Predictive coding of media signals

Prof.dr.sc. Davor Petrinović

Examples of predictive speech coding

prof.dr.sc. Davor Petrinović

Predictive speech coding

- Speech signal encoder with Vocoder structure rejects actual information about the excitation signal of the vocal tract model
  – a synthetic excitation only "roughly" mimics real (natural) vocal tract excitation.
- It is logical to expect that by encoding information about the actual excitation, better quality, i.e., more natural reconstructed speech, could be achieved.
- This can be achieved:
  – by encoding an open loop prediction error, or
  – using prediction within a closed loop.

Predictive speech coding

- Although similar structures are used, the behavior of the encoder for these two predictive coding procedures is quite different, so it is interesting to investigate it in more detail.
- The procedure will be illustrated on vowels,
  - ... these are long-lasting voices with slowly changing correlations, so the same predictor can be used for longer signal segments (approx. 0.5 seconds)
- Predictive coding procedures will be compared with direct quantization of the speech signal with the same output index entropy.

Predictive speech coding

- An entropy constrained scalar quantizer (ECSQ) will be used for quantization in both predictive procedures, as well as for direct quantization.

- The quantizer step size will be adjusted so that the same (preselected) output entropy is achieved for each of the coding structures for objective comparison.

Predictive speech coding

- Quality assessment of different speech coding structures will include:
  - objectively measured SQNR relationships, and
  - subjective assessment by listening to signals at individual points of the coding structure.
- All characteristic signals of the structures will be graphically displayed in the time and frequency domain.

An example of open-loop predictive coding of speech

prof.dr.sc. Davor Petrinović

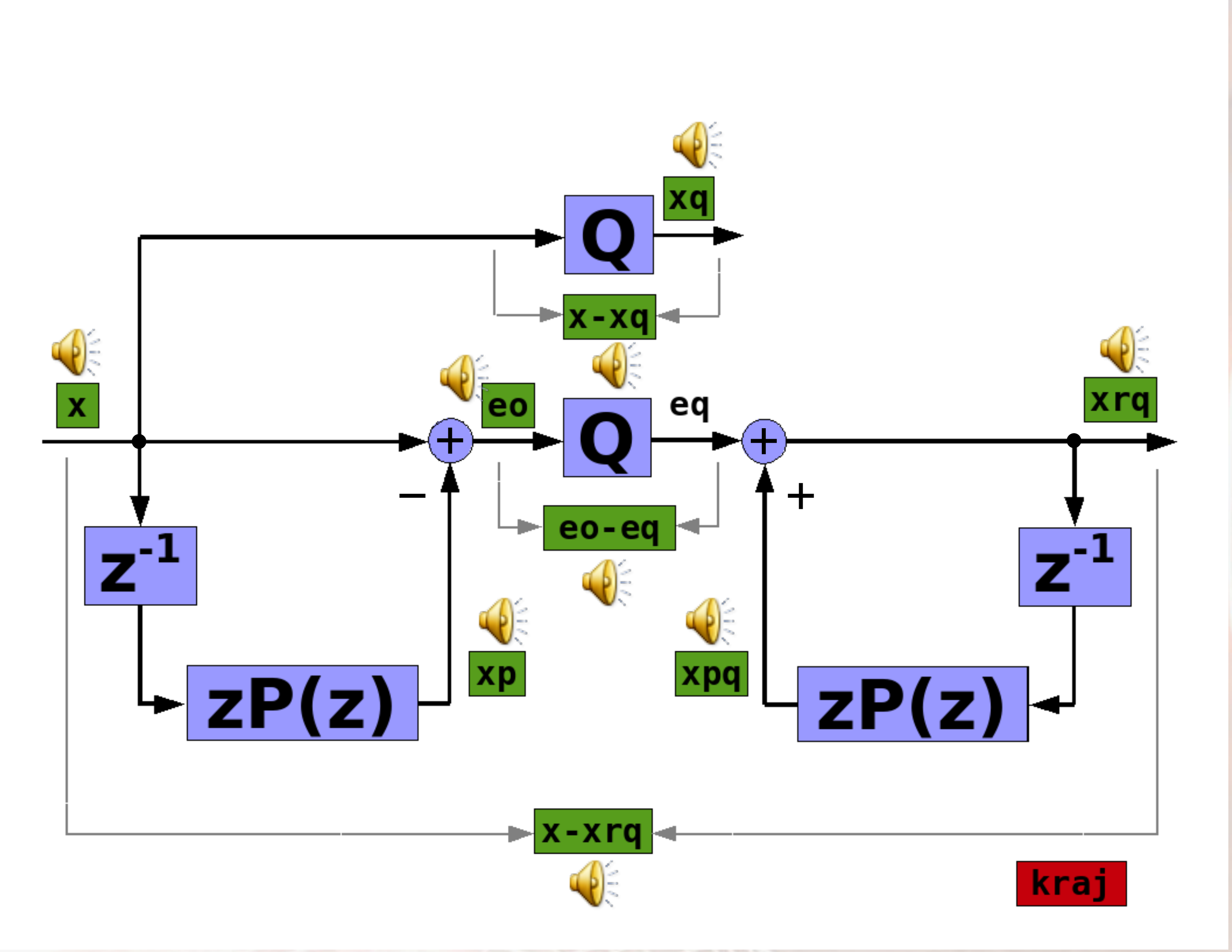
Open-loop predictive speech coding

- Simulation of open-loop speech coding is performed using the program:
  - **MT04_OLpred_govor.m**
- The selection of vowels, predictor order, and desired output entropy is done by menus at the beginning of the program execution.
- The program implements the following procedures:
  - extracts the selected vowel,
  - determines the optimal linear predictor for that voice,
  - determines the prediction error signal using an open loop predictor, ...

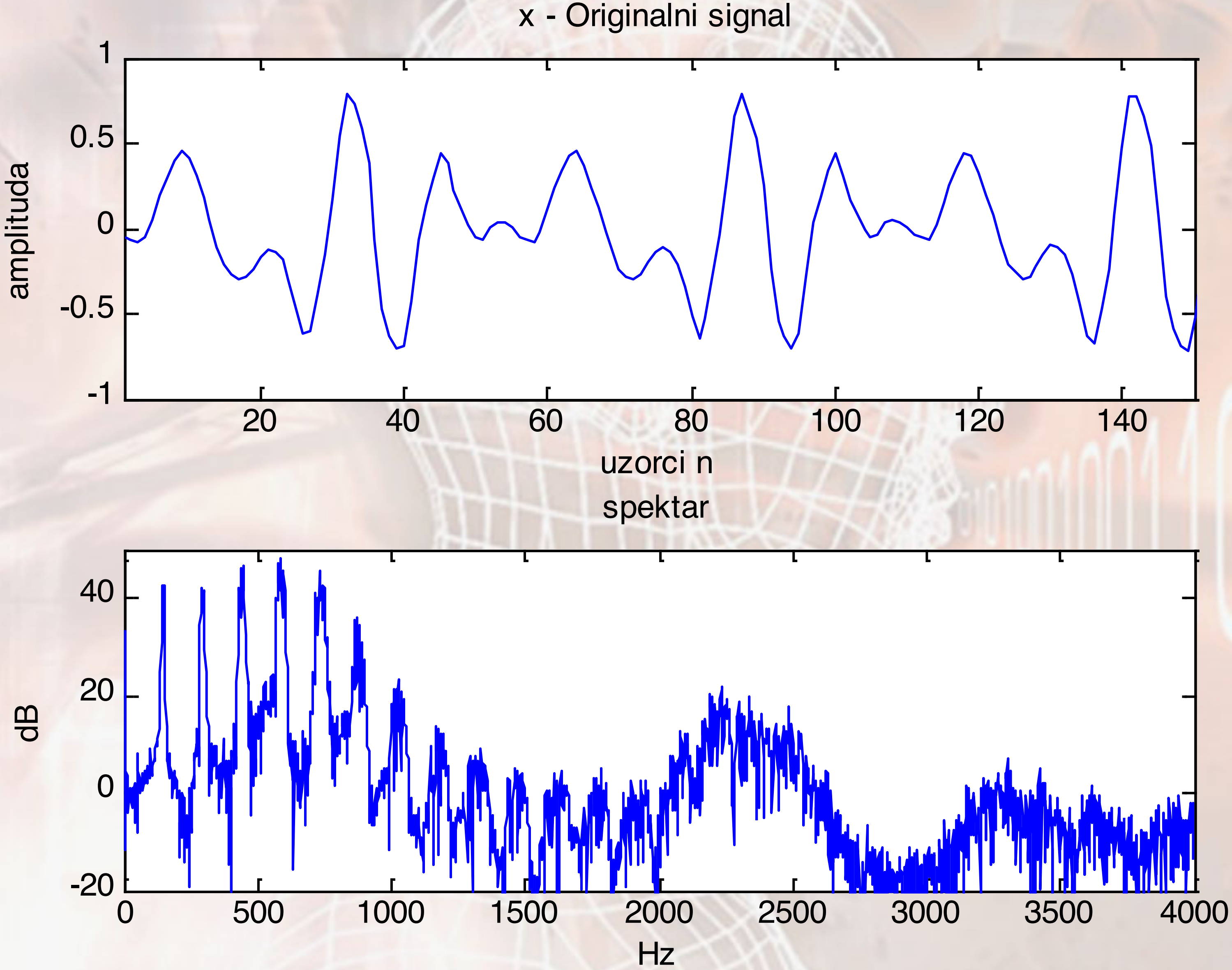
Open-loop predictive speech coding

- ... continued:
  - determines the differential entropies of the input signal and the prediction error signal,
  - determines the quantization steps for both signals that achieve the same desired output entropy,
  - performs scalar quantization with constrained entropy on both signals,
  - quantized prediction error signal is then passed through the reconstruction system,
  - calculates the errors of direct and predictive quantization and the corresponding SQNR relations, and
  - performs display of all signals.

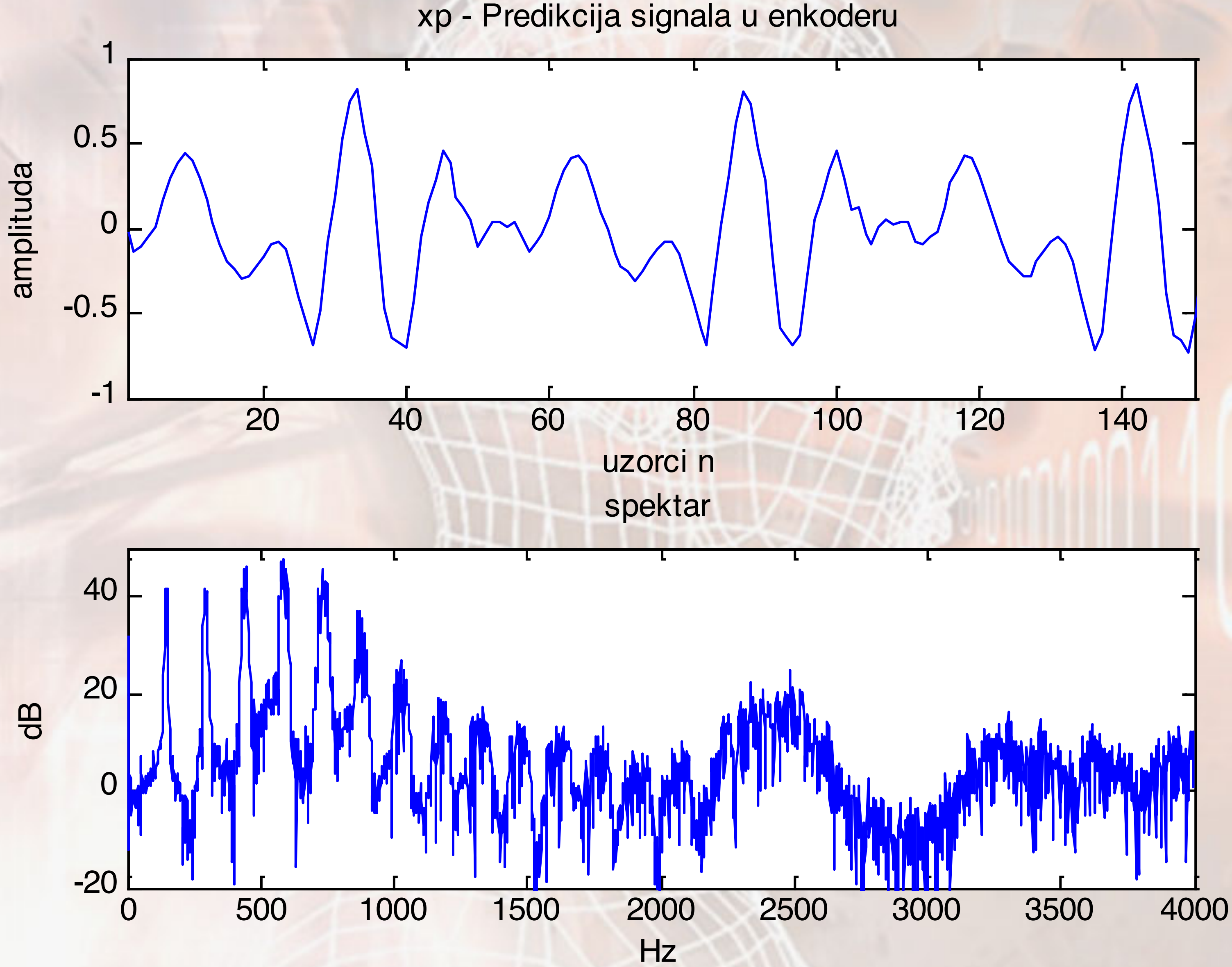
Open-loop predictive speech coding for vowel “o” p=10, $H(I)$=2 bit



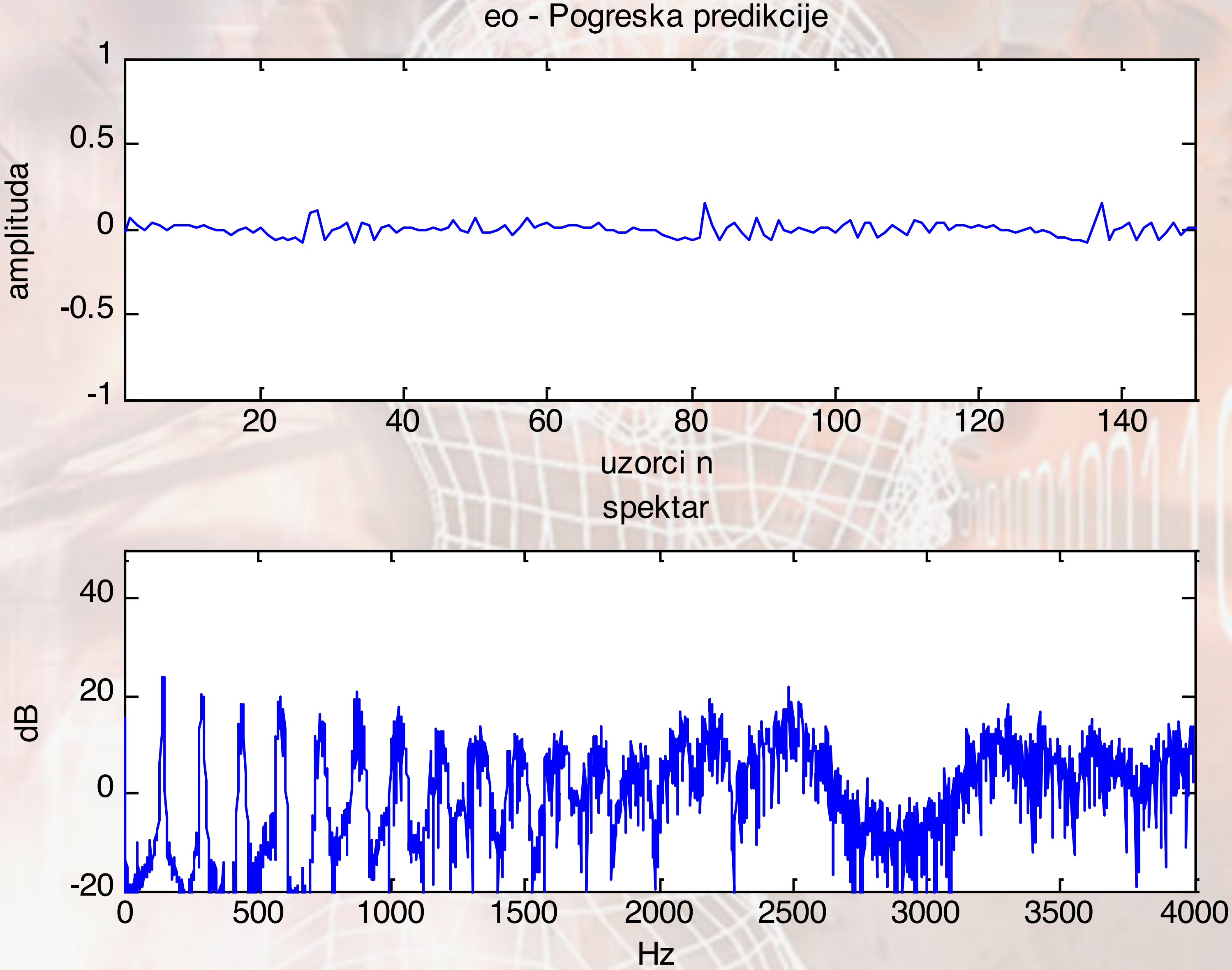
Input speech signal
vowel “o”



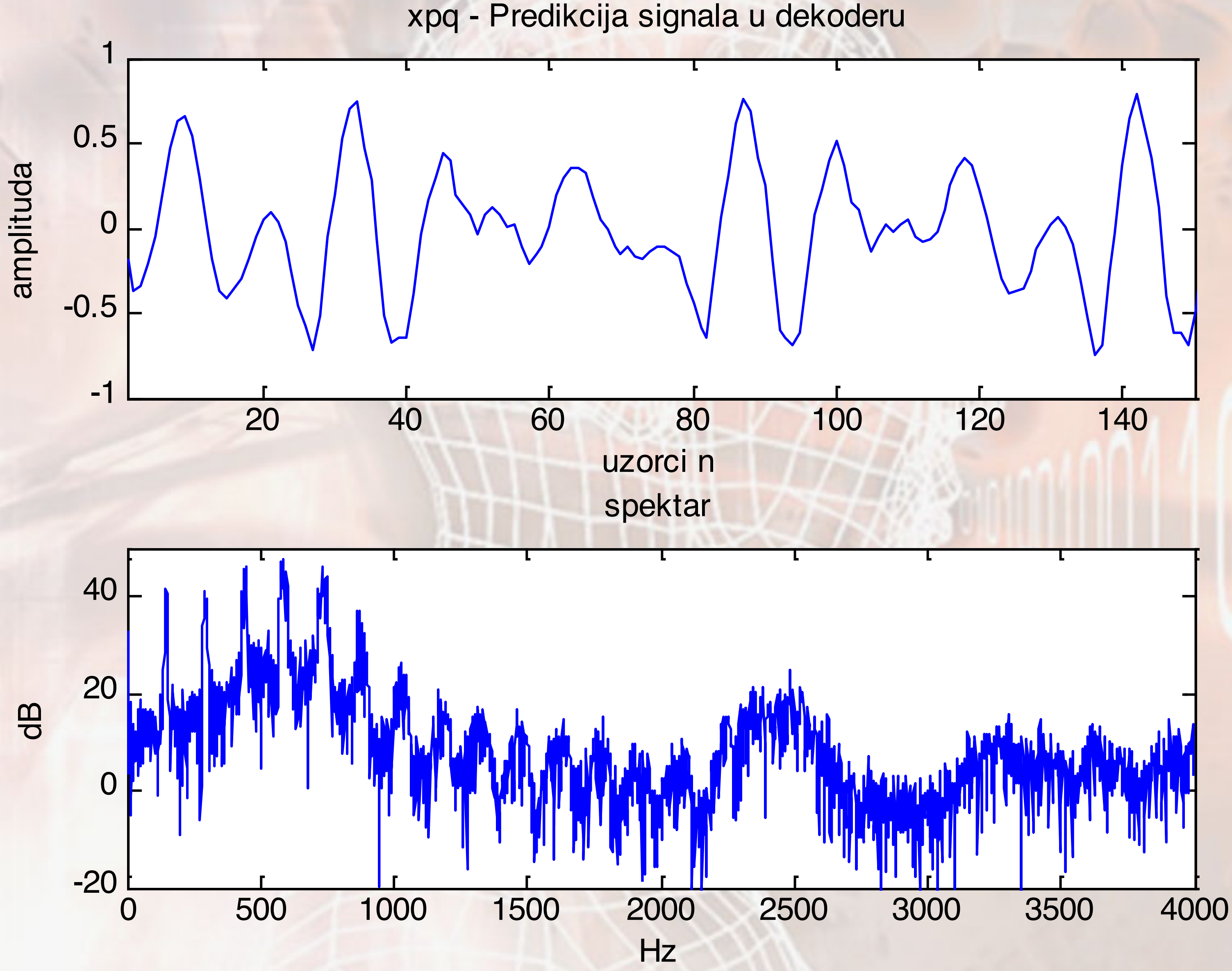
Prediction on the encoder side (open loop) vowel “o”, p=10



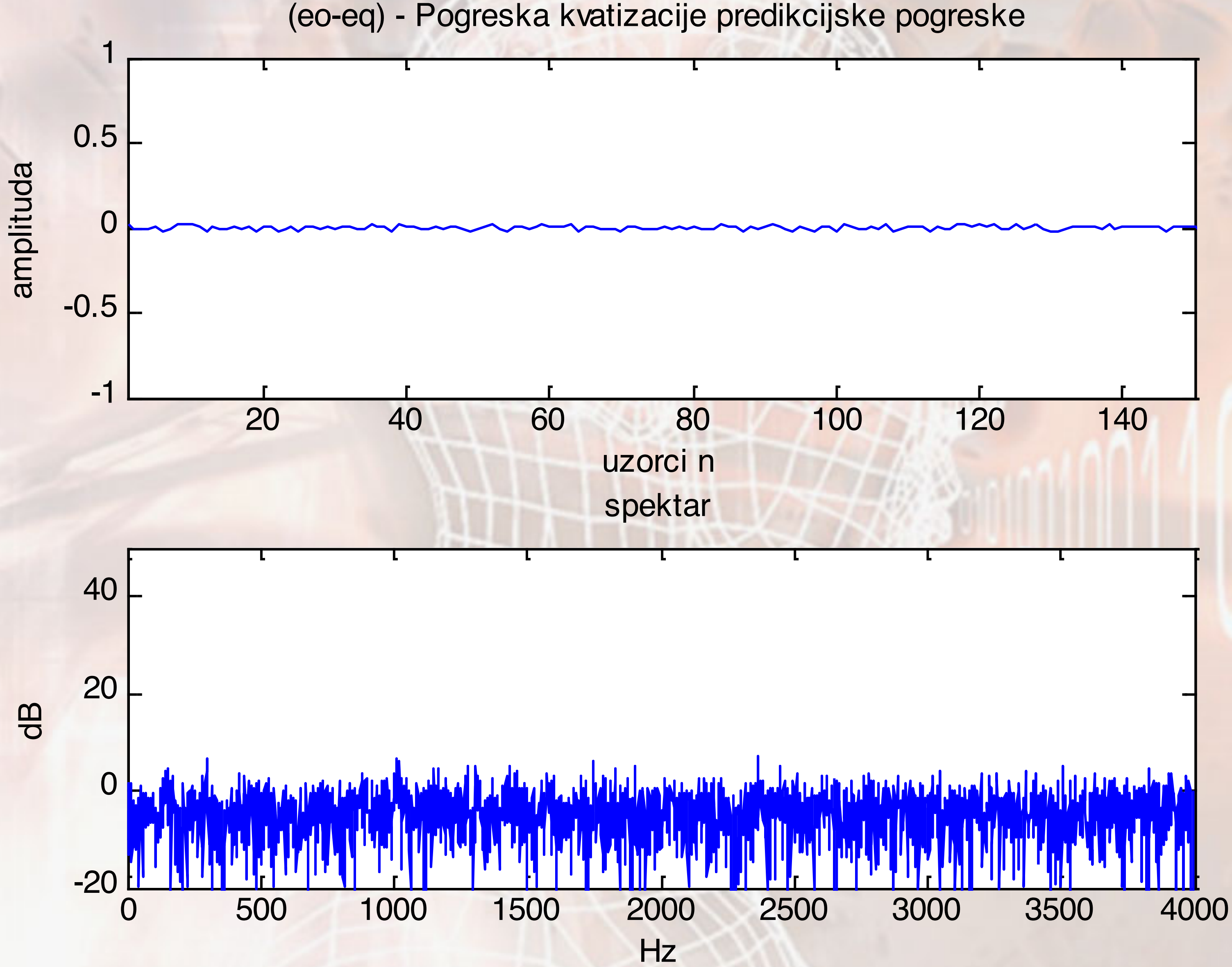
Prediction error (encoder side) vowel “o”, p=10



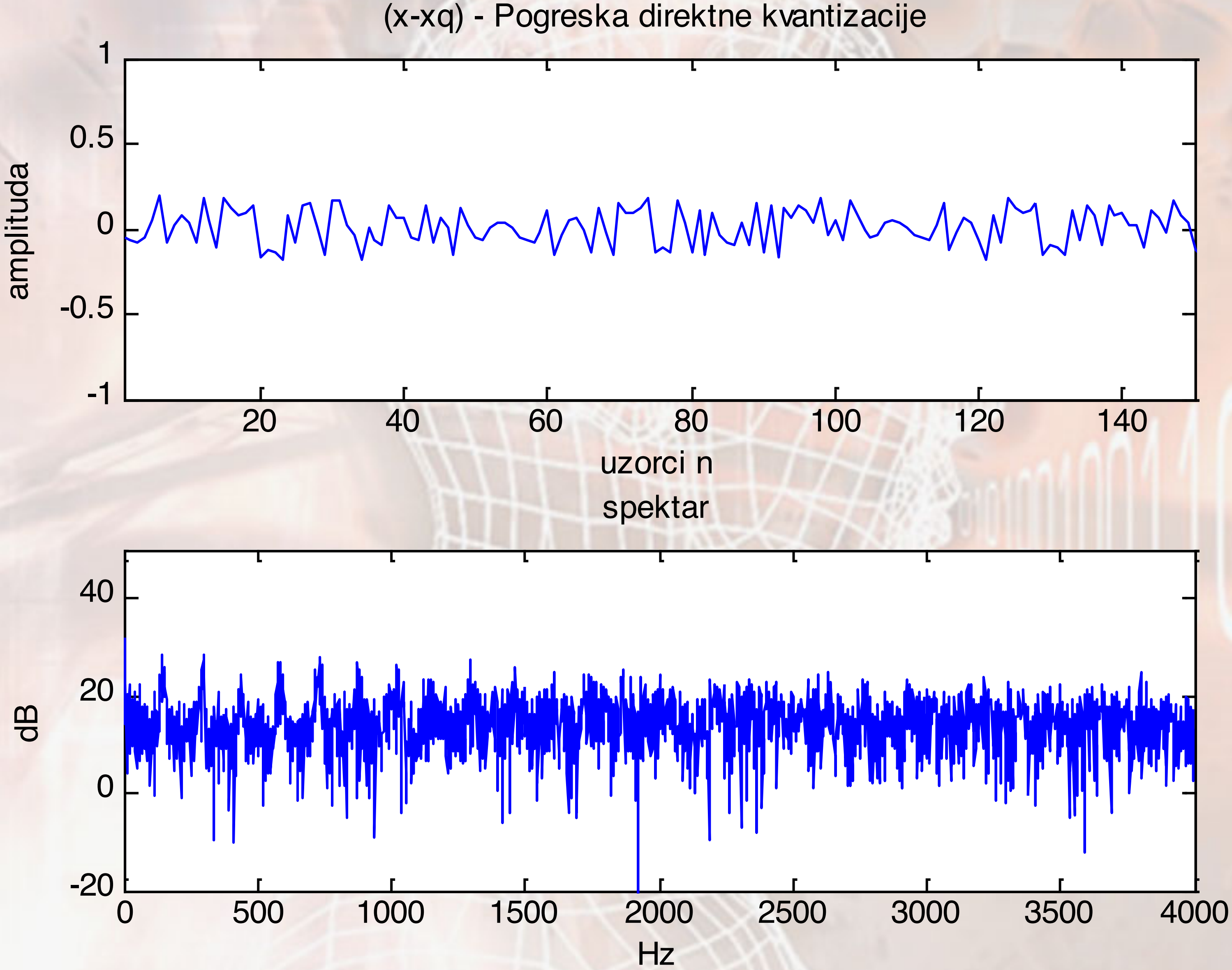
Prediction on the decoder side (open loop) vowel “o”, p=10, $H(I)$=2 bit



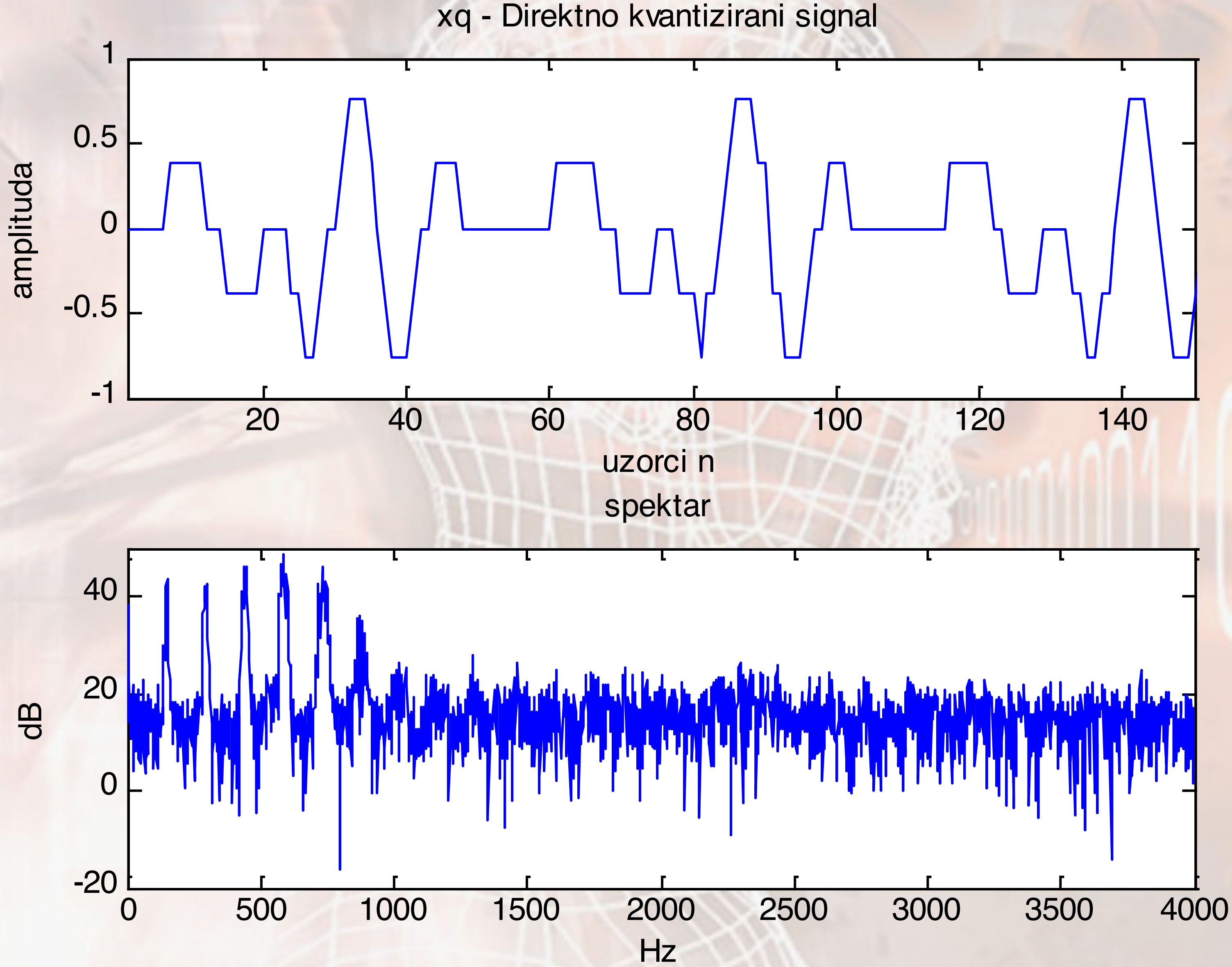
The quantization error of the prediction error vowel “o”, p=10, $H(I)$=2 bit



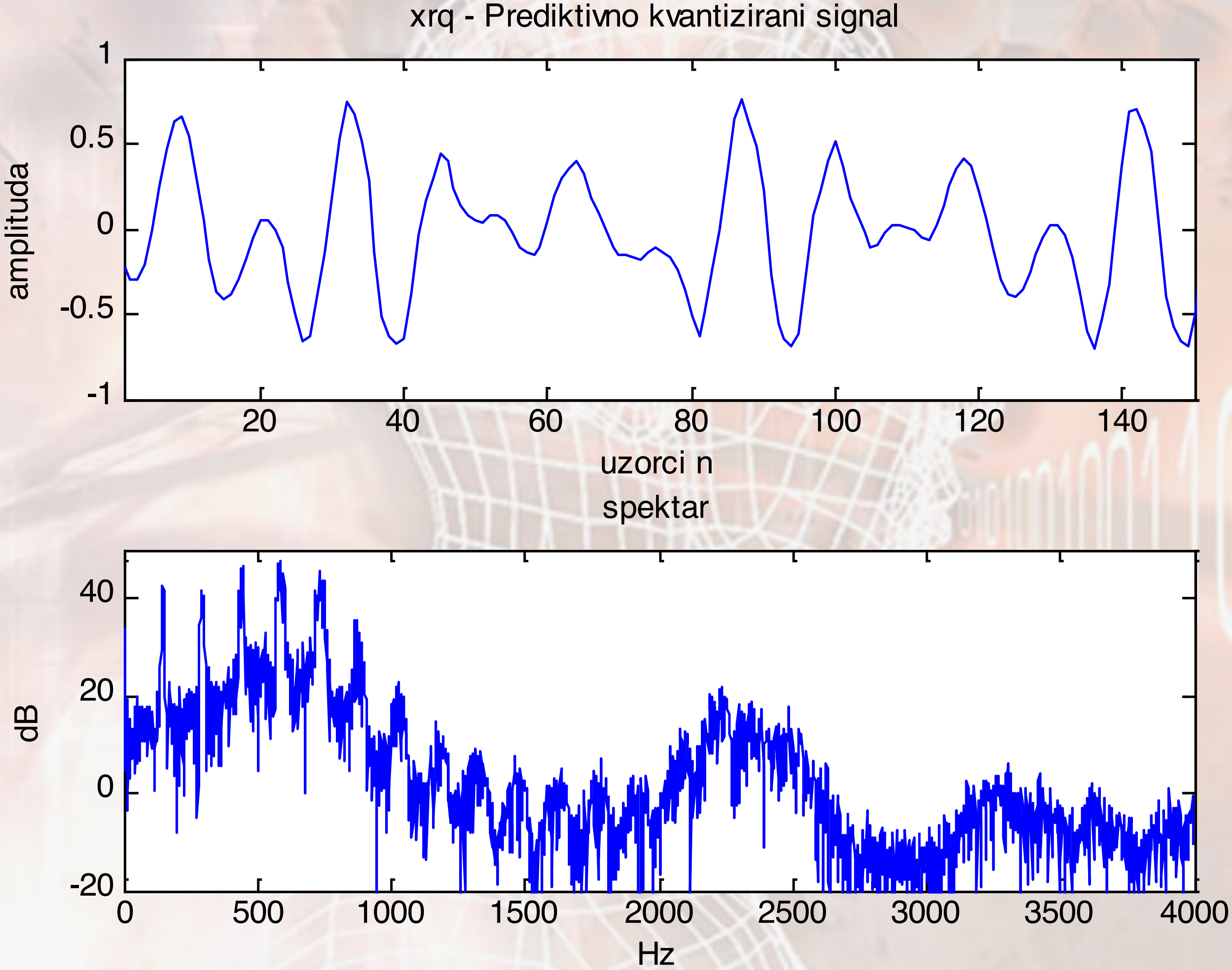
Quantization error signal of direct quantization for vowel "o", $H(I)$=2 bit



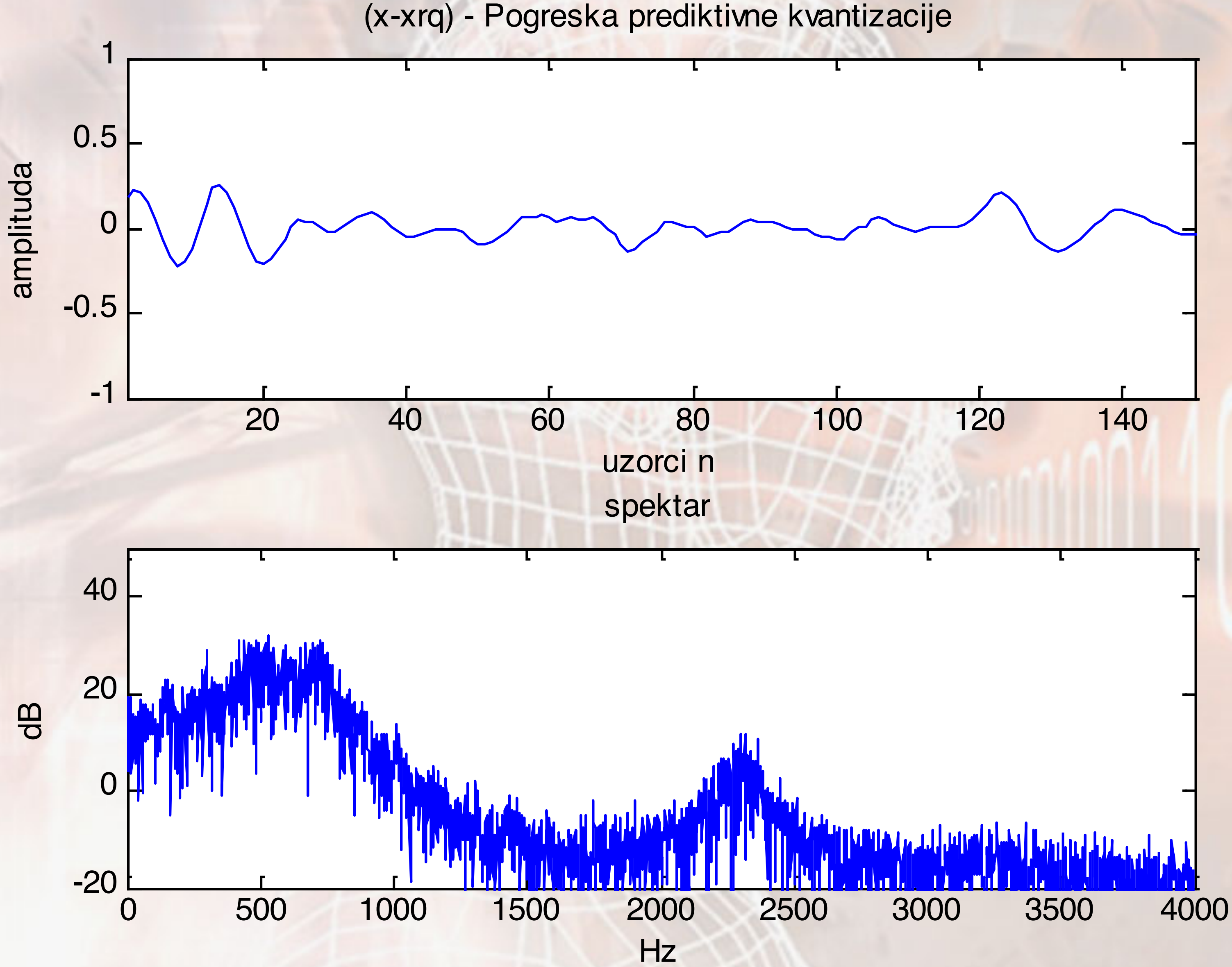
Directly quantized signal (baseline) vowel “o”, $H(I)$=2 bit



Predictively coded signal vowel “o”, p=10, $H(I)$=2 bit



Quantization error for open-loop predictive speech coding , vowel “o”, p=10, $H(I)$=2 bit



Determination of optimal predictor

- In this example, the linear predictor is determined by an autocorrelation procedure:

```
% Find a linear predictor (LPC) for the selected voice
R=xcorr(x,P);          % autocorrelation of inp. signal
R=R(P+[1:P+1]);        % only positive lag 0..P
M=toeplitz(R(1:P));    % form a system matrix
FI=R(2:P+1);           % free column
al=inv(M)*FI;          % predictor coefficients
A=[1 -al'];            % inverse filter
[HH,om]=freqz(1,A,1024);    % freq. resp. of LPC model
plot(om/pi*fs/2,20*log10(abs(HH)));
```

Determination of optimal predictor

- The coefficients of the optimal linear predictor for the selected vowel 'o' are:

	$-\alpha_1$	$-\alpha_2$	$-\alpha_3$	$-\alpha_4$	$-\alpha_5$
1.0000	-1.9266	1.1861	-0.5402	0.7915	-0.0247

$-\alpha_6$	$-\alpha_7$	$-\alpha_8$	$-\alpha_9$	$-\alpha_{10}$
-0.4785	-0.3595	0.6138	-0.0477	-0.0849

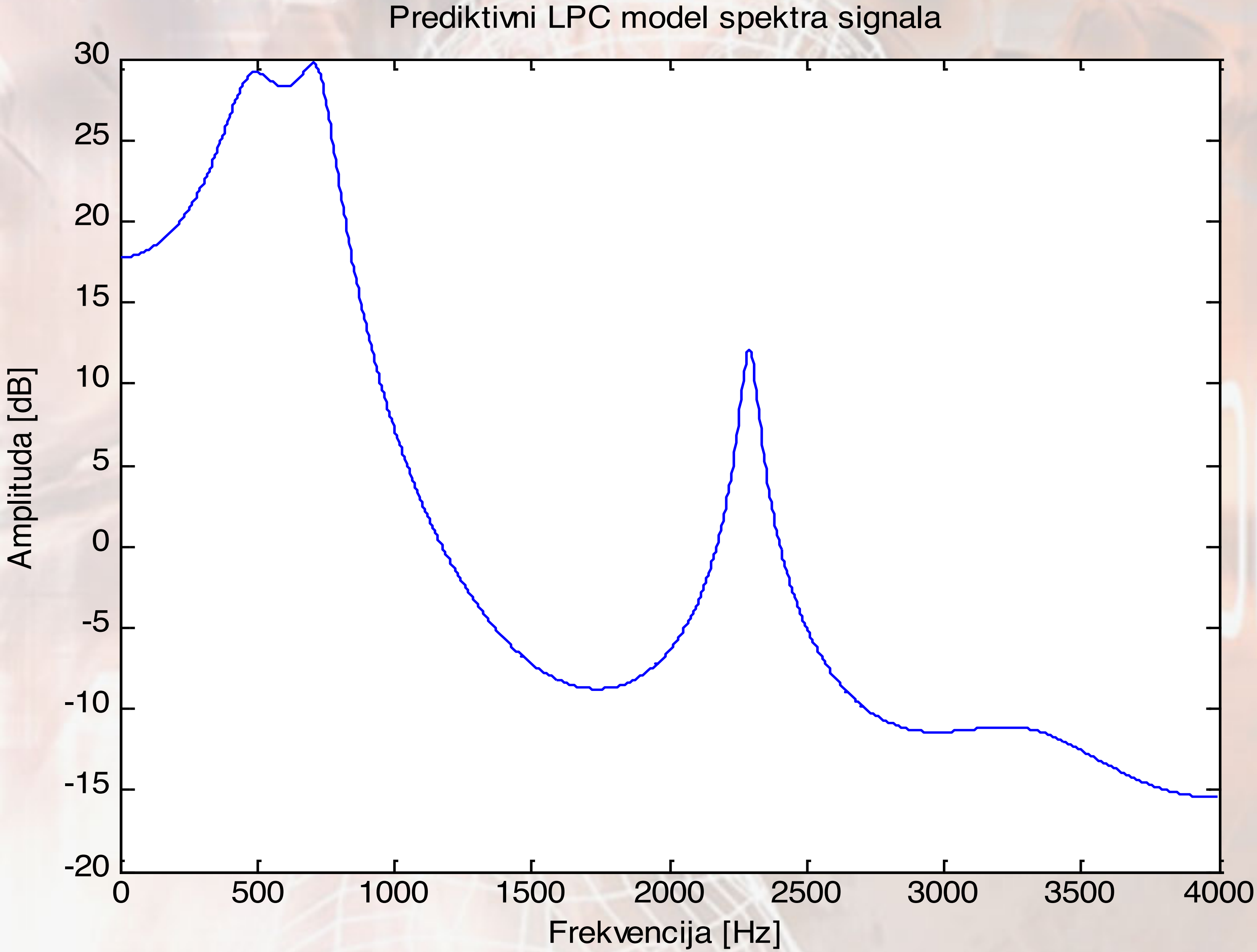
- The transfer function of the reconstruction system has the following form:

$$H(z) = \frac{1}{1 - P(z)} = \frac{1}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}$$

Open-loop predictive speech coding... discussion

- The variance of the prediction error signal is significantly smaller than the variance of the input signal ...
  - for the same entropy, a quantizer that quantizes the prediction error signal *eo* can have a much smaller step size than a quantizer for direct quantization of the signal *x*, and thus a smaller variance of the quantization error.
- Unfortunately, as part of the reconstruction on the side of the decoder, ...
  - the quantizer error (*eo*-*eq*) is amplified with the same reconstruction system that converts the ideal (unquantized) prediction error signal *eo* into the target x.

Frequency response of the reconstruction filter in the decoder $H(e^{j\omega})=1/(1- P(e^{j\omega}))$



Open-loop predictive speech coding... discussion

- So ...
  - the reconstruction error of the open-loop predictive encoder (*x*-*xrq*) is equal to the quantization error of the prediction error signal (*e*-*eq*) passed through the reconstruction filter $H(z)=1/(1-P(z))$.
- The decoder calculates the prediction from the previously reconstructed samples *xrq*, so its prediction *xpq* is different from the prediction *xp* used by the encoder.
- From the structure of the coding scheme, it is obvious that the following relation holds:

$$(x-xrq)=(eo-eq)+(xp-xpq)$$

Open-loop predictive speech coding... discussion

- We conclude that the open loop predictive coder reconstruction error consists **of two sources**:
  – prediction error signal quantization errors (*eo*-*eq*) and
  – differences between encoder-side and decoder-side predictions (*xp*-*xpq*).
- Increasing the entropy *H*(*I*) reduces the quantization error (*eo*-*eq*), but since the reconstructed samples become more accurate, it also reduces the prediction difference (*xp*-*xpq*) and thus the total predictive reconstruction error (*x*-*xrq*).

Open-loop predictive speech coding... discussion

- The quantization error (*eo*-*eq*) has an approximately uniform probability density in the interval $-\Delta/2$ to $\Delta/2$, but also a flat spectrum, i.e., it has the properties of **white (spectrally uncolored) noise**.
- Passing it through the system $H(z)$ gives a reconstruction error (*x*-*xrq*), which is spectrally colored exactly the same as the signal itself ...
  - this is equivalent to the source of white noise being added to the ideal excitation signal at the entrance to the vocal tract, i.e. at the vocal cords, so only a slightly “rougher” voice is obtained as a result!

Open-loop predictive speech coding... discussion

- For the vowel case, the reconstruction error (*x*-*xrq*) sounds very similar to the original voice, but pronounced in a whisper (silently), i.e., without vibrating the vocal cords.
- Due to the spectral similarity of the desired signal and the reconstruction error signal, the error will be harder to detect and is less annoying than in the case of direct quantization where the error (*x*-*xq*) is more similar to white noise and is more perceivable then for the predictive coding case.
- This phenomenon is called "***frequency masking***".

Open-loop predictive speech coding... discussion

- The open loop predictive encoder sounds better than the directly quantized signal precisely because of the **more favorable spectral distribution** of the quantization error!

- Because the encoder spectrally shapes quantization noise in order to obscure it "behind" the signal spectrum, such procedures are called **noise shaping**.

Open-loop predictive speech coding... discussion

- It is interesting to note that predictive coding in an open loop **will not generally increase the SQNR ratio**, but will only redistribute the same error energy better.
- Results for the example of the vowel “o” with *p*=10

```
Actual output entropies: H(I_X)=2.001 H(I_E)=2.000
Actual output SNR: SNR(X)=10.139 dB SNR(E)=10.742 dB
Actual output SNR after reconstruction: SNR(Xr)=11.486 dB
Prediction gain: PGotv =19.305 dB
Quality increase: SNR(Xr)-SNR(X)=1.347 dB
```

What have we learned?

- determination of linear predictors by autocorrelation procedure (example),
- reconstruction system on the decoder side,
- reconstruction error and its sources,
- spectral properties of quantization error for prediction error signal and for final reconstruction error on the decoder side,
- frequency masking,
- spectral shaping of quantization noise,
- SQNR encoder ratio for open loop predictive coding.

An example of closed-loop predictive coding of speech

prof.dr.sc. Davor Petrinović

Closed-loop predictive speech coding

- How to solve the problem of the difference between the predictions on the side of the encoder and the decoder?
- It is necessary to calculate the prediction on both sides **from the same signal** ...
  - from previous samples of the reconstructed signal ...
  - prediction is worse than for open loop, but at least there is no difference between predictions anymore!
- Simulation of closed-loop speech coding is performed using the program:
  - **MT04_ADPCM_govor.m**

Closed-loop predictive speech coding

- The first part of the program performs the same procedures as the open loop encoder:
  - selects the desired vowel, predictor order and output entropy,
  - extracts the voice and determines the optimal linear predictor for the case of open loop prediction,
  - calculates the prediction error signal for the open loop case,
  - determines the differential entropies of the input signal and the prediction error signal, and
  - determines the initial quantization steps for both signals which achieve the same desired output entropy,

Closed-loop predictive speech coding

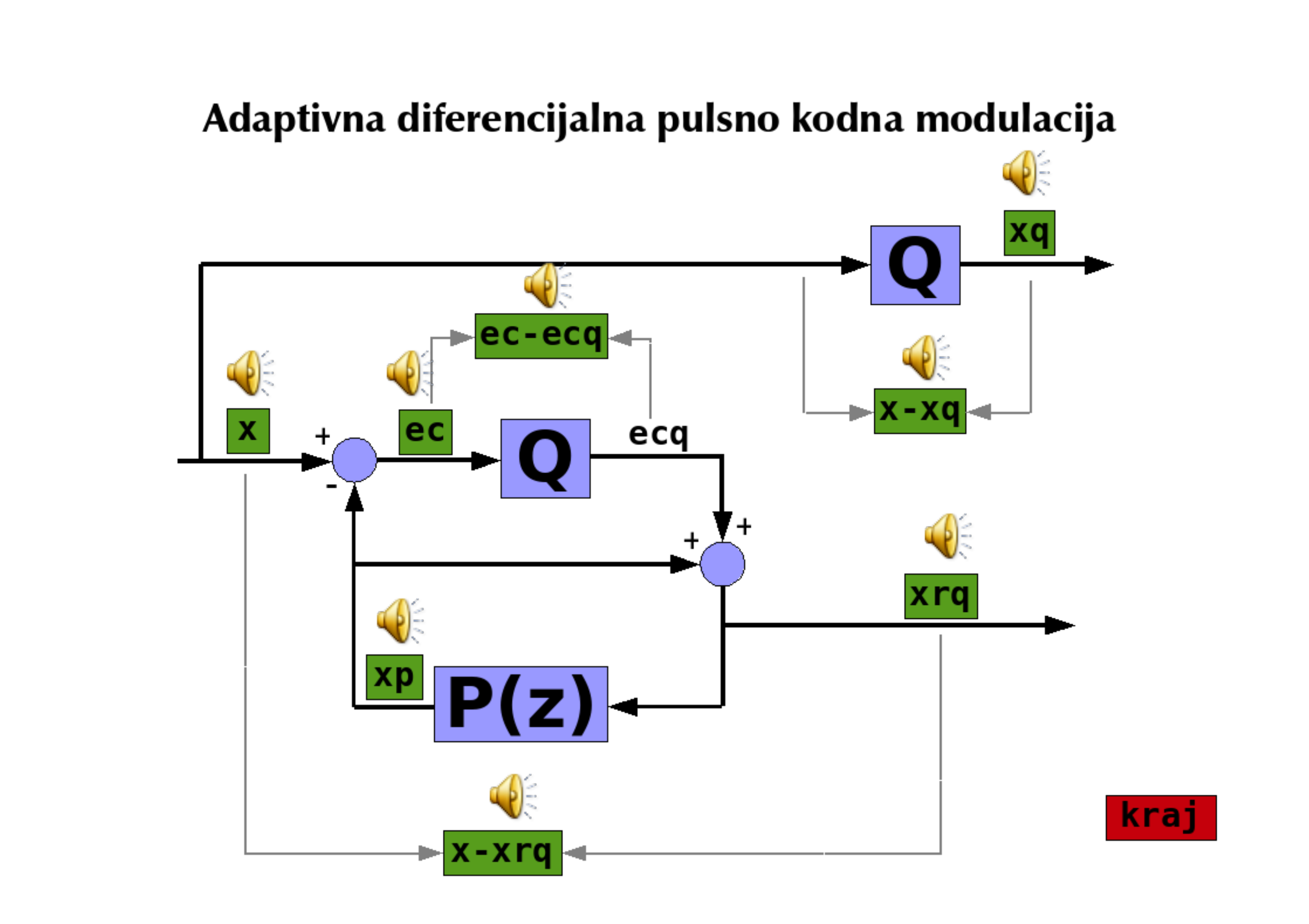
- ... in the second part of the program it performs the actual quantization of the signal in a closed prediction loop (sample by sample):
  - calculates the prediction *xp*[*n*] of the current sample *x*[*n*] from past reconstructed samples *xrq*[*n-i*], i=1, 2, … *p,*
  - subtracts the prediction from the input signal sample, thus obtaining the prediction error *ec*[*n*]
  - quantizes it by the designed quantizer into *ecq*[*n*], and
  - by summing *ecq*[*n*] back to the prediction *xp*[*n*] obtains a reconstructed output sample *xrq*[*n*].
  - An identical procedure is performed by the remote decoder as well.

Closed-loop predictive speech coding

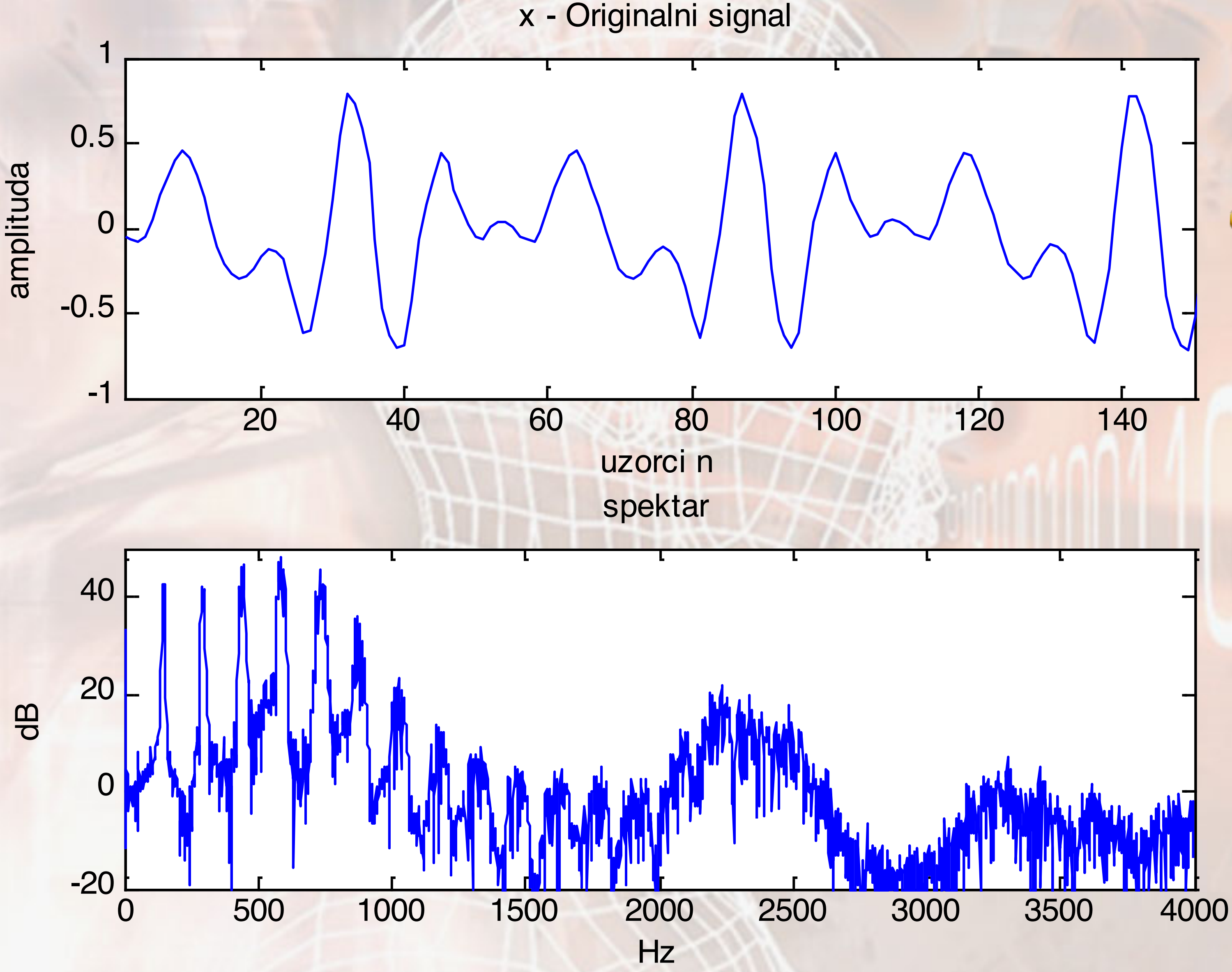
- finally ...
  - calculates the actual entropy of quantizer output indices in a closed loop operation, and ...
  - in case it differs from the desired entropy, calculates a new step size Δ
  - and performs another quantization pass for the whole signal, and
  - if necessary, repeat this several times, ... and
  - finally calculates the SQNR relationships at all characteristic points of the coding structure:

(x-xq), (x-xrq), (ec-ecq).

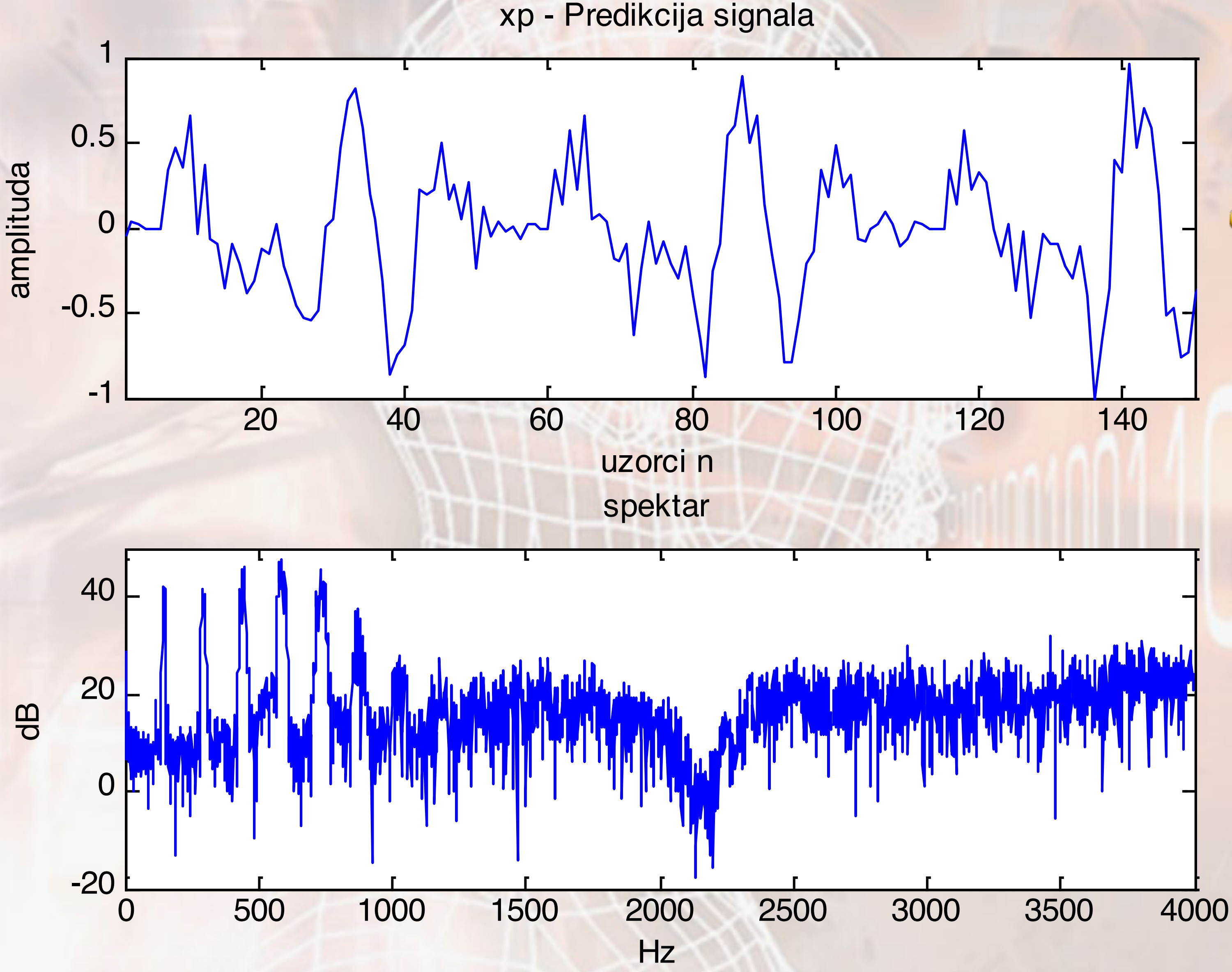
Closed-loop predictive speech coding for vowel “o” p=10, $H(I)$=3 bit



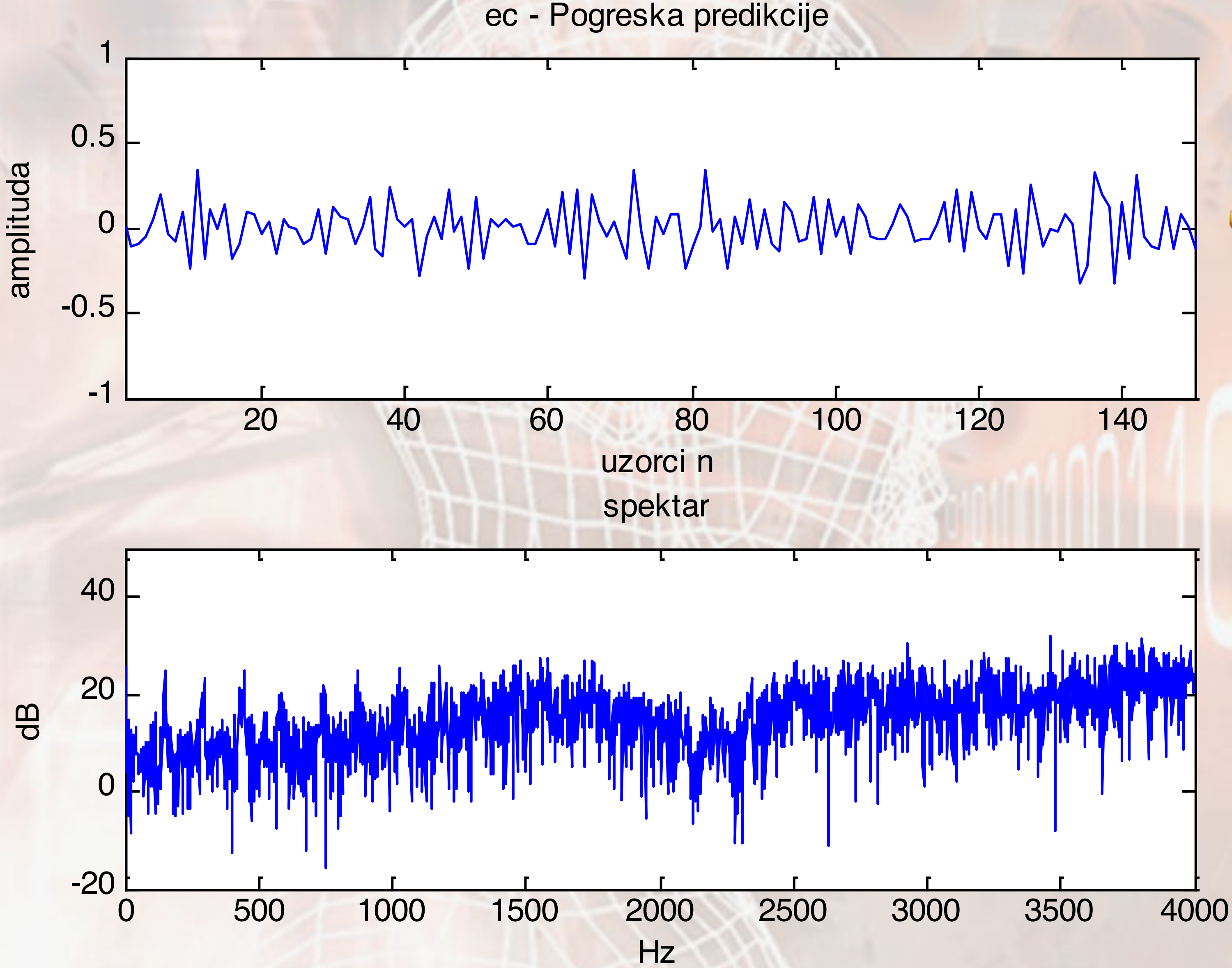
Input speech signal
vowel “o”



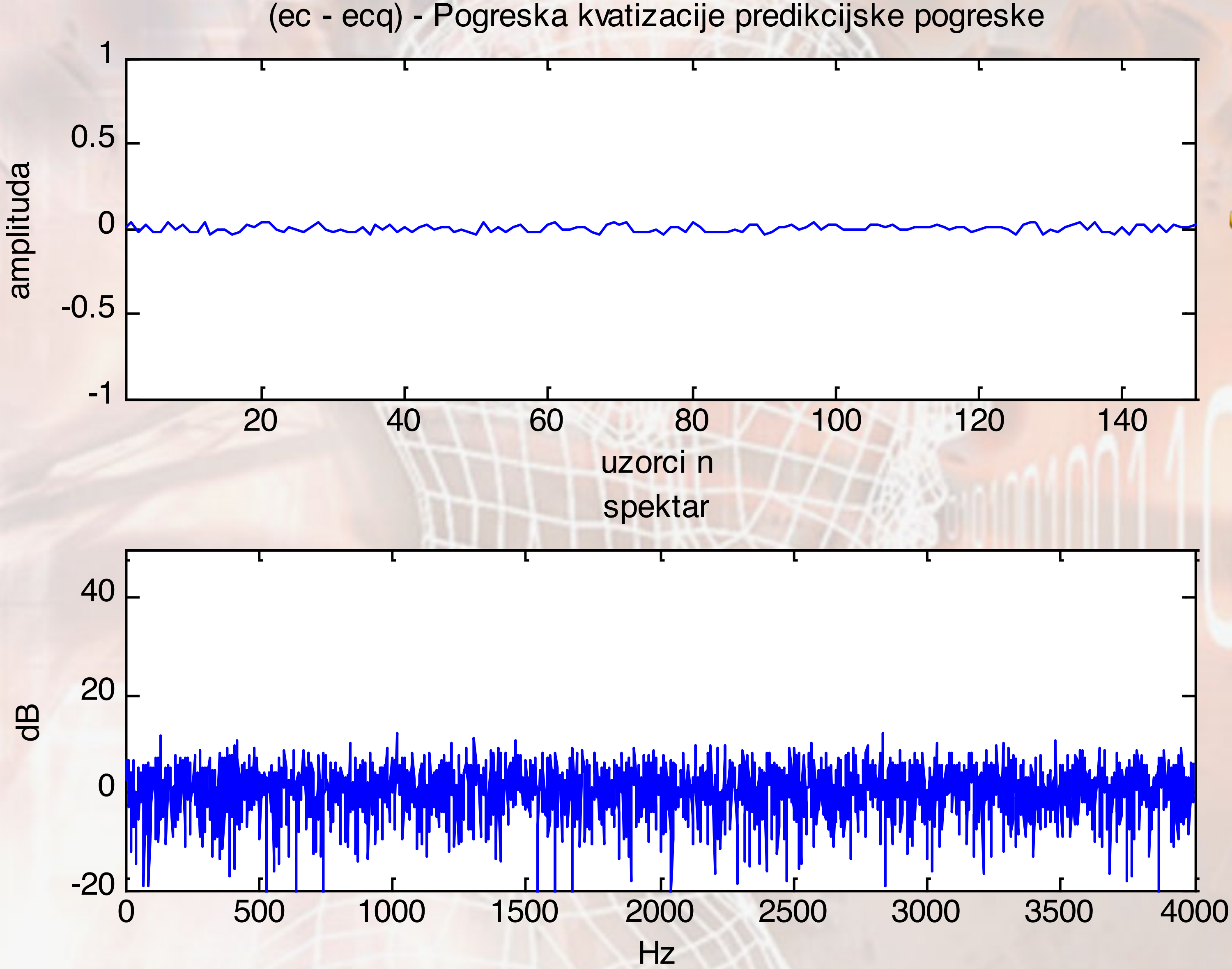
Prediction on both sides (closed loop) vowel “o”, p=10



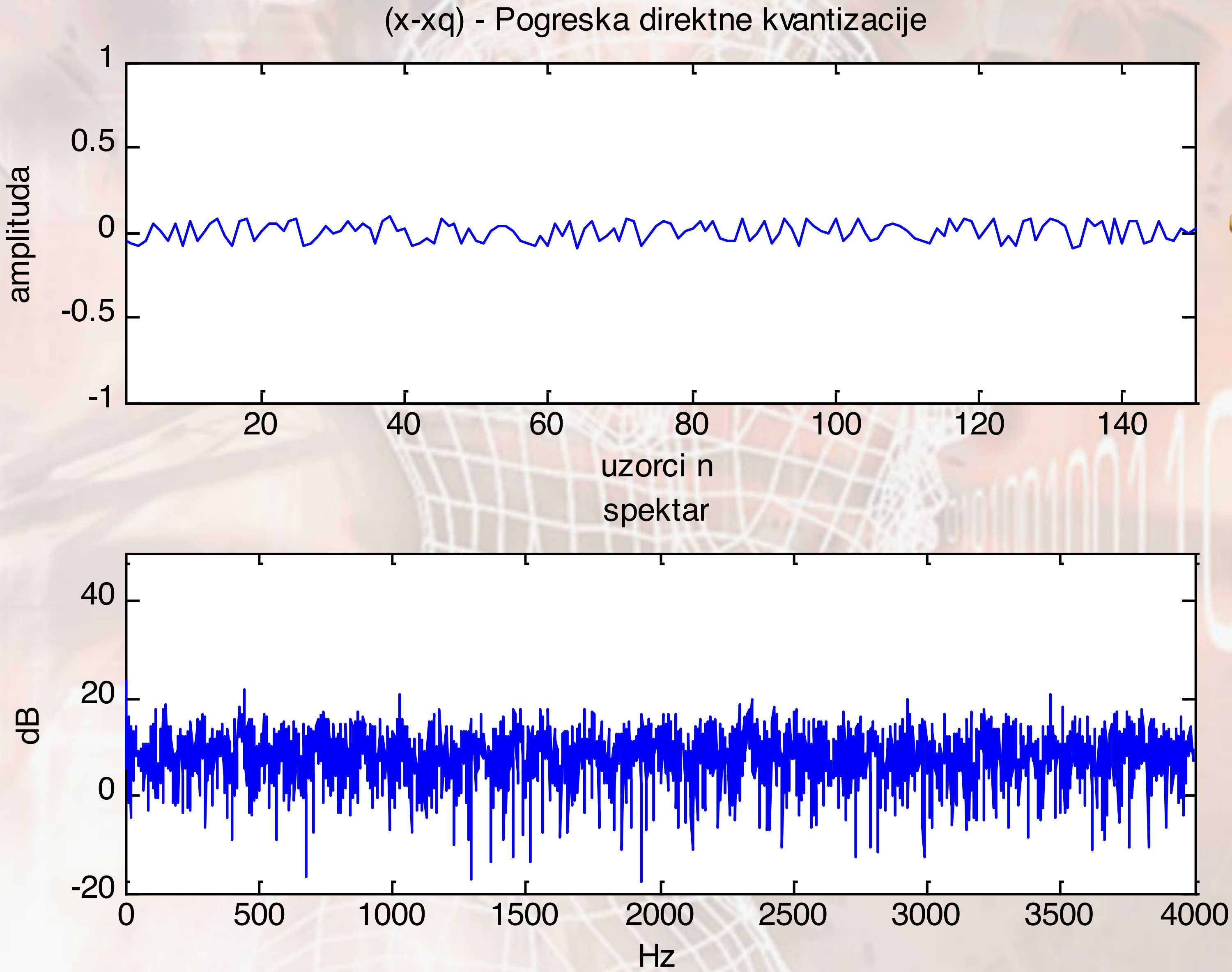
Closed loop prediction error vowel “o”, *p*=10



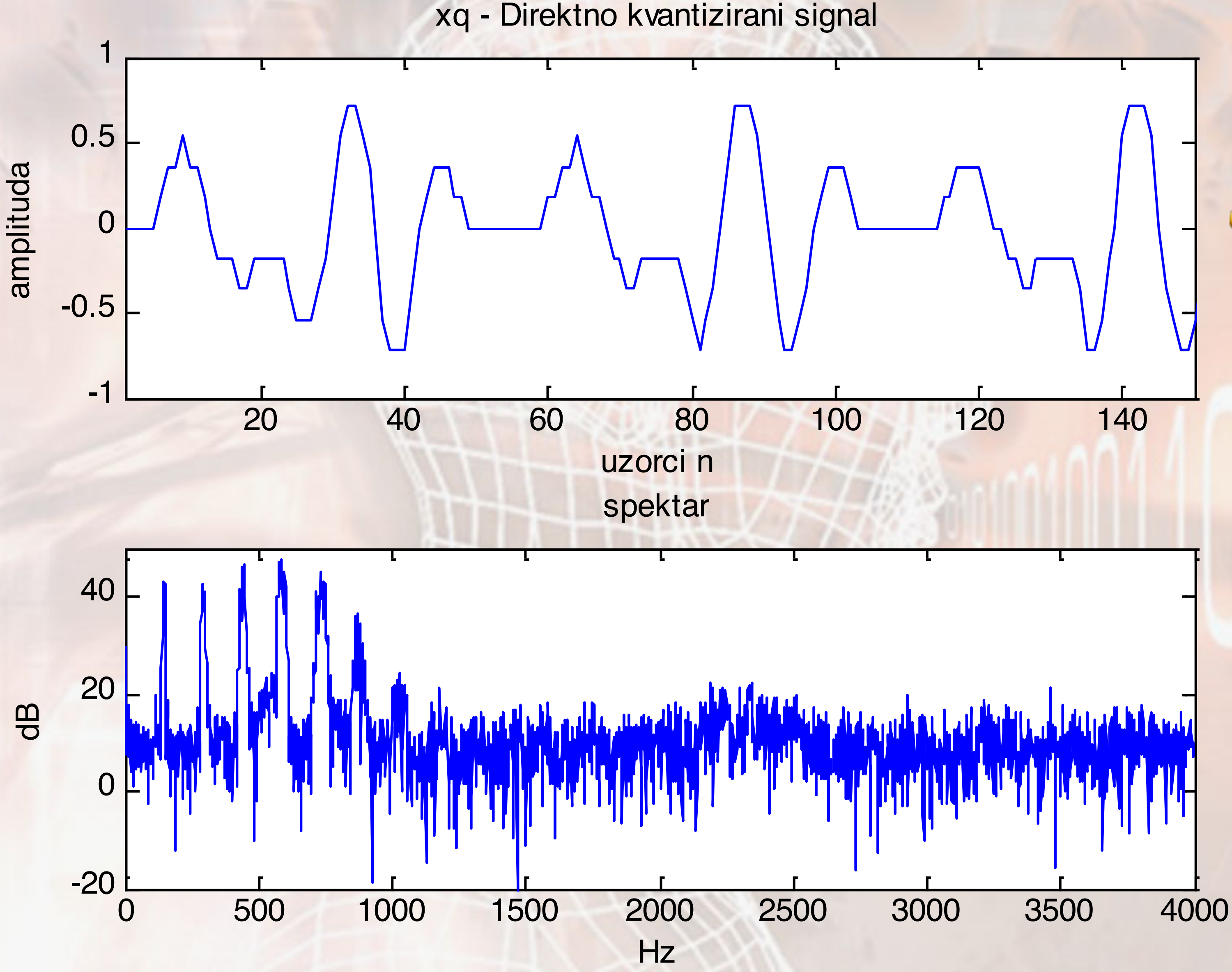
The quantization error of the closed loop prediction error, vowel “o”, *p*=10, *H*(*I*)=3 bit



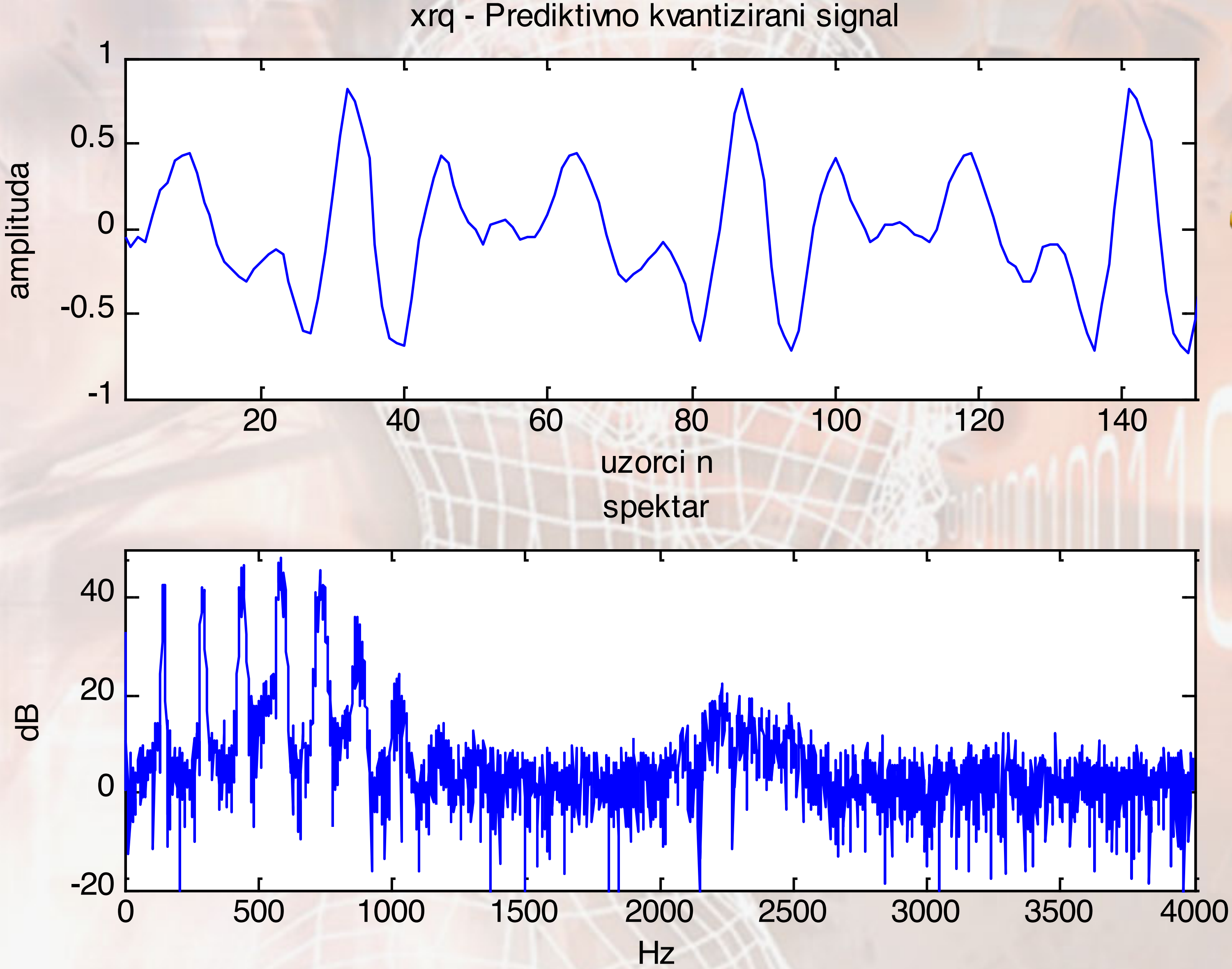
Direct quantization error signal for vowel "o", $H(I)$=3 bit



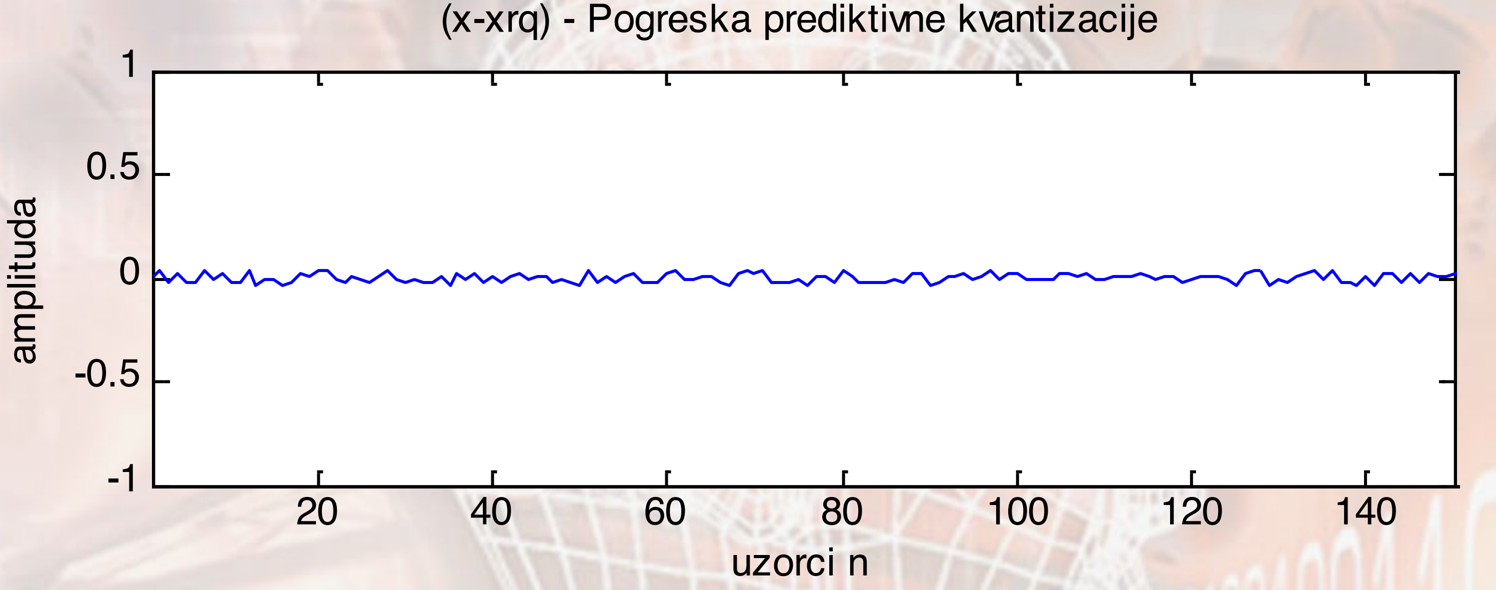
Directly quantized signal (baseline) for vowel “o”, $H(I)$=3 bit

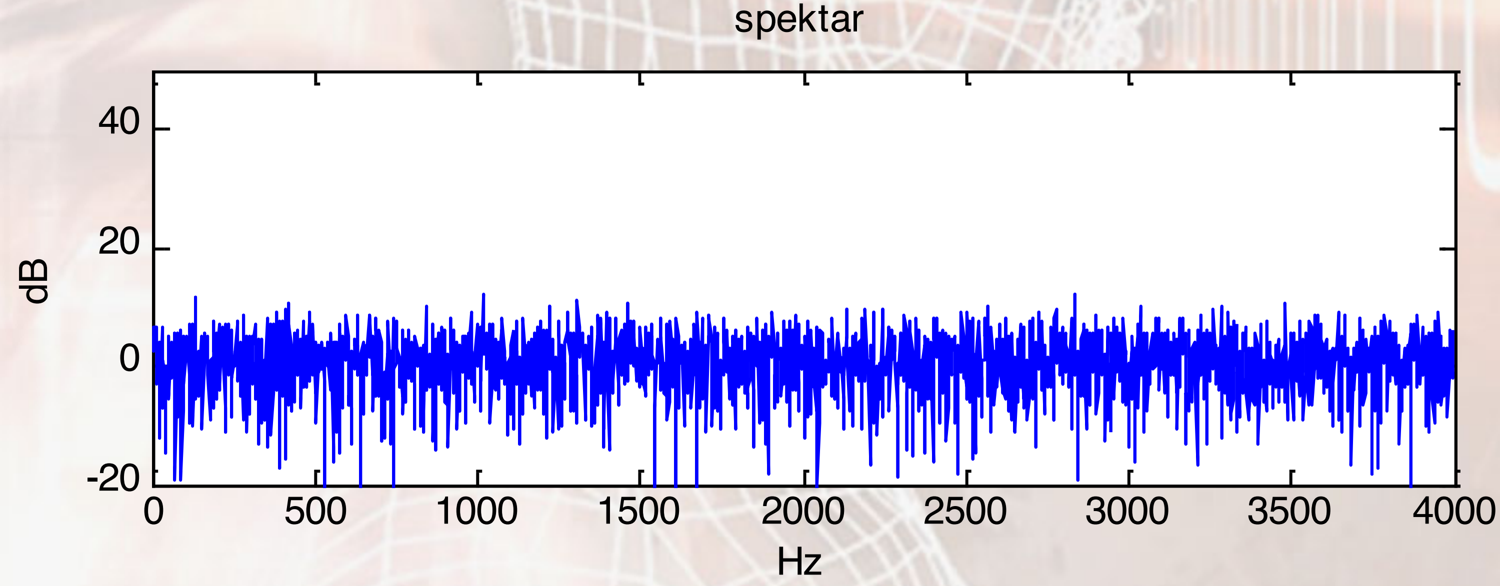


Predictively coded signal for vowel “o”, p=10, $H(I)$=3 bit



Quantization error for closed-loop predictive speech coding , vowel “o”, p=10, $H(I)$=3 bit





Closed-loop predictive speech coding... discussion

- By analysing the block diagram of a predictive encoder with a closed loop, an important fact is noticed:
  - the reconstruction error (*x*-*xrq*) is identical to the quantization error of the prediction error signal (*ec*-*ecq*), because
  - identical prediction *xp* is firstly subtracted from *ec* and then added back to *ecq* after quantization!
- Quantization of signal *ec* is easier than for signal *x*, because its variance is smaller!
  - the greater the reduction in the variance of the signal being quantized, the greater the prediction gain ... let's prove it ...

Closed loop prediction gain

- Example ...
  - the input process *X* has a Gaussian distribution of variance σ_x^2, so ...
  - the step size of the ECSQ quantizer for direct quantization (*xq*) must be ...

$$\Delta_{xq} = \sqrt{2\pi\, e\sigma_x^2} \cdot 2^{-H(I)}$$

  - For the closed-loop prediction error process, we also assume that it has a Gaussian distribution, but of reduced variance σ_{ec}^2, so
  - its ECSQ quantizer for the same output entropy has a step size ....

$$\Delta_{ecq} = \sqrt{2\pi\, e\sigma_{ec}^2} \cdot 2^{-H(I)}$$

Closed loop prediction gain

- Example ... continued
  - the mean square quantization error for a directly quantized process is:

$$D_{xq} = \frac{1}{12}\Delta_{xq}^2 = \frac{\pi\, e\sigma_x^2}{6} \cdot 2^{-2H(I)}$$

  - while the mean square error of the predictively coded process (*x*-*xrq*), D_{xrq}, is identical to the mean square quantization error of the prediction error signal D_{ecq}:

$$D_{xrq} = D_{ecq} = \frac{1}{12}\Delta_{ecq}^2 = \frac{\pi\, e\sigma_{ec}^2}{6} \cdot 2^{-2H(I)}$$

Closed loop prediction gain

- Example ... continued
  - by substituting these distortions in the *SQNR* signal-to-quantization error expressions, it follows:

$$SQNR_{xrq} - SQNR_{xq} = 10 \cdot \log_{10} \frac{\sigma_x^2}{D_{xrq}} - 10 \cdot \log_{10} \frac{\sigma_x^2}{D_{xq}}$$

$$= 10 \cdot \log_{10} \frac{D_{xq}}{D_{xrq}} = 10 \cdot \log_{10} \frac{\sigma_x^2}{\sigma_{ec}^2}$$

  - The obtained difference between the SQNR ratio of predictive and direct quantization is called the **prediction gain** of the encoder and is expressed in [dB].

Closed loop prediction gain

- Prediction gain:

$$PG = 10 \cdot \log_{10} \frac{\sigma_x^2}{\sigma_{ec}^2} \quad [dB]$$

  – The expression shows that the choice of a predictor that minimizes the variance of the prediction error is consistent with maximizing the prediction gain of the closed–loop predictive coding!
  – Unfortunately, note that the expression contains a **closed-loop prediction variance**, which cannot be determined without a predictor as well as quantizer, because ...
  – to find a predictor we need to have a quantizer, and to design a quantizer we need to have a predictor.

Closed loop prediction gain

- Determining predictor ...
  - To break this vicious circle, the predictor is first computed by assuming an ideal quantizer (*ecq*=*ec*).
  - This corresponds to the open loop predictor, because in the case of ideal quantization the input to the predictor is equal to the input signal samples (*xrq*=*x*).
  - Therefore, the predictor is optimized to minimize the variance of the open loop prediction error σ_e^2.

Closed loop prediction gain

- Determining predictor ... continued
  - The corresponding ideal prediction gain can also be defined, which is achieved with infinite output entropy $H(I)=\infty$

$$PG_{\infty} = 10 \cdot \log_{10} \frac{\sigma_x^2}{\sigma_e^2} \quad [dB]$$

  - It is sometimes called ***open-loop prediction gain.***
  - The actual prediction gain *PG* of a closed loop coder will asymptotically approach this ideal PG_{∞} value with increasing entropy.

Closed-loop predictive speech coding... discussion

- Results for the example of the vowel “o” with closed loop predictive coding with p=10 and $H(I)$=3:

```
Actual output entropies: H(I_X)=3.001 H(I_E)=3.000
Actual output SNR: SNR(X)=16.609 dB SNR(E)=16.467 dB
Actual output SNR after reconstruction: SNR(Xr)=24.539 dB
Prediction gain: PGotv =19.305 dB and PGzatv =8.072 dB
Quality increase: SNR(Xr)-SNR(X)=7.930 dB
```

- Compare the expected closed loop prediction gain (PGzatv) and the actual increase in the coding quality SNR(Xr) -SNR(X), which are almost identical.

Closed-loop predictive speech coding... discussion

- Let us also consider the spectral properties of the predictive quantization error for a closed loop coder:
  - in this case the quantization error is spectrally uncolored, i.e., it has the characteristics of white noise, just like the error of a directly quantized signal.
- Thus, unlike the open loop predictive coding, the procedure of noise shaping of the quantization error spectrum **is not used here**, but the quality is increased solely **by reducing the error energy**.

Applications of closed-loop predictive speech coding

- The described predictive coding procedure is successfully used within actual speech signal coder standards.
- The procedure is also known as: ***A****daptive* ***D****ifferential* ***P****ulse* ***C****ode* ***M****odulation,* ***ADPCM***.
- It is used within the ITU standard (*International Telecommunication Union*) **G.726**, which originated from older versions G.721 and G.723

ADPCM – G.726

- It is mostly used in digital wired telephony in the USA, digital telephone exchanges, and also in DECT wireless telephones.
- Supports speech data rates from:
  - 16, 24, 32 and 40 kbit/s.
- The most commonly used rate is 32kbit/s, which saves twice as much as the older **G.711** standard, which performs non-predictive (direct) quantization.
- The standard does not provide for the use of an entropy encoder, but the quantizer index is encoded by a fixed-length code.

What have we learned?

- reconstruction error and its relation to the quantization error of the closed-loop prediction error
- prediction gain in closed loop coding
- determination of predictor
- open loop prediction gain
- spectral properties of the reconstruction error
- adaptive differential pulse code modulation
- ADPCM - G726

Example of predictive image encoding in a closed loop

prof.dr.sc. Davor Petrinović

Predictive image coding

- The same principles can be applied to image coding:
  – instead of temporal correlations of samples, in the image there are spatial correlations of adjacent elements of the image, i.e., pixels;
  – very often the images have homogeneous regions, within which the adjacent pixels are very similar (same or similar in color and of the same or similar intensities);
  – it is logical to assume that each pixel can be successfully predicted as a linear combination of its spatial neighbors!

Predictive image coding – a problem of causality

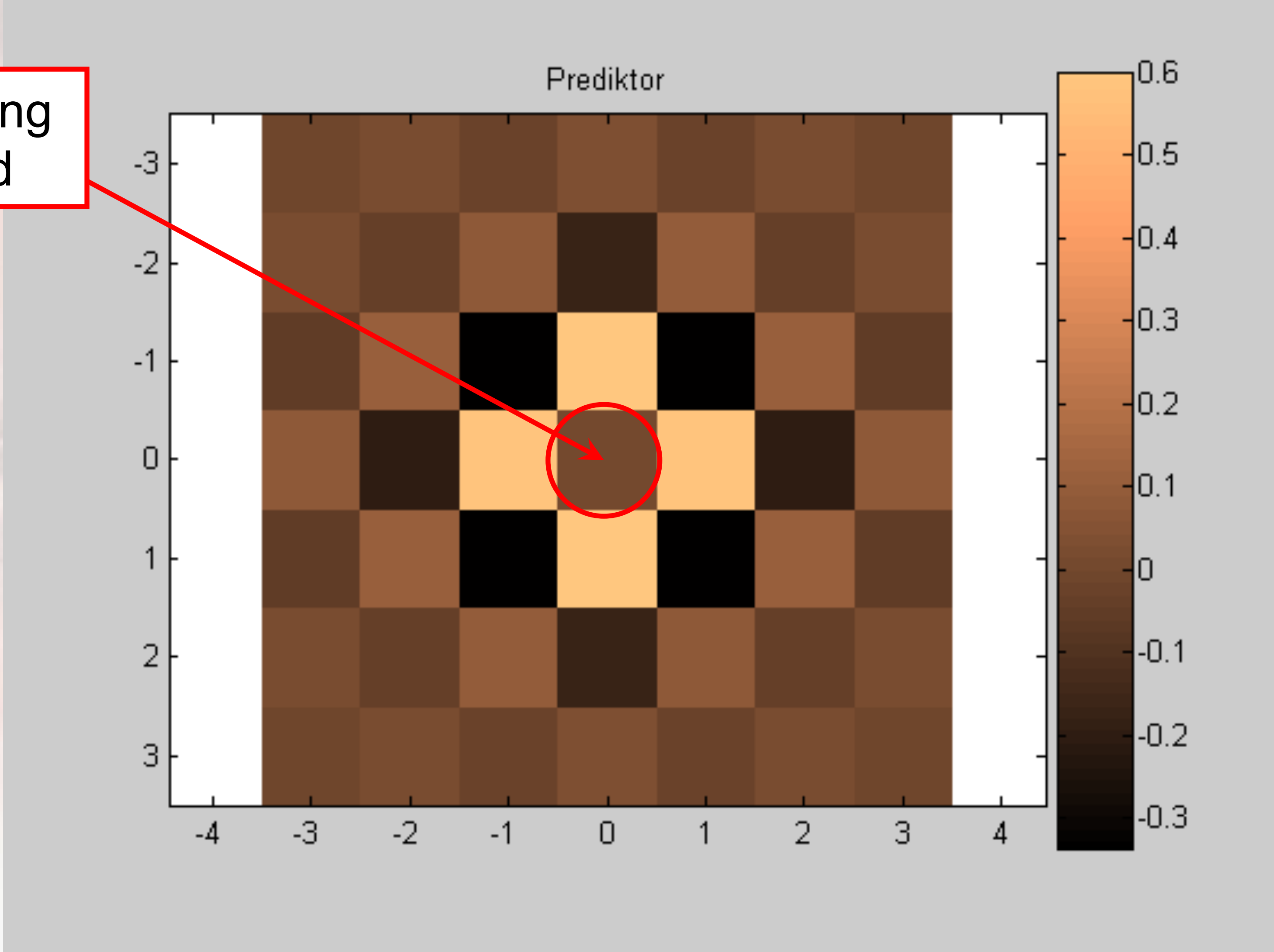
- Prediction must be performed from the part of the image that exists on both sides (encoder and decoder), hence from the "previous" samples.
- What are the previous samples?
  - there is no time ordering in the image.
- It is possible to imagine a two-dimensional image signal as if it were one-dimensional ...
  - by introducing the operation of sequential reading of image elements.

Sequential image reading

- The reading sequence can be any, and the only condition is that chosen addressing "visits" all the pixels in the image, for example:
  - reading by lines,
  - column reading,
  - a snake-shaped reading that zigzags diagonally from one corner of the image to another.
- Prediction can be made regardless of the way of chosen sequential reading ...
  - it is enough to use as a source of prediction only samples that are **“earlier”** in the reading sequence ...
  - for such a predictor we say it is **causal**.

Non-causal spatial predictor of the third order

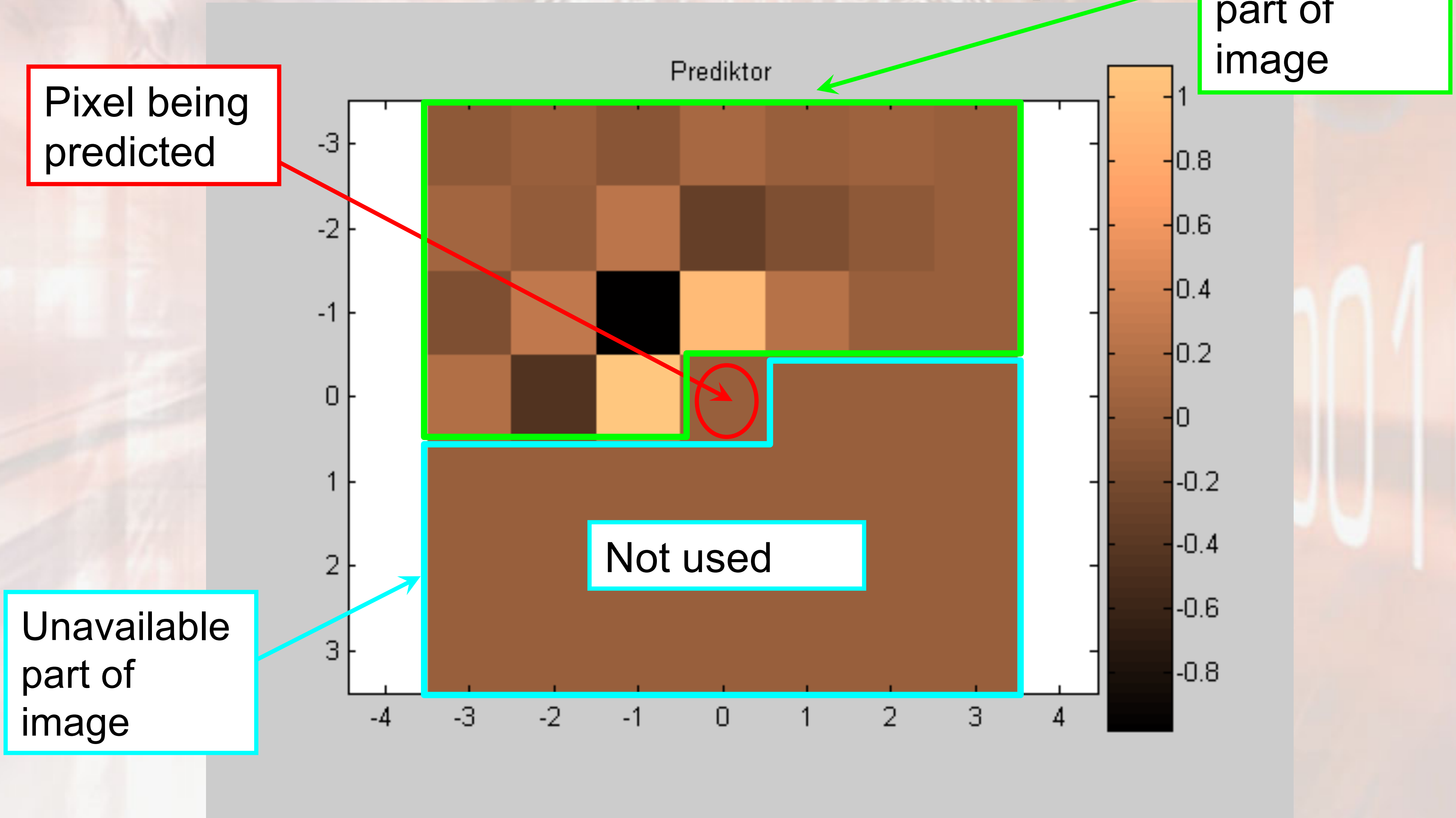
Pixel being predicted



Causal predictor

- For the case of sequential reading by lines, the causal predictor uses as the source of prediction:
  – selected pixels from all of the above (previous) lines relative to the pixel position we predict and
  – selected pixels from the same row, located to the left (before) of the pixel we predict.
- The chosen region of the previous adjacent pixels that is actually used for prediction depends on the type of spatial correlations in the image.

Causal spatial predictor of the third order - an example



Causal predictor

- The prediction gain of the causal predictor is less than the gain of non-causal predictor, since
  - for the case of sequential reading by lines predictor utilizes **only correlations from one part of the pixel neighborhood** (left & up), but
  - it can be used in a closed ADPCM predictive loop!
- Predictor coefficients are determined by the previously described procedures for determining the optimal linear predictor.

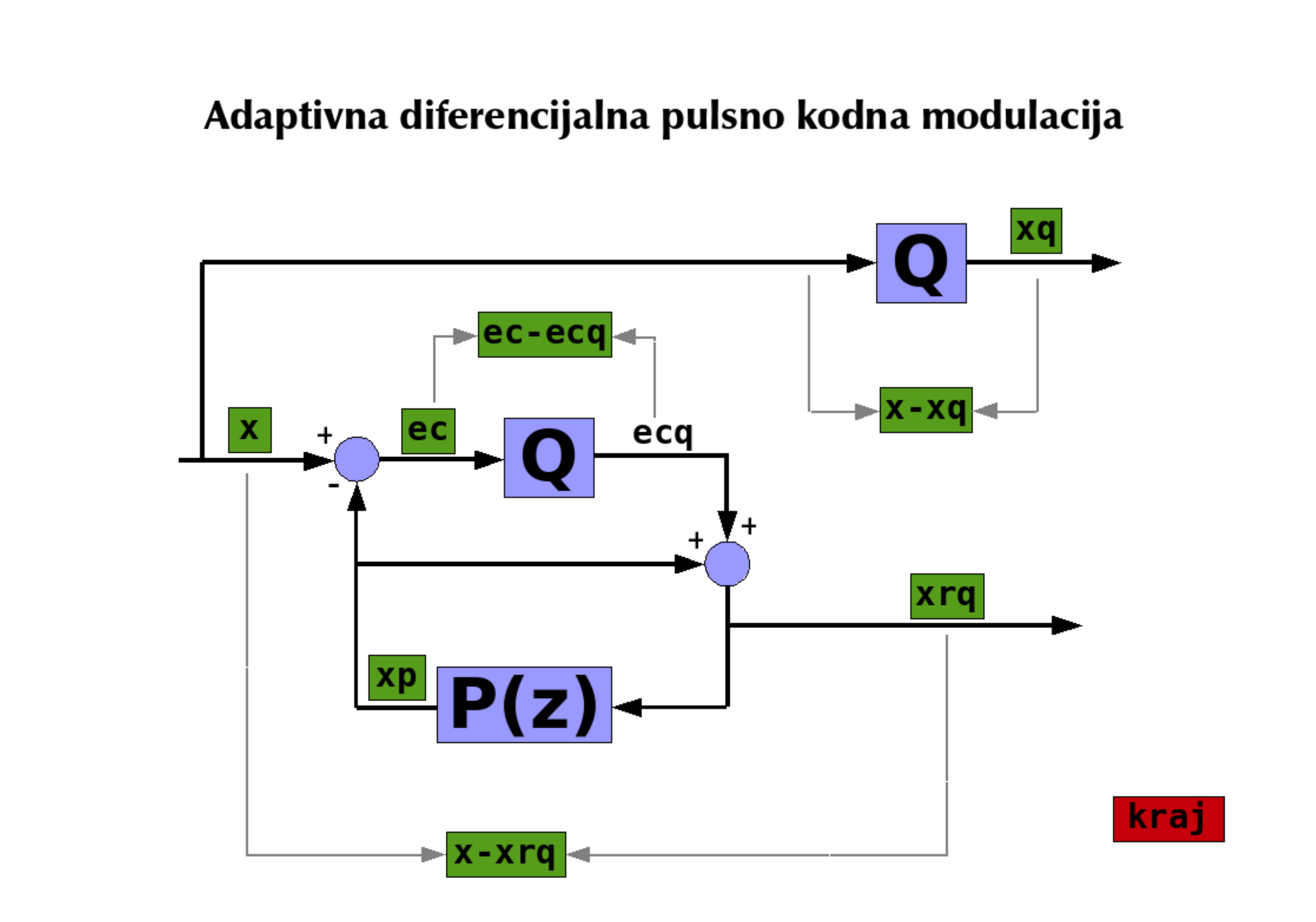
Predictive image coding - an example

- Predictive Image Coding simulation program in Matlab:
  - **MT04_ADPCM_slika.m**
  - it's very similar to a speech coding program,
  - coding is carried out on the monochromatic image in shades of gray,
  - it is possible to select the input image, the predictor order and the desired output entropy,
  - the optimal linear predictor is determined by the function: **MT04_impred.m** based on the chosen image and its spatial correlation properties.

Predictive image coding - an example

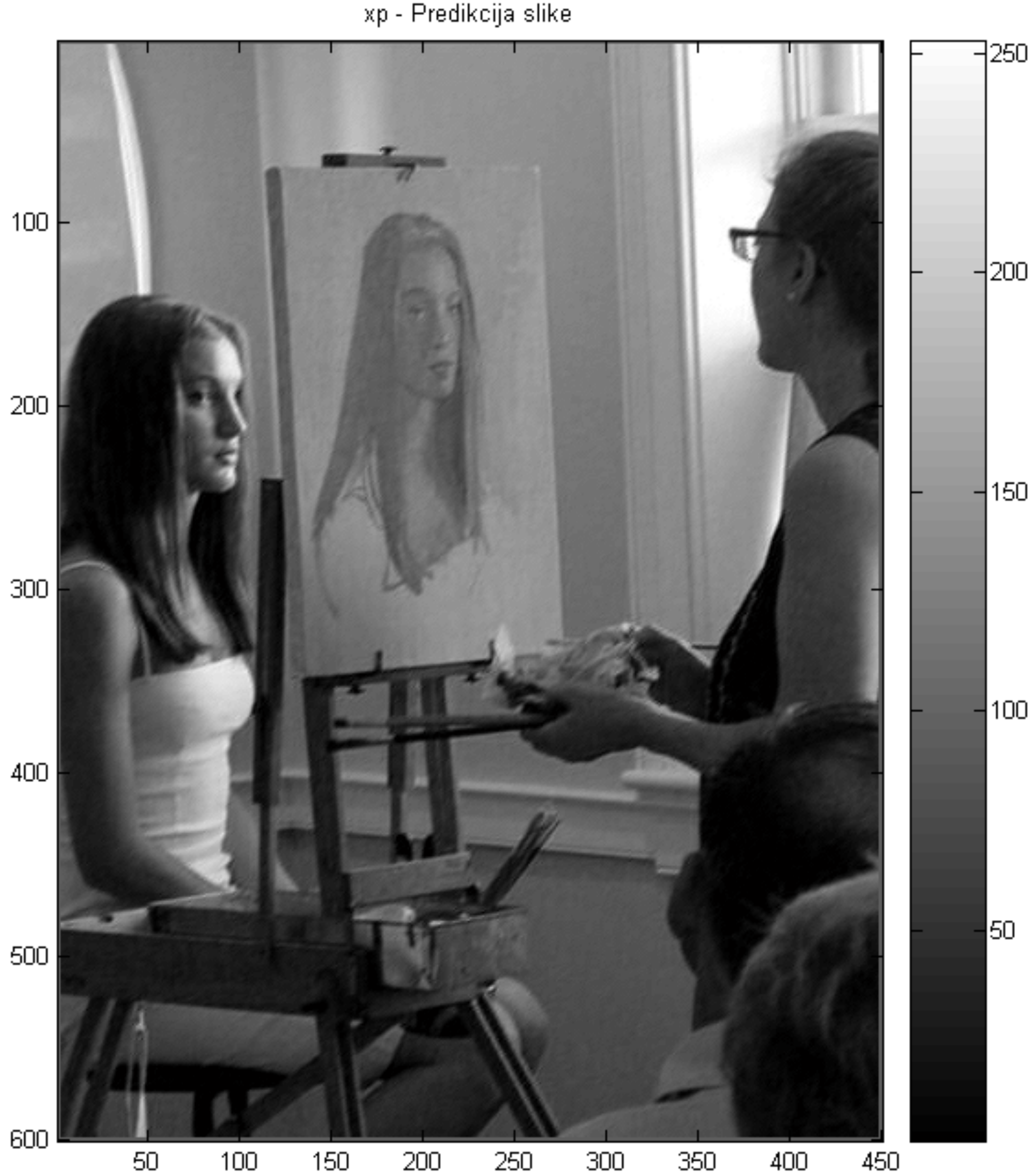
- The predictively coded image using the ADPCM structure is compared to the direct quantization process with the same output entropy.
- The quantizers are of the ECSQ type for both encoders, and the quantization steps are adjusted to achieve the chosen output entropy.
- The output quality is determined by calculating the *SQNR* ratio at the characteristic points of the encoder.
- The program displays images at all points of the coding structure, as well as the corresponding quantization errors, also displayed in the form of monochromatic images using shades of gray.

ADPCM image coding structure – same as for speech!

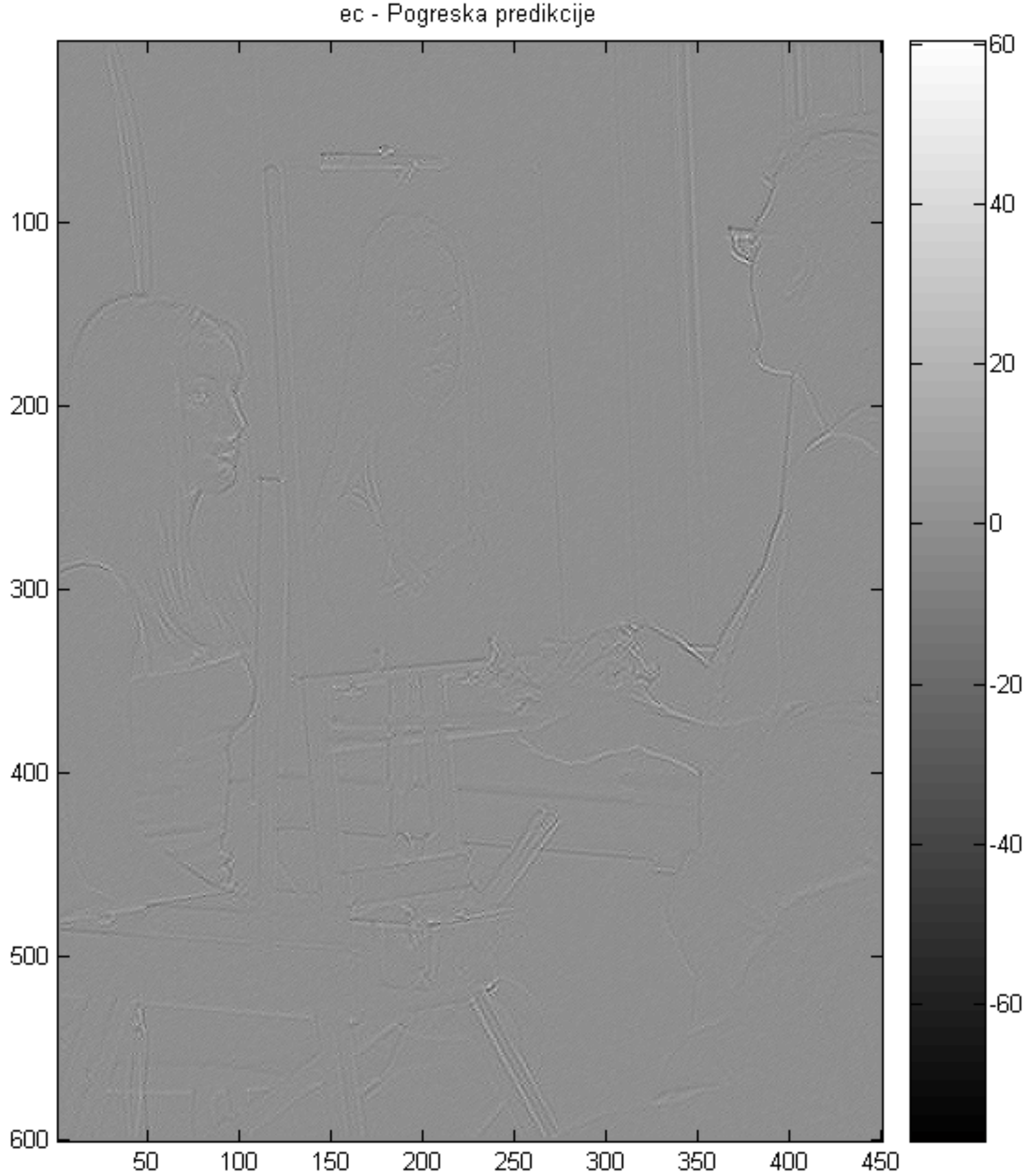


- Input image
  - "demo1"

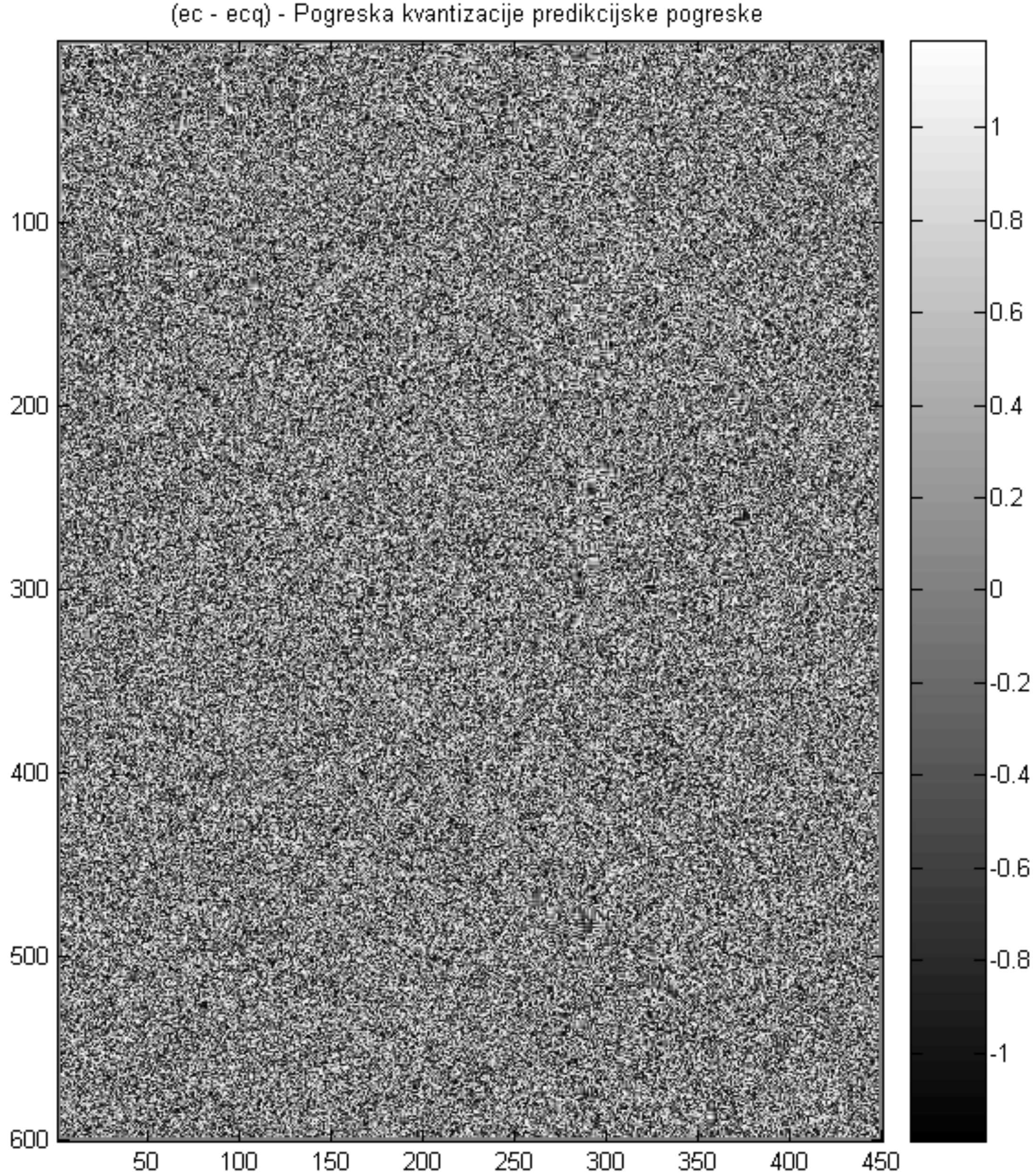
- Prediction on both sides (closed loop) *p*=2



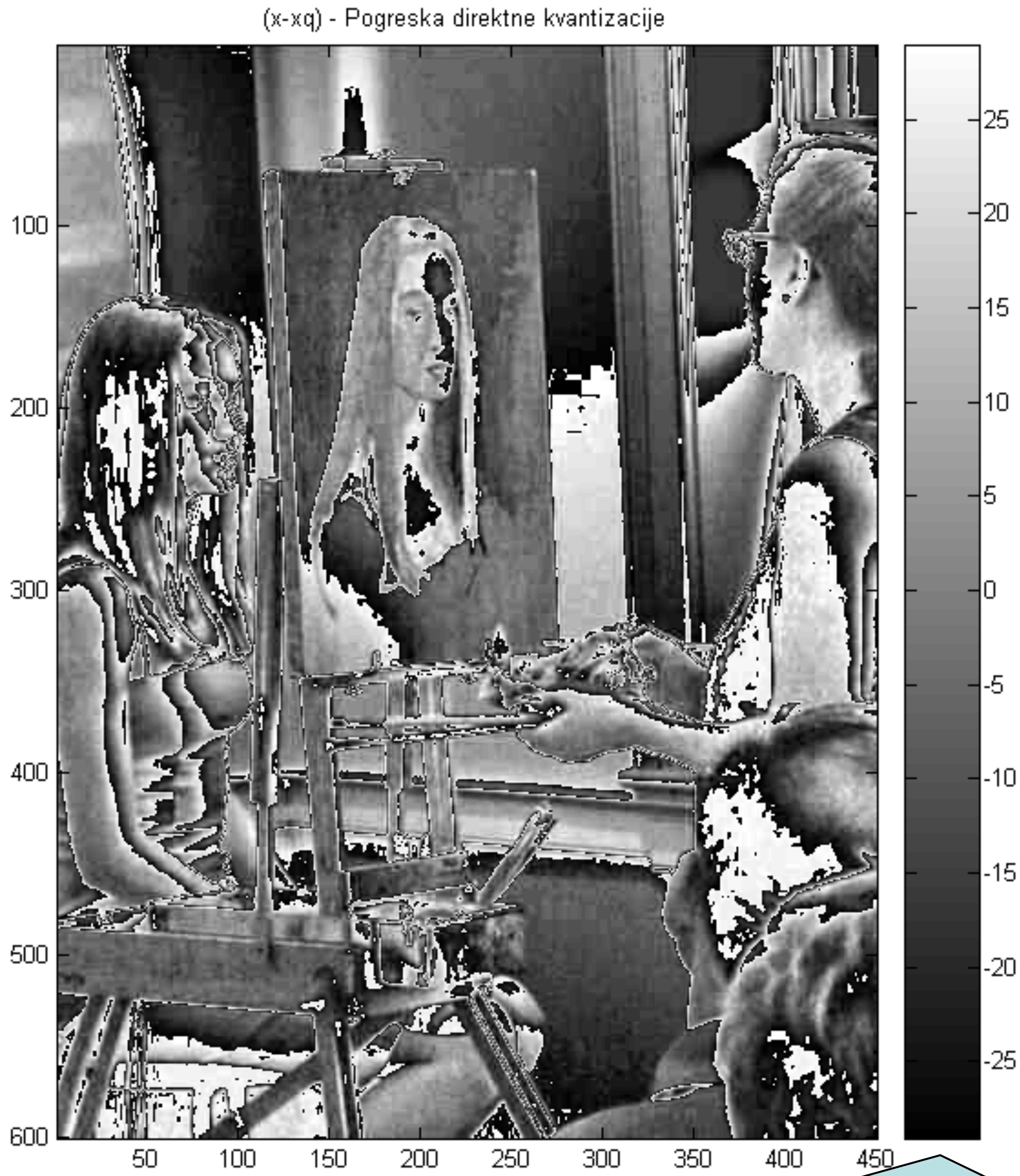
- Closed loop prediction error for p=2
- Notice the edges that could not have been predicted by the predictor!



- Prediction error quantization error , p=2, $H(I)$=2 bit

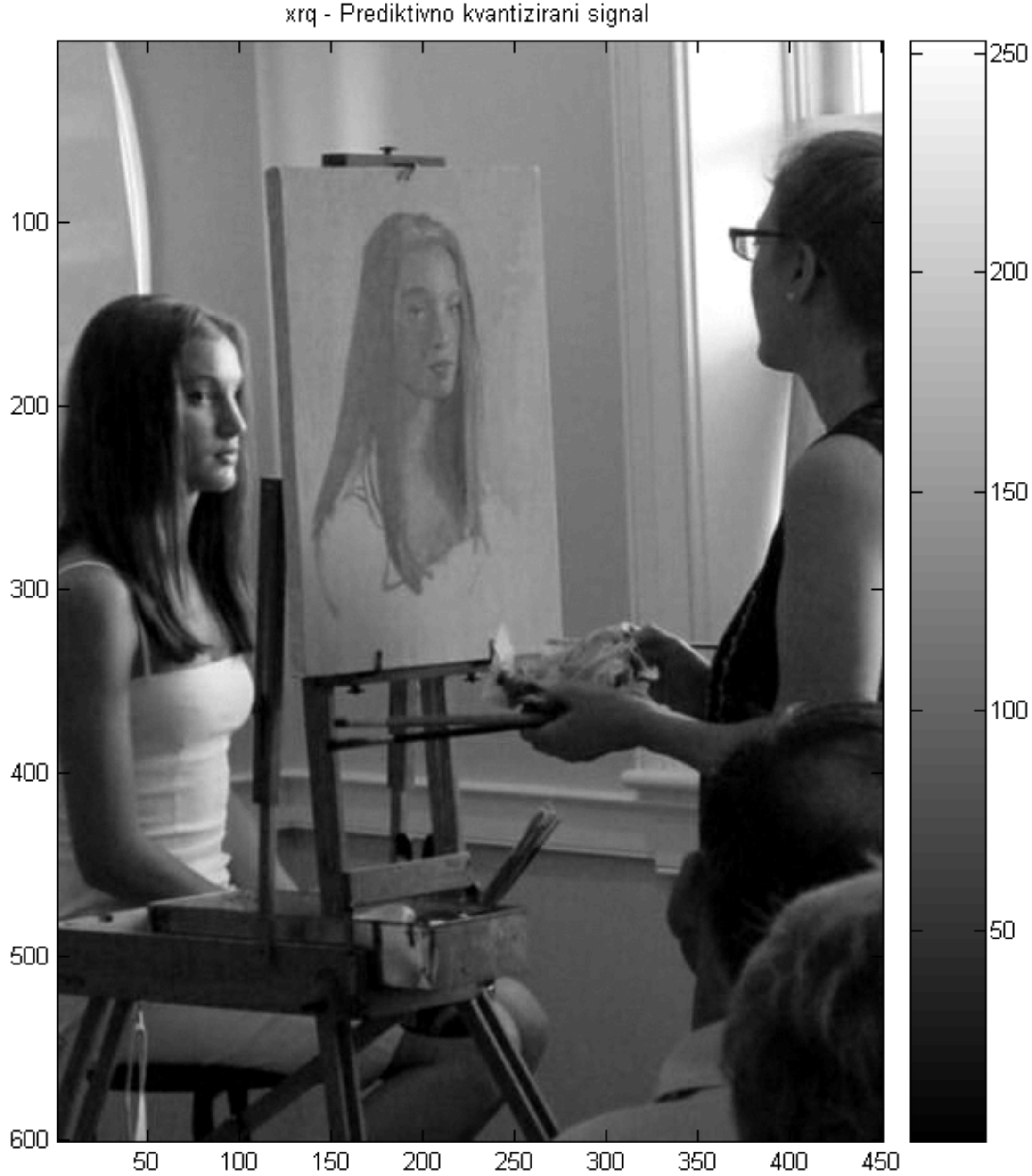


- Direct image quantization error, with *H*(*I*)=2 bit
- The error is within +/- 30 LSB of 8-bit value (0 to 255)



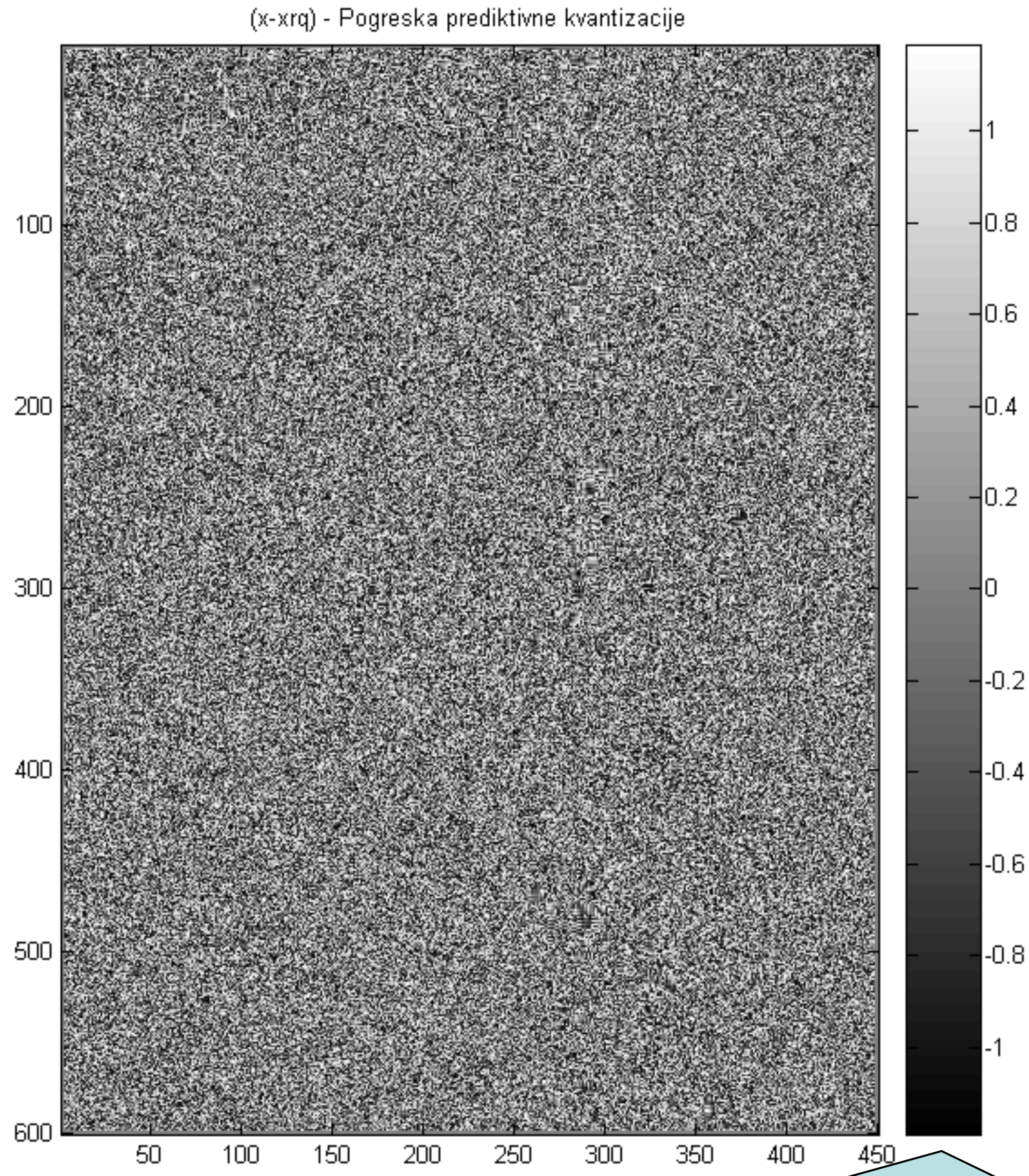
- Directly quantized image with $H(I)$=2 bit

- Predictively coded image p=2, $H(I)$=2 bit



- Predictive quantization error in closed loop, p=2, $H(I)$=2 bit

- The error is within +/- 1 LSB of 8-bit value



Predictive image encoding in a closed loop ... discussion

- Results for example “demo1.jpg” with predictive coding in closed loop with *p*=2 and *H*(*I*)=2:

```
Actual output entropies: H(I_X)=2.014 H(I_E)=2.028
Actual output SNR: SNR(X)=11.040 dB SNR(E)=11.837 dB
Actual output SNR after reconstruction: SNR(Xr)=38.601 dB
Prediction gain: PGotv =27.750 dB and PGzatv =26.764 dB
Quality increase: SNR(Xr)-SNR(X)=27.562 dB
```

- Compare the expected closed loop prediction gain (PGzatv) and the actual increase in the coding quality SNR(Xr) -SNR(X), which are very close.
- The gain is as much as 27.5dB, which corresponds to savings of as much as 4.5 bits per pixel!

What have we learned?

- spatial correlations in the image,
- the notion of causality in the image,
- sequential reading operation,
- causal and noncausal predictors,
- an example of predictive image coding within the ADPCM coder.