INVESTIGATIONS OF THE DISTRIBUTIONS OF PHONEMIC DURATIONS IN HINDI AND DOGRI

Padmini Rajput and Parveen Lehana*

Dept. of Physics and Electronics, University of Jammu, Jammu, India

*pklehanajournals@gmail.com

ABSTRACT

Speech generation is one of the most important areas of research in speech signal processing which is now gaining a serious attention. Speech is a natural form of communication in all living things. Computers with the ability to understand speech and speak with a human like voice are expected to contribute to the development of more natural man-machine interface. However, in order to give those functions that are even closer to those of human beings, we must learn more about the mechanisms by which speech is produced and perceived, and develop speech information processing technologies that can generate a more natural sounding systems. The so described field of stud, also called speech synthesis and more prominently acknowledged as text-to-speech synthesis, originated in the mid eighties because of the emergence of DSP and the rapid advancement of VLSI techniques. To understand this field of speech, it is necessary to understand the basic theory of speech production. Every language has different phonetic alphabets and a different set of possible phonemes and their combinations.

For the analysis of the speech signal, we have carried out the recording of five speakers in Dogri (3 male and 5 females) and eight speakers in Hindi language (4 male and 4 female). For estimating the durational distributions, the mean of mean of ten instances of vowels of each speaker in both the languages has been calculated. Investigations have shown that the two durational distributions differ significantly with respect to mean and standard deviation. The duration of phoneme is speaker dependent. The whole investigation can be concluded with the end result that almost all the Dogri phonemes have shorter duration, in comparison to Hindi phonemes. The period in milli seconds of same phonemes when uttered in Hindi were found to be longer compared to when they were spoken by a person with Dogri as his mother tongue. There are many applications which are directly of indirectly related to the research being carried out. For instance the main application may be for transforming Dogri speech into Hindi and vice versa, and further utilizing this application, we can develop a speech aid to teach Dogri to children. The results may also be useful for synthesizing the phonemes of Dogri using the parameters of the phonemes of Hindi and for building large vocabulary speech recognition systems.

KEYWORDS

Natural Speech Production, Text to Speech System, Indian Languages, Speech Processing.

1. INTRODUCTION

Natural speech signal generation is one of the most amazing physical processes. Various articulatory movements taking place as a result of brain's motor signals consequently regulating a dynamic vocal tract system with time varying excitations, in conjunction with the pulmonic

egressive emission of air from the lungs constitute this physical mechanism. The manner of excitation and the shape of the vocal tract may be speaker and language dependent. Thus human acoustic structure is a complicated sensory organization which generates a sequence of non-stationary sound waves, interest in its composition and effectiveness stems mainly from a general interest in the field of speech synthesis [1]. Speech signal is the most inherent kind of communication which transmits wide range of information. Among them, the value of the message being uttered is of primary importance; nevertheless, secondary information like the speaker individuality also plays vital part in the oral swap over of communication [2]. Articulation is the result of brain's activity of arranging thoughts into sequence of words. The series of indistinguishable units that add up to make a sequence of words and hence a variety of languages (according to the manner and context of utterances) are termed as phonemes. The pronunciation of phonemes depends upon contextual effects, speaker characteristics and emotions [3]. Human speech is dynamic rather than static, since the articulators keep moving during articulation this fact leads to an assumption that we begin to articulate the next segment before completing the previous one that is events are all set before they occur [4]. A human can replicate the sound of an utterance by merely visualizing the mouth movements, even if he is not aware of that particular language. This exceptional attribute of speech has constrained researchers to think of speech as the fastest and efficient method of interaction between human.

Speech signal processing has many efficient and intelligent applications, like speech recognition, speaker transformation and text-to-speech (TTS) systems, which are continuously gaining a serious attention since many years; however it is not a trouble-free task and requires that the machine should have the adequate intelligence to recognize human voices [5]. TTS systems convert arbitrary text into spoken waveform; it generally employs the processing of the text followed by speech generation. The main reason behind the improvement in text-to-speech synthesizers is the requirement of a natural sounding machine [6] [7]. Research on Indian languages has been used for developing Text-To-Speech synthesis systems for only a few Indian languages like Hindi, Tamil, Kannad, Marathi, and Bangla.

The objective of this research is to investigate the distribution of phonemic durations in Hindi and Dogri subsequently; examining the pitch, amplitude, spectrograms, formants, and bandwidths of speech waveforms. Section 2 throws light on the Indian language scripts; Section 3 describes the complete mechanism of natural speech production and types of speech signal. The knowledge about the basic characteristics and production mechanism involved for speech signal generation is employed for the development machines producing most natural sounding voices. Section 4 throws light on some of the models that are developed for speech production various models are described and compare in this section. In section 5 illustrates to a larger extent the methodology employed for the investigations. The last two sections; section 6 and section 7 present the results and conclusions of the analysis.

2. INDIAN LANGUAGE SCRIPTS

There is a wide-ranging linguistic homogeneity of Indian languages at the micro-level, however if the nation is regarded as one entity India is a linguistically diverse country with 22 official languages [8]. Language technologies play a vital role in multilingual society like India which has about 1652 dialects/native tongues. Languages of India belong to numeral racial groups like Negroids, Austrics, Mongoloids, Caucasoid Dravidians and Caucasoid Aryans but it is the last

two families that preside over the country. According to 2001 census, Dravidian languages are spoken by 24% of Indians and 74% of Indians speak Indo Aryan languages. The Dravidian languages are the languages of south India whereas Sanskrit based Aryan languages are from the north yet both have acquired their scripts from a common foundation. Marathi, Tamil and Bengali are extensively spread in outer areas, whereas, Assamese and Kashmiri have flourished in linguistic seclusion. Punjabi and Sindhi were dwarfed because of historic-geographical factors. Sindhi does not have any core in India. The other languages in India belong to Austro- Asiatic and Tibeto- Burman [9]. Hindi and Urdu are nationwide extended; still their use in the southern plain domain is insignificant. There are some peculiar characteristics in each language. Indian languages have a more sophisticated notation of a character unit or akshara that forms the fundamental linguistic unit. There are 10-12 major scripts in India. Indian languages have been derived from the prehistoric Brahmi script. Extensive use of reduplication is a particular characteristic of Indian Languages. Hindi and Dogri are one of the 22 official languages of India. Hindi is an Indo- European language of the indo Aryan subfamily. Hindi is one of the prevalent languages of India after English and Mandarin, spoken in the major regions like Himachal Pradesh, Uttar Pradesh, Rajasthan, Bihar, Haryana and Chattisgarh [8]. A distinctive character in Indian languages scripts are near to syllable and can be characteristically of the form: C, V, CV, VCV, CVC, and CCV, where C is for a consonant and V is for a vowel [10].Typical Hindi, casually spoken by people is called Manak Hindi, High Hindi, Nagari Hindi and Literary Hindi; it is derived from the Khariboli dialect of Delhi and belongs to the Devnagri script while Dogri has its own script namely Doger. Hindi and Dogri are closely related languages having their roots in Sanskrit, and belonging to the same subgroup of Indo-European family. Indian census (2001) states that about 258 million people in India are Hindi speakers, while the number of Dogri speakers is far less than that of Hindi speakers and counts to about 5 million. Mostly people of north India are familiar with Dogri; these regions include Jammu, parts of Kashmir, Himachal, and northern Punjab. This language is essentially written in Takri Script, strongly linked to Sharada script that comprises Kashmiri and Gurmukhi script for Punjabi. Dogri speakers are often called Dogras. Dogri was given the honor of national language on 22nd December, 2003. Dogri and Hindi are rather alike with diverse phonologies, and mode of pronunciation [8]. Text-To-Speech synthesis systems have also been developed for few Indian languages like Hindi, Tamil, Kannad, Marathi, and Bangla [11].

3. NATURAL SPEECH PRODUCTION MECHANISM

Speech generation is the most intrinsic phenomenon which starts with the creation a message in brain and ends up with the production of an utterance from the oral cavity. The production of speech sounds requires the integration of various information sources in order to generate the complex patterns of muscle activations required for fluency. The natural phenomenon of speech has always been an object of both general curiosity and scientific inquisition. We use speech more or less unconsciously but hardly a few people have the idea pertaining to this innate blessing. A study pertaining to human evolution revealed that human skull base, which was formerly located in the upper position gradually inclined simultaneously with the descent of the larynx. This research led to the conclusion that the time of speech acquisition and the origin of speech can be estimated by the inclination angle of the base of the skull. This theory is supported by taking the unique case of the development of human children that the enhancement in the inclination of the skull base is very much connected to the maturation of the speech organs [12]. Mechanism of speech in human beings has developed over several years yielding a vocal structure that is proficient in terms of spoken swap over of communication [13].

The mechanism of natural speech encompasses four processes: Language processing, in which the content of the utterance is converted into phonemic symbols in the brain's language centre; generation of motor commands in the brain's motor center to the vocal organs; initiation of articulatory movements for the production of speech by the vocal organs based on these motor commands; and the emission of air sent from the lungs resulting in a speech signal that we hear [14]. This whole phenomenon can be visualized as a chain mechanism passing through various levels like linguistic level, physiological level, acoustic level, and at the last linguistic level. The vocal folds change the signal originating from any source or from the vocal chord itself into intelligible speech [15] [16].

Research has shown that vocal folds in case of males are usually longer than that in females, causing a lower pitch and a deeper voice. The male vocal folds are between 17.5 mm and 25 mm (approx 0.75 to 1.0") in length. This difference in the size of vocal chords causes a difference in vocal pitch. The female vocal folds are between 12.5 mm and 17.55 (approx0.5" to 0.75") in length. There is a gap between the vocal folds which is called glottis, and the production from this place is often called glottis source. The air passing through the vocal folds, when forming a narrow opening, makes them to vibrate, producing a periodic sound. The rate of vibration of the vocal folds is called as the fundamental frequency F0. From the point of view of F0 larynx is the most important vocal organ. Fundamental frequency is between 80 to 250 Hz for a male speakers since a male can vibrate his vocal folds in between 80 to 250 times per second in comparison to this a female has F0 that is between 120 to 400 Hz [17]. The term pitch refers to the rate of vibration that is perceived by the listener. Figure 1 depicts various articulators employed in natural speech production. Above the larynx is the human pharynx situated behind the mouth, which divides into two parts one entering the mouth region and the other entering inside the nasal region [18].

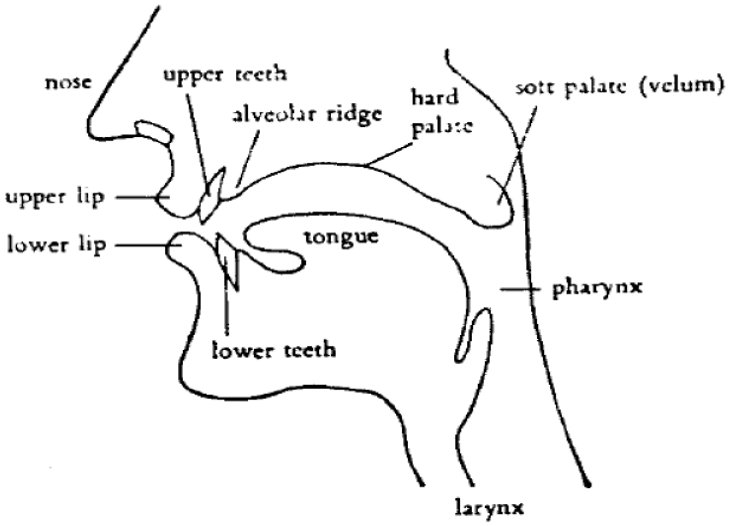

Figure 1. Natural speech production process [18].

Speech sound which is originated when the middle part of the tongue called dorsum touches the soft plate called velar consonant. While speaking, the velum is raised so as to block the air from flowing towards the nasal region, while letting it flow out of the mouth. Like the soft plate there is one more plate next to it, the hard plate above the tongue, often termed as the roof of the mouth.

Between the teeth and the hard plate is the alveolar ridge, sound made with the tongue touching this region is named alveolar. The motion of the vocal folds decides whether the sound produced will be voiced or unvoiced [15].

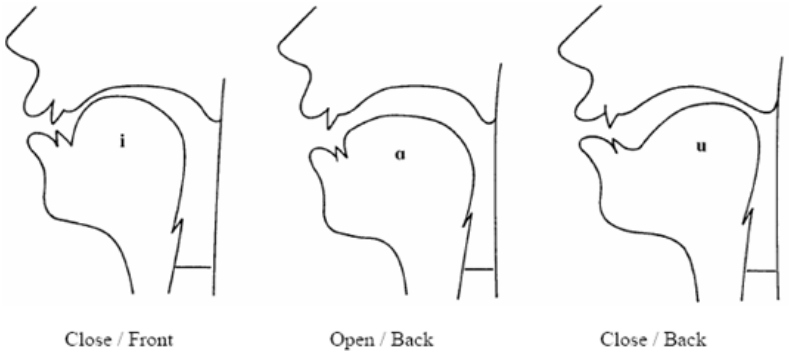

Figure 2. Shapes of mouth while articulation of vowels [15].

The voiced sounds are the result of the vibrations (depending upon the mass and tension of the cords) of the vocal chords, caused by the rapid opening and closing of the vocal folds. All vowels and some of the consonants are voiced sounds. There is also a category of sound called unvoiced sound in which there is no periodic component in the resultant speech that is a result of the condition in which the sound is produced without the vibration of the glottis. Different shapes made by lips produce different sounds. In addition to these articulators there are three more things to be taken into consideration, the larynx, the lower jaws and the nasal cavity, these three elements cannot be termed as articulators since they cannot make contact with other articulators, but they, no doubt have a very important role in natural sound production process. Various anatomical articulators work in a synchronized manner, these vocal organs collectively make an effort for sound generation, like the slight variation in the movement of tongue, can directly affect the variation in the sound produced [19].

4. MODELS FOR SPEECH SYNTHESIS

Brain process underlying the formation of spoken language involves auditor encoding, prosodic analysis, and linguistic evaluation [20]. Artificial speech production can be produced by two ways, model based method and waveform based method [21]. For waveform based methods, a set of pre-recorded statements from the database for the desired speech. Model based methods comprise source filter model (synthesis by rule), direction into velocity of articulators model (DIVA), and Lavelt' model. Model based techniques make use of the natural model of speech generation in human beings. Unit selection, concatenative, linear predictive coding (LPC), and formant synthesis, and harmonic plus noise model (HNM) techniques are used in waveform based methods [16] [22] [23].

Source-Filter synthesis also called formant synthesis; makes use of the spectral shaping of driving excitation. This model describes the speech signal as an excitation from the vocal tract by a cyclic source. Fant's and Ungeheuer's theories built two assumptions about the configuration and procedure concerned in the speech production. First, both the theories pay no attention to the

concept that the sub-glottal cavity is acoustically coupled with supra-glottal cavities for the duration of articulated phonation with the measure of coupling that varies during each cycle of the vocal chord vibration, ranging from a complete de-coupling for the duration of glottal closure, and a gradual raise and decline during the open glottis state. Second, the theories equally assume that the contributions of vocal source and the vocal tract filter can be separated for each other, with the filter having no back-coupling effect on the source. Source-filter model speech synthesizers employing LPC also take account of these assumptions [18].

DIVA model uses the sensory and motor biochemical activities of the brain, and works on the hypothesis for generating high quality acoustic speech, based on what is perceived to drive what is produced. The technique utilizes the sensory activities since for the period of speech production a great section of the cerebral cortex appears into action [16], and motor biochemical activities of the brain, since during speech production a large portion of the cerebral cortex appears into action [44]. Out of all the models created for finest synthetic speech the most extensively used model for L2 speech production is Lavelt's model originally developed for monolingual communication, based on encoding the thought into words. Levelt's model is the most current model. The Direction into Velocity of Articulators Model (DIVA) works on this assumption for generating exceptional acoustic speech [21].

Present-day model used for speech production is the Levelt's model. Lavelt predicted speech generation as a modular approach, illustrating it as an organization of models in the systems which are self-governing. He approximated two most important mechanisms: rhetorical/semantic/syntactic system and the phonological/phonetic system [20].

Under concatenative speech synthesis we have unit selection and diphone synthesis. Unit selection synthesis utilizes richer variety of speech which simply cuts out speech and rearranges it. In order to synthesize speech which is more varied in voice characteristics it makes use of a large database of pre-recorded speech, which increases the cost and difficulty. The diphone synthesis makes use of the notion that a little portion of the acoustic signal varies to minor amount, and is also less subjective by the phonetic context than others. The quality of the sound positions somewhat in between that obtained form concatenative and formant synthesis techniques but suffers from glitches. Concatenative synthesis approaches are extensively used by many systems, in which stored speech unit waveforms can be blended together to generate new speech. This is the simplest technique providing high quality and naturalness. [16] [19].

Linear predictive coding (LPC) is amongst the most powerful and latest synthesis techniques operated in signal processing for the expression of the spectral envelope of speech in compact form taking into concern the information required in the linear predictive model. It's an important technique for the accurate, economical measurement of speech parameters like pitch, formants spectra, and vocal tract area functions and for the representation of speech for low rate transmission or storage [15]. The formant synthesis also called rule-based synthesis is based on source-filter model. Formants F0, F1 and F3 are required for the production of a bit good sounding voice while the computation of up to first five formants is necessary for high quality intelligible sound production. Formant synthesis is based on source-filter model and so it has a source for speech signal and a filter for representing the resonance of the vocal tract, a two pole model resonator represents the formant frequencies and bandwidths. F0, F1 and F3 are required for the construction of a bit good sounding voice though the calculation of up to first five formants is necessary for finest precision of sound [24]. Harmonic plus noise model (HNM) has

reduced database and provides a direct technique for smoothing discontinuities of acoustic units around concatenation points which makes this method quite efficient. HNM is a pitch-synchronous system [21], unlike TD-PSOLA and other concatenative approaches, hence eliminating the problem of synchronization of speech frames and hence shows the capabilities of providing high-quality prosodic modifications without buzziness compared to other methods. HNM framework is also used in a low bit rate speech coder to increase naturalness. Research has shown that all vowels and syllables can be produced with a better quality syllables by the implementation of HNM. Results obtained from many speech signals including both male and female voices are quite satisfactory with respect to the background noise and inaccuracies in the pitch.

The reason behind the advancement of the new techniques and acceptance of improved models for speech generation is the result of need for more improved, human like utterances by a machine. Previous models produced mechanically sounding voices that are irritating to the listener, progression in models and systems improved the quality of synthetic speech greater than before, the quality of the speech generated depends on the type of model employed in the speech synthesizer however still there is not any perfect model that can generate a voice exactly like that of a human, yet the models that are discussed above have contributed their part in the generation of more natural sounding voices.

5. METHODOLOGY

The complete procedure for the analysis of the speech signal has been divided into two parts. For the analysis process the recording of 13 speakers in Dogri (3 male and 2 females) and Hindi (4 male and 4 female) language were taken. Speakers of different age group, from different regions of Jammu have been taken. The data for recording comprises 110 TIFR (Tata Institute of Fundamental Research) sentences in Hindi, and a sequence of words having all the 35 consonants in between, named VCV (vowel consonant vowel). The same data was used for recording in Dogri. After recordings in both languages the speech was segmented manually into vowels. For estimating the means and standard deviations of each vowel, ten instances were selected from similar contexts. Also the variation in duration of all the, vowels is calculated by taking the percentage of the value obtained by dividing the standard deviation with the mean of the ten vowels

6. RESULTS AND CONCLUSIONS

The averaged mean and standard deviations and the variation in duration of all the ten vowels (v1: /ah/, v2: /aa/, v3: /ih/, v4: /iy/, v5: /ey/, v6: /ae/ v7: /uh/, v8: /uw/, v9: /uh/ and v10: /ao/) for each speaker in both languages are shown in the Tables 1 and Table 2 and plotted in Fig. 3 to Fig. 5 as histograms for ease of visual interpretation. For the investigation of vowels, mean of ten instances was calculated for each speaker, and this resulting value for all the vowels was plotted. Figure 6.1 shows the average duration of vowels uttered by the first speaker sp1. The plot shows that v3 and v7 is of minimum duration of 40 ms, while v6 has the maximum duration of about 120ms, v1 and v9 and v2 and v10 have almost the same length. The height of the average bars shows the length of the utterances for a vowel. Similarly we can concude the results from the other histograms and the whole investigation can be concluded with the end result that most of the Dogri vowels have shorter duration, in comparison to Hindi phonemes. The period in milli

seconds of same vowel when spoken in Hindi was found to be of longer duration compared to when it was spoken by a person with Dogri as his mother tongue. The language we speak depends upon the region we belong to. Only Dogri is spoken in most parts of Jammu but there happens to be a change in tone, pronunciation, and utterances of phonemes with the change in region which may be the reason for different duration of the same phonemes and hence the results may differ for other regions of Jammu. There are many applications which are directly of indirectly related to the research carried out in this thesis. For instance the main application may be for transforming Dogri speech into Hindi and vice versa, and further utilizing this application, we can develop a speech aid to teach Dogri to children.

Table 1. Mean of mean of the duration of all vowels spoken by all the speakers of Dogri and Hindi language

S No.	**v1**	**v2**	**v3**	**v4**
sp1	42.8	83.04	44.88	64.83
sp2	44.81	17.24	54.59	83.06
sp3	45.32	83.12	49.81	86.88
sp4	42.54	55.58	45.49	76.73
sp5	40.07	74.47	51.58	81.51
Mean(ms)	43.11	62.69	49.27	78.60
Sd(ms)	2.09	27.78	4.11	8.52
Variation(ms)	4.84	44.31	8.34	10.83

a) Mean of mean of the duration of all vowels spoken by all the speakers of Dogri language

S No.	**v1**	**v2**	**v3**	**v4**
sp1	67.62	85.27	41.31	88.85
sp2	43.53	77.64	48.59	97.65
sp3	45.78	91.62	62.65	86.64
sp4	48.79	87.79	56.82	89.16
sp5	52.93	95.31	44.95	74.5
sp6	47.77	70.76	39.68	56.24
sp7	34.09	60.22	45.61	61.3
sp8	40.88	77.66	46.72	96.8
Mean(ms)	47.67	80.78	48.29	81.39
Sd(ms)	9.84	11.59	7.77	15.72
Variation(ms)	20.65	14.35	16.09	19.31

b) Mean of mean of the duration of all vowels spoken by all the speakers of Hindi language.

Table 2. Mean of mean of the duration of all vowels spoken by all the speakers of Dogri and Hindi language

S No.	**v5**	**v6**	**v7**	**v8**
sp1	54.53	19.91	65.77	90.43
sp2	74.59	111.06	82.47	84.75
sp3	73.98	99.91	68.07	93.68
sp4	65.32	71.57	59.69	80.84
sp5	57.13	86.46	61.74	86.58
Mean(ms)	65.11	77.782	67.548	87.26
Sd(ms)	9.27	35.56	8.96	4.98
Variation(ms)	14.24	45.72	13.27	5.70

a) Mean of mean of the duration of all vowels spoken by all the speakers of Dogri language.

S No.	**v5**	**v6**	**v7**	**v8**
sp1	111.32	126.8	41.85	116.15
sp2	79.62	109.63	62.58	93.72
sp3	96.37	123.08	57.79	110.82
sp4	101.47	108.54	48.9	72.18
sp5	74.91	97.13	60.77	106.7
sp6	81.91	82.6	47.07	93.96
sp7	64.71	73.83	47.65	98.35
sp8	81.43	101.49	56.76	83.74
Mean(ms)	86.47	102.89	52.92	96.95
Sd(ms)	15.31	18.31	7.50	14.47
Variation(ms)	17.71	17.80	14.18	14.92

b) Mean of mean of the duration of all vowels spoken by all the speakers of Hindi language.

Table 3 Mean of mean of the duration of all vowels spoken by all the speakers of Dogri and Hindi language.

S No.	**v9**	**v10**
sp1	69.12	83.75
sp2	80.19	92.59
sp3	72.12	98.07
sp4	68.37	87.33
sp5	64.35	78.27
Mean(ms)	70.83	88.00
Sd(ms)	5.92	7.67
Variation(ms)	8.36	8.72

a) Mean of mean of the duration of all vowels spoken by all the speakers of Dogri language.

S No.	**v9**	**v10**
sp1	67.62	103.45
sp2	93.74	93.45
sp3	88.65	96.06
sp4	96.44	96.45
sp5	85.5	97.49
sp6	65.56	78.01
sp7	67.59	81.18
sp8	86.41	98.94
Mean(ms)	81.44	93.13
Sd(ms)	12.56	8.87
Variation(ms)	15.42	9.52

b) Mean of mean of the duration of all vowels spoken by all the speakers of Hindi language.

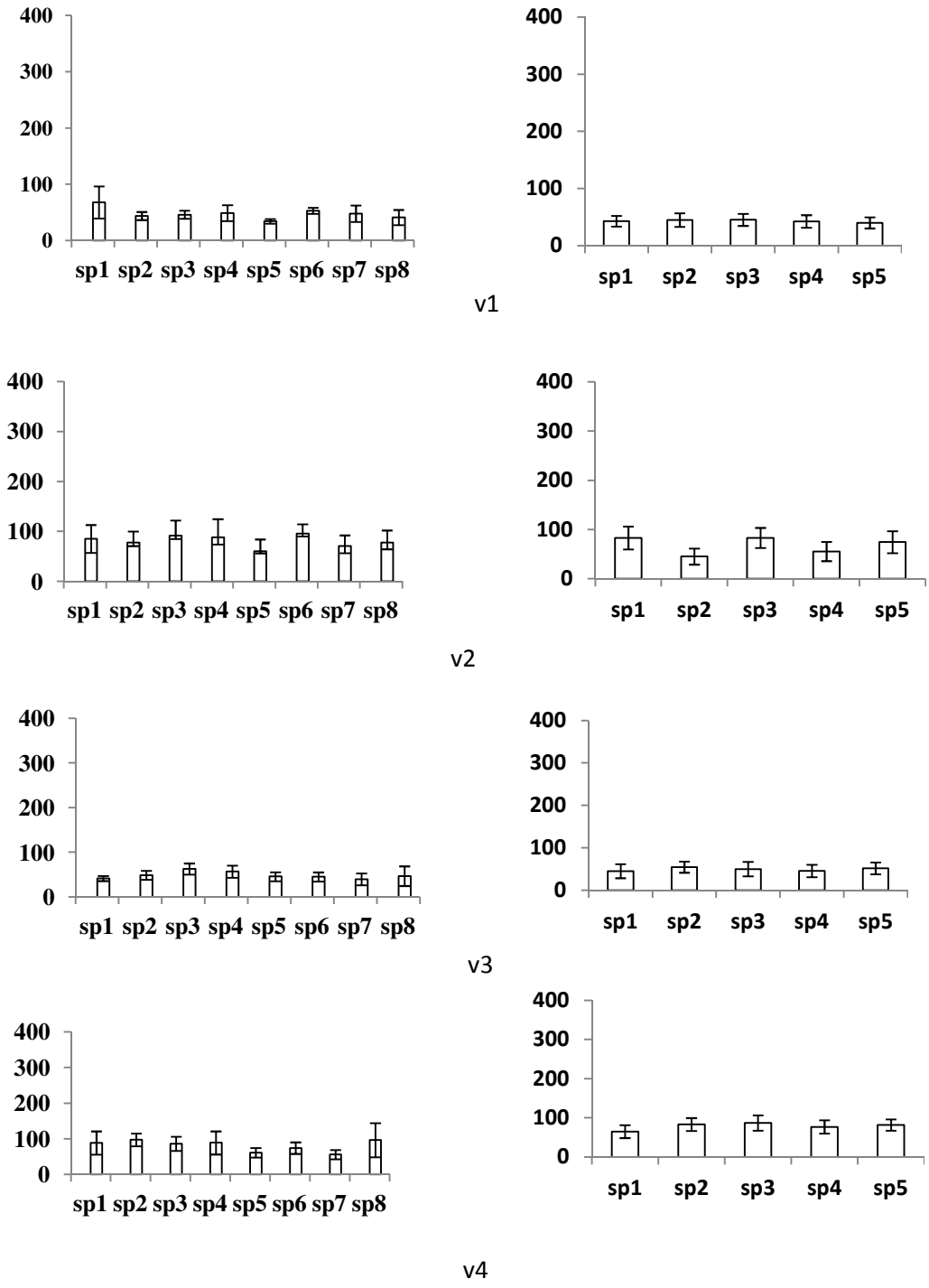

Figure 3. Comparisons of mean and standard deviation for duration of first vowels of Hindi and Dogri language articulated by respective speakers.

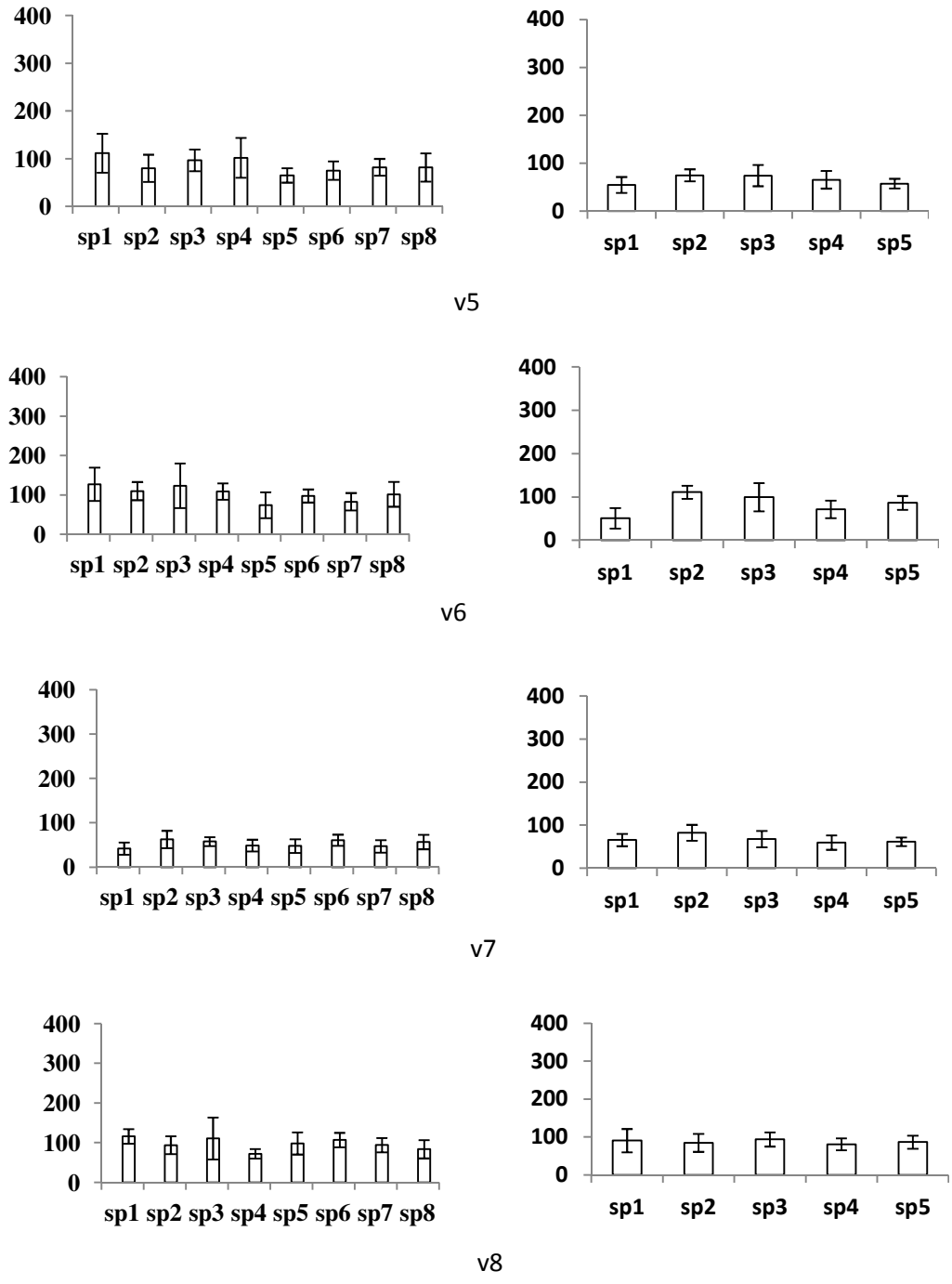

Figure 4. Comparisons of mean and standard deviation for duration ofvowels of Hindi and Dogri language articulated by respective speakers.

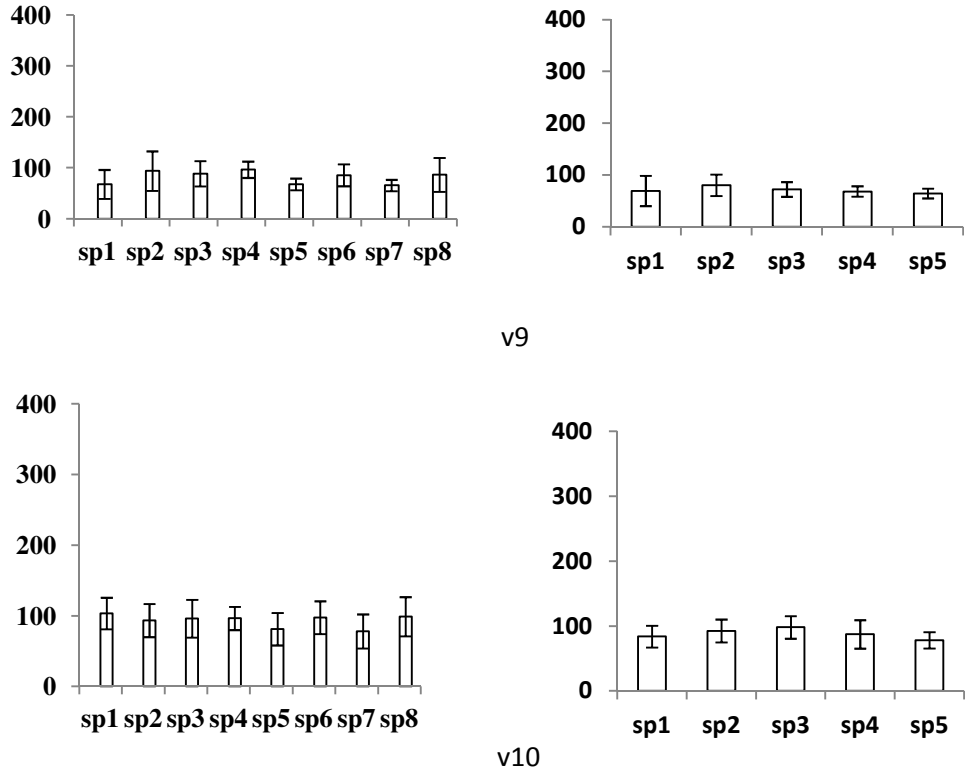

Figure 5. Comparisons of mean and standard deviation for duration ofvowels of Hindi and Dogri language articulated by respective speakers.

REFERENCES

[1] Yin Hui, Hohmann Volker, & Nadeu Climent (2011) "Acoustic feature for speech recognition based on Gammatone filterbank and instantaneous frequency", ELSEVIER Speech Communication 53, pp 707-715.

[2] Stylianou Yannis, Cappe Oliver, & Moulines Eric (1998) "Continuous probabilistic transform for voice conversion", IEEE Trans. Speech and Audio Processing, Vol. 6, No. 2, pp 131-142.

[3] Chaudhari U V, Navaratil J & Maes S H (2003) "Multigrained modeling with pattern specific maximum likelihood transformations for text-independent speaker recognition", IEEE Trans. on Speech and Audio Processing, Vol. 11, No. 1, pp 61-69.

[4] Alani Ahmed, & Deriche Mohamed (1999) "A novel approach to speech segmentation using the wavelet transform", Fifth International Symposiumon Signal Processing and its Applications (ISPRA), pp 127-130.

[5] Moataz El Ayadi, Mohamed S. Kamel, & Fakhri Karray (2011) "Survey on speech emotion recognition: Features, classification schemes, and database", ELSEVIER Pattern Recognition, pp 572-587.

[6] Park J, Delhi F, Gales M J F, Tomalin M & Woodland P C (2011) "The efficient incorporation of MLP features into automatic speech recognition system", ELSEVIER, Computer Speech and Languages, pp 519-534.

[7] Lu X, Unoki M & Nakamura S (2011) "Sub-band temporal modulation envelopes and their normalization for automatic speech recognition in reverberant environments", ELSVIER Computer Speech and Language, pp 571-584.

[8] Dubey Preeti, Pathania Shashi & Devanand (2011) "Comparative study of Hindi and Dogri languages with regard to machine translation," language in India, Vol. 11, pp 298-309 .

[9] Dutt Ashok K, Khan Chandrakanta C, & Sangwan Chanralekha (1985) "Spatial pattren of languages in India: A culture-historical analysis", International Gournal of Geography, GeoJournal, Vol 10, pp 51- 74.

[10] Kishore P.S, Kumar Rohit, & Sangal Rajeev (2002) "A data-driven synthesis approach for Indian languages using syllable as basic unit", in Conference of Natural Language Processing.

[11] Rao Sreenivasa K. (2011) "Application of prosody models for developing speech systems in Indian languages", International Journal of Speech Technology" Vol. 14, No. 1, pp 19-33.

[12] Flanagan J. L (1972) " Speech Analysis, Synthetic and Perception", Springer, New York.

[13] Dudley Homer (1940) "The carrier nature of speech", The Bell Syst. Tech. Journal. Vol. 9, No.4, pp 495- 515.

[14] Dudley Homer, Tarnozy T.H. (1950) "The speaking machine of Wolfgang von kempelen", The Joural of the Acoustic Society of America, Vol. 22, No. 2, pp151-166.

[15] Al-Akaidi Marwan, (2004) "Fractal Speech Processing", The Press Syndicate of University of Cambridge, pp 3-4.

[16] Denes P. & Pinson E. (1963) "The Speech Chain", Bell Telephone labs, Murray Hill, New Jersey.

[17] Rabiner L R & Schafer R W (1978) " Digital processing of speech signals," Prentice-Hall Inc., Englewood Cliffs. New Jersey

[18] Furui S. & Sondhi M.M. (1992) "Advance in speech Signal Processig", Marcel Dekker, New York.

[19] Gold Ben & Morgan N. (2000) "Speech and Audio Signal Processing", Willey, New York.

[20] Herrmann C S, Friederici A D, Oertel U, Maess B, Hahne A & Alter K (2003) " The brain generates its own sentence melody: A Gestalt phenomenon in speech perception," ELSEVIER Brain and Language Vol. 85, pp 396–401.

[21] Honda Masaaki (2003) "Human speech production mechanisms," NIT Technical Review, Vol. 1, No. 2, pp 24-26.

[22] Taylor Paul (2009) "Text to speech synthesis," Cambridge University Press, pp 147-1492.

[23] Desai S, Yegnanarayana B & K Prahallad (2003) "A framework for cross-lingual voice conversion using artificial neural networks," in Proc.of International Conference on Natural Language Processing (ICON).

[24] Herrmann C S, Friederici A D, Oertel U, Maess B, Hahne A & Alter K (2003) " The brain generates its own sentence melody: A Gestalt phenomenon in speech perception," ELSEVIER Brain and Language Vol. 85, pp 396–401.

[25] Alam Firoj, Nath Promila Kanti, & Khan Mumit (2007) "Text To Speech for Bangla Language using Festival", BRAC University Institutional Repository, pp 128-133.