Volume 4, Issue 6, June 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Facial Changes during the Pronunciation of Dental and Labial Consonants in IR Images

Chandroop Gupta[1], Sandeep kaushal[2], Jang Bahadur Singh[3], and Parveen Lehana[4*]

[1]M. Tech, Amritsar College of Engineering and Technology, Punjab, India
[2]Assistant Professor, Amritsar College of Engineering and Technology, Punjab, India
[3]Ph.D Scholar, Dept of Physics and Electronics, University of Jammu, J&K, India
[4*]Associate Professor, Dept of Physics and Electronics, University of Jammu, J&K, India

Abstract— Speech is a complex non-stationary signal. Its primary concern of speech processing is the production and recognition of facial expressions of emotion. Emotions play a fundamental role in human facial expression recognition. Whereas Infraraed facial images focused on differences in temperature distribution of facial muscles while speaking. Infrared cameras work by detecting radiation coming from a body through a lens and converting the points of temperature into digital form. Once the computer has the digitized image it transfers it into a picture for us to interpret. The paper is based on the changes in facial muscles during the utterance of dental and labial consonants. The average pictorial distances of the IR images at the instant of utterance of dental and labial consonants were calculated and these results can be later on used to determine that which phoneme was uttered. The average pictorial distances of labial consonants is more as compared to dental consonants. Also the pictorial distances are more speaker dependent as compared to speech dependent.

Keywords—Average pictorial distance, Infrared images, Devanagari script, Dental consonant, labial consonants.

I. Introduction

Signal is a physical quantity that is measurable. System is a physical entity that exists. Signal is produced from a system. Depending on the nature of signal, it is categorized into several classes based on some criterion. Some of the classifications include continuous v/s discrete, periodic v/s aperiodic, energy v/s power, periodic v/s aperiodic, deterministic v/s random, stationary v/s non-stationary and so on. The most of the signals, systems and signal processing concepts were taught based on the implied assumption of stationarity of the signal under consideration. Speech signal processing deviates in this aspect. This is because speech is a complex non-stationary signal [1]. The primary concern of speech processing is the production and recognition of facial expressions of emotion. Emotions play a fundamental role in human facial expression recognition [2].

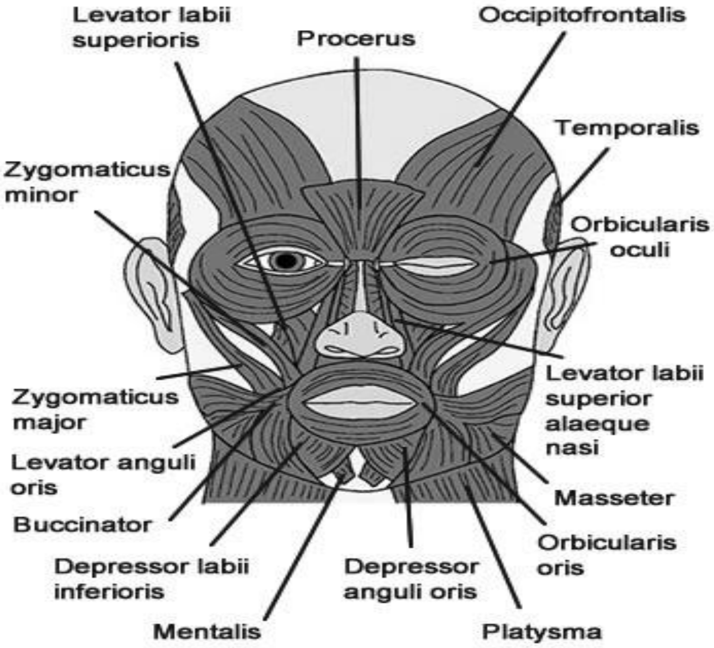

Fig.1 Different facial muscles [5]

Infraraed images focused on differences in temperature distribution on facial muscles while speaking. Infrared cameras work by detecting radiation coming from a body through a lens and converting the points of temperature into digital form. Once the computer has the digitized image it transfers it into a picture for us to interpret. The lens is constructed of germanium, one of the only materials that will allow radiation through. Glass that is transparent to the human eye is totally opaque in infrared light, which cannot therefore, pass through it. The portion of the electromagnetic spectrum perceived by the human visual system is generally called the "visible spectrum" and ranges from (about) 380 to 700 nanometers in wavelength. The near-infrared spectrum is located just after the red wavelength and comprises wavelengths that range from 700 to 1100 nanometers.Even though the NIR (near infrared images) band is located next to the visible one, there is, in general, almost no correlation between a visible and NIR signal (i.e., knowing the colour and brightness of an object gives no information about its NIR response) [3].

EMG is one of the biomedical signals that measures electrical current generated in muscles during its contraction representing neuromuscular activities. Contraction and relaxation of the muscles are controlled by the nervous system. Hence, the EMG signal is a complex signal and depends upon the anatomical and physiological properties of muscles. These signals from specific facial muscles are recorded for speech recognition and system automation. EMG signals are generally recorded using small surface electrodes placed near to each other. EMG activity is frequently recorded from specific muscles and plays a prominent role in the expression of elementary emotions and speech generation [4]. The paper is based on the changes in facial muscles during the utterance of dental and labial consonants. Section II describes about speech production and its types. Methodology of the investigations is explained in section III. The results and conclusions are presented in section IV and section V respectively.

II. Speech Production And Its Types

Some sort of air source is required for speech production. Speech sounds are produced by forcing air upwards from the lungs, an action that is used in normal breathing. But just forcing air out of the lungs does not result in speech sounds. The central organs involved in the production of speech sounds include: the lungs, larynx, and vocal tract. While each of these is used for normal physiological processes involved in breathing and eating, they also function in the production of speech [6].

The larynx, more commonly known as the voice box, is crucial in the production and differentiation of speech sounds. The larynx is located at exactly the point where the throat divides between the trachea, which leads to the lungs, and the esophagus (the tube that carries food or drink to the stomach).Over the larynx is a flap called the epiglottis that closes off the trachea when we swallow. This prevents the passage of food into the lungs. When the epiglottis is folded back out of the way, the parts of the larynx that are involved in speech production can be seen [7].

Table I. Chart of the first 25 consonants of the Devanagari script

	Unvoiced		Voiced		Nasals
	Unaspirated	Aspirated	Unaspirated	Aspirated	
Velar	क	ख	ग	घ	ङ
Palatal	च	छ	ज	झ	ञ
Retroflex	ट	ठ	ड	ढ	ण
Dental	त	थ	द	ध	न
Labial	प	फ	ब	भ	म

There are two thin sheets of tissue that stretch in a V-shaped fashion from the front to the back of the larynx. These are called the vocal folds. (You'll often hear vocal "cords," which is doesn't accurately convey the way the muscle works.) The space between the vocal folds is known as the glottis. The vocal folds can be positioned in different ways to create speech sounds. Air passes through the vocal folds. If the vocal folds are open and air passes unobstructed, the vocal folds do not vibrate. Sounds produced this way are called voiceless. But if the vocal folds are held together and tense and air doesn't pass unobstructed, the sounds produced this way are call voiced [8].

Consonants can be classified according to the place within the mouth that they are articulated.

Dental consonants are pronounced with the tip of the tongue touching the back of the upper front teeth. Examples of dental consonants in English include the "th" in "the", and the "th" in "thin". Labial consonants are pronounced with the lips. Examples of labial consonants in English include the "p" in "pit", the "b" in "boy", and the "m" in "man".

Consonants can also be classified according to their manner of articulation.

Unvoiced consonants are pronounced without vibrating the vocal cords. Examples of unvoiced consonants in English include the "s" in "sit", the "p" in "pit", the "t" in "time", etc. Voiced consonants are pronounced by vibrating the vocal cords. Examples of voiced consonants in English include the "z" in "zoo", and the "g" in "good". Unaspirated consonants are pronounced without a breath of air following the consonant. Contrast the pronunciation of the "p" in "spit" and the "p" in "pit"; the former is unaspirated, whereas the latter is aspirated..Aspirated consonants are pronounced with a strong breath of air following the consonant, as the "p" in "pit" [9]. Now, consider Table 1.which illustrate the first 25 consonants of the Devanagari script

III. METHODOLOGY

Each of the five subjects (S1, S2, S3, S4, and S5) will utter Dental and Labial consonants as mentioned in Devanagari script above and their corresponding IR facial images were captured. All the subjects taken were male having ages between 20-25 years and Average pictorial distances was obtained along with difference image and average spatial difference image for all dental and labial consonants. The complexities in the average spatial difference image helps to determine various variations while uttering different consonants.

IV. RESULTS

The Investigations were carried out to evaluate the effect of utterance of aspirated and unaspirated consonants and vowels, and then capturing corresponding IR images during the production of these consonants and vowels. Fig. 2 and Fig. 3 show IR images, difference image and average spatial difference in three different columns of two subjects for dental consonants. Fig. 4 and Fig. 5 show IR images, difference image and average spatial difference in three different columns of two subjects for labial consonants. Here we are only showing the results of two subjects (S1, S2)..

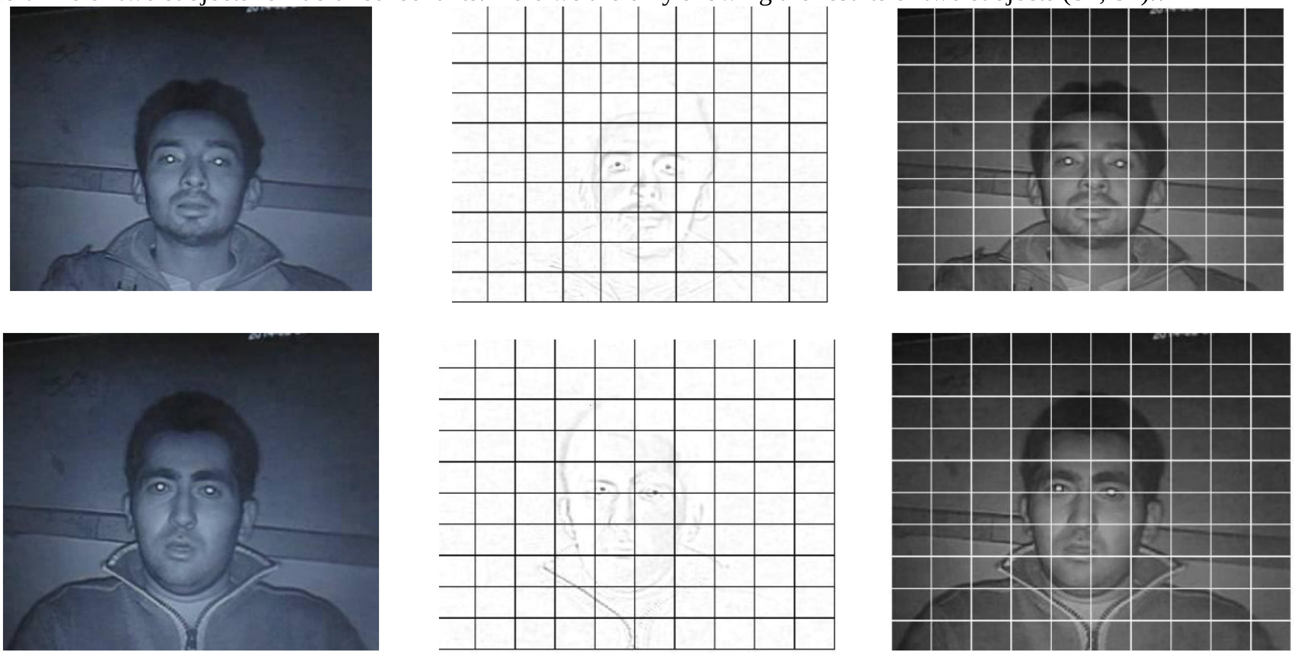

Fig. 2 IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the consonant त

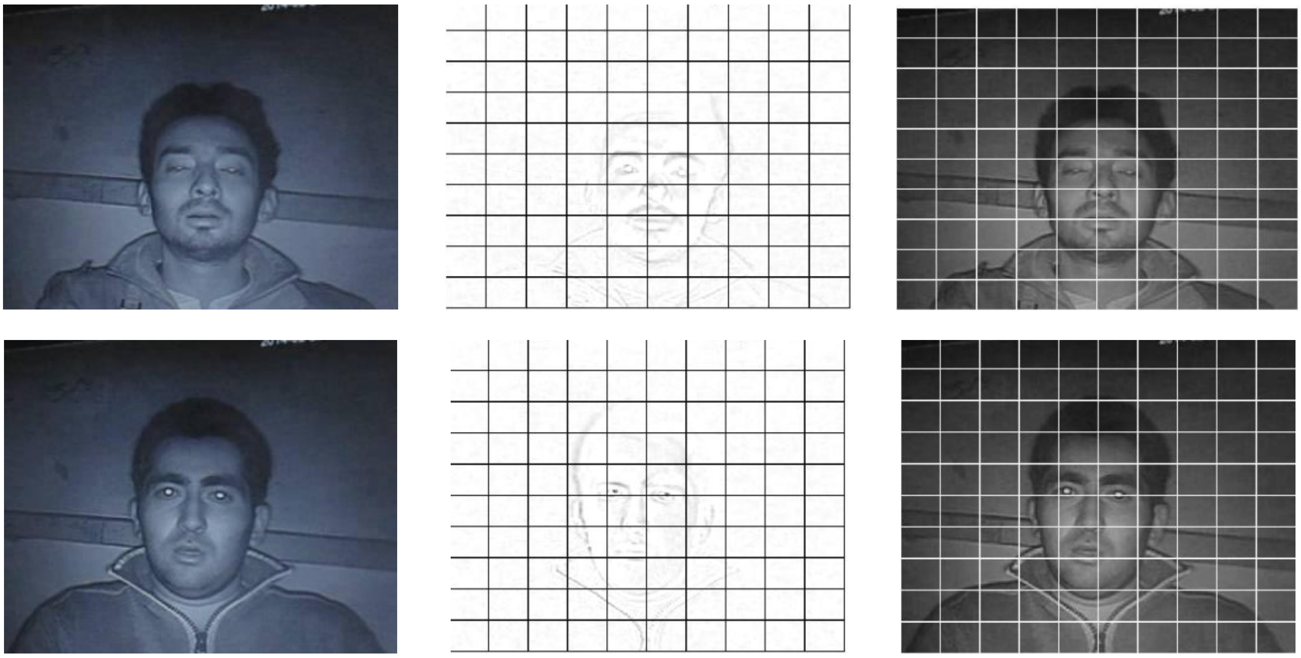

Fig. 3 IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the consonant थ

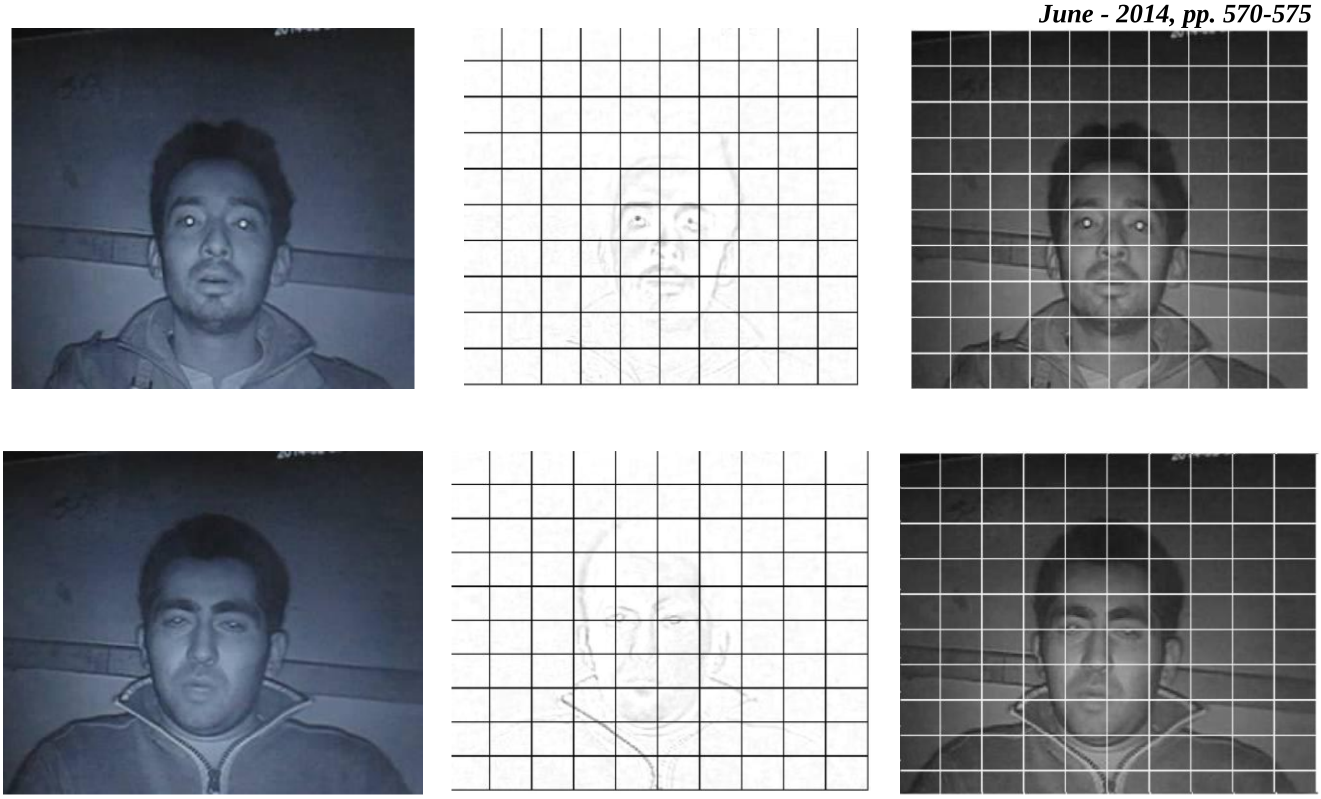

Fig. 4 IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the consonant प

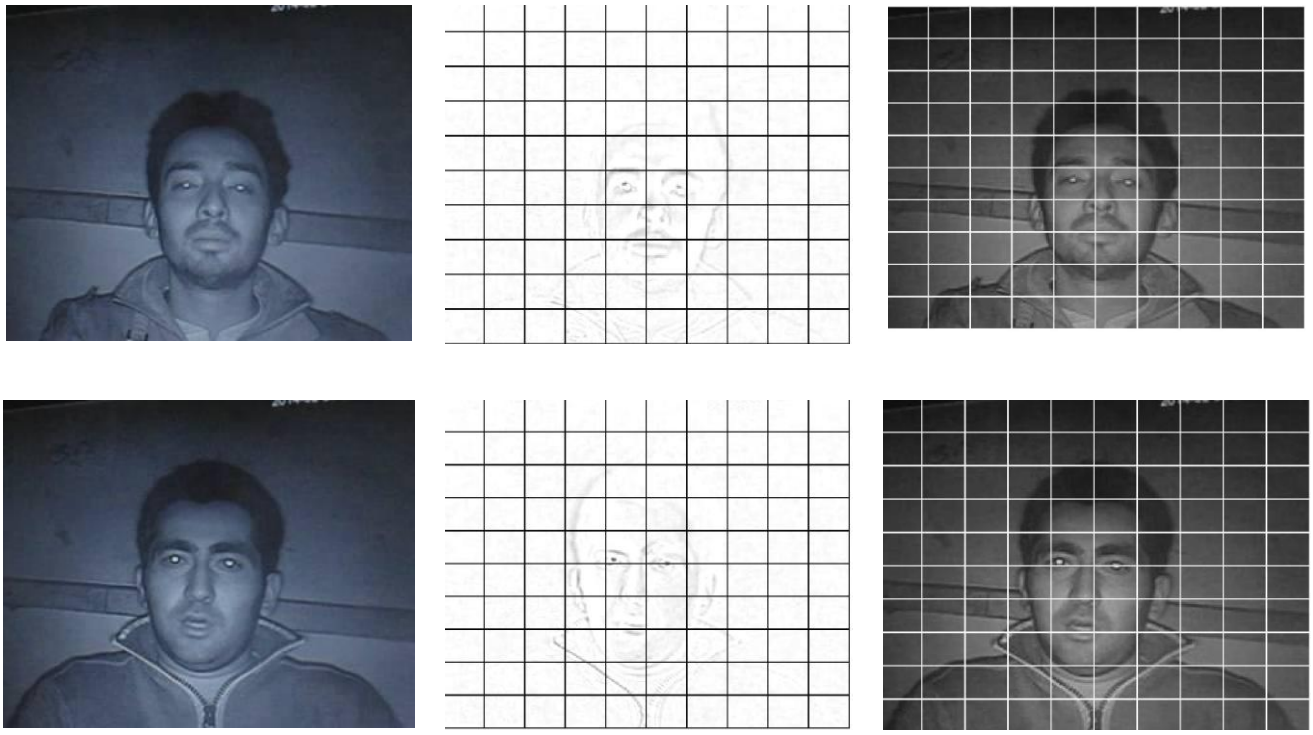

Fig. 5 IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the consonant फ

From the above results it is clear that the pictorial distances are more speaker dependent as compared to speech dependent. The minimum complexity of average spatial difference images are observed in Fig. 2 and the maximum complexity are in the case of Fig. 4.

Table II. Mean and average pictorial distances for all Dental consonants

Consonant			**Subjects**			**Mean**
	S1	**S2**	**S3**	**S4**	**S5**	
त	3.35	2.98	1.05	1.05	2.69	2.22
थ	3.05	2.60	0.90	1.29	2.66	2.10
द	3.08	2.86	1.11	1.24	2.72	2.20
ध	2.66	3.26	1.20	1.18	2.38	2.14

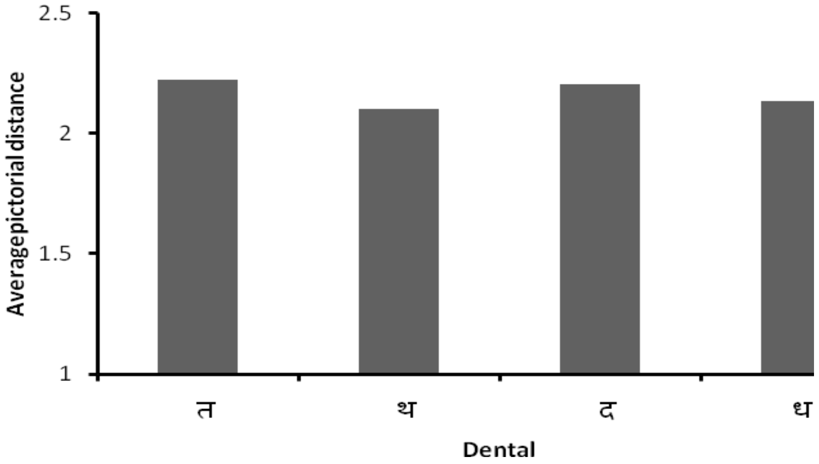

Fig. 6 Histogram of average pictorial distance for dental consonants

The Fig.6 histogram shows that the average pictorial distance is maximum for the consonant (त) i.e. 2.22 and is minimum for the consonant (थ) i.e2.10.

Table III. Mean and average pictorial distances for all Labial consonants

Consonant			**Subjects**			**Mean**
	S1	**S2**	**S3**	**S4**	**S5**	
प	2.75	3.01	0.93	0.96	2.47	2.02
फ	2.46	3.14	0.82	1.22	2.32	1.99
ब	2.56	3.44	0.84	1.32	2.42	2.12
भ	2.76	3.64	0.89	1.62	2.21	2.23

The Fig.7 histogram shows that the average pictorial distance is maximum for the consonant (भ) i.e. 2.23 and is minimum for the consonant (फ) i.e1.99.

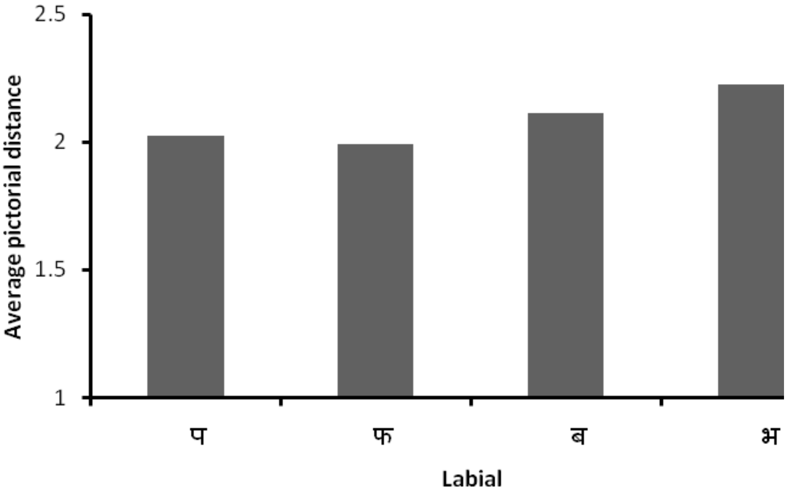

Fig. 7 Histogram of average pictorial distance for labial consonants

V. Conclusion

In this research work five male subjects were taken and their corresponding IR images at the instant of utterance of dental and labial consonants were captured and their average pictorial distances were calculated. These results can be later on used to determine that which phoneme was uttered. The maximum average pictorial distance is 2.23 for consonant (भ) and minimum average pictorial distance is 1.99 for consonant (फ). It can also be concluded that the pictorial distances are more speaker dependent as compared to speech dependent. Also the average pictorial distances of palatal consonants is more as compared to velar consonants.

References

[1] Available at http://iitg.vlab.co.in/?sub=59&brch=164&sim=371&cnt

[2] A. Martinez and S. Du, “A Model of the Perception of Facial Expressions of Emotion by Humans Research Overview and Perspectives” *Journal of Machine Learning Research*, pp. 13, 2012.

[3] D. Bhattacharjee and S. Gaungly, “Comparative study of human thermal face recognition based on Haar wavelet transform and local binary pattern” *Computational intelligence and neuroscience*, pp.6,2012.

[4] H. Riana, R. Singh, J. B. Singh and P. Lehana, “Effect of Unvoiced Consonants on EMG Signal” *International Journal of Advanced Research in Computer and Software Engineering*. pp. 3, 2013.

[5] Availablat

[6] http://droualb.faculty.mjc.edu/Course%20Materials/Elementary%20Anatomy%20and%20Physiology %2050/Lecture%20outlines/muscle_anatomy.

[7] Available at http://emedia.leeward.hawaii.edu/hurley/Ling102web/mod3_speaking/3mod3.2_vocalorgans.htm.

[8] D. Homer and T. Tarnozy, “The speaking machine of Wolfgang von kempelen”, *The Journal of the Acoustic Society America, pp* 22, 2013.

[9] R. L. Rabiner and R. W. Schafer, “Digital processing of speech signals,” *Prentice-Hall Inc. Englewood Cliffs*, New Jersey ,2012.

[10] Available online http://www.omniglot.com/language/articles/devanagari.htm