Turkish Journal of Electrical Engineering and Computer Sciences

Volume 6 | Number 2 Article 3

1-1-1998

Signal Injection With Perceptual Criteria

T. ENGİN TUNCER

ELEKTRİK, VOL.6, NO.2, 1998, ©TÜBİTAK EMO

Signal Injection With Perceptual Criteria*

T. Engin Tuncer
Electrical and Electronics Engineering Department
Middle East Technical University, Ankara - TURKEY

Abstract

In this paper, a novel method for increasing the coding performance and information transmission capacity is presented. This method is mainly based on perceptual modelling of input signal such as speech or audio. Presented approach may be seen as an alternative to transforms which dynamically change analysis window for better energy compaction. A perceptual model is established in order to obtain a global masking threshold in frequency below which sounds become inaudible. Certain criteria are developed for identifying the signal injection bands. A new multiband filter design method which is a generalization of windowing method is used to separate the inaudible spectrum. These spectral partitions are then used to send additional information. Under the condition that signal injection and synthesis after decoding is done appropriately, injected signal is not audible within the original audio signal. This type of signal injection is especially useful in audio coding and Digital Audio Broadcasting, (DAB).

1. Introduction

Classical coding techniques have been successfully employed in telecommunications leading to many different coder structures [1]. Usually an objective criteria is used for performance optimisation since it is analytically tractable and easily computed. A fundamental inadequacy of classical coders is that they disregard the human perceptual mechanism. Auditory systems have different sensitivities for different tones and perceptual evaluation is quite different than MSE type evaluation [2,3,4]. Modern coders use auditory models in order to hold the noise level below the hearing threshold. This approach yields much better results for perceptually transparent coding compared to MSE optimum coders [2].

Perceptual modelling of human auditory system has been employed in modern coders successfully [7,8]. Coders which use perceptual modelling perform much better than MSE optimum coders even for low SNR values. The advantage of perceptual coders comes from the fact that perceptual source entropy is usually lower than the classical entropy. Perceptual coders try to mask the quantization noise such that it would be inaudible at the decoder.

The importance of auditory processes during coding increases with the bandwidth of the input signal. The advantages of perceptual modelling are especially more obvious for audio coding [5,6,7]. However the signal injection method is still valid for other types of applications including speech coding.

In this paper, a similar perceptual model that is described in MPEG-1 audio standard is employed for signal injection [7,9,10,21]. The MPEG-1 audio standard describes three layers of perceptual coders with increasing performances and complexities. These coders have a dynamic bit allocation feature where they

* This work is supported by Turkish Science and Technology Research Council, TÜBİTAK project EEEAG-174.

use the perceptual model for determining the number of bits allocated to each channel. MPEG-1 audio coders use cosine-modulated filter banks to divide the audio spectrum into 32 equal bandwidth channels [11,12,13]. Third layer increases the number of channels by using an additional MDCT (Modified Discrete Cosine Transform) transform. The frequency resolution changes dynamically by changing the window size with a 6 point or 18 point MDCT depending on the characteristics of input audio signal.

The importance of time-frequency localization of the input signal is evident in coding. Although it is possible to find different transforms for obtaining good time-frequency localizations, the computational complexity of these transforms makes them less attractive in coding applications [14,15,16]. The computational burden is mainly due to the dynamic changes of the transform analysis window which depends on the input signal.

In this paper, an alternative approach is taken in order to increase the performance of an audio coder. The idea depends on the fact that one might find certain spectral regions in an audio signal where there is no contribution to the perceptual quality. These spectral regions can be identified by using a global masking threshold which is obtained from the perceptual model. After the inaudible spectral regions are found, the most suitable one for signal injection is chosen. This spectral band is filtered out with a filter designed by a multiband filter design method which is the generalization of the windowing method. An additional information signal is injected to this band carefully. Certain precautions are taken to ensure the injected signal will not be audible after encoding and decoding operations. In this sense, signal injection may be defined as the signal insertion or addition to some specified frequency region of an audio signal.

One obvious application of the presented method is in DAB. The MPEG-1 layer 2 coder has been chosen as an international standard for DAB. Certain services such as program language and service labels, special announcements, traffic information, etc. are also sent along with the audio signal in DAB [17]. The transmission rate is 180 + 12 kb/s. This side information which is 12 kb/s in particular takes another band which increases the overhead of the transmission system. Signal injection method is very suitable for avoiding this overhead by increasing the system performance.

Although the proposed signal injection method for audio coding is developed independently, there are previous works especially for data hiding and copyright protection [18,19]. These methods are employed in multimedia applications [20,21] to keep track of the original author of the multimedia product. These methods are not intended for audio coding since inserted data can not be retrieved with an acceptable amount of error after a lossy coding operation [20].

The outline of the paper is as follows. Section 2 describes the perceptual model used in this work. The multiband filter design method is explained in Section 3. Signal injection method is presented in Section 4.

2. Perceptual Model

Audible noise energy remains nearly constant at certain frequency bands which are called critical bands [22,23]. During the auditory perception, a critical band analysis is performed and a frequency-to-place transformation is done in the inner ear. It is assumed that auditory system processes the incoming signals through a series of overlapping bandpass filters which are placed at the centers of critical frequency bands. These critical bands are important in the sense that masking phenomena can be described in terms of the critical band analysis [8,9].

Human auditory system perform two types of masking, namely simultaneous and temporal masking. Simultaneous masking is the phenomena that a low level tone is made inaudible (masked) by a simultaneously occurring stronger tone if both are sufficiently close to each other in frequency. Temporal masking is a phenomena that a low level signal is made inaudible by a stronger signal which occurs before or after the low level signal in time. If there is no masker, a signal may still be inaudible as long as its sound pressure level is below the threshold in quiet (or absolute threshold) which is defined in frequency [6]. These masking phenomena should be taken into consideration in creating a perceptual model.

Masker signal may be tone-like or noise-like. The energy levels for tone-masking noise, E_T, and noise-masking tone, E_N can be given as [2],

$$E_T = E_N + 15.5 + x_i \tag{1}$$

$$E_N = E_T + 5.5 \tag{2}$$

where x_i is the Bark frequency axis value for i'th critical band. Above equations show that noise is a better masker than a tone.

Before attempting to find the masking threshold values, the unpredictability of the audio signal, U_k, should be found at each frequency. Unpredictability is a measure of noise likeliness of a signal and it has values between 0 and 1 [9]. It increases as the spectrum change increases and it is used to find the tonality which is employed to weight the equations (1) and (2). Considering a second order forward difference operator, we have defined unpredictability as,

$$U_k = \frac{||X_n(k)| - [2|X_{n-1}(k)| - |X_{n-2}(k)|]Cos(\phi_n(k) - 2\phi_{n-1}(k) + \phi_{n-2}(k))|}{|X_n(k)| + |2|X_{n-1}(k)| - |X_{n-2}(k)||} \tag{3}$$

where $X_n(k)$ is the k'th frequency component at the n'th signal frame and $\phi_n(k)$ is the k'th phase component at the n'th signal frame.

The two masking phenomena explained above describe the masking events inside the critical bands. The masking effects of critical bands between each other can be modelled by a spreading function. Spreading function relates the masking effect of a signal in a critical band onto other critical band signals [24]. It varies by sound pressure level and frequency. The spreading function, S(i,j) can be approximated as,

$$S(i,j) = 10^{(1.581+0.75(x_i-x_j+0.474)-1.75(1+(x_i-x_j+0.474)^2)^{0.5})} \qquad \begin{cases} 0 \le i \le C \\ 0 \le j \le C \end{cases} \tag{4}$$

where C is the number of critical band partitions. In this work C is taken as 63. Note that the value of C simply determines the resolution in frequency. The effect of spreading function is included through a convolution operation to account for the masking relations between critical bands. There are several steps for calculating the global masking threshold. These are

a) Critical band partitioning and energy computation

b) Calculation of spread critical band spectrum

c) Calculation of offset masking energy and spread threshold

d) Comparison with the absolute threshold and finding the global masking threshold

These steps are explained below in detail.

a) Critical band partitioning and energy computation

The human auditory system processes the sounds depending on a critical band structure, and the input audio signal spectrum needs to be represented in terms of these critical bands for establishing a perceptual model. This is done by using the critical band limits [2,22]. After the frequency axis is partitioned in terms of the critical bands, energy values, $G_e(i)$, and unpredictability measure weighted energies, $G_u(i)$, are found at each of these critical band partitions, i.e.,

$$G_e(i) = \sum_{k=0, k\in P_i}^{511} |X(k)|^2 \qquad 0 \le i \le C \tag{5}$$

$$G_u(i) = \sum_{k=0, k\in P_i}^{511} U_k |X(k)|^2 \qquad 0 \le i \le C \tag{6}$$

where P_i is the i'th critical band partition. Figure 1 shows the critical band energies, $G_e(i)$, on the Bark axis.

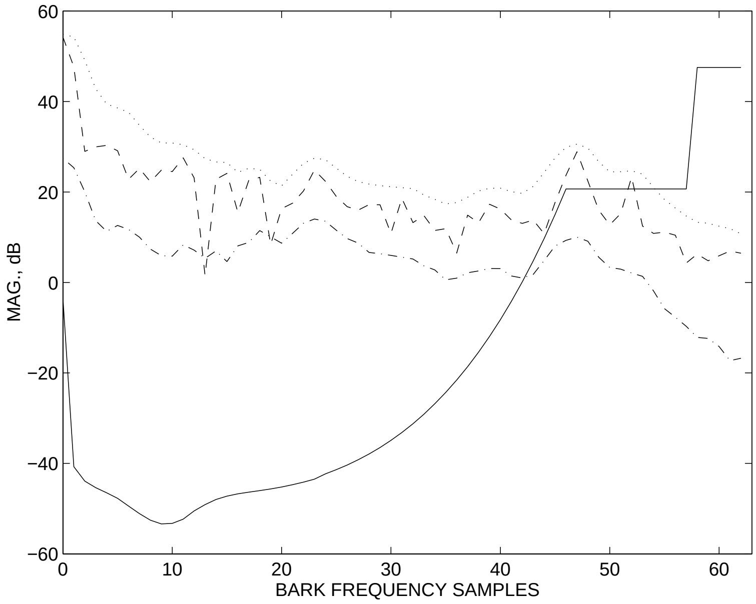

Figure 1. Magnitude responses of critical and spread spectrums, ($G_e(i)$ with .. line and $SG_e(i)$ with – line), normalized spread threshold, ($ST_n(i)$ with -.- line) and absolute threshold ($AT(i)$ with solid line). All data values are normalized with 32768^2 (90.3 dB) for convenience.

b) Calculation of spread critical band spectrum

The masking effect of critical band partitions between each other is obtained by convolving the spreading function with the partition energies. This convolution operation is done by a matrix multiplication. Spread critical band, $SG_e(i)$, and unpredictability weighted energies, $SG_u(i)$, are obtained as,

$$SG_e(i) = \sum_{j=0}^{C} S(i,j) G_e(j) \qquad 0 \le i \le C \tag{7}$$

$$SG_u(i) = \sum_{j=0}^{C} S(i,j) G_u(j) \tag{8}$$

Figure 1 shows the spread critical band spectrum, $SG_e(i)$.

c) Calculation of offset masking energy and spread threshold

The tonality of a signal, α, is a measure used to weight the energy level for tone-masking noise and noise-masking tone. This tonality measure can be obtained as,

$$\alpha(i) = 10 Log_{10}\left(\frac{SG_u(i)}{SG_e(i)}\right) \qquad 0 \le i \le C \tag{9}$$

The tonality measure is used to find the offset in masking energy. The offset masking energy, O(i), is used to identify the spread threshold and can be written as,

$$O_i = \alpha(i)(15.5 + x_i) + (1 - \alpha(i))5.5 \tag{10}$$

The spread threshold, ST(i), is obtained by using the spread critical band and offset energies,

$$ST(i) = 10^{Log_{10}(SG_e(i)) - O(i)/10} \tag{11}$$

d) Comparison with the absolute threshold and finding the global masking threshold

The spread threshold should be normalized appropriately since convolution by the spreading function increases the critical band energies. This is done by using the values of the spreading function and the normalized spread threshold. $ST_n(i)$, is obtained as,

$$ST_n(i) = \frac{ST(i)}{\sum_{m=0}^{C} S(i,m)} \qquad 0 \le i \le C \tag{12}$$

Figure 1 shows the normalized spread threshold $ST_n(i)$.

The absolute threshold or threshold in quiet, AT(i), describes the level below which any signal will be inaudible in the absence of a masker [8]. Normalized threshold should be compared with the absolute threshold since signal energy may be above the normalized threshold but still remain inaudible as long as it is below the absolute threshold. After the comparison, global masking threshold is obtained as,

$$GT(i) = \begin{cases} ST_n(i), & ST_n(i) \ge AT(i), \qquad 0 \le i \le C \\ AT(i), & ST_n(i) < AT(i) \end{cases} \tag{13}$$

The global masking threshold, $GT(i)$, is used to extract the inaudible part of the audio spectrum.

3. Multiband Linear-Phase Filter Design

Multiband linear-phase filter design is a generalization of the windowing method. This method converges to windowing design in the limit. The main advantage of the multiband design compared to other filter design techniques including windowing method is the ease of designing multiband frequency responses. Also sharp cut-off and notch type responses can be easily obtained. The design complexity is low and band limits (cut-off frequencies) in a multiband design can be easily imposed in the design process. In addition, multiband design can be extended to the multidimensional case easily.

Multiband linear phase filter design is based on the modulated prototype filter design methodology and the following fact,

Fact: Any window, w(n), which satisfies the condition

$$w(kN) = 0 \qquad \text{for k}=\pm1, \pm2, \pm3,.. \tag{14}$$

sums up to a constant in frequency, i.e.

$$\sum_{k=0}^{N-1} W(w - \frac{2\pi k}{N}) = w(0)\frac{N}{2\pi} \tag{15}$$

<u>Proof:</u> Proof of this fact is simple. Consider the following equation,

$$w(n) \sum_{k=-\infty}^{\infty} \delta(n - kN) = w(0)\delta(n) \tag{16}$$

Fourier Transform of the above equation yields eqn(15).

The requirement given in eqn(14) is satisfied by any window whose length N_w is less than N. Note that N can be increased to the desired length in order to satisfy $N > N_w$ condition. In multiband design, the summation in eqn(15) is taken within the desired band limits. This in turn gives the flexibility to design a multiband filter.

Considering a real coefficient filter, the multiband filter is written as,

$$H_m(w) = \frac{2\pi}{w(0)N} \sum_{k=-K, k\in G}^{K} W(w - \frac{2\pi k}{N}) \tag{17}$$

where G is a set of frequency points that a bandpass response is desired. In time domain, multiband filter is obtained by cosine modulation of a prototype filter, w(n), i.e.,

$$h_m(n) = \frac{2\pi}{w(0)N} \sum_{k=0, k\in G}^{511} a_k w(n) Cos(w_0 kn) \qquad a_k = \begin{cases} 0.5, & k=0 \\ 1, & \text{otherwise} \end{cases} \tag{18}$$

This prototype filter can be any symmetric window that satisfies equation (14). Figure 2 shows a multiband filter designed by considering the possible signal injection bands with a 513 point Kaiser window.

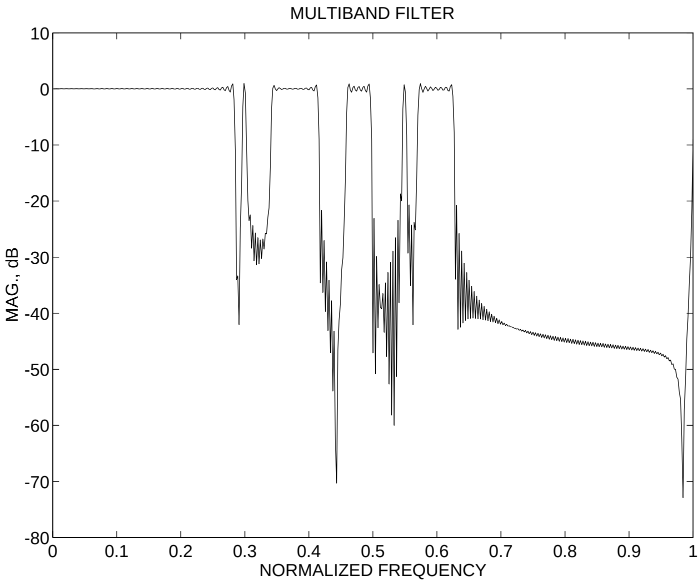

Figure 2. Multiband filter designed for filtering the inaudible frequency bands.

Now consider the filter designed by using the windowing method,

$$H(w) = \frac{1}{2\pi} \int_{-w_c}^{w_c} W(w - \theta) d\theta \tag{19}$$

where w_c is the cut-off frequency of the filter. The above equation can be obtained from equation (17) in the limit $N \to \infty$ where the summation turns into an integral. This shows that multiband filter design is a generalization of windowing method with certain advantages. In fact, multidimensional multiband filters can be easily designed by the proposed method.

4. Signal Injection Method

Perceptual coders are considered as modern coders since they use auditory models in order to cast the coding noise below the hearing threshold. Auditory models create a masking threshold in frequency for the input audio signal. The frequency components which fall below this masking threshold are not audible. Perceptual coders use the global masking threshold level, GT(i), in order to find the amount of quantization noise which will be below the threshold level and become inaudible during playback.

Although this approach yields good coders in terms of perceptual quality, there is still room for increasing the efficiency. Time-frequency localization is one of the important problems of perceptual coders as in many other transform coders. If the analysis window length is increased, better frequency resolution is obtained with a sacrifice on time resolution, and vice versa. Finding an optimum analysis window has always been a fundamental problem. The dynamical changes in input signal requires a continuously updated analysis window. However this is a prohibitively costly process. Therefore a constant bandwidth analysis window is preferred in many transform coders. For example, MPEG-1 layer 1 and layer 2 coders use 32 channel cosine-modulated filter bank for signal energy compaction. This type of frequency partitioning works well but there are alternative structures and procedures that increase the efficiency of the coder.

It is always possible to find certain frequency regions inside the subbands where audio signal energy falls below the global masking threshold. However since the overall subband signal is coded, these frequency regions are also coded and sent even though they do not contribute to the overall perceptual quality. In this paper, signal injection method is used to insert an additional information signal instead of these inaudible frequency bands. This approach increases the efficiency of the coder while it eliminates the need for a dynamically updated analysis window.

Signal injection can be used to increase either the coding efficiency or perceptual quality. The coding efficiency is increased by eliminating the need for an extra channel in order to send side information. The perceptual quality is increased by using the number of bits that are allocated to the extra channel for coding the audio signal samples. In this work, we have chosen the second approach, and coded the signal injected audio at 192 kb/s. This approach eliminates the use of 12 kb/s side information channel in DAB and codes the overall signal at 192 kb/s instead of 180 kb/s [17,25].

Signal injection method consists of different steps which can be outlined as,

a) Finding the global masking threshold level, GT(i).

b) Identifying the inaudible signal injection frequency bands.

c) Multiband linear-phase filter design.

d) Removing the inaudible frequency bands from the original audio spectrum.

e) Injection of the information signal to the inaudible frequency bands and encoding the signal injected audio with MPEG-1 layer 2 coder at 192 kb/s.

f) Synthesis of the information signal from the signal injected audio.

The block diagram of the signal injection method is given in Figure 3. The audio signal is processed frame by frame during encoding. The frame length is chosen as 1152 as in the case of MPEG-1 layer 2 coder and 1024 point FFT is used for energy calculations. There are two passes in perceptual calculations and all of the frame samples contribute to the resulting masking threshold. A masking threshold level is found for each of these frames. This masking threshold and a set of rules which are explained below are used to identify the signal injection frequency band. After signal injection is done, signal injected audio is coded and sent by using MPEG-1 layer 2 coder. During decoding, a new masking threshold is computed and signal injection bands are found from this and the side information. After the signal injection band limits are found, the injected signal is filtered out from the audio for further use. The signal injected band is deemphasized after decoding the audio in order to guarantee that injected signal is not audible during playback.

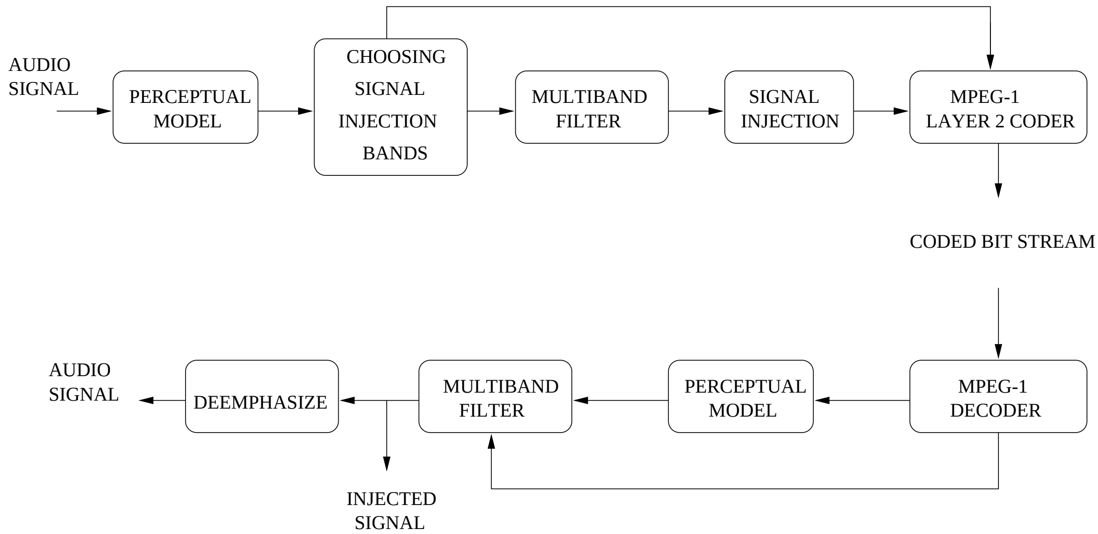

Figure 3. Block diagram of signal injection and synthesis.

Figure 4 shows an example of an audio spectrum and its global masking threshold. Note that there are certain frequency regions where signal energy falls below the global masking threshold. These are the regions which do not contribute to the overall perceptual quality.

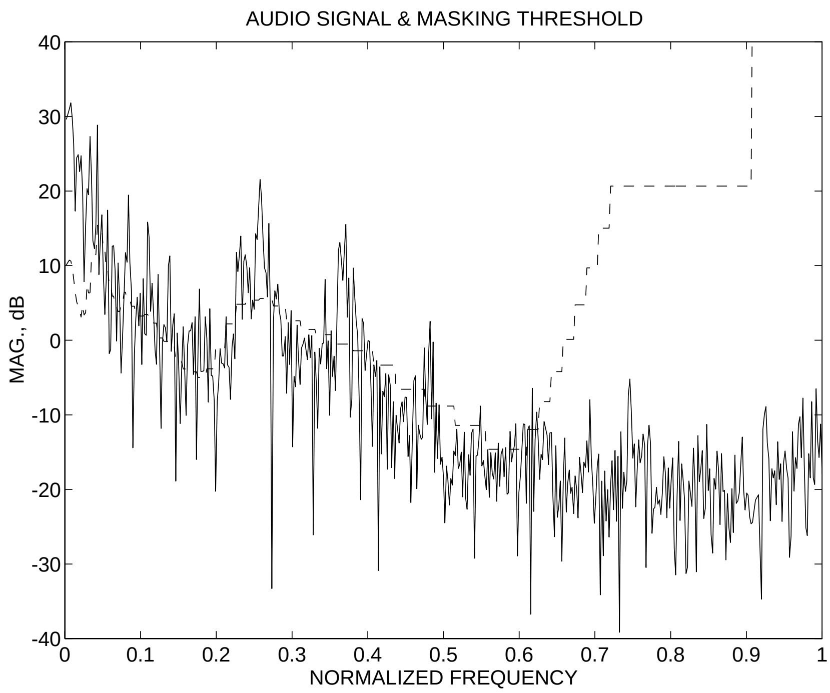

Figure 4. Spectrum of a frame of audio signal and the global masking threshold for this frame. (Masking threshold is with – line)

Given the masking threshold level, one of the most important steps of signal injection is the identification of inaudible frequency bands. These frequency bands may not be extremely narrow which makes it worthless for signal injection. Identification of the signal injection frequency bands requires a careful examination. We have developed a set of rules and procedures for this purpose which are described below.

4.1. Identification of Signal Injection Frequency Bands

Figure 5 shows the three possible cases of signal injection, inside or between subbands. Case **b** represents the mode where only one subband is used for signal injection while case **c** represents the mode where two subbands are used for that purpose. Case **a** has the rare occurrences when all the subband signal energy falls below the global masking threshold. Note that there are more than one way for the utilization of case **a** mode. A possible solution is to do signal injection as it is done for other cases. Alternatively, one can save the bits for the incoming frames by quantizing the subband samples with zero bits. In this work, we have chosen the first approach.

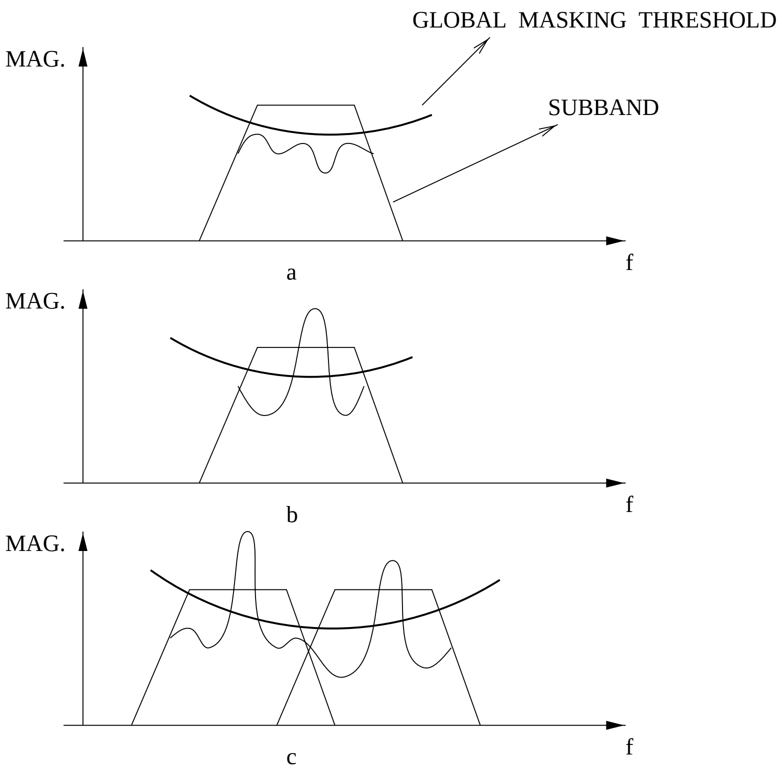

Figure 5. Possible cases of inaudible frequency bands in a subband.

Figure 6 shows an example of the spectra of an audio signal and the error made during coding where audio signal is sampled at 44.1 kHz with 16 bits. Usually the SNR is more than 25 dB for most of the frequencies for a variety of audio signals coded with MPEG-1 layer 2 coder at 192 kb/s. Considering the quality in coding, the exact location of the signal injection band within a subband can be found by using the perceptual model after decoding. The masking threshold found from the decoded signal is accurate enough for finding the band edges. The error made in finding the masking threshold in decoding is less than 0.3 dB. In order to compensate for small changes in masking threshold, signal injection is done 3 dB below the masking threshold and signal energy at injection band edges is required to be 0.5 dB above the masking threshold. Note that when SNR is 25 dB, the maximum error that can be made is in the order of %5.6. Since the errors in the order of %5.9 are controlled by a threshold of 0.5 dB, these precautions guarantee that the signal injection band edges are found correctly after decoding. Figure 7 shows the global masking thresholds found before coding and after decoding. It is clear that the quality of the masking threshold curve is preserved at 192 kb/s.

Since auditory model works with a 1024 point FFT spectrum, the audio spectrum is analyzed by considering these discrete frequency bins. In order to increase efficiency, certain rules are developed for identifying the frequency bands suitable for signal injection. These rules are also important for the design of a suitable multiband filter.

Rule 1: Only the adjacent five or more frequency bins are considered as a band.

Rule 2: There should be at least five frequency bins between two inaudible bands.

Rule 3: A subband can not include only one frequency bin of a signal injection frequency band. Otherwise it is excluded from the signal injection frequency bins.

Rule 4: Only subbands with more than 3 bits allocated can have any signal injection band.

Rule 5: The signal energy should be 0.5 dB above the global masking threshold at the edges of signal injection frequency band.

Rule 6: Signal injection should be done 3 dB below the global masking threshold.

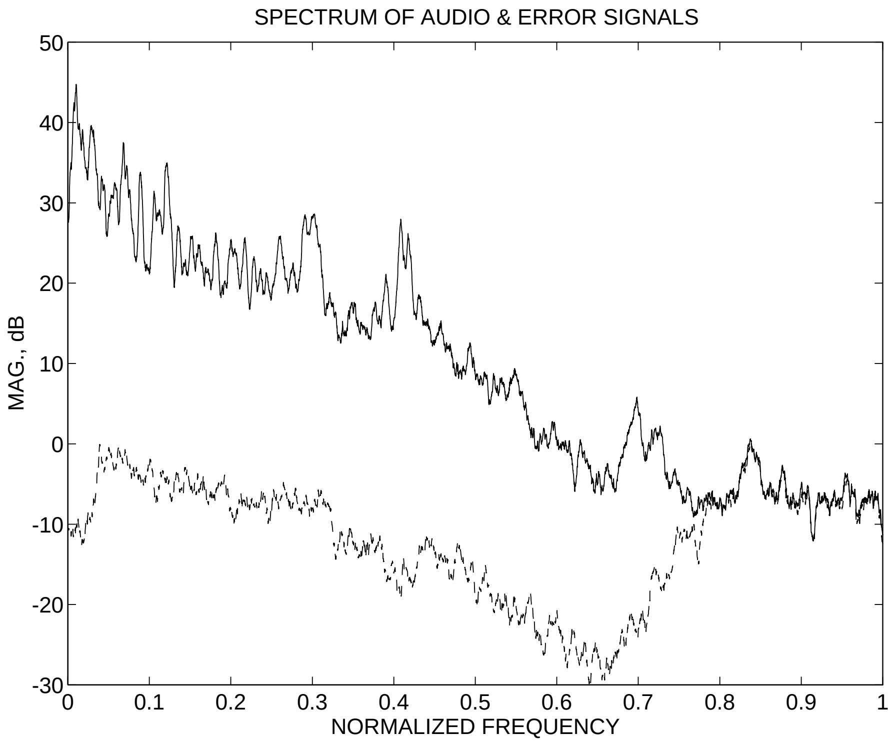

Figure 6. Spectrums of audio and error signal after decoding. Audio signal is coded with MPEG-1 layer 2 coder at 192 kb/s. Digital recording is done at 44.1 kHz and 16 bits .

Since each subband has 16 frequency bins for a 1024 point FFT, there might be only two signal injection bands present in a subband when the above rules are used for signal injection.

The location of a signal injection band is coded with 4 bits. 3 bits are used for the index of the subband where the signal injection band begins. '000' is reserved to indicate that no signal injection is done for that particular frame. Most of the audio signal energy is localized at low frequencies and more bits are allocated to the low frequency subbands. On the other hand, signal energy is usually above the global masking threshold for subbands between 0 and 5. Therefore subbands between 6 to 12 are taken as the most suitable subbands for signal injection and indexes are coded with 3 bits. Since there might be only two possible signal injection bands inside a subband, one bit is used to indicate whether signal injection band covers more frequency bins in low or high frequency sides with respect to the midpoint of subband frequency bins (low=0, high=1). If the signal injection band is centered to the midpoint, then there might be only one signal injection band inside a subband. Therefore one bit frequency side information does not affect the choice of the signal injection band location. Figure 8 shows the bit stream during coding of a frame of the audio signal.

Although one might find many possible frequency bands for a particular audio signal, it is not possible to use all of them for signal injection. Only a fraction of frequency bands do have enough bits during coding. Therefore one should verify that there are enough bits for coding before a frequency band is chosen for signal injection.

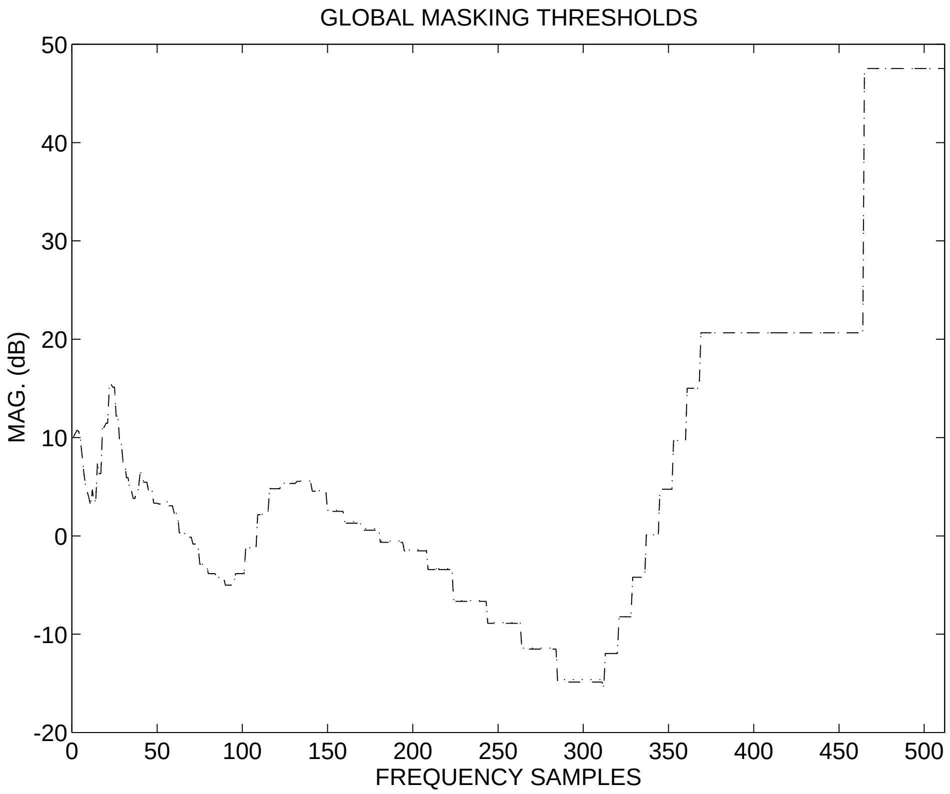

Figure 7. Global masking thresholds obtained before coding and after decoding. (.. line is used for masking threshold found before coding.)

HEADER	BIT ALLOCATION	SIGNAL INJECTION INFO	SCALEFACTOR INFO	SCALEFACTORS	SAMPLES

Figure 8. Bit stream for the encoded data during signal injection for MPEG-1 layer 2 coder.

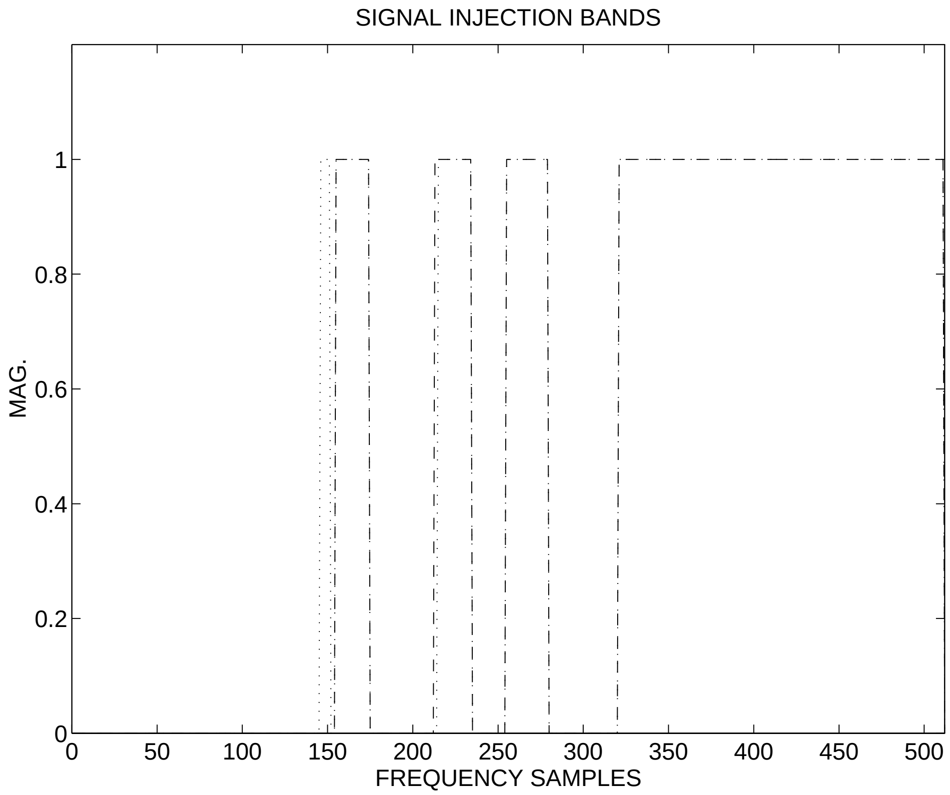

Figure 9. Signal injection frequency bands found before coding and after decoding. Signal injection is done in frequency bins between 154 and 174. (.. line for the bands found before coding)

Considering the rules presented above, signal injection frequency band is found by using the following procedure,

i) Find the candidate signal injection bands between the subbands 6 and 12.

ii) Choose the signal injection band with the largest bandwidth and find the corresponding subband where the signal injection band begins. If two bands have the same bandwidth, choose the one which is allocated more bits.

iii) Check the number of bits allocated for the subband or subbands associated with the signal injection bands. If less than 4 bits are allocated, return to step (ii) and choose the next suitable signal injection band.

iv) Continue until a suitable subband with a signal injection band is found.

v) If no suitable band is found, send '000' to indicate that no signal injection is done for that particular audio frame.

Figure 9 shows the possible signal injection bands found before coding and after decoding. Note that the second band which is chosen for signal injection for the processed audio frame is exactly found after decoding. This is because of the rules developed for signal injection and the side information. Although the remaining bands are slightly different, this does not pose any problem in terms of the signal injection procedure.

4.2. Signal Injection and Synthesis

After the global masking threshold level, GT(i), is found, suitable frequency bands for signal injection are identified as described above and a bandstop filter is designed by using the multiband filter design method considering the signal injection band limits. Figure 10 shows the bandstop filter used for signal injection.The audio signal is processed by this filter in order to separate the inaudible signal injection band. Figure 11 shows the filtered audio spectrum where the original spectrum is given in Figure 4. A new information signal which covers this frequency band is injected to the audio signal. This signal is added to the filtered audio signal such that injected signal is 3 dB below the masking threshold. This and Rule 5 ensure that the signal injection band edges are found correctly after decoding. Figure 12 shows the signal injected audio spectrum. After the signal injection, audio signal is ready for coding. MPEG-1 layer 2 coder at 192 kb/s is used for coding the signal injected audio which was sampled at 44.1 kHz with 16 bits. The side information regarding the place of the frequency region where signal injection is done is also sent with the frame information as described above. Note that this 4-bit side information is considerably small when compared to the number of samples sent through the signal injection band.

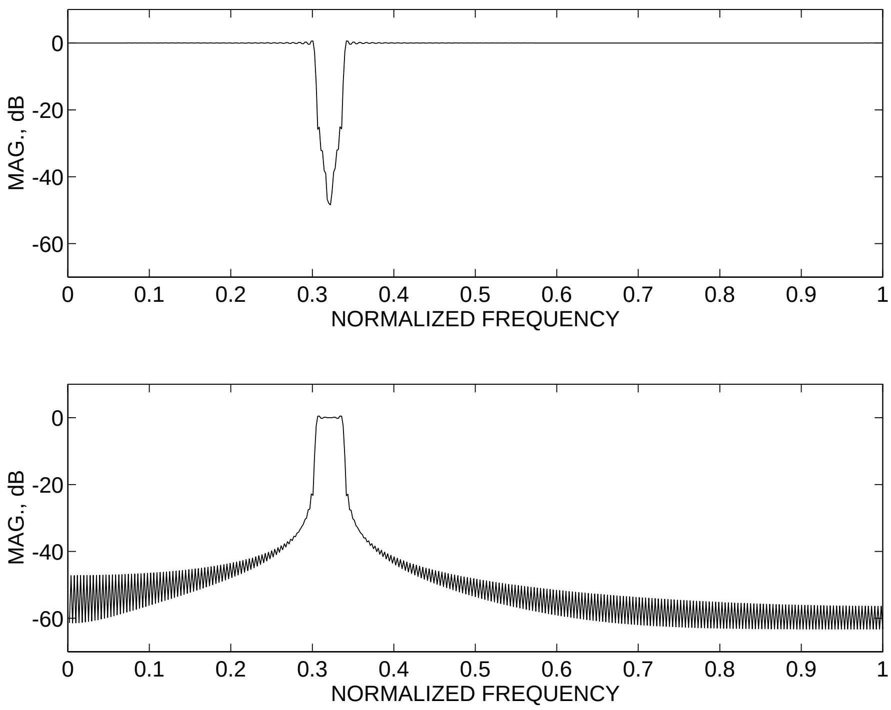

Figure 10. Bandstop and bandpass filter designed using multiband filter method. These filters are used both in signal injection and synthesis.

Synthesis of the injected signal is done at the decoding stage. The procedure for synthesis can be outlined as follows,

i) Determine the signal injection subband and the frequency side (low or high with respect to the subband midpoint) from the coded 4 bit sequence.

ii) Obtain the global masking threshold from the signal injected audio after decoding.

iii) Determine the beginning and end frequency bins of the signal injection band using the global masking threshold and 4-bit side information.

iv) Design a bandpass filter with the multiband filter design method by using the band limits obtained in (iii).

v) Obtain the injected signal by filtering it from the audio signal.

The injected signal is recovered through a process described above. Figure 13 shows the injected and reconstructed information signals. The MSE in recovering the injected signal is 4.5×10^{-3}. This shows that injected signal quality is preserved after encoding-decoding operations. MPEG-1 layer 2 coder uses scalar normalization within subbands during quantization. This and the precautions taken in signal injection are the reasons for the preservation of signal quality.

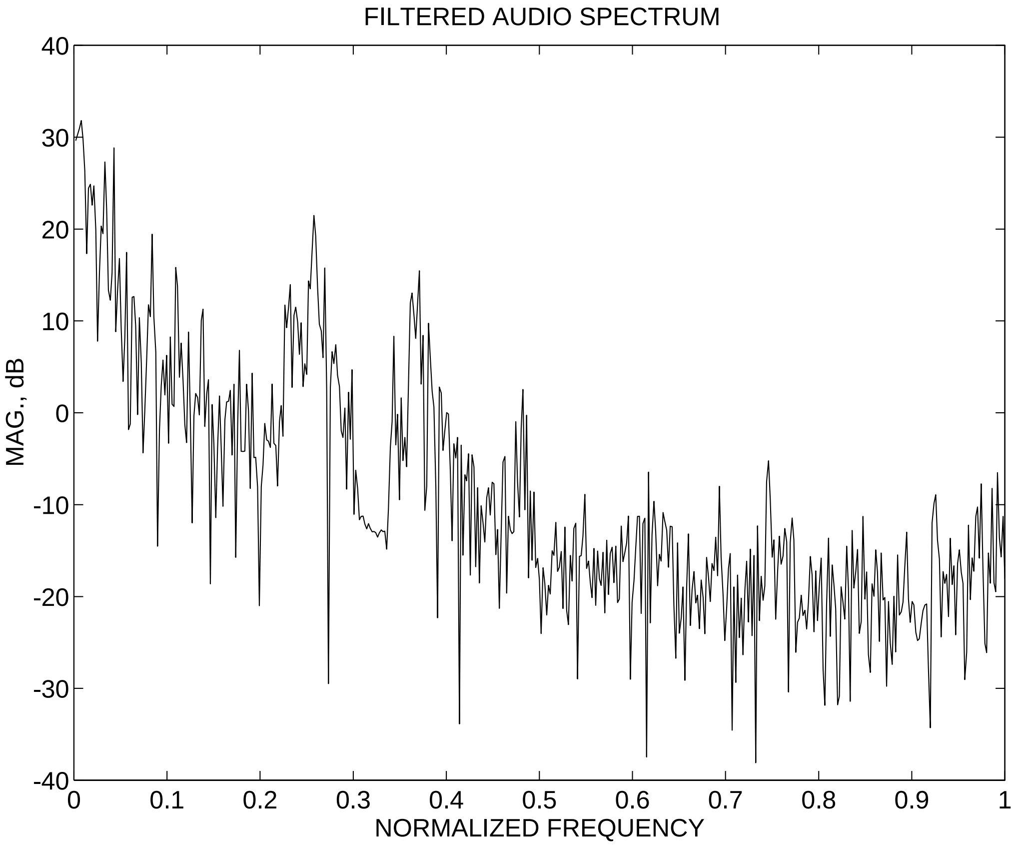

Figure 11. Audio signal spectrum where signal injection band is filtered out for further processing.

The audio signal synthesized after decoding is processed with a bandstop filter in order to deemphasize the signal injected frequencies so that injected signal does not pose a problem during playback. Assuming zero phase filters, the bandpass filter is designed by using the multiband method and bandstop filter is simply obtained from this by

$$h_{bandstop}(n) = \delta(n) - h_{bandpass}(n) \tag{20}$$

Figure 14a. shows the spectrum of the original audio signal and the error made after signal injection and decoding. Note that audio signal quality is preserved even after signal injection. Figure 14b. shows the error and global masking threshold which is found during decoding. It is clear that the error is completely below the global masking threshold and will not be audible during playback. Note that the signal injection band is deemhasized and the error is defined as the difference between the original audio and the audio recovered after deemphasis. So even though the error at the signal injection band is close to the global masking threshold, there is actually no significant signal energy at this band after deemphasis.

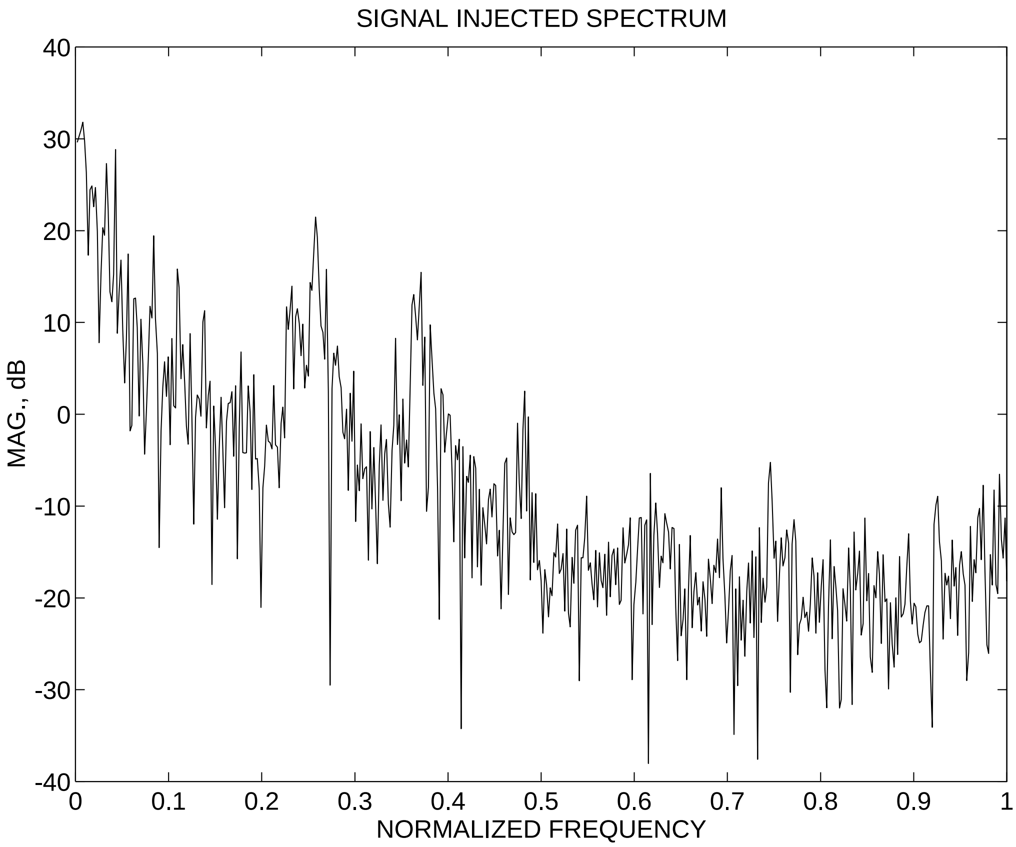

Figure 12. Signal injected audio spectrum.

5. Conclusion

A new method of signal transmission inside an audio signal is presented based on a perceptual model and multiband filtering. This approach presents an alternative to transform coders where an analysis window is dynamically adapted to the changes in input signal characteristics. Presented method is especially well suited for Digital Audio Broadcasting, eliminating the need for an extra channel for side information such as service labels, traffic info, etc.

A global masking threshold level is obtained by perceptually modelling the input audio signal. Audio signal is not audible unless its energy is above this threshold level. Usually one can find certain frequency partitions where signal energy falls below this masking threshold and becomes inaudible. These frequency components have no contribution to the overall quality of the audio signal.

Since most of the signal energy is localized at low frequencies, certain criteria which are outlined as a set of rules and procedures are developed for determining the most suitable frequency band for signal injection. These frequency bands are filtered out by using a multiband filter which is designed by a method that can be seen as a generalized windowing method. This multiband filter design method has certain advantages such as the flexibility of designing any filter characteristics, computational efficiency, and ease of extending the design concepts to multidimensions. In addition, since signal injection is done on individual frequency bins, the design method is especially well suited for this application.

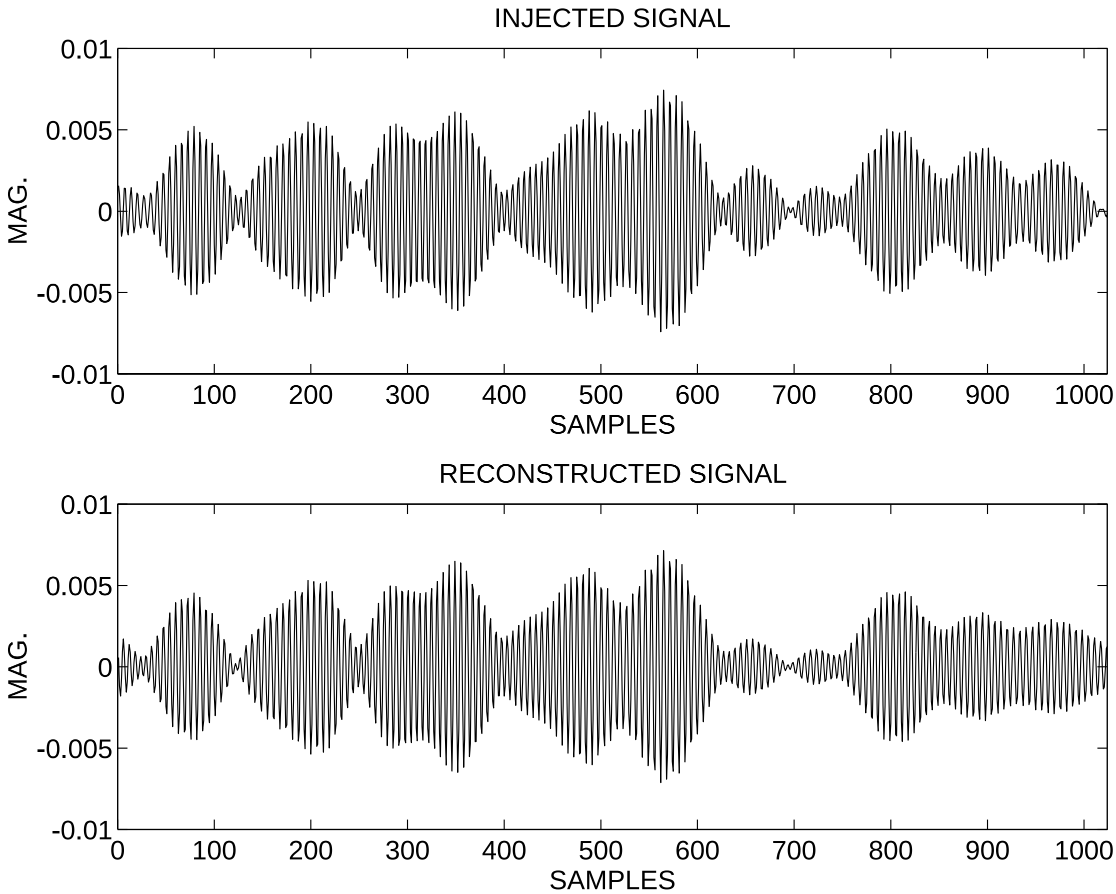

Figure 13. Signal injected to the audio and its reconstruction after decoding.

The target of the proposed signal injection method is to eliminate the need for 12 kb/s service channel in $180 + 12$ kb/s audio broadcasting where audio signal is coded with 180 kb/s. In this work, audio signal is coded with 192 kb/s which results a $2 - 3$dB SNR improvement on the average compared to the standard audio broadcasting. In addition, MSE for the injected signal after decoding is on the order of $-20dB$ without any error protection.

After the service information signal is injected to the suitable frequency bands, audio signal is coded with MPEG-1 layer 2 coder. Although there is an additional 4-bit side information for each frame of audio signal, this overhead is not significant when compared to the amount of signal transmitted inside the audio signal. Certain precautions are taken during synthesis and decoding in order to preserve the perceptual quality of the audio and injected signal. The results obtained in the experiments show that signal injection effectively increases the signal transmission efficiency while preserving the audio signal quality. Furthermore the quality of the inserted signal after decoding is high which shows the potential of the proposed approach.

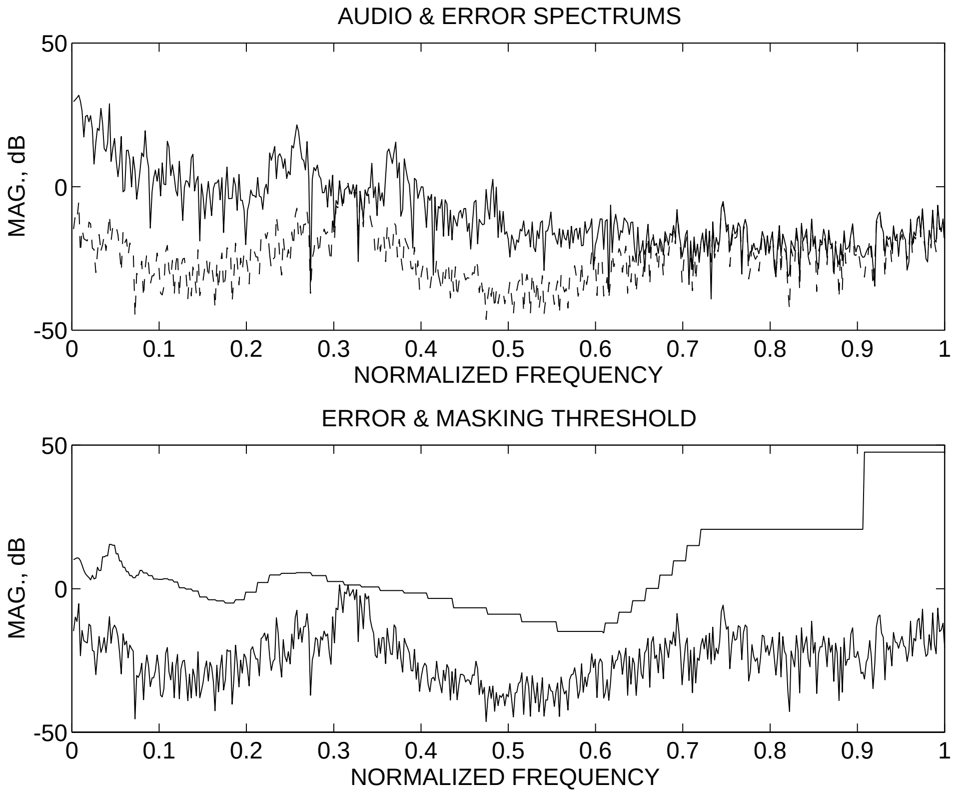

Figure 14. a. Spectrums of the original audio and error after signal injection and decoding, b. Global masking threshold and error spectrum.

References

[1] Andreas S. Spanias, “Speech Coding: A Tutorial Review”, Proc. of IEEE, Vol.82, No 10., Oct. 1994.

[2] Nikil Jayant, James Johnston, Robert Safranek, “ Signal Compression Based on Models of Human Perception”, Proc. of IEEE, Vol. 81, No 10, Oct. 1993.

[3] J.D. Johnston, K.Brandenburg, “Wideband coding: Perceptual considerations for speech and music”, Advances in Speech Sig. Proc., pp.109-140, New York:Dekker, 1992.

[4] N.S. Jayant, J.D. Johnston, R. Safranek, “Coding of wideband speech”, Speech Comm. Vol.11, pp.127-138, 1992.

[5] Allen Gersho, “Advances in speech and audio compression”, Proc. of IEEE, Vol. 82, pp.900-918, June 1994.

[6] Peter Noll, “Wide Band Speech and Audio Coding”, IEEE Comm. Mag. ,pp.34-44, Nov. 1993.

[7] Peter Noll, “Digital audio coding for visual communications”, Proc. of IEEE, Vol.83, pp.925-943, June 1995.

[8] James D. Johnson, “Transform coding of audio signals using perceptual noise criteria”, IEEE J. Sel. Ares in Comm.,Vol. 6, pp.314-323, Feb. 1988.

[9] ISO/IEC JTCI/SC29, “Information Technology-Coding of moving pictures and associated audio for digital storage media up to 1.5Mb/s-IS 11172-3 (audio)”, 1992.

[10] T. Engin Tuncer, “ An audio coding system for MPEG-1 audio standard”, Int. Conf. Telecomm., ICT-96, pp. 144-147, April 1996.

[11] R.D.Koilpillai, P.P. Vaidyanathan, “Cosine-modulated FIR filter banks satisfying perfect reconstruction”, IEEE Trans. on Signal Proc., Vol.40, No.4, April 1992.

[12] T.Q.Nguyen, R.D.Koilpillai, “The design of arbitrary-length cosine-modulated filter banks and wavelets, satisfying perfect reconstruction”, Proc. of IEEE-SP Int. Symp. on Time-Frequency and Time-Scale Analy., pp. 299-302, Oct. 1992.

[13] T.Q. Nguyen, “Near-perfect reconstruction Pseudo-QMF Banks”, IEEE Trans. on Signal Proc., Vol. 42, pp. 65-75, Jan 1995.

[14] M. Sablatash, T. Cooklev, “Compression of high quality audio signals including recent methods using wavelet packets”, Digital Signal Proc., Vol.6, pp.96-107, 1996.

[15] Z. Xiong, K. Ramchandran, C. Herley, M.T. Orchard, “ Flexible tree-structured signal expansions using time-varying wavelet packets”, IEEE Trans. on Signal Proc., Vol. 45, No 2, pp.333-345, Feb. 1997.

[16] I. Sodagar, K. Nayebi, T.P.Barnwell, M.J.T. Smith, “Time-varying analysis-synthesis systems based on filter banks and post filtering”, IEEE Trans. on Signal Proc., Vol. 43, No. 11, pp.2512-2524, Nov. 1995.

[17] K. Yamauchi, S. Kakiuchi, A. Takebe, M. Sugitomo, “ Digital Audio Broadcasting Receiver Development”, Int. Broadcasting Conv., Sept. 14, 1995.

[18] Walter Bender, Daniel Gruhl, N. Morimoto, “ Techniques for data hiding”, Proc. of the SPIE, Vol. 2420, No.40, Feb.1995.

[19] Laurence Boney, Ahmed H. Tewfik, K.N. Hamdy, “ Digital Watermarks for audio signals”, EUSIPCO-96, Italy, Sep. 1996.

[20] Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik, L. Boney, “ Robust audio watermarking using perceptual masking”, Signal Processing, Vol.66, pp.337-355, 1998.

[21] Mitchell D. Swanson, Mei Kobayashi, Ahmed H. Tewfik, “ Multimedia data-embedding and watermarking technologies”, Proc. of IEEE, Vol.86, No.6, June 998.

[22] B. Scharf, “ Critical bands”, Foundations of Modern Auditory Theory, Academic, pp.159-202, 1970.

[23] D.D.Greenwood, “Critical bandwidth and frequency coordinates of basilar membrane”, Journal of Acoust. Soc. Am., Vol.33, pp. 1344-1356, 1961.

[24] M.R. Schroeder, B.S. Atal, J.L. Hall, “ Optimizing digital speech coders by exploiting masking properties of the human ear”, J.Acoust. Soc. Am., pp.1647-1652, 1979.

[25] T.Engin Tuncer, “Duyumsal model ile isaret ekleme”, SIU-97, Kusadasi, Nisan 1997.