PROCEEDINGS OF SPEECH '88

NON-LINEAR QUANTIZATION RULE FOR AMPLITUDES PULSE CODING IN MULTIPULSE SPEECH SYSTEMS.

Maria-Gabriella Di Benedetto

University of Rome 'La Sapienza'- INFOCOM Dpt.- Via Eudossiana 18, 00184 Rome, Italy.

1. INTRODUCTION

The object of this study was to analyze the results obtained by using different possible combinations of number of pulses and number of bits for the amplitudes pulse coding in a multipulse excitation sequence.
The first step was to determine the relation between the number of pulses and the number of bits available for the amplitudes quantization of these pulses, considering speech segment of 16 msec duration (128 samples at a sampling frequency rate of 8 kHz). If one indicates by N the number of samples in a frame, by F_c the sampling frequency, and by F_{max} the maximum bit rate, it is easy to see that the number of bits available in each frame for the description of all the parameters is:

$$N_{max} = F_{max}.N/F_c$$

Let L be the number of pulses, N_i the number of bits for coding the amplitudes, N_p the number of bits for the linear prediction coefficients, and N_q the number of bits for coding the quantum. One has then:

$$N_{max} = N_p + N_q + L.N_i + \log_2\binom{N}{L}$$

In the present experimentation, the prediction coefficients were quantized on the basis of what proposed by Atal, [1]. The total number of bits considered for the representation of the prediction coefficients and of the quantum was 46 bits. Figure 1 shows how the number of pulses is related to the number of bits for the amplitudes pulse coding, at bit rates of 16, 14.8, 13.6, and 12 kbits/s. As one can see from Fig.1, at a bit rate of 16 kbits/s it is possible to consider 33 pulses with 3 bits for the amplitudes quantization, while at 12 kbits/s only 19 pulses would be available if one used 3 bits for the amplitudes description.
We examined the case of a maximum bit rate at 14.8 kbits/s. Different combinations number of bits-number of pulses were possible; particular attention was given to the cases in which a large number of pulses was available, such as the 30 pulses-3 bits, or the 33 pulses-2 bits combinations. In fact, it is these combinations which seem to be more appropriate if one wanted to make use of the multipulse speech system for coding signals other than speech.

2. EXPERIMENTAL RESULTS

In this experimentation, the maximum bit rate considered was 14.8 kbits/s. The procedure adopted for calculating the position and the amplitudes of the pulses was such to overcome the problems related to the non-stationary nature of speech signals, as proposed in [2]. One sentence spoken in Italian by one Italian male speaker was analyzed. Figure 2 shows the values of the signal-to-noise ratio (SNr) for different combinations. Note that the SNr increased when the number of pulses increased up to the limit of 30 pulses and 3 bits, even though the number of bits for the amplitudes description decreased, but that it assumed a value which was markedly lower when only 2 bits were available for the amplitudes description, even though 34 pulses per frame were considered.

The possibility of non linearly quantizing the amplitudes of the pulses was then investigated with the aim of improving the SNr in the case of combinations in which few bits for the amplitudes description were considered. Let a(n) be the amplitudes to be quantized and let A_{max} be the maximum amplitude, the transformation according to the μ law was applied as follows:

$$T[a(n)]=|A_{max}|.\mathrm{sign}(a(n)).\ln(1+\mu.|a(n)|/|A_{max}|)/\ln(1+\mu)$$

Figure 3 shows the values of the SNr in the cases of 30 and 34 pulses for different values of μ. Note that the best results were obtained with μ=2.5. For this value of μ, the SNr corresponding to the different possible values of the number of pulses were computed. Figure 4 shows the values of the SNr obtained by non linear quantization of the amplitudes (filled line) compared to the results obtained by linear quantization (dotted line) (already shown in Fig.2). As one can see, an improvement was obtained by the use of a non-linear quantization rule, in correspondance to the combinations with few quantization bits. The improvement was be very possibly due to the fact that with the non-linear quantization smaller amplitudes were available. Consequently, it was possible to represent the pulse amplitudes with better accuracy.

It seemed relevant to examine the behaviour of the SNr, frame by frame. In fact, the average SNr value could give a not sufficiently accurate indication of the details of the perturbations introduced by the use of a limited number of pulses and bits for the amplitudes quantization. As a reference, the SNr values obtained with a multipulse sequence of 15 pulses with no amplitudes quantization were considered. In fact, in this case, the ratio between number of pulses and number of samples in each analysis frame is about 12%, in agreement to what suggested by Atal and Remde ([3]).

The SNr values in the cases of 30 and 34 pulses are shown on Figs.5 and 6. In each of these figures, it is possible to compare the SNr values with the reference values. In all the cases analyzed, the amplitudes were quantized with a μ law with μ=2.5. Figures 5 and 6 show that the values analyzed and those of the reference were very similar and that there were no critical segments with significantly different SNr.

In the cases of particular interest (30 pulses-3 bits and 34 pulses-2 bits, always with μ=2.5), the

waveforms of the corresponding signals were plotted, together with the original waveforms. The approximation of the original signal was quite accurate, not only in the segments in which the signal was originally stationary, but also in the regions characterized by rapid transitions. An example of this comparison for few signal frames is shown in Fig.7.

As a last confirmation to the validity of using a non linear quantization rule for the amplitudes of the pulses, the waveform corresponding to 34 pulses and 2 bits but with linear amplitudes quantization was plotted. It was interesting to examine on these plots that the approximation to the original waveform obtained was often unsatisfactory during the transitions between consonants and vowels. The consonantal area is tipical of the phoneme [p], characterized by an abrupt tranformation of the vocal tract configuration and of the excitation signal. An example of this comparison is shown on Fig.8, in which the same frames as in Fig.7 are presented. This insufficient degree of accuracy leads to a degradation of the quality of the synthetic signal, which could be perceptually relevant. The subjective evaluation on the quality of the synthetic signal may result unsatisfactory.

3. CONCLUSION

The possibility of considering a multipulse sequence at a bit rate 14.8 kbits/s, using 34 pulses and amplitudes quantized with a μ law (μ=2.5) and 2 bits was considered. The results obtained were very satisfactory and the approximation of the original waveform was quite accurate. The results obtained showed that one can obtain excellent approximations of the original waveform by using several pulses, even though the pulses amplitudes were quantized roughly, if a μ quantization rule was applied. The approximations to the original waveform obtained were more satisfactory or comparable to those obtained with a lower number of pulses but more finely quantized. This result seems to be relevant to the aim of applying the multipulse model to the representation of data sequences.

5. REFERENCES

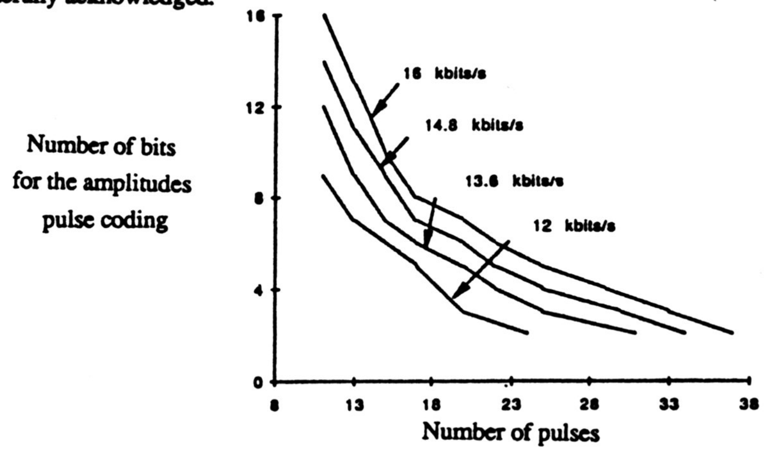
[1] B.S. ATAL, 'Predictive coding of speech at low bit rates', IEEE Trans. on Comm., Vol-Com 30, n.4, (1977), pp.600-614.

[2] M.G. DI BENEDETTO, P. MANDARINI, AND R. VIOLA, 'On critical aspects of parameters evaluation in multi-pulse LPC speech systems', Proceedings of MELECON '85, Mediterranean Electrotechnical Conference, Madrid Spain, October 8, 9, 10, (1985), pp115-118.

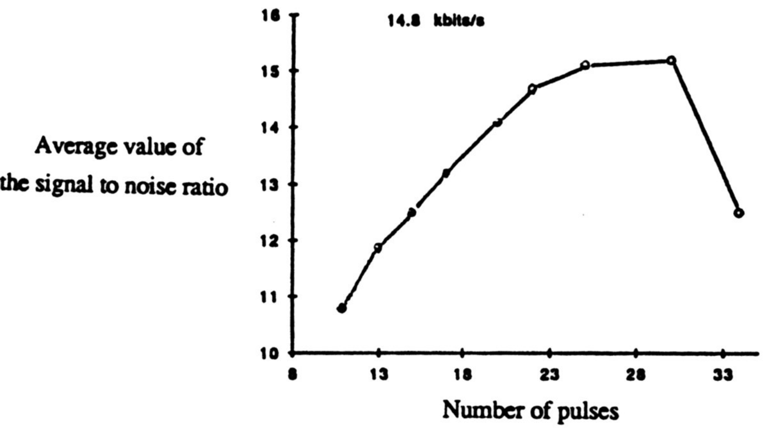
[3] B.S. ATAL, J.R. REMDE, 'A New model of LPS excitation for producing natural sounding speech at low bit rates', Proceedings Int. Conf. Acoust. Speech and Sig. Proc., Paris France, (1982), pp.614-617.

ACKNOWLEDGEMENT

The author wishes to thank Paolo Mandarini for the many discussions they have had thoughout he writing of this paper. His help and assistance in the preparation of the experiment are ratefully acknowledged.



ig.1 Relation between the number of pulses and the number of bits for the amplitudes pulse oding, for maximum bit rates at 16, 14.8, 13.6, 12 kbits/s.



ig.2 Average value of the signal to noise ratio, as a function of the number of pulses, obtained ith a linear quantization rule for the amplitudes pulse coding.

PROCEEDINGS OF SPEECH '88

NON-LINEAR QUANTIZATION RULE FOR AMPLITUDES PULSE CODING IN MULTIPULSE SPEECH SYSTEMS

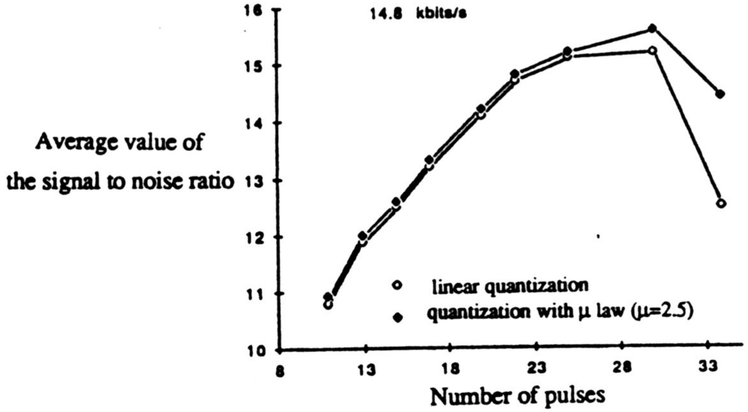


Fig.3 Average value of the signal to noise ratio, as a function of the number of pulses, obtained with linear quantization (lower curve) and non-linear quantization (higher curve) of the amplitudes of the pulses.

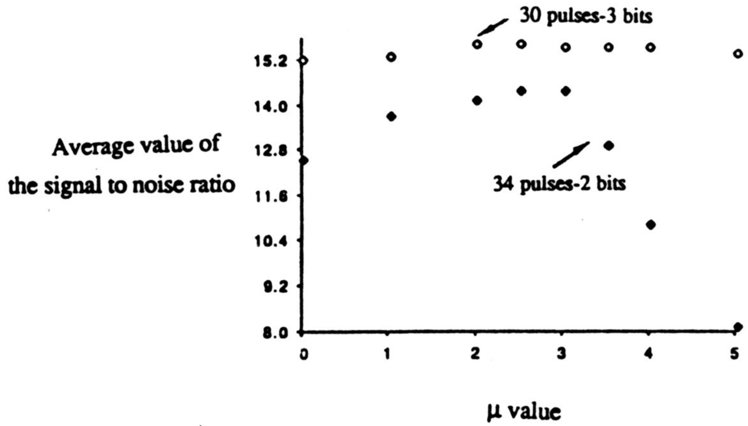


Fig.4 Average value of the signal to noise ratio, as a function of μ in the case of 30 pulses with 3 bits for the amplitudes quantization (open circles), and in the case of 34 pulses with 2 bits for the amplitudes quantization (filled circles).

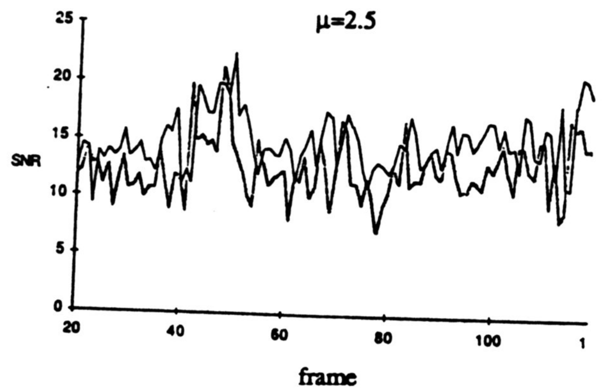
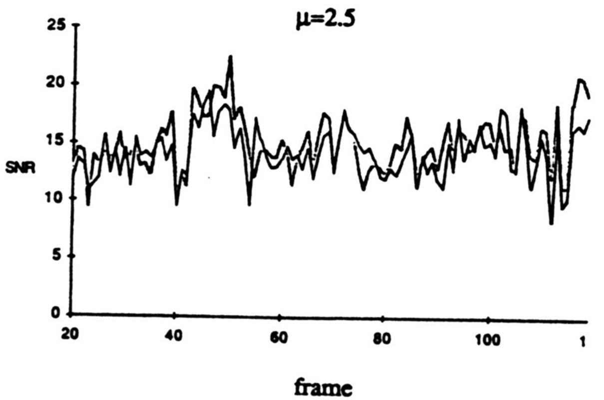


Fig.5 Signal to noise ratio values in the case of 30 pulses and 3 bits for the amplitudes quantization (continuous line) in comparison with the reference obtained with 15 pulses and no quantization of the amplitudes (dotted line).



ig.6 Signal to noise ratio values in the case of 34 pulses and 2 bits for the amplitudes uantization (continuous line) in comparison with the reference obtained with 15 pulses and no ıantization of the amplitudes (dotted line).

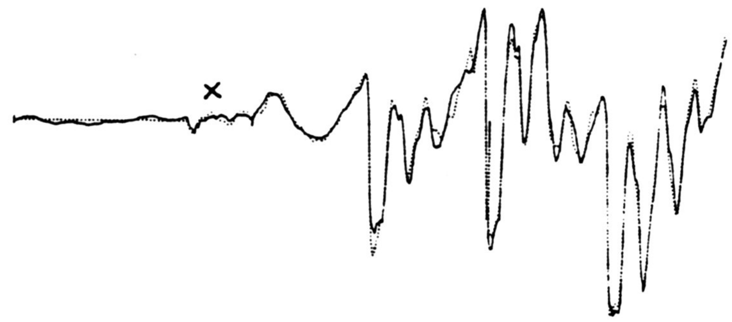


Fig.7 Example of the comparison between some frames of the original waveform (continuous line) and some frames of the synthetic waveform (dotted line), obtained with 34 pulses and 2 bits for the amplitudes non linear quantization (μ=2.5).

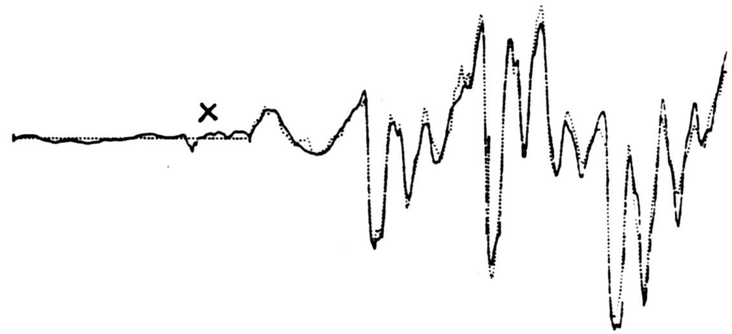


Fig.8 Example of the comparison between some frames of the original waveform (continuous line) and some frames of the synthetic waveform (dotted line), obtained with 34 pulses and 2 bits for the amplitudes linear quantization. The signal frames are the same plotted in Fig.7. Note the less accurate approximation obtained with the synthetic waveform of Fig.8, compared to Fig.7.