A Comparative Performance of Various Speech Analysis-Synthesis Techniques

Ankita N. Chadha, Jagannath H. Nirmal, and Pramod Kachare
K.J. Somaiya College of Engineering, Department of Electronics, Mumbai, India
Email: {ankita.chadha, jhnirmal}@somaiya.edu, pramod_1991@yahoo.com

***Abstract*—In this paper, we present a comparative performance of the various analysis-synthesis techniques which separate the acoustic parameters and allow the reconstruction of the speech signal, which is very close to original speech. The analysis-synthesis of speech signal is used for speech enhancement, speech coding, speech synthesis, speech modification and voice conversion. Our comparative study includes Linear Predictive Coder, Cepstral Coder, Harmonic Noise Model based coder and Mel-Cepstrum Envelope with Mel Log Spectral Approximation. The comparative performance of these vocoders is evaluated using different objective measures namely line spectral distortion, Mel cepstral distortion and signal to noise ratio. Along with objective measures, subjective measure, mean opinion score is also considered to evaluate the quality and naturalness of the resynthesized speech in term of original speech.**

***Index Terms*—acoustic parameters, complex cepstrum, harmonic noise model, linear predictive coefficients, mel-cepstrum envelope, mel log spectral approximation, vocoder**

I. Introduction

Vocoder is an intrinsic tool, in the field of signal processing and research, for speech analysis and synthesis. One of the major advantages of the speech vocoder is that it allows the separation of the segmental and supra-segmental parameters to enhance, modify and resynthesize speech signal. The analyzed parameters are used in the framework of speech recognition, speaker recognition and vocal emotion recognition. The modifications of these analyzed features are used for various applications like speech coding, speech enhancement, speech and speaker modification and voice conversion [1]-[4]. The speech signal contains acoustic and linguistic information. The language, dialect, phoneme pronunciation and social background of speaker are related to the linguistic parameters. The acoustic parameters are related to the physical structure of human speech production and perception mechanism. They are reflected at various levels such as shape of the vocal tract, shape of the glottis excitation and long term prosodic parameters. Among these the shape of vocal tract is represented using linear prediction Analysis while the glottal parameters are shown by equivalent modification of Linear Predictive Coefficients (LPC) termed as LP residual [5].

The term vocoders are classified on the basis of the type of information they yield as parametric and non-parametric vocoders. The parametric vocoders are phase vocoder, formant vocoder, LPC, Complex Cepstrum (CC) [6], Mel Frequency Cepstrum Coefficients (MFCC), Wavelet filter Bank [7], Harmonic Noise Model (HNM) and STRAIGHT [8]. The non-parametric vocoders are those which are not based on any speech processing models such as channel vocoders, Pitch Synchronous Overlap and Add (PSOLA) and its variants [9]. Another way of classifying vocoders may be on the basis of speech models namely, the source-filter and perception models. The class of source-filter model includes the LP related vocoder, cepstrum and sinusoidal model based vocoder. The LPC based analysis-synthesis may yield a very low data rate with respect to speech coding. It reduces the computational complexity and produces more natural synthetic speech. Further, the homomorphic vocoders [10], [11] are used for de-convolution of vocal tract and glottal parameters from the speech signal. The cepstrum vocoders work on the principle of homomorphic decomposition. The models based on human auditory system are the perception based models such as Mel Cepstrum Envelope (MCEP) and the HNM. The MCEP [12] overcomes the drawbacks of cepstrum coefficients and requires the Mel Log Spectrum Approximation (MLSA) [13] filter for synthesis of speech. Subsequently, the HNM has been proposed [14] to provide flexibility for speech modification and synthesis with good quality of synthesized speech. Thus, taking this into consideration, this paper covers implementation of a range of vocoders such as LPC, CC, MCEP-MLSA and HNM Vocoders. Although the vocoders have been part of speech applications for quite some time, not much work has been presented in this direction. Similar approaches have been found in [15], [16], but this paper presents a detailed evaluation and implementation of various vocoders under controlled experimental conditions. Nevertheless, the work may still offer useful insights in terms of: i) resemblances and dissimilarities between various vocoders; ii) parameters that affect the quality of speech; iii) most suitable vocoder in case of naturalness.

The paper is organized as follows: Section II describes the implementation of LPC, its analysis and synthesis.

Manuscript received March 10, 2014; revised May 6, 2014.

doi: 10.12720/ijsps.2.1.17-22

Section III comprises of Complex Cepstrum based analysis-synthesis. MCEP-MLSA based vocoder is presented in Section IV. Section V consists of HNM employed for analysis-synthesis process. The database and comparative performance using objective and subjective evaluations are discussed in Section VI. Lastly the section VII lists the concluding remarks and discussion of results.

II. Linear Prediction Analysis-Synthesis

A highly accurate analysis-synthesis scheme is LPC Vocoder [17]-[19] which is widely used due to its simplified architecture and quality of synthesized speech. For low-bit-rate speech coding applications, the LPC parameters are generally used to encode the spectral envelope. The LPC parameters form a perceptually attractive description of the spectral envelope since they describe the spectral peaks more accurately than the spectral valleys [20]. As a result, they are used to describe the power spectrum envelope not only in LPC-based coders [21], but also in some coders which are based on entirely different principles [22]-[24]. Due to issues of quantization, stability and independence of vocal tract and glottal excitation, LPC parameters are converted into LSF (Line Spectral Frequencies) which overcome these limitations leading to comparatively far better results [25]. In this work, the input speech signal is pre-processed and segmented in 30msec frame with 50% (i.e. 15msec) overlapping frames. Each frame is multiplied by hamming window which smoothness the signal and removes artifacts that will be generated during reconstruction. The LPC analysis can be represented using an all pole filter followed by an error prediction filter as shown in Fig. 1. The LPC analysis is fed to synthesizer to reconstruct the speech signal.

The predicted speech sample s(n_i) is given as

$$s(n_i) = \sum_{p=1}^{P} c_p s(n_i - p) + A\, x(n_i) \qquad (1)$$

where n_i is the discrete time instant, x(n_i) is the glottal excitation signal, c_p is the linear prediction co-efficient and p is the order of LPC filter. The synthetic speech is

$$\hat{s}(n_i) = \sum_{p=1}^{P} c_p s(n_i - p) \qquad (2)$$

The predicted error is

$$\varepsilon(n_i) = s(n_i) - \hat{s}(n_i) = s((n_i) - \sum_{p=1}^{P} c_p s(n_i - p) \qquad (3)$$

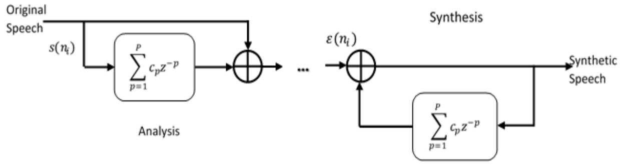


Figure 1. LPC analysis-synthesis

Generally, the order of LPC coefficients is taken as two coefficients per formants. In this work, we used the Akaike Information Criteria (AIC) [26] to compute the order of LPC as 16.

III. Complex Cepstral Analysis-Synthesis

Cepstral analysis-synthesis scheme follows the principle of homomorphic decomposition that the speech signal is a convolution of vocal tract filter response with an impulse excitation. Thus through the process of liftering, a simple and robust parametric approach is obtained which can be employed to extract fundamental frequency of speech while they show some limitations in formant estimation validating the use of LPC in case of estimation of formants. The Cepstrum may be real or complex. The real cepstrum has an infinite impulse response with a minimum phase that discards the glottal flow information of the speech and only the magnitude is considered. This contradicts to work presented by [27], [28] who suggests that the speech signal comprises of both minimum as well as maximum phase indicating that phase too contains information. Unlike the real cepstrum, the complex cepstrum vocoder takes into account the phase along with magnitude of the speech signal. This results into a stable, finite impulse response with a mixed phase vocoder. [6] has shown that the Complex Cepstrum Vocoder can be certainly used in speech processing applications like Speaker Modification and outperforms the real cepstrum vocoders. The CC co-efficient is given as

$$c_c(m) = IFFT\{\log\{FFT\{s(n)\}\}\} \qquad (4)$$

where $s(n)$ = Original Speech $c_c(m)$= Complex Cepstrum Coefficients, FFT and IFFT are the Fourier and Inverse Fourier Transform respectively

$$s_s(n) = IFFT\{exp\{FFT\{c_c(m)\}\}\} \qquad (5)$$

where $s_s(n)$ = synthetic speech signal.

The Fig. 2 shows block diagram of Complex Cepstrum based Vocoder. The input speech signal is pre-processed and segmented in 30msec frame with 50% (i.e. 15msec) overlapping frames. Each frame is multiplied by hamming window which smoothens the signal and removes artifacts that will be generated during reconstruction. The order of FFT is chosen to 1024.

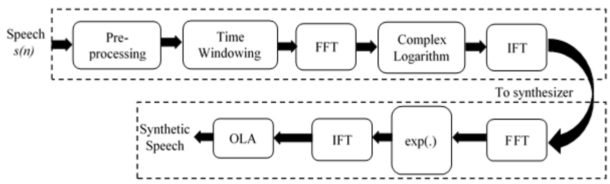


Figure 2. Complex cepstrum vocoder

Although the complex cepstrum overcomes the limitations of LPC vocoder, it is highly complex and has a higher order than the conventional LPC Vocoder.

IV. Mel-Cepstral Envelope-Mel Log Spectrum Approximation Analysis-Synthesis

The higher order of cepstral analysis-synthesis leads to computational complexity which is overwhelmed by using an extension to cepstrum on Mel-scale, termed as Mel Cepstral Coefficient [12]. The log spectrum on a Mel frequency scale is considered to be a more effective representation of the spectral envelope of speech than that on the linear frequency scale. The Mel cepstrum envelope which is defined as the Fourier transform of a spectral envelope of the Mel log spectrum has a comparatively low order; hence it is an efficient parameter. The Mel cepstrum also has the same good features as those of the conventional cepstrum. The MLSA filter is used for cepstrum synthesis on the Mel scale [13]. It has the advantages of low coefficient sensitivity and an improvement in quantization of coefficient. Pitch parameter (F0) is obtained by using peak picking algorithm for the upper quefrency cepstrum.

Fig. 3 shows MCEP-MLSA based vocoder. In the analysis step, MCEPs and the fundamental frequency (F_0) is derived for every 15 msec duration with 30% overlapping. As per [12], the frequency warping factor is taken as $\alpha = 0.375$ with filter order $(M + 1)$ as $M = 11$ and the quantization width q as 0.25. In synthesis step, the MLSA filter gives a highly precise approximation with third order modified Pade approximation 0.024 (0.2 dB) [12].

The MCEP-MLSA vocoder yields same quality speech synthesized at 60-70 % of data rates in the conventional cepstral vocoder or the LPC vocoder.

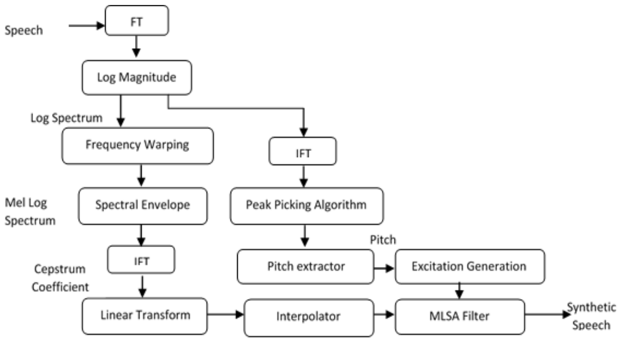


Figure 3. MCEP-MLSA vocoder

V. Harmonic-Noise Model Analysis-Synthesis

The HNM decomposes the speech signal into harmonic and noise part where the harmonic part accounts for the periodic structure of the speech signal and the noise part accounts for the non-periodic structure of the speech signal such as fricative noise, period to period variation of the glottal excitation [3], [14]. HNM has a capability of providing high quality speech synthesis and prosodic modifications. One main drawback of this model is its complexity.

Thus speech signal is given as

$$s(n) = h(n) + e(n) \tag{6}$$

where h(n) is the harmonic part while e(n) is the noise part.

$$h(n) = \sum_{m=1}^{M} G_m(n) \cos(m\theta(n)) + \gamma(n) \tag{7}$$

where $G_m(n)$ is the amplitude of m^{th} harmonic, $\theta(n) = \int_{-\infty}^{n} w_0(p)dp$ is the phase of the m^{th} harmonic, $w_0(p)$ is the instantaneous frequency and $\gamma(n)$ is the residual signal. The harmonic part is simply subtracted from the speech signal to yield the noise part. Fig. 4 shows the HNM analysis and the Fig. 5 shows HNM synthesis.

The maximum voiced frequency and the Pitch are estimated in the HNM analysis for every 10ms frame. The window length is dependent on minimum fundamental frequency. The voiced and unvoiced detection is carried out by assuming the threshold value to 5dB. The noise estimation is performed by an AR filter with an order of 10. During the synthesis, the amplitude, phase and frequency are linearly interpolated along with phase un-warping. The HNM suffers from an inter-frame incoherence between voiced frames when frames are concatenated as they are considered independent of position of glottal closure instants [4]. This issue can be resolved by post analysis like cross correlation function to estimate phase mismatches [4].

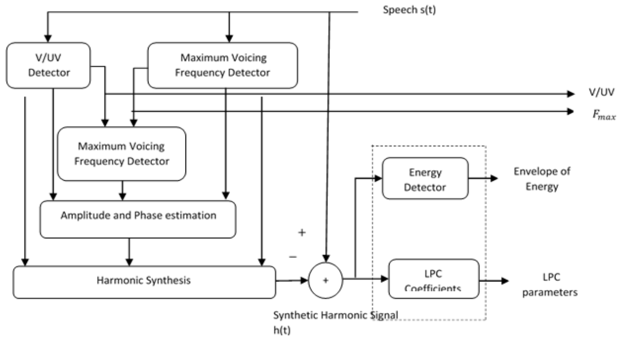


Figure 4. HNM analysis

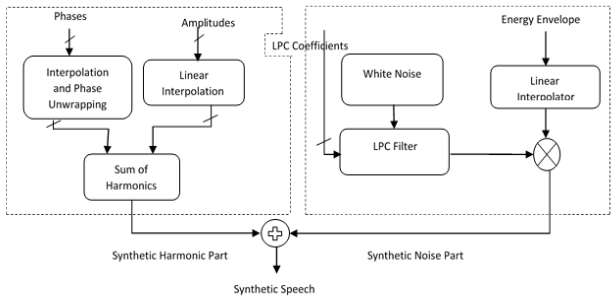


Figure 5. HNM synthesis

VI. Database and Experimental Results

For the evaluation of mentioned vocoders, the CMU-ARCTIC corpus is used [29]. The experimental training set includes phonetically balanced English utterances of seven professional narrators. The utterances in this database are sampled at 16 kHz. The corpus includes sentences of JMK (Canadian Male), BDL (US Male), AWB (Scottish Male), RMS (US Male), KSP (Indian Male), CLB (US Female) and SLT (US Female).

In order to evaluate the comparative performance of discussed vocoders the objective measures, such as Mel Cepstral Distortion (MCD), Log Spectral Distortion (LSD) and Signal to Noise Ratio (SNR) are computed. The end user of the vocoder system is a human listener, hence subjective perception is essential to confirm the objective measures. The subjective measures include rating the system performance in terms of similarity and quality of the resynthesized speech signal.

A. *Log-Spectral Distortion*

The LSD is used to find the closeness between the two speech signals. It is computed as Root Mean Square (RMS) value of the difference of the LP-log spectra of the synthesized speech and original speech signal. The frame durations is 25ms long with 60% (15ms) overlapping between the adjacent frames [30]. The RMS value of the difference between linear predictive spectra of the original speaker speech (s_n) and synthesized speaker speech (s_c) in the frame is defined as

$$\mathrm{LSD} = \frac{1}{J}\sum_{l=0}^{J-1}\left\{\frac{1}{\frac{N}{2}+1}\sum_{k=0}^{N/2}((\log(s_c) - \log(s_n))^{2}\right\}^{0.5} \mathrm{dB} \quad (8)$$

where, N is the frequency bin.

In the computation of LSD, 30 different samples of different Male and Female speakers of ARCTIC database are considered. Fig. 6 shows the LSD based comparative performance of the LPC, CC, HNM and MCEP-MLSA vocoders. The results reveal that the performance of the LPC and Complex Cepstrum vocoders are consistent.

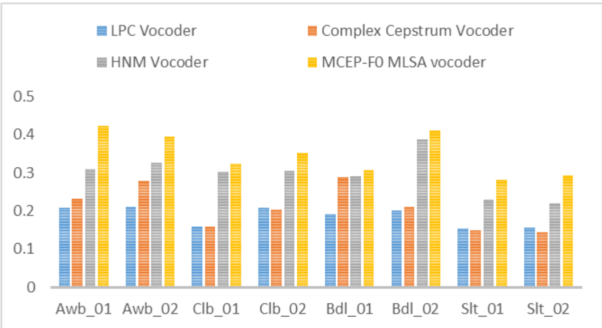


Figure 6. LSD between original and synthesized speech samples of mentioned vocoders

B. *Mel Cepstral Distortion*

Along with LSD, the Mel Cepstral Distortion (MCD) is also used as an objective error measure, which is known to have correlation with subjective test results. The MCD between the synthesized speech and original speech is calculated as [31]

$$MCD[dB] = \frac{10}{\ln 10}\sqrt{\sum_{i=1}^{D} mcc^{ta}{}_i - mcc^{tr}{}_i} \quad (9)$$

where $mcc^{ta}{}_i$ and $mcc^{tr}{}_i$ are the i^{th} Mel Cepstrum Coefficients (MCC) of the original and synthesized speech respectively and D is the order of MFCC features. The zeroth term is not considered in MCD computation as it describes the energy of the frame and it is usually copied from the source. In these experimentation 30 samples of two Male and Female each are considered. Among these the MCD of eight samples are shown in the Fig. 7 with multiple shades for individual vocoder scheme.

C. *Signal to Noise Ratio*

The SNR in dB is the ratio of signal energy to the energy of noisy speech [30]. It is defined as

$$SNR_{dB} = 10\, log \frac{\sum_n s(n)^2}{\sum_n [s(n) - s\prime(n)]^2} \quad (10)$$

where s(n) is original speech and s'(n) is the synthetic speech. The original and synthetic signal must be synchronized as the SNR value is highly sensitive to alignment of both signals.

Fig. 8 shows the signal to noise of various vocoding techniques. Due to susceptibility to noise, the SNR may not be as high as possible for analysis-synthesis method.

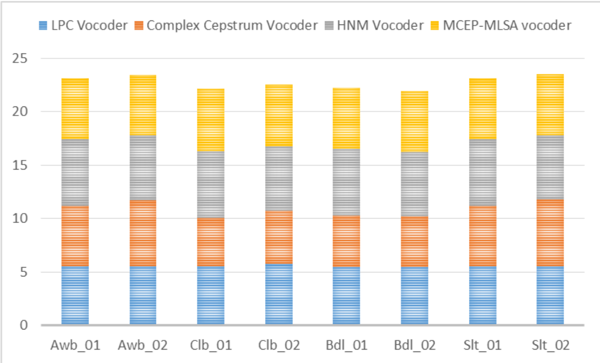
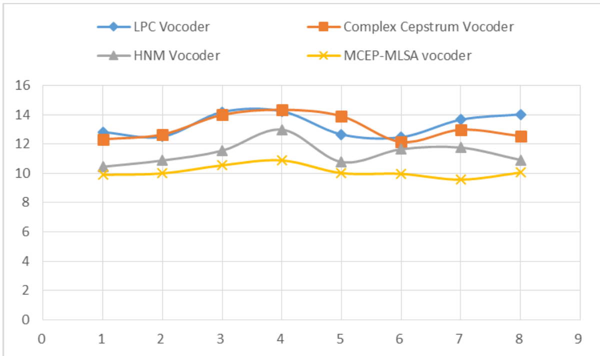


Figure 7. MCD based objective test for various vocoders



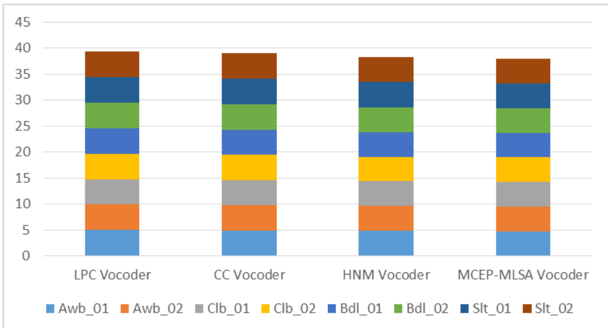


Figure 8. SNR curve for multiple vocoders

Figure 9. MOS test for vocoders

D. *Subjective Test*

The effectiveness of the algorithm can be evaluated using different subjective listening tests. The subjective tests are used to determine the closeness between the synthesized and original speech sample. Thirty synthesized speech utterances for each of vocoder and the corresponding original utterances were presented to twenty non-professional listeners. They were asked to judge their comparative performance with corresponding source and target on a scale of 1 to 5; where rating 5 specifies an excellent match between the transformed and target utterances, rating 1 indicates a poor match between the original target utterance and the transformed utterance and the other ratings indicate different levels of variation between 1 and 5. The ratings given to each set of utterances were used to calculate the Mean Opinion Scores (MOS) [32] for the mentioned vocoders and the results are shown in Fig. 9 with various colour bands indicating their respective scores piled up one after the other. The obtained MOS results show that the synthesis was effective, if the LPC vocoding scheme is employed with similar results from CC vocoder.

VII. Conclusion

In this paper we compare the performance of various vocoders namely, LPC, Complex Cepstrum, Harmonic Noise Model and MCEP-MLSA Vocoders. Evaluation of synthesized speech in terms of quality and naturalness is performed by experimental analysis. Various objective measures such as LSD, MCD and SNR are used. Along with these, the subjective measure such as MOS is also considered to measure the quality of the synthesized speech with respect to original speech signal. These objective and subjective results show that the performance of the LPC and CC vocoder is consistent for all the speech samples. However, the computational complexity of the complex cepstrum is higher than LPC vocoder. In analysis, the Mel cepstrum envelope is more robust with less computational complexity but in synthesis it loses pitch and phase of the speech signal. The results of this experiment is not stretched in all possible ways to yield very accurate answers but are precise about the performance of each individual vocoder. Lastly, the HNM vocoder although very popular for speech synthesis works profoundly well in case of highly periodic signals but in fact signals are rarely perfectly periodic in nature. It is also true that the sampling rate of speech signal affects the HNM performance. Hence there is a slight degradation in speech quality due to roll off characteristics at higher sampling rates.

Acknowledgment

The authors wish to thank Prof. Mukesh A. Zaveri, SVNIT, Surat, India for his encouragement and his continuous support during this work. The authors are grateful to all the listeners who helped in perceptual test during the research.

References

[1] A. S. Spanias, "Speech coding: A tutorial review," *Proc. of the IEEE*, vol. 82, no. 10, pp. 1541-1582, 1994.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109-1121, 1984.

[3] Y. Stylianou, O. Cappé, and E. Moulines, "Continuous probabilistic transform for voice conversion," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 131-142, 1998.

[4] Y. Stylianou, "Applying the harmonic plus noise model in concatenative speech synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 1, pp. 21-29, 2001.

[5] H. Kuwabara and Y. Sagisaka, "Acoustics characteristics of speaker individuality: Control and conversion," *Speech Communication*, vol. 16, no. 2, pp. 165-173, 1995.

[6] J. H. Nirmal, S. Patnaik, M. A. Zaveri, and P. H. Kachare, "Complex cepstrum based voice conversion using radial basis function," *ISRN Signal Processing*, vol. 2014, 2014.

[7] J. H. Nirmal, M. A. Zaveri, S. Patnaik, and P. H. Kachare, "A novel voice conversion approach using admissible wavelet packet decomposition," *EURASIP Journal on Audio, Speech, and Music Processing*, no. 1, pp. 1-10, 2013.

[8] H. Kawahara, I. Masuda, and A. Katsuse de Cheveigné, "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based f0 extraction possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, no. 3, pp. 187-207, 1999.

[9] H. Valbret, E. Moulines, and J. P. Tubach, "Voice transformation using PSOLA technique," *Speech Communication*, vol. 11, no. 2, pp. 175-187, 1992.

[10] A. V. Oppenheim, "Speech analysis - synthesis system based on homomorphic filtering," *Journal of the Acoustical Society of America*, vol. 45, no. 2, pp. 458-465, 2005.

[11] C. J. Weinstein and A. V. Oppenheim, "Predictive coding in a homomorphic vocoder," *IEEE Transaction*, vol. AU-19, pp. 243-248, Sep. 1971.

[12] S. Imai, "Cepstral analysis synthesis on the mel frequency scale," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'83*, 1983, pp. 93-96.

[13] S. Imai, T. Kitaniura, and H. Takeya, "A direct approximation technique for log magnitude response for digital filters," *IEEE Trans.*, vol. ASSP-25, pp. 127-133, Apr. 1977.

[14] Y. Stylianou, "Harmonic plus noise Model for speech, combined with statistical methods, for speech and speaker modification," Ph.D. Thesis, 1996.

[15] M. Airaksinen, "Analysis/Synthesis comparison of vocoders utilized in statistical parametric speech synthesis," Master thesis, Aalto University, Nov. 2012.

[16] Q. Hu, *et al.*, "An experimental comparison of multiple vocoder types," in *Proc. 8th ISCA Speech Synthesis Workshop*, Barcelona, Spain, 2013, pp. 177-181.

[17] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *JASA*, vol. 50, no. 2, pp. 637-655, 1971.

[18] T. Irino, R. D. Patterson, and H. Kawahara, "An auditory vocoder resynthesis of speech from an auditory Mellin representation," in *Proc. EAA-SEA-ASJ, Forum Acusticum Sevilla HEA-02-005-IP*, Sevilla, Spain, 2002.

[19] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 3-14, 1993.

[20] L. R. Rabiner and R. W. Schafer, "Introduction to digital speech processing," *Foundations and Trends in Signal Processing*, vol. 1, no. 1, pp. 1-194, 2007

[21] B. S. Atal, "High-quality speech at low bit rates: multi-pulse and stochastically excited linear predictive coders," in *Proc. International Conference on Acoustic Speech Signal Process*, Tokyo, 1986, pp. 1681-1684.

[22] P. Kroon and E. F. Deprettere, "A class of analysis-by-synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 Kbit/s," *IEEE Journal on Selected Areas Communication*, vol. 6, pp. 353-363, 1988.

[23] H. l. Yang and R. Boite, "High-quality harmonic coding very low bit rates," in *Proc. International Conference on Acoustic Speech Signal Processing*, Adelaide, 1994, pp. I181-I184.

[24] R. J. McAulay and T. F. Quatieri, "Sinewave amplitude coding using high-order all pole models," in *Signal Processing VII, Theories and Applications*, M. Holt, C. Cowan, P. Grant, and W. Sandham, Ed., Amsterdam: Elsevier, 1994, pp. 395-398.

[25] J. H. Nirmal, S. Patnaik, and M. A. Zaveri, "Line spectral pairs based voice conversion using radial basis function," *International Journal on Signal and Image Processing*, vol. 4, no. 2, pp. 26-33, May 2013.

[26] J. Rissanen, “Order estimation by accumulated prediction errors,” *Journal of Applied Probability*, pp. 55-61, 1986.
[27] T. F. Quatieri Jr., “Minimum and mixed phase speech analysis synthesis by adaptive homomorphic de-convolution,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 4, pp. 328-335, 1979.
[28] R. Maia, M. Akamine, and M. Gales, “Complex cepstrum as phase information in statistical parametric speech synthesis,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’12)*, 2012, pp. 4581-4584.
[29] J. Kominek and A. W. Black, “CMU ARCTIC speech databases” in *Proc. 5th ISCA Speech Synthesis Workshop*, Pittsburgh, 2004, pp. 223-224.
[30] A. B. Kain, “High resolution voice transformation,” PhD diss., Rockford College, 2001.
[31] J. H. Nirmal, P. Kachare, S. Patnaik, and M. Zaveri, “Cepstrum liftering based voice conversion using RBF and GMM,” in *Proc. ICCSP*, Apr. 2013, pp. 470-475.
[32] Y. Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229-238, 2008.

Ankita N. Chadha was born in Nashik, India in 1990. She has received her Diploma and Bachelor of Engineering degree in Electronics and Telecommunication from K. K. Wagh Polytechnic and K.K.W.I.E.E.R., Nashik, India in 2009 and 2012 respectively. She is currently pursuing her Master of Engineering in Electronics from K.J. Somaiya College of Engineering, Mumbai, India. Her areas of interests include Signal, Speech and Image Processing, Adaptive filtering, Multirate signal processing and wavelet transform, Machine vision and applications of speech processing.

Jagannath H. Nirmal received his B.E. and M.Tech. degrees in Electronics Engineering from SGGSIE&T, Nanded, India and VJTI, Mumbai, India in 1999 and 2008 respectively. Currently he is pursuing Ph.D. in Speech Processing at SVNIT, Surat, India. He is the author of many articles in reputed journals and conferences. His main research interest includes Speech. Processing, Patterns Recognition and Classification, Adaptive filtering and Signal Processing.

Pramod Kachare was born in Ahemadnagar, Maharashtra, India, in 1991. He received the B.E. degree in Electronics and Telecommunication engineering from the University of Mumbai, India, in 2012. He has worked in K J Somaiya College of Engineering, Mumbai, India as a lecturer. He is currently pursuing his M.Tech in Electronics and Telecommunication from VJTI, Mumbai, India. His research interests include Speech and Image Processing.