On the Nature of Granulation Noise in Uniform Quantization Systems*

ROBERT C. MAHER, AES Member

Department of Electrical Engineering and Center for Communication and Information Science, University of Nebraska-Lincoln, Lincoln, NE 68588-0511, USA

The results of several investigations into the behavior of uniformly quantized signals are collected and discussed. In particular, the characteristics of digitized low-level signals are considered that lead to the effect known in the audio vernacular as granulation noise, an undesirable correlation between the input signal and the quantization noise components. Although the paper is primarily expository in nature, practical effects and some new simulation results are also presented.

0 INTRODUCTION

The input analog signal in a conventional pulse code modulation (PCM) digital audio system must be quantized in amplitude. The resulting digitized signal is a distorted replica of the original analog input signal [1]–[3]. The amplitude quantization process, a non-linear time-invariant operation, places an up-front limit on the distortion performance of the entire digital audio system. With an appropriate sample rate, an adequate number of bits in the digital signal representation, and a suitably complex input signal, little if any perceptual degradation is present in the reproduced signal. However, since the precise meaning of the terms appropriate, adequate, and suitably complex is exquisitely vague, there remains a need to consider the deleterious effects of amplitude quantization.

Several noteworthy papers have appeared recently which consider the general behavior of amplitude quantizers [4]–[6]. In this paper low-level audio signals are considered specifically in order to address the question of why the quantization distortion has a gritty, granular sound quality. Although the development of analog-to-digital converters with noise shaping [5], [6] and dither [7], [8] is de rigueur in many audio systems, the small-signal properties of the standard linear quantizer are still interesting and worthy of consideration due to their widespread use.

A review of several existing descriptions of the quantization process in digital audio systems is presented first. The intent is to emphasize the properties of amplitude quantization of simple signals which lead to a better interpretation than the widely espoused “6-dB per bit” formulation. The deterministic nature of quantization distortion is emphasized. Finally, some new simulation results are presented, indicating the origin and behavior of so-called granulation noise.

1 AMPLITUDE QUANTIZATION

Conventional digital audio systems involve both time sampling and amplitude quantization. For properly bandlimited signals (and neglecting clock jitter) time sampling is a theoretically lossless process—the original signal can be reconstructed exactly from its samples. Amplitude quantization, however, is inherently lossy—some range of actual input signal values is associated with a single output value (a many-to-one mapping). A common representation of a bipolar uniform quantizer is shown in Fig. 1. This quantizer is chosen to be normalized, that is, the spacing of the output steps equals the spacing of the input units. Note that the zero point of the quantizer depicted in the figure occurs halfway between quantization levels, corresponding to a “round to nearest integer” mathematical operation. A quantizer could also be defined with an output level transition at the zero point, corresponding to either a round-up or a round-down operation.

Without loss of generality, the normalized rounding quantizer of Fig. 1 is used in the remainder of this
paper. It is also assumed in this paper that the input signal amplitude does not exceed the range of the quantizer. This assumption is appropriate for the low-level signals of interest here. The mathematical description of the quantization process in continuous time is presented first, followed by the practical discrete-time case.

1.1 Quantization Effects in Continuous Time

The difference between the actual continuous input signal value and the quantized output value is the quantization error $e_Q(x)$. The well-known sawtooth representation of the quantization error as a function of the input signal is shown in Fig. 2. Note that the error is a deterministic function of the input signal, that is, the error is a known quantity if the signal itself is known. A convenient mathematical form for the quantization process $Q(\cdot)$ is

$$\hat{x}(t) = Q(x(t)) = x(t) + e_Q(x(t))$$

where $\hat{x}(t)$ is the quantized value and $e_Q(x(t))$ = sawtooth$(x(t))$ [4], [5], [9]. By expanding the sawtooth function of Fig. 2 as an infinite Fourier series, namely,

$$\text{sawtooth}(x) = \sum_{n=1}^{\infty} (-1)^n \frac{\sin(2n\pi x)}{n\pi}$$

the quantized signal becomes

$$\hat{x}(t) = x(t) + \sum_{n=1}^{\infty} (-1)^n \frac{\sin(2n\pi \cdot x(t))}{n\pi}.$$  (3)

The sign alternation factor $(-1)^n$ arises from the definition of the sawtooth as continuous across the origin. The basic formulation of Eq. (3) is used in the following sections to express the spectral content of the quantized input signal.

1.1.1 Single Sinusoidal Input

Consider an input signal consisting of a single sinusoid with amplitude $V$ and angular frequency $\omega$. Using Eq. (3) the quantized signal can be written as

$$\hat{x}(t) = V \sin(\omega t) + \sum_{n=1}^{\infty} (-1)^n \frac{\sin[2n\pi V \sin(\omega t)]}{n\pi}.$$  (4)

Thus $\hat{x}(t)$ contains the true input signal $V \sin(\omega t)$ and other spectral components due to the summation term in the equation. Eq. (4) can be rewritten in terms of Bessel functions as [9]

$$\hat{x}(t) = V \sin(\omega t) + 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n\pi} \sum_{k=0}^{\infty} J_{2k+1}(2n\pi V) \times \sin[(2k + 1)\omega t]$$

or

$$\hat{x}(t) = V \sin(\omega t) + 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n\pi} \sum_{\substack{p=1 \to \infty \text{ (odd)}}} J_p(2n\pi V) \sin(p\omega t)$$

where $J_p(\cdot)$ is the Bessel function of the first kind of order $p$. The error term contains only odd harmonics, which is expected because the quantizer of Fig. 1 is an odd function.

The amplitude of the harmonics produced by the quantizer can be expressed individually using Eq. (6):

$$\omega: \quad A_1 = V + 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n\pi} J_1(2n\pi V)$$

Fig. 1. Bipolar uniform quantizer $Q(\cdot)$, which is normalized so that input and output units are the same. Output steps correspond to “round to nearest integer” operation.

Fig. 2. Representation of quantization error $e_Q(x)$ for uniform quantizer of Fig. 1. Error is difference between input signal and quantized output signal and behaves as deterministic sawtooth function of input signal.
2ω: \( A_2 = 0 \)

3ω: \( A_3 = 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n\pi} J_n(2n\pi V) \)

4ω: \( A_4 = 0 \)

and so on. Note that the amplitude of the \( p \)th harmonic is determined using a summation of values spaced by multiples of \( 2\pi V \) on the Bessel function of order \( p \).

With large arguments, such as \( 2n\pi V > p \), Bessel functions of similar order have similar amplitude ranges and decrease in extent approximately as \( \sim 1/\sqrt{x} \) (see Fig. 3). Thus detectable harmonic energy can be present over a wide frequency range due to the quantizer. Furthermore, the quantization "noise" is actually odd harmonic distortion for any amplitude of an input sinusoid. The common assumption that the quantization noise is additive, white, and uncorrelated with the input signal is simply incorrect for the case of sinusoidal input.

The formulas involving Bessel functions are compact and mathematically elegant. However, they are often computationally intractable due to the slow convergence of the summation. Instead, the infinite summation over the Bessel function terms can be represented in an equivalent finite sum formulation to ease the computational burden for simulations [5]. This procedure is outlined in the Appendix.

Example spectra calculated for four different arbitrary sinusoidal input amplitudes are shown in Fig. 4. Normalized decibel ratios \( 20 \log(\frac{A_p}{V}) \) are shown in Fig. 5. Observe that the spectral levels are intimately dependent on the amplitude of the input sinusoid. For example, the partial levels for input amplitudes of 100 and 101 are nearly the same while the levels for input amplitude 100.5 are around 10 dB greater for low partial numbers.

It must be emphasized that the quantization distortion components are present in the quantized output signal for any amplitude of the input sinusoid. However, the distortion is usually of perceptual importance only for low input amplitudes where the ratio of signal (fundamental) to total distortion is small. This characteristic is considered in more detail in the following sections.

1.1.2 Multisinusoidal Input

Amplitude quantization is a nonlinear operation. This fact prevents us from using superposition to "add up" the results obtained for several single sinusoids to represent more complex signals. However, by replacing \( x(t) \) in Eq. (3) by an arbitrary waveform in sum-of-sinusoids form,

\[ x(t) = \sum_{m=1}^{M} V_m \sin(\omega_m t) \]

the resulting quantized signal becomes

\[ \hat{x}(t) = \sum_{m=1}^{M} V_m \sin(\omega_m t) + \sum_{n=1}^{x} \frac{(-1)^n}{n\pi} \sin \left[ 2n\pi \sum_{m=1}^{M} V_m \sin(\omega_m t) \right] . \]

This signal contains spectral components at the input signal frequencies (the \( \omega_m \)) and at harmonics of the input frequencies, and intermodulation products at sums and differences of multiples of the input frequencies. The resulting spectrum is still deterministic but somewhat more complicated to calculate. Following the
concise derivation of Abuelma'atti [9], an output product of frequency \( \omega_p \) given by

\[
\omega_p = \sum_{m=1}^{M} \alpha_m \omega_m, \quad \omega_p > 0
\]  

(9)

where \( \alpha_m \) is a positive or negative integer (or zero), has the amplitude given by

\[
V_{\alpha_1, \alpha_2, \ldots, \alpha_M} = 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n \pi} \prod_{i=1}^{M} J_{|\alpha_i|} (2n \pi V_i).
\]

(10)

Thus we are able to calculate the spectral components present in the quantized output signal for any input represented by a Fourier sine series. This result can, of course, be extended to an arbitrary periodic signal. The presence of the harmonic and intermodulation terms results in a plethora of components spread over a wide bandwidth. For an input signal containing many spectral components the numerous distortion products produced by the quantizer become essentially continuous across the audible band and take on the perceptual quality of low-level white noise. In this situation the well-known formula relating signal to quantization noise (total) is applicable: SNR (dB) = 6.02b + P, where \( b \) is the

---

**Fig. 4.** Comparison of calculated harmonic distortion amplitudes for four different arbitrary input amplitudes of single sinusoid. Amplitude of each partial \( A_p \) depends on partial number and input amplitude.

**Fig. 5.** Data of Fig. 4 now shown as decibel levels relative to input amplitude \( V \). Relative level of partial is \( 20 \log(|A_p|/V) \). Again note strong dependence of partial levels on input amplitude.
number of binary bits in a quantized output sample and $P$ depends on the statistics of the input signal (peak-to-rms ratio). Under these conditions it is reasonable and more appropriate to describe the quantization error as uncorrelated with itself and the input signal [7], [8], [10];

### 1.2 Quantization Effects in Discrete Time: Aliasing

The effects of quantization in continuous time described in the previous section are of important theoretical interest. The practical situation, however, requires consideration of discrete-time system properties, particularly aliasing [11]. The frequencies present in the quantized signal [see Eq. (6) or Eq. (8)] may contain energy above the half-sampling frequency (sample rate divided by 2) which folds over into the passband because the quantization process occurs after the input anti-aliasing filter. Depending on the relationships among the input signal component frequencies and the sample rate, the aliased components of a quantized sinusoid may collide with the harmonic distortion components or appear at inharmonic frequencies. Furthermore, higher order foldover may also be present due to distortion components with frequencies exceeding the sample rate itself, as depicted in Fig. 6. The effects of aliasing can be reduced by oversampling, thereby shifting the foldover frequency up to a range where the quantization components have fallen off to very low amplitudes.

As with the continuous-time case, the spectrum produced by quantizing a complicated input signal in discrete time is broad band and essentially white. The additional components aliased into the passband can contribute to the whitening effect. The presence of quantization distortion components may also be masked to a great extent by components present in the signal itself.

### 2 GRANULATION NOISE

The results of the preceding section do not explicitly consider the situation in which quantization distortion is commonly most objectionable—the unmasked decay of a musical note played by a solo instrument. As the sound of, say, a piano note or plucked string decays, it also typically becomes more sinusoidal as higher partials are damped out. The signal presented to the quantizer starts out as a high-level signal with complex spectrum which gradually becomes a low-level sinusoid. As noted, the spectral level of the quantization distortion components varies with the amplitude of the input signal, with a trend toward decreasing the signal-to-noise ratio as the input signal amplitude decreases. Thus the quantized signal first contains a broad-band quantization noise, which becomes discrete distortion components as the input signal decays. Moreover, the decaying amplitude of the input signal translates to rapid amplitude modulation of the distortion harmonics. This is the effect known as granulation noise.

Consider a simple example of a decaying sinusoidal signal,

$$x(t) = A e^{-\tau t} \sin(\omega t)$$

(11)

where for convenience the time constant $\tau$ is assumed to be much greater than the waveform period $2\pi/\omega$. Placing this signal in Eq. (6) gives

$$\hat{x}(t) = A e^{-\tau t} \sin(\omega t)$$

$$+ 2 \sum_{n=1}^{w} \frac{(-1)^n}{n\pi} \sum_{p=1}^{w} J_p(2n\pi A e^{-\tau t}) \sin(p\omega t)$$

(12)

The author is not aware of a closed-form expression for Eq. (12), but numerical calculations are reasonably convenient using a finite-sum formulation (again, see Appendix). The time-variant spectrum of the quantized signal (neglecting aliasing) is shown in Fig. 7. Normalized decibel ratios relating the harmonic energy to the fundamental are given in Fig. 8. Note in particular that the strengths of the individual distortion components are strongly dependent on the amplitude of the decaying input sinusoid. Note also that the total power of the combined distortion components is less dependent on the exact amplitude of the input signal, but also follows the general trend toward decreasing the signal-to-noise ratio as the signal decays.

The signal-dependent fluctuation of the distortion components, perhaps combined with aliasing, results in the gritty, broad-band background noise that modulates and increases as the signal level decreases. It is this signal-dependent noise fluctuation that causes the granular perceptual sound quality.

### 3 CONCLUSION

Based on the derivations and examples collected for this paper, the following conclusions may be drawn.

1) Quantization noise is deterministic. If the input signal and quantizer characteristic are known, the error introduced by the quantizer is also known.
2) Quantization of a simple sinusoidal signal produces new discrete harmonic components for any input amplitude. Harmonics above the half-sampling frequency are aliased, resulting in possibly inharmonic components in the passband. The presence of these discrete distortion components is of perceptual importance primarily for low-level input signals.

3) The output spectrum for a known input signal and quantizer can be calculated numerically.

4) As is widely known, quantization of complicated input signals produces an essentially white background noise contribution due to the uncorrelated nature of the quantization error from sample to sample. However, decaying musical signals often become increasingly sinusoidal as their amplitude decreases, resulting in signal-correlated noise components at discrete frequencies.

5) Although the total distortion energy of a quantized sinusoidal signal is somewhat independent of the input signal amplitude, the level of any individual harmonic varies significantly with changes in the input signal. These fluctuations can be perceived as a gritty, metallic timbre as the amplitude of the input signal decays. This is the cause of granulation noise.

---

Fig. 7. Overlapped plot of calculated individual partial levels for exponentially decaying sinusoid (100e^{-10t/0.434}, decay from 100 to 1 in 2 s). Large fluctuations in both fundamental and partial levels give rise to gritty, or granular, perceptual quality to distortion. A — input signal amplitude; B — output amplitude of fundamental; C, D, E — output amplitudes of third, fifth, and seventh partials.

Fig. 8. Data of Fig. 7 now shown as decibel levels with respect to fundamental. Fundamental-to-partial ratio = 20 log(|A_i|/|A_p|); fundamental-to-total-noise ratio = 10 log(|A_i|^2/Σ|A_p|^2).
4 ACKNOWLEDGMENT

This work was supported in part the University of Nebraska Research Council (Holmes Fellowship) and the University of Nebraska Foundation (Layman Fund). The assistance of Dr. Rod Soukup, Chairman of the Department of Electrical Engineering, and Anthony Agnello of Ariel Corporation is also gratefully acknowledged.

5 REFERENCES


APPENDIX

SUMMATION OF BESSEL FUNCTIONS

Computation of the harmonic amplitudes of Eq. (6) involves the infinite summation factor containing Bessel functions of the first kind, namely,

\[
\sum_{n=1}^{\infty} (-1)^n \frac{J_p(2n\pi V)}{n}
\]

where \(p\) is the order of the Bessel function and \(V\) is the amplitude of the input sinusoid. Recall that \(p\) is always an odd integer for the cases of interest in this paper. Unfortunately the sum converges too slowly for a practical implementation of this summation on a computer. Instead, it is desired to reformulate the summation as a finite sum using an approach comparable to that presented by Gray [5].

First, two identities are used to rewrite summation (13). For odd \(p\),

\[
J_p(x) = \frac{2}{\pi} \int_{0}^{\pi/2} \sin(\rho \theta) \sin(x \sin \theta) \, d\theta.
\]

Also,

\[
\sum_{l=1}^{N} (-1)^l \sin(l\theta) = \begin{cases} \frac{-x}{2}, & 0 \leq x < \pi \\
0, & x = \pi \\
\pi - \frac{x}{2}, & \pi < x < 2\pi. \end{cases}
\]

Using Eq. (14) in Eq. (13) and noting that \(\sin(2n\pi V \sin \theta) = \sin(2\pi \langle V \sin \theta \rangle)\) (where \(\langle . \rangle\) indicates the fractional part), gives

\[
\sum_{n=1}^{\infty} (-1)^n \frac{J_p(2n\pi V)}{n} \times \sum_{n=1}^{\infty} (-1)^n \frac{\sin(2n\pi \langle V \sin \theta \rangle)}{n} \, d\theta. \tag{16}
\]

Eq. (16) can be rewritten again, using Eq. (15), giving

\[
\sum_{n=1}^{\infty} (-1)^n \frac{J_p(2n\pi V)}{n} = \frac{2}{\pi} \int_{0}^{\pi/2} \sin(\rho \theta) \left( V \sin \theta \right) \, d\theta,
\]

\[
\langle V \sin \theta \rangle < \frac{1}{2}, \quad V \sin \theta = \frac{1}{2}, \quad \langle V \sin \theta \rangle > \frac{1}{2}. \tag{17}
\]
where \( K \) is the smallest integer greater than \( V \). This lets the right-hand side of Eq. (17) be written as

\[
2 \left\{ \sum_{k=1}^{K-1} \left\{ \int_{\sin^{-1}\left(\frac{2k-1}{2V}\right)}^{\sin^{-1}\left(\frac{k}{V}\right)} \sin(p\theta) [1 - \{V \sin \theta - (k - 1)] - \int_{\sin^{-1}\left(\frac{k}{V}\right)}^{\sin^{-1}\left(\frac{2k-1}{2V}\right)} \sin(p\theta) [V \sin \theta - (k - 1)] \, d\theta \right\} + \mu(V) \right\}
\]

(18)

where \( \mu(V) \) depends on \( \langle V \rangle \) and is given by

\[
\mu(V) = \begin{cases} 
- \int_{\sin^{-1}\left(\frac{2K-1}{2V}\right)}^{\frac{\pi}{2}} \sin(p\theta) [V \sin \theta - (K - 1)] \, d\theta, & \langle V \rangle < \frac{1}{2} \\
\int_{\sin^{-1}\left(\frac{2K-1}{2V}\right)}^{\frac{\pi}{2}} \sin(p\theta) [V \sin \theta - (K - 1)] \, d\theta, & \langle V \rangle > \frac{1}{2} 
\end{cases}
\]

(19)

Evaluating the integrals of Eqs. (18) and (19) and simplifying gives the desired results. For \( p = 1 \),

\[
\sum_{n=1}^{\infty} (-1)^n J_0(2n\pi V) = 2 \left\{ \frac{-V \pi}{4} + \sum_{k=1}^{K'} \cos \left\{ \sin^{-1}\left(\frac{2k - 1}{2V}\right) \right\} \right\}
\]

(20)

and for \( p > 1 \) (odd),

\[
\sum_{n=1}^{\infty} (-1)^n J_p(2n\pi V) = 2 \left\{ \frac{\sum_{k=1}^{K'} \cos \left\{ p \sin^{-1}\left(\frac{2k - 1}{2V}\right) \right\} }{p} \right\}
\]

(21)

where \( K' \) is the input amplitude \( V \) rounded to the nearest integer.

As desired, the infinite sum over the Bessel function is thus converted to a finite sum which involves only cosine and inverse sine functions. Eqs. (20) and (21) are substituted into Eq. (6) for calculation of the quantization distortion components for sinusoidal inputs using a digital computer program. The computation savings are particularly significant for low-level inputs, that is, small values of \( V \) give a small summation range \( K' \).
Robert C. (Rob) Maher was born in 1962 in Cambridge, U.K., of American parents. He holds a B.S. degree from Washington University in St. Louis (1984), an M.S. degree from the University of Wisconsin-Madison (1985), and a Ph.D. from the University of Illinois-Urbana (1989), all in electrical engineering. He has received several prestigious academic awards, including a four-year, full-tuition Langsdorf fellowship, a National Science Foundation Graduate Fellowship, a University of Illinois Graduate Fellowship, and an Audio Engineering Society Educational Grant. While a student, he also worked as a graduate research assistant for the University of Illinois Computer Music Project. Dr. Maher is a member of the Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi honor societies, and several professional organizations including the Audio Engineering Society, IEEE, ASA, ASEE, ICMA, and IMA. He is currently an assistant professor of electrical engineering at the University of Nebraska-Lincoln, with teaching and research interests in the application of advanced digital signal processing methods in audio engineering, electroacoustics, and computer music.