



Variation in Speech Features of Different Emotions

Sandeep

Ph.D. Scholar, Dept. of Computer Science & Engineering
Singhania University, Jhunjhunu, Rajasthan
Email-id: sandeepkumarbansal@gmail.com

Dr. Anoop Sharma

Registrar, Singhania University, Jhunjhunu, Rajasthan
Email –id: sharmaanoop001@gmail.com

Dr. Sushma Bahuguna

Asstt. Professor, Banarsidas Chandiwala Institute of Information Technology, Delhi
Email – id: sushmabahuguna@gmail.com

Abstract

Speech is one of the main part of our daily communication to exchange our views, thoughts or ideas. But without knowing the emotions behind the speech it will be difficult to understand the actual meaning of the speech. There may be more than one meaning of same phrase which could be expressed through emotions attach with this phrase. With the change in emotion some prosodic features of speech also change. These features may include pitch, intensity, duration, formants etc. The aim of this paper is to analyse the variation in prosodic features of speech with emotions, and to compare the value of prosodic features of emotional speech with the features of neutral speech. This paper measures the variation in the feature pattern of the emotion i.e. the rise or fall in the value of each features of emotion.

Keywords: Emotional speech, prosodic features, speech synthesis, pitch, intensity and duration.

Introduction

Speech is an integral part of our communication. The characteristics of speech are divided into two categories namely linguistic and paralinguistic characteristics. To understand the actual meaning of words and phrases we must also know the emotions behind it. The prosodic and spectral features extracted from speech are used in emotion identification. Each speaker has unique physiological characteristics of speech production and speaking style. Speaker-specific characteristics are reflected in prosody.

The aim of this paper is to analyse the variation in prosodic features of speech with emotions, and to compare the value of prosodic features of emotional speech with the features of neutral speech. We consider the speech features like mean pitch, intensity, duration and formants of four basic emotions i.e. anger, happy, sad, surprise.

Methodology

In our study we recorded twenty five sentences of three different male speakers in different emotions and also neutral speech. From the recorded speech extracted main features like mean pitch, duration and intensity. After extraction we then compared the values of extracted features of emotional speech with

features of neutral speech. In our study we used PRAAT tool and parselmouth library of python which is used to work on functions of Prat tool. Figure 1 shows the flow chart followed by us in this study.

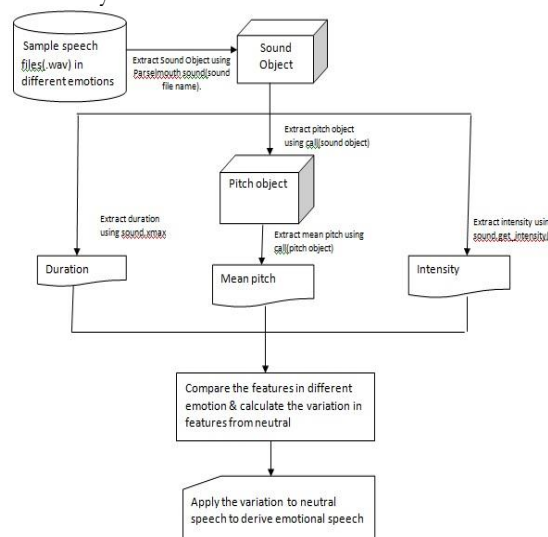


Fig 1. Flow of work

Following steps explain the Flow of work for this paper.

1. Collect Sample speech files(.wav) in different emotions like anger, sad, happy, surprise and neutral.

2. Extract Sound Object using Parselmouth sound(sound file name).
3. Extract pitch object using call(sound object). Then extract mean pitch using call(pitch object)
4. Extract Intensity using get_intensity(sound object),
5. Extract duration using Sound.xmax.
6. Compare the features in different emotion using average, sum and difference of mathematics.
7. Calculate the variation in features from neutral to emotional speech using sum average, percentage and difference formula of mathematics.

Experiments and Results

Table 1: Average value measured of mean pitch, intensity and duration under different emotions for all

Emotions	Mean Pitch (Average)	Intensity (Average)
Anger	217.48	81.69
Happy	178.21	77.42
Natural	147.62	73.71
Sad	172.51	72.03
Surprise	212.80	79.48

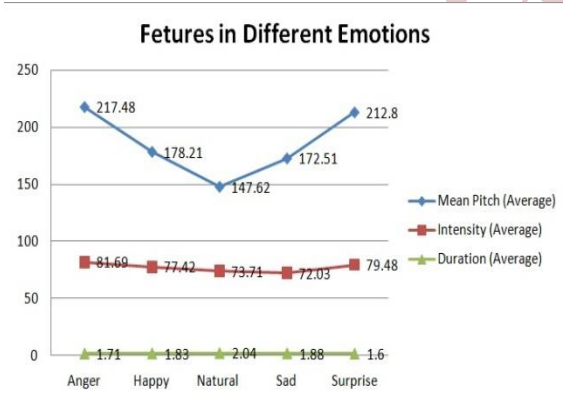


Fig 2: Graphical representation of average value measured of mean pitch, intensity and duration under different emotions for all speakers.

Prosody of speech characterizes emotions. Prosody generally refers to pitch, duration and intensity. Human beings use prosodic cues to encode and decode emotions in speech [1]. Mean F0 and speech rate are generally highest for emotions associated with high sympathetic arousal e.g., anger, fear; a slower speech rate is generally observed in passive, e.g., sad utterances [2]. In our study we recorded about 200 speech

corpora by different speakers in different emotions and tried to find the prosody for each emotion which is as stated:

Neutral: The mean pitch of neutral state is high but less than happiness[4]. Pitch value for neutral emotion is greater than sad emotion but less than all other emotions [5]. It also state that the intensity of neutral emotion is greater than fear emotion but slightly less than sad emotion[5]. As shown in Table 1 in our study for neutral emotion average of mean pitch has found 147.62 Hz and it range between 106.07 Hz to 187.33 Hz. For same set of speech signals the average duration is 2.04 sec., while the range is between 0.92 sec to 3.81 sec. The intensity measured is 73.71 dB with a range from 56.19 dB to 78.53 dB.

Anger : An increase in mean F0 and mean intensity has been found in angry speech [3]. Anger also seems to be characterised by an increase in high frequency energy that, together with the increase in intensity [3]. Some other research also shows that the mean pitch is higher with high variance in state of anger and Anger in speech causes increased intensity [4]. Pitch value for anger emotion is higher than other emotion but it is less than happiness emotion[5]. For intensity it says the anger emotion has the highest intensity value than other emotions [5]. In our experiment as shown in Table 1, for anger emotion average of mean pitch is 217.48 Hz and it range between 160.85 Hz to 296.35 Hz. For same set of speech signals the average duration is 1.71 sec., while the range is between 0.83 sec to 3.01 sec. The intensity measured is 81.61dB with a range from 78.29 dB to 85.45 dB.

Happy : There is a strong convergence of findings of increases in mean F0, F0 floor and intensity[3]. happiness causes slightly increased intensity[4]. In the state of happiness rhythm is rather fast with variance of phoneme duration [4]. Pitch value for happy emotion is highest [5]. Intensity of happy emotion is greater than all emotion but less than anger emotion[5]. As discussed in Table 1 we find that for happy emotion average of mean pitch is extracted as 178.21 Hz and it range between 108.77 Hz to 250.06 Hz. The average duration is 1.83 sec., while the range is between 1.01 sec to 3.73 sec. The intensity measured is 77.42 dB with a range from 68.87 dB to 81.37 dB.

Sad :Decreases in mean F0, F0 floor, F0 range, and intensity are usually found in sad emotion [3]. it is low pitched with little high frequency energy

in the state of sad [4]. Sad emotion shows lowest pitch value as compared to all other emotions [5]. The intensity of sad emotion is lower than happy and anger emotion but it is slightly greater than the neutral emotion[5]. From Table 1 we analyse that sad emotion shows mean pitch is 172.51 Hz and it range between 107.17 Hz to 252.59 Hz. For same set of speech signals the average duration is 1.88 sec., while the range is between 1.22 sec to 3.00 sec. The intensity measured is 72.03 dB with a range from 60.22 dB to 79.92 dB.

Surprise : The mean pitch is highest with high variance in the state of surprise[4]. Surprise emotion are characterized by fast rhythm with little variance of phoneme duration[4]. As shown in Table 1, mean pitch of surprise emotion is 212.80 Hz and it range between 165.62 Hz to 260.96 Hz. For same set of speech signals the average duration is 1.60 sec., while the range is between 1.01 sec to 2.07 sec. The intensity measured is 79.48 dB with a range from 76.77 dB to 81.48 dB.

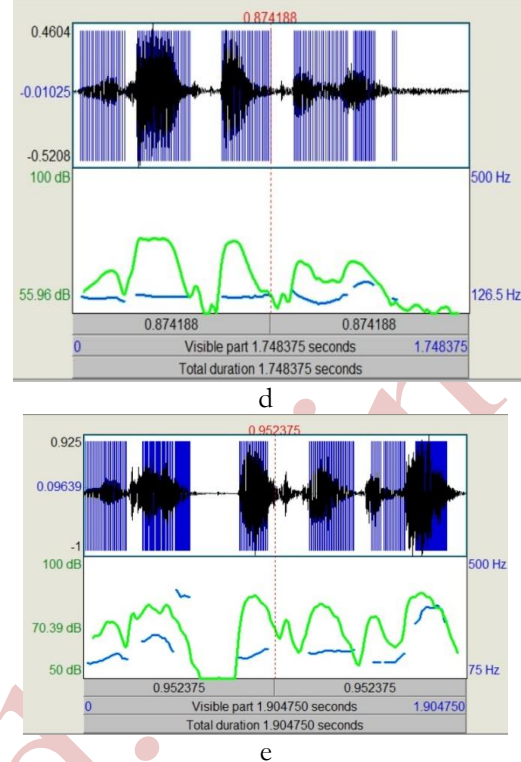
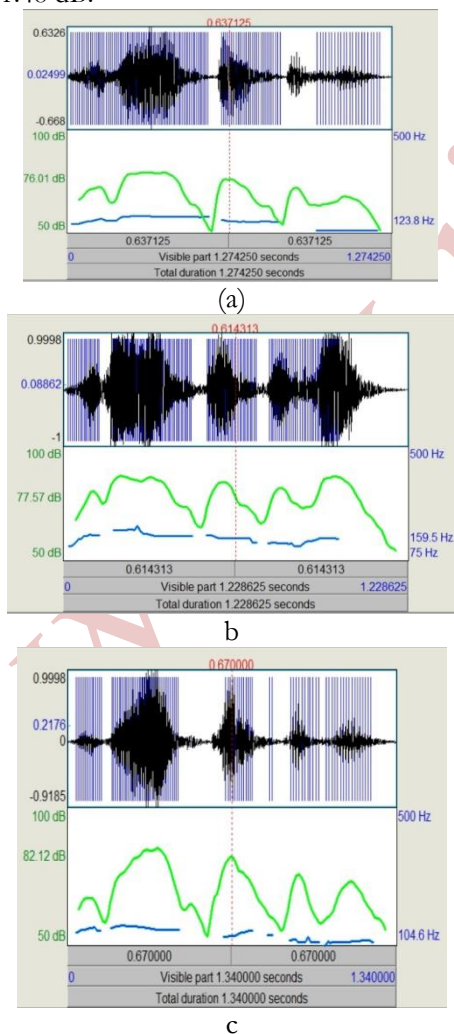


Fig 3: Samples of speech signal under different emotions a(neutral), b(anger), c(happy), d(sad), e(surprise) for hindi sentence “मुझे आज काम करना है।”

Variation in Features under different Emotions

Emotions	Mean Pitch (Average)	Intensity (Average)	Duration (Average)
Anger	47.31	10.71	-16.07
Happy	20.71	5.02	-9.97
Surprise	44.14	7.81	-21.54
Sad	16.85	-2.28	-7.77

Table-2 Variation in features from neutral speech signal to emotional speech signal

From Table-2 we find for anger emotion the mean pitch increased by approx. 47% and intensity is increased by approx. 11% but duration is decreased by approx. 16%, while for happy emotions the mean pitch increased by approx. 21% and intensity is increased by approx. 5% but duration is decreased by approx. 10%. For surprise emotion the mean pitch increased by approx. 44% and intensity is increased by approx. 8% but duration is decreased by approx. 21%. While for sad emotional speech the mean pitch increased by approx. 17% but intensity is decreased by approx. 2% and duration is decreased by approx. 8%.

Conclusion

From our study we conclude that the prosodic features play an important role to detect emotions from speech. It is also concluded that Mean pitch is highest in anger emotion and lowest in natural speech. Intensity is highest in anger emotion while lowest in sad emotion. Duration is highest in natural emotion and lowest in surprise emotion. If we want to convert our speech signal of one emotion into another emotion then we have to increase or decrease the features up to a range of the target emotion. There are some other features like formants, jitter and shimmer also which plays an important role in defining and processing the speech signal with their emotions.

References

1. A. Haque, K. S. Rao in "Modification of energy spectra, epoch parameters and prosody for emotion conversion in speech", in International Journal of Speech Technology, October 2016, DOI: 10.1007/s10772-016-9386-9.
2. C. Breitenstein, D. V. Lancker, I. Daum in "The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample", in Cognition and Emotion, 2001, 15 (1), 57-79.
3. T. Johnstone, University of Western Australia & University of Geneva in "The effect of emotion on voice production and speech acoustics" in 2001.
4. S. Bahuguna, Y. P. Raiwani in "A study of acoustic features pattern of emotion expression for hindi speech", at International Journal of Computer Engineering and Technology (IJCET), November - December (2013), Issn 0976-6367(Print), Issn 0976 - 6375(Online), Volume 4, Issue 6,
5. Jasdeep kaur, Kailash Juglan and Vishal Sharma, in "Role of Acoustic Cues in Conveying Emotion in Speech", J Forensic Sci & Criminal Inves 11(1): JFSCI.MS.ID.555803 (2018).