Language Specific Speech Feature Variation

Surbhi Dewan

CSE & IT Department
The NORTHCAP University
Gurgaon India
dewan.surbhi0250@gmail.com

Sumanlata Gautam
CSE & IT Department
The NORTHCAP University
Gurgaon India
sumanlatagautam@ncuindia.edu

***Abstract*—Speech is basically used to impart message from one person to another. There are various properties of speech that may vary from person to person or from language to language. The power of human language is found to be effected by variations in language. However, not much work has been done to analyse similarities and dissimilarities between speech features between English and Hindi language. The prosodic statistics for instance like stress and rhythm which are basically coded into intensity, pitch and formants. We have further examined the utilization of pitch and formants to study the linguistic difference of speech properties in English and Hindi Language. We clustered the speech samples into two categories and concentrated basically on pitch and formant values of speech signals. From our study we observed a significant change in the values of pitch and formants in English and Hindi language.**

Keywords— pitch; prosodic; formants; speech signal; speech features; frequency.

I. INTRODUCTION

Speech is the primarily basis for communicating the information. There features of speech signal are broadly categorized into two; Temporal features and Spectral Features. Temporal features are time sphere features. It is quite simple to extract temporal features of speech signal and they have effortless substantial elucidation. For instance energy of speech signal, maximum or minimum amplitude, and crossing rates. Whereas, the Spectral features of a speech signal are based on the frequency of signal. They are analysed primarily by transforming the time domain speech signal to frequency domain speech signal by application of Fourier Transform. For instance radical frequency, spectral density, spectral drift. These spectral features can be used to deduce the rhythm, intonation, pitch and intensity.

Majorly the most familiar spectral feature utilized in the study of sound is the domain of frequency inflection. This spectral feature assist us in examining the rhythm, intonation, pitch and formants of the speech signal. In speech audible, application of Fourier transform yields symphonic segments of signal known as spectral structure of speech signal. The symphonic segment frequencies increase as the speech signal is repeated over time. However, speech prosody is mosaic. It is mutually exclusive and utilized by every language, still it is eminently volatile among all languages. Speech prosodic features diversify on various factors such as speaker's style, emotions, environment, age and duration. Also certain symphonic features of language measures such as rhythm, tone and stress also diversify the variance in utterance. However, in our study we have concentrated on the variations in pitch and formants transition of speech signal. In the science of speech and phonemics, a formants is mostly determined as amplitude crest of spectrum of frequency in speech signal. Use of a spectrogram provides an assessment for vocal sonority. The lowest frequency formant is known as F1, second is F2, and third is F3. Usually, the formants F1 and F2 are sufficient to prescribe the peculiarity of vowels. Further the paper is divided into three sections explaining the methodology applied and concluded results.

II. RELATED WORK

This section provides the discussion on conclusions segregated from varying sources.

- In [1] the author proposed a study on identifying the utilization of pitch and intensity features of language to determine the linguistic stretch of six different South African languages. It is demonstrated that cluster analysis concluded from the pitch and intensity distance matrix that the distance matrix of pitch closely correlates with human emotive distances while on the other hand intensity matrix demonstrates no such correlation patterns about six languages respectively .

- Zhu et al. [2] about how to automatically identify the lingual stress of English. The coefficients based on frequency , energy and duration are used to verify stressed literals from that of insignificant literals by using linear distribution of overlapping coefficients .It was concluded that an appropriate directing of audial characteristic is important than association of distinctive features .

- Shrawankar et al. [3] discussed about the various methodologies and techniques used for extraction of speech prosodic features. Further the authors have discussed about the merits and demerits of these techniques.

- Farahani et al. [4] proposed a framework emphasizing the distinct transitions of pitch, to use it as a medium to identify the speaker. It was concluded that the framework is capable of presenting the progression in patterns of pitch associated with varying time scales.

- Kumar, Abhijeet, et al.[5] proposed a study on various processing mechanism used for noise reduction, channel and speaker assignment. Further the authors performed testing between MFCC and GMM dependent classifier.

- Mary, Leena, et al. [6] developed and examined a new methodology extraction and presentation of speech prosodic characteristics precisely from speech signal. They excogitate that prosody is associated with units of language like syllables. It is represented in context of transition in extended framework for instance frequency, power and time duration. Thus, syllable is selected as fundamental part to present prosodic features, which are automatically detected through vowels onset point acting as one of the essential medium for extraction of acoustic speech features.

-

III. METHODOLOGY

In this section, we will discuss about the procedure we applied to conduct our study. We performed a study to analyse the variation in speech features when the person speak in two different language they are English and Hindi. We analysed the speech sample based on various features of speech such as pitch, intensity, and formants by using PRAAT tool.

A. *Speech Sample*

We recorded and collected the speech samples of 20 adults including males and females. We categorized our data into two on the basis of language, one group for English and another for Hindi as shown below:

Table1: Speech sample details

Category	No. of Participants
Adult_English	20
Adult_Hindi	20

B. *Experimental Setup*

For collection of speech sample we considered a quiet room with as much as less noise for better results. Ideally windows of the room were closed and we switched off the fans and air conditioners as well. For recording we used a professional recorder that recorded the samples of speech at 16 bit PCM and 44 kHz sampling rate. The

participant was first made to speak in English and then in Hindi. All the participants read the same paragraphs respectively in both the languages. The samples were recorded in 16-bit PCM stereo type channel, but we converted it to 16-bit PCM mono to attain more accurate results because it provides excellent speech intelligibility. All the participants participated actively for the experiment

C. *Data Analysis*

For analyses of data we emphasized on the speech variations in pitch and formants of the speaker. We used two software tools: PRAAT and MATLAB. For calculating the mean pitch of each sample we used PRAAT software, and for determining formants we used MATLAB. For adult English speech samples we named the category as AE, and similarly we named the category for Hindi adult samples as AH. Further we performed t-pair testing on the samples to measure the significant difference in pitch of adult English speech samples and adult Hindi speech samples .The figure.1 depicts the value of P obtained after performing t-Test:

t-Test: Paired Two Sample for Means

	Variable 1	*Variable 2*
Mean	199.6491	204.0661
Variance	3059.517	3901.12
Observations	15	15
Pearson Correlation	0.993228	
Hypothesized Mean Difference	0	
df	14	
t Stat	-1.72931	
P(T<=t) one-tail	0.052865	
t Critical one-tail	1.76131	
P(T<=t) two-tail	0.10573	
t Critical two-tail	2.144787	

Fig.1 t-Test Paired Sample Means for Pitch

Here in the above figure we can see that the P value which is 0.052865 is less than t-value 1.76131. It depicts that the pitch values in English and Hindi differ significantly. Further we calculated the mean pitch value in Hz of both the language samples, to observe the significant change in pitch value. The Table.2 below presents the mean pitch value of all the samples regardless of gender. Here we can observe that there is a significant difference in mean pitch value of English and Hindi samples by a factor of 5 Hz.

Table.2 Mean Pitch Value (in Hz)

	Adult_English	Adult_Hindi
Pitch (in Hz)	199.6491	204.0661

The figure2. Below depicts the significant difference in values of pitch in context of English and Hindi language. It can be concluded from the below graphical representation that there is a significant change between English and Hindi. The prosody of speech differs when we speak two different language.

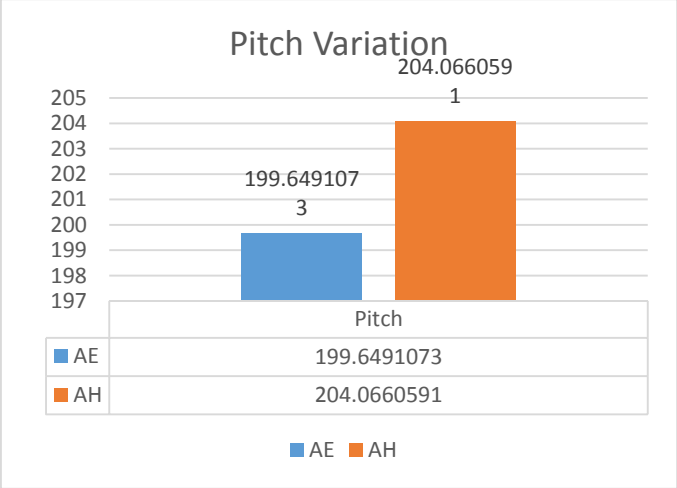

Fig.2 Graphical Representation of Variation in Pitch

The table.3 below shows the value of formants F1, F2 and F3 for both the languages. Also looking at the table we see distinct values of formants in English and Hindi language. The formant values of F1, F2 and F3 in English is more than that of Hindi by a factor of 5.67 Hz in F1, 126.38 Hz in F2 and 62.23 Hz in F3.

Table.2 Formant Values (in Hz)

Formants	F1(in Hz)	F2(in Hz)	F3(in Hz)
Adult_English	391.716	896.0807	1888.313
Adult_Hindi	397.3933	769.698	1826.08

The figure 3. Below depicts the variation in formant values of speech signal in English and Hindi language. From the graphical representation we observed that the formant values for English are more than for Hindi Language. This also concludes that English is a stress based language because the values of F1 and F2 are used to verify the quality of vowel. Since the values of F1 and F2 are greater than those in Hindi.

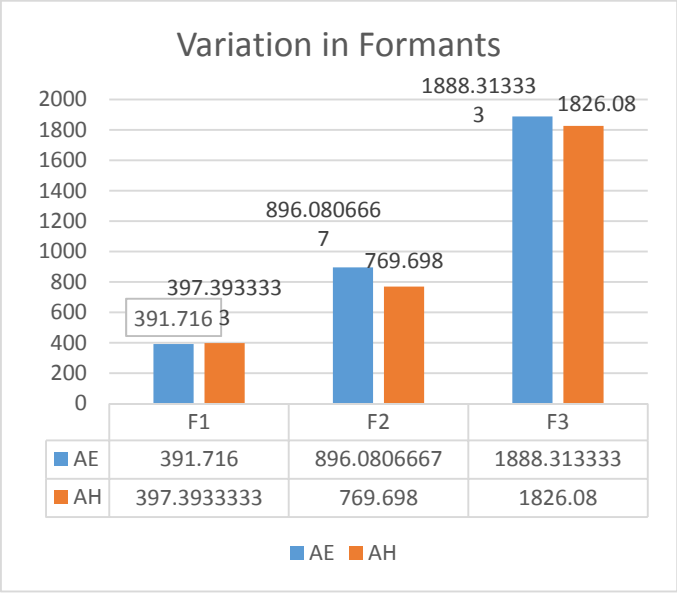

Fig.3 Graphical Representation of Variation in Formants F1, F2 and F3.

IV. CONCLUSION AND FUTURE WORK:

From our study it is observed that the speech prosodic features vary from language to language. As, the tone, stress, intensity, and rhythm changes whenever we speak different language and sometimes we may speak the language with pauses. These are some factors that differs from language to language. Apart from these factors we observed that mean pitch value for Hindi is significantly more than that of English, this concludes that Hindi is a stable language in contrast to English. Also the values for formants in English are more than that of Hindi that concludes that stress is more in speaking English in contrast to Hindi. In future we can work on speech of intellectually disable person by comparing the persons speech signal , whichever language in which less stress is required can be used for such special people to make them learn speak .

REFERENCES

[1] Zulu, Peleira Nicholas. "Language classification using prosodic features: Comparing intensity and pitch." Information Science, Computing and Telecommunications (PACT), 2013 Pan African International Conference on. IEEE, 2013.

[2] Zhu, Yun, Jia Liu, and Runsheng Liu. "Automatic lexical stress detection for english learning." Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 International Conference on. IEEE, 2003.

[3] Shrawankar, Urmila, and Vilas M. Thakare. "Techniques for feature extraction in speech recognition system: A comparative study." arXiv preprint arXiv:1305.1145 (2013).

[4] Farahani, Farhad, Panayiotis G. Georgiou, and Shrikanth S. Narayanan. "Speaker identification using supra-segmental pitch pattern dynamics."Acoustics, Speech, and Signal Processing, 2004 . Proceedings.(ICASSP'04). IEEE International Conference on. Vol. 1. IEEE, 2004.

[5] Kumar, Abhijeet, et al. "Effective preprocessing of speech and acoustic features extraction for spoken language identification." Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference on. IEEE, 2015.

[6] Mary, Leena, and Bayya Yegnanarayana. "Extraction and representation of prosodic features for language and speaker recognition." Speech communication 50.10 (2008): 782-796.