Electronic Notes in Theoretical Computer Science 82 No. 4 (2003)
URL: http://www.elsevier.nl/locate/entcs/volume82.html 12 pages

"Computing with words" Concept Applied to Musical Information Retrieval

Bożena Kostek [1]

Sound and Vision Engineering Department
Gdańsk University of Technology
Narutowicza 11/12, 80-952 Gdańsk, Poland

Abstract

The objective of the paper is to provide cognitive-based mechanisms underlying processing of musical instrument sounds. The system proposed by the author based on the rough set method and on fuzzy logic provides knowledge on how humans internally represent such notions as quality and timbre and therefore it allows for the human-like automatic processing of musical data. Therefore "Computing with words" concept can be used in musical information retrieval domain by offering better processing of subjective descriptors of musical instrument sounds and enabling the analysis of data that would result in extraction of semantic information related to musical instrument sounds. This paper shows first a review of developments in the domain of timbre mapping and classification. A decision table is built of semantic descriptors of musical instrument sounds, then rules extracted by the rough set method and the processing of musical timbre based on fuzzy logic is shown. An example of rough-fuzzy processing is given and conclusions are derived.

Key words: Computing with words, soft computing, rough sets, fuzzy logic, Musical Information Retrieval, MPEG7.

1 Introduction

One of possible examples of "Computing with words" processing is the assessment of the quality of musical instrument sounds, for which humans use criteria that are rarely quantitative but most often qualitative. Therefore, there is a need to find methods that make it possible to process such linguistic attributes describing sound timbre as: bright, dark, clear, soft, high, low, etc. Soft computing techniques originally applied to this task have been hitherto implemented and tested in many other domains and applications.

[1] Email: bozenka@sound.eti.pg.gda.pl

"Computing with words" concept introduced by Zadeh refers to the fact that humans employ words in computing and reasoning, arriving at conclusions expressed as words from premises expressed in a natural language [26], [27]. It is obvious that computing with words is generally less precise than computing with numbers. However computing with words is better solution when information is too imprecise to justify the use of numbers. It is also advantageous in situation when imprecision is in better rapport with reality [26], [27]. As an example of such imprecision a notion of timbre or sound quality can be given. Timbre or subjective quality as assessed by humans cannot in most cases be crisply defined, leaving a wide margin of uncertainty which depends on individual subjects' preferences and the unknown influences of individual parameter values on the timbre quality of musical sounds.

In the recently standardized MPEG7 it is defined that audio signals can be described at different abstraction levels: from the lowest level (primitives - such as for example: temporal or audio spectrum centroids, spectrum flatness, spectrum spread, inharmonicity, etc.) to the highest level (related to semantic information) [13]. In this context semantic information is related to textual information on audio such as titles of songs, composers, etc. Therefore there is a need to find methods that make it possible to find a relationship between objectively extracted information from sound and subjective notions of timbre. It seems a way to change primitives into higher abstraction level, namely semantics. It should be remembered that MPEG7 standard does not comprise the (automatic) extraction of descriptions/features, nor does it specify the search engine that can make use of the description [15]. Soft computing offers techniques that have been developed and tested in many other domains and applications and they are especially valuable in domains in which there is a problem of imprecision and a need of knowledge mining [10], [16], [23], [28], [26], [27], [29], [30].

2 Timbral Descriptors

Timbre is defined as the quality, which distinguishes two sounds with the same pitch, loudness and duration. The notion of musical sound timbre refers to work done by Grey [6] and later by Krimphoff et al. [12], McAdams and Winsberg [14], Wessel [25], Reuter [22] and many others [2], [3], [8], [9], [17], [18], [19], [20], [21], [24]. Three dimensions recognized by Grey were discussed in some papers and resulted in diminishing the timbral space to two dimensions. In the original paper by Grey spectral energy distribution, the presence of synchronicity in sound attacks and decays (spectral fluctuation) and low-amplitude, high-frequency energy in the initial attack represented the perceptual relationship. In other studies the first dimension is identified with the log attack time, the second one with the harmonic spectral centroid [8] and the third one with a temporal centroid. On the other hand, two dimensions are related to sharpness (or *brightness*) and velocity of attack. In most papers dealing with perceptual space it may be seen that researchers tried to limit this space to three or two parameters. It is obvious that in such a case dimensions can be easily interpreted and presented as a two- or three-dimensional projection, however in author's opinion derived from the results of processing multidimensional feature vectors describing musical sound characteristics there is no need to limit the perceptual space to such a small number [7], [10], [11]. Computer processing can easily deal with multidimensionality of feature vectors [8], [9], [10]. Moreover, such processing is now highly needed for automatic queries within sound digital archives. More dimensions can help to distinguish between particular instruments or musical instrument groups. Another vital problem related to discovering relationship between sound descriptors and objectively derived parameters remain not yet solved. Only a few parameters such as for example brightness have got their unquestioned interpretation - this subjective descriptor is related to spectral centroid [1]. Further discussion is also required for assignment of ranges of sound parameters. This also will serve to better distinguish between musical instrument sound characteristics.

3 Timbre Assessment

In order to assess sound timbre or sound quality listening test sessions are often organized. Two important assumptions were made in tests carried out in the multidimensional perceptual scaling of musical timbre. Sound stimuli were synthesized, and they were equalized for perceived pitch, loudness and duration. This was done in order to get reliable results during timbre scaling tests. However, in the presented study it is assumed that relationship between subjective descriptors and objectively derived parameters will serve for better quantization of numerical values. In such a case testing should be done using natural sound stimuli. Such tests were carried out in architectural acoustics in order to describe quality of an interior. It was possible to assign labels to certain numerical parameter values by experts. Such tests were arduous but resulted in reliable evaluation of acoustical parameters. It is much easier for experts to say that such a sound has "dark" or bright" quality, and contrarily it is difficult to assign numerical values. The problem remains how much is bright and what does mean "nasal" or "flute-like" quality as expressed in numbers.

Let us consider how such a procedure should be carried out. First, one should choose such attributes that have subjective meaning to experts. A few such parameters were already found and named in the musical acoustics domain and they are based on parameters derived from time, frequency and time-frequency domains. This can be a starting point to list some parameters suitable for creating a feature vector both in subjective and objective domain.

One can name such parameters both subjective and corresponding measures as: *pitch* (frequency in Hz or barks), *brightness* (*spectral centroid*), tone/noise-like-quality (*spectral flatness measure*), *attack asymmetry* (skewness) or *attack duration*, *overshoot* or *inharmonicity* (log ratio of the 1st harmonic to 2nd harmonic or more generally - higher frequency harmonics to the fundamental frequency ratio), *vibrato* (periodic fluctuation of pitch), *tremolo* (periodic change of sound level), *nasality* (formant positions if exist), *synchronicity* (delay of higher harmonics with relation to the fundamental during the attack), etc. that have double interpretation. In addition there are parameters on the basis of which a distinction between musical instrument group can be made. For example, skewness is a measure of data symmetry, or more precisely, the lack of symmetry. A distribution is symmetric if it looks the same to the left and right of the center point. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left (the left tail is heavier than the right tail) and positive values for the skewness indicate data that are skewed right. One can use such a statistical measure for describing distribution of harmonics, which is different for woodwind and brass instruments.

Now, the problem is not only to assign ranges to such parameters (using word descriptors) - one can easily imagine that experts would unanimously decide what does it mean high or pitch of a certain musical instrument. In such a procedure a subject has to associate presented stimuli with a set of adjective scales (semantic). The subject's task is to indicate for each sound a three- or five-point scale, which of the given terms applies to the stimulus. The drawback is that experts are forced to judge the stimuli in terms of prescribed semantic categories and scales. The preselection of scales determines the resolution of the analysis while the verbal categories may seem different from the expert's auditory sensation. In addition one should be aware that building such a set of parameters could be done only experimentally. Even an expert in musicology cannot decide as to the number of parameters and their significance to the instrument recognition without subjective tests and processing of results. Another problem is to find rules on the basis of which a chosen instrument can be qualified into adequate group with only some degree of uncertainty. For this purpose both computing with words concept and processing using soft computing methods can be applied. A discussion of main points of such an approach will be shown further on.

4 Knowledge Acquisition Phase

4.1 Rough set-based processing

Relationships between the objectively measured parameters of musical instrument sounds and their subjective quality as assessed by listeners (preferably experts) cannot in most cases be crisply defined, leaving a wide margin of uncertainty which depends on individual subjects' preferences and the unknown influences of individual parameter values on the overall timbre quality of the

Table 1
Knowledge base representation in the rough set theory

Object/attribute	A_1	A_2	A_3	$\cdots$	A_m	$D(decision)$
t_1	a_{11}	a_{12}	a_{13}	$\cdots$	a_{1m}	d_1
t_2	a_{21}	a_{22}	a_{23}	$\cdots$	a_{2m}	d_2
t_3	a_{31}	a_{32}	a_{33}	$\cdots$	a_{3m}	d_3
...	...	...	...	...	...	...
t_n	a_{n1}	a_{n2}	a_{n3}	$\cdots$	a_{nm}	d_n

tested sound. Consequently, results of subjective tests had to be processed statistically (hitherto used approach) in order to find links between preference results and concrete values of parameters representing the objective features of tested objects.

A new extended proposal of the procedure for analyzing subjective test results is formulated. In the first step of the analysis, the results of the listening test sessions should be collected into tables, separately for each expert and for each sound excerpts. Then, these tables should be transformed into the format of the decision table used in the rough set decision systems (Tab. 1) [16]. Objects t_1 to t_n from Tab. 1 represent an object, and attributes A_1 to A_m are denoted as attributes, and are used as conditional attributes. This means that a row in the *Decision Table* represents an object characterized by the feature vector, and each column corresponds to the attribute values. The last column refers to decision.

The questionnaire form that can be used in listening tests is as presented in Tab. 2. Subjects are asked to fill in the questionnaire during listening sessions. Having collected the assessments of perceptual dimensions of the tested sounds from experts, it is possible to create a decision table. Objects t_1 to t_n from Tab. 1 represent now various musical instrument sounds, and attributes A_1 to A_m are denoted as tested sound characteristics. The expert's scoring is defined by the grades a_{11} to a_{nm} (quantized values are labelled descriptively as for example *low*, *medium*, and *high*). The decision D is understood as a value assigned to the name of a musical instrument or a number referring to it. The result of the rough set-based processing is a set of rules that will be later used to recognize a musical instrument sound unseen by the system.

Then this table should be processed using the rough set method. In this way, a set of rules would be created, which may subsequently be verified by experts. A decision system based on rough set theory engineered at the Gdansk University of Technology can be used for this purpose [4], [5]. It includes learning and testing algorithms. During the first phase, rules are derived which become the basis for the second phase. The generation of decision rules starts from rules of length 1, continuing with the generation of rules of length

Table 2
Listening test results for a given sound No. i

Subj./ Grades/ Descr.	*Pitch*	*Bright.*	*Attack durat.*	*Tone/ noise quality*	$\cdots$	*Vibr.*	*Synchr.*	*Musical Instr. Class*
1	*low*	*low*	*low*	*low*	$\cdots$	*low*	*high*	No. 1
i	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	*med*	*high*	*high*	*low*	$\cdots$	*high*	*high*	No. 4

2, etc. The maximum rule length may be determined by the user. The system induces both possible and certain rules. It is assumed that the rough set measure μ_{rs} for possible rules should exceed the value 0.5. Moreover, only such rules that are preceded by some shorter rule operating on the same parameters are considered.

A rough set measure of the rule describing concept X is the ratio of the number of all examples from concept X correctly described by the rule:

$$\mu_{rs} = \frac{|X \cap Y|}{|Y|},$$

where:

X - is the concept,

Y - set of examples described by the rule.

In the testing phase the leave-one-out procedure is performed. During the jth experiment, the jth object is removed from every class contained in the database, the learning procedure is performed on the remaining objects, and the result of the classification of the omitted objects by the produced rules is saved.

In the rough set-based processing discretized data is used. This means a process of replacing the original values of input data with the number of an interval to which a selected parameter value belongs. There are many methods that can be used for such a purpose, both local and global ones operating on the whole data set [10]. However, in the proposed method data are quantized by means of labels assigned by experts in listening tests, so there is no need to discretize them at this stage of analysis (see Fig. 1). The mapping of test results to fuzzy membership functions will be presented later on but is seen on the right side of the Fig. 1.

The rules are of a form:

RULES:

if (*Pitch* = *high*) and (*Brightness* = *high*) and ..... then (*Class No.* 1) μ_{rs}=0.9

if (*Pitch* = *med*) and (*Brightness* = *high*) and ..... then (*Class No.* 1) μ_{rs}=0.8

....

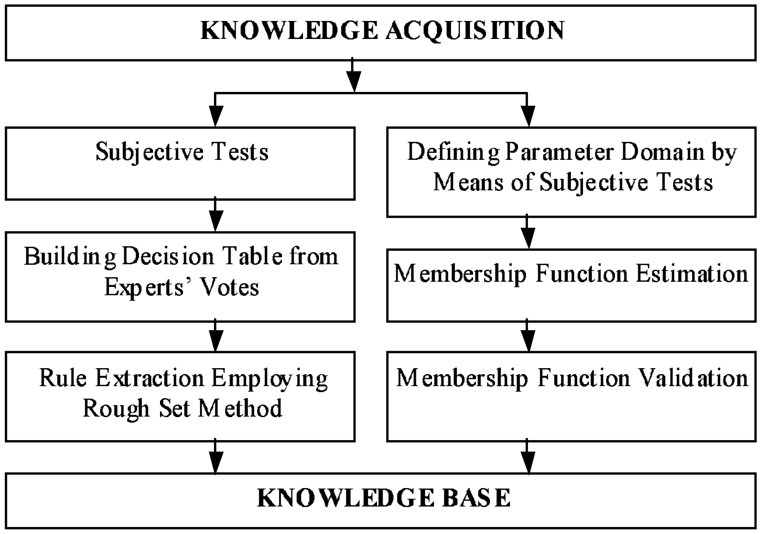

Fig. 1. Rough-fuzzy expert system - knowledge acquisition phase

4.2 Mapping Test Results to Fuzzy Membership Functions

The next step of the procedure is to obtain subjective ratings for each of objective parameters as assessed separately from the others. The mapping of objective parameter values to their subjective assessments by many experts creates some fuzzy dependencies, which can be represented by fuzzy membership functions corresponding to each parameter separately.

As was mentioned before, experts, while listening to sound, are instructed to rate their judgements using such descriptions as *low*, *medium*, and *high*, etc. This procedure uses a concept of the *Fuzzy Quantization Method* (FQM) applied to acoustical parameters [10]. This results in the relation of semantic descriptors to the particular parameter quantities. The distribution of the observed instances very often suggests the trapezoidal or triangular shape of a membership function (see sample membership functions presented in Fig. 2).

One of the important tasks is to approximate the tested parameter distribution. This can be done by several techniques. The most common technique is linear approximation, where the original data range is transformed to the interval [0,1]. Thus, triangular or trapezoidal membership functions may be used in this case [10]. Also, polynomial approximation is often used for such

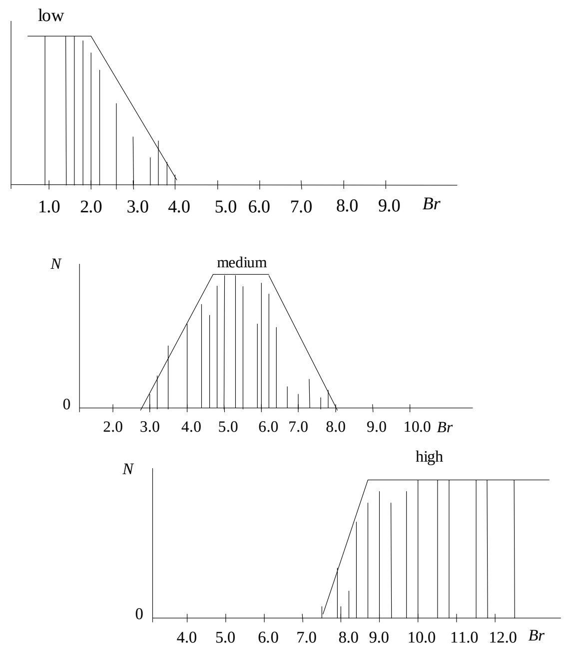

Fig. 2. Experts' vote for the *Brightness* parameter, N - number of experts voting for particular values of *Brightness* (Br)

a purpose. Another approach to defining the shape of the membership function involves the use of the probability density function. The last mentioned technique was very thoroughly discussed in previous work by the author [10].

5 Automatic Classification Stage

The second stage of the expert system (Fig. 3), namely automatic classification of musical instrument, consists of several steps. In order to enable the automatic recognition of a musical instrument class, new data representing sound parameter values is fed to the system inputs (Tab. 3). The first step is the fuzzification process in which degrees of membership are assigned for each crisp input value (as in Fig. 4). Therefore, for the data presented in Tab. 3, the degree of membership for each input value (for a given label) has to be determined.

The pointers in Fig. 4 refer to the degrees of membership for the precise value of *Brightness*=7.75. Thus, when the value of *Brightness* equals 7.75, it belongs, respectively, to the low fuzzy set with the degree of 0, to the medium fuzzy set with the degree of 0.4 and to the high fuzzy set with the degree of 0.6. The same procedure should be applied to other parameters of Tab. 3.

The next step is processing rules generated from subjective data by the RS system, using FL quantized objective data as the input. This is done as in example shown below.

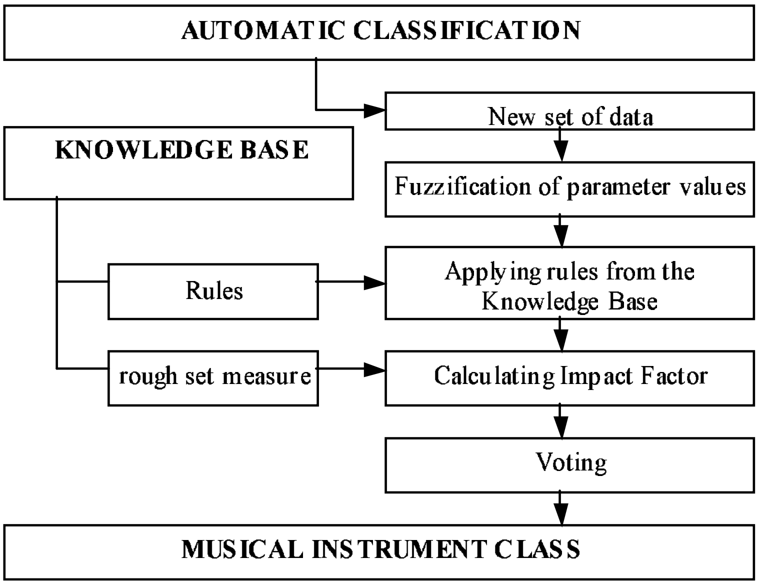

Fig. 3. Rough-fuzzy expert system - automatic classification phase

Table 3
Set of parameter values presented to the system inputs

Sound/ No. *i*	*Pitch*	*Bright.*	*Attack duration*	*Tone/ noise quality*	$\cdots$	*Vibr.*	*Synchr.*
1	440Hz	7.75	...	...	...	...	...

if *Pitch* = *high* (0.3) and *Brightness* = *high* (0.6) and ..... then (*Class No.* 1) FS=0.3 μ_{rs}=0.9 $\Longrightarrow$ $IF = 0.27$

if *Pitch* = *med.* (0) and *Brightness* = *high* (0.3) and ..... then (*Class No.* 1) FS=0 μ_{rs}=0.8 $\Longrightarrow$ $IF = 0$

....

if ..... and *Vibrato* = *high* (0.1) and *Synchronicity* = *high* (0.2) then (*Class No.* 4) FS=0.1 μ_{rs}=0.9 $\Longrightarrow$ $IF = 0.09$

where: FS - fuzzy strength, IF - *Impact Factor*

It should be remembered that after the rough set processing, only the strongest rules with rough set measure value exceeding 0.5, would be considered. The rough set measure associated with the individual rule will be contained in the knowledge base. During the classification phase a so-called *Impact Factor* would be calculated resulted from multiplication of rough set measure and fuzzy strength. The last step is the fuzzy processing of rules.

Since there might be several rules corresponding to the given musical instrument, thus a sum of rough set measures of rules associated with the individual musical instrument could be calculated. This means that even if a rule is certain for a given instrument (in this case rough set measure equals 1), but there exist also several rules that would point out another instrument, after the aggregation of rough set measures, the latter instrument name would be returned in the system decision. This procedure refers to the "Voting" in the block-diagram shown in Fig. 3.

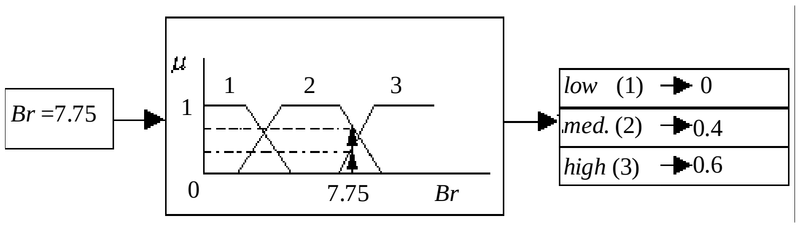

Fig. 4. Fuzzification process of the *Brightness* parameter

6 Conclusions

The presented system consisted of two stages, namely acquisition and automatic musical instrument classification. "Computing with words" concept is used both in rough set- and fuzzy logic-based processing. With rules determined from the rough set decision table (elimination of redundant attributes, discovery of hidden relations, rule derivation) and membership functions determined empirically for the studied parameters (converting of crisp data sets to fuzzy sets), one can create an expert system that provides automatic decisions on musical instrument classification each time a concrete set of parameters is presented to its inputs. Such an approach was previously applied to automatic assessment of acoustical quality, bringing reliable results and is currently implemented to automatic musical timbre recognition.

7 Acknowledgments

The research is sponsored by the Committee for Scientific Research, Warsaw, Poland. Grant No. 4 T11D 014 22.

References

[1] Beauchamp, J. W., *Unix Workstation Software for Analysis, Graphics, Modification, and Synthesis of Musical Sounds*, Proc. 94th Conv. of the Audio

Engineering Society, preprint 3479, J. Audio Eng. Soc.(Abstracts), **41** (1993), 387.

[2] De Bruijn, A., *Timbre-Classification of Complex Tones*, Acustica **40** (1978), 108–114.

[3] Cook P.R., "Music, Cognition, and Computerized Sound, An Introduction to Psychoacoustics", MIT Press, 1999.

[4] Czyzewski, A., *Speaker-Independent Recognition of Isolated Words Using Rough Sets*, J. Information Sciences, **104** (1998) 3–14.

[5] Czyzewski, A., *Intelligent Acquisition of Audio Signal Employing Neural Network and Rough Set Algorithms*, in "Rough-Neuro Computing: A Way to Computing with Words", Springer Verlag, Series on Artificial Intelligence, 2002.

[6] Grey, J.M., *Multidimensional perceptual scaling of musical timbres*, J. Acoust. Soc. Am., **61**, 5 (1977) 1270–1277.

[7] Herrera, P., X., Amatriain, E., Battle, and X., Serra, *Towards Instrument Segmentation for Music Content Description: A Critical Review of instrument classification techniques*, Proceedings Int. Symp. on Music Information Retrieval, ISMIR'2000, Plymouth, MA, (2000).

[8] Iverson, P., Krumhansl, C.L., *Isolating the dynamic attributes of musical timbre*, J. Acoust. Soc. Am., **94**, 5 (1993).

[9] Jensen, K., *The timbre model*, Workshop on current research directions in computer music, Barcelona, Spain, (2001).

[10] Kostek, B., "Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics", Studies in Fuzziness and Soft Computing, Physica Verlag, Heidelberg, New York, 1999.

[11] Kostek, B., and A., Czyzewski, *Representing Musical Instrument Sounds for Their Automatic Classification*, J. Audio Eng. Soc., **49**, 9 (2001) 768– 785.

[12] Krimphoff, J., S., McAdams, and S., Winsberg, *Caracterisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique*, J. Phys., **4** (1994) 625–628.

[13] Lindsay, A.T., Herre, J., *MPEG-7 and MPEG-7 Audio - an Overview*, J. Audio Engineering Society, **49**, 7/8 (2001) 589–594.

[14] McAdams, S., Winsberg, S., *Multidimensional scaling of musical timbre constrained by physical parameters*, J. Acoust. Soc. Am., **105** (1999) 1273 (A).

[15] MPEG-7 Overview, ISO/IEC JTC1/SC29/WG11, CODING OF MOVING PICTURES AND AUDIO, Klangenfurt, Germany, July (2002). URL:`http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm`

[16] Pawlak, Z., *Rough Sets*, J. Computer and Information Science, **11**, 5 (1982) 341–356.

[17] Pollard, H.F., and E.V. Jansson, *A Tristimulus Method for the Specification of Musical Timbre*, Acustica **51** (1982), 162–171.

[18] De Poli, G., Piccialli, A., and Roads, C., Eds., "Representations of Musical Signals", MIT Press, Cambridge, MA, 1991.

[19] De Poli, G., S., Canazza, C., Drioli, A., Roda, A., Vidolin, and P., Zanon, *Analysis and modeling of expressive intentions in music performance*, Proc. of the Intern. Workshop on Human Supervision and Control in Engineering and Music, Kassel, Germany, Sept. 21-24 (2001), URL:`http://www.engineeringandmusic.de/workshop/`.

[20] Pratt, R.L., and P.E., Doak, *A subjective rating scale for timbre*, J. Sound and Vibration, **45**, 3 (1976) 317–328.

[21] Pratt, R.L., and J.M., Bowsher, *The subjective assessment of trombone quality*, J. Sound and Vibration, **57**, 3, (1978) 425–435.

[22] Reuter, CH., *Karl Erich Schumann's principles of timbre as a helpful tool in stream segregation research*, Proc. II. Intern. Conf. on Cognitive Musicology, Belgium, (1996).

[23] Skowron A., *Data Filtration: a Rough Set Approach*, in "Rough Sets, Fuzzy Sets and Knowledge Discovery", (Ziarko W.P., Ed.), Springer-Verlag, 108–118, London 1994.

[24] Vercoe, B., Gardner, W., Schreier, E., *Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations*, Proc. IEEE, B6, 5, 1008.

[25] Wessel, D.L., *Timbre Space as a Musical Control Structure*, Computer Music J., **3**, 2 (1979) 45–52.

[26] Zadeh L., *Fuzzy logic = Computing with words*, IEEE Transactions on Fuzzy Systems, **2** (1996) 103-111.

[27] Zadeh, L., *From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions*, IEEE Transactions on Circuits and Systems, **45**, (1999) 105-119.

[28] Zadeh, L., *Fuzzy logic, neural networks, and soft computing*, Communications of the ACM, **37** (1994) 77–84.

[29] URL:`http://www.soft-computing.de/def.html`.

[30] URL:`http://www.pa.info.mie-u.ac.jp/WFSC/`.