Neuro-Rough Control of Masking Thresholds for Audio Signal Enhancement

Andrzej CZYZEWSKI, Rafal KROLIKOWSKI

e-mail: *{andrzej, rafal}@sound.eti.pg.gda.pl*
Sound Engineering Department, Technical University of Gdansk,
ul. Narutowicza 11/12, 80-952 Gdansk, POLAND

Abstract

The paper addresses the problem of neuro-rough hybridisation applied to digital processing of audio signals. Moreover, the application of some selected soft computing techniques to non-stationary noise reduction is described. Some attention is also put to a discussion of the intelligent decision algorithms performance. The noise reduction algorithm is based on the new perceptual approach exploiting some properties of the human auditory system. Furthermore, the paper introduces the engineered perceptual filter driven by an intelligent controller employing rules generated with the use of a rough set-based algorithm supported by a neural network. The goal of the intelligent controller is to estimate the current statistics of corrupting noise on the basis of the analysis of signals received from telecommunication channel. Thereafter, the noise estimate enables determining the masking threshold levels which allow making the noise inaudible in the audio signals. Since the implemented decision algorithm requires quantised data, thus the Kohonen's self-organising maps (SOM) extended by various distance metrics were used as data quantisers. Some results of the experiments in the domain of non-stationary noise reduction in speech are discussed in the paper.

Keywords: Rough sets; Self-organising maps; Non-stationary noise reduction; Vector quantisation

1. Introduction

Noise and other distortions are commonly present in audio signals transmitted via telecommunication channels. They can be introduced in the process of sound acquisition or may occur during the transmission. It should be noticed that sound acquisition systems of communication devices often work in some very poor acoustic conditions. This situation refers to microphone systems of personal computers, conference intercom devices, audio channels of video cameras and hearing aids. Consequently, there is a need for robust and efficient methods of parasite noise reduction in audio signals.

Noise reduction has been a subject of intensive research for dozens of years. The scope of the research encompassed mainly autocorrelation methods [43], methods

based on speech models [25], Wiener and Kalman adaptive filtration [39] and spectral subtraction [2] [39]. A number of various methods have been proposed referring to adaptive filtration [11] [27] [34] [42], which were commonly used in engineering [11] [16] [41]. In turn, spectral subtraction methods turn out to be both robust and simple, and are efficient in restoration of old recordings [40]. Moreover, some techniques exploiting intelligent algorithms have also been proposed recently [8]. However, the common problem related to these methods is audible artefacts becoming annoying when more serious intervention in noisy signal is made. Hence, in the proposed approach to noise suppression some masking properties [46] [47] of the human auditory system, which have been successfully exploited in some contemporary audio coding standards [4] [28] [35], are taken into account in order to remove only audible parts of the noise corrupting audio signals. This approach leads to less inference into a signal, and can presumably result in less annoying artefacts. As was confirmed by results of experiments carried out by the authors [10], the corrupting noise can be efficiently masked (made inaudible) by these audio components which convey most of the useful signal energy. The detailed mathematical foundations of this perceptual approach together with adequate algorithm proposals can be found in one of recent authors' papers [10].

Generally, in all noise reduction methods, there is a need to know at least approximated statistics of the parasite noise. This problem becomes more complex in the case of non-stationary one, since such a method requires to track an altering statistics of the noise or a useful signal in time, and to select some noise statistics from among others. Correspondingly, the need for an efficient decision system occurs, and therefore soft-computing methods were used which proved to be robust in audio signal processing [8] [9] [22] [23]. As regards non-stationary noise, it can be assumed that adequately long train of noise observations can represent the noise in a telecommunication channel. Moreover, this train forming a set of noise estimates can be referred to as a set of facts in a decision table. In this way, the problem of the decision-making can be treated as rough set-based inference task [21] [30] [31]. Furthermore, it can be noticed that the storage of the noise estimates in the decision table corresponds to vector quantisation (VQ) [12] of the altering in time noise patterns. More detailed issues related to VQ and its successful applications to digital speech processing can be found in a very abundant literature [5] [6][26] [37].

Obviously, noise estimate values produce real numbers, whereas an inference exploiting rough sets requires quantised data. As is known by practice, a selection of a quantisation method may influence the quality of the interference [23]. Despite that a number of quantisation methods have been proposed so far [7] [23] [36], the authors' proposal consists in exploitation of self-organising neural networks [45] to this task, which perform both scalar [13] and vector quantisation [15]. In the case of self-organising maps (SOMs) introduced by Kohonen [18] [19], and successfully exploited in many applications [20], a distance measure between an input data vector and competing neurons is of a paramount importance. The Euclidean metric

is applied most commonly, however the SOM algorithm have been extended by a few other distance metrics for the data quantisation purpose.

Finally, the efficiency of applications of neuro-rough hybridisation for the purposes of real audio restoration is assessed in the paper.

2. Some Principles of Psychoacoustics

One of important notions in psychoacoustics are critical bands. Their concept is related to propagation and processing of acoustic signals in the human auditory system. Well-proven experimental data reveal that the inner ear behaves as a bank of band-pass filters which analyse a broad spectral range in subbands, called critical bands, independently from others. A perceptual unit of frequency - Bark - has been introduced, and it is related to the width of a single subband. A commonly used transformation to this subjective scale of hearing is defined by the following non-linear relation proposed by Zwicker [46]:

$$b = 13 \cdot \text{arctg}\left(0.76 \cdot 10^{-3} \cdot f\right) + 3.5 \cdot \text{arctg}\left[(f/7500)^2\right], \tag{1}$$

where b, f denote frequency in Barks and in Hz, respectively.

As far as the critical bands are considered, it should be mentioned about the mel-scale which belongs also to perceptual scales of frequency. The mel-scale of auditory pitch was established basing on experiments on perception of simple tones (sinusoid), and is closely related to the critical bands. Another psychoacoustic phenomenon is related to masking which can take place in the time- as well as in the frequency domain (simultaneous masking). While masked, some tones can be inaudible in the presence of others called maskers. As was mentioned, this phenomenon is fundamental for contemporary audio coding standards, although it can be also exploited in noise reduction [10]. The simultaneous masking plays the most important role, and therefore is commonly applied. It is featured by the fact that some tones can become inaudible, especially when at least one of them is louder, and their frequencies are not too distant. In general, this relationship is dependent on intensities of maskers and masked tones, and their frequencies. This relationship is described in the spectral domain by so called masking curves which are defined for maskers of given intensity and frequency. All components laying below these curves are masked and hence become inaudible. The shape of the masking curves is complex when plotted vs. Hertz frequency, and therefore almost useless in engineering applications. However, they become almost uniform when computed vs. frequency expressed in Bark units as is presented in Fig. 1. More details can be found in psychoacoustics literature [46] [47].
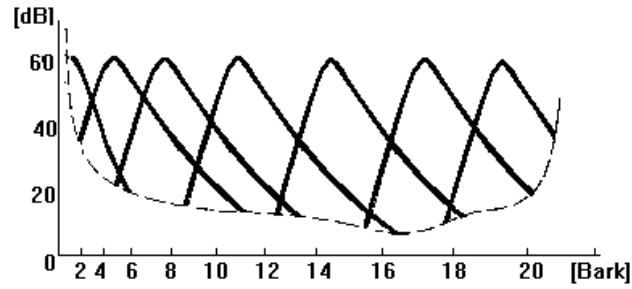
Fig. 1. Exemplary masking curves plotted versus Bark frequency [47]

3. Description Of The Perceptual Noise Reduction System

3.1. General Scheme of the System

The general lay-out of the perceptual noise reduction system is shown in Fig. 2. Two inputs of the system represent: the noise patterns $\tilde{n}(m)$ and the noisy signal $y(m)$ which consists of the original audio signal $x(m)$ corrupted by the noise $n(m)$. All distortions such as those originated from the sound acquisition process (environmental noise) or introduced during the transmission influence the useful audio $x(m)$. And these distortions constitute the noise $n(m)$. The corrupted signal $y(m)$ is transformed to the spectral representation $Y(j\omega)$ with the use of the Digital Fourier Transform (DFT) procedure. Since it is impossible to obtain the copy of noise degrading audio $x(m)$, some noise patterns $\tilde{n}(m)$ are taken from these passages of a signal transmitted in the telecommunication channel the useful information is not transmitted. It is assumed that noise patterns are correlated with the noise $n(m)$ and therefore they are its estimates. The signal $\tilde{n}(m)$ is delivered to the Noise Estimation Module which task is to collect essential information on the noise $n(m)$ in some given time intervals. The patterns $\tilde{n}(m)$ are analysed in the spectral domain, and as a result, two kinds of vectors are obtained: the short-term average power spectrum and the associated vector of coefficients related to this spectrum. Subsequently, these both vectors are stored in a table (codebook). Thus, the codebook collects changes of the noise patterns $\tilde{n}(m)$ statistics in time, and its contents can be referred to as a time-frequency noise estimate $\rho(t, j\omega)$. Both this estimate $\rho(t, j\omega)$ and the spectrum of the corrupted audio $Y(j\omega)$ are supplied to the Decision Systems. Their first task is to select one of the collected spectral estimates $\rho(j\omega) \subset \rho(t, j\omega)$ that is correlated best to the corrupting noise in a given moment of time. The mentioned selection employs a neuro-rough decision system that uses the estimate $\rho(t, j\omega)$ for the training, and returns the number of the estimate $\rho(j\omega)$ in the codebook. The output of the decision-making *index* is provided back to the codebook in order to retrieve the selected noise estimate

$\rho(j\omega)$ from this codebook. The second task of the module is to qualify the elements of the signal $Y(j\omega)$ for two disjoint sets: the set U of the useful or the set D of the useless elements in order to know, which spectral components are maskers (useful components), and which ones are to be masked (useless components). All spectral components which are to remain audible and used to mask others are called hereafter useful ones. They are assumed to convey most of the audio signal $x(m)$ energy. In turn, these elements that are to be masked will be further called useless, and they are assumed to convey mostly the parasite noise energy. This decision-making process is based on an IF-THEN reasoning. Thus, the spectral components of the corrupted signal $Y(j\omega)$ belong either to the set U or the set D, and in the case of the use of the N-point DFT the following condition is fulfilled:

$$N/2 = |U| + |D|, \tag{2}$$

where $|U|$, $|D|$ are cardinalities of sets containing useful and useless elements, respectively.

The spectrum of the corrupted signal $Y(j\omega)$ as well as the sets U, D and the selected noise estimates $\rho(j\omega)$ are fed to the Perceptual Noise Reduction Module that executes a perceptual algorithm of noise reduction. Next, the output $\hat{X}(j\omega)$ is processed by the inverse DFT procedure, and finally the restored signal $\hat{x}(m)$ is obtained, which is subjectively perceived as less noisy than the original one.
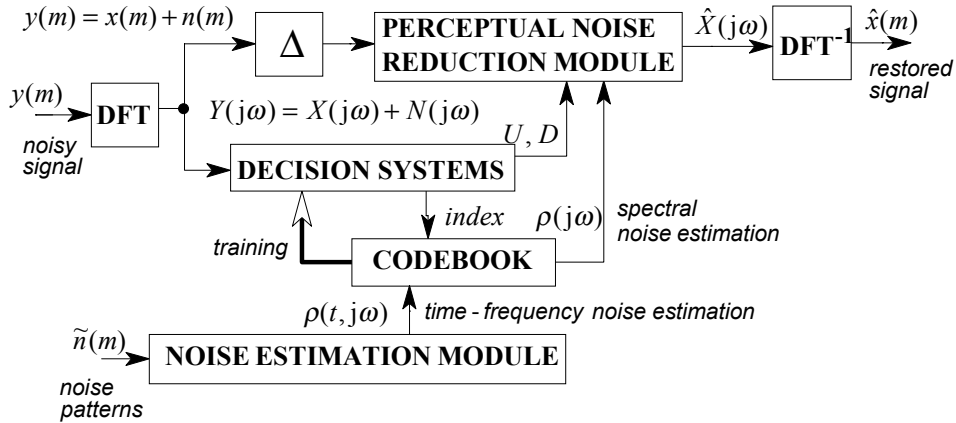

Fig. 2. General lay-out of the noise reduction system

3.2. Noise Estimation Module

As was mentioned, in the Noise Estimation Module (Fig. 2) a vector quantisation of altering in time noise patterns is performed. First, the patterns $\tilde{n}(m)$ are transformed into the spectral domain with the use of the DFT procedure. Next, the spectral representation of the patterns $\tilde{N}(j\omega)$ are fed to the *Spectral Averaging* block as well as to the *Codevector Computation* block. In the first one, the spectrum $\tilde{N}(j\omega)$ is averaged upon subsequent L frames, and at the output the average power spectrum $\hat{N}_k$ is obtained, which can be referred to as a vector of spectral power values for consecutive frequency components. In turn, in the *Codevector Computation* block the spectrum $\tilde{N}(j\omega)$ is analysed, and as a result the associated vector (codevector) $V_k^{\tilde{n}}$ of coefficients related to the spectrum $\hat{N}_k$ is computed. In the above denotations ($\hat{N}_k$ and $V_k^{\tilde{n}}$), the index k represents the time interval within which elements of these vectors are computed. Subsequently, the both vectors are delivered to the *Noise Estimation Build Up* block which sequences these data in the form of the time-frequency noise estimates $\rho(t, j\omega)$. In result, the vectors are collected in the codebook, which, in other words, can store statistics of noise changes $\tilde{n}(m)$ in time.
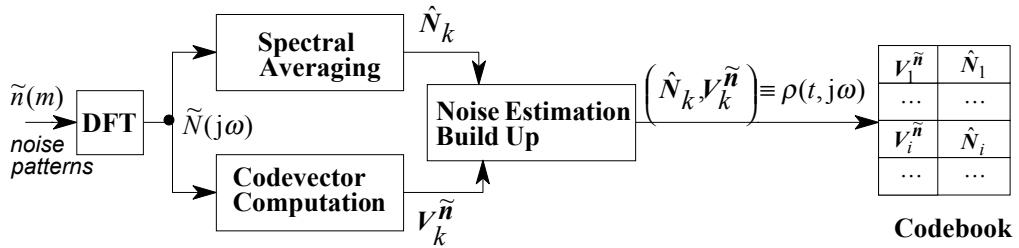

Fig. 3. Scheme of the Noise Estimation Module connected to the Codebook

In the case of use of the *N*-point DFT the vector $\hat{N}_k$ is defined as below:

$$\hat{N}_k = \begin{bmatrix} \hat{N}_{1,k} & \dots & \hat{N}_{n,k} & \dots & \hat{N}_{N/2,k} \end{bmatrix}^T, \tag{3}$$

where the *n*-th element $\hat{N}_{n,k}$ is averaged on the basis of last L values of the spectral power $\tilde{N}_n$ according to the following formula:

$$\hat{N}_{n,k} = \frac{1}{L} \cdot \sum_{l=k-L+1}^{k} \widetilde{N}_n^{(l)} \tag{4}$$

In turn, the vector $V_k^{\widetilde{n}}$ serves as an associated codevector, and thus this vector should be unique, however in practice it's uniqueness is hard to be fulfilled. Its elements are expected to represent quantitatively the noisy character of the average spectrum $\hat{N}_k$. Therefore two kinds of parameters are considered that turned out to be very robust in contemporary perceptual audio coding schemes: the *Spectral Flatness Measure* [17] [44] and the *Unpredictability Measure* [3] [28]. In these schemes, these vectors are exploited to estimate the amount of noisy (or non-tonal) components in a given moment of time, and due to this they are used in computation of the current masking threshold.

For the purposes of the codevector computation in the engineered noise reduction system, these parameters are computed in some subbands, and the width of a single subband can be constant or it can be related to a critical band. Definitions of these parameters for the *l*-th frame are given below.

- Application of Spectral Flatness Measure

  The *SFM* parameter is defined as the ratio of the geometric to the arithmetic mean of the power spectrum [17], and is expressed in dB. In the *b*-th subband, the parameter can be redefined as follows:

$$SFM_b^{(l)} = 10 \cdot \log_{10} \frac{G_m^{(l)}}{A_m^{(l)}} = 10 \cdot \log_{10} \frac{\left[\prod_{i=lower(b)}^{upper(b)} S_i^{(l)} \right]^{1/count(b)}}{\frac{1}{count(b)} \cdot \sum_{i=lower(b)}^{upper(b)} S_i^{(l)}}, \tag{5}$$

where S_i is the spectral power of the *i*-th frequency component, which is obtained by means of the *N*-point short-time Discrete Fourier Transform, whereas *lower*(*b*) and *upper*(*b*) denote indexes of the first and the last spectral component in the *b*-th subband which contains *count*(*b*) components.

Hence, the vector $V_k^{\widetilde{n}}$ can be defined in the following way:

$$V_k^{\widetilde{n}} = \begin{bmatrix} SFM_{1,k} & \dots & SFM_{b,k} & \dots & SFM_{B,k} \end{bmatrix}^T, \quad \text{where:} \tag{6a}$$

$$SFM_{b,k} = \frac{1}{L} \cdot \sum_{l=k-L+1}^{k} SFM_b^{(l)} \tag{6b}$$

- Application of the Unpredictability Measure

Introducing denotations of the spectral magnitude prediction $\hat{r}_i^{(l)}$ and the phase prediction $\hat{\phi}_i^{(l)}$ of the i-th spectral component on the basis of their last two real values as below:

$$\begin{cases} \hat{r}_i^{(l)} = r_i^{(l-1)} + (r_i^{(l-1)} - r_i^{(l-2)}) \\ \hat{\phi}_i^{(l)} = \phi_i^{(l-1)} + (\phi_i^{(l-1)} - \phi_i^{(l-2)}) \end{cases} \Rightarrow \begin{cases} \hat{r}_i^{(l)} = 2 \cdot r_i^{(l-1)} - r_i^{(l-2)} \\ \hat{\phi}_i^{(l)} = 2 \cdot \phi_i^{(l-1)} - \phi_i^{(l-2)} \end{cases}, \qquad (7)$$

the unpredictability measure $c_i^{(l)}$ is defined as the Euclidean distance between the real values of $r_i^{(l)}$, $\phi_i^{(l)}$ and the predicted ones of $\hat{r}_i^{(l)}$, $\hat{\phi}_i^{(l)}$ according to the formula [3]:

$$c_i^{(l)} = \frac{\sqrt{\left(r_i^{(l)} \cdot \cos\phi_i^{(l)} - \hat{r}_i^{(l)} \cdot \cos\hat{\phi}_i^{(l)}\right)^2 + \left(r_i^{(l)} \cdot \sin\phi_i^{(l)} - \hat{r}_i^{(l)} \cdot \sin\hat{\phi}_i^{(l)}\right)^2}}{r_i^{(l)} + \left|\hat{r}_i^{(l)}\right|} \qquad (8)$$

In such a case, the vector $V_k^{\tilde{n}}$ can be described as below:

$$V_k^{\tilde{n}} = \begin{bmatrix} C_{1,k} & \dots & C_{b,k} & \dots & C_{B,k} \end{bmatrix}^T, \qquad (9)$$

where the element $C_{b,k}$ is calculated for the b-th critical band and averaged upon last L frames in the following way:

$$C_{b,k} = \frac{1}{L} \cdot \sum_{l=k-L+1}^{k} C_b^{(l)}, \quad \text{where:} \quad C_b^{(l)} = \frac{1}{count(b)} \cdot \sum_{i=lower(b)}^{upper(b)} c_i^{(l)} \qquad (10)$$

3.3. Decision Systems

The detailed scheme of the Decision Systems is shown in Fig. 4. The module is fed by the spectral representation of the noisy signal $Y(j\omega)$. First, the input signal is processed in the *Codevector Computation* block which task is to obtain a codevector V_i^y which parameters are expected to represent the noisy character of the input $Y(j\omega)$. Therefore elements of this vector are defined by analogy to the codevector $V_k^{\tilde{n}}$ elements, and hence computed as in the formulae (5)-(10). The vector V_i^y is next supplied to the *Decision System I* which objective is to provide

the *index* value of the noise spectrum $\hat{N}_j$ (the estimate $\rho(j\omega)$) available in the codebook, that should be mostly correlated to the noise present in the noisy audio signal $Y(j\omega)$. Having received the desired vector $\hat{N}_j$ from the codebook, this estimate is compared with the spectral representation $Y(j\omega)$ in the *Decision System II* which produces two output sets: the set U of useful and the set D of useless components.
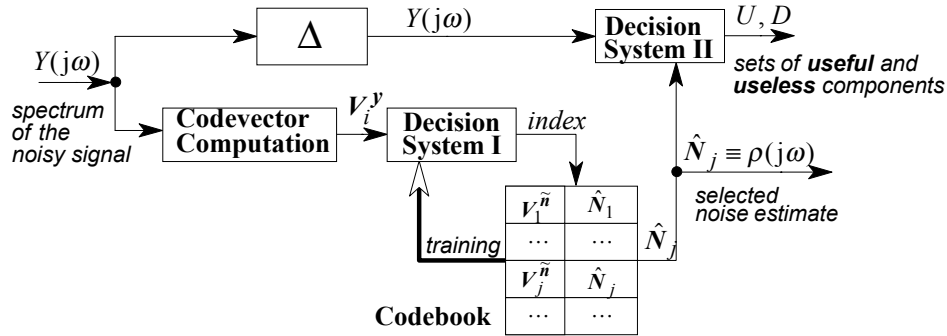

Fig. 4. Scheme of the Decision Systems connected to the Codebook

3.3.1. Implementation of the Decision System I

As was mentioned, collected vectors in the *Codebook* can be treated in the natural way as a decision table, and as a consequence - this table can be processed according to the rough set principles [21] [31]. In turn, rough set inference requires quantised data (in the *Codebook*), and for this purpose scalar- and vector quantisation is performed which is based on self organising maps (SOM). Hence, decision-making in the *Decision System I* is based on rough set inference which is additionally supported by the SOM. Thus, it can be referred to as a neuro-rough decision system. Although there are a few advanced SOM-based algorithms for vector quantisation such as LVQ1, LVQ2 and LVQ3 [18], they have been neglected in the practical approach. The LVQ1-3 algorithms are exploited for fine tuning of clustering procedure, which is unnecessary and even can be pernicious in the *Decision System I*. It was caused by the fact that the quantisation does not concern the noise $n(m)$ estimate but the noise estimates $\tilde{n}(m)$, and the additional fine tuning may lead to the selection of less accurate estimates than in the case of the SOM algorithms.

The run of the *Decision System I* can be divided into two modes: the training mode and the execution mode. In the first case, the content of the *Codebook* is exploited, which is depicted in Fig. 4 with the white arrow. In the training mode, related to rule discovery, a part of the *Codebook* is treated as a decision table, where elements of the key vector $V_k^{\tilde{n}}$ defined by the formulae (6) or (9) serve as conditional

attributes, and the vector's index in the *Codebook* is a decision attribute. Therefore the *k*-th object in the table (codebook) may be described by the following relation:

$$SFM_{1,k} \,,... \,, SFM_{b,k} \,,... \, SFM_{B,k} \Rightarrow k, \quad \text{or} \tag{11a}$$

$$C_{1,k} \,,... \,, C_{b,k} \,,... \, C_{B,k} \Rightarrow k, \tag{11b}$$

where the parameters: *SFM, C* are computed according to the expressions (6) and (10), respectively.

It can be noticed that only conditional attributes require quantisation and for this purpose the SOM-based quantiser is introduced. In the execution mode, the input vector of noisy audio parameters V_i^y is quantised, and next processed by the set of generated rules. In result, the *index* value of the noisy spectrum $\hat{N}_j$ in the *Codebook* is obtained.

Despite a number of various rough set-based inference algorithms have been developed by now [14] [29] [38], it was decided that the rule discovery procedure in the *Decision System I* will exploit the algorithm proposed by the authors [9]. More details on the neuro-rough algorithm are described in Par. 4.

3.3.2. Implementation of the Decision System II

In the *Decision System II*, the division into useful and useless elements is executed according to the following simple procedure. All these components which spectral powers *Y* exceed the double average value of the representative noise estimate $\hat{N}_j$ are assumed to be the useful elements. In turn, the remaining components are regarded as useless ones. Hence, in the case of use of the *N*-point DFT the sets *U* and *D* can defined as follows:

$$
\begin{aligned}
U &= \left\{ n, Y_n : Y_n \geq 2 \cdot \hat{N}_{n,j} \quad \text{and} \quad n = 1,...,N/2 \right\} \\
D &= \left\{ n, Y_n : Y_n < 2 \cdot \hat{N}_{n,j} \quad \text{and} \quad n = 1,...,N/2 \right\}
\end{aligned} \tag{12}
$$

where $\hat{N}_{n,j}$ is the average value of the noise estimate for the *n*-th spectral component in the *j*-th time interval (frame).

In general, it is possible to conceive many various methods of such a division [8] [9], and a selection of one of them has a significant influence on the subjective quality of restored audio signal.

3.4. Perceptual Noise Reduction Module

The task of the module is to process the spectral representation of the noisy signal as follows. All useful spectral components are reduced according to the spectral

subtraction principles [2] [39], whereas the remaining useless components are masked using the psychoacoustic masking approach. The noise suppression can be obtained either by uplifting the masking threshold above the level of the noise spectral power or by reducing this spectral power of noise to the level just below the masking threshold. However, this perceptual approach is a separate complex issue which is not related to the application of intelligent techniques. More details on the applied perceptual models, engineered methods and corresponding algorithms can be found in one of authors' publications [10].

4. Neuro-Rough Decision System

4.1. General Concept of the Neuro-Rough Decision System

As was pointed out, the part of the *Codebook* related to the key vector $V_k^{\hat{n}}$ can be considered as a decision table according to eq. (11). However, elements of the vector are real numbers, whereas rough set-based processing requires quantised data [21]. For quantisation purposes, self-organising neural net is proposed. From among a number of such networks [45], a self-organising map (SOM) introduced by Kohonen [18] [19] has been chosen which has proved to be useful in many engineering applications [20]. Usually, SOM algorithms are based on a distance measurement using the Euclidean metric. In the proposed neuro-rough algorithm which general flowchart is presented in Fig. 5, the standard SOM algorithm can be enhanced by the application of one of the following additional metrics: the inner product, the Manhattan metrics (L_1) and the Chebyshev metrics (L_∞). These metrics are reported also to be applicable to SOMs.
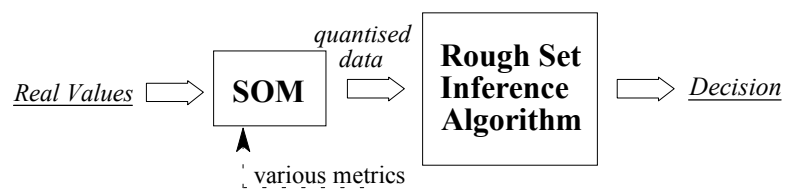

Fig. 5. Functional scheme of the neuro-rough decision system

Since the SOM is supposed to act as a data quantiser, two variants should be considered: a scalar- and a vector quantisation. In the first case, the SOM is supplied to a single element of the key vector. Owing to computational complexity of the rough set processing, this situation takes place if the dimension (number of attributes) of the vector $V_k^{\tilde{n}}$ is not too large. In turn in the second case, a few

attributes can constitute input vectors, which lowers the number of attributes in the *Codebook*, and thereby enables to avoid a vast number of attribute combinations during the rough set inference. These two approaches have been tested during the experimental phase as is described in par. 5.

4.2. Neural Algorithm Based on Self-Organising Maps

As is commonly known, the SOM of the Kohonen type defines mapping of *N*-dimensional input data onto a two-dimensional regular array of units, and the SOM operation is based on a competition between the output neurons due to any stimulation by the input vector x. As a result of the competition, this *c*-th output unit wins provided the following relations are fulfilled:

$$d(x, W_c) = \max_{1 \le i \le K \times L} d(x, W_i) \quad \text{or} \quad c = \arg \min_{1 \le i \le K \times L} \{d(x, W_i)\}, \tag{13}$$

where $d()$ is a distance between the vector x and the weight vector W_i of the output neuron, whereas $K \times L$ is the dimension of the output layer. The mathematical formulae of various metrics $d()$ are listed in Tab. 1.

The adaptation process can be described in terms of minimisation of an error function $E^{(k)}$, and hence the updating of the weight vectors in the *k*-th step is performed according to the expression:

$$\bigvee_{i=1,\dots,K \times L} W_i^{(k+1)} = W_i^{(k)} + \Delta W_i^{(k)} = W_i^{(k)} - \nabla_W E^{(k)}, \tag{14}$$

where initial values of the weight matrix $W^{(0)}$ are small random values in the range $[-1, 1]$, whereas the definition of the error function $E^{(k)}$ is related to the concept of vector quantisation, and is given by the following formula [20]:

$$E^{(k)} = \sum_{i=1}^{K \times L} h_{ci}^{(k)} \cdot \Psi[d(x, W_i)], \tag{15}$$

where h_{ci} is a spatial neighbourhood kernel for the *c*-th best matching unit. Thus, the expression updating formula can be rewritten as below:

$$W_i^{(k+1)} = W_i^{(k)} - h_{ci}^{(k)} \cdot \frac{\partial}{\partial W} \Psi(x, W_c), \tag{16}$$

for which the adequate derivatives of the function $\Psi(\boldsymbol{x}, \boldsymbol{W}_c)$ are given in Tab. 1., dependently on the various metrics.

Table 1
Definitions of various metrics and their derivatives in the *N*-dimensional space

Metric	$d(\boldsymbol{x}, \boldsymbol{W}_i)$ **Formula**	$\Psi(d)$	$\nabla_W \Psi(\boldsymbol{x}, \boldsymbol{W}_c)$ **Formula**
Euclidean metric	$\|\boldsymbol{x} - \boldsymbol{W}_i\| = \sqrt{\sum_{n=1}^{N} (x_n - W_{in})^2}$	$\frac{1}{2} \cdot d^2$	$-\|\boldsymbol{x} - \boldsymbol{W}_c\|$
Inner product	$\boldsymbol{x} \cdot \boldsymbol{W}_i = \|\boldsymbol{x}\| \cdot \|\boldsymbol{W}_i\| \cdot \cos(\boldsymbol{x}, \boldsymbol{W}_i)$	d	$\boldsymbol{x}^T \cdot \boldsymbol{W}_c$
Manhattan metric (L_1)	$\sum_{n=1}^{N} |x_n - W_{in}|$	$\frac{1}{2} \cdot d^2$	$-\sum_{n=1}^{N} |x_n - W_{cn}|$
Chebyshev metric (L_∞)	$\max_{1 \le n \le N} |x_n - W_{in}|$	$\frac{1}{2} \cdot d^2$	$-\max_{1 \le n \le N} |x_n - W_{cn}|$

In the SOM implementation, the general form of the kernel function h_{ci} is exploited, which is the Gaussian function defined as follows [19]:

$$h_{ci}^{(k)} = \begin{cases} \eta^{(k)} \cdot \exp\left[-\left(\dfrac{d(\boldsymbol{r}_c, \boldsymbol{r}_i)}{\sigma^{(k)}} \right)^2 \right] & ; \text{inside } N_c \\ 0 & ; \text{outside } N_c \end{cases} , \qquad (17)$$

where N_c denotes the set of neighbour nodes around the *c*-th winner neuron, $\boldsymbol{r}_i$ and $\boldsymbol{r}_c$ are the coordinate vectors of the *i*-th unit and the best matching one. In turn, $\eta \in [0,1]$ can be referred to as a learning rate, whereas σ corresponds to the radius of the set N_c, and is limited by the size of the array of the output neurons. The both are decreasing functions of time, which definitions are given below [18]:

• the learning rate η is expressed by the relationship:

$$\eta^{(k)} = \alpha^{(k)} \cdot \left(1 - \frac{k}{k_{\max}} \right), \qquad (18)$$

where the coefficient α varies according to the Kohonen's recommendations in the following way:

$$\alpha^{(k)} = \begin{cases} 0.95 & ;k \leq k_1 \\ 0.9 & ;k_1 \leq k \leq k_2 \\ 0.01 & ;k \geq k_2 \end{cases}, \qquad (19)$$

and the milestones are: $k_1 = 0.02 \cdot k_{max}$, $k_2 = 0.4 \cdot k_{max}$, for which the maximum number of iterations is set to: $k_{max} = \{10000; 50000; 100000\}$.

- the radius σ of the neighbourhood set N_c corresponds to the learning rate η (and the milestones) according to relationship:

$$\sigma^{(k)} = \begin{cases} \lfloor \alpha \cdot k + \beta \rfloor & ;k \leq k_1 \\ \lfloor \chi \cdot k + \delta \rfloor & ;k_1 \leq k \leq k_2 \\ 1 & ;k \geq k_2 \end{cases}, \qquad (20)$$

where the coefficients α, β, χ, δ are computed according to some Kohonen's hints [18], and it can be shown that they are defined as follows:

$$\alpha = -\frac{\sigma^{(0)} - 1}{k_1}, \quad \beta = \sigma^{(0)} \cdot \left(1 - \frac{1}{k_1}\right) - \frac{1}{k_1}, \quad \chi = -\frac{1}{k_2 - k_1}, \quad \delta = \frac{k_2}{k_2 - k_1}, \quad (21)$$

where the initial radius $\sigma^{(0)}$ is equal to the radius of the output array, i.e.:
$\sigma^{(0)} = \lfloor \max(K, L) / 2 \rfloor$.

In the structure of the implemented SOM, the input and output nodes are fully connected, whereas the output units are arranged in the hexagonal lattice. The initial values for the learning rate $\eta^{(0)}$ is equal to 0.95. For the purposes of the neuro-rough hybridisation, at the end of the weight adaptation process the output units should be labelled with some symbols. It is done in order to assign quantised input data to symbols which are to be processed in the rough set inference.

4.3. Rough Set-Based Inference System

The engineered rule induction algorithm is based on the well described in literature rough set methodology [21] [31]. Therefore only some improvements reducing the computational complexity have been introduced by the authors, which are described further in the paragraph. The functional description of the

implemented rough set-based inference algorithm has been presented in one of the recent authors' article [9].

It can be noticed that the basic rough operators (the partition of a universe into classes of equivalence, C-lower approximation of a set X and calculation of a positive region) can be performed more efficiently when objects are ordered. Therefore the proposed algorithm should execute sorting of all objects with respect to a set of attributes. It means that these objects are to be sorted according to the first attribute, than according to the second attribute and so on up to the last one. However, reducing of values of attributes and searching for reducts require so that all combinations of the conditional attributes are analysed. In general case, the decision table should be sorted as many times as is the number of all these combinations, what is computationally ineffective. However, it turns out that once sorted table can be exploited a number of times. In other words, for a given sorted table, the optimal number of sets of attributes A ($A \subseteq C = \{a_1, \ldots, a_i, \ldots, a_{|C|}\}$), subsets of the conditional attributes C, can be analysed. The scheme of the search for this optimal sequence of the subsets A on the basis of the exemplary analysis of five conditional attributes is showed in Fig. 6. Numbers in brackets indicate the order in which procedures *left*() and _*P*() are executed in the procedure *P*().
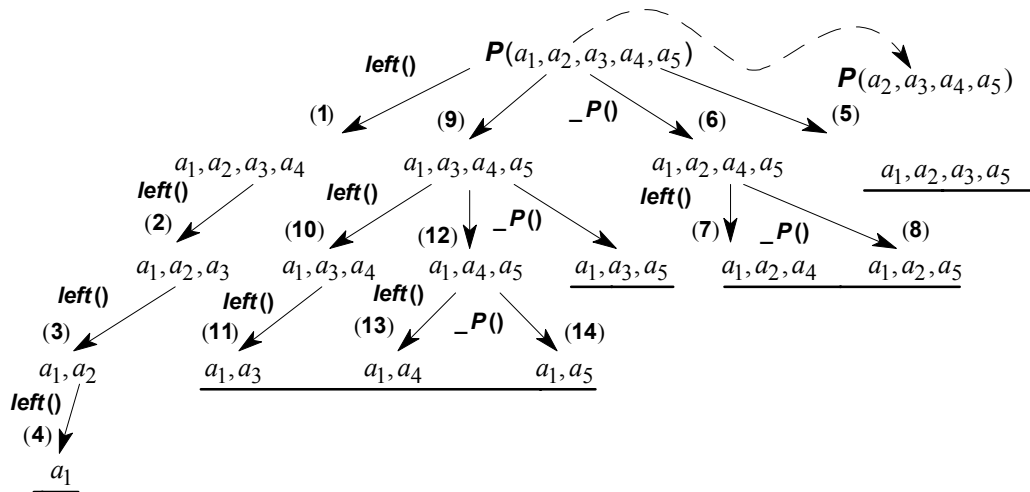

Fig. 6. The optimal sequence of a decision table analysis with regard to a given subset of attributes

Let A be a set of attributes, *depth* be a positive integer related to the recursion depth, and $|X|$ denotes the cardinality of the set X. Then the sequence of the optimal analysis of a decision table is achieved by means of the following procedures given in a pseudocode. Initially, the procedure $P(C)$ is called, where C is the conditional set.

```
{* Recurrent processing of the conditional *}
PROCEDURE P (A : set_of_attributes)
BEGIN
    sortTable(A);
    search for the reduct with respect to A;
    if |A| = 1 then exit;    {* end of the recursion *}
    left(A − a_{|A|},1);
    for i := 2 to |A|-1 do
        _P(A − a_i , |A|−i);
    P(A − a_1);
END
```

```
{* Recurrent processing of the conditional, similar to
   the procedure P() *}
PROCEDURE _P (A : set_of_attributes; depth : integer)
BEGIN
    subsortTable(A,depth);
    search for the reduct with respect to A;
    if |A| = depth then exit;    {* end of the recursion *}
    left(A − a_{|A|},depth);
    for i := 2 to |A|-1 do
        _P(A − a_i , |A|−i);
END
```

```
{* Recurrent processing of the conditional along with
   left branches of the tree from Fig. 6 *}
PROCEDURE left (A : set_of_attributes; depth : integer)
BEGIN
    search for the reduct with respect to A;
    if |A| = depth then exit;    {* end of the recursion *}
    left(A − a_{|A|},depth);
END
```

In the above algorithms, the reducts of the set A are obtained by means of the standard rough set operations which definitions are extensively described in literature [21] [30] [31] [32]. In order to make the algorithms faster, a very quick version of a sorting algorithm, namely quicksort, is exploited. Its computational complexity is of the order equal to $O(n \cdot \log_2 n)$, where n is the number of objects being sorted. The procedure $sortTable(A)$ makes the decision table sorted in a lexicographical order with respect to the set of attributes A. Analogically, the procedure $subsortTable(A, depth)$ sorts the table, however it is assumed that the table is already sorted according to the first $depth$-1 attributes.

5. Experiments

5.1. Experiment Organisation

There were two main objectives of the experiments: first, to examine various intelligent controllers employed as the *Decision System I* and to assess their efficiency; second, to remove non-stationary noise from real audio basing on the chosen decision system. Therefore, the experiments were divided into two stages. For the purposes of the comparison of results, two recordings were made: a male voice plus a non-stationary noise taken from a radio channel. Thereafter, the original audio was mixed with this additive noise and simultaneously, elements of the key vectors $V_k^{\tilde{n}}$ of the noise were computed and stored. Since the recorded sounds were defined, it was possible to determine, which part of the noisy voice was described by which key vector and by which noise spectrum vector.

The parameters of these recordings were as follows. They both were 16-bit mono, sampled with 8192 Hz, which resulted in $B = 18$ critical bands of useful spectrum. The duration of the noisy speech was 5.81 s, and the duration of the noise patterns was 6.23 s. In consequence, the *Codebook* consisted of: 193 objects for the case of which the *Spectral Flatness Measure* was exploited, and 191 objects for the *Unpredictability Measure* application (see par. 3.2).

Let for the sake of simplicity the k-th vector $V_k^{\tilde{n}}$, which the b-th element is defined as in eq. (11), be denoted as $V = \begin{bmatrix} V_1 & \dots & V_b & \dots & V_B \end{bmatrix}^T$. The number of $B = 18$ subbands, i.e. 18 conditional attributes, leads to a vast number of attribute combinations in the rough set inference, and as was pointed out in par. 4.1, there is a need of reduction of number of attributes. Thus, the following two approaches are considered:

- Scalar quantisation

Consecutive elements of k-th vector $V_k^{\tilde{n}}$ in the table are averaged and form one element of the new key vector V^{sq} in the following way:

$$V^{sq} = \left[\frac{V_1 + V_2}{2} \quad \cdots \quad \frac{V_{b-1} + V_b}{2} \quad \cdots \quad \frac{V_{B-1} + V_B}{2} \right]^T, \qquad (22)$$

and for such defined vector V^{sq}, V_i^{sq} denotes its i-th element which is afterwards quantised exploiting the SOM algorithm.

- Vector quantisation

  The vector $V_k^{\tilde{n}}$ can be regarded as a matrix composed of subvectors, and the new vector V^{vq} can be described in terms of these subvectors as follows:

$$V^{vq} = \left[\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad \cdots \quad \begin{bmatrix} V_{b-1} \\ V_b \end{bmatrix} \quad \cdots \quad \begin{bmatrix} V_{B-1} \\ V_B \end{bmatrix} \right]^T \qquad (23)$$

  For such defined vector V^{vq}, V_i^{vq} denotes the i-th row of the matrix V^{vq}, and next is quantised basing on the described SOM algorithm.

The foregoing quantisation (scalar and vector) makes the vectors V^{sq} and V^{vq} are composed of 9 elements. These vectors are processed by the neuro-rough decision system, as is illustrated in Fig. 5, and they correspond to surveys which were denoted as: V^{scalar}, and V^{vector}, respectively. Moreover, for the comparison purposes two additional cases are considered:

- first, input data are averaged as in the case V^{scalar}, next they are quantised with the use of the standard *Equal Interval Width Method*, and thereafter these data are fed to the rough set system. This procedure corresponds to the case V^{EWIM}, and its goal is to verify whether the neural net-based quantisation improves the quality of the system performance;

- second, the decision-making is based on the Euclidean distance between the noisy signal's codevector V_i^y and the codevector $V_k^{\hat{n}}$ in the *Codebook*. Given K objects in the *Codebook*, the j-th codevector wins, which satisfies the following condition:

$$\left\| V^y - V_j^{\tilde{n}} \right\| = \min_{k=1,\ldots K} \left\| V^y - V_k^{\tilde{n}} \right\| \qquad (24)$$

This procedure corresponds to the case V^{direct}, and its objective is to verify whether straightforward methods are more efficient in comparison to the intelligent ones.

In the cases: V^{EWIM}, V^{scalar} and V^{vector}, the decision table consisted of 9 conditional attributes, which made 511 various combinations of subsets of the attributes. In turn, for the case V^{direct}, the codevector was composed of 18 elements.

Taking the above into account, the experiment scheme looks as follows:

I   Preparation: recording of the male voice, the non-stationary noise and the voice corrupted by the noise.

II First stage: examination of the intelligent tools in the *Decision System I*

1. Building up the *Codebook* for the non-stationary noise with respect to various metrics for the SOMs, estimates of noisy features (SFM, Unpredactibility Measure) and the number of SOM training iterations basing on the surveys:

   a) V^{scalar}

   b) V^{vector}

   For all these cases, the size of the output layer $K \times L$ was constant and arbitrary set to 15×15 and next to 20×20.

2. Building up the *Codebook* for the non-stationary noise basing on the V^{EWIM} survey with regard to the length of an interval

3. Building up the *Codebook* for the non-stationary noise basing on the V^{direct} survey.

4. Testing of the *Decision System I* basing on the noisy speech and the various surveys (V^{scalar}, V^{vector}, V^{EWIM}, V^{direct}) - computation and comparison of the decision-making process errors (par. 5.2)

III Second stage: processing of the noisy male speech by the engineered noise reduction system

5.2. Counting Decision-Making Process Errors

For any inference system, based on rule induction, it is important to generate correct decisions with a high probability and predictability. Therefore a function, often called rule quality, is introduced for measurement of these properties. In literature [1] [33], some proposals of such a function can be found, however they are based on statistical relationships between data in a decision table (e.g. Pearson χ^2 statistic, G2 Likelihood Ratio Statistic or on the coverage factor).

The application of the statistical approach is not obvious, if the noise reduction system is considered. First, only estimates of the noise corrupting audio are available, and secondly - the computation of a codevector gives only an approximate noise estimate. Thus, these both issues introduce some additional uncertainty, and the inference process can be referred to as the approximate reasoning. Since the codevectors serve as input data for this reasoning, the definition of the rule quality function should be based on them. Therefore in order to assess the efficiency of the decision system, the quality factor q is proposed by the authors, which takes into account the uncertainty introduced by the codevector computation. This factor represents the efficiency of the decision system (with current error E) with respect to the minimum error E_{opt} and the maximum error E_{max}, and is computed as follows:

$$ q = 1 - \frac{\left| E - E_{\text{opt}} \right|}{E_{\text{max}} - E_{\text{opt}}}, \tag{25} $$

where the error measures E, E_{opt}, E_{max} are defined in the following way:

- The current error measure E

  This measure expresses the current error of the decision-making procedures. It represents the average value of the errors $E^{(i)}$ produced for all I frames of the noisy signal, and is given by the formula:

$$ E = \frac{1}{I} \cdot \sum_{i=1}^{I} E^{(i)} \quad \text{and} \quad E^{(i)} = \sum_{b=1}^{B} \left(V_{b,i}^{y} - V_{b,index}^{\tilde{n}} \right)^2, \tag{26} $$

  where: $V_{b,i}^{y}$ is the b-th element of the vector of parameters of the noisy audio, whereas $V_{b,index}^{\tilde{n}}$ is the b-th element of the key vector which is placed in the position *index* of the *Codebook*. It can be noticed that the value *index* is produced by a decision system.
- The best matching error measure E_{opt}

  This measure reflects the minimum error which occurs when the decision system selects the best matching noise estimate stored in the *Codebook*. Assuming that *i*-th frame of audio is corrupted by the noise described by the *j*-th vector in the table, the measure is expressed by the below formula:

$$E_{\text{opt}} = \frac{1}{I} \cdot \sum_{i=1}^{I} E_{\text{opt}}^{(i)} \quad \text{and} \quad E_{\text{opt}}^{(i)} = \sum_{b=1}^{B} \left(V_{b,i}^{y} - V_{b,j}^{\tilde{n}} \right)^2 \tag{27}$$

- The maximum error $E_{\max}$

  This measure corresponds to the maximum error occurring when the decision system selects the least matching noise estimate. The error $E_{\max}$ is described by the following formula:

$$E_{\max} = \frac{1}{I} \cdot \sum_{i=1}^{I} E_{\max}^{(i)} \quad \text{and} \quad E_{\max}^{(i)} = \max_{k=1,\ldots,K} \left[\sum_{b=1}^{B} \left(V_{b,i}^{y} - V_{b,k}^{\tilde{n}} \right)^2 \right], \tag{28}$$

  where K is the number of vectors in the *Codebook*.

5.3. Results

The results of the comparison tests in the first stage of the experiments are gathered in Tab. 2 - Tab. 7 in which abbreviations: *Euclid*, *Inner*, *Manh* and *Cheb* denote the Euclidean metric, the inner product, the Manhattan- and the Chebyshev metric, respectively. Moreover, in the tables various methods of noise feature estimates are taken into account: the *Spectral Flatness Measure* and the *Unpredictability Measure* (see par. 3.2). The test results are described by the quality factor q, defined by eq. (25).

Table 2

Results of the case survey V^{scalar} for 15 x 15 SOM's output nodes

noise estim.	SFM				Unpredictability Measure			
$k_{\max}$	*Euclid.*	*Inner*	*Manh.*	*Cheb.*	*Euclid.*	*Inner*	*Manh.*	*Cheb.*
10000	69.59	67.13	69.59	69.59	71.46	69.78	71.46	71.46
50000	72.00	71.57	72.00	72.00	75.87	73.43	75.87	75.87
100000	74.62	72.91	74.62	74.62	76.01	74.99	76.01	76.01

Table 3

Results of the case survey V^{scalar} for 20 x 20 SOM's output nodes

noise estim.	SFM				Unpredictability Measure			
$k_{\max}$	*Euclid.*	*Inner*	*Manh.*	*Cheb.*	*Euclid.*	*Inner*	*Manh.*	*Cheb.*
10000	68.43	67.84	68.43	68.43	73.22	65.49	73.22	73.22
50000	71.12	70.03	71.12	71.12	74.99	70.46	74.99	74.99

| 100000 | 72.86 | 71.71 | 72.86 | 72.86 | 77.87 | 73.50 | 77.87 | 77.87 |

Table 4

Results of the case survey V^{vector} for 15 x 15 SOM's output nodes

noise estim.	SFM				Unpredictability Measure			
k_{max}	Euclid.	Inner	Manh.	Cheb.	Euclid.	Inner	Manh.	Cheb.
10000	69.59	72.68	74.12	49.92	76.49	61.00	77.56	55.19
50000	84.91	82.23	83.99	53.87	86.17	85.74	87.33	53.26
100000	85.26	81.56	84.78	50.39	88.20	87.36	88.03	53.52

Table 5

Results of the case survey V^{vector} for 20 x 20 SOM's output nodes

noise estim.	SFM				Unpredictability Measure			
k_{max}	Euclid.	Inner	Manh.	Cheb.	Euclid.	Inner	Manh.	Cheb.
10000	66.28	76.28	75.12	52.33	75.09	64.73	78.98	54.26
50000	74.86	83.98	84.65	55.00	87.02	86.21	87.58	58.51
100000	86.06	85.16	84.78	58.38	88.99	89.43	88.72	59.10

Table 6

Results of the case survey V^{EWIM} (Δ denotes the interval width)

Δ	SFM	Unpredictability Measure
1	75.22	74.48
0.5	78.43	79.01

Table 7

Results of the case survey V^{direct}

SFM	Unpredictability Measure
54.37	57.99

The obtained results reveal that in general, the hybridisation of the SOM-based quantisation and rough set inference gives more accurate decisions than obtained in systems without such a hybridisation. Moreover, exploitation of the soft-computing methods improves considerably the efficiency of the decision system. Comparison between Tab. 2-3 and Tab. 4-5 shows that the vector data quantisation is more efficient than the scalar quantisation. Moreover, it turns out that the selection of an adequate metric considerably influences the quality of the SOM algorithm.

According to the results, the Euclidean metric seems to be the best one, although the Manhattan metric also proved to be efficient . It should be noticed that the latter metric is less computationally complex than the Euclidean one. In turn, the Chebyshev metric turned out to be nearly useless in the proposed neuro-rough decision algorithm. Moreover in the case of the case survey V^{scalar}, the Euclidean-, the Manhattan- and the Chebyshev metrics are compatible, since the input is reduced to one dimension. The results also suggest, that the application of the unpredictability measure improves the decision accuracy, however the computational complexity increases considerably in this case. According to some listening test, in some cases the introduction of the SFM parameter can improve results. Furthermore, it can be noticed that the size of the output layer of the SOM influences the quality factor. In general way, selection of the 20 x 20 output nodes improves the factor.

In the second stage of the experiments, the noisy male speech was processed by the engineered noise reduction system. The efficiency of the noise suppression is demonstrated by means of a time-frequency analysis in the sonograms (Fig. 7), where the horizontal axis corresponds to the time scale (in ms), the vertical axis refers to the frequency scale (in Hz), and the grey intensities are related to the spectral power of the signals' components. Fig. 7a shows the sonogram of the original signal corrupted by non-stationary noise, whereas Fig. 7b illustrates the time-frequency analysis of restored signal.

Usually for evaluation of a ratio of noise suppression, so called *Signal-to-Noise Ratio* (*SNR*) is applied, which is defined as follows [24]:

$$SNR = 10 \cdot \log_{10} \frac{\sum_{m=0}^{M-1} \hat{x}^2(m)}{\sum_{m=0}^{M-1} \hat{n}^2(m)} = 10 \cdot \log_{10} \frac{\sum_{m=0}^{M-1} (y(m) - \hat{n}(m))^2}{\sum_{m=0}^{M-1} \hat{n}^2(m)}, \qquad (29)$$

where $y(m)$ is a noisy signal, $\hat{n}(m)$ is an estimate of the corrupting noise, $\hat{x}(m)$ is an estimate of the useful signal, and *M* is the number of signals' samples.

However, in case of the perceptual noise reduction, some portion of noise is intentionally let to remain in a restored signal, although it is inaudible. Thus, the *SNR* cannot serve as a reliable measure of signal quality in this case. Therefore only subjective listening tests are applicable, and during the experiments such tests revealed a considerable reduction of subjectively perceived noise.
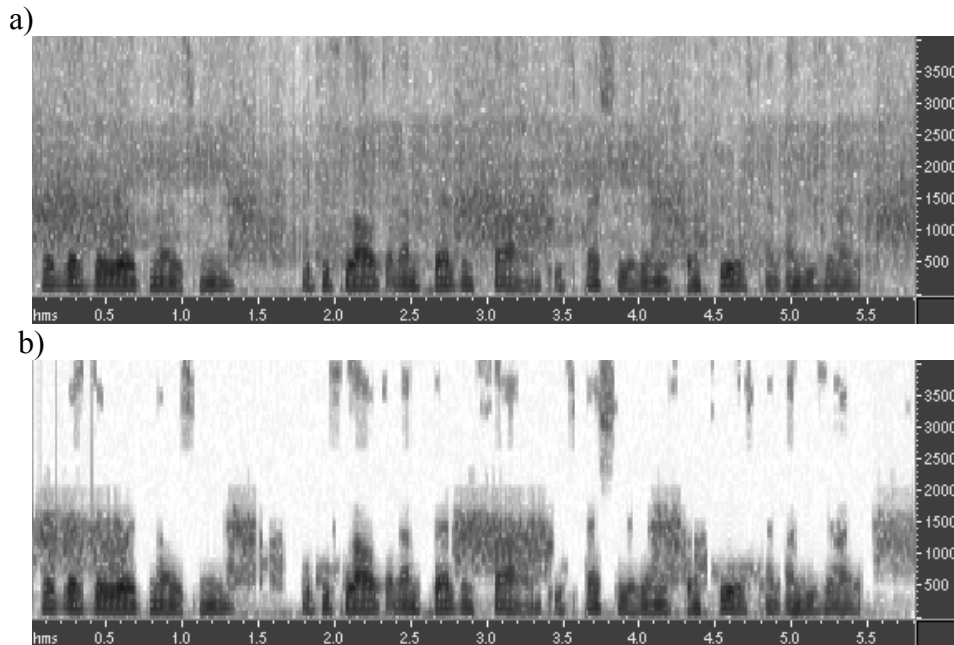
a)

b)

Fig. 7. Sonograms of: (a) noisy male voice, (b) restored speech

6. Conclusions

The engineered system for non-stationary noise reduction has been described. Two novel approaches were applied in this system, namely the auditory masking to noise suppression and the implementation of the neuro-rough controller to determination of non-stationary noise statistics estimate. A number of experiments and case surveys were carried-out with respect to various metrics, scalar- and vector quantisation applications. Different approaches to estimate of noise features were also taken into account. The obtained results show that the rough set inference system supported by the SOM-based quantisation is a robust tool for non-stationary noise reduction. A drawback of this solution is a serious computational complexity related to training of the neuro-rough algorithms. Nevertheless, as was proved by listening tests, the soft-computing approach is fully applicable to non-stationary noise suppression in practice.

References

[1] A. An and N. Cercone, An Empirical Study on Rule Quality Measures, in: N. Zhong, A. Skowron, S. Ohsuga, eds., *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing. Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, (1999) 482-491.

[2] S. Boll, Suppression of Acoustic Noise in Speech Using Spectral Subtraction, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **27** (1979) 113-120.

[3] K. Brandenburg and J. Johnston, Second Generation Perceptual Audio Coding: The Hybrid Coder, *Proceedings of the 90th Audio Engineering Society Convention*, Montreux, France, preprint No. 2937 (1990).

[4] K. Brandenburg and G. Stoll, ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio, *Journal of Audio Engineering Society*, **42** (1994) 780-792.

[5] J. Buhmann, Stochastic Algorithms for Exploratory data Analysis: Data Clustering and Data Visualization, in: M. Jordan, ed., *Learning in Graphical Models*, Kluwer Academics (1997).

[6] J. Buhmann and H. Kühnel, Vector Quantization with Complexity Costs, *IEEE Transactions on Information Theory* **39** (1993) 1133-1145.

[7] M. Chmielewski and J. Grzymala-Busse, Global Discretization of Continuous Attributes as Preprocessing for Machine Learning, *Proceedings of the 3rd International Workshop on Rough Sets and Soft Computing*, San Jose, CA, USA, (1994) 294-301.

[8] A. Czyzewski, Learning Algorithms for Audio Signal Enhancement, *Journal of the Audio Engineering Society*, **45** (1997) 931-943.

[9] A. Czyzewski and R. Krolikowski, Application of Fuzzy Logic and Rough Sets to Audio Signal Enhancement, in: S. Pal, A. Skowron, eds., *Rough Fuzzy Hybridization. A New Trend in Decision-Making*, Springer Verlag, Singapore, (1999) 397-409.

[10] A. Czyzewski and R. Krolikowski, Noise Reduction in Audio Employing Auditory Masking Approach, *Proceedings of the 106th Audio Engineering Society Convention*, Munich, Germany, preprint No. 4930 (1999).

[11] J. Gibson and B. Koo, Filtering of Colored Noise for Speech Enhancement and Coding, *IEEE Transactions on Signal Processing* **39** (1991) 1732-1742.

[12] R. Gray, Vector Quantization, *IEEE Acoustics, Speech, and Signal Processing Magazine* **4** (1984) 4-29.

[13] R. Gray and D. Neuhoff, Quantization, *IEEE Transactions on Information Theory* **44** (1998) 1-63.

[14] J. W. Grzymała-Busse, LERS - A System for Learning from Examples based on Rough Sets, in: R. Słowiński, ed., *Intelligent Decision Support - Handbook of Applications and Advances of Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, (1992) 3-18.

[15] T. Hofmann and J. Buhmann, Competitive Learning Algorithms for Robust Vector Quantization, *IEEE Transactions on Signal Processing* **46** (1998) 1665-1675.

[16] E. Ifeachor and B. Jervis, *Digital Signal Processing. A Practical Approach*, Addison-Wesley Publishing Company (1993).

[17] J. Johnston, Transform Coding of Audio Signals Using Perceptual Noise Criteria, *IEEE Journal of Selected Areas in Communication* **6** (1988) 314-323.

[18] T. Kohonen, The Self-Organizing Map, *Proceedings of the IEEE* **78** (1990) 1464-1477.

[19] T. Kohonen, *Self-Organizing Maps*, Springer Verlag, Berlin (1995).

[20] T. Kohonen, E. Oja, O. Simula, A. Visa and J. Kangas, Engineering Applications of the Self-Organizing Map, *Proceedings of the IEEE* **84** (1996) 1358-1384.

[21] J. Komorowski, Z. Pawlak, L. Polkowski and A. Skowron, Rough Sets: A Tutorial, in: S. K. Pal, A. Skowron, eds., *Rough Fuzzy Hybridization. A New Trend in Decision-Making*, Springer Verlag, Singapore, (1999) 3-98.

[22] B. Kostek, Computer-Based Recognition of Musical Phrases Using the Rough-Set Approach, *Journal of Information Sciences* **104** (1998) 15-30.

[23] B. Kostek, *Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics*, Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York (1999).

[24] J. Lim, Evaluation of a Correlation Subtraction Method for Enhancing Speech Degraded by Additive White Noise, *IEEE Transactions on Acoustics*, *Speech, and Signal Processing* **26** (1978) 471-472.

[25] J. Lim and A. Oppenheim, Enhancement and Bandwidth Compression of Noisy Speech, *Proceedings of the IEEE* **67** (1979) 1586-1604.

[26] Y. Linde, A. Buzo and R. Gray, An Algorithm for Vector Quantization Design, *IEEE Transactions on Communication* **COM-28** (1980) 28-45.

[27] A. Moghaddamjoo and L. Kirlin, Robust Kalman Filtering with Unknown Inputs, *IEEE Transactions on Acoustics*, *Speech, and Signal Processing* **37** (1989) 1166-1175.

[28] MPEG-4, *International Standard ISO/IEC FCD 14496-3, Subpart 4* (1998).

[29] A. Øhrn and J. Komorowski, ROSETTA - A Rough Set Toolkit for Analysis of Data, *Proceedings of the Third International Joint Conference on Information*, Durham, NC, USA, **3** (1997) 403-407.

[30] Z. Pawlak, Rough Sets, *International Journal of Computer and Information Sciences* **11** (1982) 341-356.

[31] Z. Pawlak, *Rough Sets - Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht (1991).

[32] Z. Pawlak, Rough Sets. Present State and the Future, *Foundations of Computing and Decision Sciences* **18** (1993) 157-166.

[33] Z. Pawlak, Decision Rules, Bayes' Rule and Rough Sets, in: N. Zhong, A. Skowron, S. Ohsuga, eds., *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing. Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, (1999) 1-9.

[34] M. Sambur, Adaptive Noise Cancelling for Speech Signals, *IEEE Transactions on Acoustics*, *Speech, and Signal Processing* **26** (1978) 419-423.

[35] S. Shlien, Guide to MPEG-1 Audio Standard, *IEEE Trans. Broadcasting* **40** (1994) 206-218.

[36] A. Skowron and S. H. Nguyen, Quantization of Real Attributes: Rough Set and Boolean Approach, *ICS Research Report* **11/95**, Warsaw University of Technology (1995).

[37] A. Spanias, Speech Coding: A Tutorial Review, *Proceedings of the IEEE* **82** (1994) 1541-1582.

[38] S. Tsumoto and H. Tanaka, PRIMEROSE: Probabilistic Rule Induction Method based on Rough Sets and Resampling Methods, *Computational Intelligence* **11** (1995) 389-405.

[39] S. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley&Teubner, New York, (1997).

[40] S. Vaseghi and R. Frayling-Cork, Restoration of Old Gramophone Recordings, *Journal of Audio Engineering Society* **40** (1997) 791-800.

[41] B. Widrow, *Adaptive Signal Processing*, Prentice-Hall International Inc., New Jersey (1985).

[42] B. Widrow et al., Adaptive Noise Cancelling, Principles and Applications, *Proceedings of the IEEE* **63** (1994) 1692-1716.

[43] K. Yoshiya and J. Suzuki, Improvement in Signal-to-Noise Ration by SPAC (Speech Processing System Using Autocorrelation Function), *Electronics and Communications in Japan* **61** (1978) 18-25.

[44] U. Zölzer, *Digitale Audiosignalverarbeitung*, B.G. Teubner, Stutgart (1996).

[45] E. Zwicker and H. Fastl, *Psychoacoustics*, *Facts and Models*, Springer Verlag, Berlin, (1990).

[46] E. Zwicker and T. Zwicker, Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System, *Journal of Audio Engineering Society*, **39** (1991) 115-126.