Mining Knowledge in Noisy Audio Data

Andrzej CZYZEWSKI

Technical University of Gdańsk, Faculty of Electronics, Sound Engineering Dept.
Narutowicza 11/12;
80-952 Gdańsk, Poland.
andrzej@next.elka.pg.gda.pl

Abstract

This paper demonstrates a KDD method applied to audio data analysis, particularly, it presents possibilities which result from replacing traditional methods of analysis and acoustic signal processing by KDD algorithms when restoring audio recordings affected by strong noise.

Introduction

Typical applications of computer technologies to audio acoustics only rarely consider the opportunities of data processing with the use of methods which stem from KDD approach. What is essential here is the fact that methods of analysis and signal processing developed on the basis of speech acoustics have not been transferred respectively so far to other related areas, e.g. as an algorithm of intelligent analysis and processing of the musical audio signal. In the meantime the area of audio acoustics has an extensive demand for applications of intelligent signal processing.

The paper demonstrates a KDD method applied to audio data analysis, which was studied at the Sound Engineering Department of the Gdańsk Technical University. Particularly, it presents possibilities which result from replacing traditional methods of analysis and acoustic signal processing by KDD algorithms when restoring audio recordings affected by strong noise.

Previously, the parallel algorithm applied to the removal of clicks has been tested (Czyzewski 1994, 1995a) and the rough set method applied to noise suppression in old audio recordings was tried (Czyzewski 1995b). A new concept of perceptual coding allowing for noise reduction in old musical recordings stemmed from a modification of KDD applications investigated previously by the author (Czyzewski 1995b, 1995c). Perceptual coding provides the way of processing audio signal in such a way that the portions of signal which are perceptible to human hearing sense are to be encoded while the remaining portions of signal or noise are to be rejected. The algorithm processes signal in subbands of the frequency scale corresponding to the critical bands of hearing. The rough set method was employed to building the knowledge base of signal and distortions in such a way that it becomes possible to automatically control the masking threshold in order to maintain the noise affecting audio signals not audible to listeners.

Details of the elaborated and tested algorithms will be presented and results of their application discussed. Some general conclusions concerning knowledge acquisition of audio signal affected by noise and distortions will be added. Potential telecommunications and multimedia applications will be quoted.

Rough set approach to mining signal and noise data

The idea of using KDD approach to the removal of continuous noise from old recordings uses the perceptual coding scheme enhanced by the intelligent decision algorithm based on the rough set method. The rule set used for the determination of thresholds for the selection of eligible components of the signal is obtained by learning from examples. Hence, the data mining process allows one to discern between signal and noise portions of the audio material. Consequently, the masking threshold level can be determined for each data frame allowing one to make the noise inaudible after the execution of the perceptual coding procedure.

Before the details of the elaborated algorithm are presented a brief introduction to the domains of rough sets and of perceptual coding of audio data will be provided.

Basic concepts of the rough set theory

The Boolean traditional logic, which is employed by computers for general use, stems from Cantor's formulation of the definition of a set and operations on sets. However, as numerous examples prove, computers which work on this basis, are not good enough to solve many practical problems which require automatic inference, especially in situations when the data being analyzed carry a certain inaccuracy or irreproducibility and when there is no possibility to create a precise enough model of the decisive process. In such cases, overcoming the axioms of Cantor's definition of sets may turn out to be of purpose and very useful. This situation obviously corresponds to the problems of discerning between signal and noise components in noisy audio patterns. Overcoming the limitations related to Cantor's definition of the set is possible by ignoring the requirement that the set boundaries have to be strictly defined, that is of a set which is precisely defined by its elements. By doing so it is possible to define the set based on its lower and upper approximations. Such a set, since it is not defined fully, may include elements which belong to it many times.

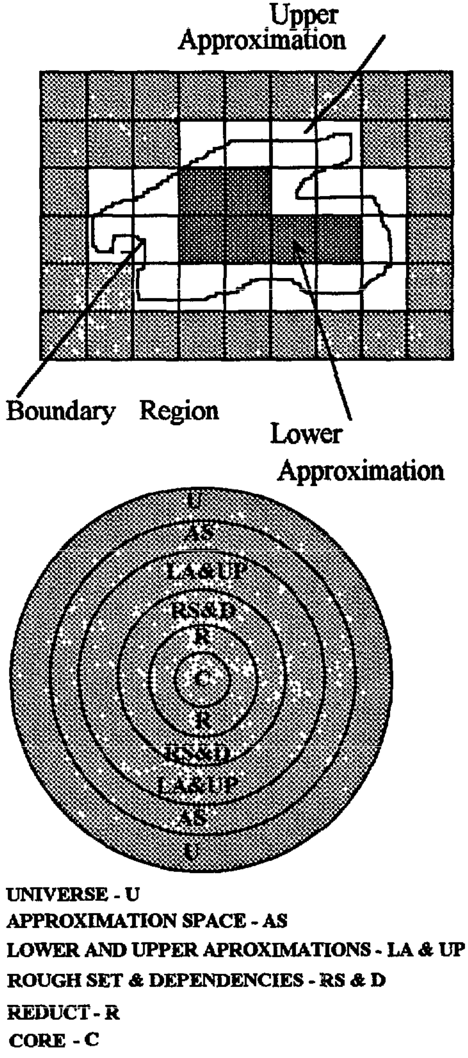

Fig. 1. Illustration of elementary notions related to the rough set theory.

Overcoming the traditional axioms applied in the case of rough sets causes the logic based on the rough set theory to acquire completely new features which make it an extremely useful tool for solving many problems which require an intelligent data analysis, searching for hidden relationships between the data and even making the right decisions in situations when incomplete and partially contrary antecedents exist.

The fundamentals of this theory were announced by Pawlak in the early 1980's (Pawlak, 1982). The theory was quickly adopted by the scientific world and is now one of the fast developing methods of artificial intelligence (Pawlak, 1993). The main terms related to defining a rough set are illustrated schematically in Fig. 1. More detailed definitions of these terms can be found in the literature (Slowinski ed., 1992).

The knowledge is represented in a system based on rough sets in a tabular form composed of the following elements:

$S_R = \langle U, P, D, V_P, V_D, F \rangle$, where U is a finite set of objects, P - is a finite set of attributes, D - is a special decisive attribute determined by the expert, $_P$ is a collection of all conditional attributes in P, V_D - represents the area of decisive attributes, F - is called the knowledge function.

An inductive learning system based on rough set consists of the learning component for the automatic rule derivation from training samples and of the inference system used for decision-making at the recognition stage. Rules are expressions in the following format:

if <condition1>and<condition2>...and<condition n>
then <decision>.

Conditions are based on attributes. Attributes should be equal to concrete values or should belong to certain ranges. The decision always use a single rule matching approach.

The knowledge base in rough set method can be conveniently represented in the form of a decision table, in which rows represent objects and columns represent attributes. In the last column decision attributes are collected. The main task for the learning phase (rule generation procedure) is to find the minimum number of maximum composed sets that cover the so called positive region $POS_A(S)$ providing one of the basic notions of the rough set theory (Pawlak, 1982). Consequently, two types of rules are to be derived from the decision table: certain rules and possible rules. An important parameter of possible rules, reflecting quality of them is the rough measure μ_{RS} defined as follows:

$$\mu_{RS} = \frac{|X \cap Y|}{|Y|} \quad (1)$$

where: X is the concept and Y is the set of all examples described by the rule.

An additional parameter was proposed by the author allowing one to optimize the rule generation process. This parameter was called the rule strength r and is defined as follows:

$$r = c(\mu_{RS} - n_{\mu}) \tag{2}$$

where: c - number of cases conforming the rule,

n_{μ} - the neutral point of the rough measure.

The neutral point of the rough measure n_{μ} provides one of some parameters of the rule generation system to be set by its operator. This parameter allows one to regulate the influence of possible (uncertain) rules on the process of decision making. More details concerning the rough set concepts are to be found in the rich literature (Slowiński ed., 1992; Wojcik, 1993).

Perceptual coding algorithm

Empirical data reveal that some signals are inaudible in the presence of others both in time and in frequency domain. The level of the barely audible pure tone depends on the difference of frequencies of the signal components and the masking tone, and on the excitation level of the masker. Usually, this dependence is presented as the function of frequency in a linear scale or in a logarithmic scale (Zwicker&Zwicker 1991). It can be observed that the shape varies significantly according to frequency changes. In the literature, some attempts to calculate that shape can be found (Krolikowski, 1996).

Another well-proven phenomenon is the fact that human auditory system analyses spectrum ranges which correspond to subbands, called critical bands. The stimuli frequencies within a critical band are perceived similarly and processed independently from those of other critical bands. Because of the need to use the critical bands scale, the Bark unit for this was defined, which is of the width of one critical band. It can be observed that Bark-to-Hertz dependence can be approximated by the straight line: up to the 8th Bark in the linear frequency scale and from the 8th Bark in the logarithmic one. The above property was taken into account when converting to Bark units in the elaborated algorithm. Moreover, it turns out that masking curves assume almost an uniform shape when calculated versus Bark units. One can notice that the slopes of the curves at the lower frequencies are constant, whereas at the higher ones, the slope depends on the excitation level of the masker.

Basing on the above indications, the elaborated perceptual algorithm for the encoder performs the following operations:

Step 1. On the basis of the sampling frequency, values of frequencies for consecutive spectral lines are computed. Next, they are converted to Bark units.

Step 2. The continuous series of 16-bit samples are grouped into blocks of the length of 512 samples. This size is assumed to be a compromise between high spectral resolution (long blocks) and time resolution requirements (short blocks). Since at the borders of blocks some audible distortions may occur, so called 'block effects', thus overlapping technique with the length of the fold equals to 32 was applied.

Step 3. Because of the Digital Fourier Transform properties, blocks containing samples are windowed before the processing is made. Rectangular-cosine shaped window was used calculated according to the following formula:

$$data(i) = \begin{cases} data \cdot \sin(\frac{\pi}{2} \cdot \frac{i}{L}), & i \in \langle 0, L-1 \rangle \\ data, & i \in \langle L, N-L-1 \rangle \\ data \cdot \cos(\frac{\pi}{2} \cdot \frac{i}{L}), & i \in \langle N-L, N-1 \rangle \end{cases} \tag{3}$$

where:

L - the size of the fold,

N - the length of the block.

Step 4. The FFT procedure is executed. Amplitudes and phases of spectral lines are calculated using complex representations. Since these values are symmetrical, only half of them is further processed. The remaining part is restored in the decoder.

Step 5. The amplitudes of spectral lines are sorted according to decreasing values of amplitude.

Step 6. Simultaneous masking procedure is executed. Spectral lines with amplitude value remaining below the masking threshold are discarded. Masking curves are represented by uniform curves. For the practical use, shapes of these curves are approximated by straight lines. The slope of the line at the lower frequencies is set to 27 dB/Bark, whereas at the higher ones, the slope S (with minus sign) is expressed as:

$$\mathrm{S = 22 - 0.2 \cdot I \left[dB / Bark\right]} \tag{4}$$

where:

I - denotes the level of the masker.

For each non-masked spectral line, there is a mask level computed, i.e. the excitation value of barely audible tone of the respective frequency.

Step 7. Within every critical band for each spectral line, the minimum mask level is chosen. This approach ensures that after the quantization of non-masked components, the quantization noise is maintained below the audible level.

Step 8. Subjective level of noise is estimated by a human listener. If its value is below the mask level in every critical band - no further actions are performed. If not, new masking thresholds are established to shift the noise below the masking level in each particular subband. This procedure is conceived as a flexible one in order to not to diminish nor to neglect eligible components within the band.

Step 9. After the perceptual processing, every spectral component is quantized or set to 0 (if masked). It would be redundant to encode masked values. Thus, only these components with non-zero magnitudes are processed. However, there is a need to know, for which frequencies spectral lines were neglected if they were. Therefore, an additional piece of information is sent, which defines whether consecutive complex components are transmitted or not. And thus, every package of data is preceded by 256 one-bit flags. If the flag is set to 0 at the kth position, it means that the magnitude and phase for the kth spectral component is not processed (the masking occurred). On the other hand, if the kth value is set to 1, it denotes that the package includes the kth complex component.

The audio samples after quantization can use from 1 to 16 bits. It would be a waste of storage capacity to utilize all 16-bit representation whereas the value can match only a part of it. Thus, it can be assumed that the number of bits used to encode these values is constant in a critical band within the time duration of several frames. In a case, when the word length changes, one needs to send only a new value for appropriate subband. Consequently, the data package is preceded by a number of one-bit flags. When the kth flag is set to 0, it means that the number of bits used to encode spectral components in the kth critical band did not change. When the flag is equal to 1, it denotes that a new number for these bits will be transmitted for the kth band. The number of one-bit flags is equal to the number of critical bands determined by the given sampling frequency.

During the masking procedure, it is evaluated the permissible noise level in each critical band. The level is used for the quantization and should be known to restore sound correctly. Thus, there is a need to store values of the levels for every subband. Fortunately, it turns out that it is essential to precede the audio package by a number of flags informing whether the noise level in consecutive critical bands changed. If so, a new 16-bit value of the level for appropriate band is transmitted. As previously, the number of the one-bit flags refer to the number of utilized critical bands.

Consequently, each 512-sample frame is stored in the format as follows:

256 of one-bit flags of non-masked spectral lines,

up to 25 of one-bit flags. They describe the change of the word length engaged to encode complex components in a critical band,

up to 25 of one-bit flags. They concern the change of permissible noise level in a critical band,

up to 25 of 4-bit values of a new value of word length used to encode spectral components in a critical band,

16-bit values of a new permissible noise level in a critical band,

values of amplitudes and phases of non-masked spectral lines. The value of phases can be encoded using 5 bits, what is recommended in the literature. Consequently, amplitude and phase of the constant complex component is encoded.

The decoder uses the above data for the additive synthesis of sound on the basis of spectral components which have been qualified as non-masked ones.

Knowledge acquisition of signal and of noise

Previously, the rough set approach to the determination of spectral components obtained in the McAulay-Quatieri analysis was tried by the author (Czyzewski 1995b, 1995c). Similar KDD procedure to the one elaborated previously was exploited in the current experiments to derive rules allowing to automatically select the optimal level of the masking threshold in the perceptual coding/filtration procedure.

The masking threshold influencing the selection of spectral components in the encoder should be updated frame by frame basing on the rule set. This rule set is to be acquired on the basis of examples processed during the learning phase. Correspondingly, three classes are to be defined: threshold low (too low), threshold medium or balanced (right) and threshold high (too high). The expressions in brackets correspond to subjective assessment of the effect of filtration. When the threshold is too low, then the noise is clearly audible. When the threshold is too high, many eligible components are removed, so the resulting sound is clean, but poor and distorted. Balanced threshold allows one to remove noise without discarding eligible signal components. As results from above indications, threshold values need to be quantized. The quantization consists in replacing real values of masking threshold by the range representations which are utilized as decision attributes (Slowiński ed., 1992). Practically, the uniform quantization was employed based on 6 dB ranges of magnitude of threshold.

Practically, the learning procedure consists in selecting some short fragments of the recording, automatically setting various threshold levels and assessing the effect subjectively when playing back those fragments after the resynthesis. Correspondingly, each sample fragment is to be represented by a set of threshold values in each critical band and the decision comes from the human expert. That is the way the knowledge base is built with regard to expert subjective assessments of individual examples. Normally, it is sufficient to choose some examples corresponding to the most characteristic fragments of the recording. Typically, up to 5 percent of the whole material should be chosen and assessed on the basis of 2 to 3 seconds portions. This produces a stream of exemplary data to be added to the relevant classes labeled as “low", "right" and too “high”. Consequently, the knowledge base is build up to be applicable to the selected fragments (certainly) and to the rest of the whole recording (possibly). The generalization capabilities of KDD algorithms the rough set belongs to proved to work well also with the new patterns representing the material not employed to the training. The rule base represents knowledge acquired during the training of the algorithm. The rule set may contain rules of the following form:

$$\left(s_{k-4}(n-5)\cap s_{k-2}(n-3)\cup s_{k-3}(n-5)\cap s_{k-3}(n-4)\right)\cap ..$$
$$\cap\left(s_{k-1}(n)=0\right)\cap\left(s_{k-1}(n-1)=0\right)\cap\left(s_{k-2}(n+1)\right)\left(s_{k-1}(n+2)=0\right)\Rightarrow\left(p=p_{\max}\right) \tag{5}$$

where: $T(b_k)=\{l_1,l_2,...,l_{16}\}$ denotes that in the frequency band No. k the threshold value is to be set to the quantized level l_i; $i\in<1,2,...,16>$; $k\in<1,2,...,24>$

Then, the magnitude of the threshold set in kth critical band is as follows:

$$L_k = 6l_i \text{ [dB]} \tag{6}$$

The exemplary rules presented above are to be determined automatically through the learning from fragments of the recording, packet after packet. As each packet represents the result of FFT transform of 512 samples, thus for each 1s portion of audio sampled with the frequency 22.05 kHz as much as 22.050/512 ≅ 43 rules may be generated.

Thus, initially after typical learning procedure employing 10 or 15 s of exemplary audio material the decision table is constructed containing several hundreds of rules. Usually, such a data collection is superfluous. This feature results from the fact that musical tones duration usually exceeds single packet length. Consequently, many rules and attributes $T(b_k)$ are discarded after the rough set based reduction of the obtained decision table and the resulting knowledge base is automatically compacted. Subsequently, the acquired rule base is used for the automatic setting thresholds for all subbands and all consecutive packets of the whole recording. For the concrete combination of input data many rules are firing, some of them certain (rough measure equal to 1) and some uncertain (rough measure lower than 1) - see eq. 1. According to the rough set method principles, the decision always comes from the single firing rule being the strongest one - see eq. 2. Consequently, the spectrum filtering thresholds are to be updated according to the winning rules. Subsequently, the current packet is processed using threshold values update controlled by the rule conditional attributes. Results of processing noisy recordings with this method are presented in the next paragraph.

Results

Music affected by strong noise was used for the described experiments. The analysis-resynthesis algorithm with the "intelligent" threshold update was applied among others to an exemplary fragment of the song performed by Edith Piaf taken from a very noisy record.

After the perceptual filtration is executed, which is supported by the rough set-based control of masking threshold, the rectified signal was obtained revealing enhanced subjective quality. Results of analyses of music material made before and after the processing are presented in Fig. 2. As is seen from Fig. 2a, spectral analysis of the musical fragment reveals that signal components are accompanied by very strong noise. The noise is broadband (so called hiss) and its components are strong within the whole range of frequency (up to 1/2 sampling freq. which was equal to 22.05 kHz).

Fig. 2b presents spectral analysis of the fragment restored with the perceptual coding algorithm with not properly selected masking thresholds - some eligible signal components are weakened while the higher components of hiss remain still beyond the masking threshold. Fig. 2c presents the result of signal restoration with the perceptual coding algorithm controlled by the rule set derived from eight characteristic portions of the whole recording (time duration of each portion: 2 to 3 s).

Conclusions

KDD algorithms should find their way to more audio applications. The previously conducted experiments related to neural network implementation to the removal of impulse distortions (Czyzewski 1994, 1995a) and the presented exemplary application related to the restoration of noisy audio recordings may support this opinion. There are many potential applications of intelligent algorithms applied to the removal of noise and distortions. Some of them might be used in telecommunications and digital broadcasting, in databases and multimedia-related systems as data reduction and noise suppression techniques.

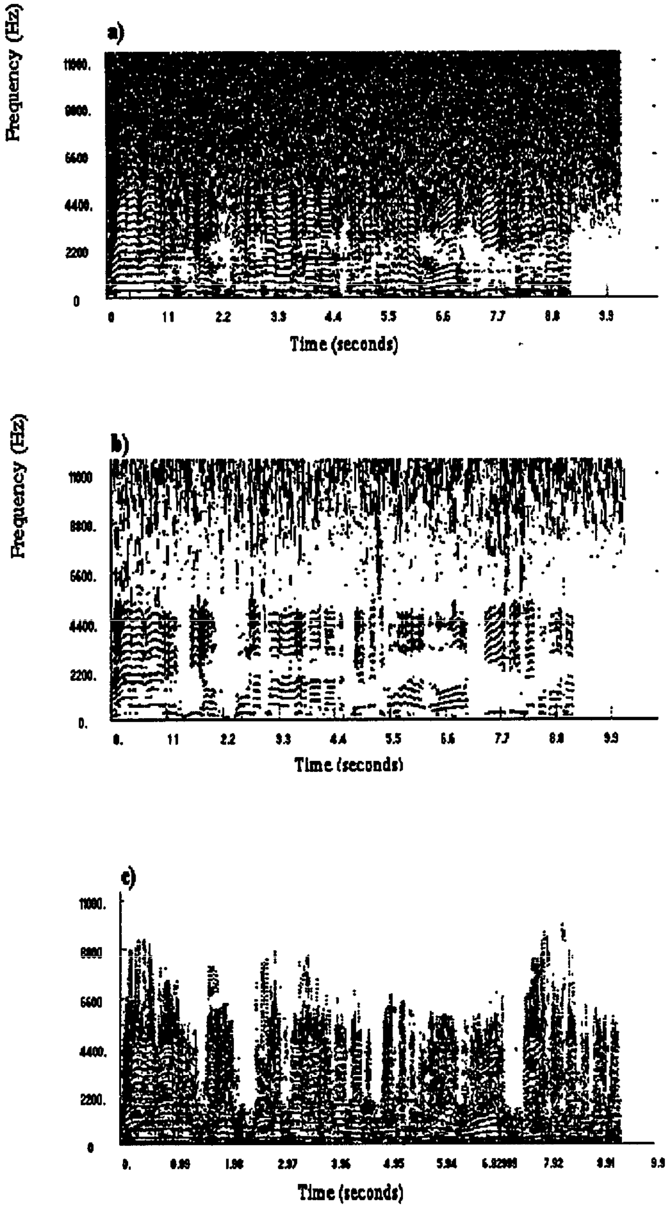

Fig. 2 Results of perceptual filtering of sound with the use of the rule base.

a./ Original fragment of an audio recording affected by strong noise,

b./ The same fragment restored with the use of insufficiently trained algorithm,

c./ Fragment as in Fig. (a) processed employing the final version of the rule set.

Acknowledgment

The presented research project was sponsored by the Committee for Scientific Research, Warsaw, Poland. Grant No. 8 T11D 002 08.

References

1. Czyzewski A. 1994, Artificial Intelligence-Based Processing of Old Audio Recordings. In Preprint of 97th Audio Engineering Society Convention. San Francisco (Preprint No. 3885).

2. Czyzewski A. 1995a. Some Methods For Detection And Interpolation Of Impulsive Distortions In Old Audio Recordings. In Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics. Mohonk Mountain, N.Y., U.S.A.

3. Czyzewski A. 1995b. Managing Noisy Data in the AI-based Processing of Old Audio Recordings. In Proceedings of International Symposium on Intelligent Data Analysis. Baden-Baden, 17-19 August, 1995.

4. Czyzewski A. 1995c. New Learning Algorithms for the Processing of Old Audio Recordings. In Preprint of 99th Audio Engineering Society Convention. New York (Preprint No. 4078).

5. Pawlak Z. (1982). Rough sets. Journal of Computer and Information Science, vol.11, No.5.

6. Pawlak Z. (1993). Rough Sets - Present State and the Future. Foundations of Computing and Decision Sciences, vol. 18, No.3-4.

7. Slowiński R., ed. 1992. Intelligent Decision Support. Handbook on Applications and Advances of the Rough Sets Theory. Kluwer Academic Publisher, Dordrecht/Boston/London.

8. Wojcik Z.M. 1993. Rough Sets for Intelligent Image Filtering, In Proc. of Rough Sets and Knowledge Discovery Workshop (RSKD), Banff, Canada.

9. Zwicker E., Zwicker U. 1991. Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System, J. Audio Eng. Soc., vol. 39 (pp. 115-126).

10. Krolikowski R. 1996. Noise reduction in old musical recordings using the perceptual coding of audio. Forthcoming.