Available online at www.elixirpublishers.com (Elixir International Journal)

Electronics Engineering

Elixir Elec. Engg. 81 (2015) 31950-31956

ISSN: 2229-712X

Investigation of the Quality of Speech with Respect to Glottal Excitation Modification in Hindi and Dogri Languages

Sonika Mahajan[1], Rajesh Mehra[1] and Parveen K. Lehana[2]

[1]Department of Electronics & Communication Engineering, NITTTR Chandigarh, India.

[2]Department of Electronics University of Jammu, Jammu, India.

ARTICLE INFO

Article history:

Received: 18 March 2015;

Received in revised form:

22 April 2015;

Accepted: 22 April 2015;

Keywords

Glottal source excitation,

LPC,

Glottal flow,

Formants,

Phonemic symbols.

ABSTRACT

Speech is an efficient mode of communication among human beings. The shape of glottal excitation may be speaker and language dependent. The objective of this paper is to investigate the quality of speech with respect to glottal excitation modification in Hindi and Dogri languages. For this, recordings of six speakers (3 males and 3 females) were carried out in Dogri and Hindi languages. Cardinal vowels (/a/, /i/, /u/) were extracted from recordings of each speaker. Investigations were carried out by modifying the glottal excitation component of speech and this modification is obtained by adding noise which is a random signal to the glottal excitation component. The analysis of the results showed that quality and intelligibity of speech changes with the modification of the glottal source component in such a manner that identity of a speaker is fully lost and the clarity is degraded though not fully lost. Further, it is perceived that with the modification clarity in male speakers is more degraded than female speakers.

Introduction

Speech is the prime mode of communication among human beings. Speech technologies are commercially available for an unlimited but interesting range of tasks [1]. Speech provides information about speaker identity, accent, emotion and state of health of speaker. The process of speech production begins with exhaling air from the lungs. Without the subsequent modulations, this exhaled air will sound like a random noise with no information. The information is first modulated onto the passing air by the frequency of closing and opening of the glottal folds. The output of the vocal fold is the glottal excitation signal to the vocal tract which is further shaped by the resonances of the vocal tract and the effects of the nasal cavities, teeth and lips. Process of speech in human beings has developed over several years yielding a vocal structure that is skilled in speech communication [2].The mechanism of natural speech involves four processes : Language processing, in which the content of the utterance is converted into phonemic symbols in the brain's language center; generation of motor commands in the brain's motor centre, to the vocal organs; initiation of articulatory movements for the production of speech by the vocal organs, based on the motor commands; and the emission of air sent from the lungs resulting in a speech signal that we hear [3].This whole phenomenon can be visualized as a chain mechanism passing through various levels like linguistic level, physiological level and acoustic level The vocal folds change the signal originating from any source or from vocal cords [4][5].Research has shown that glottal folds in case of males are usually longer than in females causing a lower pitch and a deeper voice. The male glottal folds are between 17.5 mm and 25 mm in length. This difference in the size of vocal cords causes difference in vocal pitch. The female vocal folds are between 12.5 mm and 17.55 mm in length. There is a gap between the vocal folds which is called glottis, and the production from this place is often called glottal source. Vocal folds vibrate producing a periodic sound when the air passes through it. The rate of vibration of the glottal folds is known as the fundamental frequency F0. Larynx is the most important vocal organ from F0 point of view. Fundamental frequency F0 is between 80 to 250 Hz for male speakers because a male can vibrate his vocal folds in between 80 to 250 times per second in comparison to 120 to 400 Hz in females [6]. The mechanism of natural speech also involves language processing. Hindi is one of the prevalent languages in India after English and Mandarin. Hindi belongs to Devnagri script. Another similar language is Dogri, which is an Indo-Aryan language spoken by about five million people in India and Pakistan mainly in the Jammu region of Jammu and Kashmir. Dogri has its own script named as Doger. Number of Dogri speakers are far less than Hindi speakers. Hindi and Dogri are closely related languages having their roots in Sanskrit and belongs to the same subgroup of Indo –European family

Glottal Source Excitation

Glottal source estimation aims at isolating the vocal folds vibrations showing large variations in the movement of the vocal folds from one individual to another. The vocal folds may close completely for some speakers, while for others, the vocal folds may never reach full closure. The manner and frequency in which the vocal folds close, also varies from speaker to speaker. Difference in vibration of vocal folds correspond to difference in time varying area of the opening between the folds, known as the glottis, and therefore in volume velocity air flows through the glottis. The air flow may be smooth, when the folds never close completely, corresponding to a "soft" voice, or discontinuous, when they close rapidly, resulting into "hard" voice. The flow at the glottis may be turbulent, when air passes near a small portion of the folds that remains partly open. Turbulence at the glottis is referred to as aspiration which, when occurring during vocal cord vibration, can result into a "breathy" voice.

Tele:

E-mail addresses: sonikashelly@gmail.com

Table I (a) Mean (M) and standard deviation (SD) for Experiment I in terms of identity

Vowel	Speaker	M	SD
/a/	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/i/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/u/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0

Table I. (b) Mean (M) and standard deviation (SD) for Experiment I in terms of clarity

Vowel	Speaker	M	SD
/a/	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/i/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/u/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0

Table II. (a) Mean (M) and Standard Deviation (SD) for Experiment II in terms of identity

Vowel	Speaker	M	SD
/a/	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/i/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/u/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0

Table.II. (b) Mean (M) and standard deviation (SD) for experiment II in terms of clarity

Vowel	Speaker	M	SD
/a/	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/i/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0
/u/			
	SP1 (F)	5	0
	SP2 (F)	5	0
	SP3 (F)	5	0
	SP4 (M)	5	0
	SP5 (M)	5	0
	SP6 (M)	5	0

Table III. (a) Mean (M) and Standard Deviation (SD) for Experiment II in terms of identity

Vowel	Speaker	M	SD
/a/	SP1 (F)	1	1.414214
	SP2 (F)	1	1.414214
	SP3 (F)	1	1.414214
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214
/i/			
	SP1 (F)	1	1.414214
	SP2 (F)	1	1.414214
	SP3 (F)	1	1.414214
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214
/u/			
	SP1 (F)	1	1.414214
	SP2 (F)	1	1.414214
	SP3 (F)	1	1.414214
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214

Table.III. (b) Mean (M) and standard deviation (SD) for experiment II in terms of clarity

Vowel	Speaker	M	SD
/a/	SP1 (F)	1.666667	1.728132
	SP2 (F)	1.888883	1.544224
	SP3 (F)	1.166667	1.771196
	SP4 (M)	1.333333	1.839261
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214
/i/			
	SP1 (F)	2	1.765085
	SP2 (F)	2.333333	1.510329
	SP3 (F)	2.666667	1.404695
	SP4 (M)	1	1.414214
	SP5 (M)	1.1666667	1.771196
	SP6 (M)	1	1.414214
/u/			
	SP1 (F)	1.833333	1.544224
	SP2 (F)	1.333333	1.839261
	SP3 (F)	1.666667	1.728132
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214

Table IV. (a) Mean (M) and Standard Deviation (SD) for Experiment II in terms of identity

Vowel	Speaker	M	SD
/a/	SP1 (F)	1.166667	1.771196
	SP2 (F)	1	1.414214
	SP3 (F)	1	1.414214
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214
/i/			
	SP1 (F)	1.166667	1.852668
	SP2 (F)	1	1.414214
	SP3 (F)	1	1.414214
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214
/u/			
	SP1 (F)	1	1.414214
	SP2 (F)	1	1.414214
	SP3 (F)	1	1.414214
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214

Table.IV. (b) Mean (M) and standard deviation (SD) for experiment II in terms of clarity

Vowel	Speaker	M	SD
/a/	SP1 (F)	2.666667	1.404695
	SP2 (F)	2.5	1.497765
	SP3 (F)	2.666667	1.404695
	SP4 (M)	1.5	1.821187
	SP5 (M)	2	1.06066
	SP6 (M)	1.833333	1.544224
/i/			
	SP1 (F)	2.666667	1.404695
	SP2 (F)	2.666667	1.404695
	SP3 (F)	2.666667	1.404695
	SP4 (M)	1.666667	1.728132
	SP5 (M)	1.33333	1.839261
	SP6 (M)	1.666667	1.728132
/u/			
	SP1 (F)	1.833333	1.544224
	SP2 (F)	1.5	1.821187
	SP3 (F)	1.166667	1.771196
	SP4 (M)	1	1.414214
	SP5 (M)	1	1.414214
	SP6 (M)	1	1.414214

To determine quantitatively whether such glottal characteristics contain speaker dependence, features such as the timings of vocal folds opening and closing, the general shape of the glottal flow, and the extent and timing of turbulence at the vocal folds must be extracted.

Linear Predictive Coding

LPC methods provide accurate estimates of speech parameters efficiently. Linear prediction is a very powerful modelling technique LPC represent the speech waveform directly in terms of time varying parameters related to the transfer function of the vocal tract and the characteristics of the source function [7]. LPC is operated in signal processing for the expression of the spectral envelope of speech in compact form taking into concern the information required in linear predictive mode. 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 [6]. Linear prediction is a very powerful modeling technique which may be applied to time series data.In particular, the all-pole model is used in which the the signal sn is approximately represented as a linear combination of past values.

$$s(n) = \sum_{k=1}^{\rho} \alpha_k s(n-k) \qquad \text{(i)}$$

where ρ is called the order and αk's are the LPC coefficients A LPC based speech synthesizer is shown in Fig 1. The time varying all-pole digital filter, H(z), is excited by periodic pulses for voiced speech and by white noise for unvoiced speech. The output of the filter H(z) after appropriate digital to analog conversion and low pass filtering constitutes the synthetic speech signal. The predictor coefficients (αk's) are determined by minimizing the sum of squared differences between the actual speech samples and the linearly predicted.

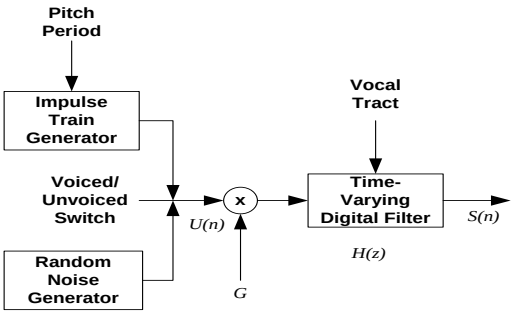

Fig 1. Block Diagram of an LPC synthesizer

Methodology

For the analysis and synthesis of the speech signal, speech of six speakers (3 males and 3 females) was recorded for sentences written in Hindi and Dogri languages at 16 kHz sampling frequency and 16 bit quantization in an acoustically treated room using Sony ICD-AX-412 digital flash memory voice recorder. The speakers were able to speak fluently both Hindi and Dogri languages. Further, the speakers belonged to same age group. After recording, the speech was segmented manually and vowels (/a/, /i/, /u/) were extracted and investigations were conducted using three experiments. Experiment I was conducted for Hindi without any parameter modification. Similarly, Experiment II was conducted using Dogri without parameter modification. In Experiment III, glottal excitation was modified with noise when the recorded speech is in Hindi. In Experiment IV glottal excitation was modified with noise when the recorded speech is in Dogri.

Results and Discussions

Identity and clarity of each speaker for Experiments I is shown in Table I(a) and Table I(b) respectively. Similarly Identity and Clarity of each speaker for Experiments II is shown in Table II(a) and Table II(b) respectively. Identity and Clarity of each speaker for Experiments III is shown in Table III(a) and Table III(b) respectively. Identity and Clarity of each speaker for Experiments IV is shown in Table IV(a) and Table IV(b) respectively. All of these experiments are conducted separately for each of the three vowels (/a/, /i/, /u/). Histograms and spectrograms are also shown corresponding to these Experiments, for each of the three vowels. Mean (M) and Standard Deviation (SD) values are calculated using the observations of six listeners. It can be seen from the Mean values of Table I(a), Table I(b), Table II(a), Table II(b) that the clarity and identity of all the speakers is almost same as the original voice of the speakers. It can be seen from the Mean values of Table III(a), Table III(b), Table IV(a), Table IV(b), where glottal excitation is modified in such a manner that identity of a speaker is fully lost and the clarity is degraded though not fully lost. Further, it is perceived that with the modification clarity in male speakers is more degraded than female speakers.

Conclusion

The investigations were carried out to study the effect on the quality of speech with respect to glottal excitation modification in Hindi and Dogri languages. This investigation was done for three cardinal vowels using four Experiments. The analysis of the results showed that LPC is able to synthesize Hindi and Dogri vowels with high quality. But modification of the glottal excitation with noise degrades the quality of speech in such a manner that identity of a speaker is fully lost and the clarity is degraded though not fully lost. Further, it is perceived that with the modification clarity in male speakers is more degraded than female speakers.

Fig 2. Comparison of mean of different speakers in terms of identity for Experiment I

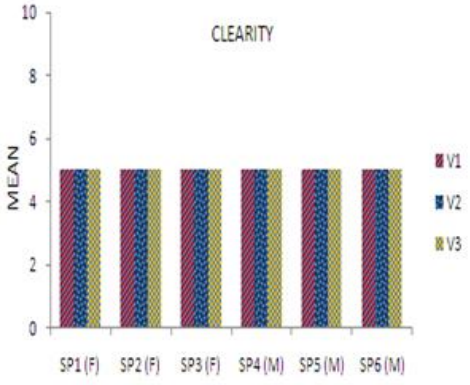

Fig 3. Comparison of mean of different speakers in terms of Clarity for Experiment I

Fig 4. Comparison of mean of different speakers in terms of Clarity for Experiment II

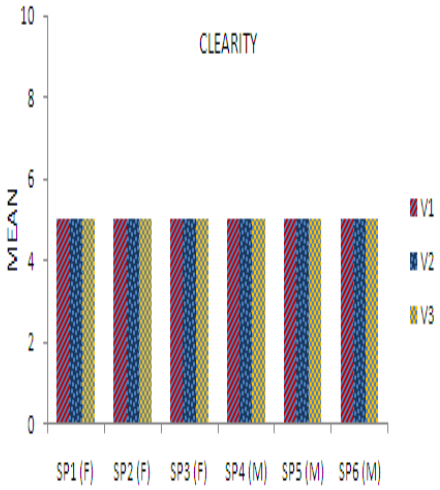

Fig 5. Comparison of mean of different speakers in terms of Clarity for Experiment II

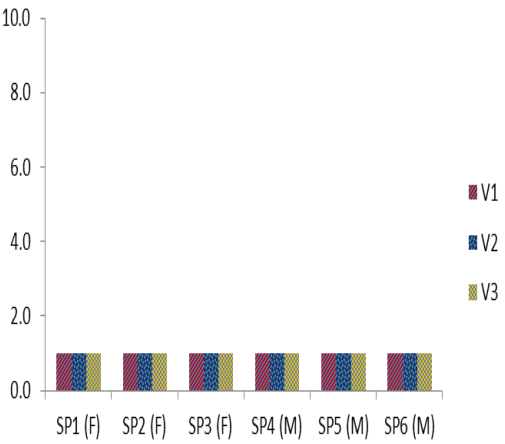

Fig 6. Comparison of mean of different speakers in terms of Identity for Experiment III

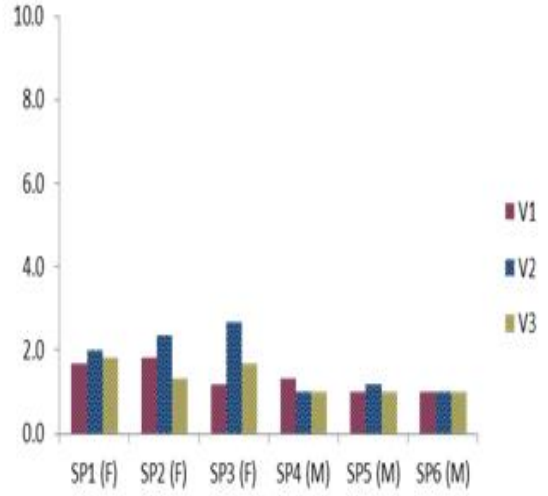

Fig 7. Comparison of mean of different speakers in terms of Clarity for Experiment III
SP –Speaker, M- Mean, SD – Standard Deviation

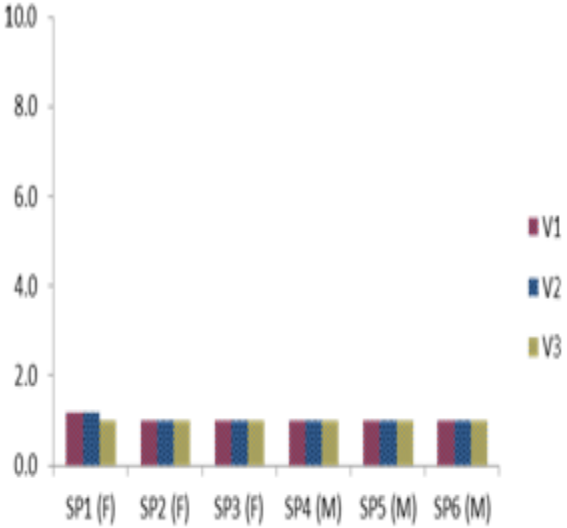

Fig 8. Comparison of mean of different speakers in

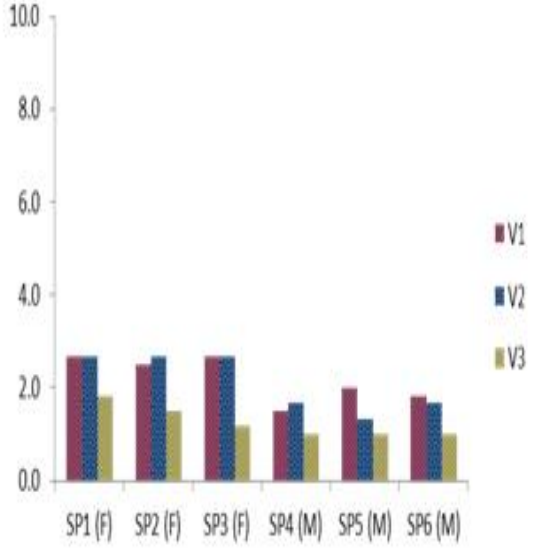

Fig 9. Comparison of mean of different speakers in terms of Clarity for Experiment II

SP –Speaker, M- Mean, SD – Standard Deviation

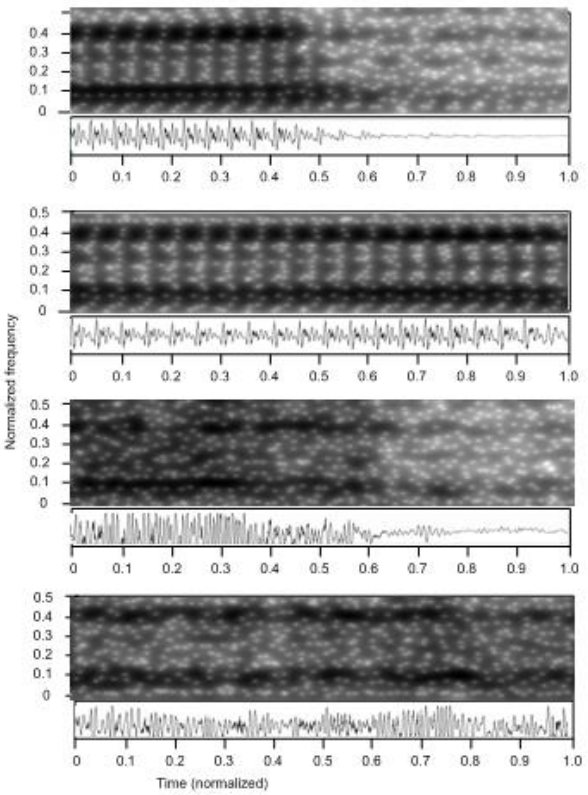

Fig 10. Spectograms obtained from recording of one of the speakers for vowel (/i/) a) Experiment I b) Experiment II c) Experiment III d) Experiment IV

References

[1] Flanagan J. L. Speech Analysis, Synthetic and Perception. Springer, New York, 1972.

[2] Dudley Homer .The carrier nature of speech. The Bell Syst. Tech. Journal,1940; 9 (4):495- 515.

[3] Dudley Homer, Tarnozy T.H. The speaking machine of Wolfgang von kempelen. The Jouralof the Acoustic Society of America, 1950; 22 (2): 151-166.

[4] Al-Akaidi Marwan. Fractal Speech Processing. The Press Syndicate of University of Cambridge, 2004: 3-4.

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

[6] Rabiner L R & Schafer R W .Digital processing of speech signals. Prentice-Hall Inc.,

[7] Englewood Cliffs. New Jersey, 1978.

[8] Furui S. & Sondhi M.M. Advance in speech Signal Processig. Marcel Dekker, New York, 1992.

[9] Rulph Chassaing and Donald Reay. Digital Signal Processing and Applications with C6713 and C6416 DSK.IEEE Press,2nd Edition Wiley Inter science Pub, London,2008.

[10] John Makhoul. Linear Prediction: A tutorial Review. In Proceedings of IEEE, April 1975;63:561-580.