INFOTEH-JAHORINA Vol. 14, March 2015.

Analysis of methods for speech signals quantization

Stefan Stojkov
Mihajlo Pupin Institute, University of Belgrade
Belgrade, Serbia
e-mail: stefan.stojkov@pupin.rs

Zeljko Djurovic
School of electrical engineering, University of Belgrade
Belgrade, Serbia
e-mail: zdjurovic@etf.bg.ac.rs

Abstract—**Digital speech processing is the important part in today's communication systems, and one of its key components is the quantization of the speech signal. There are many ways of doing so, and there are many methods which can be used to get better results. One of them is to design an optimal quantizer capable of minimizing the quantization error variance, which can be further improved by introducing differential quantization. Designing of optimal quantizer requires knowledge of the probability density function (pdf) of the magnitudes of the speech signal, which must be estimated or modeled. Since all Anglo-Saxon literature gives results only for English language, it would be interesting to see how would the results differ for Serbian speech sequences.**

Keywords-optimal quantization; differential quantization; pdf estimation

I. INTRODUCTION

The purpose of speech is communication, so in today's modern time, digital speech processing is one of the fundamental parts of digital signal processing. It has become increasingly important to transfer, store, and process speech signals quickly, and to do so in a way that would preserve its message content, and represent it in the most convenient form to enable modifications without harming its message content. As stated in the book written by Lawrence Rabiner and Ronald Schafer [1], the representation of speech signal must be such that can be easily extracted by human listeners, or automatically by machine. The mentioned book is considered to be the foundation for digital speech processing, but its models and results are given for the English language. It would be useful to check the results for Serbian language, and that was one of the goals of this paper.

First step of any digital processing algorithm is the discretization of the signal, which consists of sampling and quantization. These two processes often cannot be separated in practice, but this kind of separation is crucial for some theoretical considerations. Several quantization techniques will be shown in this paper, and the focus will be on differential optimal quantization. Since using optimal quantization requires knowledge of probability density function of the magnitude of the signal which is being processed, pdf must be estimated or modeled. Gamma and Laplace function are considered to be the best functions for this kind of problem, but, they are also tested for English language. In the paper written by Predrag Tadic [2], the third approximation is introduced, which is the combination of Gamma and Laplace distributions. He described its usage, and tested it for both English and Serbian language, using optimal quantization. That distribution will be used in this paper as well. Differential optimal quantization is one step further, and the purpose of this paper is to show that its extra complexity, which is introduced by using this algorithm, is justified. As mentioned before, the goal is also to see how would the results for English and Serbian language differ.

In order to obtain representative results, three speech sequences of English languages are used during the testing, as well as three sequences of Serbian language. They are chosen in the manner that would be representative for English and Serbian language, i.e. they contain all phonemes for the respected language. Development of the system, signal processing and displaying of results are realized using MATLAB software package.

II. STATISTICAL MODEL FOR SPEECH

In order to design a system which uses differential optimal quantization, several statistical methods must be applied to the speech signal. When doing so, it is necessary to estimate or model probability density function and autocorrelation function (or power spectrum) on the basis of the waveform of the speech signal. This can be done using assumption that the waveform of speech signal can be represented as an ergodic random process. This is a bit gross assumption, but useful results justify it.

Estimation of probability density function can be achieved using histogram method. First thing that must be done is to divide the interval which covers all possible values of the signal into n equal segments, and then count how many values fall into each interval. Value of the probability density function $f(x_i)$ at x_i, which is the center of the i-th interval can be calculated in the following manner:

$$f(x_i) = N_i / (N \cdot \Delta) \tag{1}$$

where N_i is the number of samples in i-th interval, N is total number of samples, and Δ is width of the segment.

The other approach is to model pdf with some known distributions. One of the best pdf approximation for the speech signal is the gamma function:

$$f_\Gamma(x) = \left(\frac{\sqrt{3}}{8\pi\sigma_x |x|} \right)^{1/2} \exp(-\frac{\sqrt{3}\,|x|}{2\sigma_x}) \tag{2}$$

where σ_x is the standard deviation of the signal. Simpler, but less precise approximation is the Laplace pdf:

$$f_L(x) = \frac{1}{\sqrt{2}\sigma_x}\exp\left(-\frac{\sqrt{2}\,|x|}{\sigma_x}\right) \quad (3)$$

Even though gamma approximation is better suited than Laplace pdf for the amplitudes of the speech signal, it has a big flaw - it is not defined for samples which are near the mean values of the signal (in this case at zero). That is the reason why the third approximation is introduced, a mixed pdf, mention in the Introduction, which combines gamma and Laplace distribution:

$$f_{komb}(x) = \begin{cases} f_L(x) = \dfrac{1}{\sqrt{2}\sigma_x}\exp\left(-\dfrac{\sqrt{2}\,|x|}{\sigma_x}\right), & |x| \le gr \\ f_\Gamma(x) = \left(\dfrac{\sqrt{3}}{8\pi\sigma_x\,|x|}\right)^{1/2}\exp\left(-\dfrac{\sqrt{3}\,|x|}{2\sigma_x}\right), & |x| > gr \end{cases} \quad (4)$$

where value *gr* is computed in the way which preserves the basic quality of the probability density function that its integral over the entire space is one. The mentioned pdfs are shown in Fig. 1 (for one speech sequence).

III. QUANTIZATION

A. Quantization basics

As previously stated, in many practical cases, it is impossible to separate sampling from quantization, but it is very useful for some theoretical considerations. When the process of sampling in time is finished, we get signal discrete-time signal, but still continuous-amplitude signal. Quantization of this kind of signal gives a signal which is both time and amplitude discrete. The resulting signal is then being coded, which is done by the encoder. Analogously, it is also necessary to define decoder at the receiving side, which will return the coded signal into quantized sequence. This kind of system is shown in Fig. 2, where *Δ* is the quantization step.

Quantized samples are mostly represented with binary numbers, so with *B* bits it is possible to represent 2^B quantization levels. It is important to minimize the number of bits used, because the information capacity, required to transmit or digital representation, is directly proportional to the number of bits.

There is one more thing that needs to be considered. The fact is that we must cover entire range of input signal with the finite number of symbols, so we must declare value X_{max}, for which is true that $|x_n| \le X_{max}$. It is desirable that X_{max} is infinite, but in the reality, it is not. However, it can be shown that only a small percent of speech samples will not be taken into consideration (only 0.35% if the Laplace density is assumed):

$$P\{-4\sigma_x \le x \le 4\sigma_x\} = \int_{-4\sigma_x}^{4\sigma_x} \frac{1}{\sqrt{2}\sigma_x}\exp\left(-\frac{\sqrt{2}\,|x|}{\sigma_x}\right)dx \approx 99.65\% \quad (5)$$

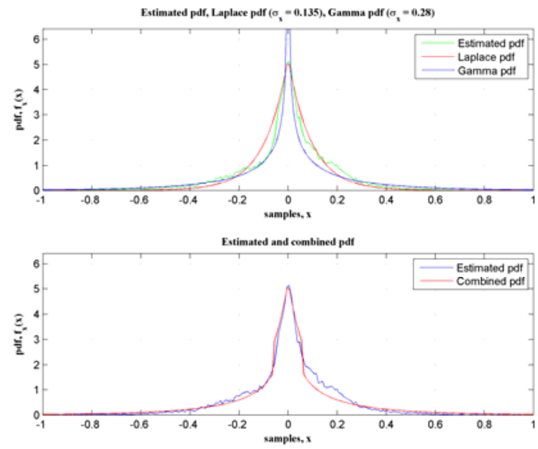

Figure 1 - Probability density functions which can be used for the speech signals

where *P* is probability that samples will fall inside the range. It can also be seen that we should bear in mind the standard deviation of the signal when deciding about the range of the input signal.

B. Optimal quantization

There are several types of quantization, and one of the most common is the uniform quantization, where quantization levels and intervals are uniformly distributed. That is also one of the simplest forms of quantization, but its signal-to-noise ratio (SNR) depends solely on the variance of the input signal, which is not desirable. It would be better if the error could be proportional to the level of the input signal, i.e. to be constant by percentage. This can be achieved using non-uniform quantizers. For example, using logarithmically spaced (*mu-law* companding) can improve dynamic range of quantization with relatively small loses in SNR ratio. Another approach is to choose quantization levels in order that would minimize the quantization error variance, and maximize SNR. In order to achieve this goal, we can design optimal quantizer.

To do this, two sets of equations most be solved simultaneously [3]:

$$\int_{x_{j-1}}^{x_j} \left(\hat{x}_j - x\right) f_x(x)\,dx = 0,\; j = 1, 2, \ldots, \frac{M}{2} \quad (6)$$

$$x_j = \frac{\hat{x}_j + \hat{x}_{j+1}}{2},\; j = 1, 2, \ldots, \frac{M}{2} - 1 \quad (7)$$

where $f_x(x)$ is the pdf of the amplitude of the speech signal. By solving these equations, we will get the set of parameters $\{x_i\}$ and $\{\hat{x}_i\}$, which will minimize the quantization error variance. Several assumptions have been made during this process:

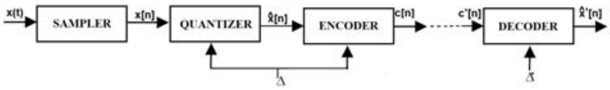

Figure 2 - Process of signal discretization

- The central boundary point is set to zero:

$$x_0 = 0 \tag{8}$$

- With the assumption that pdf is nonzero for large amplitudes (which is the case for Laplace, gamma and mixed distribution), the edge values are set to ±∞, i.e.:

$$x_{M/2} = \pm\infty \tag{9}$$

Characteristic of the optimum quantizer will be antisymmetric, since $f_x(x) = f_x(-x)$:

$$\hat{x}_i = -\hat{x}_{-i},\ x_i = -x_{-i} \tag{10}$$

We can conclude from (6) that the optimum location of the quantization level $\hat{x}_i$ is at the centroid of probability density over the input range $[x_{i-1}, x_i]$. Equation (7) states that the limit points must be in the middle of the range $[\hat{x}_i, \hat{x}_{i+1}]$. As mentioned before, these two sets must be solved at the same time in order to get the parameters for optimum quantization. However, due to the nonlinearity of these equations, in most cases it is not possible to obtain close form solution. Therefore, iterative methods must be used. For the purpose of this paper, the optimal quantizer is designed using MATLAB software package. The function is written which is being minimized over the variable $\hat{x}$ using MATLAB's function *fminsearch*. The mentioned function uses (6) and (7) in order to obtain the optimal solution. The following criterion function is used during this process:

$$J = \left[\int_{x_{j-1}}^{x_j} \left(\hat{x}_j - x\right) f_x(x)\,dx\right]^2 + vp \tag{11}$$

where value *vp* is used to ensure that values of the vector $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_{M/2}]$ are arranged in the ascending order. It has one other purpose, and that is to make sure that the first and the last value does not fall outside the specified values. Put it simply: The first element must not be less than the minimum allowed value, and the last element must not exceed the maximum allowed value. By minimizing (11), we can obtain optimal set of parameters which will minimize quantization error variance, and thus maximize SNR. Keeping in mind the symmetry of the quantizer, (8) and (10) are valid here.

A characteristic of one optimum quantizer is shown in Fig. 3. Even though usage of optimum quantizers ensures minimal quantization error, its results are not always satisfactory, due to the non-stationary nature of the speech communication process.

C. *Differential quantization*

By utilizing differential quantization, the results obtained using optimal quantization can be improved. This type of quantization is based on a fact that speech signal does not change so quickly from sample to sample. Consequently, correlation is high between neighboring samples, and the difference variance of those samples is lower than the variance of the signal itself. General scheme of differential quantization is shown in Fig. 4. Here, input signal to the quantizer, $d[n]$, is the difference between the unquantized input signal, $x[n]$, and the prediction of the input signal $\tilde{x}[n]$. The input signal to the quantizer is also called prediction error signal, and that is the signal that will be quantized. By looking at Fig. 4, it can be shown [1,4] that the input to the predictor is basically quantized input signal:

$$\hat{x}[n] = x[n] + \varepsilon[n] \tag{12}$$

where $\varepsilon[n]$ is the quantization error. So, we can come to the conclusion that, if the prediction is good, it is possible to design quantizer in the manner which would lower quantization error in comparison with the direct quantization. In this case, SNR is:

$$SNR = \frac{\sigma_x^2}{\sigma_\varepsilon^2} = \frac{\sigma_x^2}{\sigma_d^2}\cdot\frac{\sigma_d^2}{\sigma_\varepsilon^2} = G_p \cdot SNR_Q \tag{13}$$

where SNR_Q is the SNR of the used quantizer, and G_p is the gain we get by introducing differential quantization. Any type of quantizer can be used fixed or adaptive, uniform or non-uniform, etc. SNR_Q depends on properties of the chosen quantizer, as well as the properties of its input signal. This value can be maximized in many ways, for example, by using optimal quantization, described in the previous section. We can also maximize G_p in order to further improve the results of quantization.

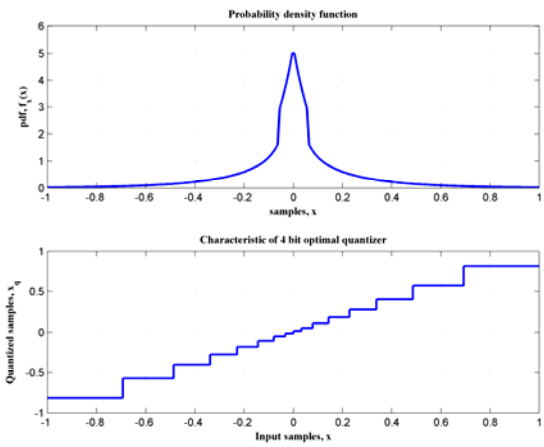

Figure 3 - pdf and characteristic of 4-bit optimum quantizer

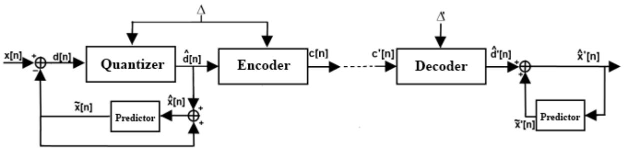

Figure 4 - Differential quantization

This can be achieved by choosing the right predictor. One way of doing so is to minimize the variance of the prediction error by designing linear predictor. In that case, the output of the predictor is a linear combination of past quantized values:

$$\tilde{x}(n) = \sum_{k=1}^{p} \alpha_k \hat{x}(n-k) \tag{14}$$

where α_k are the coefficients of the predictor. In order to achieve the minimum variance prediction error, we must differentiate this variance over the α_k and set the derived equation to zero. Variance of the difference signal $d[n]$ can be computed using the following expression:

$$\sigma_d^2 = E\left\{d^2(n)\right\} = E\left\{\left[x(n) - \sum_{k=1}^{p}\alpha_k \hat{x}(n-k)\right]^2\right\} \tag{15}$$

By differentiating (15), combining the result with (12), and with the assumption that the input signal $x[n]$ and quantization error signal $\varepsilon[n]$ are not correlated, the following expression is produced:

$$\alpha = C^{-1} \cdot \rho \tag{16}$$

where:

$$\rho = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix}, C = \begin{bmatrix} 1+\frac{1}{SNR} & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & 1+\frac{1}{SNR} & \cdots & \rho(p-2) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1+\frac{1}{SNR} \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \tag{17}$$

In (17), ρ is normalized autocorrelation function, obtained by dividing correlation function of the input signal with its variance, p is the order of the predictor. Matrix C is Toeplitz matrix, and there are many numeric methods to compute its reverse matrix. The problem here are elements on the main diagonal which depends upon SNR. One way of dealing with this is to assume that SNR>>1, and therefore neglect the term 1/SNR. For example, with that assumption and with the usage of the first order predictor (where $\alpha_1 = \rho_1$), G_p is greater than 2.77 (or 4.43 dB) for the lowpass filtered speech samples.

IV. Experimental Results

As stated in the Introduction, the goal of this paper is to check how the results of quantization would differ for Serbian and English when utilizing differential optimal quantization. Since differential quantization is far more complex than instantaneous quantization, the objective was also to show that introducing that kind of complexity is justified. Instantaneous (direct) optimal quantizer is designed using the procedure described in IIIB and utilizing experimental and mixed pdf (4). Optimal quantizer is also used for differential quantization, but in that case, the input to the quantizer are samples of the error prediction signal, and therefore its distribution of magnitudes is used (again experimental and mixed pdf). The results are presented in tables I and II.

These results clearly show the advantage of differential quantization in comparison with the instantaneous one. The SNR increase is quite obvious. Differential configuration depends on the correlation of samples and, according to (16), when using first and second order predictors, values $\rho(1)$ and $\rho(2)$ are of interest. Accordingly, the advantage of differential quantization is clear when these values are high. For example, for one speech sequence where $\rho(1) = 0.926$ and $\rho(2) = 0.756$ (using the experimental pdf), the following values for SNR are obtained: 12.42, 20.71 and 26.51 dB for 2, 3 and 4 bits, respectively, when utilizing first order predictor. If second order predictor is used, the results are even better: 16.37, 22.39 and 28.95 dB, respectively. In comparison, instantaneous optimal quantizer for the same sequence gives: 7.55, 12.82 and 18.81 dB, so the gain is obvious, and the introduction of extra complexity is therefore justified.

In tables I and II, we can also see how the average results differs for Serbian and English language.

V. Conclusion

The purpose of this paper is to present some of the basic speech quantization technics with special emphasis on differential optimal quantizer. Knowledge of pdf of speech signal (precisely, magnitudes of its samples) is required for this type of quantization. Bearing in mind that the results in Anglo-Saxon literature are given for English language only, the pdf is modeled for Serbian language, too. Estimated pdf is used, obtained by utilizing histogram method, as well as mixed pdf, which combines advantages of gamma and Laplace distributions. These two distributions are considered to be the best suited for the approximation of the distribution of the speech signal.

Table I. Average SNR in dB for Serbian language

Number of bits		B=2	B=3	B=4
Optimal quantizer	Experimental pdf	7.58	13.15	18.4
	Mixed pdf	6.97	12.52	18.17
Differential optimal quantizer[a]	Experimental pdf	11.22	17.79	23.22
	Mixed pdf	9.56	17.48	22.13
Differential optimal quantizer[b]	Experimental pdf	12.81	18.79	24.04
	Mixed pdf	11.82	18.84	22.88

a. 1st order predictor
b. 2nd order predictor

TABLE II. AVERAGE SNR IN DB FOR ENGLISH LANGUAGE

Number of bits		B=2	B=3	B=4
Optimal quantizer	Experimental pdf	7.33	12.7	18.42
	Mixed pdf	7.09	12.65	18.45
Differential optimal quantizer[a]	Experimental pdf	10.94	18.1	24.11
	Mixed pdf	8.81	17.7	22.53
Differential optimal quantizer[b]	Experimental pdf	13.61	20.19	26.15
	Mixed pdf	13.14	20.47	24.68

a. 1st order predictor

b. 2nd order predictor

Although gamma distribution is a better approximation, it has a major flaw - it is not defined for samples which are near their mean value, so the usage of the mixed pdf is justified. Tables I and II show that it gives satisfactory results.

The advantage of differential optimal quantization is also shown in this paper in comparison with instantaneous optimal quantization. Depending on the correlation of the samples of the speech signal, it is shown that differential optimal quantizer can obtain SNR better up to 10 dB than the one obtained by using instantaneous optimal quantizer. Therefore, extra complexity of the differential configuration is more than justified.

In this paper, it is also shown how the results differ for Serbian and English speech sequences.

REFERENCES

[1] L. R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signals", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978

[2] Predrag R. Tadić, Željko Đurović, Branko Kovačević, "Analysis of Speech Waveform Quantization Methods", Journal of automatic control, University of Belgrade, Vol. 18(1): 19-22, 2008

[3] M.D. Paez, T.H. Glisson, „Minimum Mean Squared Error Quantization in Speech," IEEE Trans. Comm., Vol. Comm-20, pp. 225-230, 1972

[4] C. C. Cutler, "Differential Quantization of Communications", U. S. Patent 2, 1952