

Effect of LPC Based Synthesis on the Vocal Calls of Indian Ringneck and African Grey Species of Parrots

Kamaljit Singh Arora¹, Randhir Singh², Jang Bahadur Singh³, Parveen Lehana^{4*}

¹M.tech (ECE) student, Department of Electronics and Communication Engineering, SSCET, Punjab, India

²Assistant Professor & Head, Department of Electronics and Communication Engineering, SSCET, , Punjab, India

³M.Phil student, Department of Physics and Electronics, Jammu University, Jammu, J&K, India

⁴Associate Professor, Department of Physics and Electronics Department, Jammu University, Jammu, J&K, India

Abstract: Speech signal processing is one of the interesting research areas now-a-days and it has been explored for various applications. For example, speech synthesis is a synthetic signal processing method to produce exactly same replica of output speech as that of original input speech using various parameters derived from the recorded speech. Vocal signals generally represent an important part in the life of birds since they do not have a strong smelling sense. These signals help birds to protect themselves from any threat or danger and are needed for finding food, water and shelter. Synthesis of calls has many applications for e.g. synthesizing an alarm call may be used to save the bird from any danger or to prevent bird flu by providing a proper habitation to the birds, etc. The vocal signals which comprised of some special sounds usually given for a particular communicating activity, known as a bird call. Since the basic mechanism of sound production in birds is almost similar to that of humans, LPC model may be used for the synthesis of vocal calls of birds. In this paper, the validity of LPC model have been successfully investigated by estimating speech parameters, visual analysis of spectrograms, subjective evaluation, MOS scores, etc. for the synthesis of the calls of Indian Ringneck and African Grey species of parrots.

Keywords: Speech signal processing, Speech synthesis, Linear predictive coding (LPC), Vocal calls, MOS scores.

I. INTRODUCTION

Speech synthesis is a signal processing method in which simulation is needed for the synthesis of speech using different models of speech generation and synthesis. LPC model is one of the excellent models for speech synthesis as it reduces bit rate, reduces error, and increases accuracy [1], [2].

Bird calls make a unique role in human's life. Bird calls keep us in touch with the nature beauty. Their calls suddenly make our mind fresh and stress free. Instead of villages and mountains, even in big cities, people can recognize and enjoy different calls and songs of bird's species especially in the early morning. Bird calls give a lot of inspiration to children, poets, artists, writers, and music composers [3], [4]. The basic composition of speech production mechanism in bird's species is almost similar to that of the human beings. As we know that, human's speech sounds are composed of small units called phonemes, the bird calls are also made up of small units. Humans use vocal cords for excitation of their vocal tract, but there are no vocal cords present in birds. Instead, they have a special important organ called syrinx and its position can be naturally present in the intersection of the trachea and the two bronchi or entirely in the trachea or in the bronchi's [5]. The main internal parts of speech for the sound production mechanism in bird's species include lungs, bronchi, syrinx, trachea, larynx, mouth, and beak. Syrinx is an important sound organ used for the production of calls in birds and it also gives information about the internal structure of different birds because different bird's species have different anatomy or structure of syrinx, for e.g. parrots have mostly tracheal syrinx [6]. Different interiors of the syrinx can be used by different parrots to produce calls and songs, and sound production or mechanism is controlled by non homologous part of the brain. Tracheal syrinx is generally found in parrots. The parrot tracheal syrinx generally composed of syringeal muscles and a pair of lateral tympaniform membranes. The tracheal syrinx composition is shown in Fig.1. Parrots can also speak out complex vocalizations, it occurs due to the anatomical structure of the syrinx with the use of two pair of syringeal muscles. They can twist their tongues very easily. Parrots have two intrinsic syringeal muscles and labium is generally not present in them [7], [8]. Some of the researchers have already put a lot of effort on the calls of parrot species of birds. Irene M. Pepperberg experimentally found out excellent results for the vocal learning of grey parrots (*Psittacus*

Erithacus) [9]. P. Skripa explained that cepstral transformation and self organizing maps (S.O.M) can be successfully and easily performed on parrot vocalizations or calls [10]. The objective of this research is to investigate the validity of LPC model, focusing on the synthesis of parrot calls by taking different observations, parameters, plots and spectrograms. This research paper is organised as follows: Section I gives the introduction about bird’s vocal communication and their sound production mechanism. Section II is helpful to understand the LPC model of speech in brief. Section III explains the categories of Parrot’s vocal calls. Section IV shows the research methodology i.e. a step by step process for the synthesis of vocal calls. The last section V comprised of results and conclusions and followed by the references.

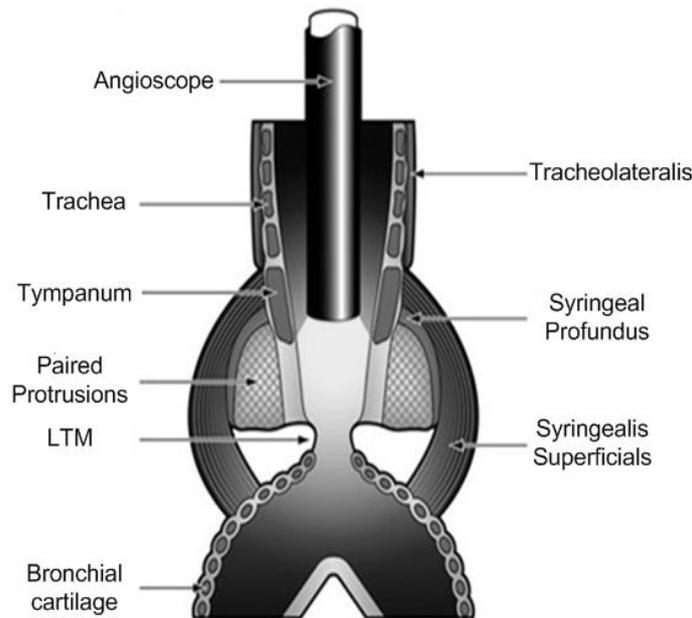


Fig.1. Syringeal anatomy of a Parrot (*Psittacine*) [11].

II. LPC MODEL OF SPEECH

LPC model has the ability to find out the parameters of a speech signal for synthesizing the speech. This model has been used for speech coding applications and is useful for reducing transmission rate (low bit rate speech coding). [12].

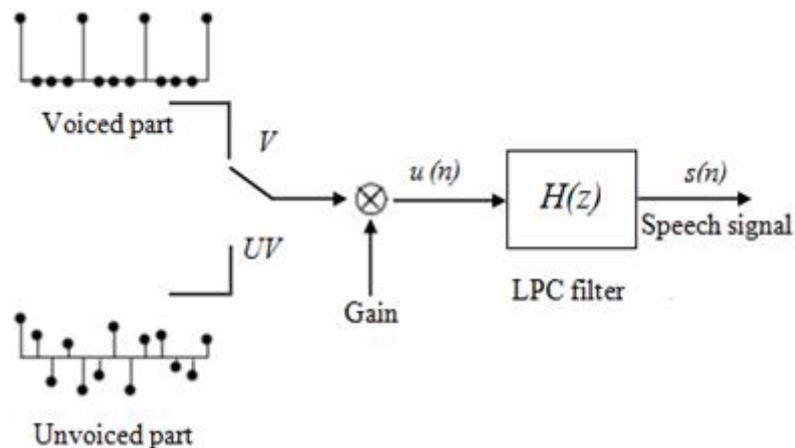


Fig.2. LPC model of speech

It can be used to produce efficient speech synthesis output. It can be used to find out vocal tract area functions, fundamental frequency ‘*F_o*’, bandwidths ‘*B*’, intensity of the sound and frequencies of spectral poles and zeros (for e.g., formants of

analyzing sound), but it is mostly used to determine small set of parameters of speech which represents vocal tract functions. The LPC model is splitted into two parts, one part is analysis part and the other part is synthesis part. The analysis part is used to analyze the speech signal and to estimate the error signal. The Speech analyzer is used for determining the coefficients of LPC filter, the pitch, voiced (*V*) and unvoiced (*UV*) part for each speech frame of a particular sound [2], [13]. The synthesis part of the model can be used for producing synthesized speech output by taking an input as error signal [14]. The LPC model is shown in the Fig.2.

The relation between speech signal and excitation signal can be written as

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n),$$

where *G* is gain and *s*(*n*) is the speech signal.

III. PARROT'S SPECIES AND THEIR CALLS

The Parrots are also named as Psittacines and can produce complex vocalizations. There are around 372 species of parrots around the world. They can twist their tongue in order to produce complex vocalizations. There are different categories of parrots depending on their size, weight, age, calls, etc [15]. The most commonly parrot found in India is Indian Ringneck parrot. Its biological name is *Psittacula krameri Manillensis*. They usually measure on an average size of 16 inches in length including their feathers. The average single wing length of Ring neck is about 15-17.5 cm and they live a life of about 20-30 years [16]. Also, African grey is the famous parrot of Africa mainly found in Congo and can produce vocalizations much more than the Indian parrot. Their mimicry activity is really very impressive. They are about 11-14 inches long and live a life of about 50 years [17]. It was found out that the calls of the parrots can be sub-divided into almost nine meaningful categories are explained as:

Pair Duets: Many pair of birds ranging from songbirds to parrot's species can make duet with each other for their mating nature. Only some of the parrot species have pair duets. Some species can also vocalize together in a particular place, but mostly they call sequentially in nature. Pair duets birds can be observed in several species of parakeets, Lovebirds, Galahs and a number of Amazon species, etc.

Warbles: Some parrot's species have long rambling sounds with some notes highly variant in tone and can be originated near nesting area or during resting time and can also be produced at late noon. Budgerigar's male species of parrots use these calls for stimulation in females for reproduction process.

Begging call: This call is mostly given by young parrots or baby parrots. This call is indicating that the young parrots or parrot babies are hungry. This type of call is usually given in the morning.

Alarm call: It is a loud and sharp cry call and the parrots use these calls when they feel threatened. This call is also used to warn or protect other birds from danger.

Pre-flight call: This call is generally produced just before the flight take off or before the time when parrot start to fly in the sky. It is mostly audible as a loud and harsh flight call. Pre-flight call can be produced by Orange Fronted Conures, Sulphur Rosellas and Budgerigars etc.

Distress call: This call is produced or given by an injured or distress parrot.

Soft contact call: This call makes coordination with the movement of flock members when they moved for particular vegetation in a group. This call listens to be very low in amplitude and can be repeated mainly with or without any responses by other members of the flock species. Examples of these types of calls include soft buzz sounds of orange chinned Parakeet, mumble sounds of yellow Amazons, chet sounds of the Galah and chatter sounds of monk parakeets, etc.

Agonistic protest: Some birds produce some form of sound like squawk during flights. It signifies a highly pitched squawk or high in tone, mostly produced by a parrot of angry or disturbed nature.

Loud contact call: Mostly all species of parrots can produce a particular type of call that is given during flight and can be mutually transferred by some other species of flying birds and some separated members. It is used to create a vocal connection between some specific birds. It is generally the loudest of all calls. Examples of these types of calls include tinkle of orange chinned parakeets, chet of the Galah, etc. [18], [19], [20].

IV. METHODOLOGY

The methodology of investigations carried out may be divided into four sub-parts: material recording, estimation of parameters, visual perception of speech using spectrograms and subjective evaluation. The corresponding flow chart is shown in Fig.3.

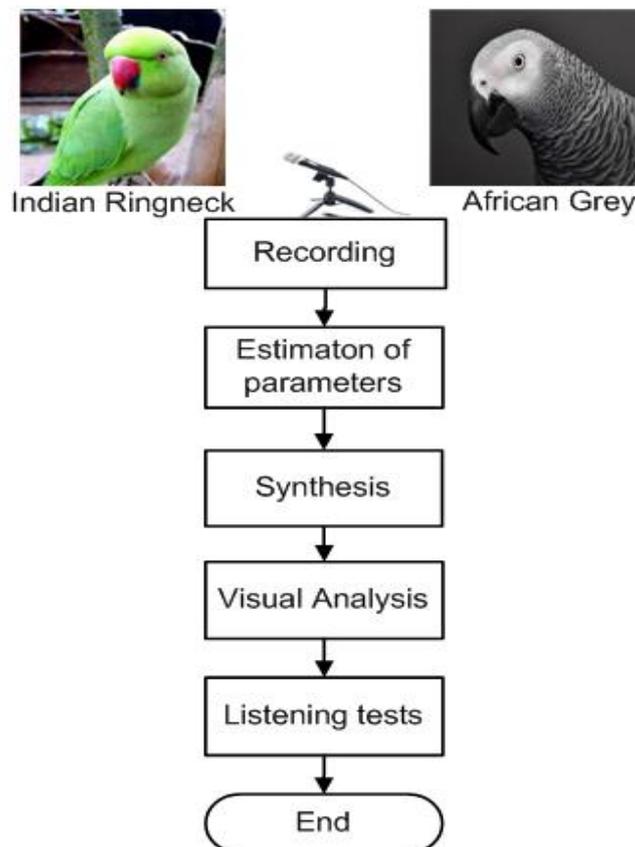


Fig.3. Research Methodology

A. Material Recording

Investigations were carried out by recording six calls of Indian Ringneck and African Grey (Parrot's species) at different places using a high quality microphone of Sony digital voice recorder (ICD-UX513F). It is a 4GB UX digital voice recorder with expandable memory capabilities and provides recording with high voice quality [21]. The recording was done in an acoustically shielded and noiseless environment. The total recorded calls were of about 15 minute-duration. The sampling frequency and number of bits used for quantization were 16 KHz and 16 bits, respectively. After that, recorded calls were labeled, processed and stored in wav format.

B. Estimation of Parameters

LPC model have been used for the analysis and synthesis of the recorded calls. Since LPC parameters exhibit significant effect on the synthesized speech quality. So, the recorded calls were analyzed for different parameters of speech. The parameters like size of time frame, in ms (fr), size of window, in ms (fs) and order of LPC (L) were adjusted manually at different values in order to carry out efficient speech synthesis after listening original and synthesized calls of both the species. Size of time frame, in ms (fr) represents the speed or rate at which input samples of speech are analyzed and output speech is reproduced [22]. Size

of rectangular window, in ms (fs) represents the frame size which is useful for windowing the speech. The order of LPC model (L) represents the number of poles in the filter. So, speech synthesis has been carried out at a fixed range of values for these significant parameters.

C. Visual Analysis of Call Spectrograms

In order to make efficient synthesis, the frequencies, amplitudes and duration of the synthesized speech must be similar with the original recorded speech. The performance and effectiveness of LPC model for the synthesis of recorded calls were analyzed using voice spectrograms. The spectrograms of both i.e. original recorded and synthesized calls of the two species were observed for making a visual comparison between recorded and synthesized call quality keeping in view different factors like formants, bandwidths and amplitudes. The spectrogram is said to be a computer generated plot which shows different frequencies present in a speech signal at each instant of time. Spectrogram is a useful method to recognize acoustic signals on the basis of their amplitude, frequency and duration. These spectrograms represent speech signals as the square of their absolute value of short time Fourier transform (STFT). The spectrogram can be plotted as a three dimensional plot for frequency versus time domain. The original and synthesized plots in frequency versus time domain which give an indication about the changed frequency view of a speech signal [23], [24].

D. Subjective Evaluation

Subjective evaluation was carried out for comparing the natural and synthesized speech quality. Subjective evaluation is a method which instructs all the listeners to hear the speech signal carefully for rating or scoring all the processed or synthetic speech signals so as to recognize the perceived call quality [25]. Mean opinion score is a method to provide a numerical value for the quality of speech output. This method uses subjective tests in which opinionated scores are mathematically averaged so as to get an indication for the speech quality [26], [27], [28]. It is defined as the arithmetic mean of all the individual scores that have value between 1 and 5 where 1-Bad, 2-Very Annoying, 3-Annoying, 4-Good and 5-Excellent. In order to evaluate the overall accuracy for the synthesis process, a Mean Opinion score (MOS) test was carried out for evaluating the similarity the synthesized and natural speech quality to find out the synthetic performance of LPC model. Five listeners give scores in the range from 1 to 5 on different vocal calls of both the species so as to compare the natural and synthesized speech quality. Since all the calls were randomly listened by all the listeners. So, each listener scored four times a call and the average of the no. of times a call listened was calculated for each listener. The MOS scores and standard deviations for all the calls were also calculated. The listener must be good in hearing and capable of differentiating the synthesized signal from any distortion or background noise in comparison to natural signal. The listener must sit in a quiet room and the noise level of the room must be below 30 dB. The listening test was performed using high quality headphones (Sony MDR-XD200).

V. RESULTS AND CONCLUSIONS

Investigations were carried out for determining the validity of LPC model for the analysis and synthesis of the calls of Indian Ringneck and African Grey species of parrots. From the investigations, it has been found out that LPC parameters affect the synthesized quality of the speech. When size of window ' fs ' become more greater than size of time frame ' fr ', then the synthesized speech gets distorted and perceived as annoying or unacceptable in audible form. Hence, the parameters was adjusted of the values as $fs = 30$ ms, $fr = 20$ ms and $L = 21$ so to make the synthesized speech acceptable.

From the visual perception of speech through spectrograms, it has been found out that the frequencies or formants (frequency bands), amplitudes and duration of the recorded and synthesized calls are almost similar. In other words, it can be seen that the spectrograms of synthesized calls were analyzed and found out as almost similar to the natural or recorded calls spectrograms for both the species. The spectrograms have been plotted and analyzed for different calls of the above mentioned species. For example, one of the spectrograms for the original and synthesized calls of Indian Ringneck and African Grey species of parrots are shown in Fig.4 and Fig.5, respectively.

Mean opinion score (MOS) was scored by five different listeners representing from L1 to L5 for different calls C1 to C6 and C7 to C12 for both the species. MOS scores and standard deviations of the calls of Indian Ringneck and African Grey species of parrots are shown in Table I, and their corresponding block plots are shown in Fig.6 and Fig.7, respectively where MOS scores are represented by blocks (indicated by light black colour and their standard deviations are represented by plus bars above the blocks). The overall average MOS score of Indian Ringneck and African Grey species of parrots are calculated as 4.23 and 4.39, respectively. Hence, from the estimated parameters, visual perception of speech through spectrograms, listening tests or subjective evaluation, MOS scores and block plots, it is concluded that the perceived synthesized quality of the calls is excellent

in audible form. So, LPC model can be efficiently used for the synthesis of the calls of Indian Ringneck and African Grey species of parrots.

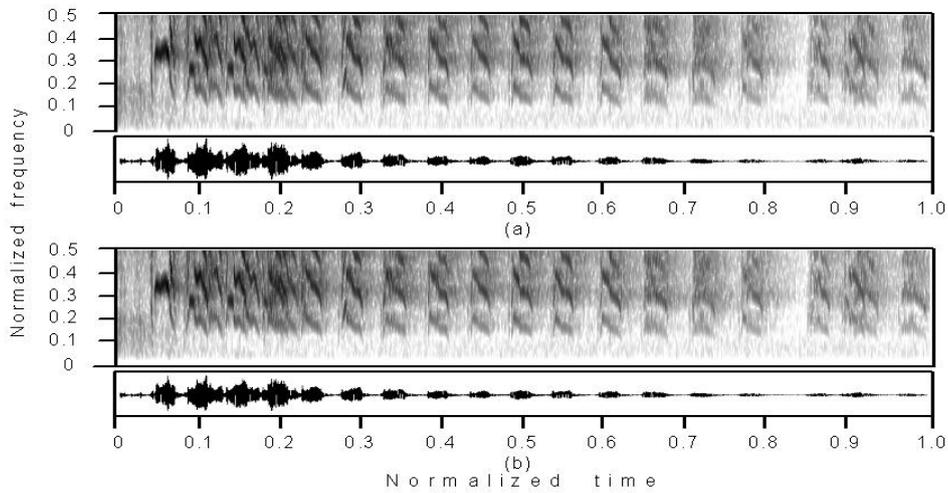


Fig.4. Spectrograms of Ringneck Vocal Call
(a) Original call (b) Synthesized call

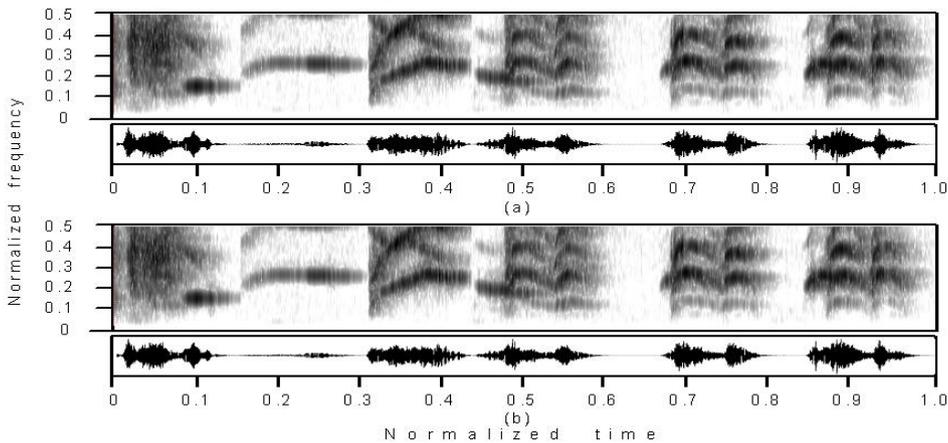


Fig.5. Spectrograms of African Grey Vocal Call
(a) Original call (b) Synthesized call

Table I: MOS scores for vocal calls

Listeners	Vocal calls of Indian Ringneck						Vocal calls of African Grey					
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
L1	4.5	5	4	4.25	4	4.25	4.5	4.25	4	4.21	4	5
L2	4	4.5	4	4.25	5	4.5	4	5	4.25	5	4	5
L3	4.25	4.5	4	4	4.75	3.75	4.25	4.5	5	4.5	4.25	4.5
L4	4	4.75	3.75	3.5	4.75	4	5	4	4.5	4	4.25	4
L5	4.21	4	4.25	4	4	4.25	4.5	4.25	4.21	4.21	4.5	4.25
MOS	4.19	4.55	4	4	4.5	4.15	4.45	4.4	4.39	4.38	4.2	4.55

S.D	0.185	0.370	0.176	0.273	0.418	0.285	0.331	0.379	0.343	0.387	0.187	0.447
-----	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

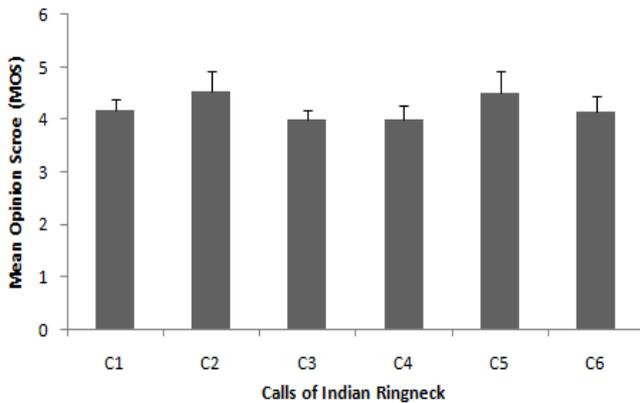


Fig.6. Block plot of MOS scores for Ringneck calls

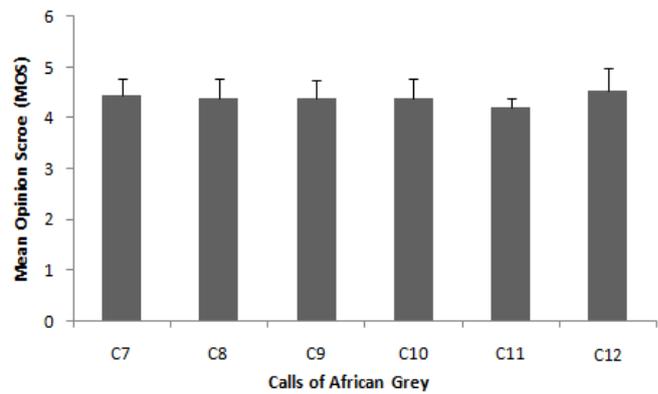


Fig.7. Block plot of MOS scores for African Grey calls

REFERENCES

[1] Rabiner, L., "Fundamentals of Speech Recognition" Pearson, India, 1993.

[2] Kura, V.B., "Novel pitch detection algorithm with application to speech coding," M.S.Thesis, Dept. of Elect. Engineering, University of New Orleans, 2003.

[3] Harma, A., "Automatic identification of bird species based on sinusoidal modeling of syllables," in Proc. ICASSP, Hong Kong, 2003.

[4] Somervuo, P., Harma, A. and Fagerlund, S., "Parametric representations of bird sounds for automatic species recognition," IEEE Trans. Speech and Audio Processing, vol. 14, no. 6, pp. 2252 – 2263, 2006.

[5] Singh, R., Lehana, P. and Singh, G., "Investigations of the phonemes in the calls of little owls using vector quantization," Int. journal of Information technology and knowledge management, vol. 2, no. 2, pp. 337-342, 2009.

[6] Fagerlund, S., "Acoustics and physical models of bird sounds," in Seminar in acoustics, HUT, Laboratory of Acoustics and Audio Signal Processing, (Espoo), April 2004.

[7] Moustaki, N., "Bird Brains: Parrot intelligence," in Parrot for Dummies, Wiley publishing, Hoboken, ch. 15, 2005.

[8] Marler, P. and Slabbekoorn, Hans, "How birds sing and why it matters," in Nature's music: the science of birdsong, vol.1, Elsevier Academic Press, California, ch.9, 2004.

[9] Pepperberg, M. I., "Vocal learning in Grey parrots (Psittacus Erithacus): Effect of social interaction reference and context" in Auk, USA, vol. 111, pp.300-313, 1994.

[10] Skripal, P., "Analysis of acoustic communication in parrots" Diploma Thesis, Dept. of computer science and Engg., Czech Technical University.

[11] <http://www.2ndchance.info/parrotalk-structureofsyriximage.htm>

[12] Shaughnessy, D.O., "Speech Analysis," in Speech Communications, 2nded., Hyderabad, Universities Press Private Limited, ch.6, pp.192-209, 2001.

[13] Madane, A.R., Shah, Z., Shah, R, and Thakur, S., "Speech Compression using linear predictive coding," in Proc. Machine Intelligence Research, Nagpur, 2009.

[14] Pham, T.D. and Wagner, A., "A geostatistical model for linear prediction analysis of speech," Pattern Recognition Society, Published by Elsevier Science, vol. 31, no. 12, 1998.

[15] Jonker, R.M.V. and Tamis, W.L.M, "Introduction, Breeding and Poaching of Scarlet Macaws in a temperate country" The open ornithology journal, vol. 5, pp. 1-4, 2012.

[16] <http://www.birdtricks.com/psittacula-krameri.html>

[17] [http://www.bestwest.ca/images/AFRICAN%20GREY%20PARROT\[2\].pdf](http://www.bestwest.ca/images/AFRICAN%20GREY%20PARROT[2].pdf)

[18] Waal, F.B.M.D. and Tyack, P.L., "Vocal communication in wild parrots," in Animal social complexity, Harvard University Press, ch.11, pp.293-303, 2003.

[19] http://www.acguanacaste.ac.cr/loras_acg/parrots.home.html

[20] Taylor, S. and Michael, R.P., "Vocalizations of the brown headed parrot Poicephalus cryptoxanthus: their general form and behavioural context", OSTRICH, pp. 61-72, 2005.

[21] <http://www.sony-asia.com / product/ icd-ux523f>

[22] Gupta, R., Mehta, A.K. and Tiwari, V., "Vocoder (LPC) analysis by variation of input parameters and signals" ISCA journal of Engineering sciences, vol. 1, no.1, March, pp.57-61, 2012.

[23] Chassing, R., "Digital Signal Processing and Applications with the C6713 and C6416 DSK" John Wiley and sons, 2004.

[24] Kammler, D.W., "A first course in Fourier analysis", Cambridge University Press, 2008.

[25] Philipos, C.L., "Speech quality Assesment", Multimedia, Processing, Analysis and communications, University of Texas-Dallas, U.S.A, pp. 623-654, 2011.

[26] Deepa, D., "Dual channel speech enhancement using HADAMARD-LMS algorithm with DCT preprocessing technique" International journal of Engineering science and technology, vol.2, no. 9, pp.4418-4423, 2010.

- [27] Lehana, P.K and Pandey, P.C, "Transformation of short time spectral envelope of speech signal using Multivariate Polynomial Modeling" in Proceedings of 17th National conference on Communications (NCC 2011), Bangalore, India, 28-30 January, 2011.
- [28] Othman, K. and Ahmad, Z.H, "Standard Malay text to speech system" International journal of computer science, vol. 2, no.4, pp. 285-293, 2007.

BIOGRAPHY



Er. Kamaljit Singh (*M.Tech Engineer*), resident of Jammu (J&K), received his B.Tech degree in Electronics and Communication Engineering with an aggregate percentage of 72% (First division) from M.B.S College of Engineering and Technology (MBSCET) in the year 2009 affiliated to University of Jammu. He also have training experience in Rural Electrification Corporation (REC) of India Ltd., Bharat Sanchar Nigam Limited (Central Govt.) and passed his M.Tech degree in Electronics and Communication Engineering with an aggregate percentage of 68% (First division) from Sri Sai College of Engineering and Technology (SSCET), Punjab, affiliated to Punjab Technical University, Jalandhar (Punjab) in the year 2012. His research interests include Speech signal processing, Digital signal Processing, Electrical systems, Power and Distribution transformers, Electrical Transmission and Distribution and having some publications in national conferences/international journals.



Er. Randhir Singh (*Assistant Professor and Head*) received his M.Tech. Degree in Electronics and Communication Engineering from Beant College of Engineering and Technology, Gurdaspur (Punjab) affiliated to Punjab Technical University, Jalandhar (Punjab). He is presently working as Assistant Professor (H.O.D), Electronics and Communication Engineering Department, Sri Sai College of Engineering and Technology, Pathankot (Punjab) and pursuing in Ph.D. Electronics and Communication from Punjab Technical University. His research interests include Speech signal processing, Digital signal Processing, Image processing, Analog and Digital Communication, Electronics and control systems, etc. and having more than 10 publications in national/international conferences and journals and a lot of experience in guiding M.Tech students.



Mr. Jang Bahadur Singh (*MPhil. Research scholar*) received his Master's degree in Electronics from University of Jammu and presently pursuing as an M.Phil. Research Scholar under *Prof. Parveen Lehana* in University of Jammu. His research interests include Speech signal processing, Digital signal processing and having more than 10 publications in national / international conferences.



Dr. Parveen Lehana (*Associate Professor*) received his Master's degree in Electronics from Kurushetra University in 1992. He worked as lecturer in Guru Nanak Khalsa College, Yamuna Nagar (Haryana) for next two years. He qualified NET-JRF in Physical science in 1994 and got selected as permanent lecturer in A. B. College, Pathankot, where he worked for one year. He also qualified NET-JRF in Electronic Science and presently working as Associate Professor in Physics and Electronics Department, University of Jammu and did Ph.D. degree from IIT, Bombay in Speaker Transformation. He also invited in many colleges for attending national and international conferences and also guided many students for short term certification programs in embedded systems. He also invited for conducting workshops on MATLAB/simulinks in different esteemed institutions/colleges. His research interests include Speech recognition, Speaker transformation, Signal processing, Speech signal processing, Analog and Digital signal processing, Nanowires characterization, Robotics, Image processing, Analog communication,

Digital communication, Microwaves and Antennas, Electronics and control systems, Instrumentation, Electronics system designing, etc. and having more than 100 publications in national/international conferences and journals. He has a lot of experience in guiding M.Tech, M.Phil and Ph.D. students and other researchers.