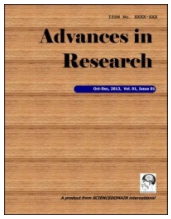

Advances in Research
3(4): 417-426, 2015, Article no.AIR.2015.037
ISSN: 2348-0394

SCIENCEDOMAIN *international*
www.sciencedomain.org

Facial Changes during the Pronunciation of Velar and Palatal Consonants in IR Images

Chandroop Gupta[1], Sandeep Kaushal[1], Jang Bahadur Singh[2] and Parveen Lehana[2*]

[1]Amritsar College of Engineering and Technology, Punjab, India.
[2]Department of Physics and Electronics, University of Jammu, J&K, India.

Authors' contributions

This work was carried out in collaboration between all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AIR/2015/12236
Editor(s):
(1) Anonymous.
(2) Anonymous.
Reviewers:
(1) Anonymous, Universidade Federal De Mato Grosso Do Sul- Brazil.
(2) Anonymous, Fourth Military Medical University, China.
(3) Anonymous, Lanzhou University of Technology, China.
(4) Anonymous, Institute of Technology Ladkrabang, Thailand.
(5) Anonymous, University of Tampere, Finland.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=756&id=31&aid=6987

Original Research Article

Received 24th June 2014
Accepted 14th October 2014
Published 18th November 2014

ABSTRACT

Speech is one of the most complex non stationary signals and is also the easiest way of communication between human beings. Speech signals can be exaggerated by the behavior and emotions of speakers. Facial expression can be recognized through electromyography (EMG) signals. These signals from specific facial muscles are recorded for speech recognition and system automation. The human face is an area for displaying different emotions. The temperature distribution on facial muscles can be captured through R camera, which may be regarded as texture features of images. In this research work five male subjects were taken and their corresponding IR images at the instant of utterance of velar and palatal consonants were captured and their average distances were calculated. The analysis of the results showed that the average of difference images were more speaker dependent as compared to speech content. The average distances of palatal consonants are relatively more as compared to velar consonants.

**Corresponding author: E-mail: pklehana@gmail.com;*

Keywords: Average difference images; infrared images; phonemes.

1. INTRODUCTION

Sound is a vibration that propagates as a mechanical wave of pressure and displacement, through some medium. Sometimes sound refers to only those vibrations with frequencies that are within the range of hearing for humans or for a particular animal. Speech, a form of sound, is a natural and efficient means of communication for human beings. These signals can be exaggerated by the behavior and emotions of speakers [1]. The human face is an area for displaying different emotions. In our skin there are large numbers of muscles which help us to express different emotions. The top facial muscles are linked bilaterally, whereas the bottom facial muscles are connected unilaterally to the opposite hemisphere. Information of gender, attractiveness and age, may be assumed through facial expression. Emotions play a fundamental role in the recognition of human facial expression [2]. Fang et al. [3] investigated that the facial expression recognition does not require manual specification of where the particular behavior occurs or the subjective imposition of thresholds and system uses the dynamic information of facial feature extracted from video sequences. The dynamic response of those features for the subject performing in given expression was obtained which does not rely on action unit and therefore eliminate the error due to incorrect initial identification of action units [3]. Gupta et al. [4] investigated that the facial expression changes during the pronunciation of dental and labial consonants which can be captured using IR camera.

Facial expression can be recognized through EMG signals. The shape and position of different facial muscles are shown in Fig. 1. 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. 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 signals play a prominent role in the expression of elementary emotions and speech generation [1].

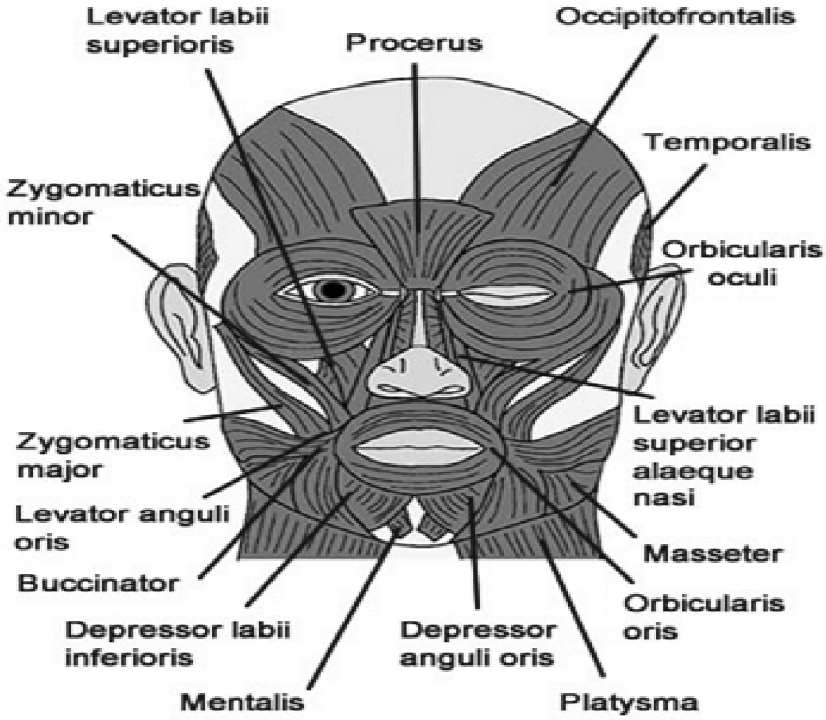

Fig. 1. Facial muscle [5]

The changes of temperature distribution on facial muscles and blood vessels may be captured through IR camera. The temperature changes may be regarded as texture features of images [6]. Infrared images reflect the temperature distribution of the facial muscles while speaking and should focus on changes of thermal distribution on facial muscles and blood vessels, which is caused by mood variety [7].

Yasunari et al. [8] investigated the effect of sensor fusion for recognition of emotional states using voice, face, and thermal images of the human faces. The IR radiations from the human body comes from a well known law given by Stefan and Boltzmann, which is expressed as $W = \varepsilon\sigma T^4$, where W is radiant emittance, ε is emissivity, σ is Stefan- Boltzmann constant and T is temperature. The author had taken the value of emissivity of the human skin as 1.0 as suggested in [9]. IR images do not depend upon skin colour, darkness, and lighting conditions [10]. IR images characteristics are extracted from the input image and its surroundings depending upon the temperature distribution.

The objective of this paper is to investigate the effect of voiced and unvoiced consonant on the IR images of facial muscles during their utterance. Section II represents the speech production mechanism. The methodology of the investigations is presented in Section III. The results and conclusions are presented in Section IV.

2. SPEECH PRODUCTION AND ITS TYPES

The mechanism of speech production consists of following processes. The first one is language processing, in which phonemic cipher generated in the brain is the result of conversion of the content of the utterance, generation of motor commands in the brain. For the production of speech articulatory movements are initiated; and the speech signal is produced [11]. This entire process can be considered as a chain mechanism passing through various levels like linguistic level, physiological level, and acoustic level Signal originating from the vocal folds when passing through the vocal tract acting as a linear filter generates speech at the output [12,13]. Vocal folds in case of males are usually longer than that in females, causing a lower pitch and a deeper voice. The length of male vocal folds is between 17.5 mm and 25 mm. and the length of female vocal folds is between 12.5 mm and 17.55 mm. The glottis is a gap between the vocal folds. The air passes through the vocal folds, makes them to vibrate, producing a periodic sound. The rate of vibration of the vocal folds is called as the fundamental frequency *F0* (pitch). From the point of view of *F0*, the larynx is the most important vocal organ. Fundamental frequency is between 80 Hz to 250 Hz for male speakers since a male female has *F0* that is between 120 Hz to 400 Hz [14]. The term pitch refers to the rate of vibration that is perceived by the listener. Fig. 2 shows various articulators working in natural speech production. Above the

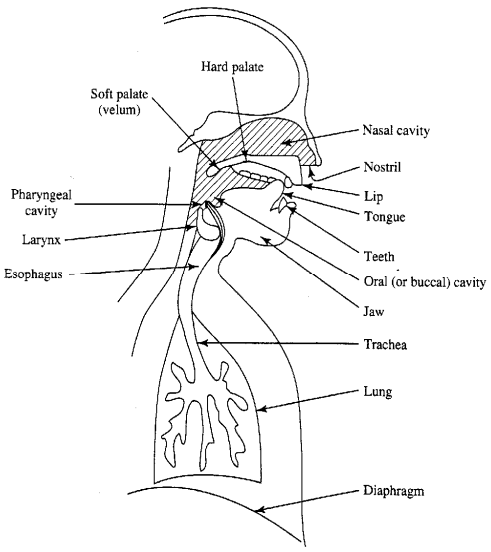

Fig. 2. Speech production mechanism [16]

larynx the human pharynx is situated behind the mouth, which is divided into two paths, one entering the mouth region and the other entering inside the nasal region [15]. The consonants are produced when the air flow through the vocal tract is constricted or stopped.

The consonant are grouped in two different categories-velar and palatal consonants. Velar consonants are pronounced with the back of the tongue touching the soft palate. Examples of velar consonants in English include 'k' as in keep, and 'g' as in good. Palatal consonants are pronounced with the tongue touching the hard palate. Examples of palatal consonants in English include 'ch' as in change and 'j' as in job. Depending upon their manner of verbalization the consonants can also be categorized as: voiced and unvoiced consonants Voiced consonants are pronounced by vibrating the vocal cords. Examples of voiced consonants include the 'z' in zoo, and the 'g' in good. Unvoiced consonants are pronounced without vibrating the vocal cords. Examples of unvoiced consonants includes 's' in sit, 'p' in pit, and 't' in time, etc. Consonants are grouped in two categories aspirated and Unaspirated consonants Aspirated consonants are pronounced with a strong breath of air following the consonant, as the 'p' in pit Unaspirated consonants are pronounced without a breath of air following the consonant [17].

3. METHODOLOGY

In order to investigate the effect of facial changes during the pronunciation of velar and palatal consonants five subjects (S1, S2, S3, S4, and S5) having age between 20-25 years were selected and each subject was asked to read a passage written in Hindi. The passage was designed in such a manner that it contained at least 10 instances of each consonant and vowel. For IR recording, the camera (Company name-Enter (IR camera), Model no –W600/R30, with VGA resolution 800*600) was placed at a distance of 3 feet from the subjects. A video of the faces of the subjects while speaking was recorded along with the speech signal. The length of the video of each subject was of about 10 minutes. From the IR video, images corresponding to the consonants listed in Table 1 were extracted. For the average pictorial distances corresponding intensity difference of image during silent and the image captured during the utterance of constant. Each difference image was divided into 100 blocks and mean of the intensity for each block was computed. Intensity of each block was replaced by the corresponding mean intensity for proper visualization of the blocks. Although investigations were carried out for each consonant, the results of only velar and palatal consonants are reported here.

4. RESULTS AND DISCUSSION

Investigations were carried out to evaluate the effect of utterance of velar and palatal consonants, and then capturing corresponding IR images during the production of these consonants and vowels. Figs. 3 to 6 shows IR images (first column), difference image (second column) and average spatial difference (third column) of two subjects for all velar consonants. Figs. 7 to 10 shows IR images (first column), difference image (second column) and average spatial difference (third column) of two subjects for all palatal consonants. Analysis of the figures shows that the difference images are more speaker dependent as compared to speech content. Further, the average spatial difference for S2 spreads over large area as compared to S1, also the outer boundaries of difference image for S2 is clearer than S1 as observed from the images in the last column. Mean and averages of difference images of all velar consonants are shown in Table 2 and plotted as histogram in Fig. 11. Similarly mean and averages of difference images of all palatal consonants are shown in Table 3 and plotted as histograms in Fig. 12. The histograms in Fig. 11 show that the average of difference images is maximum for the consonant (ख) and is minimum for the consonant (क). The histograms in Fig. 12 show that the average of difference images is maximum for the consonant (च) and is minimum for the consonant (छ).

Table 1. List of velar and palatal consonants

	Unvoiced		Voiced	
	Unaspirated	**Aspirated**	**Unaspirated**	**Aspirated**
Velar	क (ka)	ख (kha)	ग (ga)	घ (gha)
Palatal	च (ca)	छ (cha)	ज (ja)	झ (jha)

Referring Table 2 and Table 3 the overall mean of palatal consonants is more as compared to the overall mean of velar consonants. It may be due to fact that during the pronunciation of palatal consonants more acoustical energy is consumed as compared to velar consonants. This also may be the reason that palatal consonants show more facial changes than velar consonants.

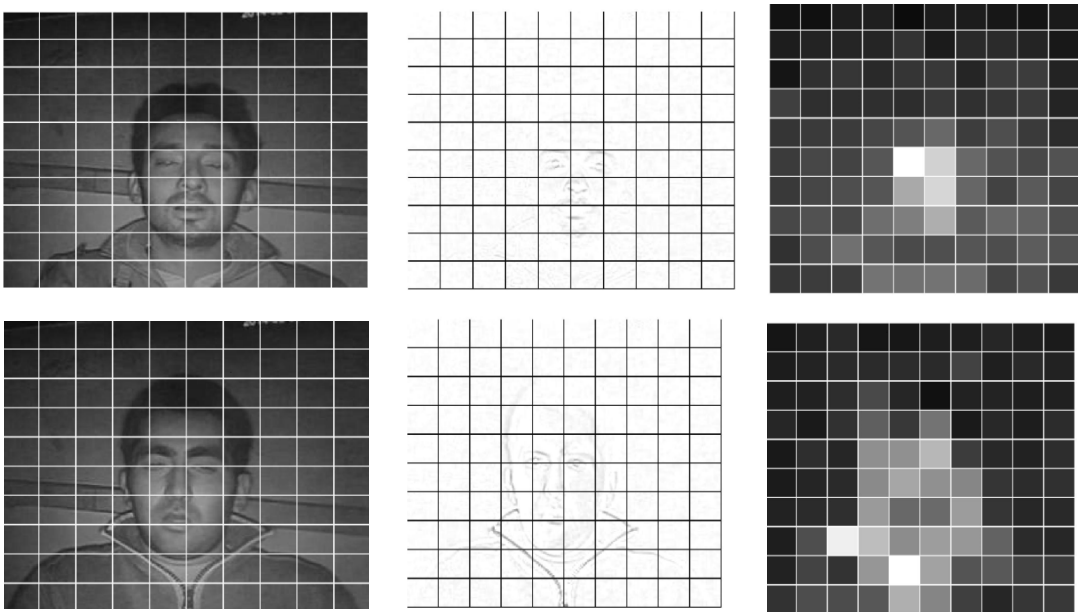

Fig. 3. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the velar consonant क (ka)

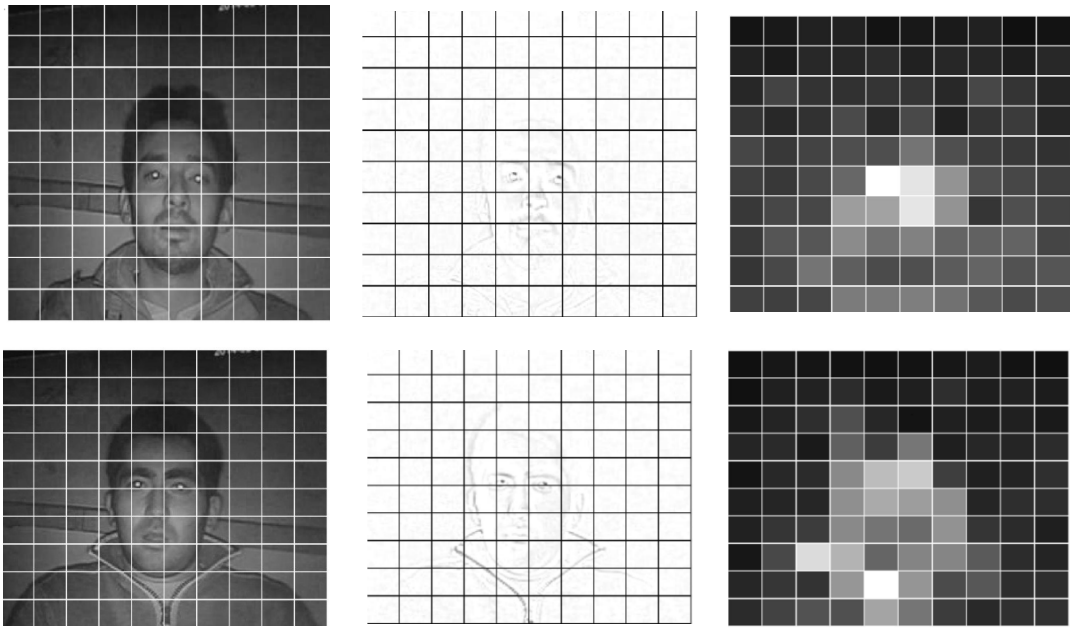

Fig. 4. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the consonant ख (kha)

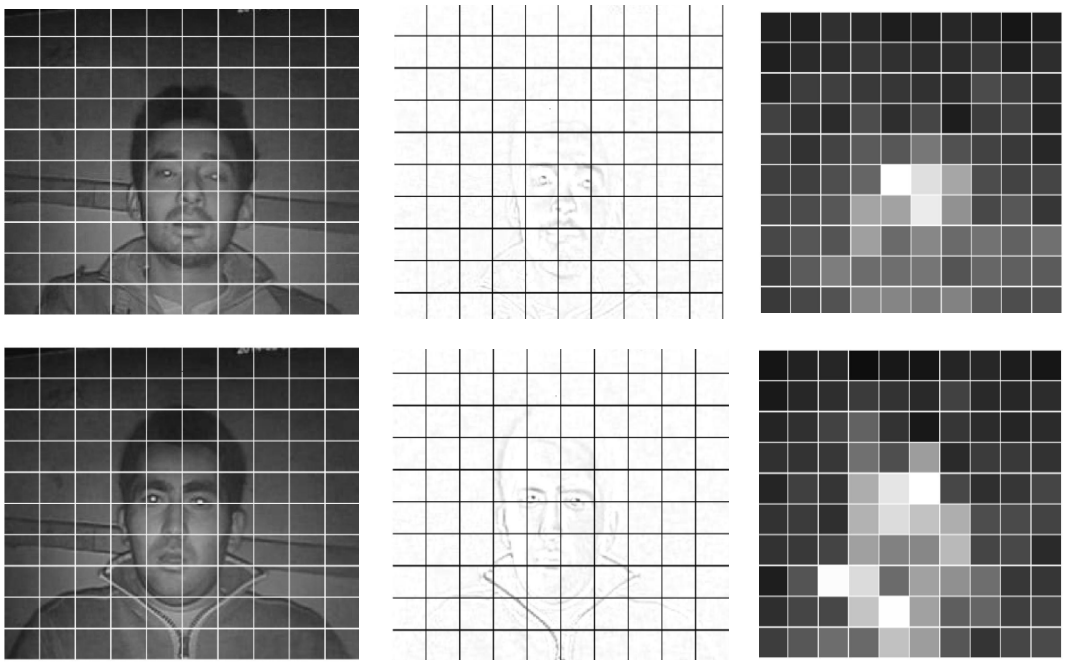

Fig. 5. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the velar consonant ग (ga)

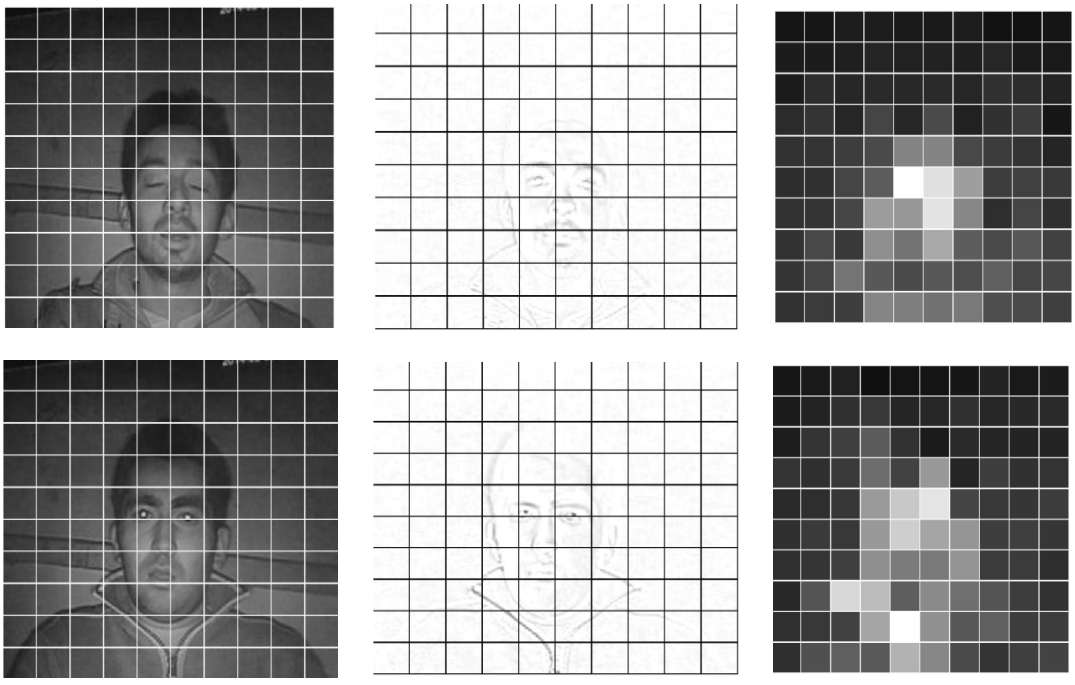

Fig. 6. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the velar consonant घ (gha)

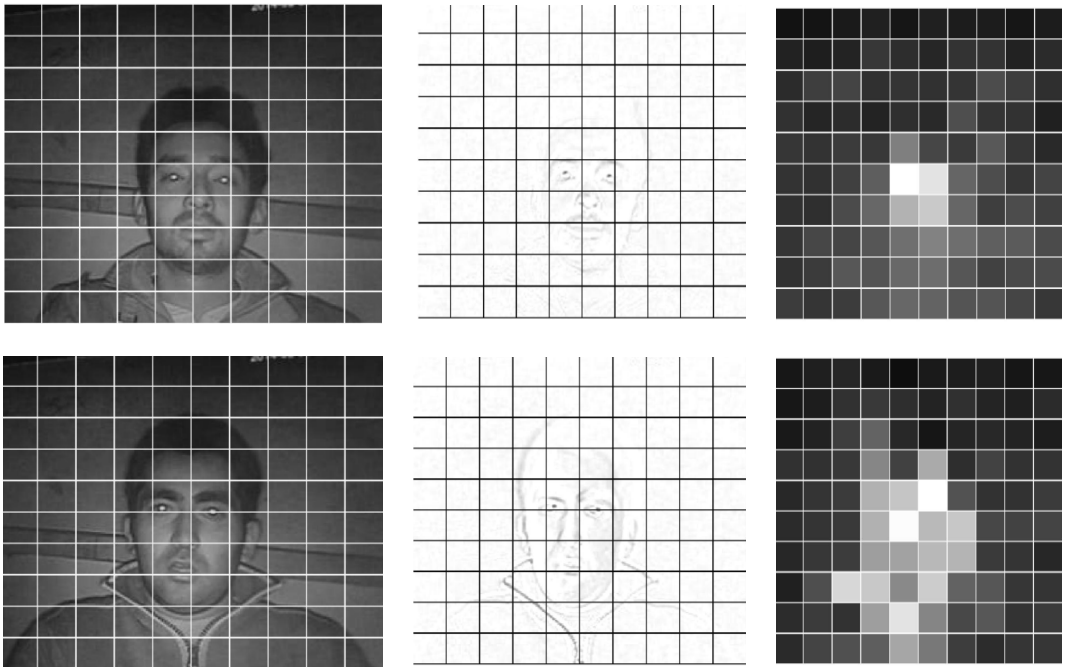

Fig. 7. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the velar consonant च (ca)

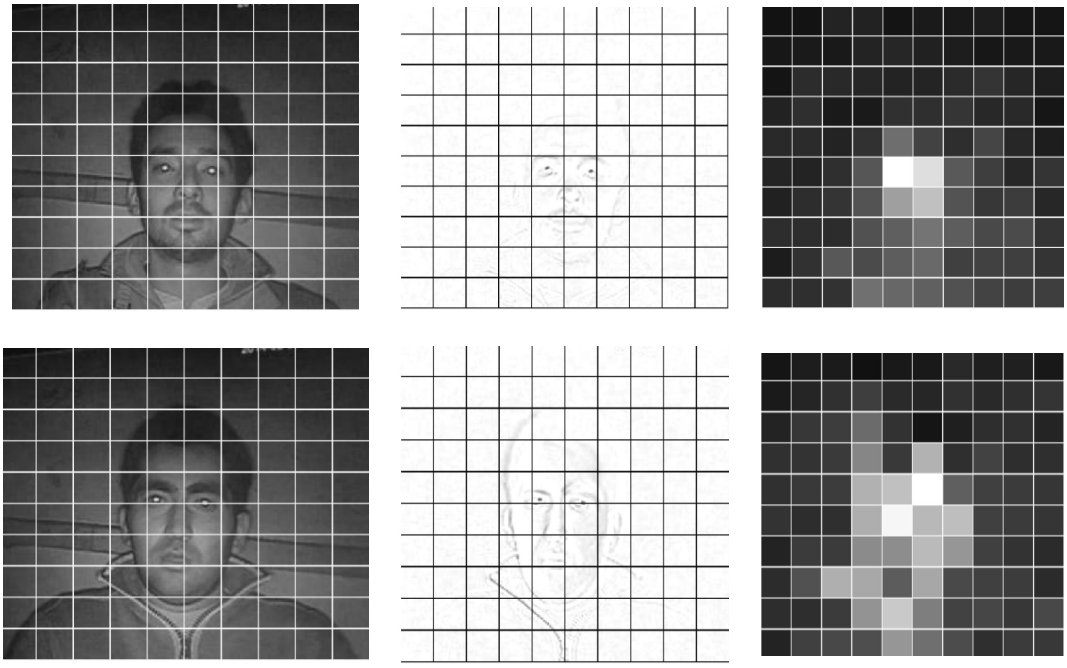

Fig. 8. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the palatal consonant छ (cha)

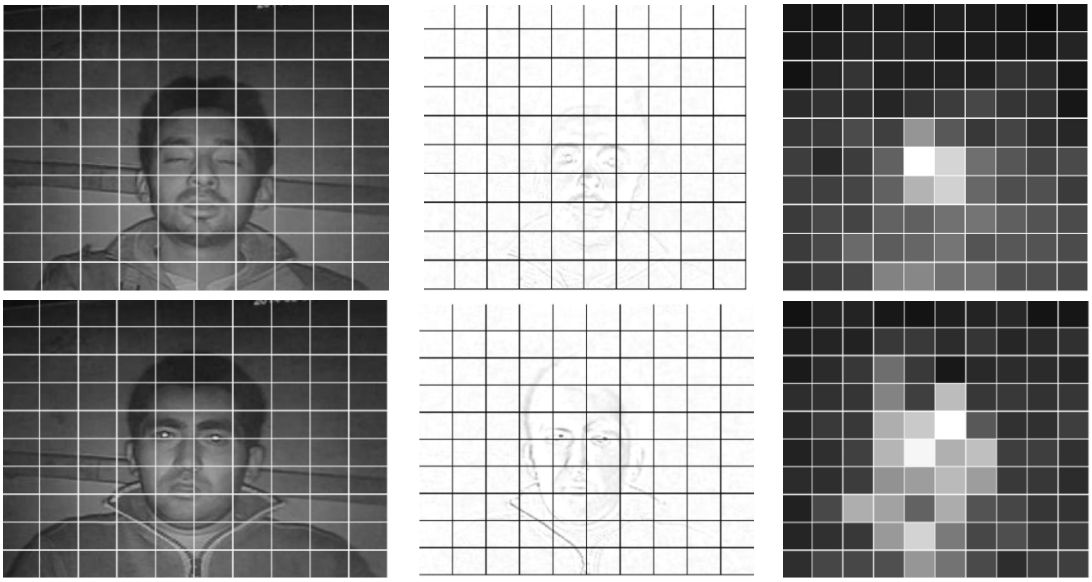

Fig. 9. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the palatal consonant ज (ja)

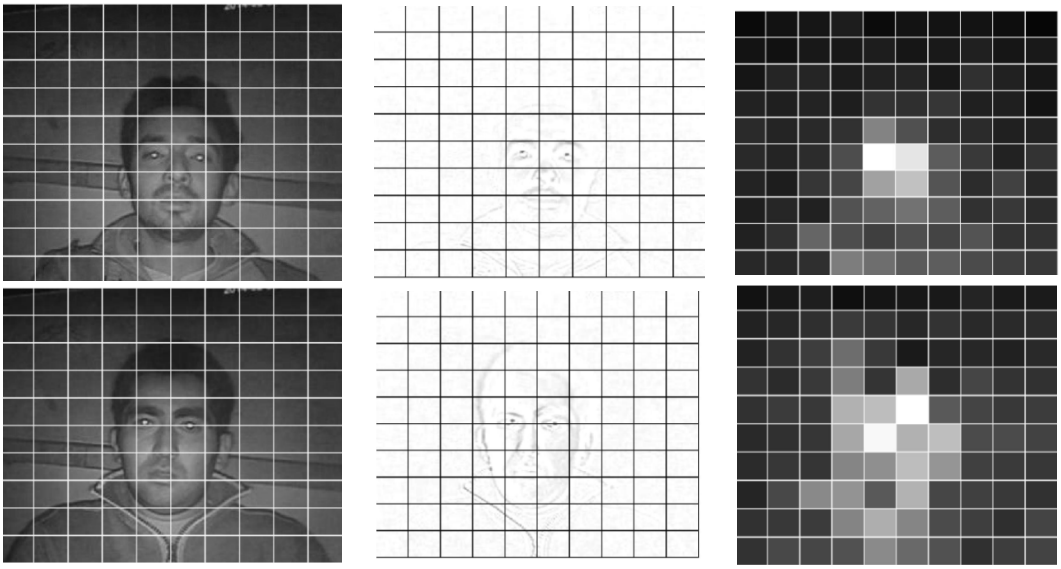

Fig. 10. IR images (first column), difference images (second column), and average spatial difference (third column) for two subjects for the palatal consonant झ (jha)

Table 2. Mean and average difference images of all velar consonants

Consonant	Subjects					Mean
	S1	S2	S3	S4	S5	
क	2.33	3.08	0.98	1.31	2.00	1.94
ख	2.80	3.04	0.84	1.33	2.69	2.14
ग	2.96	3.27	0.90	1.02	2.31	2.09
घ	2.58	3.48	1.00	1.37	2.07	2.10
Mean	2.66	3.21	0.93	1.25	2.26	2.06

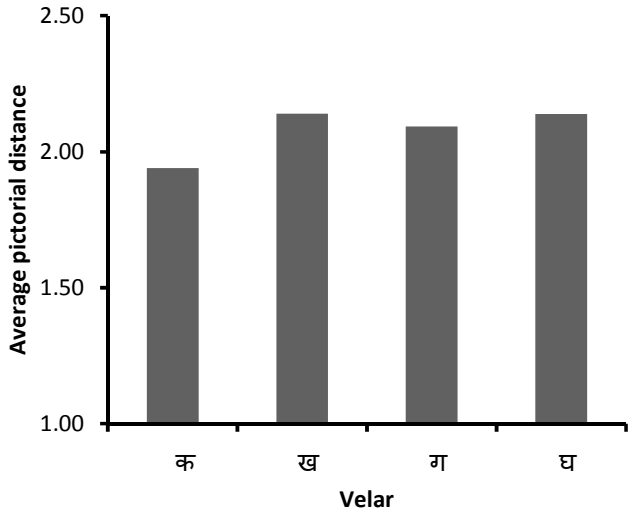

Fig. 11. Histogram of average of difference images for velar consonants

Table 3. Mean and average of difference images for all palatal consonants

Consonant	Subjects					Mean
	S1	S2	S3	S4	S5	
च	2.69	3.67	1.09	1.23	2.84	2.30
छ	2.26	3.53	1.16	1.38	2.36	2.13
ज	2.88	3.52	0.89	1.32	2.48	2.21
झ	2.30	3.60	1.27	1.30	2.58	2.22
Mean	2.53	3.58	1.10	1.31	2.56	2.23

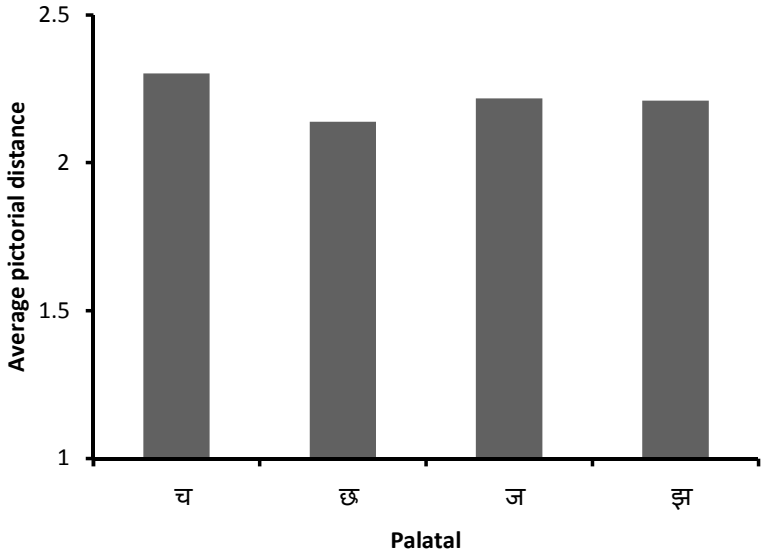

Fig. 12. Histogram for average of difference images for palatal consonants

5. CONCLUSION

Investigations were carried out to study the effect of speech on the facial expressions captured in IR images. For the recording, five male subjects were taken and their corresponding IR images at the instant of utterance of velar and palatal consonants were recorded. The difference images were obtained from the video frames corresponding to speaking and idle faces for each consonant. The maximum average of difference images was found as 2.30 for consonant (च) and minimum average as 1.94 for consonant (क). Difference images were observed as more speaker dependent as compared to speech content. Average of difference images of palatal consonants was observed relatively more as compared to that of velar consonants. These results can be later on used to determine that which consonant (velar or palatal) was uttered.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Riana H, Singh R, Singh JB, Lehana P. Effect of unvoiced consonants on EMG signal. International Journal of Advanced Research in Computer and Software Engineering. 2013;3:135.
2. Martinez A, Du S. A model of the perception of facial expressions of emotion by humans research overview and perspectives. Journal of Machine Learning Research. 2012;13:1589.
3. Gupta C, kaushal S, Singh JB, Lehana P. Facial changes during the pronunciation of dental and labial consonants in IR images. International Journal of Advanced Research in Computer Science and Software Engineering. 2014;4:570.
4. Fang H, Pathalain NM, Bargo R. Facial expression recognition in dynamic sequences and pattern recognition. 2013;47:1271.
5. Available:http://www.kidport.com/reflib/science/HumanBody/MuscularSystem/HeadFaceMuscles.htm
6. Bhattacharjee D, Seal A, Gaungly S. Comparative study of human thermal face recognition based on Haar wavelet transform and local binary pattern. Computational Intelligence and Neuroscience. 2012;2012:6.
7. Li SZ, Zhang L, Liao SC. A near infrared image based face recognition system. Institute of Automation, Chinese Academy of Sciences. 2006;38:455.
8. Yasunari Y, Sung I, Takako I. Effect of sensor fusion for recognition of emotional states using voice, face and thermal image of face. Proc. Roman. 2000;178.
9. Yoshitomi Y, Miyawaki N, Tomita S. Facial expression recognition using thermal image processing and neural network. International Workshop on Robot and Human Communication. 1997;380.
10. Kuno H, Kougaku S. IEICE; 1994.
11. Homer D, Tarnozy T. The speaking machine of Wolfgang von kempelen. The Journal of the Acoustic Society of America. 2013;22:151.
12. Marwan A. Fractal speech processing. The Press Syndicate of University of Cambridge; 1999.
13. Rajput P, Lehana P. Investigations of the distributions of phonemic durations in Hindi and Dogri. International Journal on Natural Language Computing. 2013;2:17.
14. Rabiner LR, Schafer RW. Digital processing of speech signals," Prentice-Hall Inc. Englewood Cliffs. New Jersey.
15. Furui S, Sondhi M. Advance in speech signal processig. New York. 2012;1:453.
16. Palo P. A review of articulatory speech synthesis. Department of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing. 2006;4:126.
17. Available:http://www.omniglot.com/language/articles/devanagari.html

© 2015 Gupta et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=756&id=31&aid=6987