Rough-Neuro Approach to Testing Influence of Visual Cues on Surround Sound Perception

Bozena Kostek

Technical University of Gdansk, Sound & Vision Engineering Department,
Narutowicza 11/12,
80-952 Gdansk, Poland
e-mail: bozenka@sound.eti.pg.gda.pl

Abstract. The paper aims at revealing in which way and how the surround sound interferes or is associated with the visual context. Such parameters as distance, angle or level of sound source were tested with and without video image presence in the screen. For that purpose subjective testing was applied. Processing of the obtained results has been done on the basis of the combined neural network and rough sets based algorithm. The main task of experiments was the application of modular neural networks for the purpose of quantization of the surround sound parameter values. The rough set algorithm made decisions showing the influence of visual cues on the perception of surround sound.

1 Introduction

The presented study shows the methodology of testing the influence of video image on the surround sound localization perception. Discovering such a relationship may result in formulation of some rules of surround sound production accompanied by video image. One can find references to the literature concerning audio-visual perception, but they are mostly related to the classical studies on this subject including stereo sound systems for HDTV (High Definition Television) [4][10][15][34][35]. At present, digital video, film or multimedia presentations are often accompanied by the surround sound. Home Theater Systems win popularity. Meanwhile, only few researchers made an effort to explore influence of video on surround source and vice versa [2][6][19][28][36]. However, there is still no clear answer as to the question how the video influences the localization of virtual sound sources in multichannel surround systems (e.g. DTS – Digital Theater System) and in most references one can find a list of problems to be solved while testing relevant inter-modal relations [7][8][28][29]. Therefore, several problems should be addressed, i.e. what is the optimum width of surround panorama for individual kinds of music? What kind of audio material rear loudspeakers should transmit? What changes in sound mix (if any) should be made when video zoom is modified? Such experiments should be based on the subjective testing procedures [7][17] in which experts should listen to the sound with- and without video image and provide assessments.

Experiments related to examination of human perception of surround audio and accompanying visual cues may be divided into several categories: testing without or with video image presence, low-level or high-level multimodal interrelation between audio and video signal. The first category of tests may be used for calibration tasks. In the second category of experiments one can use abstraction audio-visual objects, e.g. synthesized graphical primitive objects (numbers, circles, lines, ping-pong ball, etc.) and synthesized artificial sounds. To the third category, more complex objects may belong, for example an audio-video recording of a speaker, a soloist, etc. The reason for using simple objects instead of complex ones is a need of discovering and describing basic mechanisms underlying the audio-visual perception of human beings. In a simple way one may examine the influence of the shape, color and movement of the visual object on the localization of the sound in surround system. On the other hand, the high-level multi-modal interrelation experiments employing more complex objects may be used as a basis for adjustments in audio panorama settings during the audio-visual postproduction. Results of such experiments can show in which cases and in what way the video will affect the localization of virtual sound sources. In most cases video "attracts" the attention of the listener and, as a consequence, he or she localizes the sound closer to the screen center. Therefore, this effect is called image proximity effect or the sound localization shifting effect [7].

The main goal of this research study was to discover dependencies between reactions of sight and hearing senses due to perception of visual stimuli accompanied by surrounding sound. In order to achieve this aim a number of psycho-acoustic experiments had been conducted on a group of properly trained experts and as a result of those experiments a collection of data has been created. Those data were then analyzed by means of modern techniques of intelligent data exploration and knowledge discovery. The problem concerns mainly finding hidden relations between semantic descriptors of subjective impressions (in the form of words - adverbs). Thus for the purpose of this research study various soft computing techniques employing genetic algorithms, modular neural networks and rough sets were used. The detailed description of the results of those experiments is given in this paper.

2 Experiment Layout

2.1 Test Principles

Testing the influence of video image on the surround sound localization perception is a subjective process. It can be assigned to the category of the object evaluation processes. This means that during the evaluation process a number of properly trained and experienced in critical listening experts should take part by filling in a given questionnaire [17].

There are several issues to be checked during subjective test sessions. First of all, the software package along with the sound system should be calibrated. This is due to the fact that some phase shifting or delays may be present in electroacoustic channels.

The calibration process will be presented in the next paragraph. The second factor while testing is to check whether a so-called precedence effect influences sound perception. First of all, tests should be carried out without video image presence in the screen. Then, in the next step, sound may precede video image or vice-versa. The third case is when video image appears together with sound. Also, differences in sound level presentations may affect sound perception. Other factors that should be checked are image size and the stability of image. Obviously, in order to maintain tests reliable, also subjects' preferences should be taken into account. To that end, factors such as subject's gender, experience etc. are of importance. Since there is a large number of interrelated factors underlying tests, thus they will result in a huge number of experts' answers in some case contradictory to each other. To discover dependencies between obtained data some techniques belonging to the soft computing domain will be necessary.

Subjective testing sessions took place in two rooms, which are acoustically separated (see Fig. 1). The expert's seat is positioned in a so-called "sweet spot", the best place for listening.

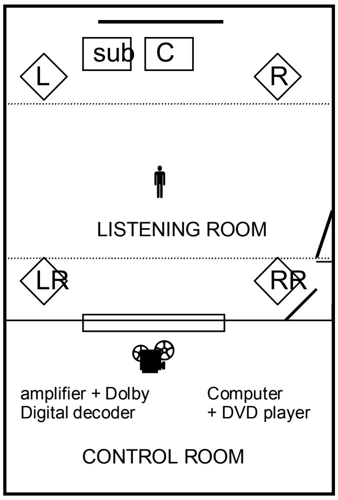

Fig. 1. Setup used during listening tests

The whole experimental setup consists in software package allowing for audio and video encoding, AC-3 encoder, computer with built in DVD player, amplifier with Dolby Digital decoder, video projector, screen, loudspeakers working in 5.1 system. During the tests files were used with audio encoded in the AC-3 (Dolby Digital) format and video encoded in MPEG2 standard. All audio files contained five channels (without ultra low frequency channel) and bit-rate equal to 448 kbit/s. Sound files were prepared with the *Samplitude 2496* software and then exported to the AC-3 encoder (*A.pack*). The video files were prepared with the *Adobe Premiere* software. All video files had resolution 720x576 and relevant quality. It prevented possible influence of video quality on experts' judgment [3][12][11]. After encoding, audio and video files were multiplexed into *VOB* files.

2.2 Test Procedure Setup

The calibration procedure consisted among others in checking whether loudspeakers can be replaced by the phantom sources created in the software. In Fig. 2 two diagrams are shown pertaining angle and distance sound source localization in the presence of video employing both really existing loudspeakers and phantom sound sources. The arrangement of loudspeakers was as follows: four loudspeakers were aligned along the left-hand side of the screen (angle localization) and in the second case positioned between the listener and the screen (distance localization). In the first case, loudspeaker No. 1 was placed at the edge of the room, whereas the fourth one was positioned directly under the screen. In the second case loudspeaker No. 1 was the closest to the listener. While using phantom sources the arrangement of loudspeakers was as shown before in Fig. 1. First experts listened to sound samples (no video) and their task was to determine from which loudspeaker a particular sound sample was heard. Then, in the second phase of experiments an object was displayed in the screen with a synchronously generated sound sample.

As is seen from Fig. 2, results lying on the diagonal of the diagrams refer to the situation in which there was no video present while listening to sound samples. On the other hand, there is a shift caused by the image appearance. This means that sound sample transmitted for example from the loudspeaker No. 1 (the most distant from the screen) was perceived as the one transmitted closer to the screen. As is indicated by the standard deviation measures seen in diagrams, experts differently localized perceived sound samples, thus it may concluded that this effect is both expert- and sound type-dependent. On the other hand, experts have no difficulties while perceiving a sound sample either from a loudspeaker or listening to the phantom source.

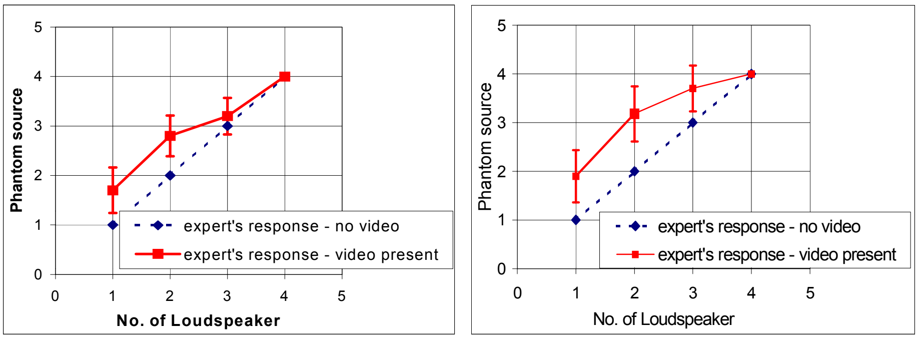

Fig. 2. Comparison of answers for two types of experiments: sound source angle localization shift caused by the image appearance (left-hand), sound source distance localization shift caused by the image appearance (right-hand); loudspeaker No. 4 was the closest to the screen

The list of audio-visual signals used in experiments is presented in Table 1. They are both low- and high-level. The first stage of experiments consisted of a series of five audio presentations without accompanying video. These tests called mapping tests aimed at checking both correctness of directional hearing of an expert and experts' reliability. As was said, sounds were presented without image – the screen was blank.

If the listener did not localize sounds properly, his or her results were acknowledged as not qualifying for further processing. The number of experts participating in experiments was 34. This group consisted of staff members and students of the Sound and Vision Engineering Department. After statistically checking experts' answers it appeared that four of them should be excluded from this group due to some mistakes in localization of sound arrival direction.

Table 1. Description of audio-visual signals used in tests

Low-level inter-modal relations - Abstraction tests	High-level inter-modal relations - Thematic tests
video: blinking circle	video: talking speaker
audio: amplitude-modulated tone	audio: speaker' voice
video: circle with modulated colors	video: musician playing solo
audio: amplitude-modulated tone	audio: musical excerpt
video: bouncing ping-pong ball	video: musical band playing
audio: ping-pong ball sound	audio: musical excerpt
video: metronome	video: musical video-clip
audio: metronome sound	audio: music from a video-clip
video: vertical bar moving from right to left	video: film excerpt
audio: filtered noise	audio: sound track from the movie

The questionnaire form for assigning arrival directivity of the sound in the mapping tests is shown in Fig. 3. While considering a spherical space around the listener head, this space can be sampled at different elevations (from below the horizontal plane to directly overhead). In addition, at each elevation the full 360 degrees of azimuth can be sampled in equally sized increments. A total of some hundred of locations can be obtained in this way. However, in the experiments only horizontal plane was considered, of which the division of angles is seen in Fig. 3.

Apart from mapping tests, in the experiments 10 abstraction tests and 15 or 20 high level-abstraction tests were presented to experts'. The abstraction tests used simple objects instead of complex ones. That was due to the need of discovering and describing basic mechanisms underlying the audio-visual perception.

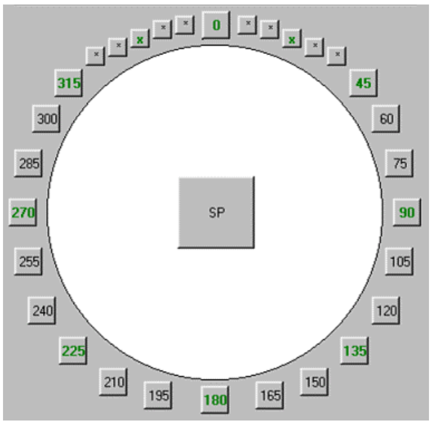

Fig. 3. Questionnaire form for assigning directivity of sound arrival

Surround sound systems allow creating phantom sound sources in 360^0 range. However using too many sound sources may introduce some errors due to inaccuracy of phantom sound sources positioning. Thus, the number of sources was limited to the following angles: 0 (central loudspeaker), 22.5^0, 45^0 (front right loudspeaker), 90^0, 135^0 (rear right loudspeaker), 180^0, 225^0 (rear left loudspeaker), 270^0, 315^0 (front left loudspeaker), 338^0. For the purpose of increasing number of possible answers experts could choose also other angles: 7.5^0, 15^0, 30^0, 37.5^0, 60^0, 75^0, 105^0, 120^0, 150^0, 165^0, 195^0, 210^0, 240^0, 255^0, 285^0, 300^0, 322.5^0, 330^0, 345^0, 352.5^0. Furthermore, in order to allow an expert expressing more spatial-like impressions - not only those angle-oriented, but also some angle-group oriented entities were added, such as: L+C+R – wide central base (315^0+0^0+45^0), WF – wide front base (315^0+45^0), WR –wide right base (45^0+135^0), WB – wide back base (135^0+225^0), WL – wide left base (225^0+315^0), SS – Sweet Spot, ALL – all five channels playing simultaneously. In this way attributes defining the sound domain space were assigned.

The visual domain space was described with only one attribute assigned to thematic tests indicating whether video was present or not. In the abstraction tests several attributes were added describing for example how the line was moving on the screen: L2R – from left to right, R2L – from right to left, D2U – up, U2D – down. All those given attribute sets served as a basis for determining the structure of decision rules discovered by the data mining system.

It is important to point out that the assumption was made that all the parameters in both visual and sound domains could contain only binary data. This means that a given angle could be either completely included in the perception of a surrounding sound or completely excluded from it. Similarly, images could be used for a given test or not.

3 Genetic Algorithm-Based Processing of Listening Test Results

3.1 Knowledge Base

Methodology based on searching for repetitive patterns existing in data and association rule generation from those patterns was used for data mining in this research study. Data were represented as a simple information system. An example of a data record from the information system is shown in Fig. 4. A record in the abstraction test database (Fig. 4) contains values of 1 at 4th, 11th, 13th and 22nd positions (0's elsewhere). This means that a sound stimulus presented at 90° (4th attribute) accompanied by an image (11th attribute) of a vertical line moving from right to left side of the screen was actually localized by an expert at 45° (22nd attribute).

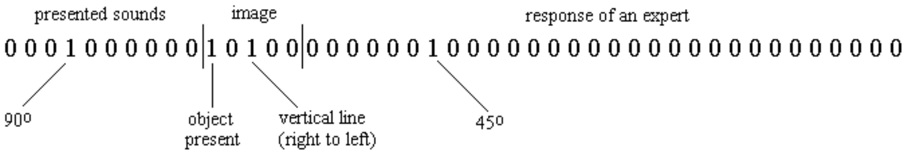

Fig. 4. Example of a record in the database (abstraction tests case)

After creating the appropriate data sets, it was possible to explore and analyze the data. The aim was to discover the influence of visual stimuli on the perception of a sound in surround space, thus searching for association rules was performed. The genetic algorithm was employed to this task. Since genetic algorithms belong to the most often used soft computing methods, thus their principles will be not reviewed here.

In this research study, the chromosomes that are being produced and modified during the evolution process represent patterns covering records in the data set. Each one of them has the length of the number of attributes describing the data (specific for the type of the tests – abstraction vs. thematic), and the alleles of the chromosome are constrained by the domains of those attributes. An allele of such a chromosome can either contain a value that is valid for a corresponding attribute in the data set (in this case 1's, all 0's can be omitted since such a testing is aimed at interrelation of angle and image) or a "don't care" asterisks which means that this attribute is not important and will not be used for generation of a rule [8]. An example of a chromosome is presented in Fig. 5.

```
* * * 1 * * * * * * 1 * 1 * * * * * * * * 1 * * * * * * * * * * * * * * * * * * * * * * *
```

Fig. 5. Example of a chromosome (set positions – 4th, 11th, 13th, 22nd)

Each of such patterns has a possible coverage in the data (support) which is given by the number of records matching the pattern (i.e. having given values at the set position). For the above example it will be all records containing "1" at 4th, 11th, 13th and 22nd positions regardless of other values. Obviously, one should look for patterns that have relatively high support and this can form the basis of the fitness function used for this algorithm. The desired level of support in data can be adjusted by setting the *epsilon* value, which stands for the percent-based, maximum allowed error in terms of pattern coverage (the higher epsilon, the lower minimum support required) [8][32].

Although the support of a pattern is a basic feature of the fitness function implemented in the algorithm, it cannot be its ultimate characteristic. The number of "set" positions (not the "*don't care*" asterisks) is also very important. For example, a pattern consisting only of asterisks will gain support of 100% of the data records, but it has no meaning in terms of knowledge discovery. The structure of the *IF–THEN* rules generated afterwards is also very important, and from the practical point of view patterns must contain at least two (or even three) set attribute values in order to stand as a basis for any useful association rules. Such a rule should have the following structure:

$$\{\text{presented sound}\} \cup \{\text{image}\} \Rightarrow \{\text{response of an expert}\},$$

Obviously, not all the chromosomes will have a physical coverage in the available data set. Some of them (especially the ones with relatively large number of set positions) might not have a support at all, however some parts of them (subsets of values) still can be very useful and after an application of some genetic operators (i.e. crossover and mutation) may produce desired result. It is crucial then to appropriately treat all those chromosomes and assign them some "credit" in terms of the fitness function even though they do not have support in data as a whole.

Based on the above discussion all the chromosomes (potential solutions) should be awarded or punished according to the criteria during the evolutionary process. Thus the fitness function can be completely described as a multi-layer estimation of the fitness of the chromosomes in terms of their partial support in the data at first, and then total coverage of the data weighted by the number of set positions.

Another very important feature of the genetic algorithm used here is a multi-point crossover option. In many experiments mining patterns in different types of data, this approach was found to be much more effective with regard to both the number of discovered patterns, and the time of convergence. On the basis of empirical premises the maximal number of cuts (crossover points) was set to 1 for every 10 attributes. In the example given in Fig. 6 there are three crossover points and the arrows point out to the genetic material that will be exchanged and thus will create two new chromosomes.

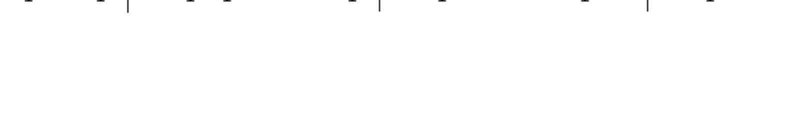

Fig. 6. Example of a multi-point crossover (three point)

As an outcome of several evolutions modeled by this genetic algorithm, a set of data patterns was created. Those patterns along with the information about the level of their support were then used as an input to the application generating association rules.

Association rules determine existence of some relations between attributes in data or values of those attributes. Basically they are simple *IF–THEN* type rules that, for binary domain of values, can be considered as statements [8]:

> *"if attributes from the premise part of the rule have values of 1 then the attributes included in the consequent part also tend to have value of 1".*

In the discussed case, rules should be of the following type:

> *"if a given set of angles was used for the reproduction of a sound and image was/was not present then the experts tended to localize the sound source at a particular angle/set of angles"*,

Association rules are characterized both by their support in data (number of cases that a given rule applies to – how "popular" the rule is) and the confidence (ratio of the support of the rule to the number of cases that contain its premise part – revealing how sure one can be that judging on the basis of the values from the premise part of the rule the rule is correct).

An algorithm of searching for association rules consists of two parts: searching for patterns hidden in data (in this project this was achieved initially by the application of the genetic algorithm) and generating rules based on those patterns. The idea of the algorithm for rule generation in this research study is relatively simple. Basically it takes "not asterisk" values of each of the patterns, divides them into subsets, and by moving those subsets from the premise to the consequent part (according to the specified constraints) creates all possible rules based on the given pattern. The algorithm is quite resource consuming, thus it removes all records that are covered by any others (i.e. those that are subsets of another set). This decreases the computational complexity of the algorithm and together with the support and confidence parameter limits the number of generated rules.

3.2 Pattern Searching

In order to increase the variety of patterns, the algorithm was launched on several computers simultaneously. Because of the randomness aspects of genetic algorithms, the results differed from each other. However, some of those results were duplicated.

The support threshold of desired patterns was lowered to 5%. This seems to be extremely low, but it is valid because rules based on patterns with relatively small support in data still may have quite a large level of confidence. As a result of several evolutions of the Genetic Algorithm, a total of 806 distinct patterns for the abstraction test, and 890 for the subject test were found. Some of those patterns were characterized by including set values only in the range of generated locations (angles), and not the ones that were a response from an expert. This was quite obvious, taking into consideration the fact that a big part of the generated tests consisted of different angles at the same time (e.g. WF, L+C+R, etc. – see the description of the sound space), and an appropriately engineered algorithm should definitely find them. However some patterns that were satisfactory in terms of the rule definition were also discovered (i.e. they consisted of generated locations that were perceived by an expert, as well as the information about the image presented). Some examples of those patterns are given below:

(Abstraction tests; *ABSTRACT* space – 45 attributes):
{1 * 1 * * * 1 * 1 * * * * * * * 1 * * * * * * * * * * 1 * * * 1 * * * * * 1 * * * * * * *}
support: [*18/300*]
(Thematic tests; *THEMATIC* space – 41 attributes):
{1 1 * * * 1 1 * 1 * * * * * * * * * * * * * * 1 * * * * * * * * * * * * * * * * *}
support: [*16/465*]

3.3 Rule Generation

Patterns discovered and prepared in the previous step were then used as a basis for associative rule generation. At this level, sets of attributes were divided into premise (generated angles along with the information about an image) and consequent (response from an expert) parts. After removal of duplicated patterns, 49 effective patterns for abstraction and 23 for thematic tests were preserved. On the basis of this final set of patterns, a number of rules of a given support and confidence was generated. Some of rules indicated a lack of any influence of the image on the perceived localization of sound, and this was usually connected to the sounds perceived from behind of the listener. Nevertheless, most rules proved an existing interrelation between the auditory and visual senses. A sample of such a rule is presented below for the case of abstraction tests; this is a rule with clear indication of audio-visual dependencies:

IF i045=1 AND i135=1 AND P=1 THEN 45=1 [s=6%] [c=66%]
(IF sound is presented at the angles of 45° and 135° and the image is present THEN the perceived angle is 45° WITH support of 6% and confidence of 66%)

IF i225=1 AND i315=1 AND P=1 AND D2U=1 THEN 315=1 [s=4%] [c=75%]
(IF sound is presented at the angles of 225° and 315° and there is an image of a horizontal line moving from down up THEN the perceived angle is 315° WITH support of 4% and confidence of 75%)

Based on the performed experiments it may be concluded that rules generated by the genetic algorithm proved an existence of the proximity effect while perceiving sound in the presence of video image. However the support for these rules is so low that it is difficult to conclude whether these rules are valid, even if the confidence related to such rules is quite high. That is why in the next paragraph another approach to processing of the data obtained in subjective tests will be presented. It concerns a hybrid system consisted of modular neural network and rough set-based inference system.

4 Rough-Neuro Hybridization

4.1 Hybrid Neural Networks

Enumerated applications of artificial neural networks to various fields imposed on development in theory. Due to this fact some new trends in this domain appeared. One of these trends involves compound structure of neural networks, so-called hierarchical neural networks. The basic network structure is composed of numbers of subnetworks. These subnetworks have common input layer. Their middle layers are independent from one another. Every subnetwork has an assigned output node [18].

Another trend that differs much from the *All-Class-One-Network* is a modular neural network concept. In this case information supplied by the outputs of subnetworks can be fused by applying either fuzzy or rough set approach. Hybrid methods have been developed by integrating merits of various paradigms to solve problems more efficiently. It is often pointed out that hierarchical or modular neural networks are especially useful while discussing complex classification tasks involving large number of similar classes. In such a case one can refer to some sources that appeared lately in the literature [5][20][21][30][33].

Feature subset selection by neuro-hybridization was presented as one of the most important aspects in machine learning and data mining applications by Chakraborty [5]. He engineered the neuro-rough hybrid algorithm that uses rough set theory in the first stage in order to eliminate redundant features. Then a neural network used in the second stage operates on a reduced feature set. On the other hand, Auda and Kamel proposed a modular neural network that consists in unsupervised network to decompose classification task over a number of neural subnetworks. Then information from the outputs of such modules are integrated via a multi-module decision-making strategy that has the ability to classify a tested sample as "vague class" or boundary between two or more classes [1]. The paper by Peters *et al.* reviews the design and application of neural networks with two types of rough neurons: approximation neurons and decider neurons [26]. The paper particularly considers the design of rough neural networks based on rough membership functions, the notion introduced recently by Pawlak and Skowron [25]. A so-called rough membership neural network consists of a layer of approximation neurons that construct rough sets. The output of each approximation neuron is computed with a rough membership function. Values produced by the layer of approximation neurons provide condition vectors. The output layer is built of decider neuron that is stimulated by each new condition vector. A decider neuron compares the new condition vector with existing ones extracted from decision tables and returns the best fit. The decider neuron enforces rules extracted from decision tables. Information granules in the form of rules are extracted from decision tables using rough set method [26]. Also other approaches based on modular and complex integral neural networks are widely used in various problems as robust search methods, especially in cases of uncertainty and redundancy in data [24][27][31].

4.2 Rough-Neuro System Principles

As was mentioned before, at least three factors should be taken into account while testing surround sound perception accompanied by video. They are such as follows: sound arrival angle, distance, and level of the sound. It is obvious that all these three might be interrelated, however as was shown in previous study, employing subjective tests based on fuzzy logic technique [17], it was sufficient to work with a single function separately and then, to interrelate these factors in some rule premises [7]. Rule premises contained the above mentioned factors and assessed descriptors assigned to them during subjective testing sessions, and the consequence (decision) resulted from these test data. However rules that were formulated were hard to verify by experts. Therefore, in this study a new concept of rule discovering was conceived. For this purpose a modular rough-neuro system was engineered that is described further on (Fig. 7).

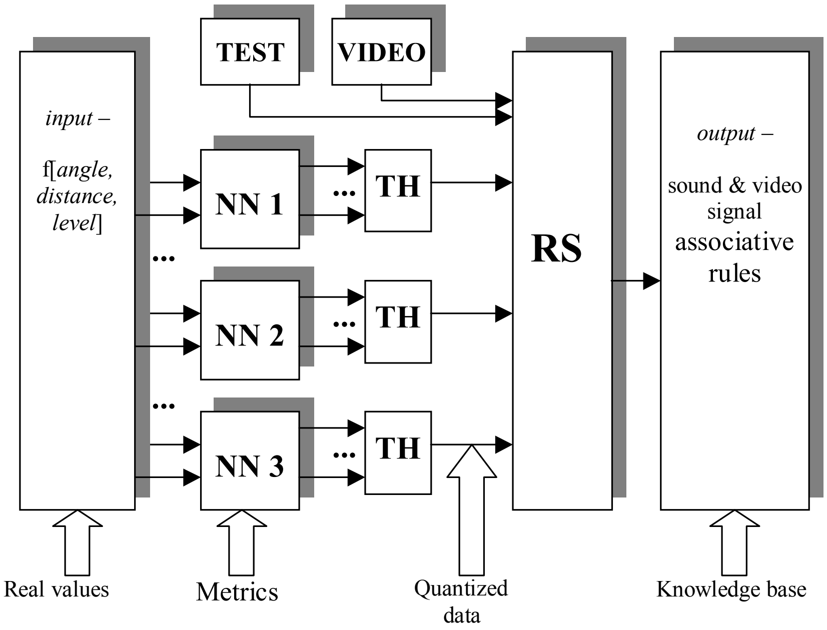

Fig. 7. Rough-neuro system lay-out

As seen from Fig. 7, the main two blocks of the rough-neuro system are related to data processing. These are neural network modules that quantize numerical data and the rough set-based engine that extract rules from data. The elements of the input vector shown in Fig. 7 are numbers representing realm of angles, distances and sound loudness values, whereas rough set-based decision system requires quantized data. Consequently, for quantization purposes, self-organizing neural net is proposed, similarly like in other experiments [9]. For this purpose self-organizing map (SOM) introduced by Kohonen has been chosen [13]. This is one of the best known neural network clustering algorithms, which assigns data to one of specified subsets according to the clusters detected in a competitive learning process. During this learning only the weight vector which is most similar to a given input vector is accepted for weight building. Since data can be interpreted as points in a metric space, thus each pattern is considered as a N-dimensional feature vector, and patterns belonging to the same cluster are expected to possess strong interval similarity according to the chosen measure of similarity. Typically, the Euclidean metric is used in SOM implementations [13].

Using the SOM as a data quantizer, a scalar- and a vector quantization can be taken into account. In the first case, the SOM is supplied to a single element of the key vector. In the second case, a few attributes can constitute input vectors, which lowers the number of attributes helping to avoid a large number of attribute combinations in the rough set inference. 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 [13]. As a result of the competition, this c-th output unit wins provided the following relations are fulfilled.

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 value for the learning rate $\eta^{(0)}$ was equal to 0.95. For the purposes of the rough-neuro hybridization, at the end of the weight adaptation process the output units should be labeled with some symbols. It is done in order to assign quantized input data to symbols, which are to be processed by the rough set rule induction algorithm.

The engineered rule induction algorithm is based on the rough set methodology well described in the literature [16][22][23]. The employed algorithm aimed at reducing the computational complexity [9]. This concerns the values reduction of attributes and searching for reducts, so that all combinations of the conditional attributes are analyzed using reasonable computational cost. Particularly, for a given sorted table, the optimum 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 analyzed using special way of attribute sorting [9]. The algorithm splits the decision table into two tables: consisting of only certain rules and of only uncertain ones. There is an additional information associated with every object in them. The information concerns the minimal set of indispensable attributes and the rough measure μ_{RS}. The latter case is applied only for uncertain rules. More details corresponding to the rough set algorithm can be found in the literature [9].

4.3 Experiments

Results from test sessions gathered in a database (see Table 2) are then further processed. The type of the test provides therefore one of attributes that are contained in the decision table. Other attributes included in the decision table are: "*angle*", " *distance*", "*level*", '*video*" and a decision attribute called "*proximity_effect*". To differentiate between attributes resulting from experts' answers and actual values of angle, distance and level known to the experimenter two adjectives, namely “subjective” and “objective' were added to attribute names and from this resulted six attributes that were contained in the table (Tab. 2). As was mentioned before, during the test session experts are asked to fill in questionnaire forms, an example of which was shown in Fig. 3. As there are numerical values gathered using questionnaires, thus values indicated by experts are then forming a feature vector that is fed to the neural networks modules. The neural network module assigns each numerical value indicated by an expert to one of clusters corresponding to semantic descriptors. The selection of the strongest output of the neural network is done by adding a threshold function operating in the range of (-1,1) to the system shown in Fig. 7. These threshold filters connected to outputs of the NN can be realized in practice by the output neuron transfer function. Their role is to choose only the strongest values obtained in the clustering process. Therefore Table 2 contains descriptors resulted from neural network-based quantization related to “*angle_subjective*”, “*distance_subjective*”, “*level_subjective*” attributes. Semantic descriptors related to the *angle_subjective* attribute are as follows: “*none*”, “*front*”, *left_front*”, “*left*”, *left_rear*”, “*rear*”, *right_rear*”, *right*”, “*right_front*”. All but one such a descriptor is obvious. The “*none*” descriptor is related to the case when distance equals 0, thus a phantom source is positioned in a so called “*sweet-spot*” (expert's head position). This means that sound is subjectively perceived as directly transmitted to the head of an expert so there is no angle of its arrival defined. In addition, *distance_subjective* is quantized by the NN module as: “*none*”(“sweet-spot” location), “*close*” (large distance from the screen), “*medium*”, “*far*” (smaller distances from the screen), and correspondingly *level_subjective* is denoted as: *low*, *medium*, *high*.

Table 2. Decision table (fragment)

Experts' answers	*Angle_ subj.*	*Dis tance_ subj.*	*Level_ subj.*	*Angle_ objective*	**...**	*Test*	*Video*	*Decision - Proximity_ effect*
e_1	*front*	*close*	*medium*	0^0	**...**	“*abstraction*”	“*static_ image*”	*no_ shift*
e_2	*left_front*	*close*	*high*	60^0	**...**			*medium_ shift*
...	**...**	**...**	**...**	**...**	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**	**...**	**...**	**...**	**...**
e_n	*left_rear*	*far*	*medium*	315^0	...	“*thematic*”	*dynamic _image*	*strong_ shift*

Values of angle, distance and level of the phantom sound source are given numerically by the experimenter, however they are quantized values (angles in degrees, distance in centimeters and level in dB). The range of angle attribute was already shown in Paragraph 2.2. The quantization resolution of angles and distance was directly related to the practically available resolution of phantom sound sources in the applied

Samplitude 2496 software (see Fig. 8). Level values were quantized in the range of (50 dB to 100 dB) with 10 dB step. The problem of quantization of level, distance and level attributes is further complicated because of some acoustical principles, which will be however not reviewed here. The values of these attributes were left in the numerical form, because in this case rules will be easily understandable. On the other hand, descriptors related to "*test*", "*video*" attributes and *proximity effect* attribute were set as semantic descriptors. Therefore "*sound*" and "*video*" attributes can have values such as follows: "*abstraction*", "*thematic*" and correspondingly: "*no_video*", "*static_image*", "*dynamic_image*". The decision attribute can be read as "*no_shift*", "*slight_shift*", "*medium_shift*" and "*strong_shift*" and these descriptors will appear in the consequence part of a rule.

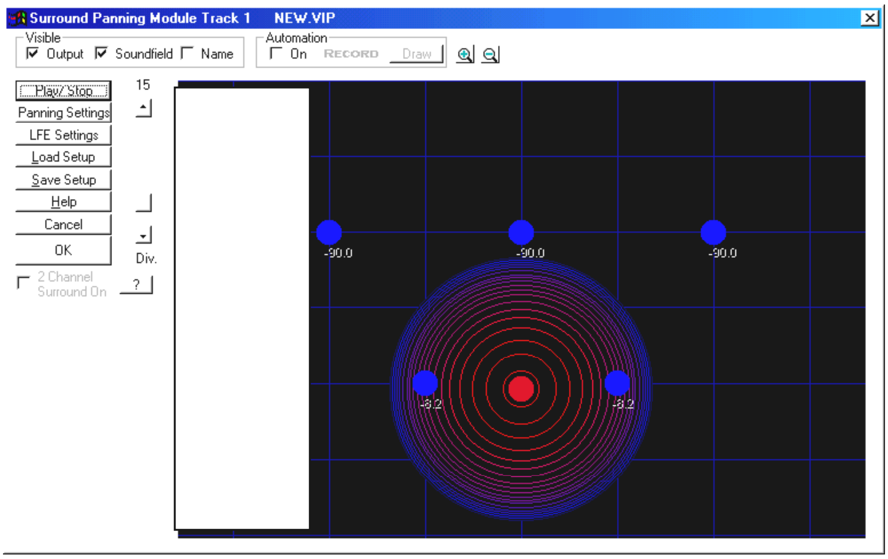

Fig. 8. Phantom sound source created by the *Samplitude 2496* application

Sample rules that can be derived from the decision table are presented below:

if Angle_subjective=*front* AND Distance_subjective=*close* AND Level_subjective=*medium* AND Angle_objective= 0^0 AND Distance_objective=*20* AND Level_objective=*70* AND Test=*abstraction* AND Video=*static* than Proximity effect= *no_shift*
if Angle_subjective=*left_front* AND Distance_subjective=*close* AND Level_subjective=*medium* AND Angle_objective= 60^0 AND Distance_objective=*20* AND Level_objective=*70* AND Test=*thematic* AND Video=*static* than Proximity effect=*slight_shift*
if Angle_subjective=*front* AND Distance_subjective=*far* AND Level_subjective=*high* AND Angle_objective= 45^0 AND Distance_objective=*20* AND Level_objective=*90* AND Test=*abstraction* AND Video=*static* than Proximity effect= *strong_shift*
....

Rules that will have a high value of the rough set measure can be included in the knowledge base to be used for investigating psychological principles of sound & vision interaction.

7 Conclusions

The subjective listening tests proved that visual objects could influence the subjective localization of sound sources. Measurement data showed that visual objects may “attract” the listeners’ attention thus in some cases sound sources are then localized closer to the screen. It was found that the image proximity effect is listener-dependent, what is probably related to some psychological processes occurring in the individual human brains.

As is seen from the presented concepts and experiments, numerical values and subjective descriptors gathered in the decision table can be processed by the hybridized rough-neuro algorithm. In this way a concept of simultaneous computing with numerical data and with words was applied allowing for the processing of data obtained from both: objective values and their subjective counterparts.

On the basis of the experiments described in this paper and opinions of experts taking part in them, it can be stated that subjective listening tests and soft computing processing of their results seem appropriate for the analysis of hearing and sight hidden relations. It creates an environment for the automatic exploration of data derived from psychoacoustic experiments with surround sound and accompanying vision, employing knowledge discovery based on soft computing–oriented methodologies. The results of such experiments and their analysis could yield the recommendations to sound engineers producing surround movie sound tracks, digital video and multimedia content.

Acknowledgements

The research is sponsored by the Committee for Scientific Research, Warsaw, Poland. Grant No. 8 T11D 00218.

References

1. Auda, G., Kamel, M.: A Modular Neural Network for Vague Classification. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC’2000). Banff, Canada, October 16-19 (2000) 545-550
2. Bech, S., Hansen, V., Woszczyk, W.: Interactions Between Audio-Visual Factors in a Home Theater System: Experimental Results. 99th Audio Eng. Soc. Conv., New York, Preprint No. 4096, October (1995)

3. Beerends, J.G., de Caluwe, F.E.: The Influence of Video Quality on Perceived Audio Quality and Vice Versa. J. Audio Eng. Soc., Vol. 47, No. 5, (1999) 355-362
4. Brook, M., Danilenko, L., Strasser, W.: Wie bewertet der Zuschauer das stereofone Fernsehes?. 13 Tonemeistertagung; Internationaler Kongres, (1984) 367-377
5. Chakraborty, B.: Feature Subset Selection by Neuro-Rough Hybridization. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 481-487
6. Czyzewski, A., Kostek, B., Odya, P., Zielinski, S.: Influence of Visual Cues on the Perception of Surround Sound. 139th Meeting of the Acoustical Society of America, J. Acoust. Soc. Amer., No. 5, V ol. 107, No. 3aPP14, p. 2851, Atlanta, GA USA (2000)
7. Czyzewski, A., Kostek, B., Odya, P., Zielinski, S.: Determining Influence of Visual Cues on the Perception of Surround Sound Using Soft Computing. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 507-514
8. Czyzewski, A., Kostek, B., Odya, P., Smolinski, T.: Discovering the Influence of Visual Stimuli on The Perception of Surround Sound Using Genetic Algorithms. In: Proc. 19th Int. Audio Eng. Soc. Conference. Germany, (2001) 287 – 294
9. Czyzewski, A., Krolikowski, R.: Neuro-Rough Control of Masking Thresholds for Audio Signal Enhancement. In: Neurocomputing, Vol. 36 (2001) 5-27
10. Gardner, M.B.: Proximity Image Effect in Sound Localization. J. Acoust. Soc. Amer. Vol. 43 (1968) 163
11. Hollier, M.P., Voelcker, R.: Objective Performance Assessment: Video Quality as an Influence on Audio Perception. 103rd Eng. Soc. Conv., New York, Preprint No. 4590, (1997)
12. Kaminski, J., Malasiewicz, M.: Investigation of influence of visual cues on perceived sound in the surround system. M.Sc. Thesis, Sound and Vision Eng. Dept., Technical Univ. of Gdansk, Poland (*in Polish*) (2001)
13. Kohonen, T.: The Self-Organizing Map. Proc. IEEE, 78 (1990) 1464-1477
14. Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J.: Engineering Applications of the Self-Organizing Map, Proc. IEEE, 84 (1996) 1358-1384
15. Komiyama, S.: Subjective Evaluation of Angular Displacement Between Picture and Sound Directions for HDTV Sound Systems. J. Audio Eng. Soc. Vol. 37 (1989) 210
16. Komorowski, J., Pawlak, Z., Polkowski, L. and Skowron, A.: Rough Sets: A Tutorial. In: Pal, S.K., Skowron, A. (eds.): Rough Fuzzy Hybridization. A New Trend in Decision-Making. Springer Verlag, Singapore (1999) 3-98
17. 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, Heilderberg New York (1999)
18. Liqing, Z.: A New Compound Structure of Hierarchical Neural Networks. In Proc. of IEEE World Congress on Computational Intelligence, ICEC'98. Anchorage, Alaska (1998) 437-440
19. Meares, D.J.: Perceptual Attributes of Multichannel Sound. Proc. AES 12th International Conference, Copenhagen, Denmark (1993) 171-179
20. Mitra, S., Pal, K. S., Banerjee, M.: Rough Fuzzy Knowledge-based Network – A Soft Computing Approach. In: Pal, S.K., Skowron, A., (eds.): New Trend in Decision-Making. Springer-Verlag, Singapore Berlin Heidelberg (1999) 428-454
21. Pan, Y., Shi, H., Li, L.: The Behavior of the Complex Integral Neural Network. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 585-592
22. Pawlak, Z.: Rough Sets. In: J. Computer and Information Science. Vol. 11, No. 5, 1982 341-356

23. Pawlak, Z.: Rough Sets - Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
24. Pawlak, Z.: Rough Sets and Decision Algorithms. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 1-16
25. Pawlak, Z., Skowron, A.: Rough Membership Functions. In: Yager, R., Fedrizzi, M., Kacprzyk, J. (eds): Advances in the Dempster-Shafer Theory of Evidence. John Wiley & Sons New York (1994) 251-271
26. Peters, J.F., Skowron, A., Han. L., Ramanna S.: Towards Rough Neural Computing Based on Rough Membership Functions: Theory and Application. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 572-579
27. Polkowski, L., Skowron, A.: Rough-Neuro Computing. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 25-32
28. Sakamoto, N., Gotoh, T., Kogure, T., Shimbo, M.: Controlling Sound-Image Localization in Stereophonic Reproduction. J. Audio Eng. Soc., Vol . 29, No. 11, (1981) 794-798
29. Sakamoto, N., Gotoh, T., Kogure, T., Shimbo, M.: Controlling Sound-Image Localization in Stereophonic Reproduction: Part II. J. Audio Eng. Soc., Vol . 30, No. 10 (1982) 719-721
30. Sarkar, M., Yegnanarayana, B.: Application of Fuzzy-Rough Sets in Modular Neural Networks, In: Pal, S.K., Skowron, A., (eds.): New Trend in Decision-Making. Springer-Verlag, Singapore Berlin Heidelberg (1999) 410-427.
31. Skowron, A., Stepaniuk, J., Peters, J.F.: Approximation of Information Granule Sets. In: Proc. 2nd Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, October 16-19 (2000) 33-40
32. Smolinski, T., Tchorzewski T.: A System of Investigation of Visual and Auditory Sensory Correlation in Image Perception in the Presence of Surround Sound. M.Sc. Thesis, Polish – Japanese Institute of Information Technology, Poland (in Polish) (2001)
33. Szczuka, M.S.: Rough Sets and Artificial Neural Networks. In: Pal, S.K., Skowron, A., (eds.): Rough Sets in Knowledge Discovery: Applications, Case Studies and Software Systems. Physica-Verlag, Heidelberg New York (1998) 449-470
34. Thomas, G.J.: Experimental Study of the Influence of Vision of Sound Localization. J. Exp. Psych., Vol . 28, (1941) 163-177
35. Wladyka, M.: Examination of Subjective Localization of Two Sound Sources in Stereo Television Picture. M.Sc. Thesis, Sound Eng. Dept., Technical Univ. of Gdansk, Poland (*in Polish*) (1987)
36. Woszczyk, W., Bech, S., Hansen, V.: Interactions Between Audio-Visual Factors in a Home Theater System: Definition of Subjective Attributes. 99th Audio Eng. Soc. Conv., New York, Preprint No. 4133 (1995)