INVERTING DYNAMICS COMPRESSION WITH MINIMAL SIDE INFORMATION

Benoit Lachaise and Laurent Daudet

UPMC Univ. Paris 06
IJLRA / Lutheries, Acoustique, Musique
11 rue de Lourmel, 75015 Paris, France
benoit.lachaise@etu.univ-nantes.fr;daudet@lam.jussieu.fr

ABSTRACT

Dynamics processing is a widespread technique, both at music production and diffusion stages. In particular, dynamic compression is often used in such a way that the “average” listener can best enjoy the music. However, this may lead to an excessive use of compression, especially with respect to listeners in quiet listening conditions. This paper presents estimates on the amount of extra data that is needed to invert the effects of such non-linear processing, using simple blind identification techniques. We present two simple test cases, first in the case when perfect reconstruction is needed, and second when the ancillary data rate is constrained, leading to an approximate reconstruction.

1. INTRODUCTION

A common complaint amongst artists and sound engineers is that dynamics compression (hereafter just called compression) is often being overly used. This is not only true at the production stage: as extra compression is almost systematically added by radio stations. There are many reasons for the use of compression: first, in a noisy environment (e.g. listening to music in a car or on an iPod-type portable device in the street), or with a constrained transmission channel (e.g. the maximum modulation for FM radios has to obey national regulations, such physical limitations were also present on vinyl LPs), compressed music renders most of the music content without having to constantly change the volume between soft and loud passages. Second, this gives a timbral identity to certain types of music and / or radio stations. Third, sound engineers / producers use high compression because they don’t want their music to sound dull compared to their competitors’ highly compressed music ; a typical vicious circle that, according to many, has gone too far. This feeling is shared by a number of listeners who like to listen to music on medium- to high-end equipment, in low background noise levels. For these listeners, in large numbers though not the majority, there is the feeling that this is “too late”. Indeed, due to the non-linearities of dynamics processing in finite precision, some information has been irreversibly lost. Setting a dynamics expander after a compression does not recover the original sound, and often results in the so-called “pumping” sound; which is often disgraceful when unintentional.

Most of the techniques that have been proposed (see [1]) so far to weaken these effects are based on the idea that listeners would be given the combination { original sound + processing parameters }: in this way each user can decide whether or not to apply dynamics processing. However, this idea has never caught up in practice. Amongst many explanations, it can be suggested that this is due to a simple off-balance between the number of users and the extra “cost” (in terms of transmission / storage bandwidth) required for sound processing parameters: such a system imposes on everyone for the benefit of few. In other words, for the majority of people for which high compression is useful / desirable, an extra amount of info is needed ; this system is perfectly designed (i.e. with no overhead) only for the minority of “audiophiles”. Another drawback of this system is that it often restricts the type of processing that can be used, and/or allows the explicit transmission of all processing operations, which is often considered as a “trademark” of a sound engineer / music label / radio station.

In this paper, we investigate the opposite scheme: the data transmitted / stored is { compressed sound + reverse-processing parameters }. The main advantage of such a scheme is that it is entirely backwards compatible: it doesn’t require any change for the majority of users who are happy with the compressed sound, who can just ignore the extra parameters. The “hi-fi” listeners, with properly equipped devices -and possibly higher transmission / storage bandwidth-, can choose to cancel the dynamics processing thanks to the extra set of parameters. Also, it is totally independent of the details of the processor used, since our system is based on a “blind” estimation of non-linearities [2]: the ancillary parameters are derived with no *a priori* knowledge on the processor, apart the fact that this is based on an instantaneous level-dependent gain. In this way, the fine “trademark” details of the successive hardware or software stages in studio compression techniques also remain hidden. The drawback of this technique is that the amount of ancillary data for reverse-processing is *a priori* significantly higher than in the first scenario.

The goal of this paper is to present preliminary results that quantify the amount of extra data for dynamics compression reverse-processing. This data is divided in two parts: first an estimate of the instantaneous gain, that has to be subsampled and quantized at finite precision. Second, using this approximate gain one only gets an estimate of the original signal: one should encode also the residual between estimated ant true original. Two test cases have been studied. In the “lossless” scenario, we have investigated how much extra data is needed to exactly recover the original signal. This scenario is appropriate in a digital storage / transmission context. In the “lossy” scenario, we are given stringent bitrate constraints for the ancillary data, and try to get as close as possible to the original sound. This may be relevant for instance when trying to invert compression in FM radio using ancillary data transmitted in the RDS channel.

The rest of this paper is constructed as follows. In section 2 we briefly recall the principle of dynamics compression. Section 3 introduces the way to estimate the reverse-processing parameters. Section 4 gives the details of numerical experiments. Results regarding the sound quality at constrained bitrate are presented section 5. In section 6, we evaluate on typical soundfiles the amount of ancillary data in the "lossless" scenario. Finally, we conclude (section 7) on limitations and possible improvements.

2. PRINCIPLES OF DYNAMIC PROCESSING

Dynamic processing is an amplification system whose gain is automatically controlled by the input signal level. This level is calculated by an envelope follower [3] (see figure 1), which calculates the mean of the absolute value of input signal $x(n)$ on a certain time interval. The relation between output and input signals is defined by the function $y(n) = f(x(n)) = x(n).g(n)$, which in general is non-linear. Time signals $x(n)$, $y(n)$ and instantaneous gain $g(n)$, can be formulated with level values X, Y, and G in dB. These values are the logarithm of the root mean square $x_{rms}(n)$ or peak value $x_{peak}(n)$ of the time signals according to $X = 20.\log(x)$. The multiplication $y = x.g$ can be regarded as an addition in the logarithmic domain: $Y = X + G$. This way, the dynamics range controller can be illustrated by static functions as in figure 2.

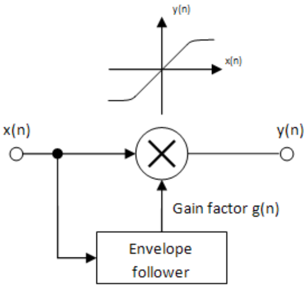


Figure 1: Block diagram of the nonlinear operations performed in dynamics processing (from [3]).

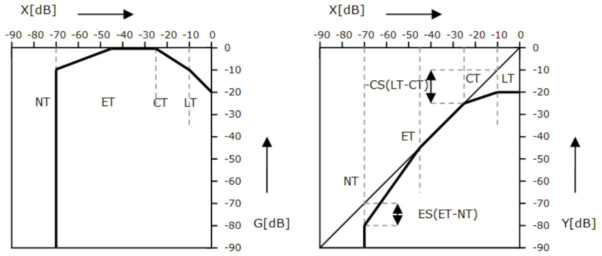


Figure 2: Static characteristic of a dynamic range controller (from [3])

A simple compressor can be described by four principal thresholds:

- LT (limiter threshold): to limit output signal at a maximum level
- CT (compressor threshold): to reduce the dynamic of the output signal in a certain interval of the input signal level.
- ET (expander threshold): to increase the dynamic of the output signal in a certain interval of the input signal level.
- NT (noise gate threshold): to cut the very low sound, especially the noise.

Dynamics processing is often based on multi-band frequency selection. In this case, each frequency band is affected by a different static curve. For the sake of simplicity, only single band compression is studied in this article, but the methods presented here can be extended in a straightforward way to the case of multi-band dynamics processing. Moreover, we only work with mono signals.

We will note x the original signal, y the compressed signal, and g the instantaneous gain.

3. ESTIMATION OF COMPRESSION PARAMETERS

In order to invert the compression, we are going to transmit as side information two sets of data:

- The approximate gain factor $\tilde{g}(n)$, which is the ratio between the compressed signal amplitude and the original signal amplitude (see part 3.1).
- A residue, which is the error between estimated and true original signal: $r(n) = x(n) - y(n)/\tilde{g}(n)$. r is not equal to zero due to roundoff errors and bad estimation of the true gain $g(n)$ (see part 3.2).

Because we want our system to be as generic as possible, and in particular independent of particular type / brand / fine tuning of compressor, we assume that the only signal we have at hand are the original signal x and the compressed signal y. Figure 3 shows a general process of a compression inverter. The encoder creates the two parameters, and encodes them as described in sections 5 and 6.

This process can be used in two scenario:

- In the case of a limited capacity metadata channel, such as the RDS digital metadata channel for FM radio, we have to think the process as a lossy process. In such a low rate constrained channel, we can choose either to transmit only the down-sampled and quantized gain factor, or to transmit also a quantized residue which implies an even cruder quantization of the gain. For this scenario, we want to minimize the error between the estimated uncompressed signal and the true original uncompressed signal, under a maximum rate constraint (see part 5).
- In the case of high rate (potentially unbounded) metadata channel, we can think the process as a lossless process. Here, we study what is the minimum rate of metadata for exact recontruction (see part 6).

3.1. Estimation of the gain

To find the gain factor between compressed and original signal, we have to measure their levels (see figure 3 - Encoder), as given by an RMS method. RMS levels are estimated as the mean of the signal's absolute value, on a sliding window with a fixed size. Then, as shown in the global process scheme, the gain factor is down-sampled by a ratio corresponding to the window length. Alternatively, it could have been possible to use peak values (local maxima) instead of RMS to estimate the gain - this option has been discarded here for the sake of simplicity, as it provides irregularly-spaced values of the gain. Figure 4 shows the temporal form of the gain factor, where different regimes can be obsverved.

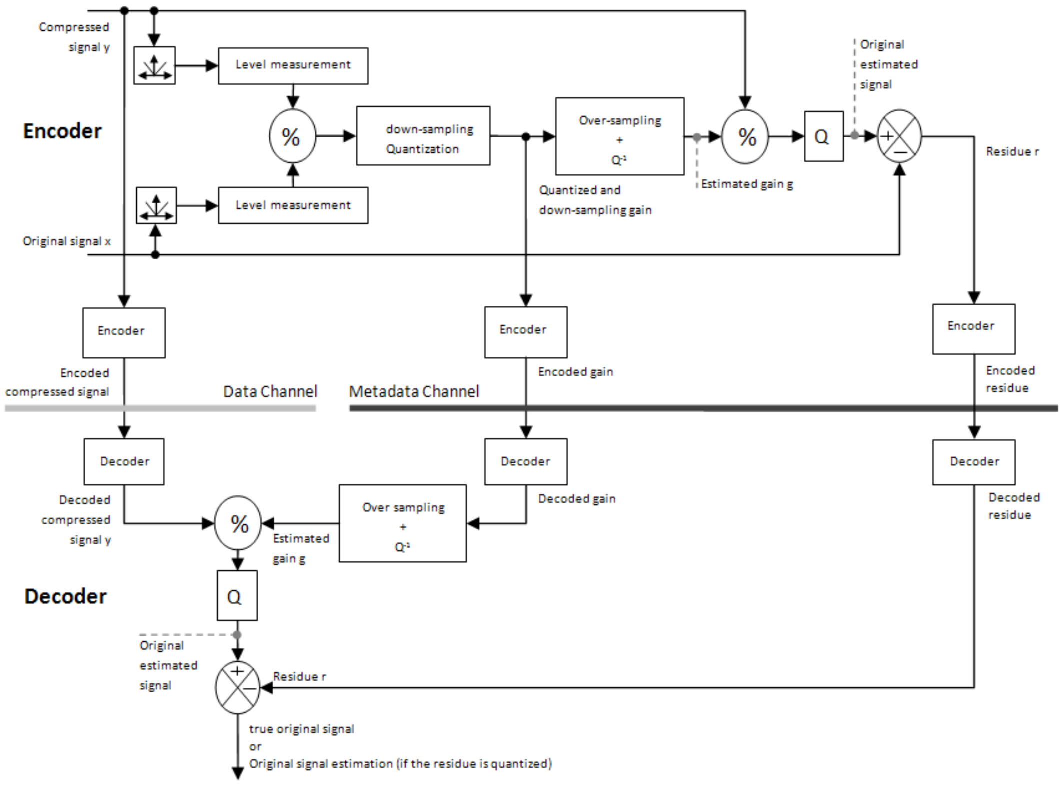


Figure 3: Global diagram of the compression inverter.

3.2. Estimation of the residue

The residue represents a parameter used to reduce or cancel the signal reconstruction error. It is transmitted at the same sampling frequency as the audio signal (44100 Hz). Figure 4 shows the temporal form of the residue when the gain has been coarsely estimated and down-sampled. As expected, the energy of this residue is higher in transient zones.

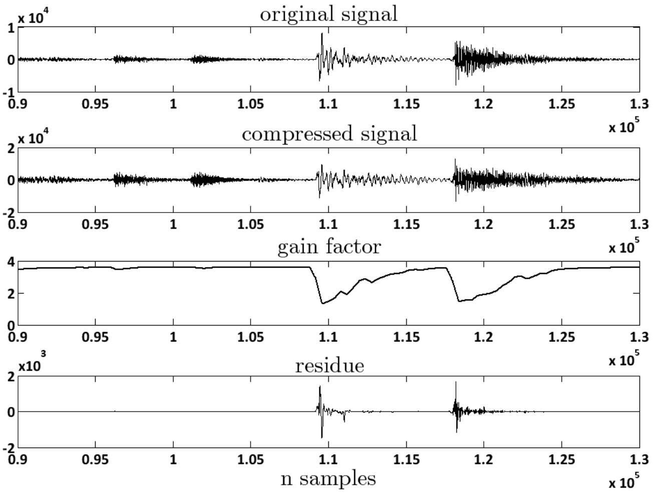


Figure 4: Variations of the gain and residue parameters.

4. EXPERIMENTAL DETAILS

In parts 5 and 6, we want to show relations between rate capacity and reconstruction quality. These two quantities are estimated as follows:

- Rates are estimated with a theoretical entropy calculation and real arithmetic encoding. In order to have precise rate calculations, the signal must be stationary, because encoding algorithms depend on signal statistics (e.g. entropy coding). Since we work on short sound sequences, it is unlikely that the asymptotic regime is reached. Therefore, the numbers given below are only rate estimates that provide a global behavior of our system (see [4]).
- The quality of the estimated original signal is measured by SNR. Especially for high distortions, this is a poor estimate of the perceived quality. Further studies will consider more complex estimates based on psycho-acoustic models, such as PEMO-Q [5]. Moreover, losses of the transmitted audio signal y (bitrate compression algorithms such as MP3, or analog channels) are not considered in the quality of the estimated original signal.

Measurements are made on two types of sounds:

- a drum sound, with sharp transients,
- a piano sound, whose level has smaller variations.

5. LOSSY ALGORITHM

Here, we study the lossy process with a simple gain transmission. The estimated gain factor is quantized with a uniform quantizer (see [6]) and transmitted to the decoder. The decoder makes the inverse quantization and interpolates the gain (here linearly) in order to create an estimation of the original signal:

$$\tilde{x}(n) = \frac{y(n)}{\tilde{g}(n)} \tag{1}$$

It can be noted that higher-order interpolations could easily be used, potentially improving the quality of the reconstructed signal.

Before showing Global SNR vs. rate results, it is interesting to show the temporal evolution of the instantaneous SNR. Figure 5 shows the error created by the gain estimation, in a temporal representation, for the two types of sound considered here. The temporal error is computed this way:

$$eps(n) = \tilde{x}(n) - x(n) = \frac{y(n)}{\tilde{g}(n)} - x(n) \tag{2}$$

and the corresponding SNR on time intervals T is:

$$segmentSNR_{dB}(k) = 10\log_{10}\left(\frac{\sum_{n=kT}^{(k+1)T} x^2(n)}{\sum_{n=kT}^{(k+1)T} eps^2(n)}\right). \tag{3}$$

If we look at the drum sequence example, we can observe that transient zones are more critical for the quality of the reconstructed signal, because of errors in the gain parameter (due to quantization and / or down-sampling). Also, we can see that the quality of the transmission decreases when the input and output levels decrease. This quality loss appears because the gain parameter is a ratio between two 16-bits quantized signals. The piano example has a higher SNR than the drum example, because the gain factor dynamics is lower. Generally, as it was expected, the higher the dynamics of the signals, the poorer the quality of the reconstructed signal.

As an overall measure, figure 6 shows which quality the process can have for a maximum of 1.2 kbit/s, which is the capacity of the metadata RDS channel in FM radio. Global SNR represents the quality as measured by the average of the local SNR. The rate is first estimated according to the signal's entropy, and then computed with a real arithmetic encoder [7, 8]. This arithmetic function gives a vector of 8 bits elements.

The estimated rate is defined this way:

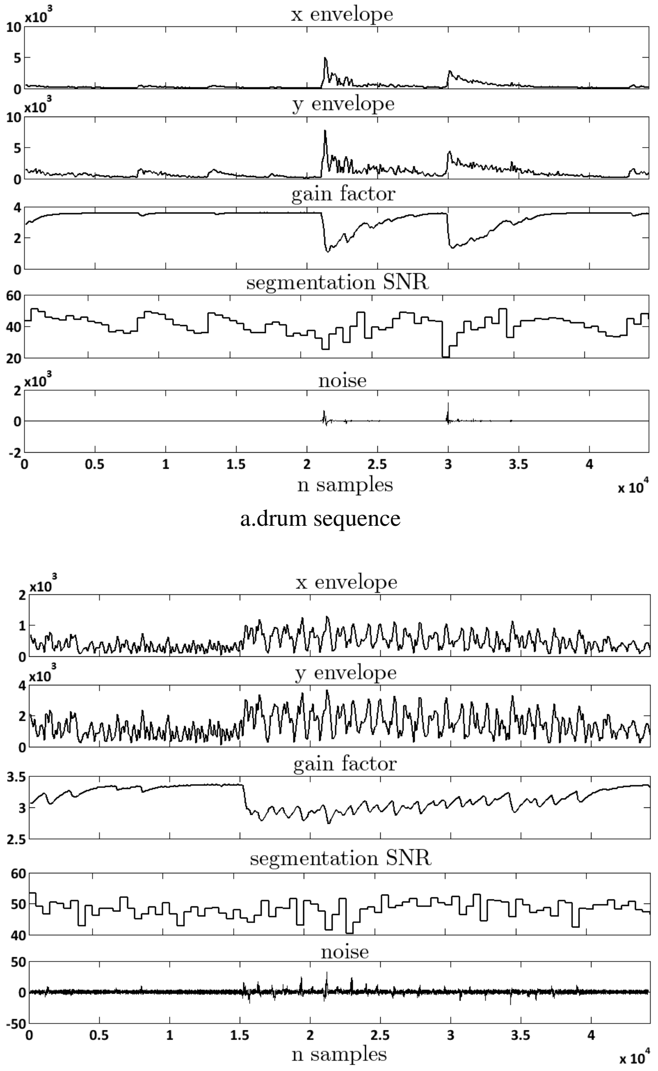
$$E = -\sum_{i=1}^{n} p(i)\log_2(p(i)) \tag{4}$$

n being the number of possible states

$$rate_{estimated} = \frac{44100 \cdot E}{N} + \frac{32+32}{T} \tag{5}$$

N being the gain elements number
T being the length of the sound(seconds)
And the real rate this way:

$$rate_{real} = \frac{8 \cdot L + 32 + 32}{T} \tag{6}$$



a.drum sequence

b.piano sequence

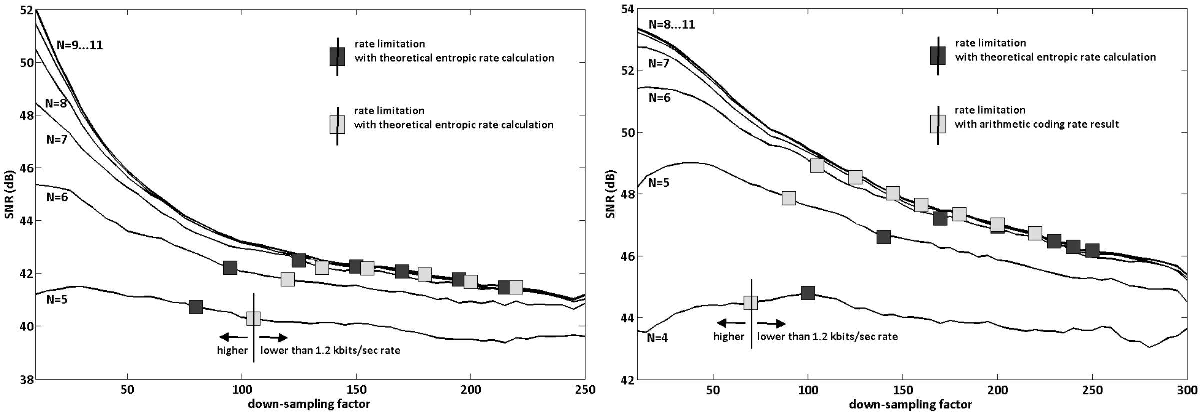
Figure 5: Instantaneous SNR

L being the length of the arithmetic coded vector
T being the length of the sound(seconds)
The two therms of 32 bits correspond to the transmission of the maximum and minimum values of the gain factor. We chose 32 bits to have a very good precision, this having an influence on the precision of all the frame.

Squares on curves represent the rate limitations of 1.2 kbit/s (RDS channel): For example, in figure 6, if the gain factor is quantized on 5 bits (N=5), the square shows that the process can only downsample by up to 105 factor for the drum sequence, and by up to 70 factor for the piano sequence.

Curves show the results with entropy estimates (dark squares), and real averaged rate of an arithmetic encoder (white squares).

These results give us some indications on the expected behavior of our system for very low rate transmission channels such as the RDS channel, if we transmit only the gain factor with very simple quantization / encoding. Several other encoding methods could be used that could potentially bring a significant benefit in terms or



a.drum sequence b.piano sequence

N = nb of bits for the quantization of the gain

Figure 6: Quality estimates of the inverted signal for a 1.2 kbit/s channel

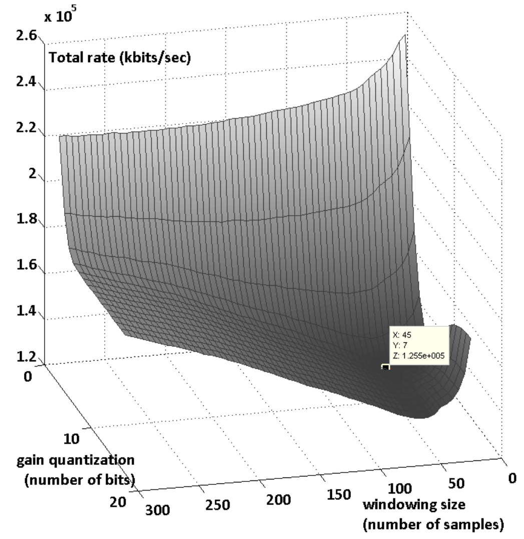
rate / SNR, such as predictive differential encoding [4].

6. LOSSLESS ALGORITHM

In a lossless process, the two parameters gain and error are sent, without quantizing the residue, in order to reconstruct exactly the original signal. The goal in this context is to find the minimum global rate allowing a lossless reconstruction, by studying the entropy of the two transmitted parameters. The total rate is the sum of the gain rate and the residue rate. If the gain down-sampling factor and the gain quantization decrease, the gain rate decreases ; however, because of precision loss, the residue entropy, so its rate, will increase. The total rate admits therefore a minimum value, with a good balance between quantization and down-sampling of the gain.
Figure 7 show a global approximation of the rate by entropy estimates. It provides us with an estimation of the optimal parameters, around a factor 45 for the down-sampling, and a 7 bits gain quantization.
The minimum rate allowed by this process is about 125 kbit/s, which is the average rate of an audio transmission. With the real arithmetic coder, the allowed minimum rate for the same parameters is similar, at 124.4 kbit/s. Given the nominal rate of PCM data at CD quality of around 705 kbit/s, these rate estimates can be considered as relatively high (about 18 %). Note that for stereo signals, this number would likely be significantly reduced, as the gain can be considered roughly equal between channels.

7. CONCLUSION

This article presents preliminary results on the amount of extra data that is needed to invert, exactly or approximately, a typical dynamics compression applied on music signals. Simple scenarios have been studied, that give us a first estimate on the potential of these methods. In a tight rate-constrained channel such as the RDS channel that has a capacity of 1.2 kbit/s, we have shown that it is possible to reverse the compression with an overall SNR quality of approximately 50dB. For a lossless transmission, the data rate can be as high as 125 kbit/s.



drum sequence

Figure 7: Rate minimization for a lossless process

The algorithms that have been presented in this article are extremely simple. Future work will focus on more efficient methods, for instance with a better prediction of the gain. More importantly, as the relatively high rates for exact reversibility makes it inconvenient for practical uses, our work will focus on approximate reconstruction, with much smaller data rates. In this case, listening tests with audio experts will be necessary to tune the systems for the better perceived accuracy.

8. ACKNOWLEDGMENTS

The authors wish to acknowledge the help of Camille Dianoux (FSMS) and Edwige Roncière (Radio France) for the insightful discussions on the uses of dynamics compression by audio engineers. We also wish to thank Emmanuel Ravelli for his help on implementation of entropy coding.

9. REFERENCES

[1] *Dolby Digital and Dolby Volume Provide a Comprehensive Loudness Solution*, Dolby - White Paper, January 2007.

[2] Uwe Simmer, Denny Schmidt, and Joerg Bitzer, *Parameter Estimation of Dynamic Range Compressors: Models, Procedures and Test Signals*, Proc. AES 120th Convention, Paris, France, May 2006.

[3] DAFX-Digital Audio Effects, Udo Zölzer ed., John Wiley & sons, 2003.

[4] *Codage Audio Stéréo Sans Perte*, student internship report, Jean-Luc Garcia, ENST Bretagne option Signal et Communications, 5 septembre 2003.

[5] Rainer Huber and Birger Kollmeier, *PEMO-Q-A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception*, IEEE transactions on audio, speech, and language processing, vol. 14, no. 6, november 2006.

[6] Allen Gersho And Robert M. Gray, Vector quantization And Signal Compression, Kluwer Academic Publishers, 1992.

[7] Introduction to Arithmetic Coding - Theory and Practice, Amir Said, Imaging Systems Laboratory HP Laboratories Palo Alto,hpl-2004-76,April 21, 2004.

[8] Lossless Compression Handbook, K. Sayood ed., Chapter 5: Arithmetic Coding (A. Said), pp. 101-152, Academic Press, 2003.