QUANTIZATION NOISE ESTIMATION FOR LOG-PCM

Mohamed Konaté and Peter Kabal

McGill University
Department of Electrical and Computer Engineering
Montreal, Quebec, Canada, H3A 2A7
e-mail: mohamed.konate2@mail.mcgill.ca, peter.kabal@mcgill.ca

ABSTRACT

ITU-T G.711.1 is a multirate wideband extension for the well-known ITU-T G.711 pulse code modulation of voice frequencies. The extended system is fully interoperable with the legacy narrowband one. In the case where the legacy G.711 is used to code a speech signal and G.711.1 is used to decode it, quantization noise may be audible. For this situation, the standard proposes an optional postfilter. The application of postfiltering requires an estimation of the quatization noise. In this paper we review the process of estimating this coding noise and we propose a better noise estimator.

Index Terms— Postfilter, quantization noise

1. INTRODUCTION

Noise suppression methods have routinely been used to reduce acoustic background noise and coding noise. The process to reduce these two types of noise has traditionally been different due to their different nature. Usually, when reducing coding noise, a postfilter is used and when reducing background noise, a speech enhancer is used (often as a prefilter before coding).

Conventional linear prediction based postfilters are used in many speech coding standards today to reduce the perceptual effect of coding noise. They are composed of short-term and long-term adaptive filters. The short-term filter emphasizes the formants and deemphasize the spectral valleys of a given speech frame. The long-term filter emphasizes the fine structure of the speech.

Speech enhancers are used to reduce acoustic background noise. The noise is estimated during non-speech intervals. This estimation usually occurs in the frequency domain. A filter is then estimated in the frequency domain based on the estimated noise amplitude at each frequency. Typical speech enhancers use simple spectral subtraction or Wiener filtering.

ITU-T G.711.1 [1] is a multi-rate wide-band extension for the well-known ITU-T G.711 pulse code modulation of voice frequencies. It is interoperable with the legacy G.711 at 64kbps. This means that a signal that is encoded with the legacy narrowband G.711 can be decoded by G.711.1 and vice-versa. It is easy to notice that a signal that was encoded by G.711.1 and decoded by G.711 is qualitatively better than one that was encoded and decoded by the legacy codec. This is due to the quantization noise shaping feature offered by G.711.1 coder. On the other hand though, the noise of a signal that was encoded by G.711 is not shaped. Therefore, the quantization noise can be heard when the signal is decoded by either system. G.711.1 proposes an optional postfilter to remedy this problem. This postfilter borrows ideas from enhancement systems. It estimates the coding noise using the noisy speech received at the decoder and generates adaptive Wiener filters to attenuate it. Such an estimation is possible because the quantization methods (A-law or μ-law) used by the coder have quantization noise with known properties.

The estimation of the quantization noise is important to optimize the performance of the postfiltering process. In this paper, we analyze the noise estimation method proposed in the G.711.1 standard. We then propose an improved quantization noise estimator.

2. LOG-PCM

Let $x(n)$ be the input signal to the quantizer and $y(n)$ be its output. The quantization error can be defined as:

$$q(n) = x(n) - y(n). \tag{1}$$

The variance of a signal $u(n)$ will be denoted σ_u^2. The signal $u(n)$ can be $x(n)$, $y(n)$ or $q(n)$. The signal to noise ratio (SNR) is defined by:

$$\text{SNR} = \frac{\sigma_x^2}{\sigma_q^2}. \tag{2}$$

2.1. Uniform Quantizers

For a *uniform* quantizer, the SNR is:

$$\text{SNR}_{\text{unif}} = 3\frac{\sigma_x^2}{x_{\max}^2}2^{2b}, \tag{3}$$

where b is the number of bits used by the quantizer. In dB, the SNR is:

$$\text{SNR}_{\text{unif}} \approx 6.02b + 4.77 - 20\log\Gamma \quad \text{[dB]} \tag{4}$$

where $\Gamma = x_{max}/\sigma_x$ is the load factor. We can see from Eq. (4) that if the standard deviation of the signal is such that the signal uses the full dynamic range of the quantizer, the SNR is maximized. On the other hand, if the standard deviation is small relative to x_{max}, the SNR quickly decreases. Uniform quantizers do not handle different intensity levels within a speech signal and different speaker volumes too well.

2.2. Nonuniform Quantizers

If one knows the probability distribution function (PDF) of the input signal, one can design a quantizer that will generate a better SNR than the simple uniform quantizer. The resulting quantizer is nonuniform: the quantization intervals are smaller where the signal's probability is the highest and they are bigger where the signals probability is smallest. A model that achieves such a nonuniform quantization is one that consisted of a compressor function $C(x)$ and a uniform quantizer at the encoder and then a dequantizer and an expander function at the decoder to recover the signal. The effect of applying the compressor on the input signal is that it renders its PDF uniform within its dynamic range. Jayant and Noll have shown in [2] that when the PDF $p(x)$ of the input is smooth, the quantization noise variance is given by:

$$\sigma_q^2 \approx \frac{x_{max}^2}{3 \cdot 2^{2b}} \int_{-x_{max}}^{x_{max}} \frac{p(x)}{|\dot{C}(x)|^2} dx \tag{5}$$

where $\dot{C}(x)$ represents the derivative of $C(x)$.

One can also find the companding function $C(x)$ that minimzes σ_q^2. The resulting SNR is maximized in this case but it still depends on the variance of the signal. For signals such as speech where the variance is time-varying, this is not always the best approach. This led to the development of methods where the SNR is constant for a large range of the signal variance. Two popular examples of such quantizers, A-law and μ-law, are logarithmic quantizers. They were standardized by ITU as G.711.

2.3. A-law and μ-law Quantizers

G.711 is a standard for a log-pcm speech coder which uses either A-law or μ-law quantizers. The compression function for the A-law compander is given by:

$$C(x) = \begin{cases} \dfrac{A|x|/x_{\text{max}}}{1 + \ln A} \, \text{sgn}\, x & 0 \le \dfrac{|x|}{x_{\text{max}}} < 1/A \\ x_{\text{max}} \dfrac{1 + \ln A|x|/x_{\text{max}}}{1 + \ln A} \, \text{sgn}\, x & 1/A \le \dfrac{|x|}{x_{\text{max}}} \le 1 \end{cases} \tag{6}$$

The compression function has a linear portion for small signals and a logarithmic portion for signals whose norms are greater than x_{max}/A. In the standard, $A = 87.56$.

The compression function for the μ-law compander is given by:

$$C(x) = x_{\text{max}} \frac{\ln(1 + \mu|x|/x_{\text{max}})}{\ln(1 + \mu)} \, \text{sgn}\, x \tag{7}$$

We can notice that the μ-law companding function is linear for small signals since $\ln(1 + ax) \approx ax$. It is logarithmic for large signal values. When $\mu|x| \gg x_{max}$, Eq. (7) becomes:

$$C(x) = x_{\text{max}} \frac{\ln(\mu|x|/x_{\text{max}})}{\ln(1 + \mu)} \, \text{sgn}\, x \tag{8}$$

3. QUANTIZATION NOISE ESTIMATION IN G.711.1

In G.711.1, the A-law properties are used to estimate both the quantization noise generated in the A-law case and the quantization noise generated in the μ-law case[1][3].

Using Eq. (5), we can approximate the SNR when A-law is used. For small signals (linear portion of the companding law), we get:

$$\dot{C}(x) = \frac{A}{1 + \ln A} \tag{9}$$

Therefore,

$$\sigma_q^2 \approx \frac{x_{\text{max}}^2}{3 \cdot 2^{2b}(\frac{A}{1+\ln A})^2} \tag{10}$$

which gives us an SNR of:

$$\text{SNR} = 3 \cdot 2^{2b} \left(\frac{A}{1 + \ln A}\right)^2 \frac{\sigma_x^2}{x_{\text{max}}^2} \tag{11}$$

In dB, the SNR when $b = 8$ bits for the uniform portion is given by:

$$\text{SNR}_{\text{unif}} \approx 77.02 - 20\log\Gamma \quad \text{[dB]} \tag{12}$$

For large signals (logarithmic portion of the companding law), we get:

$$\dot{C}(x) = \frac{x_{\text{max}}}{(1 + \ln A)x} \tag{13}$$

This gives us a constant SNR:

$$\text{SNR} = 3 \cdot 2^{2b} \left(\frac{1}{1 + \ln A}\right)^2 \tag{14}$$

In dB, the SNR when $b = 8$ bits for the logarithmic portion is given by:

$$\text{SNR}_{\text{log}} \approx 38.16 \quad \text{[dB]} \tag{15}$$

The transition between the two portion can be obtain by equating equations Eq. (12) and Eq. (15) and solving for Γ. This gives us the transition threshold $\Gamma_{\text{th}}^2 = 38.86$ dB.

Unfortunately, a mistake in [3] has propagated to the standard and to the reference code accompanying the standard. The error is the omission of the square on the bracketed term in Eq. (11). The consequence of this mistake is a different transition threshold. An additional error in the standard and the reference code is the presence of a factor of 40 (the frame length) in the computation of the SNR in the uniform portion. This creates a discontinuity of the SNR at the transition point. In Fig. 1, we represent the SNR in the standard specifications with a thin line and the correct version of the A-law SNR with a bold and dotted line. The change shown at the -50 dB signal level will be explained below.
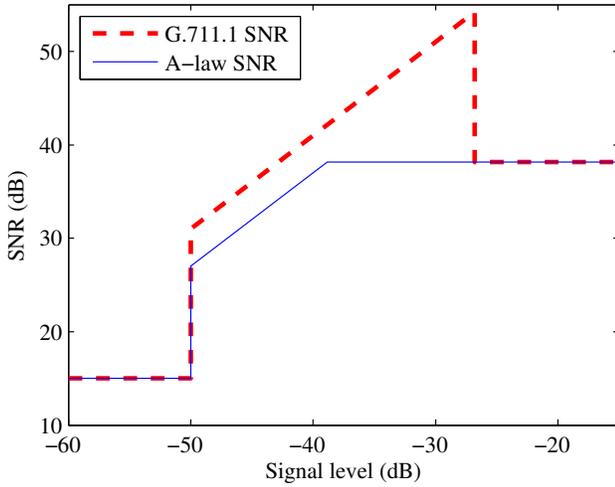
Fig. 1. Comparison of G.711.1 SNR a the correct A-law SNR

We see from Fig. 1that with a good estimate of the signal, the SNR can be determined. Given the latter and an estimate for the signal variance, we can estimate of the noise variance. This is the approach taken in the G.711.1 postfilter.

We will use "ˆ" to denote estimated values. In A-law and μ-law, the quantization noise level is usually lower than the signal. An estimate for the variance of the signal can be achieved by getting an estimate of the variance of the decoded signal: $\hat{\sigma}_x^2 \approx \hat{\sigma}_y^2$.

The variance of the decoded signal is estimated on a frame by frame basis. For a frame of length L, the estimation is given by:

$$\hat{\sigma}_x^2 = \frac{1}{L} \sum_{n=0}^{L-1} y^2(n) \tag{16}$$

Assuming $x_{\max} = 1$, the load factor is then estimated by:

$$\hat{\Gamma}^2 = \frac{1}{\hat{\sigma}_x^2} \tag{17}$$

Now that a load factor has been estimated, we can easily determine the portion of the signal that was used to code the signal. When the estimated load factor $\hat{\Gamma}^2$ is greater than Γ_{th}^2, we conclude that the uniform part of the quantizer was used. This means that the SNR in this case is SNR_{unif} and this leads to a constant noise variance:

$$\hat{\sigma}_q^2 = \frac{1 + \ln A}{3 \cdot 2^{2b} \cdot A} \tag{18}$$

When the signal energy becomes comparable to that of the quantization error, the approximation $\hat{\sigma}_x^2 \approx \hat{\sigma}_y^2$ is no longer valid. In such cases, the postfilter in G.711.1 forces the estimated $\hat{\sigma}_q^2$ to be 15 dB lower than the signal variance. This explains the discontinuity at the -50 dB in Fig. 1.

When the estimated load factor $\hat{\Gamma}^2$ is smaller than Γ_{th}^2, we conclude that the logarithmic part of the quantizer was used. This means that the SNR in this case is SNR_{log} and this leads the following noise variance:

$$\hat{\sigma}_q^2 = \frac{\hat{\sigma}_x^2}{3 \cdot 2^{2b} \left(\frac{1}{1 + \ln A} \right)^2} \tag{19}$$

4. IMPROVED QUANTIZATION ESTIMATION

A better estimate of the quantization noise can be obtained. Due to space constraints, we will only explain the method based on A-law coding. A similar approach can be done with μ-law.

In practice, the compression function is not directly used when coding with A-law or μ-law. Rather, a piecewise linear approximation to the function is used. For A-law, the approximation consists of 16 linear segments. To each segment is associated a uniform quantizer of 16 levels (4 bits). The quantization is symmetric. Therefore, 3 bits are used to identify one of 8 segments and 1 bit is used to identify the sign which result in the 8-bit representation for a coded level:

- Bit 1: sign
- Bit 2 to 4: segment number
- Bit 5 to 8: level within segment (mantissa)

The decoded signal is available at the input of the postfilter. Using it, one can easily determine the segment in which each sample was coded. Each segment corresponds to a uniform quantizer with uniformly distributed noise on its dynamic range. Therefore, it is easy to estimate the quantization noise energy for each decoded sample.

For A-law, each segment corresponds to a uniform coder with a step size. For segment $i_s[0-7]$, the step size is:

$$\Delta(i_s) = \begin{cases} 1 & i_s = 0, 1 \\ 2^{i_s - 1} & i_s > 1 \end{cases} \tag{20}$$

For each segment, this yields to an estimated noise variance of:

$$\sigma_q^2(i_s) = \frac{\Delta^2(i_s)}{12} \qquad (21)$$

For dynamic signals, it is reasonable to assume that noise is independent on a sample-to-sample basis. Therefore, for a frame of length L, the noise variance in G.711 can simply be estimated as:

$$\hat{\sigma}_q^2 = \frac{1}{L} \sum_{n=0}^{L-1} \sigma_q^2(i_s(n)) \qquad (22)$$

where $\sigma_q^2(i_s(n))$ is the variance from Eq. (21) for each sample n in the frame.

5. WINDOWING EFFECT

The postfilter is implemented in the frequency domain. The decoded signal is first windowed in time domain and then transformed into its frequency form:

$$y_w(n) = w(n)y(n) \qquad (23)$$
$$Y_w(k) = \text{FT}\{y_w(n)\} \qquad (24)$$

where y is defined as in Eq. (1), $y_w(n)$ is the windowed decoded signal, $w(n)$ is the window and $\text{FT}\{\cdot\}$ is the Fourier Transform. The window thus affects both the signal portion and the quantization noise.

In frequency domain, the postfilter gain $G(k)$ is computed by the two-step noise reduction method [1] [3]. This gain calculation is based on the SNR:

$$\text{SNR} = \frac{|Y_w(k)|^2}{|\hat{N}(k)|^2} \qquad (25)$$

where $|\hat{N}(k)|^2$ is the estimated Power Spectral Density (PSD) of the noise. Since the windowed decoded signal is used, windowing effects must be accounted for in the noise estimate.[1]

Assume that we have obtained an estimate for the variance of the noise $\hat{\sigma}_q^2$ through one of the methods discussed above. This estimate is for the "unwindowed" signal. Since the noise is white, the corresponding PSD has a constant value across all frequencies in frequency domain. From Parseval's theorem, one can show that for any white signal (and here particularly for the quantization noise), we get:

$$|Q(k)|^2 = \sum_{n=0}^{L-1} \text{E}\{|q(n)|^2\} \qquad (26)$$

Therefore, the estimated quantization noise is given by:

$$|\widehat{Q}(k)|^2 = \sum_{n=0}^{L-1} \hat{\sigma}_q^2 \qquad (27)$$

---

[1]The postfilter in G.711 does not account for this windowing effect

Let $q_w(n)$ be the windowed version of $q(n)$ i.e. $q_w(n) = w(n)q(n)$. We then get:

$$\begin{aligned} \text{E}\{q_w^2(n)\} &= \text{E}\{w^2(n)q^2(n)\} \\ &= w^2(n) \cdot \text{E}\{q^2(n)\} \end{aligned} \qquad (28)$$

Therefore,

$$\sum_{n=0}^{L-1} \text{E}\{q_w^2(n)\} = \sum_{n=0}^{L-1} w^2(n) \cdot \text{E}\{q^2(n)\} \qquad (29)$$

Since q_w is a windowed version of q, it is also white. So, from Eq. (27) and Eq. (29), we have:

$$|Q_w(k)|^2 = \sum_{n=0}^{L-1} w^2(n) \cdot \text{E}\{q^2(n)\} \qquad (30)$$

For $\text{E}\{q^2(n)\}$, we can use the estimated $\hat{\sigma}_q^2$ by either method discussed previously:

$$|\hat{Q}_w(k)|^2 = \hat{\sigma}_q^2 \sum_{n=0}^{L-1} w^2(n) \qquad (31)$$

For both methods, the window energy can be pre-computed and stored. The complexity of each method therefore depends on the computation of $\hat{\sigma}_q^2$. In the method proposed in G.711.1, one has to compute the energy of the decoded signal has shown in Eq. (16). This operation takes L multiplications and $L-1$ additions. Having that value, one can immediately get the estimated $\hat{\sigma}_q^2$. In our method, one needs to compute $\hat{\sigma}_q^2$ as shown in Eq. (22). The variances associated to each segment can all be pre-computed and stored in a table. Therefore, the computation of our estimate takes L additions.

6. RESULTS AND DISCUSSION

We implemented both the noise estimation proposed in the G.711.1 standard and the noise estimation we proposed in Section 4. We applied both methods on a speech signal (8kHz sampling frequency). Fig. 2 shows the two noise estimates relative to the true noise which was computed on a frame by frame basis as:

$$\sigma_q^2 = \frac{1}{L} \sum_{n=0}^{L-1} (x(n) - y(n))^2 \qquad (32)$$

We can see that the noise estimate that we get with our method is more accurate than the one proposed by the G.711.1. In Fig. 2, windowing is not taken into consideration.

The second experiment we ran accounted for the window. We used the same window that is used in the G.711.1 standard. The results are shown in Fig. 3. Here, we observe that the windowed noise has less energy than the unwindowed signal. This is expected as the window is tapered. Our estimate
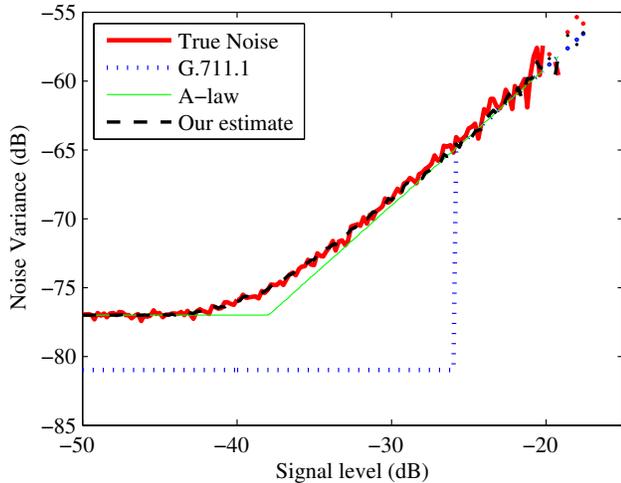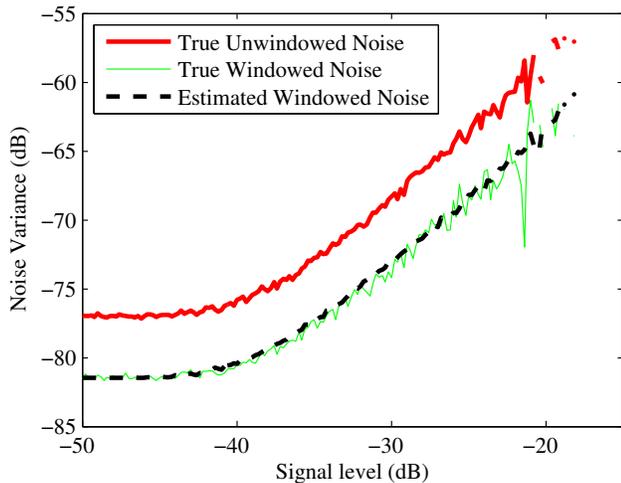
Fig. 2. Estimated noise comparison

Fig. 3. Estimated noise with windowing

Table 1. Results of PESQ Test

Attenuation	No Postfilter	G.711.1	A-law	Windowed Estimate
dB	MOS	MOS	MOS	MOS
None	4.359	4.372	4.375	4.374
20	3.415	3.559	3.561	3.560
40	1.740	1.822	1.822	1.822

ation level, we computed 4 values of PESQ: No postfilter, G.711.1 postfilter, A-law postfilter without windowing, and our estimate with windowing. The results are summarized in Table 1.

The 3 postfilters studied give MOS values which are close. For the 40 dB case, all three postfilters give the same result because they all use the 15 dB "fix" when the signal energy is below -50 dB. For the no-attenuation and 20 dB cases, we do note a slightly better result for the A-law and our version compared to the G.711.1 postfilter as we would expect. Fig. 2 and Fig. 3 show that this postfilter tends to underestimate the noise. However, this underestimation is partly offset by the failure to consider the effect of the windowing. As we have shown in experiment 2, the noise estimation in our system is more accurate. The scores we obtained can be explained by the fact that we used the same handling procedure for low energy signals that the one used by G.711.1.

7. CONCLUSION

This paper has suggested a noise estimation process which is demonstrably a better estimate that the one proposed in the G.711.1 standard. Additionally, the suggested method has a smaller complexity. However, the effective benefit in term of perceptual quality is small.

8. REFERENCES

[1] ITU-T, *Wideband Embedded Extension to G.711 Pulse Code Modulation*, Recommendation G.711.1, March 2008.

[2] N. S. Jayant and P. Noll, *Digital coding of waveforms*, Prentice Hall, 1990.

[3] J. L. Garcia, C. Marro, and B. Kosvesi, "A PCM coding noise reduction for ITU-T G. 711.1," *INTERSPEECH (Brisbane, Australia)*, pp. 57–60, September 2008.

[4] ITU-T, *Perceptual evaluation of speech quality:*, Recommendation P.862, February 2001.

coincides well with the true windowed quantization error variance. This experiment confirms that the window should be taken into account. Otherwise, the noise would be overestimated.

The third experiment we conducted consisted of the replacement of the noise estimator in G.711.1 with the one we proposed in this paper. Our test signals consisted of 6 different speakers (3 females and 3 males). The original signals use most of the dynamic range of the quantizer. For the purpose of plotting, we gathered statistics of the signal energy for each frame. These values are assigned to 2 dB bins. For each bin of signal values, we calculated the average noise variance. To test for the case of quiet talkers, we also attenuated the signal by 20 dB and 40 dB. We computed the average MOS (Mean Opinion Score) using the PESQ (Perceptual Evaluation of Speech Quality) methodology [4]. For each attenu-