QUANTIZATION NOISE SPECTRAL SHAPING IN INSTANTANEOUS CODING OF SPECTRALLY UNBALANCED SPEECH SIGNALS

Gaël Mahé and André Gilloire

France Télécom R&D / DIH / IPS,
2, avenue Pierre Marzin, 22307 Lannion cedex, France
gael.mahe@rd.francetelecom.com

ABSTRACT

In the context of centralized spectral equalization of speech in a telephone network, the signal is spectrally strongly unbalanced at the output of the equalizer, before being quantized, which results in low SNR at the reception. We propose and evaluate experimentally two methods to reshape the quantization noise, in order to make it less perceptible in reception. The first one consists in finding the most probable quantization sequence, given the desired noise spectrum. In the second one, the filtered quantization error is added to the signal to be quantized.

1. INTRODUCTION

In a PSTN telephone link, the voice spectrum is affected by two kinds of distortions in the analog part of the network: the band-pass filtering modeled by the "Intermediate Reference System" (IRS) [1], and the low-pas filtering of the analog lines.

These distortions may be corrected by an equalizer placed in the digital part of the network, as shown in Fig. 1. The principles of such an equalizer are presented in [2]. The frequency response of this equalizer is the inverse of the total response of the analog channel in the band [F_c-3150 Hz], where F_c is a cut-off frequency below 300 Hz.

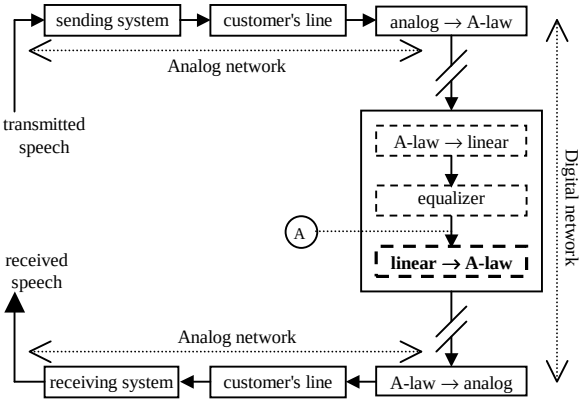

Figure 1: Telephone link with equalizer.

Since the equalizer is placed before the receiving part of the IRS, it must enhance the low frequencies (LF) components *vs* the other components in order to perform an anticipated equalization of the loss of LF components at the reception. The resulting unbalance is all the stronger as F_c is low. Figure 2 represents, for different values of F_c, the frequency responses of the global filtering applied to the speech signal between the input of the chain and the linear to A-law conversion (point A in Fig. 1).

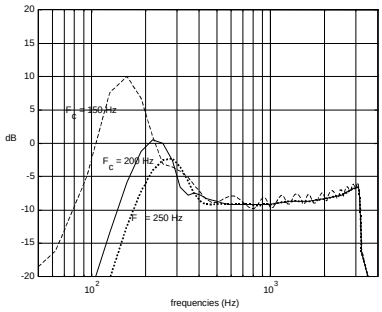

Figure 2: Frequency response of the filtering applied to voice before linear to A-law conversion (point A in Fig. 1).

Because of this level difference, the quantization noise level is close to the level of the medium and high frequencies components. After the LF loss in the receiving system, this results in a low SNR at the reception. The decrease of SNR is all the stronger as F_c is low.

In [2] we proposed to choose F_c = 250 Hz as a tradeoff between restoration of timbre and quantization noise. In this paper, we present methods to reduce perceptually the noise in the received signal while decreasing F_c, for example to 200 Hz. The principle is to reshape the spectrum of the quantization noise, so that the received noise spectrum is below the masking threshold of the received signal. A method [3] consists in whitening the signal being quantized, and reshape its spectrum at the reception. We propose to reshape the quantization noise without any additional filter at reception, simply by using differently the A-law quantization (or any instantaneous quantization) after equalization. Section 2 presents a method based on a probabilistic approach, consisting in finding the optimal quantization sequence. We present in Section 3 a second method, based on a recursive filtering of the quantization error.

2. PROBABILISTIC METHOD

2.1. Principles

Instead of quantizing each sample by the closest quantization level, we search, among all the quantization levels, the most probable sequence, given the desired noise spectrum.

Assuming the quantization is equivalent to the addition of a noise b, and given a quantization sequence $C(0…n\text{-}1)$ from

sample 0 to n-1, the conditional probability of quantizing the n^{th} sample $x(n)$ by the quantization level Q_k is:

$$\begin{aligned} &\mathrm{P}\left(Q(n)=Q_k \mid C(0...n-1), \text{spectrum of } b\right) \\ &= \mathrm{P}\left(S_k < x(n)+b(n) < S_{k+1} \mid C(0...n-1), \text{spectrum of } b\right) \end{aligned} \quad (1)$$

where P(X|A,B) denotes the conditional probability of X, given A and B, S_k and S_{k+1} are the lower and upper thresholds corresponding to the quantization level Q_k,.

The spectrum of b can be represented by an ARMA model:

$$b(n) = w(n) - \sum_{i=1}^{p} a_i b(n-i) + \sum_{j=1}^{q} d_j w(n-j), \quad (2)$$

where w is a white noise with zero mean and standard deviation σ. So $x(n)$ + $b(n)$ is a random variable with the same distribution as $w(n)$ around the mean value:

$$x(n) - \sum_{i=1}^{p} a_i b(n-i) + \sum_{j=1}^{q} d_j w(n-j). \quad (3)$$

Given the distribution of w, we can now compute the conditional probability of Q_k. The probabilities of the possible quantization sequences are computed step by step, using:

$$\mathrm{P}\left(C(0...n-1)\circ Q(n)\right) = \mathrm{P}\left(C(0...n-1)\right)\mathrm{P}\left(Q(n) \mid C(0...n-1)\right), \quad (4)$$

where ° denotes concatenation.

2.2. Application

The most probable sequences are selected by a Viterbi algorithm. For each possible continuation $C(n+1\ldots N)$ of a sequence $C(0\ldots n)$, where N is the number of samples to be quantized,

$$\mathrm{P}\left(C(0...n)\circ C(n+1...N)\right) = \mathrm{P}\left(C(0...n)\right)\mathrm{P}\left(C(n+1...N) \mid C(0...n)\right) \quad (5)$$

From the precedent sub-section (and particularly from Equations (1) and (3)), it can be deduced that the second factor of the right part of (5) only depends on the $x(i)$ and $Q(i)$ later than n-L, where L = max(p,q). Consequently, at the n^{th} sample, among all the sequences with the same L latest samples, we select the one with the highest P($C(0\ldots n)$).

For a A-law quantizer, with 256 quantization levels, this process requires to keep in memory 256^L sequences, with the corresponding noises and probabilities. In order to reduce the complexity and the amount of memory, we simplify the algorithm as follows. Considering a gaussian distribution for w, because of the quick decay of the distribution, only the 4 most probable continuations $Q(n)$ of each sequence $C(0\ldots n\text{-}1)$ are taken into account, instead of the 256 possible quantization levels.

The masking threshold is obtained by the Johnston's method [4] and updated every 16 ms. We model the noise spectrum by an ARMA model of orders p=5 and q=4, with σ so that the frequency representation of the model is 5 dB below the mask. The amplitudes of the poles are reduced to avoid rough noise.

2.3. Results

Figure 3 presents, for a frame of speech, the reception noise spectra, with and without shaping, compared to the masking threshold of the received speech signal and the ARMA model used in the algorithm. The observation of these curves for all the frames of various speech signals confirmed the validity of our algorithm in the same way, as far as the shape of the noise spectrum is concerned: it is actually reshaped according to the ARMA model.

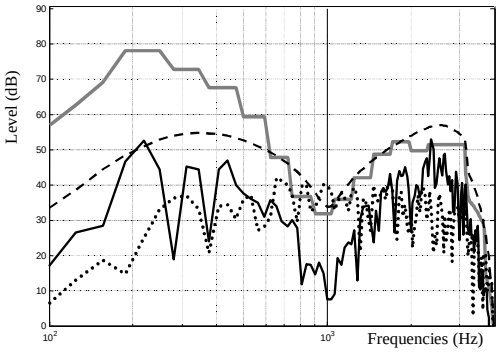

Figure 3: Reception noise spectra, with (continuous line) and without (dotted line) shaping, masking threshold (grey line) of the received speech signal, ARMA noise model (broken line).

On the other hand, the noise level does generally not match the level σ of the model, as shown in Fig. 3: the algorithm fixes the actual noise energy, whichever standard deviation we choose. Consequently, the noise cannot be masked for some frames, which depend on the pronounced words and the speakers. The masking can be objectively measured by the difference, in dB, between the actual noise power spectral density (PSD) and the PSD given by the ARMA model. This difference, which should ideally be below 0 dB, is represented in Fig. 5 for two speakers.

Subjectively, the noise is irregularly masked and appears "rough" when not masked. Moreover, a weak permanent musical high frequency noise appears. For some speakers, the white quantization noise is more comfortable.

2.4. Discussion

This method allows to mask the noise with variable performance, depending on phonemes and speakers. The limitation of the results should be moderated by the fact that they were obtained with a sub-optimal algorithm, where only the 500 most probable quantization sequences are selected at each sample, in order to reduce the complexity.

3. NOISE SHAPING WITH FEEDBACK LOOP

3.1. Principles

This method, inspired by the adaptive noise spectral shaping [3], consists in adding to the signal to be quantized s (point A in Fig. 1) the quantization error e filtered by B, as presented in Fig. 4, so that the final noise spectrum is below the masking threshold.

According to Fig. 4, with evident notations,

$$\tilde{S}(z) = S(z) + (1+B(z))E(z). \quad (6)$$

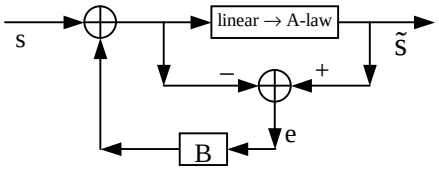

Figure 4: Noise shaping using a feedback loop.

The spectral shaping of the quantization noise involves:

$$\gamma^{R}_{noise}(f) = \lambda^2 \mathrm{Mask}(s * r), \quad (7)$$

where γ^{R}_{noise} is the PSD of the reception noise, Mask is the masking function computed using the Johnston's method, *r* is the filter equivalent to the reception line and the receiving system and λ^2 is the noise to mask ratio in reception. According to (6):

$$\gamma^{R}_{noise}(f) = |R(f)|^2 \, |1 + B(f)|^2 \, \sigma_e^2, \quad (8)$$

where σ_e denotes the standard deviation of *e*. (7) becomes:

$$|1 + B(f)|^2 \sigma_e^2 = \lambda^2 \frac{\mathrm{Mask}(s * r)}{|R(f)|^2} = \lambda^2 \gamma_{mask}(f). \quad (9)$$

So the loop filter is defined by:

$$B(z) = \frac{\lambda}{\sigma_e} H(z) - 1, \quad (10)$$

where *H* is a filter which frequency response corresponds to γ_{mask}. Since the loop has to include a delay for causality,

$$\lambda = \frac{\sigma_e}{h(0)}, \quad B(z) = \frac{H(z)}{h(0)} - 1 \quad (11)$$

3.2. Loop structure

Attention has to be paid to the stability of the loop. Using:

$$E(z) = \frac{\tilde{S}(z) - S(z)}{1 + B(z)} = \left(\frac{h(0)}{H(z)} \right) \left(\tilde{S}(z) - S(z) \right), \quad (12)$$

the system is stable if 1/*H* is stable. Since a stable IIR structure for *H*, corresponding to the previous ARMA model of noise, experimentally led to plateaus in the output, we use a FIR structure. We derive *B* directly from an AR model $\{(a'_i)_{1 \le i \le p'}; \sigma'\}$ of the inverse of γ_{mask}:

$$H(z) = \frac{1 + \sum_{i=1}^{p'} a'_i z^{-i}}{\sigma'} \rightarrow b(0) = 0, \quad b(n > 0) = a'_n. \quad (13)$$

Equation (12) becomes:

$$E(z) = \frac{1}{1 + \sum_{i=1}^{p'} a'_i z^{-i}} \left(\tilde{S}(z) - S(z) \right) \quad (14)$$

Since the AR model is naturally stable, the loop is stable.

3.3. Results

The subjective results are similar to the result of the probabilistic method, with an improvement: no musical high frequency noise appears. Parameter λ, representing the same objective measure of the noise masking as in Section 2, is presented in Fig. 5, compared to the result of the probabilistic method.

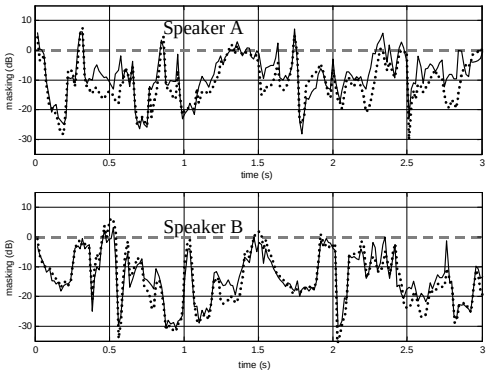

Figure 5: Measure of the noise to mask ratio: probabilistic method (continuous line) *vs* feedback loop (dotted line).

3.4. Discussion

The main advantage of this method compared to the probabilistic approach is its simplicity. Theoretically, the probabilistic algorithm can lead to a better masking, since the feedback loop method is sub-optimal: the minimization of λ, which involves the maximization of $h(0)$, is constrained by the loop stability. Both methods were simulated with a synthetic MA signal of order 1, which allowed taking into account all the most probable sequences in the probabilistic method. The resulting λ was 1 dB lower with the probabilistic method than with the loop method.

4. CONCLUSIONS

The presented methods for reshaping the quantization noise of an instantaneous coder allow to reduce the perceived noise. The noise masking is however not perfect, depending on speakers and phonemes. Their practical interest depends on the preference between a sporadic rough noise and a permanent white noise.

5. REFERENCES

1. ITU, Recommendation P.48, *"Specification for an intermediate reference system,"* 1988.
2. G.Mahé and A.Gilloire, "Correction of the voice timbre distortions on telephone network," *Proc. Eurospeech*, pp 1867-1870, 2001.
3. J.Makhoul and M.Berouti, "Adaptative noise spectral shaping and entropy coding in predictive coding of speech," *IEEE Transactions on acoustics, speech, and signal processing*, vol. ASSP-27, n° 1, pp. 63-73, February 1979.
4. J.D.Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on selected areas in communications*, vol. 6, n° 2, pp. 314-323, February 1988.