Sub-band coding for audio signals using Matlab

D Cardozo\textsuperscript{1}, S Campo\textsuperscript{2}, J Manjarres\textsuperscript{2} and W Percybrooks\textsuperscript{2}

\textsuperscript{1} Fundación de Estudios Superiores Comfanorte, San José de Cúcuta, Colombia
\textsuperscript{2} Universidad del Norte, Barranquilla, Colombia

E-mail: do_cardozo@fesc.edu.co

Abstract. This work presents the results of a research project aimed at studying the coding of audio signals in order to obtain a compact and efficient digital representation for transmission or storage purposes. A psychoacoustic model based on sub-band coding is implemented in MATLAB\textsuperscript{\textregistered}, which identifies the type of audio: voice or music signal, based on a model of the human auditory system. Our focus was on measuring the relationship between reduced transmission/storage bitrate and audio quality.

1. Introduction

Since the decade of 1980, different encoding methods have been developed for audio signals, in order to represent the signals with the lowest possible transmission rate without affecting their perceptual quality. At that time, digital audio coding systems had good quality and low bit rate. Contributions like those of the MPEG (Moving Pictures Expert Group) have intensified this line of research from which several international audio coding standards have evolved as: MPEG-1 audio (ISO/IEC 11172-3, 1992), MPEG-2 audio (ISO/IEC 13818-3, 1994) and MPEG-4 (ISO/IEC 14496, 2000) \cite{1}.

During the last 20 years, the need to reduce the binary rate of the audio signal in order to transmit the largest possible number of audio channels through digital networks (internet or mobile networks) has increased. To achieve this goal, new coding technologies have been developed, starting with the use of time-frequency transforms for coding the waveform of signals, then moving on to signal models to extract parameters from the audio signal and finally encode them. These coding technologies present very good performance due to the great compaction of energy by supposing a model of three components: tones, transients and noise. On the other hand, the application of these parametric encoders minimizes the complexity of making changes to the signal from its parameters, facilitating the recognition and separation of sources \cite{1}.

There are two types of parametric encoders: sub-band coder (compression with loss), which divides the band of signals into frequency sub-bands using filters, then the output of each filter is sampled and coded independently; and transform encoder (compression without loss), which encodes the components of a unitary transform of the signal and the decoder performs the inverse transform \cite{2}.

Some loss-compression schemes focus on using the peculiarities of the human ear to make imperceptible loss of signals, for example the Yamaha VQF, Musepack \cite{3}, Vorbis \cite{4}, AAC \cite{5} and the well-known MP3\cite{6}. Since the entropy of the signals varies over time, the sampling rate is accommodated according to the demands of the signals, and this is because some parts contain more information than others.

The work presented here includes an algorithm developed in MATLAB\textsuperscript{\textregistered} which applies the sub-band comprehension technique for digital audio and voice signals, in which, depending on the input,
different sub-band schemes are defined. By means of a statistical analysis, the efficiency of the developed algorithm is evaluated.

2. Conceptual framework

2.1. Digital audio
The audio signals are continuous due to their nature, leading to the need of using sampling to obtain the value of the wave amplitude in established time periods. Quantization is used for the discretization of an amplitude value whose precision is finite [1].

2.1.1. Uniform quantizers. This type uses a fixed quantizer step. If the wave values are not integers, they are rounded to the nearest value [1].

2.1.2. Audio encoders. The audio encoding combines sampling, quantization and storage. Some modern methods such as that used in CDs record 1 bit of data with sampling frequency near 2822.4kHz, giving a dynamic range of 120dB [7].

2.1.3. Objective quality of digital audio representation. A metric to measure the similarity between the original signal and its digital version is the signal-to-noise ratio (SNR), which is defined in Equation (1) and Equation (2) [8]:

\[
SNR = 10 \log_{10} \frac{x_{in}^2}{q^2} \, dB
\]  

(1)

\[
q(t) = x_{out}(t) - x_{in}(t)
\]  

(2)

If the error (noise) \(q(t)\) is small, the SNR will have a large value. The CD audio has a SNR of around 96dB, while the DVD audio is capable of achieving a maximum SNR of 144dB [7]. To determine the SNR in digital audio, the following Equation (3) is used [8].

\[
SNR = 10 \log_{10} \left( \frac{P_{in}}{q^2} \right)
\]  

(3)

Where \(P_{in}\) is the power of the total input signal, \(q\) is the constant that divides the size of the signal.

2.2. Basic principles of perceptual coding
The objective of perceptual coding is to achieve a minimum transfer rate with high quality, by exploring the peculiarities of the human ear to compress part of the signal. There are models that describe the behavior of the human ear and discard information that is not perceptible. Figure 1 describes the audio perceptual coding model, composed of blocks as analysis, quantification and coding filter bank, perceptual model and binary flow coding [9].

![Figure 1. Model of audio perceptual coding [9].](image-url)
2.2.1. **Sub-band coding.** It consists of using a bank of filters to divide the total frequency band of the signal into a series of sub-bands, where each sub-band is adaptively coded. This filter bank presents no overlapping and covers the whole signal bandwidth, so the resulting sub-bands can be combined to recover the original signal obtaining a little perceptible distortion [9].

2.2.2. **Perceptual model (psycho-acoustic).** Simultaneously with the bank of filters, the signal goes through a block designated with the perceptual model, as shown in Figure 1. The function of this block is to apply the rules of psycho-acoustic behavior of the human ear and establish the necessary masking levels in each sub-band. This signal enters the quantifier by adjusting the quantization level of each sub-band, so that all the quantization noise does not exceed the masking thresholds [9]. The psycho-acoustic characteristics establish that, in the human perceptible frequency range from 20Hz to 20kHz, the hearing threshold is low at frequencies close to the voice frequencies and high at low and high frequencies. The human ear can perceive sound waves with intensity greater than or equal to 1pW/m², calling this intensity an auditory threshold, which compared to the pain threshold (10W/m²) is very small.

2.2.3. **Auditory threshold.** The auditory threshold varies depending on the frequency, mainly affected by the sensitivity of the human ear, as can be seen in Figure 2. This threshold indicates the level of pressure necessary for sounds with certain frequencies to be perceived [10,11]. For the methodology proposed in this work, the channel coding was established to overcome an amplitude according to the threshold of silence described in Figure 2.

![Figure 2. Auditory threshold [11].](image)

3. **Proposed methodology**

The algorithm for the sub-band coding model determines whether the digital input signal is music or voice, as described in Figure 3. It should be noted that the coding is not in real time.

![Figure 3. Diagram of algorithm functions.](image)

In the analysis section of the sub-band coder, for the digital music audio input, the configuration of Figure 4 was used.
Figure 4. Sub band encoder for music signals [11]

The configuration of Figure 5 was used for a digital audio voice input.

Figure 5. Sub band encoder for voice signals [11].

To make a uniform decomposition of the bands, a low-pass filter and high pass filter is applied and a decimator of value 2, as described in Figure 6.

Figure 6. Decomposition of the signal accompanied by a filter and a decimator [11].
The digital filter was obtained MATLAB® filter design tool, as shown in Figure 7. The frequency response on Figure 7 corresponds to the low-pass filter, while the high-pass filter is obtained by complementing its coefficients to 1.

\[ h_0(n) = h(n) \]  
\[ h_1(n) = (-1)^n h(n) \]

Where \( h(n) \) is the designed low pass filter, \( h_0(n) \) and \( h_1(n) \) are the low pass and high pass filters, respectively.

\[ g_0(n) = 2h(n) \]
\[ g_1(n) = -2(-1)^n h(n) \]

In this case, \( g_0(n) \) and \( g_1(h) \) are the low pass and high pass filters, respectively, for the synthesis section.

4. Results
For the tests of the designed sub-band encoder, 5 audio files were used. Each file had 16bits per sample and were compared in two ways. The first comparison is performed using an established compression ratio for each file. The second test was a perceptual poll for the audio quality, by randomly assessing people in a university space. The results of the output bits versus the input and the compression ratio are shown in Table 1, this ratio was calculated for each audio by referencing the input bits as 100%.

| File | Input bits | Output bits | Compression |
|------|------------|-------------|-------------|
| Music |            |             |             |
| Music 1 | 14112000  | 8820000    | 38%         |
| Music 2 | 11995200  | 7590753    | 37%         |
| Music 3 | 11289600  | 7232400    | 36%         |
| Voice |            |             |             |
| Voice 1 | 3211264   | 1605632    | 50%         |
| Voice 2 | 6062080   | 3220480    | 47%         |
4.1. Music signal
Tests were performed with three music signals, with length between 10 and 20 seconds, to a group of people in the same space (Universidad del Norte library). The people listened to the original version and the compressed version and gave their rating from five possible options. The results of the survey for the three signals are shown in Table 2. Figure 8 describes, in percentage values, the rating of the compressed signals.

| Scale | Quality | Classifications |
|-------|---------|-----------------|
| 5     | Excellent | 9               |
| 4     | Good     | 14              |
| 3     | Regular  | 1               |
| 2     | Enough   | 0               |
| 1     | Bad      | 0               |
| Total |          | 24              |

Figure 8. Percentage evaluation according to the music signal poll.

4.2. Voice signal
Tests were performed with two voice signals, between 5 and 8 seconds, to a group of people in the same space (Universidad del Norte library). The people listened to the original version and the compressed version and gave their rating, see Table 3, according to the survey applied. Figure 9 describes, in percentage values, the rating of the compressed signal.

| Scale | Quality | Classifications |
|-------|---------|-----------------|
| 5     | Excellent | 0               |
| 4     | Good     | 5               |
| 3     | Regular  | 15              |
| 2     | Enough   | 3               |
| 1     | Bad      | 1               |
| Total |          | 24              |

Figure 9. Percentage evaluation according to the voice signal poll.

5. Conclusion
The performance tests of the algorithm, for different types of digital signal, reveals that it is possible to compress the digital signals while maintaining a good subjective quality according to the psychoacoustic levels, suitable for the human ear.

The designed encoder has a perceptual quality for audio files of good music, with compression between 36% and 38% of the bits, evidencing a pleasant audio quality to people, but of very low compression. As for the voice audios, compression was around 50%; however, the audio quality perceived by the people in the tests is mostly regular.
References

[1] Vera P 2006 Development of audio coding techniques based on parametric signal models (Spain: University of Alcalá)
[2] España Boquera M C 2003 Servicios avanzados de telecomunicación (Madrid: Diaz de Santos S.A.)
[3] Musepack Living audio compression 2018 consulted on: http://musepack.net/
[4] The ogg vorbis codec project 2017 consulted on: http://www.vorbis.com/
[5] International Organization for Standardization 1999 Coding of audio-visual objects-Part 3: Audio ISO/IEC 14496-3 (Switzerland: International Organization for Standardization)
[6] International Organization for Standardization 1993 Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s-Part 3: audio ISO/IEC 11172-3 (Switzerland: International Organization for Standardization)
[7] Konstantinides K July 2003 An introduction to super audio cd and DVD-audio IEEE Signal Processing Magazine 20 71-82
[8] Bosi M and Goldberg R E 2003 Introduction to digital audio coding and standards (Boston: Springer)
[9] Gold B Morgan N and Ellis D 2011 Speech and audio signal processing: Processing and perception of the speech and music (New Jersey: Jhon Wiley & Sons)
[10] Terhardt E. 1979 Calculating virtual pitch Hear. Res. 1 155-182
[11] Espinosa Camacho L 2003 Compresion de voz para su transmission en redes de datos (México: Universidad de Nuevo Leon)
[12] Pavel Z and Miroslav V 2012 Perfect decomposition narrow-band FIR filter banks audio IEEE Transactions on Circuits and Systems II: Express Briefs 59 805-809