Behavior of quantization noise for sinusoidal signals. A revision.

P. R. Pérez-Alcázar[1]*, A. Santos[2]

[1]Departamento de Ingeniería Electrónica
Facultad de Ingeniería, UNAM, Mexico
*pperezalcazar@fi-b.unam.mx

[2]Departamento de Ingeniería Electrónica
Universidad Politécnica de Madrid, Spain

ABSTRACT

This paper presents a brief revision of several studies presented in the literature about the behavior of quantization noise for sinusoidal signals and for uniform quantizers. From this revision, the conclusion is that quantization noise has been assumed to be additive and has a white spectrum, although some published studies, considering the problem either from a deterministic point of view or from a stochastic one, have shown a different noise behavior for some specific cases. Some of these cases are related with the parameters that characterize the sinusoidal signal and other with the conditions under which the process of conversion is realized. There are some cases that have not very been well considered in the previous literature and about which it is convenient to call attention. By this reason and using computer simulations with sinusoidal input signals, it is illustrated here that the quantization noise spectrum can show a discrete or complex structure depending on the relation between the sampling rate used and the frequency of the signal. Moreover, some points to consider in order to get a better description of the quantization noise are presented.

Palabras clave: Analog-digital conversion, quantization noise, noise spectrum.

RESUMEN

Este artículo presenta una breve revisión de varios estudios presentados en la literatura sobre el comportamiento del ruido de cuantización señales sinusoidales y cuantificadores uniformes. A partir de esta revisión, la conclusión es que generalmente se ha asumido que el ruido de cuantización es aditivo y tiene un espectro blanco; aunque algunos estudios publicados, considerando el problema desde un punto de vista determinístico o desde un punto de vista estocástico, han mostrado un comportamiento diferente del ruido para algunos casos específicos. Algunos de estos casos están relacionados con los parámetros que caracterizan la señal sinusoidal y otros con las condiciones bajo las cuales se realiza el proceso de conversión. Hay algunos casos que no han sido muy bien considerados en la literatura previa y sobre los cuales conviene llamar la atención. Por esta razón, utilizando simulaciones mediante computadora, con señales de entrada sinusoidales, se muestra aquí que el espectro del ruido de cuantización puede mostrar una estructura discreta o compleja dependiendo de la relación entre la velocidad de muestreo utilizada y la frecuencia de la señal. Además, algunos puntos a considerar para obtener una mejor descripción del ruido de cuantización son presentados.

Keywords: Analog-digital conversion, quantization noise, noise spectrum.

1. Introduction

When we were working in the acquisition process of nuclear magnetic resonance signals, making the conversion of the signal from analog to digital, taking it directly in the radiofrequency range (about 200.36 MHz), and using undersampling (5 Megasamples per second), we found that in the spectrum of such digital signal appeared copies of the original signal in odd multiples of the original frequency [1]. For this reason, we decided to study the information about the two processes involved in the analog to digital conversion.

We found that quantization error has attracted considerable attention as a result of its unavoidable presence in the conversion of analog to digital signals. However, the analysis of quantization error is usually simplified by considering it as additive white noise, with the output of the quantizer, y_n, assumed to be the addition of the input signal, x_n, plus white noise,

e_n, as shown in Figure 1. n is an integer. Quantization error is then considered as a sequence of uniformly distributed random variables, uncorrelated with the input signal and uncorrelated with each other. This result was first obtained using assumptions about high-resolution quantization that were initially made in classical papers such as that on pulse-code modulation (PCM) by Oliver, Pierce and Shannon [2], that on quantization error spectra by Bennett [3] and that on optimum quantizers by Panter and Dite [4]. The theory is supported by the following assumptions:
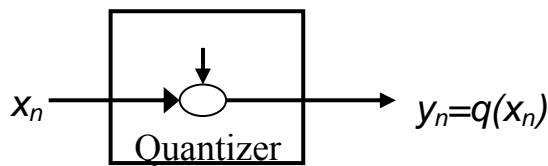

Figure 1. Model of quantizer using an input sequence of infinite precision x_n and additive white noise e_n.

1. The number of quantization levels, M, is high.
2. The difference between levels or quantization steps, Δ, is small.
3. The probability density distribution, f_x, of the input is smooth.

The simplified quantization model is known as the quantization theorem [5] and has been used as a general model, usually without detailed consideration of the work by Widrow [6,7]. Widrow proved that for a random input variable, with a band-limited characteristic function that is uniformly quantized with an infinite number of levels, the moments of the quantized signal can be calculated with the moments of the input signal plus a random additive variable that is independent of the signal and distributed uniformly in the interval (-Δ/2, Δ/2). The band-limited nature of the characteristic function implies that the probability density function must have an infinite support, because a signal cannot be time-limited and at the same time has a band-limited transform.

Previously, Clavier et al. [8] studied quantization error from a deterministic point of view. He showed that when the number of samples in a cycle of a signal is an integer, then the spectrum of the quantization error consists of odd harmonics of the frequency of the signal being quantized.

Based on simplified quantization models, partial descriptions of quantization noise behavior have been obtained, although few researchers have studied quantization noise behavior from either a deterministic or a stochastic point of view. In this paper, after presenting a brief revision of several classical papers that analyze quantization noise and indicating some contradictory and incomplete conclusions from these works, computer simulations of the sampling and quantization processes are presented, for which the generally accepted assumption of additive white noise is clearly not valid and other previous results need to be reconsidered.

2. Background

Following the studies of Bennett [3] and Widrow [6, 7], Sripad and Snyder [5] extended the class of input functions for which the model of uniform noise is valid. They showed that this model is adequate for any random variable, x, with characteristic function, ϕ_x, which satisfies Equation (1):

$$\phi_x\left(\frac{2\pi n}{\Delta}\right) = 0 \quad \text{for } n = \pm 1, \pm 2, \cdots, n \neq 0 \quad (1)$$

This condition is less restrictive than the requirements of the quantization theorem, which can be considered a special case of Equation (1). Thus, the uniform quantization error model is

obviously adequate for describing the quantization of a uniform random variable, which has a no band-limited characteristic function. This result can be extended to uniform probability densities in the interval $\left[-(2k+1)\dfrac{\Delta}{2}, (2k+1)\dfrac{\Delta}{2} \right)$, where k is an integer. The model is also valid for a random variable with a triangular probability density function.

Sripad and Snyder [5] also indicated that the quantization error is not statistically independent of the input function. Considering an input modeled as a Gaussian process, they showed that the difference between the model of uniform noise and the model including quantization noise decreases as the ratio, $\dfrac{\sigma}{\Delta}$, of the standard deviation of the quantization noise to the quantization cell width increases.

Apart from the asymptotic theories, exact solutions have been found for some cases of uniform quantization. For example, Clavier et al. [8] analyzed quantization noise for sinusoidal signals using the characteristic function or Rice transform and obtained a result consisting of complex additions of Bessel functions. Another example is the derivation by Bennett [3] of the power spectral density of a uniformly quantized Gaussian random process. For high resolution, he showed that a uniform quantizer with a small quantization step, Δ, produces an average distortion, $D(q) \cong \dfrac{\Delta^2}{12}$.

Claasen and Jongepier [9], following the work by Sripad and Snyder [5], analyzed the error spectrum for uniform quantization of deterministic signals. They initially considered that the white noise model was adequate if the signal was sufficiently complex, as remarked by Oppenheim and Schafer [10]. However, they also observed that until then,

no-one had established the acceptable degree of variation or complexity needed to validate the model. They also showed that there are well-known examples for which the model is not valid (such as for a step function or a dc input). It is convenient to note here that the consideration by Oppenheim and Schafer also appears in a later version of their book [11].
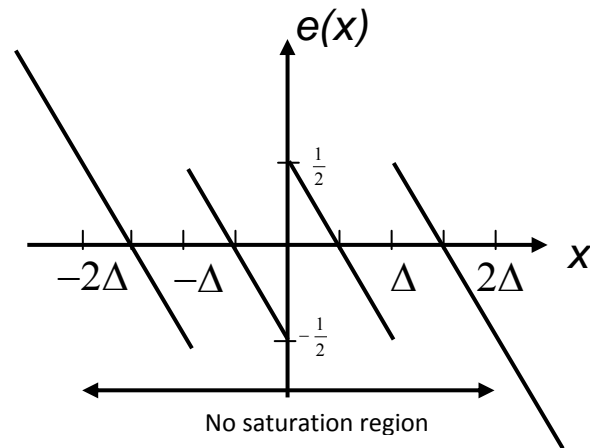

Figure 2. The transfer function between input x and quantization error e.

Claasen and Jongepier [9] obtained the power spectral density of quantization error by decomposing the error function (Fig. 2) as a Fourier series, as done by Clavier et al. [8]. Taking into account the relationship between the power spectral density of a phase-modulated signal and the amplitude distribution of the derivative of the modulating signal, and the fact that the spectral components add on a power basis, they obtained an approximation of the power spectral density of the quantization noise as:

$$S_e(\omega) = \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{p_x\left(\dfrac{\omega}{2\pi k}\right)}{k^3}, \qquad (2)$$

where $p_{\dot{x}}(a)$ is the amplitude distribution function of the signal derivative and ω is the angular frequency. a is the argument of the distribution function, which in Equation (2) corresponds to $\dfrac{w}{2\pi k}$. For a sinusoidal function of amplitude A, it is given by the equation:

$$p(a;A) = \begin{cases} \dfrac{1}{A\pi}\dfrac{1}{\sqrt{1-\left(\dfrac{a}{A}\right)^2}}, & |a| < A \\ \\ 0 & |a| > A \end{cases} \qquad (3)$$

For $x(t) = A\sin\omega_0 t$, $p_{\dot{x}}(a) = p(a;\omega_0 A)$ and the quantization noise spectrum is given by Equations (2) and (3). As can be seen from Equation (3), p(a; A) tends to infinity or has a singularity at a = A, which means that the model given by Equation (2) has peaks or singularities in the noise spectrum for frequencies:

$$\omega = (2kA\pi)\omega_0 \qquad (4)$$

This is how the position of the spectrum peaks and the amplitude of the quantified signal are related. From this result, the authors conclude that if a flat noise spectrum is desired, the band of interest has to be restricted to $|\omega| \leq A\omega_0\pi$. In the digital case, Claasen and Jongepier considered that, "due to the sampling process and the associated folding of the corresponding spectrum around the sampling frequency, the spectrum of the quantization noise will be white if the sampling rate is taking well below the first discontinuity in $S_e(\omega)$. If, however, the sampling rate is taken appreciably larger than this value, a non-white spectrum results"[9].

As will be illustrated, the previous condition does not guarantee that the quantization noise

spectrum is white, and (4) gives only some of the spectrum peaks.

In 1990, Gray [12] published a more in-depth analysis of uniform quantization of deterministic signals without saturation, which was followed in 1998 [13] by a complete revision of quantization. He also represented the error sequence as a Fourier series, from which he could compute the second moment and the autocorrelation. This Fourier series for a signal, x, is given by:

$$e = e(x) = \sum_{l \neq 0}\frac{1}{2jl\pi}e^{j\frac{2l\pi}{\Delta}x} = \sum_{l=1}^{\infty}\frac{1}{l\pi}\sin\left(2l\pi\frac{x}{\Delta}\right). \qquad (5)$$

This approximation considers the error signal as periodic, which is only true in the no-overload region. Outside this region, the Fourier series is not valid.

As Gray remarks, this approximation is valid only for *simple* inputs (e.g. sinusoids), but even so, the Fourier series cannot converge if e_n is not a periodic function of *n*. As an example, $x_n = A\cos(2\pi f_0 n)$, with f_0 not a rational number, gives functions, x_n and e_n, which are not periodic [14]. In this case, it is still possible to use the Harald Bohr generalization of the Fourier series for almost-periodic functions.

Then he proceeded to use a sinusoid, $x_n = A\sin(n\omega_0 + \theta)$, with an initial phase θ, in Equation (5). At this point he remarked that the discrete-time signal can be related to the continuous-time signal, by defining $\omega = \psi T$, where *T* is the sampling period and $\psi = 2\pi f$, where *f* is the continuous-time frequency of the original signal. A further assumption was that $A \leq \dfrac{M}{2}$, so that the quantizer would not saturate. Finally, after some manipulations and considerations, and with the assumption that " f_0 is not a rational number (and hence that in the

corresponding continuous time system, the sinusoidal frequency is incommensurate with the sampling frequency)" [12], Gray concluded that the spectrum of the error has components of amplitude:

$$S_m = \left(\frac{1}{\pi} \sum_{l=1}^{\infty} \frac{1}{l} J_{2m-1} \big((2\pi) l\gamma \big) \right)^2 \qquad (6)$$

located at frequencies $(2m-1)\omega_0 (\mathrm{mod}\, 2\pi)$. Where J_m is the Bessel function of order m and $\gamma = \dfrac{A}{\Delta}$. Thus, the power spectrum density may be written as

$$S_e(f) = \sum_{m=-\infty}^{\infty} S_m \delta\big(f - \langle (2m-1) f_0 \rangle \big). \qquad (7)$$

Gray [12] concludes that "it suffices here to point out that the spectrum of the quantizer noise is purely discrete and consists of all odd harmonics of the fundamental frequency of the input sinusoid. The energy at each harmonic depends in a very complicated way on the amplitude of the input sinusoid. In particular, the quantizer noise is decidedly not white since it has a discrete spectrum and since the spectral energies are not flat. Thus here the white noise approximation of Bennett and of the describing function approach is invalid, even if M is large."

Therefore, Gray established that the discrete spectrum with odd harmonics of the input sinusoid is valid for any ratio of the frequency of the input signal to the sampling frequency in the corresponding continuous time system.

Another deterministic approach to quantization in the continuous time, which confirms the results obtained by Claasen and Jongepier [9] and also those obtained by Gray [12], is presented by Bellan et al. [15]; that is, the quantization error spectrum

shows repeated peaks at frequencies given by (4) and other peaks in odd harmonics of the frequency of the input sinusoid. However, these results are obtained when a uniform mid-tread quantizer is considered and when the signal crosses equal number of positive and negative levels of quantization. In this case, Bellan et al. characterize the spectrum shape of the quantization error as follows: it increases up to every peak given by Equation (4) and then falls to the beginning of the following increasing section. In this work, the Fourier coefficients also depend of the ratio between the amplitude of the signal and the quantization step size. This behavior is not obtained from the study realized by Gray [12].

Besides, Bellan et al. considered the case of no-symmetric crossed levels or that of a sinusoidal wave with offset. In this case they conclude that the spectrum is going to have a dc component and all the harmonics of the frequency of the signal, even and odd ones.

For both cases, Bellan et al. said "if we are dealing with a time shifted sinusoidal waveform, we must multiply the Fourier coefficients by $e^{-j2\pi n \frac{t_o}{T}}$, where t_0 is the time shift." Here, n is the number of harmonic and T is the period of the signal being quantized. It is important to note here that Bellan et al. have not considered the process of sampling, which is important in the process of conversion from analog to digital. Then, their work is incomplete in this sense.

On another front, some theoretical studies have shown that in order to reduce the harmonic distortion due to the quantization noise, it could be convenient to use dithering quantization. This is a technique in which a signal called a dither is added to an input signal prior to quantization, trying to force the Bennet approximations to hold [16] and to have a flat spectrum of the quantization noise, and then it could be subtracted

or not before reconstruction. Therefore, it is possible to have subtractive dithered quantizer and nonsubtractive dithered quantizer. Schuchman [17] studied the conditions a dither signal must meet so that the quantizer noise can be considered independent of the signal being quantized for a quantizer having a finite number of levels. He found that an infinite class of dither signals satisfied these conditions, but he remarked that the most useful member of this class is one whose probability density function is uniformly distributed over a quantizing interval. As it is highlighted by Gray and Stockhman [16], the subtractive dither result is nice mathematically, but it is impractical in many applications. For this reason, the most commonly used technique is the nonsubtractive one.

However, as it is said by Wannamaker et al. [18], nonsubtractive dithered quantizers can not render the total error statistically independent of the input and can not temporally separated values of the total error statistically independent of one another. Besides, this technique involves the addition of a corrupting noise and decreases the dynamic range of the input if quantizer overload is to be avoided.

In a practical case, there are other factors which can modify the structure of the quantization noise obtained in the process of conversion of a signal from analog to digital. These factors are studied in papers like [19], [20] and [21].

Taking into account the brief revision of the studies about the quantization noise presented above, the aim of the current work is to confirm or to reconsider the conclusions reached by the authors of those previous works. Some of them consider that under certain conditions the quantization error could be white, like Claasen and Jongepier [9], and others consider that it is discrete with tones related harmonically, like R.M. Gray in his two papers [12, 13], and Bellan et al. [15].

3. Simulations

To study the behavior of quantization noise produced by the process of uniform quantization of a sinusoidal input signal, computer simulations with MATLAB were performed under various conditions. The computer simulations allowed the use of classical round-off, avoiding completely the effect of other factors appearing in the operation of an analog-to-digital converter, such as jitter, integral and differential non-linearity, thermal noise, etc. The process of rounding was considered taking into account the definition of quantization given by Proakis and Manolakis [22]: a process of rounding or truncation. Before rounding, it is necessary to increase the amplitude of the sinusoidal signal from the range -1 to 1 up to an adequate range. After the process of rounding, the signal is adjusted to have an amplitude equal to that of the original one and, then, the error of quantization is obtained as the difference between the signal generated at the resolution given by MATLAB and the quantized signal, as defined by Oppenheim [11]. Apart from these implementation issues, the quantization noise for sinusoidal inputs depends on the ratio between signal and sampling frequencies (analyzed in Sections III.A and III.B), the ratio between the amplitude of the input signal and the quantization step size (analyzed in Section III.A), the offset (analyzed in Section III.A) and the phase of the input signal (analyzed in Section III.C).

In all the simulations presented, the sampling of the analog signal was simulated using the signal and sampling frequencies specified in the following sections. The samples are then quantized and the quantization error is computed. Finally, the spectrum of the original signal, the quantized signal and the error signal are computed. In all the simulations the amplitude of the input signal was 0.24, so that the quantization step size was 0.48/M, where M is the number of quantization levels. The ratio of the amplitude and the quantization step size was M/2. As it was

presented in the studies briefly described above, the important parameter in the behavior of the quantization noise is the ratio between the amplitude of the input signal and the quantization step size and it is not the amplitude by itself. For this reason, an arbitrary value for the amplitude has been chosen. The values of the several parameters used in the simulations were chosen to exemplify or illustrate some cases where the theoretical conclusions need to be revised. The examples presented in the next paragraphs are typical cases of what one can find during the combination of the processes of sampling and quantization.

A. Integer and Non-Integer Relation Between Sampling and Signal Frequencies

The aim of this simulation is to show the behavior of quantization noise when the ratio of the sampling frequency to the signal frequency is an integer or a non-integer number, and will include illustrating the effect observed by Claasen and Jongepier [9]. The simulation has been done using the following steps:

1. A sinusoidal signal with a frequency of 21.5 Hz is generated and oversampled at 40 k-samples/s. Here is used an oversampling factor of 1860.46… (a rational number, having a repetition of digits after the 18th decimal digit. A total of 4096 samples are produced.
2. The signal is quantified to 94 levels (in an attempt to present a similar effect to that predicted by Claasen and Jongepier [9]).
3. This analysis is repeated for a signal with a frequency of 800 Hz but using first a sampling rate of 40 ksamples/s and then 39995 samples/s, corresponding to oversampling factors, respectively, of 50.00 and 49.99… (a rational number, as is any number in a computer).
4. The previous step is repeated but adding a dc component to the 800 Hz sinusoidal signal.

Figures 3(a) to 3(d) show the results obtained when oversampling a sinusoidal signal (21.5 Hz) by a rational factor of 1860.46…. The power spectrum of the signal using the maximum resolution given by MATLAB can be seen in Figure 3(a), while Figure 3(b) shows the spectrum when the signal is quantized with 94 levels. As expected, with the high resolution used to generate Figure 3(a), quantization noise cannot be discerned, but with 94 levels this noise can be seen as shown in Figure 3(b).

It can also be seen that the quantization process generates several components distributed from 0 to the Nyquist frequency. An analysis of the spectrum of the quantization error, presented in Figure 3(c), shows the discontinuities predicted by Claasen and Jongepier [9] in the regions between 6 and 7 kHz and between 12 and 13 kHz. This is in agreement with the results of Equation (4), which predicts a first peak at 6,349 Hz and a second at 12,698 Hz. Although the spectrum seems almost flat or white, if we study it in detail, enlarging its first section, as it is presented in Figure 3(d), we can appreciate that the quantization error is concentrated in odd harmonics of the frequency of the sinusoid being quantized (there are peaks at frequencies $(2n + 1) \times 21.5$ Hz, with n = 0, 1, 2,...). This effect was predicted by Gray [12] and obtained by applying Equation (7). As can be seen, in this case, although the condition of Claasen and Jongepier is satisfied, a white noise spectrum is not obtained. The problem here is that Claasen and Jongepier established their condition considering only the presence of the peaks given by Equation (4), but they did not take into account the harmonics of the frequency of the signal being quantized as they are considered by Gray [12] and Bellan [15]. However, as it will be illustrated later in the next section, an approximation to a flat spectrum can be obtained depending of the relation between the sampling frequency and the frequency of the signal.

Spectrum of the 21.5 Hz sinusoidal signal without quantization

(a)

Spectrum of the quantized 21.5 Hz sinusoidal signal

(b)

Spectrum of the quantization error
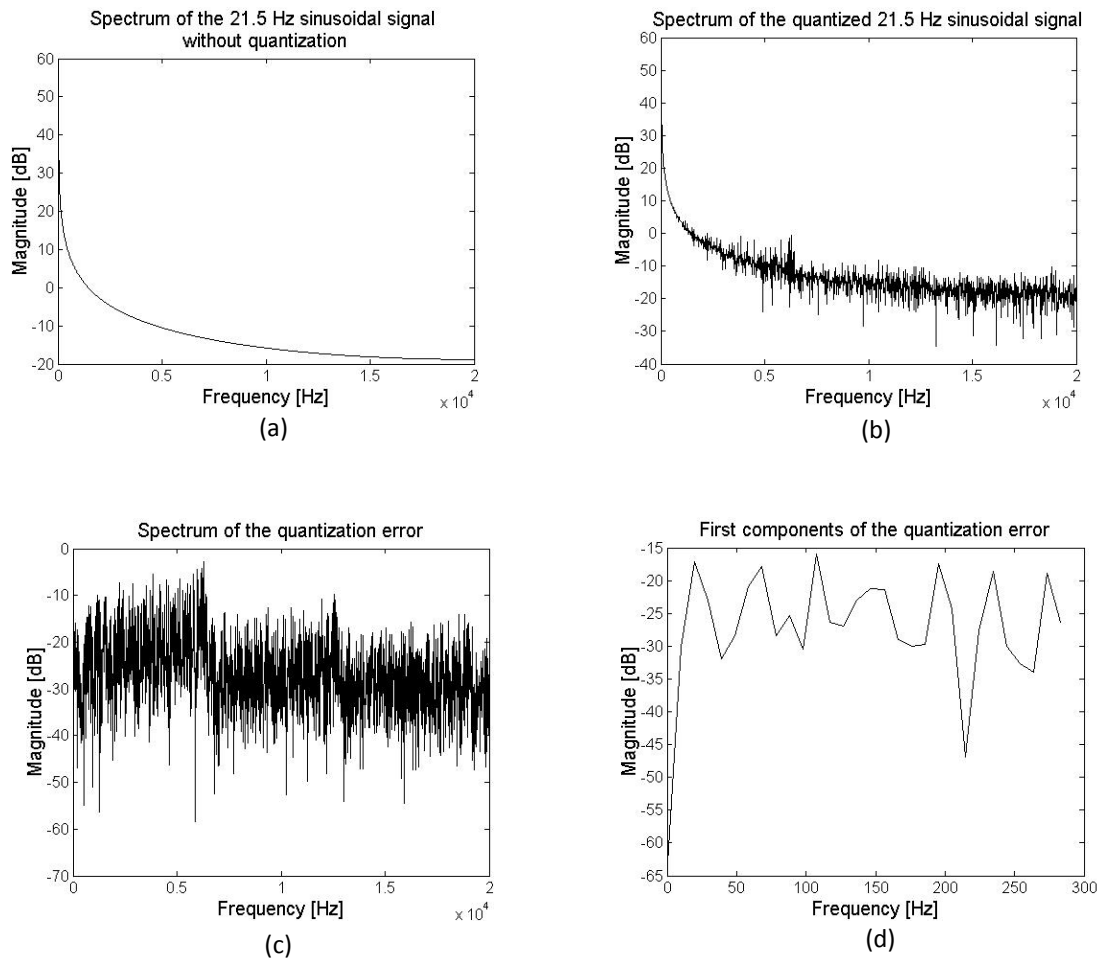
(c)

First components of the quantization error

(d)

Figure 3. (a) Spectrum of an unquantized 21.5 Hz sinusoidal signal, sampled at 40 k-samples/s; (b) spectrum of the 21.5 Hz sinusoidal signal, quantized with 94 levels and sampled at 40 k-samples/s; (c) spectrum of quantization error in (b); (d) detail of (c), showing the first components of the spectrum.

It is important to remark here that Gray, in his theoretical study, demonstrated that the amplitude of the signal relative to the step size only affects the magnitude of the spectrum peaks. However, Claasen and Jongepier have demonstrated, and we have illustrated here, that this relative magnitude causes the presence of peaks that are not considered by Gray.

Power spectrum analysis did not yield useful information with this sample, as the FFT resolution, approximately 10 Hz, is too close to the harmonic separation. In this case the harmonics are separated by approximately 43 Hz, that is, 4 points in the graphic of the spectrum, but each harmonic is represented by 4 points. Then the harmonics are touching each other or separated at most by 1 point. The analysis was therefore

repeated for the 800 Hz signal, but at 40 k-samples/s and 39995 samples/s, and their spectra presented in Figures 4(a) and 4(b), respectively. These simulations show that Gray's prediction in his work of 1990 [12] is not valid for all ratios between the sampling rate and signal frequency (let us denoted it by RoF). There exist some rational ratios for which the power of the harmonics that constitute the quantization noise is spread out in the Nyquist band. When the ratio is 50, the odd harmonic components appear, as it is illustrated in Figure 4(a), but when this ratio changes to 49.9937... (a change of around 0.013%), the quantization error contains other components, as shown in Figure 4(b), and part of the power of the harmonics is transferred to the new components; that is, the power of the harmonics starts spreading over the frequency band. This effect is not due to the folding.

An additional interesting case is obtained when the sinusoidal signal has a d.c. component or an offset. In this case, as can be deduced from the work by Bellan et al. [15], the quantization error signal is going to have even and odd harmonics of the frequency of the sinusoidal input. The important point here is that the behavior of the quantized error also depends of RoF, as it is illustrated in Figures 5(a) and 5(b). This point was not established by Bellan et al. because they only considered the process of quantization without taking into account the process of sampling. They considered that the sampling process only introduces the folding of the spectrum of the quantization noise. With the examples presented here, we are illustrating that is necessary to consider simultaneously both processes, sampling and quantization. Figure 5(a) presents the spectrum of the quantization error obtained after
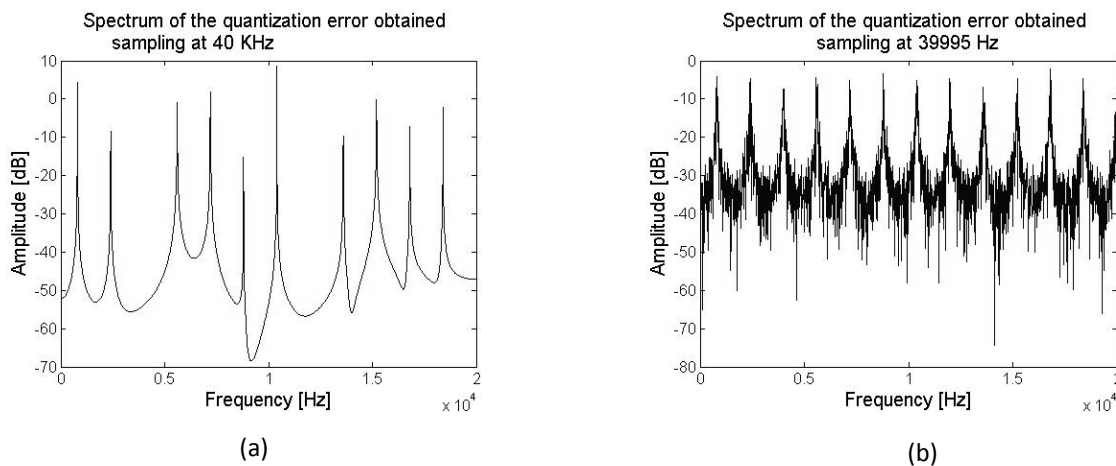

(a)

(b)

Figure 4. Spectra of the quantization error obtained from quantizing a 800 Hz sinusoidal signal with 94 levels; (a) using a sampling rate of 40 k-samples/s; (b) using a sampling rate of 39995 samples/s.

sampling the signal of 800 Hz using a sampling rate of 40 KHz; that is, the process was done considering an integer RoF. In this case, the spectrum presents both the odd and even harmonics of the frequency of the sinusoidal input signal. On the other hand, Figure 5(b) presents the spectrum of the quantization error when there is a noninteger RoF and the other characteristics of the sinusoidal signal similar to those used in the case of Figure 5(a) were kept. Now, it is possible to observe the presence of components that are not harmonically related with the frequency of the signal and, also, the transference of power from the harmonics to the other components that appear in the spectrum; that is, the spreading of the power of the harmonic components in all the band of interest.

B. Relation between quantization error and a sinusoidal input signal

The aim of these simulations is to determine whether Gray's analysis is still valid when the sampling frequency is not considerably larger than the input signal frequency, and to illustrate that the discrete quantization error is not periodic even though the input signal is periodic, except in the case in which RoF is an integer.

The simulation requires the following steps:
 1. A sinusoid of 800 Hz is oversampled at 8 k-samples/s, generating 4096 samples.
 2. The signal is quantized to 94 levels. This analysis is repeated for the same input signal but using a period of sampling of 0.000125625. In these cases we have oversampling factors of 10.00 and 9.95..., respectively. The last factor has no repetition of digits in the 31 digits used by the PC.

   Figure 6(a) shows the behavior of quantization noise in the time when there is an integer ratio (in this case, 10) of the sampling rate to the signal frequency. As can be seen, the error is periodic with frequency components in the odd harmonics of the input signal as shown in Figure 6(c); that is, in 800 Hz and 2400 Hz. Note that in Figure 6(c), because of the sampling rate, we only can appreciate the first and third harmonics of the sinusoidal signal being quantized; the other components fold back into the positions of the visible components. However, when the ratio of the two frequencies is not an integer, there is no such periodic behavior in the error, as shown in Figure 6(b). In this case, as presented in Figure 6(d), the error spectrum becomes distributed over the whole frequency range and, therefore, it is possible now to say that the spectrum tends to be flat or white, complementing the condition by Claasen and Jongepier.
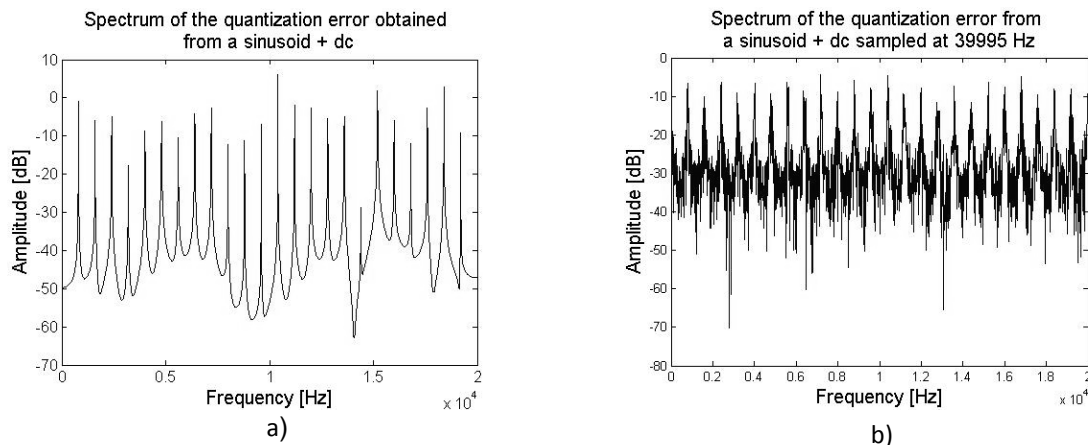

Figure 5. Spectra of the quantization error obtained from quantizing a 800 Hz sinusoidal signal with a dc component using 94 levels; (a) sampled at 40 k-samples/s; (b) sampled at 39995 samples/s.

This presentation can be completed if we consider Figures 7(a) and 7(b). Figure 7(a) presents the relationship between the discrete quantization error and discrete input signals when the *RoF* takes on an integer value. It is possible to observe that the quantization error takes on the same values in each of the four cycles of the input signal shown in the figure, and then it is possible to say that if the input signal is periodic the error will be periodic. This would be the case obtained following the well-known analog relationship shown in Figure 2. On the other hand, Figure 7(b) presents the relationship between the discrete quantization

error and the discrete input signal when the RoF is not an integer value and only for four cycles of the input signal. It is possible to appreciate that the quantization error does not take on the same values in each cycle of the input signal and then it is not periodic. The results of this section do not follow Gray's prediction, as he considered that the error could be approximated by an almost periodic signal and for this reason he always presented only odd harmonic components of the signal being quantized. The periodic or quasi-periodic behavior of the quantization noise cannot be assumed for all ratios of sampling rate to signal frequency.
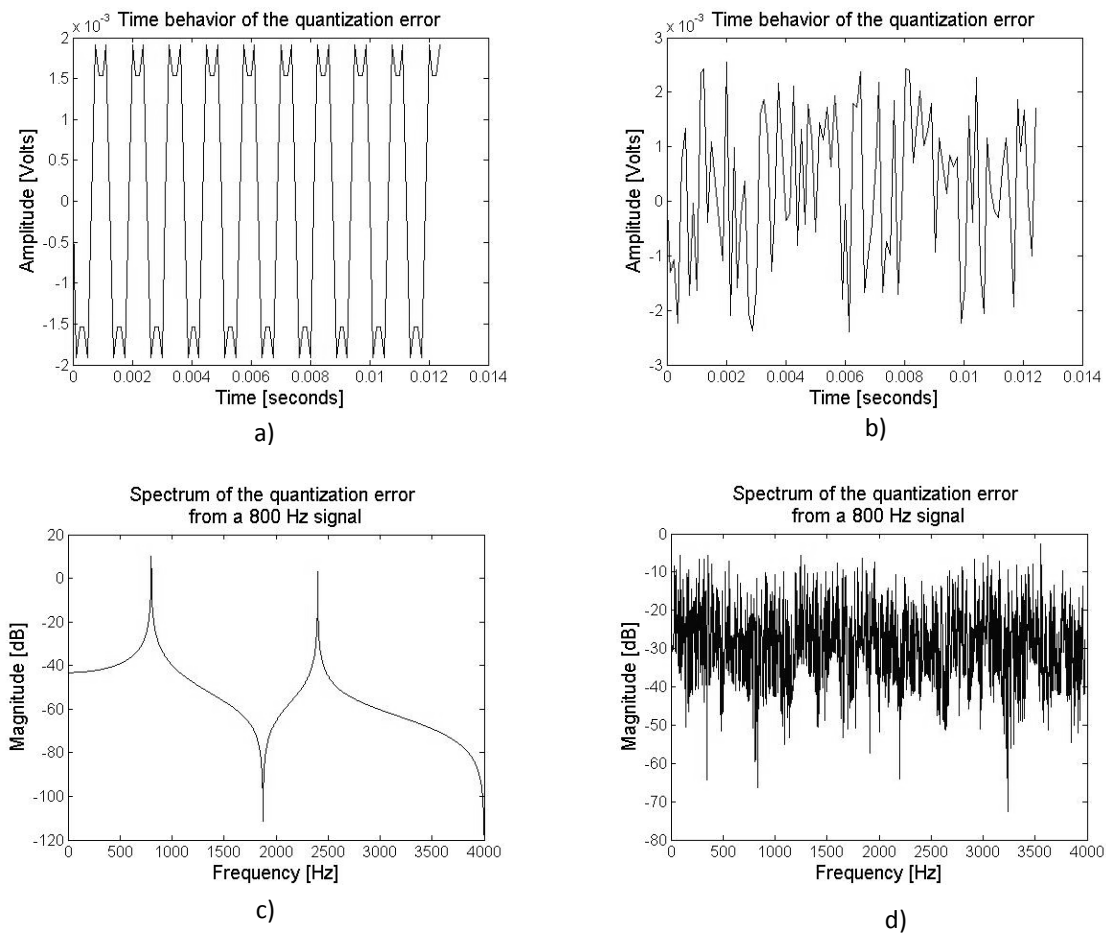

Figure 6. Behavior of the quantization error obtained from quantizing with 94 levels a sinusoid of 800 Hz in time domain and frequency domain. (a) In the time domain using an integer *RoF*=10. (b) In the time domain using a noninteger *RoF*=9.95…. (c) In the frequency domain using an integer *RoF*=10. (d) In the frequency domain using a noninteger *RoF*=9.95.
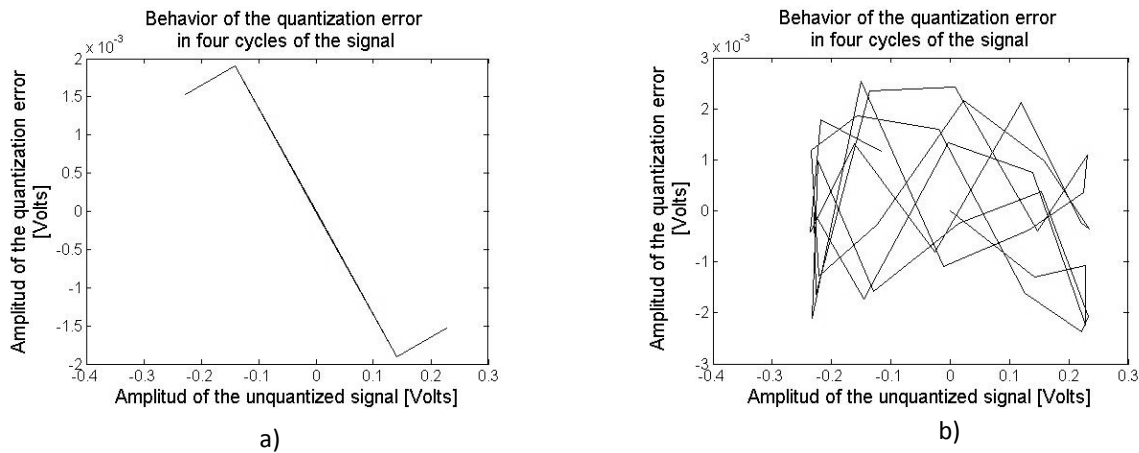
Figure 7. Relationship between the discrete quantization error and the discrete 800 Hz sinusoidal signal at the input of the quantizer (*RoF*=10).(a) Integer ratio between the sampling rate and the frequency of the input signal; (b) noninteger ratio (*RoF*=9.95…).

In this example, a small change in the ratio of frequencies, that is, in the conditions of operation, causes a large change in the structure of the quantization noise spectrum. Specifically, small changes in the frequency ratio also induce changes in the error spectrum from periodic to non-periodic and then again to periodic. This behavior occurs undoubtedly because in the digital domain a small change in the frequency of the signal can cause a great change in the period.

C. Relation Between the Signal Phase and the Spectrum

The analysis of the section A is repeated for four input sinusoids of 800 Hz, which have a $\dfrac{\pi}{4}$ radian phase difference, starting from zero. They were sampled at 40 K-samples/s.

Figure 8 shows the spectrum of the quantization error of four quantized signals, having a different for a sinusoidal signal with phase equal to zero does not present the third and eighth harmonics, however the spectrum of the error for the signal phase. As can be seen, the signal phase has an influence on the quantization error. Figure 8(a) shows that the spectrum of the quantization error with $\dfrac{\pi}{8}$ phase has those harmonics. This is the most drastic change in the spectrum; in the other cases, the change is basically in the magnitude of the odd harmonics, as can be seen in Figure 8(b). This influence has not been considered either by Gray [12] or by others. A related point to consider here is that the value of the phase determines whether or not there are discontinuities in the initial and final acquired samples. These possible discontinuities affect the spectrum computed by a discrete Fourier transform, as is well known since the publication of the work of Harris [23], but not in the way shown here. The discontinuities at the ends of the record introduce an imaginary sinusoidal component of period $2\pi/(NT_s/2)$, where N is the number of samples in the record and T_s is the sampling period.
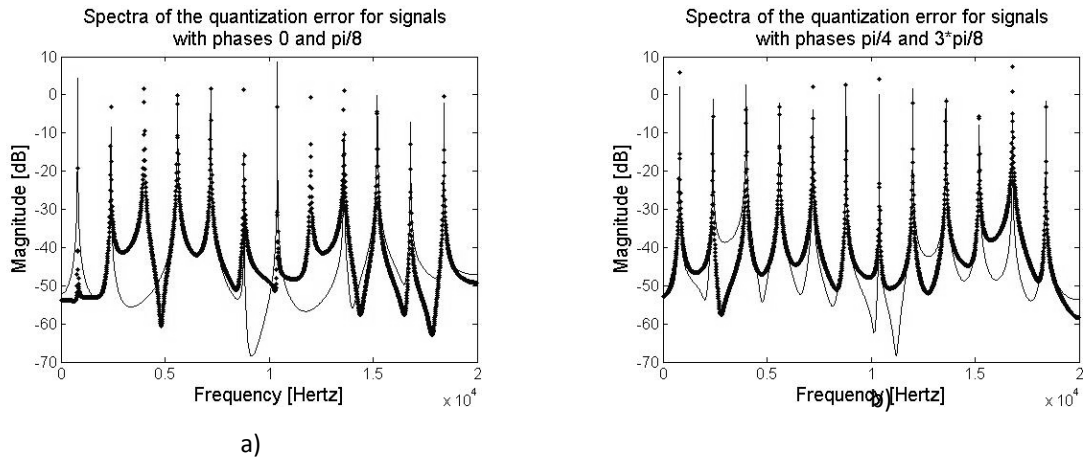
a)

Figure 8 . Modification of the quantization error spectrum with phase.(a) Spectra of the quantization error obtained from quantizing with 94 levels a 800 Hz sinusoidal signal with phase 0 (solid line) and $\frac{\pi}{8}$.
(b) Spectra of the quantization error obtained from quantizing with 94 levels a 800 Hz sinusoidal signal with phase $\frac{\pi}{8}$ (solid line) and $\frac{3\pi}{8}$.

4. Discussion and conclusions

The behavior of the quantization error, under different conditions, is obtained modeling the process of quantization as a rounding step and then calculating the quantization error as the difference between a signal generated at the resolution given by MATLAB and the same signal but quantized. The models described in the literature were not used because they do not describe completely the behavior of the quantization noise.

The work described here shows, through simulations, that some statements or conclusions published about the structure and behavior of the spectrum of the quantization error generated when a sinusoidal signal is quantized are not valid. First, it is not true that a discrete spectrum with only odd harmonics is obtained, as established by Gray [12], following both the deterministic and

stochastic points of view. Gray showed that this type of spectrum is obtained independently of the relationship between the sampling rate and signal frequency. The simulations presented here (Sections III.A and III.B) show that Gray's conclusion is only true when the sinusoidal signal and the sampling frequencies have an integer—and in a large number of cases, a rational—relationship in the continuous time domain. However, when this relationship changes by even a small amount (0.013 % in the case shown), the harmonics disappear and their power is spread over the entire band established by the sampling frequency. Then, with a further small change in the relationship between frequencies, the spectrum evolves from a periodic structure to a non-periodic one. An additional point to be highlighted is that the spectrum of the quantization error has even harmonics if the sinusoidal signal being quantized presents a component of dc.

The simulations also show that the model usually applied in initial analyses of quantization error (periodic or almost periodic quantization error) is not always adequate. This is only true, as has been illustrated here, when the ratio of the sampling rate to signal frequency in the continuous time domain is either an integer number or a particular rational number.

Another conclusion obtained is that the condition established by Claasen and Jongepier [9], to ensure that the spectrum is white, is not sufficient. There is a white spectrum, that is, a flat spectrum with components not related harmonically when the ratio between the sampling rate and the frequency of the signal is not an integer value or some rational values (these values need to be determined theoretically). There are many cases in which the condition established by Claasen and Jongepier is satisfied and, however, the error is not white, that is, it has not a flat spectrum with uncorrelated components. In these cases, the spectrum presents multiple peaks or poles that are located at odd harmonics of the frequency of the signal being quantized. The problem in the paper by Claasen and Jongepier is that they did not consider the presence of those odd harmonics. Claasen and Jongepier were right, however, in establishing their condition based on the fact that there are peaks in the signal spectrum that depend on the frequency and amplitude of the signal, as has been illustrated here.

Gray also proposed that the only effect of the signal amplitude relative to the quantization step size is in the amplitudes of the odd harmonics, which is not correct. This ratio, combined with the signal frequency, modifies the position of the quantization error spectrum peaks.

Finally, it is interesting to note that there is a dependence of quantization error on the signal phase. The harmonic amplitudes depend, to a certain extent, on the signal phase, although the total error energy is obviously constant. Both Gray [12] and Bellan et al. [15] introduced the phase factor as a complex exponential, which did not affect the spectrum magnitude but, as has been shown here, the phase does affect the magnitude.

It is necessary to note here that albeit Gray considered the process of sampling in the study of the behavior of the quantization noise, he did not considered the process of folding due to the sampling [12] and that Bellan et all [15] considered neither the process of sampling nor folding. However, we have illustrated that it is necessary to consider simultaneously the processes of sampling and quantization in order to approximate better the practical situations.

The previous discrepancies indicate that it is necessary to study the behavior of the quantization error in more detail, perhaps using the tools of nonlinear dynamic systems theory, as the studies based on deterministic and stochastic points of view seem incomplete or inaccurate.

The very simple procedure of simulation presented here permits to evaluate the theoretical results obtained about the behavior of the quantization error with different parameters used in the process of sampling and quantization.

The presence of harmonic components in the quantization error must be considered when designing digital systems that process analog signals. Although quantization error is something that is usually undesirable, as it distorts the digital signal, the harmonic component effect could be used to evaluate ADCs, as the quantization noise could be well localized, and any additional noise could be attributed to other sources. For example, the presence of a dc component in the signal can be immediately detected if the spectrum presents the even harmonics of the sinusoidal signal being quantized. As it was highlighted before, if the

sampling frequency is selected adequately, the quantization noise might be white and it is not necessary to use a dither signal and reduce the dynamic range.

References

[1] Pérez P., Adquisición de imágenes de resonancia magnética nuclear mediante técnicas de submuestreo, PhD Thesis, Universidad Politécnica de Madrid, Madrid, 1999, pp. 122-125.

[2] Oliver B., Pierce J., & Shannon C., The philosophy of PCM, Proceedings of the IRE, Vol. 36, No. 11, November, 1948, pp. 1324–1331.

[3] Bennett W., Spectra of quantized signals, Bell System Technical Journal, Vol. 27, No. 4, July, 1948, pp. 446–472.

[4] Panter P. & Dite W., Quantization distortion in pulse-count modulation with nonuniform spacing of levels, Proceedings of the IRE, Vol. 39, No. 1, January, 1951, pp. 44–48.

[5] Sripad A. & Snyder D., A necessary and sufficient condition for quantization errors to be uniform and white, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 25, No. 5, October, 1977, pp. 442–448.

[6] Widrow B., A study of rough amplitude quantization by means of Nyquist sampling theory, IRE Transactions of the Professional Group on Circuit Theory, Vol. 3, No. 4, December, 1956, pp. 266–276.

[7] B. Widrow, Statistical analysis of amplitude quantized sampled data systems, Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry, Vol. 79, No. 52, January, 1961, pp. 555–568.

[8] Clavier A., Panter P & Grieg D., Distortion in a pulse Count modulation system, Transactions of the American Institute of Electrical Engineers, Vol. 66, November, 1947, pp. 989–1005.

[9] Claasen T. & Jongepier A., Model for the power spectral density of quantization noise, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 29, No. 4, August, 1981, pp. 914–917.

[10] Oppenheim A. & Schafer R., Digital Signal Processing, Prentice Hall, 1975, pp. 413-418.

[11] Oppenheim A. & Schafer R., Discrete-time Signal Processing, Prentice Hall, 1989, pp. 119-123.)

[12] Gray R., Quantization noise spectra, IEEE Transactions on Information Theory, Vol. 36, No. 6, November, 1990, pp. 1220–1244.

[13] Gray R. & Neuhoff D., Quantization, IEEE Transactions on Information Theory, Vol. 44, No. 6, October, 1998, pp 1–63.

[14] Oppenheim A. & Willsky A., Signals and systems, 2d Ed., Prentice Hall, 1996, pp. 25-28.

[15] Bellan D., Brandolini A., & Gandelli A., Quantization theory—A deterministic approach, IEEE Transactions on Instrumentation and Measurement, vol. 48, No. 1, February, 1999, pp. 18–25.

[16] Gray R. & Stockham T., Dithered quantizers, IEEE Transactions on Information Theory, Vol. 39, No. 3, May, 1993, pp 805-812.

[17] Schuchman L., Dither signals and their effect on quantization noise, IEEE Transactions on Communication Technology, Vol. 12, No. 4, December, 1964, pp. 162-165.

[18] Wannamaker R., Lipshitz S., Vanderkooy J., & Wright J., A theory of nonsubtractive dither, IEEE Transactions on Signal Processing, Vol. 48, No. 2, February, 2000, pp. 499-516.

[19] Louzon P., Decipher high-sample-rate ADC specs, Electronic Design, Vol. 43, No. 5, March, 1995, pp. 91-100.

[20] Schweber B., Converters restructure communication architectures, EDN, Vol. 8, No. 8, August, 1995, pp. 51-64.

[21] Wepman J., Analog-to-digital converters and their applications in radio receivers, IEEE Communications Magazine, Vol. 33, No. 5, May, 1995, pp. 39-45.

22] Proakis J. & Manolakis D., Digital signal processing: principles, algorithms, and applications, 3d Ed., Prentice Hall, 1996, pp. 33-35.

Authors Biography

Pablo R Pérez-Alcázar

He received his B. Eng. degree in mechanical and electrical engineering, B. Sc. degree in physics, and the M.Sc. in electronics engineering from the Universidad Nacional Autónoma de México, in 1979, 1995, and 1982, respectively. Later, in 1999, he received his Ph. D. degree in telecommunications engineering from the Universidad Politécnica de Madrid. He joined the Faculty of Engineering of the Universidad Nacional Autónoma de México in 2004 where he is an associate professor. His current research activities include signal and image processing, digital electronics design as well as microelectromechanical systems design.

Andrés SANTOS

He is a professor in the Department of Electronic Engineering of Universidad Politécnica de Madrid, where he is the director of the Biomedical Imaging Technology laboratory. He has been coordinator of several national and international projects and author or co-author of more than 150 publications in international journals or peer reviewed conference proceedings. His main research areas of interest are medical image analysis and processing, specially, functional and molecular imaging.