

Speech Signal Coding Using Forward Adaptive Quantization and Simple Transform Coding

Milan Z. Tancic¹, Zoran H. Peric¹, Stefan Tomic¹, Nikola Simic¹

¹*Faculty of Electronics Engineering, University of Nis,
Aleksandra Medvedeva 14, 18000, Nis, Serbia
zoran.peric@elfak.ni.ac.rs*

Abstract—The paper proposes a novel speech signal coding scheme that implements a simple transform coding and forward adaptive quantization. The proposed scheme is adapted to the input signal variance, providing highly efficient bandwidth usage, whereas implemented transform coding provides sub-sequences with more predictable signal characteristics, so that more suitable signal processing can be performed. The aforementioned transform coding precedes adaptive quantization, providing additional compression. The objective quality measure used for system performance estimation is SQNR (signal-to-quantization-noise ratio), which represents a standard measure for lossy coding types. The influence of transform coding is discussed by comparing the obtained results with the corresponding one achieved by applying only the same adaptive quantization. Furthermore, the comparison with system performance of PCM (pulse-code modulation) coding system confirms that the proposed coding scheme has a lot of potential for further implementation, since that the proposed system ensures SQNR gain up to 4.0983 [dB] for various values of system parameters.

Index Terms—Speech signal coding; transform coding; forward adaptive quantization; pulse-code modulation.

I. INTRODUCTION

Speech signal presents a non-stationary process, whose average power is significantly varying in time domain, resulting in a wide dynamic range. However, occurrence of a wide set of frequencies is rare. Consequently, signal compression methods can be applied successfully, whereas input signal can be described with Laplacian source. This way, signal compression makes storing and transmission of the digital signal easier, since it requires less memory resources and narrower bandwidth for transmission while customers' experience is satisfactory [1], [2]. Although natural speech signal processing is the mostly researched, traditionally, with the growth of information technologies a lot of papers are dedicated to synthetic speech signal processing, due to its' importance in education (distance learning, foreign languages, blind individuals) and automatic recognition [3], [4].

This paper proposes a novel speech signal coding scheme based on transform coding with application of forward adaptive quantization. The main idea is to exploit simple transformations for signal decomposition, providing narrower bandwidth usage for signal transmission over the

channel and lower bit-rates. The proposed transformations decompose input signal into two independent sequences with more predictable statistical characteristics. Moreover, those facts have motivated us to adapt the coding scheme to individual sequence and perform more efficient signal quantization. Thus, special attention is paid to support range discussion [5]–[6].

There are two main kinds of adaptive quantizers: forward adaptive and backward adaptive quantizers. The coding scheme proposed in this research implements forward adaptive quantization, due to its' less sensitivity to transmission errors – a common problem in voice over IP applications. Moreover, it is well-known from literature that forward adaptive quantization provides SQNR gain of 1[dB] in comparison to backward adaptation [7]. To sum up, the forward adaptation was introduced to further adapt range of varying subsequences of the input signal [8].

The paper is organized as follows. In Section II it is provided an overview of transform coding and it is proposed quantizer design. The numerical results and performance analysis will be presented in Section III. In the end, conclusions and future research plans will be presented in Section IV.

II. TRANSFORM CODING AND QUANTIZERS DESIGN

Transform coding is one of the most popular methods for data compression, and certainly an indispensable part of systems that perform high compression rate. Nowadays, a lot of transformations such as wavelet, DCT (Discrete Cosine Transform) or Hadamard are widely used in different systems – not only for speech signal processing, but also for image and video processing. However, with increasing the level of transformation complexity, required processing time is also increased, producing higher delay. On the other hand, one of the most important characteristics of systems that transmit speech signal over the channel is very low processing delay, aiming real-time transmission. Consequently, in this paper we propose a novel speech signal coding scheme that implements a simple transform coding. The applied transformations divide the input signal into two branches with different support range. Next, forward adaptive quantization in each branch is applied in order to make additional signal compression. Speech signal is sampled at 8 [kHz], which is standard sampling frequency for digital telephone communication and VoIP. The proposed coding scheme is shown in Fig. 1.

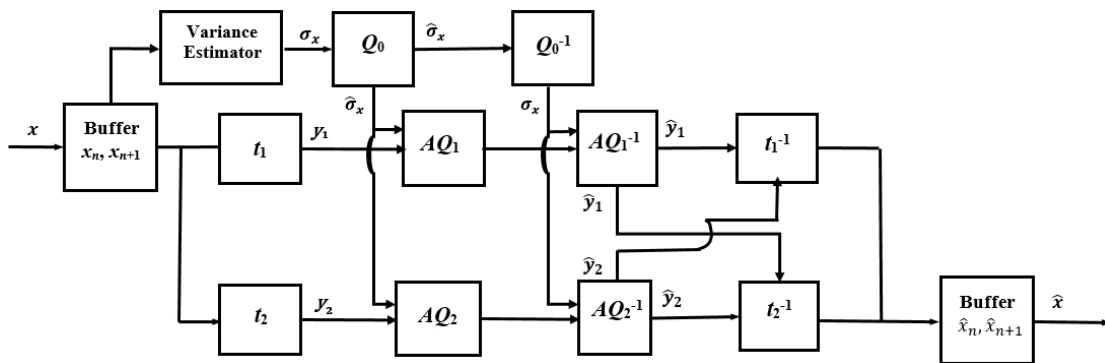


Fig. 1. The proposed speech signal coding scheme.

Simple transformations applied into this coding scheme are defined by following expressions [8]:

$$y_1 = \frac{x_n + x_{n+1}}{2}, \quad (1)$$

$$y_2 = \frac{x_n - x_{n+1}}{2}, \quad (2)$$

where x_n and x_{n+1} are samples of the input signal, whereas y_1 and y_2 represent transformed signals. Signals obtained after transformations have variances σ_1^2 and σ_2^2 , which depend on the input signal variance σ_x^2 and correlation coefficient ρ [8]:

$$\sigma_1^2 = \frac{\sigma_x^2}{2}(1 + \rho), \quad (3)$$

$$\sigma_2^2 = \frac{\sigma_x^2}{2}(1 - \rho). \quad (4)$$

Transformed sequences are independent and they are further coded separately by using quasi-logarithmic quantizers with forward adaptation (quantizers Q_1 and Q_2).

Quantizers Q_1 and Q_2 are designed using μ -logarithmic compression law, whose compression function is defined with [7]–[10]

$$c_\mu(x) = \frac{x_{\max}}{\ln(1 + \mu)} \ln\left(1 + \mu \frac{|x|}{x_{\max}}\right) \text{sgn}(x), \quad (5)$$

where $|x| \leq x_{\max}$, x_{\max} is support range of quantizer whereas μ is compression factor. According to μ -logarithmic compression function, decision thresholds x'_i and representation levels y'_i are obtained as [7]–[10]:

$$x'_i = \frac{x_{\max}}{\mu} \left(\exp\left(\frac{2i}{N} \ln(1 + \mu)\right) - 1 \right), \quad (6)$$

$$y'_i = \frac{x_{\max}}{\mu} \left(\exp\left(\frac{2(i-0.5)}{N} \ln(1 + \mu)\right) - 1 \right), \quad (7)$$

where $i = 1, 2, \dots, N$. In order to achieve higher reconstructed signal quality, adaptive quantization on signal's variance has been performed. The scheme of adaptive quantization is shown in Fig. 2 [10].

After reception of transformed signal, it is lead to the

buffer at adaptive quantizer entrance.

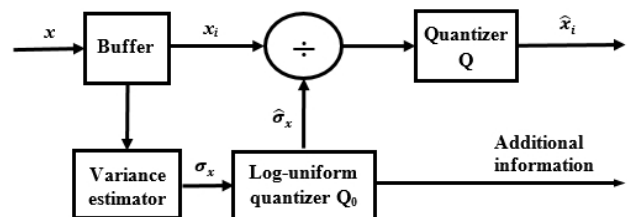


Fig. 2. Forward adaptive quantization.

This way, frames of M input samples are being loaded. Variance estimator calculates variance of M received samples (the whole frame) that is later quantized and transmitted to the receiver. This can be done by using log-uniform quantizer that deals with corresponding values of variance in logarithmic domain [10]:

$$20 \log(\hat{\sigma} = \hat{\sigma}_k) = 20 \log_{10}(\sigma_{\min}) + \frac{2k-1}{2} \Delta^{\text{lu}}, \quad (8)$$

$$\Delta^{\text{lu}} = \frac{20 \log_{10}\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)}{N_g}, \quad (9)$$

where $k=1, \dots, N_g$, Δ^{lu} represents the quant width, whereas N_g is a number of representation levels [3], [11]. Log-uniform quantizer is designed for low and middle bit-rates (number of quantization levels (N_g) is 2, 4, 8 and 16). Variance estimator sends calculated standard deviation of frame to the log-uniform quantizer. This way, quantized standard deviation is used for determining support range values ($x_{\max 1}$ and $x_{\max 2}$) for quantizers Q_1 and Q_2 , respectively. This way, quantizers Q_1 and Q_2 are adapted to the input signal for each frame. These values are calculated with:

$$x_{\max 1} = x_{\text{ml}} \times \hat{\sigma} \times \sqrt{\left(\frac{1+\rho}{2}\right)}, \quad (10)$$

$$x_{\max 2} = x_{\text{ml}} \times \hat{\sigma} \times \sqrt{\left(\frac{1-\rho}{2}\right)}, \quad (11)$$

where $\hat{\sigma}$ represents quantized standard deviation, ρ is correlation coefficient of the input signal whereas x_{ml} is support range of Q_1 and Q_2 for Laplacian source with the unit variance [9], [12], [13]:

$$x_{\text{ml}} = \frac{1}{\sqrt{2}} \ln\left(\frac{3\mu N^2}{\ln^2(1 + \mu)}\right), \quad (12)$$

where N is the average number of quantization levels of quantizers Q_1 and Q_2 , whereas μ is compression factor.

SQNR for proposed coding scheme is calculated by averaging SQNR value for all frames [7]–[13]

$$SQNR_{av}[dB] = \frac{1}{N_f} \sum_{j=1}^{N_f} SQNR_j, \quad (13)$$

where SQNR per frame is obtained by

$$SQNR_j[dB] = 10 \times \log_{10} \left(\frac{\sigma^2}{D_j} \right). \quad (14)$$

In previous equation, with D_j is denoted signal distortion, which can be calculated with

$$D_j = \sum_{i=1}^M (x_i - \hat{x}_i)^2, \quad (15)$$

where with $\hat{}$ are denoted quantized values that are transmitted to the decoder.

III. NUMERICAL RESULTS

In this section, experimental results of applying the proposed coding scheme are presented. The benchmark speech signal is recorded in Laboratory of Acoustics, Faculty of Electronic Engineering, University of Nis. The signal is natural, it is sampled at 8 [kHz] and its' variance is equal to 0.9086. The obtained results, after processing it with the proposed scheme, are compared to the corresponding one obtained by using coding scheme without transformation coding that precedes forward adaptive quantization as well as with performance of PCM coding system. The experiments are performed for different frame sizes $M \in \{40, 80, 160, 240\}$, low and middle bit-rates for log-uniform quantizer Q_0 , $N_g \in \{2, 4, 8, 16\}$, different average bit-rates of quantizers Q_1 and Q_2 , and for compression factor $\mu = 16.9227$, taken from paper [4].

Table I shows SQNR values obtained by using the proposed system, for various value of average bit-rate:

$R = 5.5$ [bits/sample] (bit-rate of quantizer Q_1 is $R_1 = 6$ [bits/sample], whereas $R_2 = 5$ [bits/sample] for Q_2), $R = 6$ [bits/sample] and $R = 7$ [bits/sample], depending on all of the aforementioned frame sizes and N_g values. The obtained results are compared with the corresponding SQNR values achieved by using PCM for the same bit-rates and $\mu = 16.9227$. In Table I and Table II, the total average bit-rate is calculated with $R_{av} = R + R_v$, where $R_v = N_g / (2M)$ represents the required number of bits/sample to transmit signal variance for observed frame.

By observing results shown in Table I, it can be concluded that the performance achieved by using the proposed scheme for all presented combinations of frame size and number of representational levels of log-uniform quantizer Q_0 , ensures higher SQNR than PCM system – the gain is reaches even 4.0983[dB] for the most complex design. It should be noted that SQNR for PCM system is a constant value depending on M and N_g , and it is equal to 25.5040[dB] for $R = 5.5$ [bits/sample], 32.0016[dB] for $R = 6.5$ [bits/sample], whereas for $R=7$ [bits/sample] it is equal to 35.1832[dB]. Furthermore, it is evident that the proposed system ensures different gain depending on the total average bit-rate.

Table II shows the comparison of SQNR values obtained by using the proposed scheme without transformation coding included and performance of PCM system. The comparison is provided for various values of system parameters. By observing Table II, it can be seen that for $R = 6$ [bits/sample], obtained results show gain for 4, 8 and 16 representational levels (N_g), whereas for $N_g = 2$, PCM system provides better performance.

However, it should be noted that the achieved gain is much lower, unlike using transformation coding, and it is less than 0.3[dB]. Furthermore, it can be concluded that for $R = 7$ [bits/sample], PCM system ensures better performance for the all observed cases. Taking aforementioned discussion into account, we can conclude that the proposed transformation coding ensures much better performance, and that relative gain is higher than 4[dB]. Next, if we compare the best obtained performance ($R = 7$, $M = 240$, $N_g = 16$) with the log-PCM based on G. 711 [14], achieved gain is about 10[dB].

TABLE I. EXPERIMENTAL RESULTS OF APPLYING THE PROPOSED SCHEME AND COMPARISON WITH PCM FOR VARIOUS VALUES OF SYSTEM PARAMETERS ($\mu=16.9227$).

M	N_g	SQNR[dB]	R_{av} [bits/sample]	SQNR _{PCM} [dB]	SQNR[dB]	R_{av} [bits/sample]	SQNR _{PCM} [dB]	SQNR[dB]	R_{av} [bits/sample]	SQNR _{PCM} [dB]
40	2	27.9149	5.525	25.5040	33.8550	6.525	32.0016	38.3156	7.025	35.1832
	4	28.5620	5.55		34.4667	6.55		38.9715	7.05	
	8	28.8159	5.75		34.4614	6.575		39.0745	7.075	
	16	28.8591	5.6		34.4662	6.6		39.0891	7.1	
80	2	27.9252	5.5125		33.8304	6.5125		38.2624	7.0125	
	4	28.7203	5.525		34.4613	6.525		39.0860	7.025	
	8	28.7881	5.5375		34.4600	6.5375		39.0420	7.0375	
	16	28.8526	5.55		34.5109	6.55		39.2575	7.05	
160	2	27.5856	5.5062		33.5942	6.5062		38.0441	7.0062	
	4	28.7812	5.5125		34.5695	6.5125		39.2138	7.0125	
	8	28.8411	5.5187		34.4568	6.5187		39.2231	7.0187	
	16	28.8874	5.5250		34.5740	6.525		39.2815	7.025	
240	2	27.1206	5.5042		32.9279	6.5042		37.3456	7.0042	
	4	28.6600	5.5083		34.2717	6.5083		38.9784	7.0083	
	8	28.8097	5.5125		34.3880	6.5125		39.2063	7.0125	
	16	28.8875	5.5167		34.6034	6.5167		39.2283	7.0167	

TABLE II. SQNR FOR THE PROPOSED CODING SCHEME WITHOUT TRANSFORMATION CODING INCLUDED AND PCM FOR DIFFERENT VALUES OF THE TOTAL AVERAGE BIT-RATE ($\mu = 16.9227$).

M	N_g	SQNR[dB]	R_{av} [bits/sample]	SQNR _{PCM} [dB]	SQNR[dB]	R_{av} [bits/sample]	SQNR _{PCM} [dB]
40	2	28.3134	6.025	28.8200	33.9203	7.025	5.1832
	4	28.9524	6.05		34.8007	7.05	
	8	28.9529	6.075		34.8087	7.075	
	16	28.9780	6.1		34.7276	7.1	
80	2	28.3221	6.0125		33.9591	7.0125	
	4	29.0424	6.025		34.6495	7.025	
	8	29.0610	6.0375		34.6625	7.0375	
	16	29.0862	6.05		34.8167	7.05	
160	2	28.1932	6.0062		33.9213	7.0062	
	4	29.0077	6.0125		34.5579	7.0125	
	8	29.0561	6.0187		34.6691	7.0187	
	16	29.0920	6.025		34.8151	7.025	
240	2	27.8857	6.0042		33.7235	7.0042	
	4	29.0193	6.0083		34.3047	7.0083	
	8	29.0954	6.0125		34.6475	7.0125	
	16	29.1053	6.0167		34.9055	7.0167	

IV. CONCLUSIONS

In this paper, it has been proposed a new two-branch speech signal coding scheme, which implements a simple transform coding and forward adaptive quantization. The implemented transformations divide signal, making two independent streams with more predictable characteristics that are exploited for more accurate signal reconstruction. Forward adaptation is applied in both branches, providing additional signal compression of previously transformed input signal. The proposed system was tested for benchmark speech signal, recorded in Laboratory of Acoustics, Faculty of Electronic Engineering, University of Nis. The obtained results have shown that the proposed simple transform coding is very suitable for speech signal processing and it ensures gain even more than 4[dB], comparing to the same forward adaptive scheme, without transform coding. Furthermore, it has been shown that the proposed scheme has better performance than PCM system and that for all discussed system configurations achieves gain between 1.85 [dB] and 4.0983 [dB].

In the future, we will intend to research not only the natural speech signal, but the synthetic speech signal due to its' rising importance in pattern recognition systems. Different quantizers' design will be one of the most challenging tasks, due to different dynamics of synthetic speech signal. We can expect that design may include some simpler coding solutions, and possibly dual-mode quantization instead of forward adaptation will provide the same quality level, as the proposed scheme for natural speech signal in this paper.

REFERENCES

- [1] L. R. Rabiner, R. W. Schafer, "Introduction to digital speech processing", *Foundations and Trends in Signal Processing*, vol. 1, pp. 1–194, 2007. [Online]. Available: <http://dx.doi.org/10.1561/20000000001>
- [2] A. Anskaitis, A. Kajackas, "The tool for quality estimation of short voice segments", *Elektronika ir Elektrotechnika*, no. 8, pp. 103–108, 2010.
- [3] C. Valentini-Botinhao, M. Toman, M. Pucher, D. Schabus, J. Yamagishi, "Intelligibility of time-compressed synthetic speech: Compression method and speaking style", *Speech Communication*, vol. 74, pp. 52–64, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.specom.2015.09.002>
- [4] J. E. Garcia, A. Ortega, A. Miguel, E. Lleida, "Low bit rate compression methods of feature vectors for distributed speech recognition", *Speech Communication*, vol. 58, pp. 111–123, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.specom.2013.11.007>
- [5] S. Na, "On the support of fixed-rate minimum mean-squared error scalar quantizers for a Laplacian source", *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 937–944, 2004. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2004.826686>
- [6] S. Na, D. L. Neuhoff, "On the support of MSE-optimal, fixed-rate, scalar quantizers", *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2972–2982, 2001. [Online]. Available: <http://dx.doi.org/10.1109/18.959274>
- [7] N. S. Jayant, P. Noll, *Digital Coding Of Waveforms, Principles and Applications to Speech and Video*. New Jersey: Prentice Hall, 1984, ch. 4, pp. 115–220.
- [8] K. Sayood, *Introduction to data compression*. 2005.
- [9] Z. Peric, D. Aleksic, M. Stefanovic, J. Nikolic, "New approach to support region determination of the μ -law quantizer", *Elektronika Ir Elektrotechnika*, vol. 19, no. 8, pp. 111–114, 2013. [Online]. Available: <http://dx.doi.org/10.5755/j01.eee.19.8.5406>
- [10] J. Nikolic, Z. Peric, "Lloyd-Max's algorithm implementation in speech coding algorithm based on forward adaptive technique", *Informatica*, vol. 19, no. 2, pp. 255–270, 2008.
- [11] G. Petkovic, Z. Peric, L. Stoimenov, "Switched scalar optimal μ -law quantization with adaptation performed on both the variance and the distribution of speech signal", *Elektronika Ir Elektrotechnika*, vol. 22, no. 1, pp. 64–67, 2016. [Online]. Available: <http://dx.doi.org/10.5755/j01.eee.22.1.14111>
- [12] J. Nikolic, Z. Peric, D. Aleksic, "Optimization of μ -law companding quantizer for Laplacian source using Muller's method", *Przeglad Elektrotehniczny*, vol. 89, no. 3A, pp. 206–208, 2013.
- [13] D. Aleksic, Z. Peric, J. Nikolic, "Support region determination of quasi-logarithmic quantizer for Laplacian source", *Przeglad Elektrotehniczny*, vol. 88, no. 7A, pp. 130–132, 2012.
- [14] Z. Peric, M. Tancic, S. Tomic, D. Ciric, "Subband coding of audio signal with logarithmic companders", in *Proc. Telsiks 2015*, Serbia, Nis, 2015, pp. 19–22. [Online]. Available: <http://dx.doi.org/10.1109/telsks.2015.7357728>