Evaluation of Measurement Performance in Averaging Quantization System with Noise

Miroslav KAMENSKÝ, Karol KOVÁČ, Eva KRÁLIKOVÁ, Anton KRAMMER

Dept. of Measurement, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovak Republic

miroslav.kamensky@stuba.sk, karol.kovac@stuba.sk, eva.kralikova@stuba.sk, anton.krammer@stuba.sk

Abstract. *Statistical description of quantization process is common in the theory of quantization. For the case of nonsubtractive dither theoretical analyses of the dithered quantizer have been confronted with experimental results. As a quantization system one-chip microcomputer with the analog-to-digital converter on a chip has been used. Generally valid criteria for dithered system performance have been practically applied for Gaussian dither. Interaction of natural noise present in the signal with an added Gaussian noise of several different disperses and influence of differential nonlinearity of the converter has been observed.*

Keywords

Quantization, nonsubtractive dither, averaging, Gaussian noise.

1. Problem Description and Notification

Striving for ever higher precisions in measurement technology naturally leads to methods, which can shift the resolution limit of the quantizer usually represented by the analog-to-digital converter (ADC) bellow the lowest significant bit (LSB). Dither is a random noise added to a signal prior to its (re)quantization in order to control the statistical properties of the quantization error [1]. There is also a possibility of addition of the deterministic signal for similar purpose in [2] called deterministic dither. But the statistical theory of quantization better suits for stochastic type of the added signal. Then two dither types are distinguished. The term subtractive dither (SD) is used for the case, when the dither is subsequently subtracted from the output signal after quantization. Unfortunately SD is difficult to use in many practical systems because the dither signal must be available at both ends of the channel and furthermore, any digital processing of the dithered signal would necessitate processing of dither prior to subtraction [1]. Nonsubtractive dither (ND) is the second dither type and is not subtracted from the output. As shown in [1], ND in contrast with SD cannot render the total error statistically independent of the input neither it can make temporally separated values of the total error statistically independent of one another. These could be critical drawback for digital audio or video applications. In measurement applications the output square error is usually of interest.

1.1 Notifications

Schematic of used nonsubtractive dithered system is shown in Fig.1. The quantization system input is denoted

$$w = s + d \tag{1}$$

where s is measured value and d is added noise. If there is null dither, the output of the ADC is

$$\beta = Q(s)\,. \tag{2}$$

According to the Fig.1 $Q(.)$ means static transfer characteristic of the quantizer. The error function Q_e determines behavior of error e (error if null dither is applied)

$$e = Q(s) - s = Q_e(s)\,. \tag{3}$$

For general ND the quantizer output is

$$v = Q(w) \tag{4}$$

and the error after quantization is

$$\varepsilon = Q(s+d) - s = Q_e(s+d) + d\,. \tag{5}$$

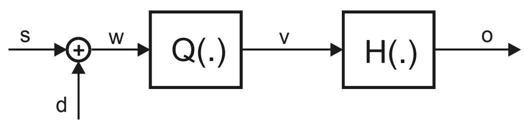

Fig. 1. The nonsubtractive dithered system with averaged output.

The total error ξ will be evaluated for static input values s (therefore $v=v|s$, $o=o|s$) and for averaged quantizer output denoted o

$$\xi = H(v|s) - s = o|s - s \tag{6}$$

where $H(.)$ is low passed filter implemented (in this case) through averaging

$$H(x) = \frac{1}{N}\sum_{i=1}^{N} x_i\,. \tag{7}$$

1.2 Ouantizer

In the paper mid-tread quantizer [1] is considered (Fig.2). When the input is confined in no-overload region the transfer characteristic can be analytically expressed by equation

$$Q(w)=\left\lfloor \frac{w+\frac{q}{2}}{q} \right\rfloor q \qquad (8)$$

and the corresponding error characteristic is

$$Q_e(w)=Q(w)-w=\frac{q}{2}-q\left\langle \frac{w+\frac{q}{2}}{q} \right\rangle \qquad (9)$$

where the „floor“ operator $\lfloor\ \rfloor$ returns the greatest integer less than or equal to its argument and $\langle\ \rangle$ means the fractional part, e.g. $\langle a\rangle = a - \lfloor a\rfloor$ [3].

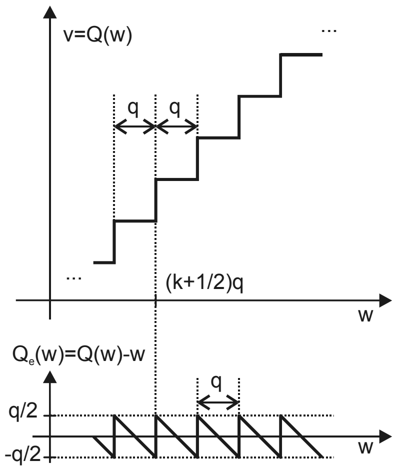

Fig. 2. Characteristic of mid-tread quantizer or of an ideal ADC.

According to theorem of equivalent nonlinearity for system with ND [4] if the conditional mean value is of interest, the quantizer transfer (static) characteristic could be replaced by its convolution with probability density function (PDF) of dither. Therefore ND could be used for correction of nonlinerity errors – i.e. of quantization error or differential nonlinearity (DNL).

1.3 Statistical Description

For statistical description of a general signal x the PDF $f_x(x)$ will be used, e.g. for the dither it is $f_d(d)$. The characteristic function (CF) is the Fourier transformation of PDF in the form [5], [6]

$$\Phi_x(u)=\int_{-\infty}^{\infty} f_x(x)e^{jux}\,dx\,. \qquad (10)$$

Then the back transformation should be

$$f_x(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_x(u)e^{-jux}\,du\,. \qquad (11)$$

Profitably the Fourier transform of multiplied signals could be used because it leads to a simple relation

$$f_1(x)*f_2(x)\leftrightarrow\Phi_1(u)\Phi_2(u)\,. \qquad (12)$$

It is possible to evaluate moments of the signal from CF

$$E\left[x^k\right]=\frac{1}{j^k}\left.\frac{d^k\Phi_x(u)}{du^k}\right|_{u=0}. \qquad (13)$$

2. Theory of Nonsubtractive Dither

2.1 Area Sampling

Fig.3 is a sketch of a typical quantizer input and output. The input PDF $f_w(w)$ is smooth, and the output PDF $f_v(v)$ is discrete, because each input value is rounded towards the nearest allowable discrete level. The probability of each discrete output level equals to the probability of the input signal occurring within the associated quantum band.

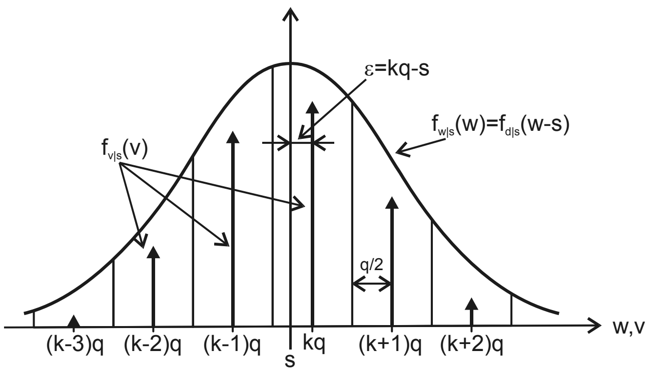

Fig. 3. Area sampling - in the output only integer (k is integer) multiples of quantization step q are possible.

The PDF's of input and output signals are related to each other through a special type of sampling called area sampling [7]. Cutting up the input PDF into strips as in Fig.3., the area of each strip is compressed into an impulse in the center of the strip when forming the output PDF. The output PDF is a string of Dirac delta functions, whose areas correspond to areas under the input PDF within the bounds of each quantum box.[7].

In [3] or [1] the process is described mathematically if ND is used. If the input value is in the range of $\{(k - 0.5)q \le w \le (k + 0.5)q\}$, the output is $v=s+kq$. For generalization of this result for dither signal the following set is defined

$$S_{k,d|s}=\left\{\left(k-\frac{1}{2}\right)q-s\le d\le\left(k+\frac{1}{2}\right)q-s\right\} \qquad (14)$$

which is set of dither values d for constant signal s leading to the same output v. The probability, that there is such a dither is

$$P_{d|s}\left(S_{k,d|s}\right)=\int_{S_{k,d|s}} f_{d|s}(d)dd=\int_{-\frac{q}{2}+kq}^{\frac{q}{2}+kq} f_{d|s}(w-s)dw\,. \tag{15}$$

For given s the error value could achieve only discrete levels $\varepsilon=kq-s$ [1]. Therefore the conditional PDF (CPDF) of error is series of delta functions separated by quantization step q and weighted by proper probability of occurrence

$$f_{\varepsilon|s}(\varepsilon)=\sum_{k=-\infty}^{\infty}\delta[\varepsilon-(kq-s)]P_{d|s}\left(S_{k,d|s}\right). \tag{16}$$

It could be shown that [6]

$$f_q(x)*f_{d|s}(x)=\int_{x-\frac{q}{2}}^{x+\frac{q}{2}}\frac{1}{q}f_{d|s}(\alpha)d\alpha \tag{17}$$

where

$$f_q(x)=\begin{cases}\dfrac{1}{q} & -\dfrac{q}{2}<x<\dfrac{q}{2}\\ 0 & else\end{cases} \tag{18}$$

then

$$f_{\varepsilon|s}(\varepsilon)=\left[f_q(\varepsilon)*f_{d|s}(\varepsilon)\right]\sum_{k=-\infty}^{\infty}\delta[\varepsilon-(kq-s)]. \tag{19}$$

Occurrence of PDF of uniform noise f_q in convolution indicates possible modeling of quantization trough addition of such a uniform noise, because PDF of sum of independent noises could be evaluated through convolution of PDF's of summands. This convolution is in (19) multiplied by uniform impulse train, in other words the result of convolution is conventionally sampled.

2.2 Conditional Mean Error

Taking the transform of (19) one can get the CF of error CPDF. The desired moment - mean value - could be found employing (13). According to (12) the image of convolution in PDF domain is multiplication in CF domain. Then product of Φ_q and $\Phi_{d|s}$ should occur in the final CF. For CF of a uniform noise it could be written

$$\Phi_q=\mathrm{sinc}\left(\frac{qu}{2}\right). \tag{20}$$

The summation from (19) should be found also in CF domain. Similarly to discrete Fourier transform for sampling in time, the CF should be periodic and infinite sum of replicas (the member $e^{j\psi s}$ in (20) is caused by shifting of dither PDF with measured signal s) and from [6] it is

$$\Phi_{\varepsilon|s}(u)=\sum_{k=-\infty}^{\infty}\mathrm{sinc}\left(\frac{q(u+k\Psi)}{2}\right)\Phi_{d|s}(u+k\Psi)e^{jk\Psi s} \tag{21}$$

what suggest possibility of aliasing-like effect. The variable

$$\Psi=\frac{2\pi}{q} \tag{22}$$

can be thought of as "quantization radian frequency". Mean value is the first moment and with respect to (13) [1] it is

$$m_{\varepsilon|s}=E[\varepsilon|s]=\sum_{k=1}^{\infty}\frac{q(-1)^k}{\pi k}\mathrm{Im}\left[\Phi_{d|s}(k\Psi)e^{ik\Psi s}\right]. \tag{23}$$

3. Quantitative Criteria

In measurement applications, the output squared error is usually of interest to describe the performance of the overall system. Written in the conditional form, it is

$$E\left[\xi^2|s\right]=VAR[\xi|s]+E^2[\xi|s] \tag{24}$$

where the variance is known as

$$VAR[\xi|s]=E\left[(\xi-E[\xi|s])^2|s\right]. \tag{25}$$

The mean value could be introduced also for sampled data in following form

$$E[x]=\lim_{L\to\infty}\frac{1}{L}\sum_{i=1}^{L}x_i\,. \tag{26}$$

Notations

$$\mu^2=E\left[\xi^2|s\right], \tag{27}$$

$$\sigma^2_{\xi|s}=VAR[\xi|s], \tag{28}$$

$$m_{\xi|s}=E[\xi|s]=E[\varepsilon|s] \tag{29}$$

help to simply suggest dependency of observed final parameters from parameters of measurement system. Error is sum of the disperse of error and squared mean error

$$\mu^2=\sigma^2_{\xi|s}+m^2_{\xi|s}\,.$$

For ND if averaged

$$\mu^2(s,\sigma_d,N)=\frac{1}{N}\sigma^2_{\varepsilon|s}(s,\sigma_d)+m^2_{\varepsilon|s}(s,\sigma_d) \tag{30}$$

where N is the number of values to be averaged. The total error depends on the type of dither signal, on its disperse (variance), on the number of averaged values N and on the measured signal s (dependency from s could be easy replaced by dependency from e). Because occurrence of measured value s in general measurement is unknown, it is better to eliminate s [8] by the expression

$$\mu_{\rm a}^2 = \frac{1}{q} \int_{-\frac{q}{2}}^{\frac{q}{2}} \mu^2(s) ds \,. \tag{31}$$

According to [9] this mean-square error could be evaluated from

$$\mu_{\rm a}^2(\sigma_d, N) = \frac{\frac{q}{12} + \sigma_d^2}{N} + \left(1 - \frac{1}{N}\right)\mu_{\rm a}^2(\sigma_d, \infty) \tag{32}$$

where $\mu_{\rm a}^2(\sigma_d, \infty) = \lim_{N\to\infty} \mu_{\rm a}^2(\sigma_d, N)$ and it could be shown that (* means complex complement)

$$\mu_{\rm a}^2(\sigma_d, \infty) = \frac{q^2}{2\pi^2} \sum_{k=1}^{\infty} \frac{\Phi_{d|s}(k\Psi)\Phi_{d|s}^*(k\Psi)}{k^2} \,. \tag{33}$$

3.1 Gaussian Noise

Gaussian noise - in the role of dither - exhibits the following PDF [3]

$$f_{d|s}(d) = \frac{1}{\sqrt{2\pi}\sigma_d} e^{-\frac{d^2}{2\sigma_d^2}} \tag{34}$$

and its CF is

$$\Phi_{d|s}(u) = e^{-\frac{u^2\sigma_d^2}{2}} \,. \tag{35}$$

The conditional mean value (23) for this type of dither is then

$$m_{\varepsilon|s} = q \sum_{k=1}^{\infty} \frac{(-1)^k}{\pi k} \exp\left[-2\pi^2 k^2 \left(\frac{\sigma_d}{q}\right)^2\right] \sin\left(\frac{2\pi k s}{q}\right) . \tag{36}$$

For big enough standard deviation of the noise $\sigma_d \geq 0.3$ it is sufficient to keep only the first term in the series expansion. The mean-square error is then [9]

$$\mu_{\rm a}^2(\sigma_d, \infty) \cong \frac{q^2}{2\pi^2} e^{-4\pi^2\left(\frac{\sigma_d}{q}\right)^2} \,. \tag{37}$$

For given N it achieves its optimal value for disperse

$$\sigma_{d,\rm opt}(N) = \frac{q}{2\pi}\sqrt{\ln[2(N-1)]} \,. \tag{38}$$

4. Experiments

4.1 Workplace

The block diagram of the realized workplace is depicted in Fig 4. A standard microcontroller with a 10-bit ADC has been selected as an object of experiment. Input voltage was adjusted by a precise data acquisition (DAQ) card while dither was added in PC software. Averaged output numbers were sent from microcontroller via serial link RS232 to the PC and then displayed.

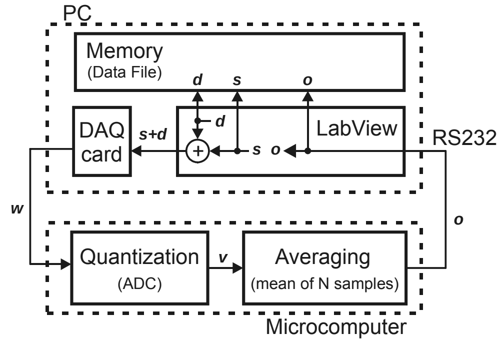

Fig. 4. Block scheme of the workplace.

In the analyses the quantization and other errors on the PC side has been neglected.

4.2 Experimental Results

In experiments average of $N=16$ 10-bit values were calculated in the microcomputer for each single input (i.e. quantizer output)

$$o = H(v) = \frac{1}{16} \sum_{i=1}^{16} v_i \tag{39}$$

which was then sent to the output of the microcomputer as a 14-bit number. Resolution of DAQ card in the input range of ADC (0÷5V) is 14-bit. Therefore quantization errors in other part of channel has been 16-times smaller than in the tested quantizer and so neglected. Measurements were held in M static quantizer input levels repeated P times for every level.

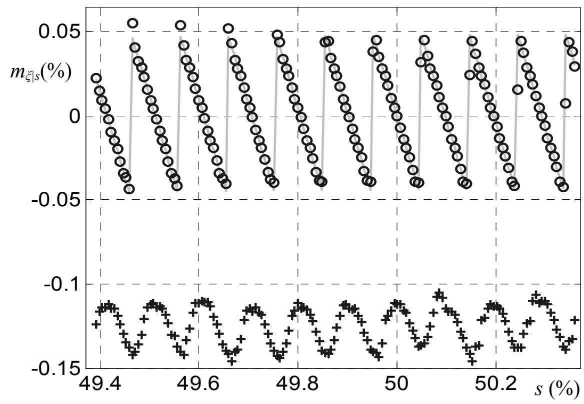

Fig. 5. Conditional mean error of measurement evaluated for the range of 10 quantizer steps – approximately 16 points in each of them – as a mean value from P=20 processes. (+ – experiment, o – theory, gray line – simulation)

In the Fig.5 conditional mean value of the error as a function of measured s is depicted. The + marker signifies this value estimated from measurements as

$$m_{\xi|s} \doteq \frac{1}{P} \sum_{i=1}^{P} o_i - s \,. \tag{40}$$

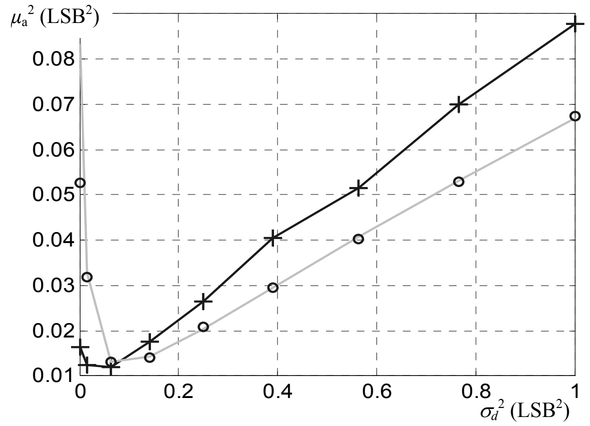

Fig. 6. Mean-square error averaged for quantization step μ_a^2 evaluated for different disperses of added Gaussian noise. (+ – experiment, o – theory, gray line – simulation).

Parameters of the first experiment (Fig.5, 6.) were P=20 (averaged) outputs in each of M=160 input levels with uniform step approximately 1/16 LSB of the tested ADC – said approximately because LSB of the ADC is not a precise integer multiple of LSB of this DAQ card. The grey line is based on simulation results, i.e. no integral nonlinearity (INL) and no noise is present. The **o** markers represent theoretical values according to (29) resp. (36) where 20 summands were used. The offset between the simulated or theoretical values and the measurement results is caused by offset or INL of the ADC. Local INL changes should have been seen from changing shift (phase-like) of the measured curve towards the theoretical waveform. The reason for difference in shape of waveforms is natural noise present in real signal while simulating with no noise.

Measurements similar to Fig.5 were acquired after adding Gaussian noise to the signal s. For each disperse of added noise σ_d the mean-square error as the mean value for quantization step (31) was estimated as

$$\mu_a^2 \doteq \frac{1}{M}\sum_{j=1}^{M}\frac{1}{P}\sum_{i=1}^{P}\left(o_{i,j} - s_j - offset\right)^2 \tag{41}$$

where *offset* means average offset of measurement outputs towards ideal outputs, i.e. mean error caused by other imperfections of ADC than quantization had to be subtracted. The results are compared to theoretical ones in Fig.6. The + markers and solid line belongs to μ_a^2 from measurements, the grey line is obtained also with (41) but from simulations, the **o** marks theoretical values according to (32) with (37).

In the next case the aim was to show influence of INL. Therefore M=160 levels were spread through a wider input range of 80 quantizer steps. The conditional mean value of the error $m_{\xi|s}$ as a function of measured value s is shown in Fig.7. For every quantization step they are now approximately two input levels. The nonlinearity is apparent.

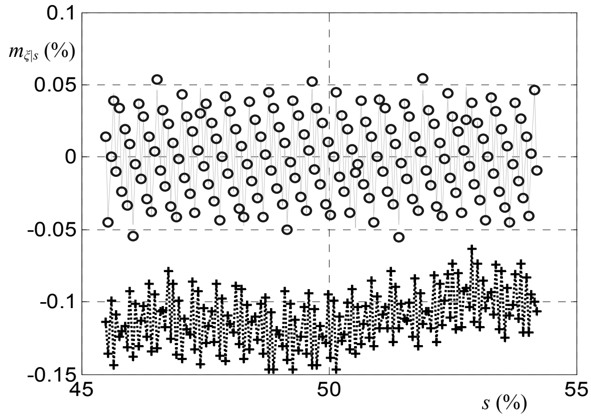

Fig. 7. Conditional mean error of measurement evaluated for the range of 80 quantizer steps – approximately 2 points in one step – as a mean value from P=20 processes. (+ – experiment, o – theory, gray line – simulation).

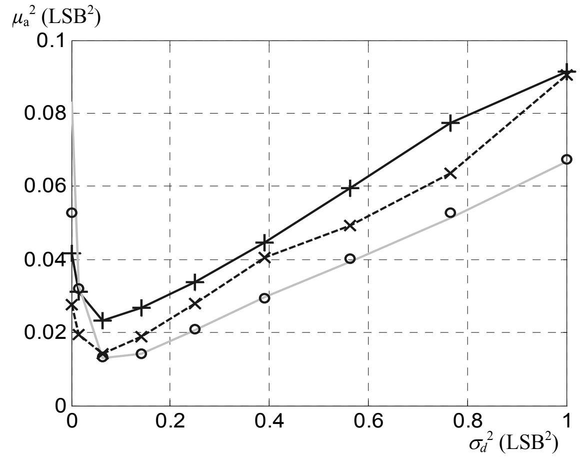

Fig. 8. Mean-squared error averaged for quantization step μ_a^2 evaluated for different disperses of added Gaussian noise. (+ – experiment, o – theory, gray line – simulation). The added **x**-line responds to the reduced number of input levels to the first M=16.

Fig.8 shows resulting theoretical, simulation and experimental values of μ_a^2 similarly to the first case from Fig. 6. But in addition there is also included a line with **x** markers corresponding to the reduced number of levels M=16. These are the first 16 numbers from the second experiment thus covering shorter input range. The less the range the less influence of DNL is expected. This assumption is reflected in graphs in Fig.8. As it could be seen, the DNL shifted up the line of μ_a^2.

4.3 Discussion

The simulations showed proper theoretical description of considered performance parameters for ideal quantizer and signal behavior. In the experiments the natural noise present in signal shifted the curve of mean squared error averaged for a quantization step μ_a^2 in the way, that less dither disperse should be added to achieve minimal μ_a^2. According to (38) this optimal disperse is

$\sigma_{d,\ opt}^{2}$=0.0862 LSB^2. In experimental results the optimum is shifted to lower disperse values. But if PDF of the natural noise is known it could be easily involved into the dither. From other point of view disperse of common Gaussian natural noise could be estimated from the shift of theoretical optimum for disperse of added noise.

In real measurements other error sources than quantization influence result. DNL should negatively change μ_a^2. Its impact rises for wider ranges on the input side of ADC. Shift of μ_a^2 to upper values is apparent for wider input range (s). Similarly for rising disperse of dither the DNL changes come into the role. Further investigations of interactions between dither and DNL seem to be challenging. It is possible to suppress local DNL through dithering. But special correction methods are available for suppression of nonlinearities and dither could be optimized for correction of quantization error.

Acknowledgements

The work presented in this paper was supported by the Slovak Ministry of Education under grant No. 2003SP200280802 and by the Slovak Grant Agency VEGA under grant No. 1/3101/06.

References

[1] WANNAMAKER, R. A., LIPSHITZ, S. P., VANDERKOOY, J., WRIGHT, J. N. A theory of nonsubtractive dither. *IEEE Transactions on Signal Processing*, February 2000, vol.48, no.2.

[2] CARBONE, P., PETRI, D Performance of stochastic and deterministic dithered quantizers. *IEEE Transactions on Instrumentation and Measurement*, April 2000, vol.49, no.2.

[3] CARBONE, P., PETRI, D. Effect of additive dither on the resolution of ideal quantizers. *IEEE Transactions on Instrumentation and Measurement*, June 1994, vol.43, no.3.

[4] ĎAĎO, S., VEDRAL, J. *Digital Measurement. Devices and Methods.* Prague: ČVUT, 2002. Lecture notes (in Czech), 232 pp., ISBN 80-01-02492-X.

[5] LAMOŠ, F., POTOCKÝ, R. *Probability and Mathematical Statistics. Statistical Analyses*. Bratislava: Alfa, 1989. (In Slovak.)

[6] WIDROW, B., KOLLÁR, I. *Quantization Noise in Digital Computation, Signal Processing, and Control*. Home page of the book http://www.mit.bme.hu/books/quantization/, prepared for publication by the fall of 2007.

[7] WIDROW, B., KOLLÁR, I. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, April 1996, vol.45, no.2.

[8] CARBONE, P. Quantitative Criteria for Design of Dither-Based Quantizing Systems. *IEEE Transactions on Instrumentation and Measurement*, Vol.46, No.3, June 1997.

[9] SKARTLIEN, R., ØYEHAUG, L. Quantization error and resolution in ensemble averaged data with noise. *IEEE Transactions on Instrumentation and Measurement*, June 2005, vol.54, no.3.

About Authors...

Miroslav KAMENSKÝ (Ing.) was born in Bratislava in 1977. He graduated in automation from the Faculty of Electrical Engineering and Information Technology of the Slovak University of Technology in Bratislava in 2002. His research interests include sensor techniques with emphasis on signal processing.

Karol KOVÁČ (Assoc. Prof., Ing., PhD.) was born in Bratislava, Slovakia on June 9, 1952. He received the MSc. degree with honors in 1976 and the PhD degree in electrical engineering from the Faculty of Electrical Engineering of the Slovak Technical University, Bratislava. Since 1976 he has been with the Department of Measurement of the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, now as Associate professor for Electric Measurement. His research interests are in the area of digital signal processing for measurement purposes.

Eva Králiková (Ing.) was born in Bratislava, Slovakia on February 23, 1955. She received the Ing. (MSc) degree with honors in 1979 in electrical engineering from the Faculty of Electrical Engineering of the Slovak Technical University, Bratislava. Since 1988 he has been with the Department of Measurement of the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology.

Anton KRAMMER (Ing.) was born in Levice in 1979. He studied electronics at the Faculty of Electrical Engineering and Information Technology of the Slovak University of Technology and graduated in 2004. Now he works as a researcher at the Department of Measurement especially in laboratory of electro-magnetic compatibility.