Identification of Emotion Stress Agents in Hindi and English Sentences

Varun Gupta
G.G.S.I.P.U.
Delhi, India
varunguptaa@hotmail.com

Anurag Jain
G.G.S.I.P.U.
Delhi, India
Anurag_jain76@yahoo.com

Neeraj Kumar
I.I.I.T.
Allahabad, India
Neerajmaurya2000@gmail.com

***Abstract*—Identification of emotion plays a significant role in human machine interaction. Our research focuses on identifying the emotion stress agent in a spoken sentence in Hindi and English. These agents help us in quickly identifying the emotion in speech. It greatly increases the capability of machine by just analyze the stress agent. We consider the emotions "Happy", "Anger", "Surprise" and "Sad" along with "Neutral" in speech and examine the difference between the agents. At the end we come up with a table showing the comparison between various stress agents.**

Keywords- stress agent, emotion, fundamental frequency, harmonics and sub-harmonics.

I. INTRODUCTION

Speech considered one of the most reliable and further more comfortable modalities to automatically identification of person's emotion [1]. Emotions can be expressed in various ways; therefore there are also multiple methods to detect them. Emotional expressions include three main categories verbal expressions, facial expressions as well as semantic cues. Verbal expressions are usually detected only from human speech and rely mainly on linguistic and acoustic information analysis. Emotive information should be extracted by analyzing the linguistic and semantics of the utterance as well as classifying it to a specific pattern of acoustic correlates [2]. In our approach we use tonal information for emotion identification. Studied proposed that pitch contour (F0) or fundamental frequency is one of the measure foundation of speech identification [3]. In our approach we use F0 as a main focus of our study. Various statistical feature of F0 termed as stress agents and help in identification of emotion. Although various systems proposed by different organization for speech emotion recognisation but still no one check it works fine for Indian English and Hindi. English and Hindi are structurally different from each other and nature of speaking is also different. English is stress-timed language while Hindi is syllable timed language. Average duration of speech in Hindi is more than that of English. One probable reason is average size of words which is more in Hindi. Intensity of speaking Hindi is more than English in a definite emotion. Section II describes the speech corpora used in this research. Section III describes the stress agents used for identification. Section IV describes the comparison of various stress agents. Section V describes the conclusion and future work of after this research.

II. SPEECH CORPORA

We created our own speech corpora due to requirement of same sentences in both language and no freely available database provides this feature in targeted language. Though there are many different categories of emotion, people have the common view that there are six different basic emotions: happiness, sadness, surprise, fear, anger and disgust [4]. In our approach, we used only four primary emotions i.e. happiness, anger, surprise and sadness along with neutral speech style to make comparison. To avoid overstated emotion nonprofessional speaker did recordings.2 native Indian speakers (both male) uttered 10 sentences in both languages in five different emotions, resulted a sum of 200 utterances. These utterances are recorded in quite environment in *.wav format having a sampling rate of 16 kHz with a precision of 16 bits per sample. To make the speech corpora accurate six listeners independently evaluate these utterances. Finally rejected utterances would be re-recorded again and same procedure repeated until database is approximately corrected.

III. FEATURE SELECTION

Speech prosody is an important feature for emotion modeling. In our study we use F0 as a central focus. For calculating the F0 of the speech signal we use "sub harmonics to harmonic pitch determination algorithm" [5]. Basic reason for using this algorithm is its performance is superior to other algorithm [6]. Studies show that the typical Pitch contour range for male is (50,250) and for female is (50, 400). Therefore, for removal of creaky voice we take the F0 range from 50 to 500 Pitch contour calculated by using overlapped frame length of 40ms on every 10 ms short-term time analysis. We are using statistical features for the uniquely identification of emotion. We termed various statistical component of F0 as stress agents. These stress agents are shown in table 1.

TABLE 1. STATISTICAL COMPONEMT OF F0

F0_mean	F0_standard deviation
F0_maximum	F0_mainimum
F0_range	F0_median
F0_kurtosis	

IV. COMAPRISON OF STRESS AGENTS

We compute the statistical feature of F0 both at voiced and sentence speech. Voiced speech is one in which vocal cord

vibrates and unvoiced speech is one in which vocal cord doesn't vibrate. Sentence level speech is a collection of both voiced and unvoiced speech. Ratio of these statistical value calculated using neutral speech values as a reference one. We compute all the ratios and draw the tables 2 to 5 showing the comparison between above mentioned statistical component in Hindi and English both sentence and voiced level.

TABLE 2. RATIO OF VOICED LEVEL COMPARISON OF VARIOUS F0 COMPONENT IN ENGLISH LANGUAGE

	Emotion			
F0 value	Happy	Anger	Surprise	Sadness
Mean	1.02858	1.10466	1.19494	0.95779
Std. deviation	1.18159	1.65974	1.90621	0.97703
Maximum	1.22022	1.52072	1.53184	0.94364
Minimum	1.05261	1.06277	1.23550	1.24816
Range	1.31522	1.76005	1.68944	0.81307
Median	0.98177	1.05646	1.14027	0.94326
Kurtosis	0.79170	1.02988	0.78281	0.68828

TABLE 3. RATIO OF SENTENCE LEVEL COMPARISON OF VARIOUS F0 COMPONENT IN ENGLISH LANGUAGE

	Emotion			
F0 value	Happy	Anger	Surprise	Sadness
Mean	1.29671	1.31131	1.27804	1.00996
Std. deviation	1.39323	1.49292	1.71154	1.01031
Maximum	1.21171	1.26219	1.30084	0.95745
Minimum	0.93842	1.05261	0.98826	0.98711
Range	1.27284	1.31856	1.37333	0.95267
Median	1.39406	1.44166	1.17850	1.02681
Kurtosis	0.66314	0.85641	0.57063	0.80730

TABLE 4. RATIO OF VOICED LEVEL COMPARISON OF VARIOUS F0 COMPONENT IN HINDI LANGUAGE

	Emotion			
F0 value	Happy	Anger	Surprise	Sadness
Mean	1.06272	1.07835	1.10539	0.98335
Std. deviation	1.03502	1.24624	1.22316	0.81933
Maximum	0.92749	0.98410	0.92001	0.77267
Minimum	1.36038	1.18672	1.06798	1.22815
Range	0.84532	0.94726	0.90045	0.69535
Median	1.04839	1.09013	1.08086	0.98362
Kurtosis	0.70871	1.12177	0.48953	0.66850

By theses, result of corpora a more general insight to show the result that voiced level tonal feature is as important as that of sentence level as not all the emotion are easily identified using sentence level or voice level. Our research also shows that not a single stress agent is used for examination of emotion. So we use a series of emotion for identification. At the end, we can say that same stress agents are used for identification in both the language but with different values. By analyzing these parameters we found that its quite easy identify the anger and sadness emotions while happy and surprise are so overlapped we have to need more parameters for identify them.

TABLE 5. RATIO OF SENTENCE LEVEL COMPARISON OF VARIOUS F0 COMPONENT IN HINDI LANGUAGE

	Emotion			
F0 value	Happy	Anger	Surprise	Sadness
Mean	1.14161	1.13170	1.17006	0.984503
Std. deviation	0.99252	1.10070	1.31470	1.08561
Maximum	1.00052	1.06035	1.08452	1.11452
Minimum	1.01286	1.02665	1.00499	0.98749
Range	0.99858	1.06550	1.09706	1.13407
Median	1.33174	1.14322	1.20946	0.97491
Kurtosis	0.77748	0.74410	0.58839	1.20737

V. CONCLUSION AND FUTURE WORK

In this paper we identified selected statistical component of F0 has taken by us for evaluation and comparison. We analyze that an existing system would not be able to identify the emotion, as its statistical features are different for both the languages. We try to develop an automatic recognisation model for detecting emotion using supervised machine learning as a future work.

ACKNOWLEDGMENT

I thank my research guide Mr. Anurag Jain for their excellent suggestion for me. I forget to thanks to the staff of University School of Information Technology for their valuable help

REFERENCES

[1] Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J. G., Jan. 2001. Emotion recognition in human-computer interaction, IEEE Signal Processing magazine, vol. 18, no. 1, 32–80.

[2] Dellaert, F. Polzin, and T. Waibel, "Recognizing emotions in speech", in Proc. ICSLP 96, vol. 3, pp. 1970-1973, USA, Oct. 1996.

[3] Carlos Busso, Sungbok Lee, Shrikanth Narayanan "*Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection*", IEEE Transaction of Audio and Speech of Language Processing, Vol. 17, No. 4, May 2009

[4] R.Cowie, "Describing the Emotional States Expressed In Speech" [C], In Proceedings of the ISCA, Belfast, 2000, Pp.11-18..

[5] Sun, X, "A pitch determination algorithm based on sub harmonic-to-harmonic ratio," the sixth International Conference of Spoken Language Processing, Beijing, China, 2000, 4, pp. 676-679.

[6] Sun, X.,"Pitch determination and voice quality analysis using sub harmonic-to-harmonic ratio" To appear in the Proc. of ICASSP2002, Orlando, Florida, May 13 -17, 2002.