Volume 5, Issue 3, March 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Comparison of LPC and HNM Model on the Quality of Children Speech in Dogri Language

[1]Varun Sharma[*], [2]Randhir Singh, [3]Parveen Lehana
[1,2]Department of ECE & PTU, Punjab,
[3]Dept. of Physics and Electronics, University of Jammu, Jammu, India.
India

Abstract— ***Speech is the most innate and fastest means of communication between humans. Speech is so familiar a feature of daily life that we rarely pause to define it. Speech is a natural form of communication in human beings and seems as natural to humans as walking, and only less so than breathing. The research work is carried out to synthesis and analyses children speech in Dogri language using Liner Predictive Coding (LPC) and Harmonic Plus Noise Model (HNM) models of speech synthesis. For the analysis of speech signal we have carried out the recording of seven children speakers (3 male and 4 female) in Dogri language between the age group of 3-6 years. LPC and HNM have been employed as the analysis-synthesis platform as it outperforms almost all models of speech production in terms of important characteristics like naturalness, intelligibility, and pleasantness. Quality of the synthesized children speech has been tested using ITU-T standard perceptual evaluation of speech quality (PESQ). Mean and standard deviation (SD) is estimated for original and synthesized speech. PESQ score obtained for all the speakers in case of LPC synthesis is around 4.5 and in case of HNM synthesis PESQ score is around 3.5. Hind is one of the prominent languages of India while Dogri is spoken in the regions like Jammu, parts of Kashmir, Himachal and northern Punjab. Dogri was given the honour of the national language on 22nd December, 2003***

Keywords— speech production, HNM, PESQ, LPC, Indian Languages

I. INTRODUCTION

Speech is the most natural kind of communication provided different forms of information to the listener. Information like the message contents, emotion, gender and identity of speak, accent, expression, style of speech, emotion and the state of health of the speaker is also an essential part in the oral swap over of communication [1]. The speech produced by vocal organs of the speaker reaches not only the ears of the listener but also those of the speaker himself. Speech sounds are sensations of air pressure vibrations produced by air exhaled from the lungs and modulated and shaped by the vibrations of the glottal cords and the resonance of the vocal tract as the air is pushed out through the lips and nose

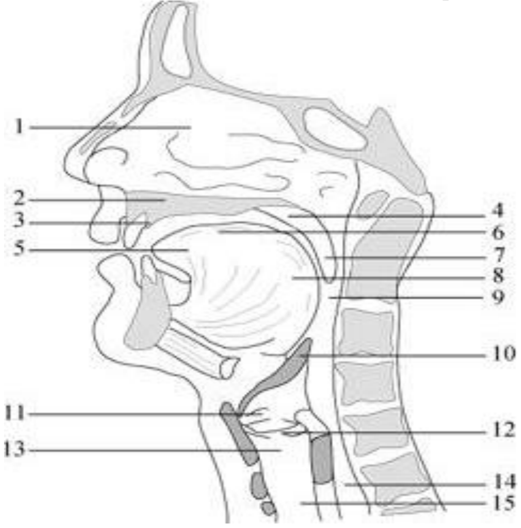

Fig.1 The human vocal organs. (1) Nasal cavity, (2) Hard palate, (3) Alveolar ridge,(4) Soft palate (Velum), (5) Tip of the tongue (Apex), (6) Dorsum, (7) Uvula, (8) Radix,(9) Pharynx, (10) Epiglottis, (11) False vocal cords, (12) Vocal cords, (13) Larynx,(14) Oesophagus, and (15) Trachea[2]. Harmonics plus Noise Model (HNM) are widely used to model spectra in many synthesis and conversion systems. HNM model composed of harmonic part (accounts for quasi periodic component of speech) and noise part (accounts for non periodic components such as fricative or aspiration). These parts are separated in frequency domain by time varying parameter [3]. Harmonics are represented by the lower

band and modulated noise is represented by upper band. These validations are useful from perception point of view which leads to simple speech model, providing high quality synthesis and modification of the speech signals [4].

Linear predictive coding, LPC, has been widely used especially in speech signal processing [5].In linear prediction an estimate f(n) for a sample value x(n) is given as a linear combination of previous sample values. There are infinitely many alternative ways to form a linear combination of signal history and use it to predict the next signal value. A weighted sum of a finite number N of previous signal values, is typically utilized in, e.g., coding applications. In traditional LPC filter assumes conventional all-pole filter, and the optimal coefficients, in a least squares sense, can be obtained from the celebrated Yule-Walker equations which obey a more general orthogonality principle of linear prediction [6, 7]. The Indian subcontinent consists of a number of separate linguistic communities each of which share a common language and culture. The people of India speak many languages and dialects which are mostly varieties of about 15 principal languages. Some Indian languages have a long literary history--Sanskrit literature is more than 5,000 years old and Tamil 3,000. Dogri is part of the Indo-European language family, meaning that it is related to many of the languages spoken across the broad region from India to Europe. It is more specifically part of the Indo Aryan family, which includes most languages of North India and Pakistan–Hindi, Urdu, Punjabi, Marathi, Oriya, and Bengali, to name a few. Dogri is also classified as one of the Pahari languages. Dogri was given the honor of the national language on 22nd December, 2003[8-9].

II. HARMONICS PLUS NOISE MODEL

HNM model composed of harmonic part (accounts for quasi periodic component of speech) and noise part (accounts for non periodic components such as fricative or aspiration). The harmonic part and noise part constitute the quasi-periodic components and non-periodic part respectively [10]. HNM decomposes speech into two components: a harmonic component and a noise component

Harmonics parts in lower band are modelled as sum of harmonics

$$s_h(t) = \sum_{k=-L(t)}^{L(t)} A_k(t) e^{jk\omega_0(t)} \qquad (1.1)$$

Where $L(t)$ denotes the number of harmonics included in the harmonic part, $\omega_0(t)$ denotes the fundamental frequency while $A_k(t)$ can take on one of the following forms:

$$A_k(t) = a_k(t_i) \qquad (1.2)$$

$$A_k(t) = a_k(t_i) + tb_k(t_i) \qquad (1.3)$$

$$A_k(t) = a_k(t_i) + tc_k(t_i) + t^2 d_k(t_i) \qquad (1.4)$$

Where $a_k(t_i), b_k(t_i), c_k(t_i)$, $d_k(t_i)$ and are assumed to be complex numbers with $\arg\{a_k(t_i)\} = \arg\{c_k(t_i)\} = \arg\{d_k(t_i)\}$ where, arg, denotes the phase angle of a complex number [11-12].

Modulated noise is dominates the voiced speech spectrum in the upper band of HNM. The noise part is given by following expression

$$s_n(t) = e(t)[h(\tau, t) * b(t)] \qquad (1.5)$$

Where $*$ denotes convolution and $b(t)$ is white Gaussian noise. The synthetic signal is given by

$$\$ = s_h(t) + s_n(t) \qquad (1.6)$$

It is important that the noise part $s_n(t)$, be synchronized with the harmonic part $s_h(t)$ [13].

III. LINEAR PREDICTIVE CODING

Linear Predictive Coding (LPC) of speech is one of the most powerful speech analysis techniques. Its main property is its computationally efficient ability to extract sufficiently accurate estimates of the spectral envelope *H(z)* in the form of an all pole filter. Due to this property LPC is a useful technique for estimating many basic speech parameters such as formants and spectra, and for low bit rate coding.LPC is one of the main methods used to extract the filter parameters of the source-filter model of speech production. . The block diagram of a simple LPC vocoder is presented

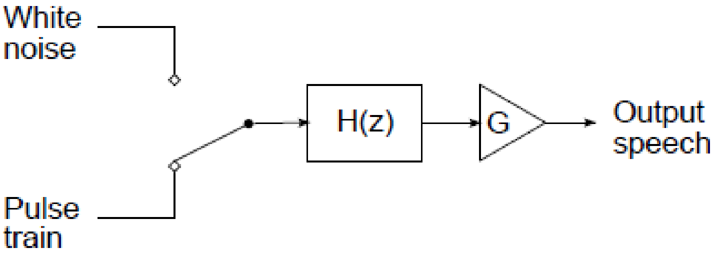

Fig.2 the block diagram of LPC vocoder

In linear prediction an estimate f(n) for a sample value x(n) is given as a linear combination of previous sample values. There are infinitely many alternative ways to form a linear combination of signal history and use it to predict the next signal value. A weighted sum of a finite number N of previous signal values, is typically utilized in, e.g., coding applications

IV. METHODOLOGY

The research work is divided into two major parts. In first part speaker selection, speech recording and segmentation is done, while in the second part analysis-synthesis of speech has been performed by using HNM model and objective evaluation of speech quality has been estimated by perceptual evaluation of speech quality (PESQ). Eight different phrases in Dogri language are recorded using Goldwave software at the sampling rate of 16,000 KHz. The material was recorded in an acoustically treated environment and segmented and labelled manually using Praat software. HNM based speech synthesis of Dogri language is carried out in this research work taking seven speakers in the age group of 3-6 years using HNM algorithm. The deviation between the original and HNM synthesized speech were analysed Fig.3 shows the block diagram of the research methodology

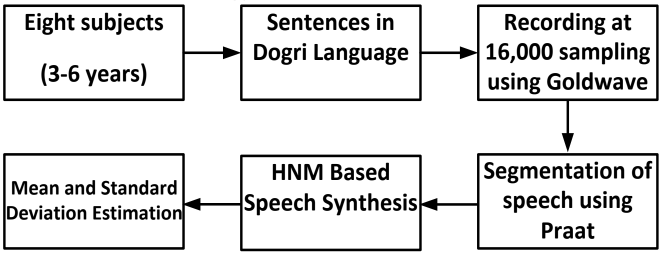
Eight subjects (3-6 years)
Sentences in Dogri Language
Recording at 16,000 sampling using Goldwave
Segmentation of speech using Praat
HNM Based Speech Synthesis
Mean and Standard Deviation Estimation

Fig. 3 Block diagram representation of proposed methodology

V. RESULTS AND DISCUSSION

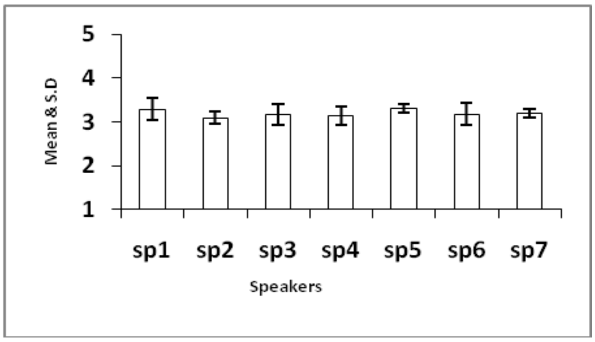

Fig. 4 Mean and standard deviation of all the HNM at v100n100 synthesized speech with respect to original speech

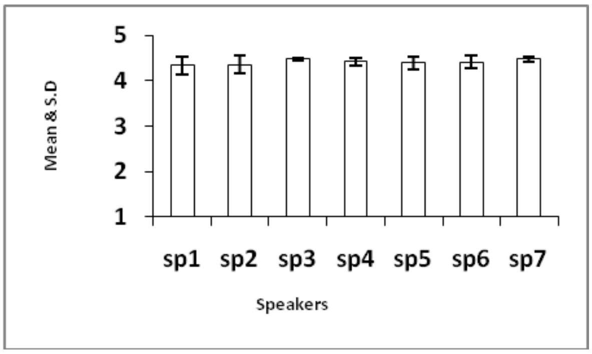

Fig. 5 Mean and standard deviation of the LPC at level 14 of synthesized speech with respect to original speech

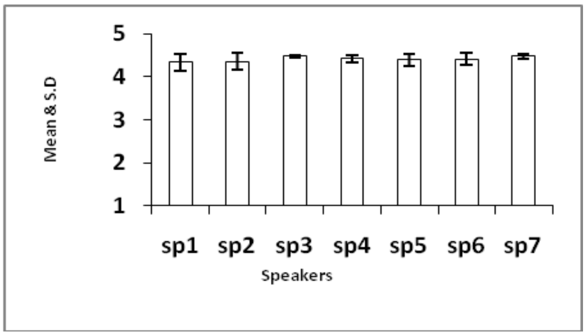

Fig.6. Mean and standard deviation of the LPC at level 18 of synthesized speech with respect to original speech

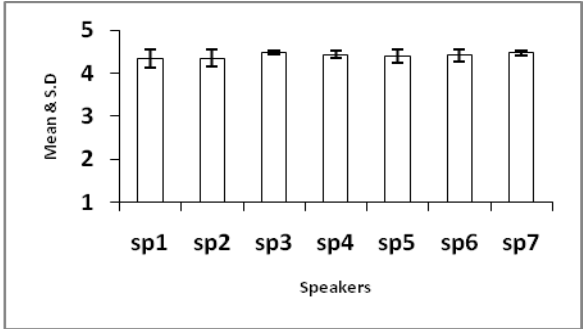

Fig.7. Mean and standard deviation of the LPC at level 22 of synthesized speech with respect to original speech

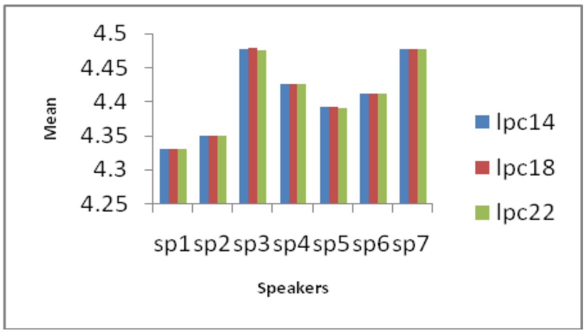

Fig.8. Mean of all the LPC at different level 14,18,22 of synthesized speech with respect to original speech

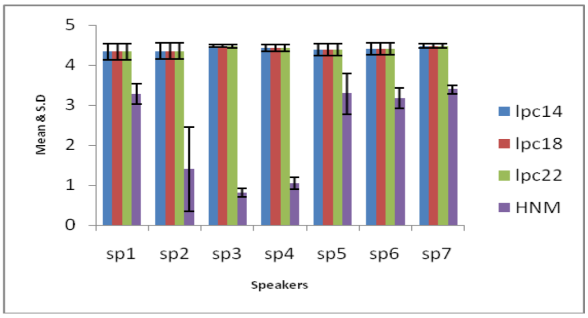

Fig.9. Mean and standard deviation of all the LPC at different level 14,18and 22 and HNM at v100n100 of synthesized speech with respect to original speech. There is no considerable effect on the quality of children speech by using different order of LPC that is 14,18 and 22. PSEQ scores obtained by LPC model is higher than the HNM model

VI. CONCLUSION

Research work is carried out to evaluate and compare the quality of synthesized speech of children in Dogri language. The effect of the HNM at 100% voice part and at 100% noise part on the synthesized speech quality and compare it with LPC at different order at 14,18 and 22 intelligibility has been discussed. HNM and LPC has been used as analysis and synthesis platform and the PESQ as the evaluation method for the quality. There is no considerable effect on the quality of children speech by using different order of LPC that is 14,18 and 22.PSEQ scores obtained by LPC model is higher than the HNM model From the results it is quite apparent that LPC model proves a robust model for children speech as it synthesizes all the voices quite clearly

REFERENCES

[1] P. B. Denes and E. N. Pinson. The Speech Chain: The Physics and Biology of Spoken Language, New York: Anchor Press, 1973

[2] M. Berouti, R. Schwartz, and J. Makhoul,"Enhancement of speech corrupted by acoustic noise,"in *Proc. ICASSP*, pp. 208-211, 1979.

[3] Y. Stylianou, Applying the Harmonic Plus Noise Model in Concatenative Speech Synthesis, *IEEE Trans. Speech and Audio Process*.,vol. 9, 2001

[4] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Commun*., vol. 9, pp. 453–467, 1990.

[5] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals," in *Proc. IEEE Conf. on Commun. and Process.*, pp. 360–361, 1967.

[6] A. N. Kolmogorov, "Stationary sequences in Hilbert space," *Bull. Math. Univ. Moscow*, vol. 2, no. 6, 1941.

[7] N. Wiener. Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, Technology Press and John Wiley & Sons, Inc., New York, 1949.

[8] Grierson, Sir George, ed. Linguistic Survey of India. Volume IX, Part 4. 1906. Reprinted by MotilalBanarsidas, 1967.

[9] Pushp, P. N., and K. Warikoo, eds. Jammu, Kashmir, and Ladakh: Linguistic Predicament. 2004.

[10] Moulines E and Charpentier F. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones Speech Commun; 1990, 9(453).

[11] Erro D, Sainz I, Navas E, Hernaez I. HNM-based MFCC+F0 extractor applied to statistical speech syn-thesis. Proc. ICASSP; 2011.

[12] Stylianou Y. Applying the harmonic plus noise model in concatenative speech synthesis. IEEE transactions on speech and audio processing; 2001, 9(1).

[13] Stylianou Y, Laroche J, and Moulines E. High-quality speech modification based on a harmonic + noise model in Proc. Eurospeech; 1995, (451).

[14] L. Rabiner and R. Schafer. Digital Processing of Speech Signals Prentice Hall, 1978.

[15] F. Soong, and B. Juang, "Line spectrum pair (LSP) and speech data compression. In Acoustics, Speech, and Signal Processing,"*IEEE International Conference on ICASSP*, vol. 9, pp. 37-40, 1984.