International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 3, Issue 1, January 2014

A Literature survey on Psychoacoustic models and Wavelets in Audio compression

Jagadeesh B.Kanade[#1], Assist Prof, HKBKCE, Bangalore,

Dr.Sivakumar B[#2] Prof & HOD TC dept, Dr AIT, Bangalore

A.

Abstract— **this paper presents the review of different psychoacoustic models and wavelets techniques used in the area of digital audio compression. This paper includes a brief technical summary of the psychoacoustic models and wavelets for audio compression. Furthermore, this paper outlines recent advancements in the general area of audio quality assessment since the publication of the ITU standard.**

Index Terms— **WT, SMR, Bitrtae, DFT, DCT, AAC, FFT, MPEG, WFB, PEAQ**

II. INTRODUCTION

The development of high-quality audio compression methods [1, 2] have benefited greatly from the successful integration of psychoacoustic models. Audio compression methods try to represent the original audio with as low bit rate as possible. High audio quality is achieved by rendering quantization noise inaudible. The exploitation of psychoacoustic models in the design of audio coders has led to high compression ratios while keeping audible degradation in the compressed signal to a minimum. In the basic approach, the signal is represented on a critical-band scale in the frequency domain and then quantized. Frequency-domain masking properties of the human auditory system are exploited in the quantization process to maximize perceived fidelity of the signal transmitted (or stored) at a given bit-rate. The frequency representation of the signal is typically accomplished using a filter bank implemented as a frequency transform or sub-band filter.

A number of methods have been proposed for the digital compression of audio signals. Accordingly, audio coders are commonly categorized as either *parametric caders* or *wavefarm caders.* 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 signaIs 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) 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 signaIs 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

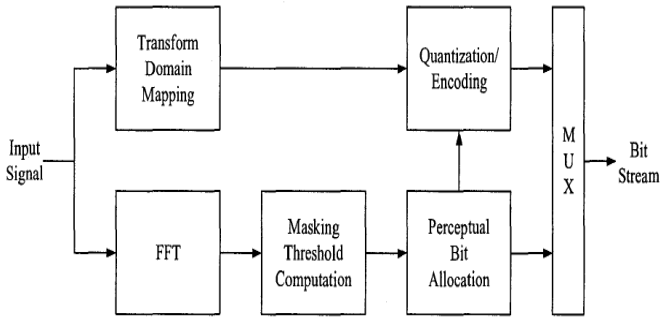

Fig. **1.1** Basic structure of a perceptual audio coder

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 [3]. Firstly, effective transforms compact the information of the signal into fewer coefficients, ensuing in a more efficient usage of quantizes Transform coefficients are less correlated than

temporal samples of the input signal. Secondly, the desired frequency resolution is achievable through judicious selection of the transformation. Auditory masking effects are significantly influenced by the frequency composition of the input signal. As such, transform domain coding is ideal for the application of auditory perception characteristics. The transformation is applied to temporal frames of the input

ISSN: 2278 – 909X

signal during intervals for which the signal is considered stationary. Audio coders typically segment the input signal into frames ranging from 2 ms to 50 ms, depending on the desired temporal and frequency resolution [4].

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. Audio signaIs have complex spectra, composed of multiple masking components. Masking components are extracted from the spectrum of the input signal and individual masking effects are combined to yield an overall masking threshold. Auditory models deliver a masking threshold along with the amount of allowable distortion in the frequency domain. Classical masking applications assume that signal energy lying below the masking threshold is inaudible. As many as 50% of transform coefficients are masked in transform coding of music and speech signaIs [2]. A frequently cited example of masking is the 13 dB miracle. Noise added to an audio signal, having a spectral structure that is adapted to that of the signal, is inaudible for signal-to-noise ratios as low as 13 dB [4]. A common output of the masking threshold computation stage is the *Signal-to-Mask Ratio* (SMR), which represents the ratio of the signal input to the amount of masking produced by the signal. Quantization is defined as the process of transforming the sample amplitude of a message signal into discrete amplitude taken from a finite set of possible amplitudes.

The allocation of information bits to the different Quantizer is performed adaptively, based on the computed masking threshold. Firstly, spectral components lying beneath the masking threshold need not be represented. Such components do not contribute to the perception of the audio signal according to classical masking principles. Secondly, the noise that is introduced by the quantization process is shaped in frequency such that it becomes inaudible. For instance, more noise is allowed where the masking threshold is high, resulting in the allocation of fewer information bits to those regions. The process is referred to as *spectral noise shaping.*

All perceptual coders share a similar structure in that they contain a filter bank, psychoacoustic model, and an encoding and quantization stage, The filter bank stage provides a decomposition of the input signal that makes the application of perceptual criteria possible and also provides some decorrelation of the input signal. Many types of filter banks and time-to-frequency transforms exist where each other's a different set of trade-offs in its design. Many have been considered and explored in the context of audio coding [2, 26] and one in particular called the Wavelet Transform (WT) has shown to be interesting and potentially very useful. The wavelet transform, or more generally the wavelet filter bank (WFB), is an iterated filter bank that provides a flexible way of analyzing a signal at various resolutions and across various frequency regions. This flexibility is especially appealing in audio coding since the WFB can provide an analysis of the input signal according to the critical band (CB) resolution of the inner ear and, more generally, provide a scheme that can adapt to the time-varying nature of the audio signal . However, the WFB has also been found to provide poor localization properties that can be a drawback in audio coding. The application of wavelets in perceptual audio coding, therefore, requires us to explore ways to maximize its benefits and minimize its drawbacks.

The performance of audio coding schemes is evaluated using objective and/or subjective quality measures that compare the coded signal with the input reference signal. Typical objective measures, such as Signal-to-Noise Ratio (SNR) or Mean-Square Error (MSE), do not accurately represent the perceived quality of the reconstructed signal. Assessing the perceptual quality of wideband audio signals is an important consideration in many audio and multimedia networks and devices . This paper gives a brief idea of accessing perctuptual quality of the compressed audio using the various codecs. The detailed study of perceptual quality evolution is done in [5].

II. HISTORICAL OVERVIEW OF AUDIO CODING

Early work on signal compression dates back to the information-theoretic foundation that was laid out by Shannon [3]. Shannon introduced the idea of entropy as a quantity expressing the information content of a signal and showed that a source could be coded with zero error if encoding was done at a bitrate equal to or greater than the entropy of the signal (and with coding delay that approached infinity). An implication of this was that sources with infinite alphabets, such as analog audio, required infinite bitrates for error-free coding.

In practice, however, audio signals are first digitized before any meaningful processing is done. This digitization of a signal from analog to digital domain, typically done through the use of an analog-to-digital (A/D) convertor, can actually be thought of as a coding stage that reduces the entropy of a signal to a infinite level while introducing some distortion or coding noise. The type of coding done at this stage is usually simple and results in a high bitrate so that complexity and coding noise can be minimized, e.g. pulse code modulation (PCM). In order to further reduce the bitrate and still maintain high signal quality, removal of statistical redundancy and perceptual irrelevancy is required [1].

A group of coding algorithms developed early on, commonly referred to as entropy or lossless coders, were designed to exploit the statistical redundancy of the source signal. Although the entropy provided a measure of the bitrate required to encode a signal, practical coders were only able to approach this theoretical limit. Examples of lossless coding schemes developed for both speech and audio have appeared in [1, 4]. Since most of the early coding work was done in speech, wideband audio coding finds its root in speech coding. A number of differences can, however, be noted. Wideband audio generally has a wider sampling range, wider dynamic range, and higher expectation of quality by the listener. In terms of coding, the most notable difference could be the use of a production model in speech coding that leads to highly efficient ways to encode speech signals, whereas nothing similar exists for general audio signals. The most significant advances in audio coding came with the introduction of perceptual coders. Perceptual coders are designed to take advantage of the masking phenomena that occurs in the ear so that coding noise can be introduced in a way that minimizes or eliminates perceived distortion. It has

ISSN: 2278 – 909X

been noted that many of the innovations in perceptual coding came from people closely familiar with audio applications rather than those involved in research, and this has caused the technology and literature of audio coding to evolve somewhat independently [7]. A number of notable examples of perceptual coders are mentioned next.

The earliest examples of perceptual coders were developed in the 1970's by Crochiere [8], Schroeder [9], and Zelinski and Noll [9]. These algorithms utilized a time to frequency transformation stage, e.g. Short-time Fourier Transform in [9] and 4-channel non-uniform filter bank [8] that allowed noise shaping in the frequency domain according to some well known psychoacoustic principles. They were followed by the work of several other people in the 1980's who tried to improve on the choice of the transformation stage, accuracy of the psychoacoustic model, and use of other coding techniques that further improved coding efficiency. Most notable of these were the works by Schroeder [10], Brandenburg [11], Johnston [12], and Mahieux [13]. One in particular, an algorithm called MUSICAM developed by Dehery et al. [14], was adopted in the application of digital audio broadcasting (DAB) in Europe and also became part of the well known MPEG-1 audio coding algorithm. Another algorithm called ASPEC [15] also became part of the MPEG-1 audio coding algorithm as the basis to Layer III. The MPEG-1 audio coding algorithm, perhaps the most well known audio coding algorithm, was developed in the early 1990's through a collaborative effort led by the International Standardization Organization (ISO) [16, 17] and was designed to provide three layers of complexity and performance. Layer I provided the lowest complexity and lowest performance, layer II provided medium complexity and medium performance, and layer III provided the highest complexity and highest performance. Layer III, also commonly referred to as \MP3", became popular and widely used on the Internet. Subsequent development of the MPEG audio coding standard appeared as the MPEG-2 and MPEG-4 standard where several improvements were made over the original algorithm in terms of performance, scalability, and functionality [18, 19]. Variants of the MPEG algorithm outside the standard also appeared from other groups, e.g. MPEGplus and MP3 Pro [20]. Other well known audio coders have appeared more commercially, including the AC-3 family of coders developed by Dolby [21], the ATRAC coder developed by Sony [22], and the PAC coder developed by Lucent (formerly AT&T) [23]. More recently, an open-source and patent-free audio codec called Ogg Vorbis [24, 25] appeared as an alternative to the popular but somewhat proprietary MP3 algorithm. Many of these algorithms are used in a variety of applications that include transmission and broadcasting on the Internet, portable audio players and recorders, and multichannel digital sound system in DVD and movie theatres.

III. PSYCHOACOUSTIC PRINCIPLES

Psychoacoustics is the scientific study of sound perception. More specifically, it is the branch of science studying the psychological and physiological responses associated with sound (including speech and music). It can be further categorized as a branch of psychophysics. Hearing is not a purely mechanical phenomenon of wave propagation, but is also a sensory and perceptual event; in other words, when a person hears something, which something arrives at the ears a mechanical sound wave travelling through the air, but within the ear it is transformed into neural action potentials. These nerve pulses then travel to the brain where they are perceived. Hence, in many problems in acoustics, such as for audio processing, it is advantageous to take into account not just the mechanics of the environment, but also the fact that both the ear and the brain are involved in a person's listening experience. The inner ear, for example, does significant signal processing in converting sound waveforms into neural stimuli, so certain differences between waveforms may be imperceptible. In addition, the ear has a nonlinear response to sounds of different intensity levels; this nonlinear response is called loudness. Telephone and audio noise reduction systems make use of this fact by nonlinearly compressing data samples before transmission, and then expanding them for playback. Another effect of the ear's nonlinear response is that sounds that are close in frequency produce phantom beat notes, or intermediation distortion products. The two major techniques used by psychoacoustic principles are absolute threshold of hearing and masking effect.

Absolute threshold of Hearing

Absolute threshold of hearing (ATH) characterizes the amount of Energy needed in a pure tone such that it can be detected by a listener in a noiseless environment. Absolute threshold is expressed in db sound pressure level (dB SPL). Fig 1.2 shows the ATH curve . The absolute threshold of hearing determines the amount of energy required for people to hear the sound with certain frequency. However, music is created by sound signals each with different frequency and sound pressure level. Each sound signal can influence the neighbor sound signals as well. The quiet (absolute) threshold is well approximated by the nonlinear function

$T_q(f)=3.64(f/1000)^{-0.8}-6.5e^{-0.6(f/1000-3.3)2}+10^{-3}(f/1000)^4$ (dB SPL)

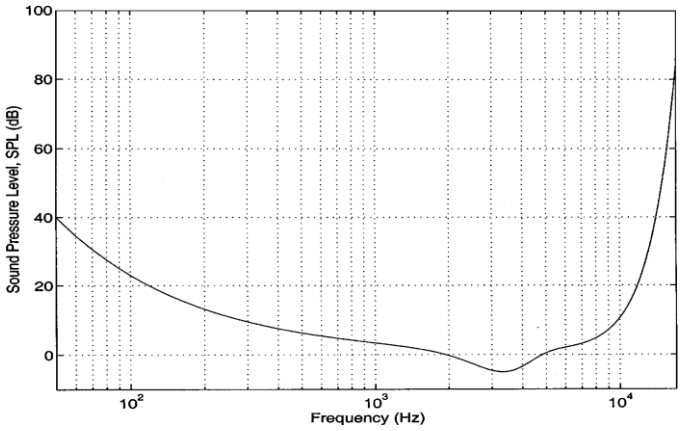

Fig.1-2 The absolute threshold of hearing in quiet

Across the audio spectrum, it quantifies the SPL required at each frequency such that an average listener will detect a pure tone stimulus in a noiseless environment

Masking refers to one sound that is inaudible because of the presence of another strong sound. The masking effect allows us to further remove the sound signals that are masked. By removing the signal, we reduce the size of a music file and achieve a better compression result. Masking can be further

ISSN: 2278 – 909X

divided to simultaneous masking and non-simultaneous masking.

Critical Band

In audiology and psychoacoustics the term critical band, introduced by Harvey Fletcher in the 1940s, refers to the frequency bandwidth of the "auditory filter" created by the cochlea, the sense organ of hearing within the inner ear. Roughly, the critical band is the band of audio frequencies within which a second tone will interfere with the perception of a first tone by auditory masking.

Simultaneous Masking, Masking Asymmetry, and the Spread of Masking

Masking refers to a process where one sound is rendered inaudible because of the presence of another sound. Simultaneous masking may occur whenever two or more stimuli are simultaneously presented to the auditory system. From a frequency-domain point of view, the relative shapes of the masker and maskee magnitude spectra determine to what extent the presence of certain spectral energy will mask the presence of other spectral energy. From a time-domain perspective, phase relationships between stimuli can also affect masking outcomes. A simplified explanation of the mechanism underlying simultaneous masking phenomena is that the presence of a strong noise or tone masker creates an excitation of sufficient strength on the basilar membrane at the critical band location to block effectively detection of a weaker signal. Although arbitrary audio spectra may contain complex simultaneous masking scenarios, for the purposes of shaping coding distortions it is convenient to distinguish between only three types of simultaneous masking, namely, noise-masking-tone (NMT) [2], tone-masking-noise [2] and noise-masking-noise (NMN) [2].

Another important concept of psychoacoustics is spreading function. Many psychoacoustic model uses a spreading function to compute an auditory spectrum. It is straight forward method to estimate masking levels within a critical band by using a component in the critical band .More details on spreading function can found in [4].

IV. PSYCHOACOUSTIC MODELS

The psychoacoustic model transforms the time domain input signals into a basilar membrane representation (i.e. a model of the basilar membrane in the human auditory system) and after this transformation the signals are processed in the frequency domain with the use of a fast Fourier transform (FFT). A transformation to the pitch scale (Bark scale) takes place (where the pitch scale is the psychoacoustic representation of the frequency scale). The goal of a psychoacoustic model for an audio coder is to determine the masking threshold that is generated by a time-localized input signal where the maskee can be assumed to be (white) quantization noise. Ideally, this would require finding the Masking curve of the input signal that takes both spectral and temporal masking into Account so that the resulting curve can be represented in the time-frequency plane. Furthermore, the psychoacoustic model would need to handle all types of signals that contain any combinations of tonal and noise components at various frequencies and at various intensities. Johnston proposed an auditory masking model in [6] that was largely based on the work of Schroeder [27]. Johnston's model was used to derive a short-term spectral masking threshold, from which quantization noise was shaped in a transform coder. The model operates on 64 ms frames of audio signaIs sampled at 32 kHz, yielding a spectral resolution of 15.625 Hz per frequency bin.

The Moving Pictures Expert Group (MPEG) [1] draft provides two informative psychoacoustic Models that compute the just-noticeable level of noise for signal coding. The output of the auditory model is a Signal-to-Mask Ratio for each coder sub-band.

Model1 performs these nine steps[56]

1. Perform FFT analysis: A 512 or 1024 point FFT transform with a hanning window with adjacent overlapping of 32 or 64 samples respectively to reduce edge effects is used to transform time aligned time domain data to the frequency domain.
2. Determine the sound pressure level: The signal is normalized to a maximum of 96dB SPL. The maximum SPL is calculated for each subband by choosing the greater of the maximum amplitude spectral line in subband.
3. Consider the threshold in quiet: An absolute hearing threshold in the absence of any signal is given; this forms the lower masking bound. An offset is applied depending on the bit rate
4. Finding tonal (sinusoidal) and nontonal (noise like) components: Tonal and nontonal components in the signal are identified.
5. Decimation of tonal and non tonal masking components: The number of maskers is reduced to obtain only the relevant maskers. Relevant maskers are those with magnitude that exceeds the threshold in quiet and those tonal components that strongest within ½ bark
6. Calculate individual masking thresholds: The total number of masker's frequency bins is reduced and maskers are relocated. Noise masking thresholds for each subband accounting for tonal and nontonal components and their different downward shifts are determined by applying a masking (spreading) function to the signal. Calculations use a masking index and masking function to describe masking effects on adjacent frequencies.
7. Calculate the global masking threshold: The powers corresponding to the upper and lower slopes of individual subband masking curves as well as a given threshold of hearing (threshold in quiet) are summed to form a composite global masking contour.
8. Determine the minimum masking threshold: the minimum masking level is calculated for each subands
9. Calculate the signal to mask ratio: SMR ratios are determined for each subband based on the global masking threshold.

Psychoacoustic model 2 performs a more detailed analysis at the expense of greater computational complexity; it is designed for lower bit rates than model 1.As in model 1.

Model 2 performs these 14 steps[56]:

1. Reconstruct input samples: A set of 1024 input samples is assembled
2. Calculate the complex spectrum: The time aligned input signal is windowed with a 1024 point Hanning window .An FFT is computed and output represented in magnitude and phase

ISSN: 2278 – 909X

3. Calculate the predicted magnitude and phase: The predicted magnitude and phase are determined by extrapolation from the two preceding threshold blocks.
4. Calculate the unpredictability measure: The unpredictability measure is computed using the Euclidian distance between the predicted and actual values in the magnitude /phase domain. To reduce complexity the measure may be computed only for lower frequencies and assumed constant for higher frequencies.
5. Calculate the energy and unpredictability in the partitions: The energy magnitude and the weighted unpredictability measure in each threshold calculation partition is calculated.
6. Convolve energy and unpredictability with the spreading function.
7. Derive tonality index: The unpredictability measures are converted to tonality indices ranging from 0 to 1.
8. Calculate the required signal to noise ratio: An SNR is calculated for each threshold calculation part ion using tonality to interpolate an attenuation shift factor between noise masking tone (NMT) and tone masking noise(TMN).
9. Calculate power ratio: The power ratio of the SNR is calculated for each threshold calculation partition.
10. Calculate energy threshold: the actual energy threshold is calculated for each threshold calculation partition.
11. Spread threshold energy: The masking threshold energy is spread over FFT lines corresponding to threshold calculation partitions to represent the masking in the frequency domain.
12. Calculate final energy threshold of audibility: The spread threshold energy is compared to values in absolute threshold of quiet tables and the higher value is used as the energy threshold of audibility.
13. Calculate pre echo control: A narrow band pre echo control used in the layer III encoder is calculated, to prevent audibility of the error signal spread in time by the synthesis filter. The calculation lowers the masking threshold after a quiet signal. The calculation takes the minimum of comparison of the current threshold with the scaled thresholds of two previous blocks.
14. Calculate signal-to-mask ratios: Threshold calculation partitions are converted to codec partitions (scale factor bands).The SMR (energy in each scale factor band divided by noise level in each scale factor band) is calculated for each partition and expressed in decibels. The SMR values are forwarded to allocation algorithm.

An auditory model was developed by the International Telecommunications Union (ITU) within the framework of the Perceptual Evaluation of Audio Quality (adopted as ITUR BS.1387) [28]. PEAQ provides advanced metrics for the assessment of the calculation perceptual quality of audio signaIs. Among other model output variables, a masking threshold is estimated from the auditory model. The two psychoacoustic ear models used in PEAQ Firstly, the FFT-based model is described, followed by the filter bank-based model .In the Advanced Version of PEAQ a second ear model is used in conjunction with the FFT based ear model already used in the Basic Version of PEAQ. In the filter bank based ear model, processing is carried out in the time domain rather than in short frames as with the FFT based peripheral ear model. Prior to the standardization of PEAQ there were few audio codecs or audio quality assessment algorithms containing a filter bank based ear model due to issues of complexity and computational inefficiency.

V. WAVELETS IN AUDIO CODING

All perceptual coders share a similar structure in that they contain a filter bank, psychoacoustic model, and an encoding and quantization stage, the filter bank stage provides a decomposition of the input signal that makes the application of perceptual criteria possible and also provides some decor relation of the input signal. Many types of filter banks and time-to-frequency transforms exist where each other's a different set of trade-offs in its design. Many have been considered and explored in the context of audio coding [2] and one in particular called the Wavelet Transform (WT) has shown to be interesting and potentially very useful.

A number of audio coders based on the WFB [54] have been proposed over the past decade in order to demonstrate the feasibility of such a scheme and to explore various configurations that lead to a better design. A brief description of several examples as well as a summary of design approaches for the decomposition tree structure and wavelet basis filter is given next.

One of the earliest examples of a wavelet audio coder was proposed by Wickerhauser in [30], where the well known Best Basis algorithm was used. The wavelet analysis was done by selecting the \best" tree from a library of tree structures through the use of a simple entropy criterion. The resulting decomposition provided many coefficients that fell below a certain threshold, where such coefficients were simply discarded so that coding requirements were reduced. Furthermore, it was found that Huffman coding in the wavelet domain provided better performances than applying Hu Huffman man coding in the time domain, indicating that the wavelet transformation did provide a good decorrelation property. The proposed coder was tested using speech signals only and results indicated that the algorithm provided modest compression ratios of between 2 and 3. Other non-perceptually-based wavelet audio coders have also followed, e.g. [31, 32, 33] but were generally found to provide lower performances compared to the perceptually-based audio coding schemes.

The first extensive study using a perceptually based scheme was done by Sinha and Tewfik in [29]. The coder that they proposed was comprised of two parts, namely, a perceptual part and a dynamic dictionary part. The two were designed to work in conjunction so that one removed the perceptual irrelevancies and the other removed the Statistical redundancies. The perceptual part consisted of a wavelet filter bank The WFB was based on a fixed 29-band CB resolution tree structure and an adaptive basis filter. The filter selection was done by computing the bitrate required for perceptual transparency with each filter from a library of basis filters, and choosing the filter that provided the best performance. The filter library was limited to wavelets with the maximum number of vanishing moments, e.g. Daubechies, which only differed in their phase responses. Filters with the maximum number of vanishing moments were indicated as being "near optimal" among different classes of filters.

ISSN: 2278 – 909X

Other variations on Sinha and Tewfik's wavelet audio coder have also appeared and a few are mentioned here. In [34], Black and Zeytinoglu proposed a simpler wavelet coder based on the same fixed CB tree structure as [29] but using a fixed 16-tap Daubechies filter. The psychoacoustic model employed the output from the WFB stage rather than re-computing a Fourier domain representation, which was essentially less accurate but also less computational load. The coding quality of the algorithm was reported to be comparable to MPEG-1 Layer I algorithm, which provided near-transparency for bitrates above 128 kbps [36]. Other wavelet audio coders that tried to use a wavelet analysis inside the psychoacoustic model appeared in [37, 38]. These audio coders have reported encoding rates that ranged anywhere between 70 and 110 kbps. In [39], Leslie and Sandler proposed a coder that used a fixed uniform 32-band tree structure, similar to the Polyphase filter bank that was used in MPEG-1 Layer I and II algorithm, and a fixed Daubechies filter. Listening results indicated that the coder was comparable to the MPEG-1 Layer I coder even though the frequency localization property of the WFB was found to be poorer than that of the Polyphase filter bank.

In [35], Srinivasan and Jamieson proposed a wavelet audio coder with a fixed basis filter and an adaptive tree structure. The MPEG-1 Psychoacoustic Model 2 [16] was used for computing the masking threshold. The algorithm essentially tried to adapt the tree structure to match the frequency resolution of the resulting masking threshold while satisfying some computational constraint.

In [40], Philippe et al. experimented with a WFB coding scheme that offered a great deal of flexibility in terms of the tree structure and basis filter. This scheme was used to determine how various choices of the tree structure and basis filter affected the overall performance.

In summary, the application of the WFB in perceptual audio coding has been shown to be feasible by several proposed coders and bitrates of between 48 and 110 kbps have been reported. We note that the wide range of performances in these wavelet coders can be attributed to the choice of the WFB, but also to the differences in the other stages of the coder, as well as the testing procedure used to carry out the evaluations. As a result, an objective comparison between various WFB strategies is difficult. But we can still analyse the various strategies and determine if any consensus exists among them

VI .AUDIO QUALITY ASSESSMENT

Perceptual audio quality assessment has been investigated for several decades. Most audio quality models are designed for handling coding distortions only Traditional objective measurement methods, such as signal-to-noise ratio (SNR) or total harmonic distortion (THD), have never really been shown to relate reliably to the perceived audio quality[55]. A number of methods for making objective perceptual assessment of audio quality have been developed as the ITU identified an urgent need to establish a standard in this area. The level difference between the masked threshold and the noise signal is evaluated in a noise-to-masked ratio (NMR) Measurement method presented by Brandenburg [43]. In the method proposed by Beerends and Stemerdink. [44], the difference in intracranial representations of the reference and distorted audio signals was transformed with a cognitive mapping to the subjective perceptual audio quality. A perceptual evaluation developed by Paillard et al. [45] first modeled the transfer characteristics of the middle and inner ear to form an internal representation inside the head of the subject, which is an estimate of the information being available to the human brain for comparison of signals, and the difference between the representations of the reference and distorted signals was taken as a perceptual quality. By comparing internal basilar representation of the reference and distorted signals, a perceptual objective measurement (POM) proposed by Colomes and Rault. [46] Quantified a certain amount of degradations including the probability of detecting a distortion and a so-called basilar distance. Sporer introduced

a filter bank with 241 filters to analyze and compare the reference and distorted signals in [47]. A perceptual measurement method (DIX: disturbance index) proposed by Thiede and Kabit. [48] is based on an auditory filter bank that yields a high temporal resolution and thus enables a more precise modeling of temporal effects such as pre- and post-masking. These six perceptual models [43–48] combined with some toolbox functions were integrated into the ITU recommendation BS.1387 [49].

However, some limitations have been discovered in PEAQ. Most notably PEAQ is shown to be unreliable for signals with large impairment resulting from low bitrate coding [50]. Furthermore, PEAQ is limited to a maximum of two channels. Consequently, improvements in PEAQ have been developed. Barbedo and Lopes. [51] Proposed a new cognitive model and new distortion parameters. The limitation of PEAQ up to a maximum of two channels was addressed by the development of an expert system to assist with an optimization of multichannel audio system [52].

VIII. CONCLUSION

In this paper we have reviewed the different psychoacoustic models and different wavelets techniques used in the area of digital audio compression. Also we have reviewed the Perceptual audio quality assessment techniques.

REFERENCES

[1] ISO/IEC 11172-3, “Information technology coding of moving picture and associated audio for digital storage media at up to about 1.5 Mbits—part 3: audio,” 1993..

[2] T. Painter and A. Spanias, “Perceptual coding of digital audio,” Proceedings of the IEEE, vol. 88, no. 4, pp. 451–512, 2000

[3] H. Najafzadeh-Azghandi, “Perceptual Coding of Narrowband Audio Signals”. PhD thesis, McGill University, Montreal, Canada, ApI. 2000.

[4] T. Painter and A. Spanias, "Perceptual coding of digital audio," Proc. IEEE, vol. 88, pp. 451~513, Apr. 2000.

[5] Dermot Campbell, Edward Jones, Martin Lavin “Audio quality assessment techniques—A review, and recent developments”

[6] J. D. Johnston, "Transform coding of audio signaIs using perceptual noise criteria," IEEE J. Select. Areas Commun., vol. 6, pp. 314-323, Feb. 1988

[7] N. Jayant, \Signal compression: Technology targets and research directions," IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, pp. 796-818, June 1992.

[8] R. Crochiere, S. Webber, and J. Flanagan, “Digital coding of speech in subbands,"Bell Syst. Tech. Journal, pp. 1069-1085, 1976.

[9] M. Schroeder, B. Atai, and J. Hall, “Optimizing digital speech coders by exploiting masking properties of the human ear," Journal Acoust. Soc. Am., vol. 66, no. 6, December 1979.

ISSN: 2278 – 909X

[10] E. Schroeder and W. Voessing, "High quality deigital audio encoding with 2.0 bits/sample using adaptive transform coding," in Proc. of the 80th. AES Convention, 1986. preprint 2321.
[11] K. Brandenburg, \OCF- A new coding algorithm for high quality sound signals," in ICASSP-97, pp. .1.1-5.1.4, 1987
[12] J. Johnston, \Transform coding of audio signals using perceptual noise criteria," IEEE Journal on Selec. Areas in Comm., vol. 6, no. 2, pp. 314-323, 1988.
[13] Y. Mahieux, J. Petit, and A. Charbonnier, \Transform coding of audio signals using correlation between successive transform blocks," in ICASSP-89, pp. 2021-2024, 1989
[14] Y. Dehery, M. Lever, and P.Urcun, "A MUSICAM source codec for digital audio broadcasting and storage," in ICASSP-91, vol. 1, pp. 3605-3609, 1991
[15] K. Brandenburg, J. Herre, J. Johnston, Y. Mahieux, and E. Shroeder,\ASPEC: Adaptive spectral perceptual entropy coding of high quality music signals," in 90th AES convention, 1991. preprint 3011 (A-4).
[16] ISO/IEC, JTC1/SC29, Information technology- Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s-IS 11172-3 (audio), 1992
[17] K. Brandenburg, \ISO-MPEG-1 Audio: A generic standard for coding of high quality digital audio," J. Audio Eng. Soc., vol. 42, no. 10, pp. 780-792, October 1994
[18] J. Johnston, S. Quackenbush, G. Davidson, K. Brandenburg, and J. Herre, "Mpeg audio coding," in Wavelets, Subband, and Block Transform in Communications and Multimedia (A. Akansu and M. Medlyey, eds.), Kluwer Academic Publishers, 1999.
[19] ISO/IEC, Overview of the MPEG-4 Standard. http://mpeg.telecomitalialab.com/standards/mpeg-4/mpeg-4.htm, May 2002.
[20] P. Stokas, "Which is the best low-bitrate audio compression algorithm? OGG vs. MP3 vs. WMA vs. RA", http://http://ekei.com/audio/, March 2002.
[21]] L. Fielder, M. Bosi, G. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC- 3: Low-complexity transform-based audio coding," in Collected Papers on Digital Audio Bit-Rate Reduction, Audio Engineering Society, 1996.
[22]] K. Tsutsui and et al., \ATRAC: Adaptive transform acoustic coding for MiniDisc," in Collected Papers on Digital Audio Bit-Rate Reduction, Audio Engineering Society, 1996
[23] J. Johnston and et al., \AT&T Perceptual Audio Coding (PAC)," in Collected Papers on Digital Audio Bit-Rate Reduction, Audio Engineering Society, 1996.
[24] J. Moffitt, "Ogg vorbis - open, free audio - set your media free," Linux Journal, pp. 146-50, January 2001.
[25] Xiph.Org, OggVorbis: open, free audio. http://www.vorbis.com, April 2003.
[26] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting Masking properties of the human ear," 1. Acoust. Soc. Am., vol. 66, pp. 1647-1652, Dec. 1979.
[27] J. Johnston, \Audio coding with filter banks," in Subband and Wavelet Transforms (A. Akansu and M. Smith, eds.), pp. 287-307, Kluwer Academic, 1996.
[28] International Telecommunication Union, Method for Objective Measurements of Perceived Audio Quality, July 1999. ITU-R Recommendation B.S.1387.
[29] D. Sinha and A. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," IEEE Trans. Signal Processing, vol. 41, no. 12, pp. 3463-3479, December 1993.
[30] M. Wickerhauser, \Acoustic signal compression with wavelet packets," in Wavelets: A Tutorial in Theory and Applications (C. Chui, ed.), Academic Press, 1992.
[31] S. Chan and et al., "A hybrid coder using the wavelet transform," in Proceedings of the IEEE International Symposium on Time-Frequency and Time-Scale Analysis, pp. 463-466, 1992.
[32] W. Kinsner and A. Langi, "Speech and image signal compression with wavelets," in IEEE WESCANEX 93, Communications, Computers and Power in the Modern Environment Conference Proceedings, pp. 368-375, 1993.
[33] A. Erdemir and et al., "Data compression using wavelet transforms and vector quantization," in Proceedings of 1994 Midwest Symposium on Circuits and Systems, pp. 965-968, 1994.
[34] M. Black and M. Zeytinoglu, "Computationally efficient wavelet packet coding of wide-band stereo audio signals," in ICASSP-95, pp. 3075-3078, 1995.
[35] P. Srinivasan and L. Jamieson, "High-quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling," IEEE Transactions on Signal Processing, vol. 46, no. 4, pp. 1085-1093, 1998.
[36] D. Pan, "A tutorial on MPEG/Audio compression," IEEE Multimedia, vol. 2, no. 2, pp. 60-74, 1995.
[37] W. Dobson and et al., \High quality low complexity scalable wavelet audio coding," in ICASSP-97, pp. 327-330, 1997
[38] T. Blu, "An iterated rational filter bank for audio coding," in IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis", pp. 81-84, 1996.
[39] B. Leslie and M. Sandler, "Audio compression using wavelets," in IEE Colloquium on Audio and Music Technology: The Challenge of Creative DSP, 1998.
[40] P. Philippe and et al., "Wavelet packet filterbanks for low time delay audio coding," IEEE Transactions on Speech and audio processing, vol. 7, no. 3, pp. 310-322, May 1999.
[41] Recommendation ITU P.911, Subjective audiovisual quality assessment methods for multimedia application, ITU Telecommunication Standardization Sector, December 1998.
[42] ITU Recommendation J.148, Requirements for an objective perceptual multimedia quality model, ITU Telecommunication Standardization Sector, May 2003.
[43] K. Brandenburg, Evaluation of quality for audio encoding at low bit rates, in: Proceedings of the Contribution to the 82nd AES Convention, preprint 2433, London, United Kingdom, 1987.
[44] J.G. Beerends, J.A.J.A. Stemerdink, A perceptual audio quality measure based on a psychoacoustics sound representation, J. Audio Eng. Soc. 40 (1992) 963–978 December.
[45] B. Paillard, P. Mabilleau, J. Soumagne," perceptual evaluation of the quality of audio signals", J. Audio Eng. Soc. 40 (1992) 21–32.
[46] C. Colomes, M. Rault, A perceptual model applied to audio bit-rate reduction, J. Audio Eng. Soc. 43 (1995) 233–240 April.
[47] T. Sporer, Objective audio signal evaluation – applied psychoacoustics for modeling the perceived quality of digital audio, in: Proceedings of the 103rd AES-Convention, preprint 4512, New York, United States of America, October 1997.
[48] T. Thiede, E. Kabit," A new perceptual quality measure for bit rate reduced audio", in: Proceedings of the Contribution to the 100th AES Convention, preprint 4280, Copenhagen, Denmark, 1996.
[49] ITU-R Recommendation 1387-1," Method for objective measurement of perceived audio quality", ITU Telecommunication Standardization Sector, 1998–2001
[50] C.D. Creusere, K.D. Kallakuri, R. Vanam," An objective metric for human subjective audio quality optimized for a wide range of audio fidelities", IEEE Trans. Audio Speech Lang. Process. 16 (1) (2008) 129–136 January.
[51] J. Barbedo, A. Lopes, A new cognitive model for objective assessment of audio quality, J. Audio Eng. Soc. 53 (1/2) (2005) 22–31.
[52] S. Zielinski, F. Rumsey, R. Kassier, S. Bech, Development and initial validation of a multichannel audio quality expert system, J. Audio Eng. Soc. 53 (1/2) (2005) 4–21.
[53] Christopher R. Cave, Perceptual Modelling for Law-Rate Audio coding. MS thesis, McGill University, Montreal, Canada, ApI. 2002
[54] Peter Lee, Wavelet filter banks in Perceptual Audio coding. MS thesis, Waterloo University, Ontario, Canada, 2003.
[55] Junyong You, Ulrich Reiter,Miska M. Hannuksela,Moncef Gabbouj,Andrew Perkis," Perceptual-based quality assessment for audio–visual services: A survey", Elesvier journal on Signal Processing: Image Communication 25 (2010) 482–501
[56] Ken C.Pohlmann " Principles of digital audio" 5th edition Mcgraw hill publication

Jagadeesh B.Kanade holds the Master degree from VTU, Karnataka, and currently pursuing PhD at PRIST University Thanjavur, India, and Earlier worked at several software industries in audio and video compression algorithms on different hardware platforms currently working as Assistant professor in HKBK College of engineering, 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.

ISSN: 2278 – 909X