Introduction

In everyday life, we often come in contact with compressed signals: when using mobile telephones, mp3 players, digital cameras, or DVD players. The signals in each of these applications, telephone-band speech, high fidelity audio signal, and still or video images are not only sampled and quantized to put them into a form suitable for saving in mass storage devices or to send them across networks, but also compressed. The first operation is very basic and is presented in all courses and introductory books on signal processing. The second operation is more specific and is the subject of this book: here, the standard tools for signal compression are presented, followed by examples of how these tools are applied in compressing speech and musical audio signals. In the first part of this book, we focus on a problem which is theoretical in nature: minimizing the mean squared error. The second part is more concrete and qualifies the previous steps in seeking to minimize the bit rate while respecting the psychoacoustic constraints. We will see that signal compression consists of seeking not only to eliminate all redundant parts of the original signal but also to attempt the elimination of inaudible parts of the signal.

The compression techniques presented in this book are not new. They are explained in theoretical framework, information theory, and source coding, aiming to formalize the first (and the last) element in a digital communication channel: the encoding of an analog signal (with continuous times and continuous values) to a digital signal (at discrete times and discrete values). The techniques come from the work by C. Shannon, published at the beginning of the 1950s. However, except for the development of speech encodings in the 1970s to promote an entirely digitally switched telephone network, these techniques really came into use toward the end of the 1980s under the influence of working groups, for example, “Group Special Mobile (GSM)”, “Joint Photographic Experts Group (JPEG)”, and “Moving Picture Experts Group (MPEG)”.

The results of these techniques are quite impressive and have allowed the development of the applications referred to earlier. Let us consider the example of

a music signal. We know that a music signal can be reconstructed with quasi-perfect quality (CD quality) if it was sampled at a frequency of 44.1 kHz and quantized at a resolution of 16 bits. When transferred across a network, the required bit rate for a mono channel is 705 kb/s. The most successful audio encoding, MPEG-4 AAC, ensures "transparency" at a bit rate of the order of 64 kb/s, giving a compression rate greater than 10, and the completely new encoding MPEG-4 HE-AACv2, standardized in 2004, provides a very acceptable quality (for video on mobile phones) at 24 kb/s for 2 stereo channels. The compression rate is better than 50!

In the Part 1 of this book, the standard tools (scalar quantization, predictive quantization, vector quantization, transform and sub-band coding, and entropy coding) are presented. To compare the performance of these tools, we use an academic example of the quantization of the realization $x(n)$ of a one-dimensional random process $X(n)$. Although this is a theoretical approach, it not only allows objective assessment of performance but also shows the coherence between all the available tools. In the Part 2, we concentrate on the compression of audio signals (telephone-band speech, wideband speech, and high fidelity audio signals).

Throughout this book, we discuss the basic ideas of signal processing using the following language and notation. We consider a one-dimensional, stationary, zero-mean, random process $X(n)$, with power σ_X^2 and power spectral density $S_X(f)$. We also assume that it is Gaussian, primarily because the Gaussian distribution is preserved in all linear transformations, especially in a filter which greatly simplifies the notation, and also because a Gaussian signal is the most difficult signal to encode because it carries the greatest quantization error for any bit rate. A column vector of N dimensions is denoted by $\underline{X}(m)$ and constructed with $X(mN) \cdots X(mN+N-1)$. These N random variables are completely defined statistically by their probability density function:

$$p_X(\underline{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \mathbb{R}_X}} \exp(-\frac{1}{2}\underline{x}^t \mathbb{R}_X^{-1} \underline{x})$$

where $\mathbb{R}_X$ is the autocovariance matrix:

$$\mathbb{R}_X = \mathbb{E}\{\underline{X}(m)\underline{X}^t(m)\} = \begin{bmatrix} r_X(0) & r_X(1) & \cdots & r_X(N-1) \\ r_X(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_X(1) \\ r_X(N-1) & \cdots & r_X(1) & r_X(0) \end{bmatrix}$$

Toeplitz matrix with $N \times N$ dimensions. Moreover, we assume an auto-regressive process $X(n)$ of order P, obtained through filtering with white noise $W(n)$ with variance σ_W^2 via a filter of order P with a transfer function $1/A(z)$ for $A(z)$ in the form:

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_P z^{-P}$$

The purpose of considering the quantization of an auto-regressive waveform as our example is that it allows the simple explanation of all the statistical characteristics of the source waveform as a function of the parameters of the filter such as, for example, the power spectral density:

$$S_X(f) = \frac{\sigma_W^2}{|A(f)|^2}$$

where the notation $A(f)$ is inaccurate and should be more properly written as $A(\exp(j2\pi f))$. It also allows us to give analytical expressions for the quantization error power for different quantization methods when quadratic error is chosen as the measure of distortion. Comparison of the performance of the different methods is thereby possible. From a practical point of view, this example is not useless because it is a reasonable model for a number of signals, for example, for speech signals (which are only locally stationary) when the order P selected is high enough (e.g. 8 or 10).