

Speech Coding Development for Audio fixing Using Spectrum Analysis

S. Nageswara Rao¹, C.D. Naidu², K. Jaya Sankar³

Assistant Professor, Department of ECE, Sri Venkateswara Engineering College, Suryapet¹
Professor, Department of ECE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad²
Prof. & Head, Department of ECE, Vasavi College of Engineering, Hyderabad³

Abstract

A new method for the enhancement of speech signals contaminated by speech-correlated noise, such as that in the output of a speech coder, is presented. This module is based on numerical speech processing algorithms which modelise the infected ear and generates the stimulus signals for the cilia cells (brain). The method is also based on constrained optimization of a criterion. This interface uses a gamma chirp filter bank constituted of 16 band pass filters based on IIR filters. The implemented method is on a block by- block basis and uses two constraints. A first constraint ensures that the signal power is preserved. A modification constraint ensures that the power of the difference of the enhanced and unenhanced signal is less than a fraction of the power of the unenhanced signal. The applied method is to increase the periodicity of the speech signal. Sounds that are not nearly periodic are perceptually unaffected by the optimization because of the modification constraint. The results demonstrated a degree of discrimination and interferences between different sounds especially in multi speaker environment.

Keywords

Frequency Analysis, speech coding, human auditory system, filtration, pole placement.

1. Introduction

Many studies developed auditory models such as Flanagan model which is based on the physiological data measured by Bekesy, and a mathematical-computational model for the auditory mechanism. Most hearing deficiencies of the human auditory system affect the internal ear (cochlea) and then requires a specific cochlea implant [15]. It is common practice to reduce audible speech-dependent (often called speech-correlated) noise in the output of speech-coding algorithms [1]. Such enhancement systems can be motivated using high-rate source-coding theory for stationary Gaussian processes with a mean-squared-error distortion criterion [2]. The power spectrum of

the ideal reconstructed signal equals the power spectrum of the original signal minus the mean squared quantization error. This means that the decrease in the signal power spectrum is proportionally strongest in regions of low energy. It is based on the conversion of the vocal message in electric impulses to the stimulation of the nerve cells. This prosthesis is composed of on a speech processing module which models and replaces the internal ear and an electronic interface for transmission, wave generation and signal reception [1]. In speech-coding algorithms, the analysis and synthesis models are usually identical and quantization errors often lead to a de-emphasis of the spectral shape. Thus, the results of source coding theory for Gaussian signals suggest an emphasis of the spectrum of the reconstructed signal by means of an adaptive post-filter. The action of the active outer hair cells is model by Automatic Gain Controls (AGC) which simulated the dynamic compression of the intensity range on the basilar membrane. In 1986, [10] developed a model of inner hair cells based on its physiology. This model was used by [12] to include a stage of neural transduction using the meddis hair cell. Finally, [4] has presented, a new model using a temporal speech analysis based on a gamma tone and gamma chirp filter bank decomposition. This qualitative statement remains correct if masking is considered. For good performance, the coder and the adaptive post-filter are generally optimized as a complete system. Post-filtering, which was originally heuristically motivated, and then leads to a coding structure that resembles the optimal coding structure for Gaussian signals? Post-filters for speech coders can be traced back to the work of Ramamurthy and Jayant [7], who introduced an adaptive post filter structure for the enhancement of coded speech. Chen and Gersho [8, 5] introduced the now ubiquitous adaptive pole-zero post-filter structure. The spectral fine structure offers particular large potential for enhancement because of the large dynamic range of the harmonic structure of voiced speech. However, this potential for enhancement of the fine-structure is difficult to achieve because of implementation problems. Conventional adaptive post-filters are based on the coding parameters, and contain no feedback of the properties of the enhanced signal other than the signal power. This generally leads to a spectral emphasis that is too

strong or too weak within different segments of a signal. Furthermore, the time synchronization between the spectral envelope and the spectral fine structure is generally incorrect in current fine-structure post-filters [9] because the inherent delay of the post-filter is ignored. In the present paper, we propose a robust speech-enhancement procedure to reduce speech-dependent noise based on constrained optimization. The new technique avoids the problems of current post-filters in the enhancement of spectral fine-structure of speech.

2. The Speech Processing Algorithm

The principle of our speech processing strategy is given by Figure 1.

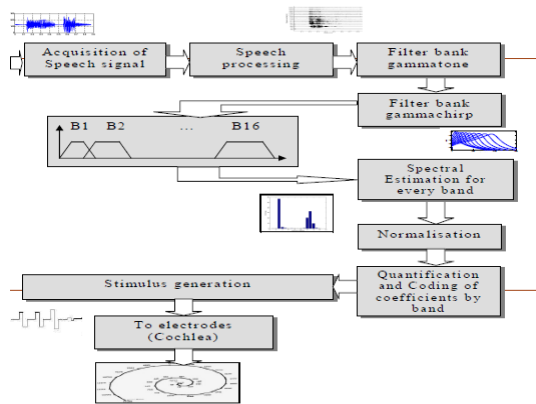


Fig. 1: The Speech Processing Algorithm.

The voice signal processing and coding algorithms are based either on temporal or on frequency representation and modeling of the human auditory system. After signal pre-emphasis and segmentation into overlapped hamming windows, our developed algorithm uses gamma tone filter bank decomposition. Each signal of the sixteen outputs is analyzed in order to compute its energy and envelope. The most significant bands (3 to 5) are selected to be coded according to CIS strategy and then transmitted to the basilar membrane electrodes.

Constrained Enhancement: Emphasis of the coded speech spectrum should lead to signal enhancement for almost all currently-used coding structures since they employ identical analysis and synthesis models and suffer from spectral deemphasize resulting from quantization. Emphasis of the spectrum can be achieved by constrained optimization of suitable measures. In particular, the decreased periodicity

observed in coded speech can be mitigated by constrained optimization of a periodicity measure.

2.1. The Constraints

Let x_j be a discrete speech segment of dimension K , with time label j . That is, x_j is a sample sequence consisting of K subsequent speech samples. It will assume that K is sufficiently large that averaging over K is meaningful. Furthermore, let \tilde{x}_j be the sample sequence that replaces x_j upon enhancement. It is reasonable and customary to make sure that the Euclidean norm of x_j is preserved in the enhancement procedure:

$$\|\tilde{X}_j\| = \|X_j\| \quad (1)$$

In our optimization-based enhancement procedure, the signal-norm preservation condition (1) becomes a first constraint. The signal-norm constraint corresponds to the energy correction made in existing post-filtering procedures. Our optimization based enhancement procedure makes the introduction of additional constraints natural. In particular, we can introduce a second constraint that the difference between x_j and \tilde{x}_j is relatively small:

$$\|X_j - \tilde{X}_j\|^2 \leq \beta \|X_j\|^2 \quad (2)$$

Where $\beta \in [0, 1]$ we refer to this inequality constraint as the modification constraint. The modification constraint prevents that the enhancement procedure modifies the signal more than desirable. In the present paper, we enhance the fine-structure of the speech signal. That is, we perform a constrained maximization of a periodicity measure. It is interesting to consider the effect of the constraints qualitatively. For voiced speech, the signal-norm constraint leads to a simultaneous reduction of energy in the spectral valleys between the speech harmonics and increase in energy of the spectral peaks (harmonics). Since the signal in the valleys has low energy per definition, the modification constraint either is not active or it affects performance of the enhancement procedure little. Thus, the audible enhancement of periodicity is strong. For signal segments that are not nearly periodic, the modification constraint prevents a change in the perceived quality of the signal.

3. Implementation

The proposed a temporal model deduced from the impulse responses measured from the electric impulses of the nervous fibers of the internal ear. [9] proposed a new model of the auditory filter called gamma chirp, to

introduce dependence opposite the level of intensity of resonant hard working stimulus .The impulse response of the gamma chirp filter is given by the following expression [10]:

$$g_c(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_r)t) \times \exp(j2\pi f_r t + jc \ln t + j\phi) \quad (3)$$

Where: n is a filter order, f_r is the modulation frequency of the gamma function, as is the carrier normalization parameter, c is the asymmetry coefficient of the filter, ϕ is the initial phase; BERB is the filter envelope, ERB represents the equivalent rectangular band given by [11, 14]:

$$\text{ERB}(f_r) = 24.7 + 0.108 f_r \quad (4)$$

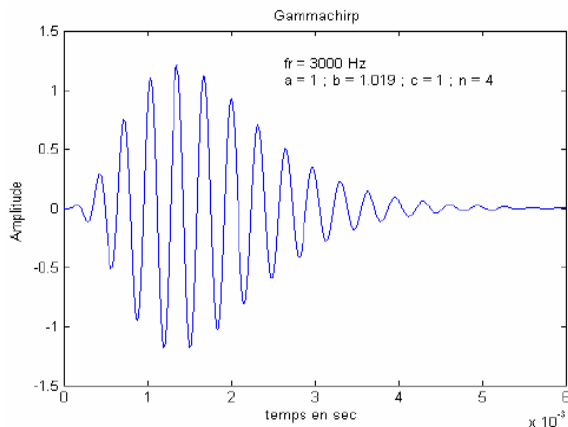


Fig. 2: Temporal impulse response of the gamma chirp filter ($a=1, b=1.019, c=1, n=4$).

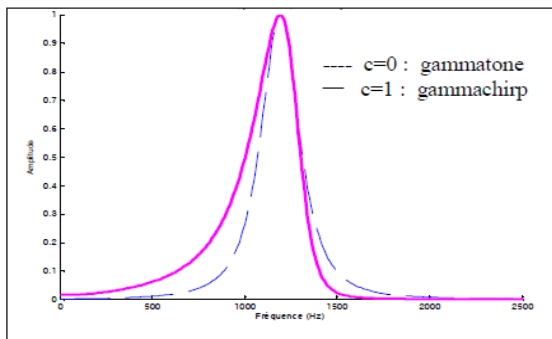


Fig. 3: Frequency response of the gamma chirp filter

The frequency response of the gamma chirp filter can be expressed as:

$$G_c(f) = \frac{a\Gamma(n+jc)}{\Gamma(n)} \frac{\Gamma(n)}{\left[2\pi\sqrt{(b\text{ERB}(f_r))^2 + (f-f_r)^2}\right]^n} e^{c\phi} \quad (5)$$

Figures 2 and 3 represent respectively the temporal impulse response and the frequency response of the gamma chirp filter. The ERB is calculated in function of the central frequency (f_r) according to [11]. If we use the formula and if we suppose that the signal band is between f_H and f_L with a filter recovery ratio (V) hence, the N number of filters is selected like this [12]:

$$N = \frac{9.26}{V} \ln \frac{f_H + 228.7}{f_L + 228.7} \quad (6)$$

However, the central frequencies (f_r) can be premeditated by the expression:

$$f_r = -228.7 + (f_H + 228.7) e^{\frac{Vn}{9.26}}$$

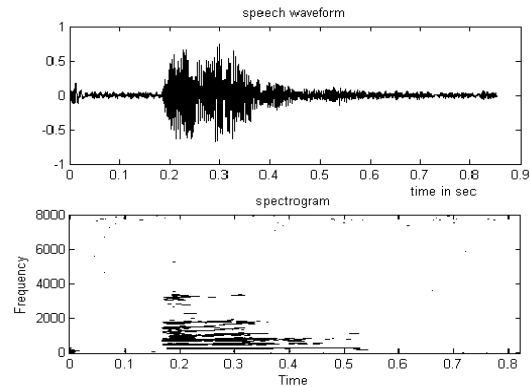


Fig. 4: The speech signal and its spectrogram of the vowel "A": female sound

Figure 4 represents the speech signal and its spectrogram of the vowel /a/ pronounced by a female speaker.

4. Calculation of the Filters Coefficients

The main problem is therefore the calculating of the a_k and b_k coefficients of the GFB. We can use Butterworth, by using Irino model and according to relation (3), the frequency response of the GFB is [7]:

$$\begin{aligned}
 G_c(f) &= \frac{a\Gamma(n+jc)e^{j\phi}}{\left\{2\pi bERB(f_r) + j2\pi(f-f_r)\right\}^{n+jc}} \\
 &= \frac{\bar{a}}{\left\{2\pi\sqrt{b^2+(f-f_r)^2}\cdot e^{j\theta}\right\}^{n+jc}} \\
 &= \bar{a} \cdot \left[\frac{1}{\left\{2\pi\sqrt{b^2+(f-f_r)^2}\right\}^n \cdot e^{jn\theta}} \right] \left[\frac{1}{\left\{2\pi\sqrt{b^2+(f-f_r)^2}\right\}^{jc} \cdot e^{-jc\theta}} \right] \\
 &= \bar{a} \cdot \left[\frac{1}{\left\{2\pi\sqrt{b^2+(f-f_r)^2}\right\}^n \cdot e^{-jn\theta}} \right] \cdot \left[e^{c\theta} e^{-jc \ln\left\{2\pi\sqrt{b^2+(f-f_r)^2}\right\}} \right]
 \end{aligned}$$

The last equation will be written as

$$G_c(f) = G_T(f) \cdot H_A(f)$$

And

$$|G_c(f)| = |G_T(f)| \cdot |H_A(f)| = \frac{1}{\left\{2\pi\sqrt{b^2+(f-f_r)^2}\right\}^n} \cdot e^{c\theta}$$

$$\text{with } \theta = \arctan \frac{f-f_r}{b},$$

$$\bar{a} = a\Gamma(n+jc)e^{j\phi} \text{ and } \bar{b} = bERB(fr) \quad (7)$$

As the Z transform of the GFB can be written as

$$H_c(z) = \prod_{k=1}^N H_{Ck}(z), \quad (8)$$

Hence

$$H_{Ck}(z) = \frac{(1-r_k e^{j\phi_k} z^{-1})(1-r_k e^{-j\phi_k} z^{-1})}{(1-r_k e^{j\phi_k} z^{-1})(1-r_k e^{-j\phi_k} z^{-1})} \quad (9)$$

Where [7]

$$r_k = \exp\left\{-k \cdot p_1 \cdot 2\pi bERB(fr) / fs\right\}$$

IR filter K^{th} pole module

$$\phi_k = 2\pi \left\{ fr + p_0^{k-1} \cdot p_2 \cdot c \cdot bERB(fr) \right\} / fs$$

IIR K^{th} pole argument

$$\varphi_k = 2\pi \left\{ fr - p_0^{k-1} \cdot p_2 \cdot c \cdot bERB(fr) \right\} / fs. \quad (10)$$

p_0, p_1 and p_2 are of the positive coefficients, fs : is the sampling frequency. We can adopt the next values [7]: $p_0 = 2$; $p_1 = 1.35 - 0.19 |c|$ and $p_2 = 0.29 - 0.004 |c|$. By identification of expressions 10 of the GFB with the RII expression 9, we obtain the next coefficients values: $b_0 = 1$, $a_1 = 2 r_1 \cos(\mu_1)$, $b_1 = -2 r_1 \cos\Phi_1$, $b_2 = r_{12} a_2 = -r_{12} \cos\mu_2$ the $r_1, \mu_1, \Phi_1, r_2, \mu_2, \Phi_2, \dots$ values can be deduced from last expressions by putting $k = 1, 2$.

5. Coding Strategy

The spectral estimation of the filter bank output signals is used to extract the stimulus parameters which are: the excited electrodes (or channels) number and their order then the stimulation speed (or spikes) deduced from the channel amplitude or envelope. These parameters once normalized, will be quantized according to a uniform quantification before coding.

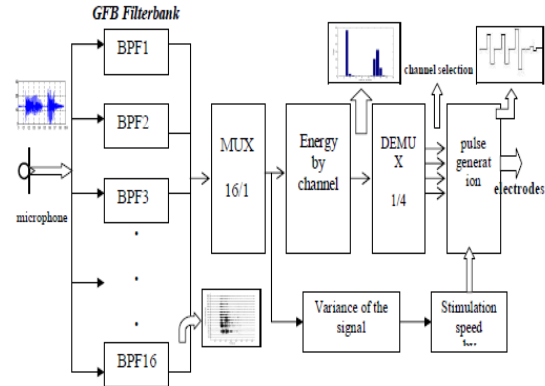


Fig. 5: Coding Strategy

Performance and Discussion

Figure 1 illustrates the operation of the constrained periodicity enhancement procedure. For this example, we set which is a typical value for noisy-sounding coded speech. This value of $_$ corresponds to a signal to modification power ratio of about 13 dB. The enhancement procedure is operating at all time and does not have any information about whether the signal is voiced or unvoiced. The figure shows that, for voiced speech, the audible noise present in the valleys between the signal harmonics is reduced by the enhancement procedure. This is possible because the signal-to-noise ratio is more than 13 dB. On the other hand, the method does not change the reconstructed-signal quality for unvoiced speech because the modification is no more than 13 dB. The continuously changing pitch track also implies that the periodicity is

not enhanced during signal regions that do not contain nearly periodic signals. The perceptual quality resulting from the constrained periodicity enhancement procedure was evaluated with formal and informal testing procedures. In general, for signals with a noisy or rough character, the enhancement obtained from the procedure is immediately clear even from casual listening. The method is also capable of enhancing signals with very high signal-to-noise ratios. Informal tests show that the low-level audible noise that can be heard in a signal encoded with the ITU G.711 standard can be removed with the proposed constrained enhancement procedure.

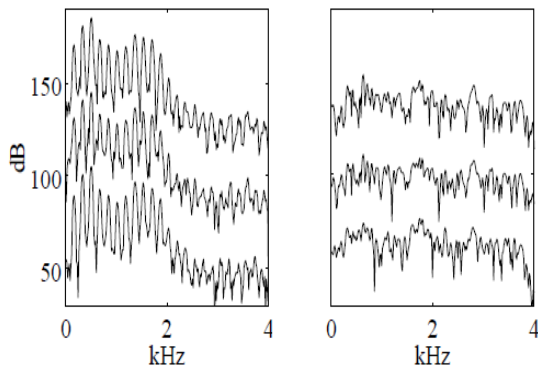


Fig. 6: Illustration of the operation of the enhancer with a maximum modification ratio of 13 dB. On the left power spectra of voiced speech and on the right power spectra of unvoiced speech. Lower spectra are original signals, middle spectra are coded signals, and top spectra are enhanced coded signals.

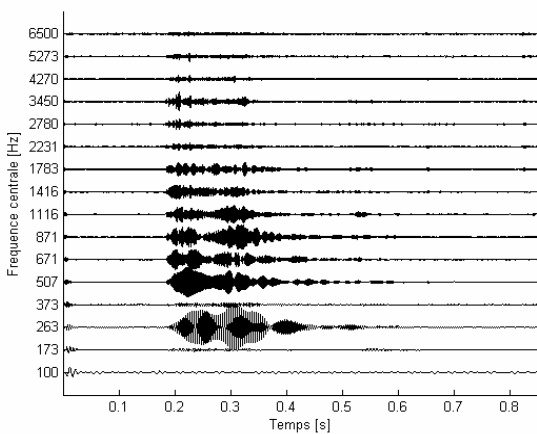


Fig. 7: Filter bank outputs for the vowel /a/ (N=16 channels).

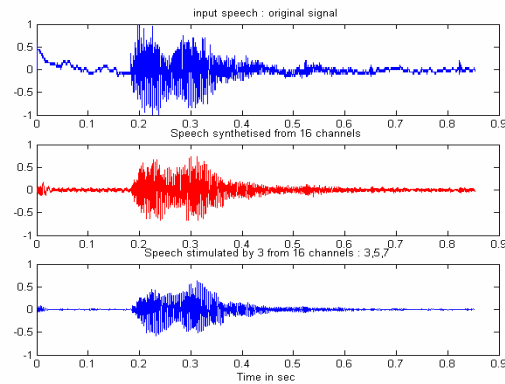


Fig. 8: Speech input and reconstructed speech using 16 then 3 channels for the vowel /a/: female sound.

6. Conclusions

In this study, a new implementation method of speech processing and coding which is intended for cochlea implants. This strategy is based on an adaptive parameters extraction of the speech signal. The first and second parameters are chosen after a spectral energy analysis by channel of the 16 filter bank output signals. However, the last parameter is chosen in function of the envelope and amplitude signal of every stimulated channel and the vocal and acoustic information. An introduced a method for enhancing coded speech based on constrained optimization. The method is very effective at both low and high signal-to-noise ratios and forms an alternative to conventional post-filtering. The method is inherently robust because it cannot introduce large changes to the signal. The described implementation of the method is sufficiently powerful that post filtering of the spectral envelope can be avoided. This means that the fidelity of the spectral envelope is maintained, thus facilitating tendering. This technique is implemented and simulated under MAT lab under several environments and speech database. The simulation results of the stimulation channels and their interferences in different words, demonstrated good discrimination between this information

References

- [1] J. Chen and A. Gersho, "Adaptive post filtering for quality enhancement of coded speech," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 59–71, 1995.
- [2] S. V. Andersen and W. B. Kleijn, "Reverse water-filling in predictive encoding of speech," in *IEEE*

- Speech Coding Workshop, (Porvoo), pp. 105–107, 1999.
- [3] Meddis R., “Simulation of Mechanical to Neural Transduction in the Auditory Receiver,” *Computer Journal of the Acoustical Society of America*, vol. 79, no. 3, pp. 702-711, 1986.
- [4] Patterson D., Allerhand H., and Giguere C., “Time Domain Modelling of Peripheral Auditory Processing: A Modular Architecture and a Software Platform,” *Computer Journal of the Acoustical Society of America*, vol. 98, no. 3, pp. 1890-1894, 1995.
- [5] Irino T. and Patterson D., “A Compressive Gammachirp Auditory Filter for both Physiological and Psychophysical Data,” *Computer Journal of Acoustical Society of America*, vol. 109, no. 5, pp. 2008-2022, 2001.
- [6] Ahmed H., Samet M., Benmessaoud M., Ghriani H., Lakhoua N., Drira M., and Mouine J., “Algorithme de Traitement de la Parole Basé sur un Banc de Filtrés Numériques Programmable dédié à la Prothèse Cochleaire,” in proceedings of the International Electronic Conference Jefferson Township Education Association, Tunis, pp. 159-163, 1998.
- [7] V. Ramamoorthy and N. S. Jayant, “Enhancement of ADPCM speech by adaptive post is filtering,” *AT&T Bell Labs. Tech. J.*, pp. 1465–1475, 1984.
- [8] J. Chen and A. Gersho, “Real-time vector APC speech coding at 4800 bps with adaptive post filtering,” in *Proc. Int. Conf. Acoust. Speech Sign. Process.*, (Dallas), pp. 2185–2188, 1987.
- [9] Irino T. and Patterson D., “Temporal Asymmetry in the Auditory System,” *Computer Journal of Acoustical Society of America*, vol. 99, no. 4, pp. 126-129, 1997.
- [10] Irino T. and Patterson D., “A Time-Domain, Level Dependent Auditory Filter: The Gammachirp,” *Computer Journal of Acoustical Society of America*, vol. 101, no. 1, pp. 412-419, 1997.
- [11] Irino T. and Unoki M., “An Analysis Auditory Filterbank Based on an IIR Implementation of the Gamma chirp,” *Computer Journal Acoustical Society of Japan*, vol. 20, no. 6, pp. 397-406, 1999.
- [12] Waleed A., “Signal Processing and Acoustic Modeling of Speech Signal,” PhD Thesis, 2002.
- [13] W. B. Kleijn, “Improved pitch prediction,” in *Proc. IEEE Workshop on Speech Coding for Telecomm.*, (Sainte-Adele, Quebec), pp. 19–20, 1993.
- [14] Zwicker E. and Feldkeller R., *Psychoacoustique ‘Oreille Recepteur d’ information*, Masson Press, 1981.
- [15] Johannesma P., “The Stimulus Response of Neurons in the Cochlear Nucleus,” *Computer Journal of Symposium on Hearing Theory*, vol. 1, no. 2, pp. 58-69, 1972.



S. Nageswara Rao is currently working as Assistant Professor in SVEC, Suryapet, Andhra Pradesh. He received his M.Tech degree in Computer and Communication (ECE) from the Jawaharlal Nehru Technological University, Hyderabad in 2004. His current research interest is Digital Signal Processing Based Global Non-Parametric Frequency Domain Audio Testing



Dr. C.D. Naidu received B.Tech degree in Electronics & Communication Engineering from JNTU Anantapur, M.Tech in Instrumentation and Control Systems from S.V. University, Tirupathi and Ph.D from JNTU Hyderabad. He is currently working as Principal of VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad.



Dr. K. Jaya Sankar obtained bachelor's degree in ECE from S.V. University, Tirupati in the year 1988. M.E., (Microwave and Radar Engineering) from Osmania University Hyderabad, in year 1994 and Ph.D. degree in 2004 from the same University. After around 4 years of industrial experience, he joined the Department of ECE at Vasavi College of Engineering in 1992. Presently he is Professor and Head of Department. He has around 20 Research Publications to his credit. His areas of interest include Electromagnetics, Antennas, Digital and Wireless Communication and Signal Processing.