

Analysis of Masking Energy for PEAQ Model using Lifting Technique

Jagadeesh B. Kanade
Research Scholar, PRIST
University, Thanjavur
Tamilnadu, India

B.Sivakumar, Ph.D
Prof&Head (TCE), Dr.AIT
Bangalore, Karnataka
India

Nandeesh M.D
Assistant Prof (IT), MSRIT
Bangalore, Karnataka
India

ABSTRACT

Audio compression is the lossy compression technique of converting audio signal into an efficiently encoded bitstream that can be decoded to produce a close approximation of the original signal. For the purpose of improving the coding this work attempts to combine psychoacoustic model for perceptual evaluation of audio quality (PEAQ) in BS.1387 with perceptual audio coder. The implementation of this new psychoacoustic model is done using wavelet lifting method instead of FFT; Later the computation of the masking energy for the different bands will be done.

Keywords

ATH, DWT, FFT, PEAQ, Psychoacoustic, Wavelets, MPEG, SMR, Wavelet Lifting, CELP

1. INTRODUCTION

A number of methods have been proposed for the digital compression of audio signals. Accordingly, audio coders are A number of methods have been proposed for the digital compression of audio signals. Accordingly, audio coders are commonly categorized as either parametric coders or waveform coders. The concept of perceptual audio coding is relevant in the latter case, where auditory perception characteristics are applicable. Parametric coders represent the source of the signal rather than the waveform of the signal. Such coders are suitable for speech signals since accurate speech production models are available. More specifically, the vocal tract is modeled as a time-varying filter that is excited by a train of periodic impulses (voiced speech) or a noise source (unvoiced speech). The parameters that characterize the filter are encoded and then used by the decoder to synthesize speech segments. More advanced parametric coders also include the error signal resulting from the reconstruction using the extracted speech parameters. The error signal generally represents the excitation to the vocal tract filter, as implemented in Code-Excited Linear Predictive (CELP) coders.

On the other hand, waveform coders attempt to accurately replicate the waveform of the original signal. Such coders provide a more perceptually agreeable reconstruction of general audio signals than parametric coders. Efficient waveform coders remove redundancy within the coded signal by exploiting the correlation between signal components, either in time or transform domain. Perceptual waveform coders additionally remove information that is irrelevant to the perception of the signal. The block diagram of a generic perceptual audio coder is illustrated in Figure 1.1. The encoding of the input signal is performed in the upper branch

of the diagram, whereas the lower branch determines the bit assignment per signal component

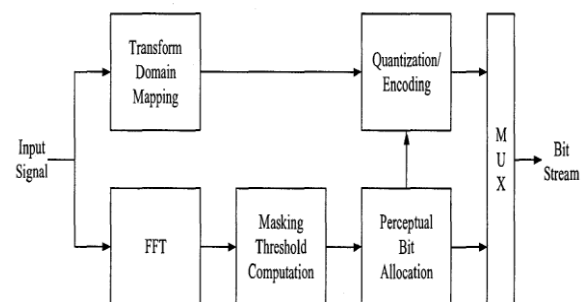


Fig. 1 Basic structure of a perceptual audio coder

A transformation is applied so as to obtain the spectral representation of the input signal. The transformation typically corresponds to a unitary transform or a bank of critically sampled band pass filters. Several Advantages result from encoding the input signal in a transform domain [2]. A masking threshold is computed based on the frequency representation of the signal. More specifically, the Discrete Fourier Transform (DFT) coefficients are used to evaluate the masking threshold. In audio compression standards like MPEG the DFT and Masking threshold computation will be done in perceptual model in generic MPEG audio encoder we can observe block FFT and Masking threshold computation is part of the perceptual model. The figure 1 shows the generic block diagram used in mpeg audio encoders.

The inchoation of wavelets are many and multi-disciplinary. It seems that the conception of wavelet analysis subsisted even before it was first introduced in the mathematical literature (under its current name) in sundry forms and in different fields, including pristine mathematics, quantum physics, geophysics, artificial vision, and signal processing. For example, the Haar wavelet was first proposed as early as 1910 by A. Haar. The designation “wavelet” as we utilize it today derives its meaning from the works of Goupillaud, Morlet, and Grossman from the early 1980’s, albeit the term was utilized by others afore. A number of audio coders predicated on the wavelet have been proposed over the past decade in order to demonstrate the feasibility of such a scheme and to explore sundry configurations that lead to a more preponderant design. The psychoacoustic model was predicated on frequency domain masking; the sundry psychoacoustic models are Psychoacoustic model 1, Psychoacoustic model 2 and PEAQ model. First two psychoacoustic models are utilized in MPEG standards Later

one is utilized for some proprietary audio codecs and it is additionally utilized for the evaluation test of audio contrivances. In this work author is implementing PEAQ model utilizing the different wavelet hoisting schemes and comparing the results with the FFT.

2. PSYCHOACOUSTIC PRINCIPLES

Psychoacoustic models are predicated on psychoacoustic quantifications of the masked threshold. The main question in perceptual coding is: how much noise can be introduced into the signal afore it becomes audible? Answers to this question are derived from psychoacoustic. Psychoacoustic describes the relationship between acoustic events and resulting perceived sounds [2] Masking depends on the spectral composition of both the masker and masked signal and on their variations with time.

Psychoacoustics describes the cognations between the physical representation of sounds and the auditory perception sensations elicited by them.

In other words psychoacoustic gives the cognation between sound stimuli and aurally perceiving perception this cognation includes the effects of both the physiology of the auditory perceiver and the cognitive processing of auditory stimuli. For sound quality evaluation, psychoacoustic magnitudes like loudness or sharpness have proven prosperous [9].

In elongated psychoacoustic studies an optimal sound for a concrete product is modified and it is the responsibility of the engineers to modify the physics of the sound generation in such a way as to arrive at a sound which is as proximate as feasible to the target sound. The psychoacoustic model transforms the time domain input signals into a basilar membrane representation and after this transformation the signals are processed in the frequency domain with the utilization of an expeditious Fourier transform (FFT). A transformation to the Bark scale takes place. The field of psychoacoustics has made consequential progress toward characterizing human auditory perception and particularly the time-frequency analysis capabilities of the inner auditory perceiver. most current audio coders achieve compression by exploiting the fact that “irrelevant” signal information is not detectable by even a well trained or sensitive listener. Impertinent information is identified during signal analysis by incorporating into the coder several psychoacoustic principles, including absolute aurally perceiving threshold (ATH), critical band frequency analysis, simultaneous masking, the spread of masking along the basilar membrane, and temporal masking.

3. WAVELET TECHNIQUES

The standard DWT involves a dyadic tree structure in which the low-channel side is successively split down to a certain depth. We obtain the detail coefficients from the right-leaf node of each level and the approximation coefficients from the left-leaf node at the lowest level. This is illustrated in Figure 2(a) where the nodes represent the wavelet coefficients (at various decomposition stages) and the left and right branches represent the low- and high-pass filtering operations, respectively. If we allow the tree to also split on the right-hand side at each node, then we obtain the more general Discrete Wavelet Packet Transform (DWPT) as shown in Figures 2(b).

In terms of the wavelet basis function (in perpetual domain), the additional degree of liberation in the DWPT comes in the form of frequency, where the standard DWT uses wavelets that are shifted and dilated in time, but the DWPT uses

wavelets that can withal be modified in terms of the number of oscillations in the substructure function.

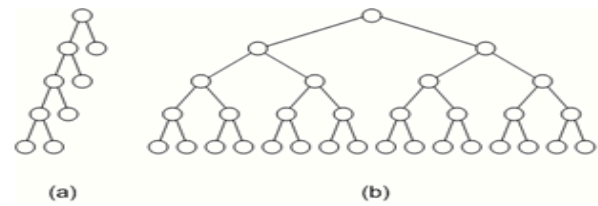


Fig 2: Decomposition structure for (a) DWT (b) DWPT

Wavelet decomposition [4] provides a solution that makes possible for a finer an adjustable resolution of frequencies at high frequencies. This makes adaptation to particular signals [5]. Psychoacoustic model achieves an amended decomposition of the signal into critical bands utilizing the discrete wavelet decomposition transform. This results in a spectral partition which approximates the critical band distribution much more proximate than afore. Furthermore, the Masking thresholds are computed entirely in the Wavelet domain to get approximation of critical bands utilizing Wavelets analysis should meet the spectral resolution requisites. The wavelet substratum additionally plays paramount role to satiate temporal resolution of the signal.

The Lifting scheme is an algorithm to calculate wavelet transforms in an efficient way. They are much more flexible and can be acclimated to define a wavelet substructure in an interval. Furthermore all classical wavelets can be engendered by the hoisting scheme. Several exordiums to the lifting scheme are available; The lifting scheme is an incipient method for constructing wavelets. The main difference with classical constructions [1-3] is that it does not rely on the Fourier transform. This way hoisting can be habituated to construct second generation wavelets, wavelets which are not compulsorily translates and dilates of one function. The latter we refer to as first generation wavelets. It sanctions a more expeditious implementation of the wavelet transform. Traditionally, the expeditious wavelet transform is calculated with a two-band subband transform scheme, visually perceive Figure 3. In each step the signal is split into a high pass and low passes band and then sub sampled. Recursion occurs on the low pass band. The hoisting scheme makes optimal utilization of homogeneous attributes between the high and low pass filters to expedite the calculation. The number of operations can be reduced by a factor of two. The lifting scheme allows a fully in-place calculation of the wavelet transform. In other words, no auxiliary memory is needed and the original signal (image) can be replaced with its wavelet transform.

The lifting scheme at its highest level is very simple. The lifting scheme can be performed via two operations: Predict and Update. Suppose we have the one dimensional signal a_0 . Lifting is done by performing the following sequence of operations:

1. Split a_0 into Even₁ and Odd₁
2. $d_{-1} = \text{Odd}_{-1} - \text{Predict}(\text{Even}_{-1})$
3. $a_{-1} = \text{Even}_{-1} + \text{Update}(d_{-1})$

These steps are reiterated to construct multiple scales of the transform. The inverse transformation is simple as well. We only transpose operations and transmuted the denotements. The even and odd sequences are then merged together to compose the pristine Signal. The wire diagram in Figure 3 shows the forward transform visually. The coefficients are represent the averages in the signal, while the coefficients in d represent the

differences in the signal. Thus, these two sets withal correspond to the low- pass and high- pass frequencies present in the signal. To mirror the operations of the wavelet transform, we must formulate the operations Presage and Update.

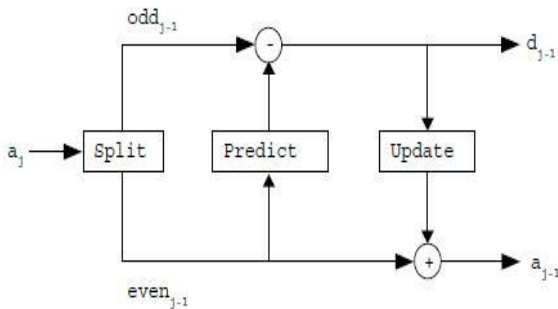


Fig 3: Wired diagram of forward transformation with the lifting scheme [13]

4. PROPOSED DESIGN OF PEAQ MODEL USING WAVELET LIFTING

PEAQ (Perceptual Evaluation of Audio Quality) is the objective quantifications Recommendation Standard of perceived audio quality established by ITU in 1998, which is additionally called BS.1387. It utilizes software to simulate perceptual properties of human auditory perceiver, then integrates multiindices to evaluate subjective quality of test audio PEAQ quantification method models fundamental properties of the auditory system [11]. In MPEG psychoacoustic model 2, SMR (ratio between signal energy and masking threshold) is resolute by experiential value of examination observation. Fundamental version of PEAQ adopted a design approach different from MPEG audio standard. It endeavors to amalgamate physiological structure of human auditory perceiver with masking effect of simple signal represented from examination to find the inherent consequence, and then use mathematic model to emulate the structure of human auditory perceiver. Figure. 4 illustrate the block diagram of this design [10], in which the function of outer/middle auditory perceiver, inner auditory perceiver, and audio perception cognate nerve cell and brain are emulated. This kind of psychoacoustic model could be elongated conveniently to acoustic masking of complex signals.

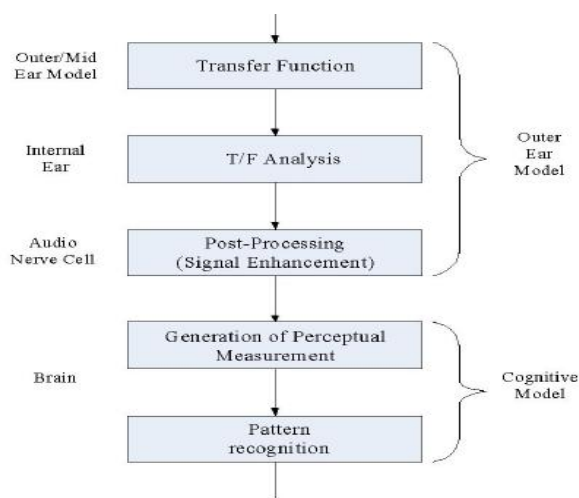


Figure 4: psychoacoustic model based on PEAQ (Basic version) [11]

PEAQ was initially designed to quantify audio quality without special consideration for requisites of audio coder, such as window switching, unification of critical band scale, estimation of masking properties on transient signal, etc. Thus, we must take into account these requisites and redress the psychoacoustic model of PEAQ fundamental version so as that it could be utilized in audio coder. In this work we are performing the analysis with the rudimental PEAQ model utilizing wavelet hoisting method. The signal spectrum will be obtained by applying wavelet packet decomposition instead of FFT and whose connections are culled in such a way that sub bands correspond to the best possible one to the critical bands.

5. PROPOSED DESIGN OF PEAQ PSYCHOACOUSTIC MODEL

The detailed design of the proposed method is as shown in figure 4. The implementation steps as below,

Step 1: Framing

The “.wav” file is actually an uncompressed audio signal, the sampling rate chosen here is 48000Hz. Divide the audio signal into different frames each frame of size 2048.

Step 2: Apply Wavelet Lifting scheme to each Frame

Wavelet lifting scheme is applied for each frame of the incoming data.

Step 3: Outer and Middle Ear Modelling

The outer and middle ear response is modeled by the frequency dependent weighting function[11].The implementation of outer and middle ear modelling is in compliance with ITU-R BS.1387.

Step 4: Frequency Grouping

In this step take a frame of frequency domain samples and group them into the frequency bands these frequency bands will be grouped according to the critical bands. These critical bands are defined by the frequency bins. The frequency grouping is performed in the range from 80 Hz to 18 kHz. The bandwidth is 0.25 Bark in the basic version resulting in 109[11][14][15].

Step 5: Adding of internal Noise

Internal noise is generated by blood flow within the human ear and this will be modeled by adding the frequency dependent Energies of the frequency groups [11][14][15].

Step 6: Frequency Spreading

The spreading function is adopted from an auditory model developed by Terhardt [14][15][16]

Step 7: Time domain spreading

Temporal masking effect is incorporated in time domain spreading. The time domain spreading depends on the previous frame in addition to the current frame. In this model temporal masking effect is considered by means of first-order smooth filtering [11].

Step 8: Calculation of Masking Parameters

The masking threshold is calculated by weighting the excitation patterns [11][14][15].

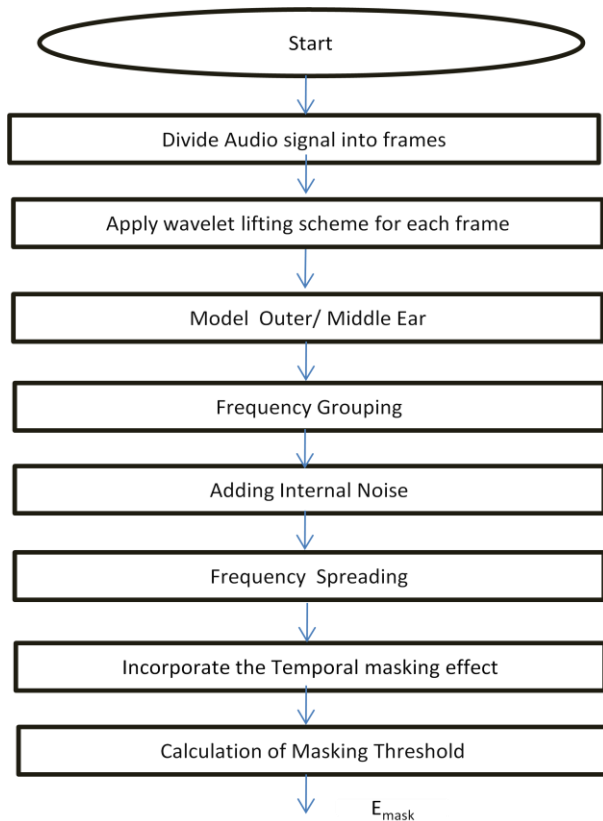


Fig 5: Flow chart for Psychoacoustic model of PEAQ (Basic version) using wavelet lifting method

6. EXPERIMENTAL RESULTS AND CONCLUSION

In this work we have implemented PEAQ based Psychoacoustic model for audio using wavelet lifting scheme instead of FFT to analyze the masking energy of the audio signal. This new PEAQ psychoacoustic model gives more accurate masking threshold compare to PEAQ basic model which uses the FFT for calculating masking threshold. The result of the proposed method is shown in the fig7. Which indicates that masking threshold for critical bands is more compare to FFT method. Table. 1 shows the comparison of estimated time for executing (in seconds) frames using wavelet packet decomposing and wavelet lifting method.

The PEAQ based psychoacoustic model implemented using wavelet packet decomposition takes an account for the critical bands also takes an account of Spectral and temporal masking phenomenon. The specialty of the proposed model is that it estimates masking threshold more accurately compare to masking threshold obtained from FFT as in the standard BS.1387 standard. The results of the proposed models is almost similar the results of the wavelets decomposition techniques [15]. Figure.6 shows the distribution curve of the masking threshold using FFT (PEAQ basic version) and figure.7 shows the distribution curve of masking threshold by wavelet packet decomposition method. Figure .8 indicates the distribution curve of masking threshold by proposed method. From the figure 6, figure.7 and figure 8 it is evident that for the same number of critical bands masking threshold has increased compared to that of FFT method. From figure.7 and figure.8 for the lower bands masking energy high in lifting method compare to the wavelet packet decomposition and for higher critical bands the masking energy less in lifting method

compare to the packet decomposition method. From the table.1 it is evident that wavelet lifting will take less time to compute a frame comparing to the wavelet packet decomposition

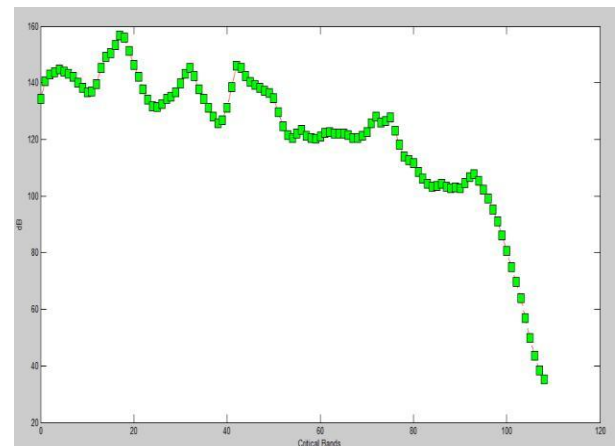


Fig 6: Distribution curve of masking threshold Using FFT (PEAQ Basic Version)

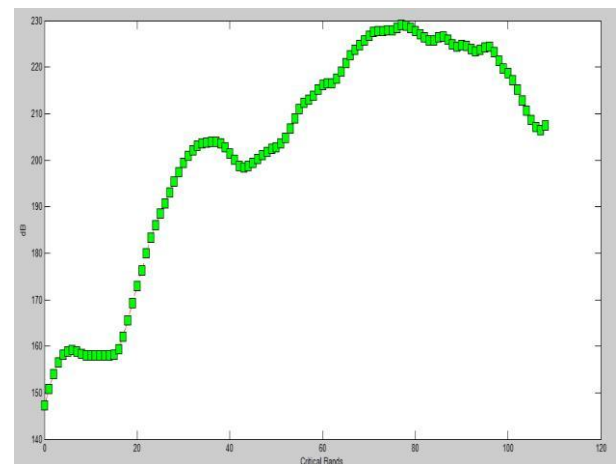


Fig 7: Distribution curve of masking threshold Using Wavelet Packet decomposition (PEAQ Basic Version)

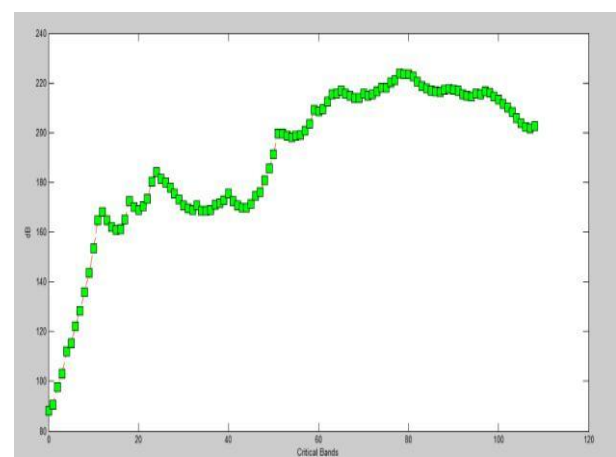


Fig 8: Distribution curve of masking threshold using Lifting scheme (PEAQ Basic Version)

Table 1: Time consumed in seconds

Frames Num	Wavelet Packet Decomposition	Wavelet Lifting
Frame 1:	0.982	0.437
Frame 2:	0.905	0.4222
Frame 3:	0.874	0.437
Frame 4:	0.873	0.421
Frame 5:	0.874	0.421
Frame 6:	0.89	0.422
Frame 7:	0.873	0.437
Frame 8:	0.967	0.406
Frame 9:	0.874	0.422
Frame10:	0.858	0.452

7. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of this work.

8. REFERENCES

- [1] T. Painter and A. Spanias, "Perceptual coding of digital audio," Proc. IEEE Trans.Sig, vol. 88, Apr. 2000, pp. 51-513
- [2] Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s-Part3: Audio, 1999
- [3] D.Sinha and A.Tewfik, "Low Bit Rate Transparent Audio Compression using adapted wavelets", IEEE Trans.Sig.Proc, pp.3463-3479, Dec 1993
- [4] C.S.Burrus, R.A. Gopinath and H.Guo, "Introduction to wavelets and wavelets transforms: A Primer, "Prentice Hall, 1998.
- [5] M. V. Wickerhauser, "Adapted wavelet analysis from theory to software," Wellesley, Massachusetts, 1994
- [6] Zwicker, E., Feldkeller, R., "Das Ohr als Nachrichtenempfänger"research area. Stuttgart: Hirzel Verlag, 1967.
- [7] M. R. Zurera, F. L. Ferreras, M. P. J. Amores, S.M. Basc' on, and N. R. Reyes, "A new algorithm for translating psycho-acoustic information to the wavelet domain," Signal Processing, vol. 81, no. 3, pp. 519–531, 2001.
- [8] B. Lincoln, "An experimental high fidelity perceptual audio coder," Project in MUS420 Win97, March 1998
- [9] Zwicker. Fastl, (1999) Psychoacoustics-Facts and Models.2nd updated edition, Springer Verlag, Heidelberg, New York.

- [10] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," IEEE Transactions on Signal Processing, vol. 47, no. 6, pp. 1622–1635, 1999
- [11] Kabal, P. 2002. An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality. Montreal, Department of Electrical & Computer Engineering, McGill University, TSP Lab technical report. 89 p
- [12] Dermot Campbell, Edward Jones, Martin Lavin " Audio quality assessment techniques—A review and recent developments" Elsevier Signal Processing89 (2009)1489–1500
- [13] Wade Spires "Lossless Image Compression via the Lifting Scheme" University of Central Florida (2005)
- [14] Jagadeesh B.K, Dr.B.Sivakumar "PEAQ based psychoacoustic model implementation using wavelet packet decomposition" Elsevier ACEEE conference Proceedings series -01
- [15] Jagadeesh B.K, Dr.B.Sivakumar "Analysis of PEAQ Model Using Wavelet Decomposition Techniques", ACEEE Int. J. on Recent Trends in Engineering and Technology, Vol. 10, No. 1, Jan 2014
- [16] International Telecommunication Union, Method for Objective Measurements of Perceived Audio Quality, July 1999. ITU-R Recommendation BS.1387.

9. AUTHOR'S PROFILE

Jagadeesh B.K holds the Master degree from VTU, Karnataka, and currently Working as Associate professor, Dept of ECE, HKBKCE, Bangalore, Karnataka INDIA, Before entering in to teaching profession he has worked at several software industries like Toshiba, SASKEN communication and NXP semiconductors. He has worked on Audio, Video codecs and Audio post processing algorithms on different DSP Platforms.

Dr B Sivakumar holds PhD degree from Anna University Tamilnadu, and currently working as Professor and head of the Telecom dept at Dr AIT, Bangalore.**Dr B Sivakumar** holds PhD degree from Anna University Tamilnadu, and currently working as Professor and head of the Telecom dept at Dr AIT, Bangalore.

Nandeesh M.D holds Master degree from VTU, Karnataka, and currently working as Assistant professor, Dept of Instrumentation Technology, MSRIT, Bangalore, Karnataka INDIA.