] C. Ratner, "A cultural-psychological analysis of emotion (electronic version)," Culture and Psychology, Vol. 6, pp. 5-39, 2000.

[11] Behaviorism-wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Behaviorism.

[12] M. Kuremastsu et al, "An extraction of emotion in human speech using speech synthesize and classifiers for each emotion," WSEAS Transaction on Information Science and Applications, Vol .5(3), pp.246-251, 2008.

[13] P. Delattre, " L'accent final en fran¸cais : accent d'intensit´e, accent de hauteur,accent de dur´ee'," French Review, Vol. 12(2), pp.141–145, 1938
[14] J. Atkinson, "Correlation analysis of the physiological factors controlling fundamental voice frequency," Journal of the Acoustic Society of America, Vol. 63(1), pp.211-222, 1978.
[15] M. Beckmann, (1986). Stress and non-stress accent.Foris, Dordrecht.
[16] N. Campbell, "Prosodic encoding of English speech," In International Conference on Spoken Language Processing, pp. 663-666, Edmonton, Canada, 1992.
[17] C. Tseng, and Y. Lee, "Intensity in relation to prosody organization," In International Symposium on Chinese Spoken Language Processing, pp.217-220, Hong-Kong, China.2004
[18] D. Beaver, B. Zack Clarck, E. Flemming, T.F. Jaeger and M. Wolters, "When semantics meets phonetics: Acoustical studies of second-occurrence focus," Journal of the Linguistic Society of America, Vol. 83(2), pp. 245-276, 2008.
[19] Formant -Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Formant.
[20] Pitch-Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Pitch_(music).
[21] M.D.Pell, L. Monetta, S. Paulmann, and S.A. Kotz, "Recognizing emotions in a foreign language," Journal of Nonverbal Behavior, Vol. 33, pp. 107-120, 2009.
[22] P. Ekman, "An argument for basic emotions,"Cognition and Emotion, Vol. 6, pp. 169-200, 1992.
[23] H.G. Wallbott and K.R. Scherer, "Cues and channels in emotion recognition,"Journal of personality and social psychology, Vol. 51, pp. 690-699, 1986.
[24] L. Anolli, L. Wang, F. Mantovani and A. De Toni, "The Voice of Emotion in Chinese and Italian Young Adults," Journal of Cross-Cultural Psychology, Vol. 39, pp. 565-598, 2008.
[25] Praat: doing phonetics by computer. http://www.fon.hum.uva.nl/praat/.
[26] Advanced analysis speech tools: Praat. http://homepage.ntu.edu.tw/~karchung/Phonetics%20II%20page%20twentythree.htm