Work Directions and New Results in Electronic Travel Aids for Blind and Visually Impaired People

VIRGIL TIPONUT, DANIEL IANCHIS, ZOLTAN HARASZY, MIHAI-EMANUEL BASCH, IVAN BOGDANOV
Department of Applied Electronics
POLITEHNICA University of Timisoara
Bvd. Vasile Parvan 2, 300223 Timisoara
ROMANIA
virgil.tiponut@etc.upt.ro, daniel.ianchis@etc.upt.ro, zoltan.haraszy@etc.upt.ro,
bash_mihai@yahoo.com, ivan.bogdanov@etc.upt.ro
http://www.ea.etc.upt.ro/index.php/en/personal

Abstract: - Many efforts have been invested in the last years, based on sensor technology and signal processing, to develop electronic travel aids (ETA) capable to improve the mobility of blind users in unknown or dynamically changing environment. In spite of these efforts, the already proposed ETAs do not meet the requirements of the blind community and the traditional tools (white cane and guiding dogs) are still the only used by visually impaired. In this paper, research efforts to improve the main two components of an ETA tool: the Obstacles Detection System (ODS) and the Man-machine Interface (MMI) are presented. Now, for the first time, the ODS under development is bioinspired from the visual system of insects, particularly from the locust and from the fly. Some original results of the author's team, related to the new concept of Acoustical Virtual Reality (AVR) used as a MMI is then discussed in more detail. Some conclusions and the further developments in this area are also presented.

Key-Words: - visually impaired persons, electronic travel aids, obstacles detection system, man-machine interface, acoustical virtual reality.

1 Introduction

In the last years, the traditional tools used by visually impaired to navigate in real outdoor environments (white cane and guiding dogs) are to be substituted with electronic travel aids (ETA). These devices are capable to improve the mobility of blind users in unknown or dynamically changing environment.

An ETA tool includes the following main components [1]:

- An obstacles detection system (ODS),
- A path planning module,
- A man-machine interface (MMI) and,
- A monitoring system.

The monitoring system is tracking the movement of the blind persons in order to be sure that they are in progress and capable to reach the target. Moreover, it is important to know in every moment the actual position of the subjects, in order to be able to help them in case of dynamic changing environments or more importantly, in case of emergency.

The path-planning module is responsible for the generation of the path to the desired target, with obstacles avoidance. The positions of the obstacles in front of the subject are determined by a 3D obstacles detection system. The last two mentioned components should meet requirements similar to the requirements for the global path planning and obstacles detection in mobile robotics.

The MMI is capable of offering in a friendly way the information extracted from the surroundings, assisting the visually impaired individuals with hands-free navigation in their working and living environment.

It should be mentioned here that there are known solutions for the path planning module and monitoring system, ready to be practical implemented, while the ODS and MMI are still under development. Both these problems are addressed in the next two sections. New, bioinspired solutions for the ODS are presented in section 2; the problem of MMI is then addressed in section 3. Some conclusions and suggestions for further research are included in the section 4.

2 Bioinspired Obstacles Detection System

The most, known solutions for ODS are using ultrasonic transducers in order to detect obstacles in the 3D space. The low cost and easy implementation are the main advantages of this solution. The commonly used ultrasonic systems, which are based on pulse-echo method and can include up to 32 sensors, are to be replaced now by biomimetic systems [2], [3].

In order to improve the ODS performances other principles have to be investigated, among them visual sensors inspired from insects [4].

Insects are capable to navigate and avoid predators using a simple visual system. Systems imitating the characteristics of the biological model have been created, making evident the advantages of such systems. From the elementary motion detector inspired from the flies [5] to more complex systems that imitate the human eyes [6], the research in the field of bio-inspired artificial vision are in continuous development.

2.1 Obstacles detection system inspired from the locust LGMD neuron

We propose here a collision detection sensor inspired form the Lobula Giant Motion Detector (LGMD) found in locusts. LGMD is a large neuron found in the optical lobule of the locust [7] that mainly responds at the approaching objects [8]. Using electro-physical techniques, C. Rind formulated the functional structure and a mathematical description of the LGMD neuron [9]. This neuron is tuned to respond to objects on direct collision course and gives few or no response for receding objects [8]. The output of the neuron is a burst of spikes that increase in frequency with the approaching of the object.

The system proposed in this study is inspired by the computation that takes place in the locust visual neurons. One approach to develop such a system is to use the neural network model proposed by Rind et al. [9]. Four hierarchical levels of a neural network constitute the model, obtained after experimenting with locusts.

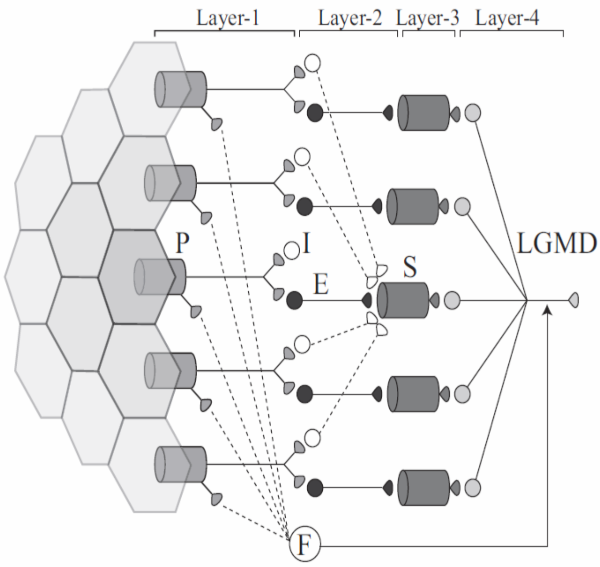

Fig. 1. The model of the LGMD neural network proposed by Rind et al. [9].

Fig. 1 presents the four retino-topical layers of the model. Level 1 represents the input of the neuronal circuit and includes “P” type units that respond at illumination changes in the scene, which is mainly induced by the motion of edges. In these units, a high-pass temporal filtering is taking place by subtracting the value of the illumination corresponding to a previous moment, from the current value of the illumination.

Then, the excitation is sent to the next hierarchical level, where “E” type units transmit excitatory signals to the next level, and “I” type units transmit delayed inhibitory signals to the nearest neighbors. In the “S” type units, the inhibitory signals are being subtracted from the excitatory signal corresponding of the same retino-topical level. If the difference exceeds a given threshold, the “S” type unit is excited and this signal is passed to the next level. Exceeding the threshold takes place only if the excitation surpasses the lateral inhibitory spread, for example near the moment of an imminent collision. The output of each “S” type units is summed in the LGMD neuron of the fourth level, and when the value of the sum exceeds a threshold value, the LGMD neuron outputs a spike. An imminent collision is signalized depending of the number of spikes emitted in a specific time interval.

In the “F” type unit, a feed forward inhibition is taking place, by summing the outputs of the “P” units. This is done to prevent the LGMD neuron to respond to the global excitation as in suddenly changes in background illumination.

Using the equations that describe the original LGMD neuron functionality [10], we developed a MATLAB version of the neuron, to simulate its behavior with different stimuli. Two types of cell have been used in our simulation: linear threshold cells and integrate and fire. In both cases, the membrane potential v(t+1), at moment t+1, is define by a discrete time equation:

$$v_i(t+1) = P_i \cdot v_i(t) + g_i^{Exc} \cdot \sum_{j=1}^{N_{Exc}} w_{ij} \cdot a_j(t-\delta_{ij}) - g_i^{Inh} \cdot \sum_{k=1}^{N_{Inh}} w_{ik} \cdot a_k(t-\delta_{ik}) \quad (1)$$

where $P_i(t) \in [0,1]$ is the persistence of the membrane potential, g_i^{Exc} and g_i^{Inh} are the gains of the excitatory and inhibitory inputs respectively, N_{Exc} is the number of excitatory inputs, N_{Inh} is the number of inhibitory inputs, w_{ij} and w_{ik} are the strengths of the synaptic connections between cells *i* and *j* and cells *i* and *k* respectively, δ_{ij} and δ_{ik} are the delays along the connections between cells *i* and *j* and cells *i* and *k* respectively.

Linear threshold cells are used to model graded potential neurons. The output activity of a linear cell *i* at time t+1 is given by:

$$a_i(t+1) = \begin{cases} v_i(t+1) & , v_i(t+1) \geq \theta \\ 0 & , otherwise \end{cases} \quad (2)$$

where θ is the membrane potential threshold.

Integrate and fire cells generate a discrete amplitude spike whenever the membrane potential exceeds the threshold, after which the membrane potential is hyperpolarized. The output activity of integrate and fire cell *i* at the time t+1is given by:

$$a_i(t+1) = \begin{cases} \beta & , \ v_i(t+1) \geq \theta \\ 0 & , \ otherwise \end{cases} \quad (3)$$

where β is the amplitude of the output spikes and θ is the threshold membrane potential.

After producing a spike, the membrane potential of cell *i* is hyperpolarized, such that $v_i(t+1) \rightarrow v_i(t+1) - \alpha$, where α is the amplitude of the hyperpolarization.

The number of inputs of the LGMD neuron was limited to 150x100; this corresponds to motion pictures with a resolution of 150x100 pixels. This resolution is considered high enough to detect obstacles in the real life scenes, and also needs a reasonable time to process the information.

By analyzing the number of consecutive spikes obtained at the output, we can determine if a collision is imminent or not. We determined experimentally three states in which the neuron can be found (Table 1): *safe, attention and danger*. Only the *danger* state signalizes an imminent collision. For this state, the necessary corrections of movement should be taking to avoid the collision. If the neuron is found in the other states, there is no danger of an imminent collision.

No. of consecutive spikes	1	2	3	4	5
State	Safe		Attention	Danger	

Table 1. Definition of the collisions states for the system.

A series of experiments have been performed to see how the algorithm responds to different stimuli. Considering the particular cases that a visually impaired person can encounter in his movement, the following situations were tested:

- The obstacle is approaching the subject on direct collision course;
- The obstacle appears from the side (left or right);
- The obstacle is approaching the subject, but the obstacle is not on the collision course;
- The obstacle is moving away from the subject.

For very simple scenes as stimuli, (a black box as an obstacle on a white background), the obtained results showed that the LGMD neuron can be successfully used to detect imminent collisions [11].

To examine the system response in a real-life situation, the environment shown in the images presented in Fig. 2 have been used. In the considered situation, a person is approaching on direct collision course to the system. The background for this particular case is not very complex, and only the person is moving, while the system is stationary.

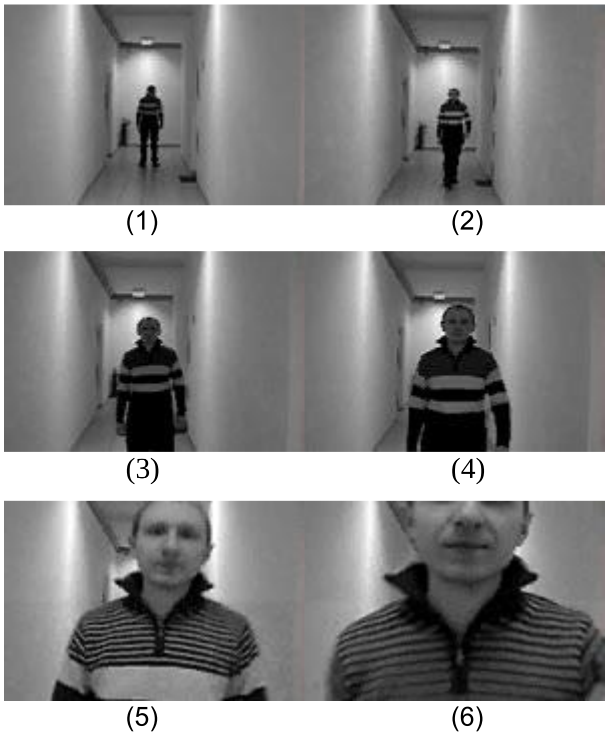

Fig. 2. Some of the real-life images applied to the input of the neuron.

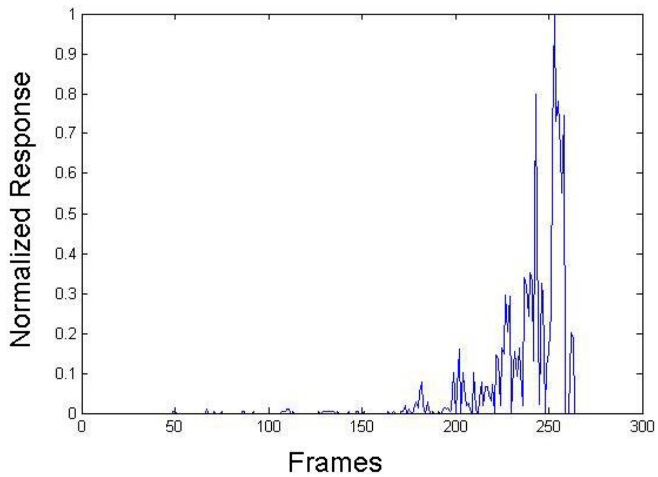

Fig. 3. System response to movie images from a real life situation, shown in Fig. 2.

Fig. 3 shows the system response for this particular case. Using the convention given in Table 1, the system detected an imminent collision at the frame 208, corresponding with the fourth image in Fig. 2.

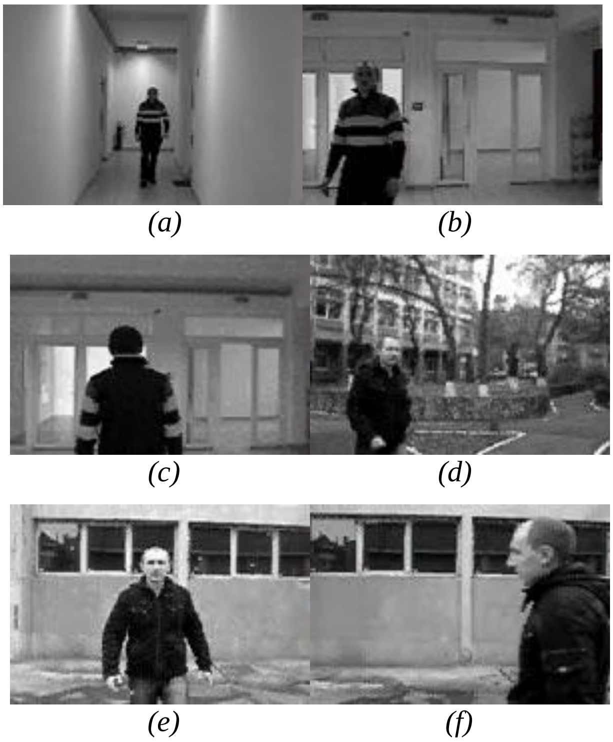

Fig. 4. Examples of test videos:
(a) - (c) were taken inside, (d) – (f) were taken outside.
All the videos have a resolution of 150x100 pixels.
(a) Obstacle is approaching.
(b) Obstacle is approaching from a non collision course.
(c) Obstacle is moving away from the subject.
(d) Obstacle is appearing from the left.
(e) Obstacle is on collision course.
(f) Obstacle is appearing from the right.

The environment in which blind people have to navigate is much more complex than the cases mentioned above. Real-life situations were developed to test the capability of the neuron to detect collisions. Movies containing images taken inside of the buildings and outside like in Fig. 4 were applied to the input of the neuron. We had success in detecting collisions in the most cases we tested, but in some cases (environments with very complex backgrounds, or moving objects in the background) we had false collision detections.

One solution, to improve the behavior of the system in real-life situations, is to filter the input images before to be processed by the LGMD neuron. This solution is under development and the already obtained results are very promising.

2.2 Motion detection system inspired from the fly

Insects generally use many sensors for their orientation in space, but their behavior is largely dominated by visual inspection.

They have a hierarchical organization of the visual system (as seen in Fig. 5) and by using visual feedback, they manage to do altitude control, flight stabilization, speed control, depth perception and collision avoidance.

Ability to detect objects in their vicinity is very important to avoid collisions and to avoid being caught by other predators.

The fly is the insect to which we will refer next. Even if the fly has a compound eye build with multiple lenses, the pattern projected on the retina, which is located under the lens is a single visual image of the scene. The signal received from the retina is transferred to the lamina, which contains cells that act like a high-pass filter. In the lamina layer, the luminance signal is normalized using a logarithmic compression and a contrast intensification is made.

The motion perception is made in the lobula plate layer, where are the so called VS and HS Tangential Cells, a wide-field horizontal and vertical neurons that respond to stimuli moving in a certain specific direction.

There are also two levels of priority, direction and altitude control, collision avoidance respectively, the later having priority.

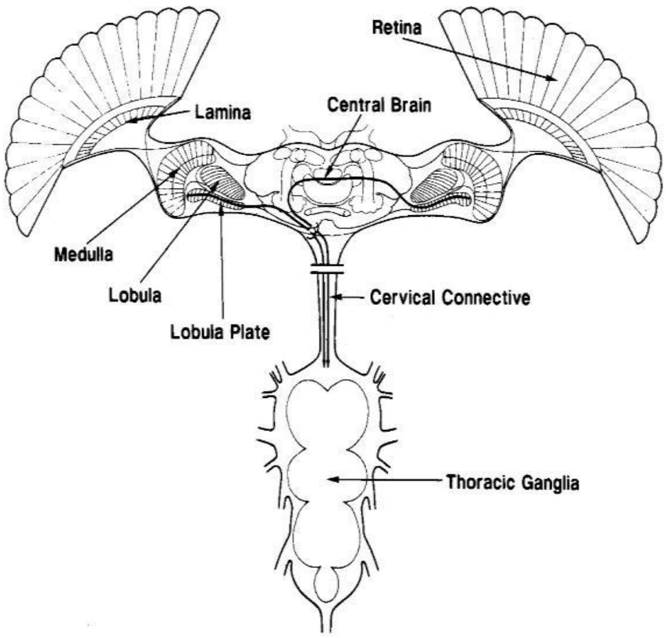

Fig. 5. Horizontal cross-section through the nervous system of the fly. Hierarchical organization (lamina, medulla, lobula, lobula plate) of the visual system of the fly.

2.2.1 The simplified Hassenstein-Reichardt correlator model

Hassenstein and Reichardt explained the mechanism of the insect vision and proposed an alternative to motion detection with an intensity-based spatiotemporal correlation algorithm [12], [13]. This type of algorithm is considered to be the fundamental part in all insect motion processing.

The Reichardt detector, also known as Hassenstein-Reichardt detector or Elementary Motion Detector (EMD) has the block diagram from Fig. 6.

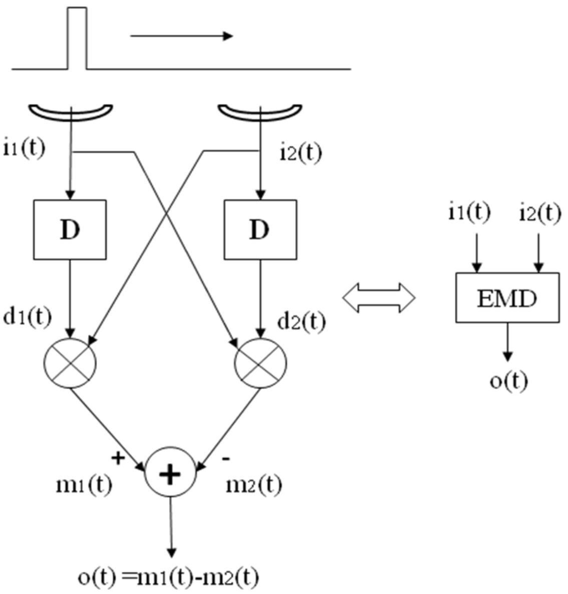

Fig. 6. The elementary motion detector (EMD) block diagram.

The block-diagram from Fig. 6 consists in two symmetrical sub-units and each sub-unit has a photo-receiver sensor, a delay line and a multiplication unit. The signals received from their neighboring entries are multiplied each other after one of them was delayed against the other. The combination of time delay and a multiplication is why this type of detector measures the degree of coincidence of signals in the input channels, making a space-time inter-correlation.

If an object passes through the detector, the channel that will activate first will be the channel that received first signal from the sensor and after a time will activate the other channel. This time depends on the speed of the object and spatial distance between the two input channels. Therefore, using the correlation between the two inputs, the EMD will give a strong response when a visual stimulus moves in a preferred direction and weak response when the stimulus moves in the opposite direction.

2.2.2 An elaborated motion detector based on Reichardt correlator model

Using as main unit the EMD block from Fig.6, it can be achieved a more complex motion detector which can distinguish between an object moving towards the front direction of an object moving horizontally.

The preprocessed image received from the image sensor is computed differently by two channels as shown in Fig.7.

The first channel divides the image received by the sensor into two symmetrical parts. The EMD cell computes the correlation between this parts pixel by pixel. It is necessary to sum all the pixels provided by the EMD's output to obtain response from the entire visual filed, which in this case is an excitation signal that contains the information about the sense of the motion in the horizontal line (motion from right to left or from left to right).

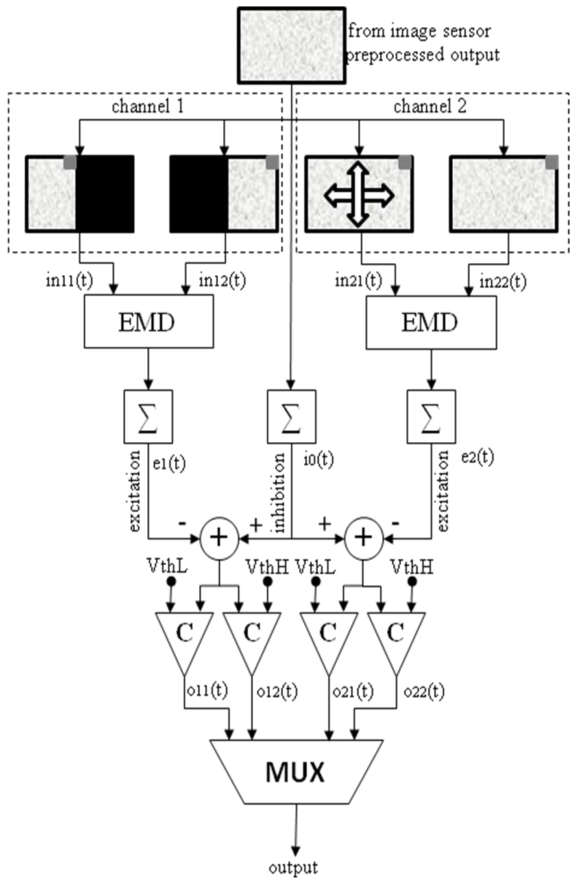

Fig. 7. The elaborated motion detector block diagram.

In the second channel, the image is divided also in two parts, but in this case, one part remains unchanged and the other part enlarge the image by both axes. The same correlation process accomplished by the EMD will give the information about an object that is approaching or it moves away in front of the sensor.

In both cases, it is necessary to have a normalized signal (or an inhibition signal) against which to compare the excitation signal that results from the EMD unit. This inhibition signal is obtained from the summation of the pixels that comes from the image sensor preprocessed output.

The output results of the comparators from Fig. 7, which represent the sense, and the direction of the moving object are gathered in one multiplexor to be available for more computing processes.

2.2.3 Simulation results of the elaborated motion detector

To highlight the proprieties of this motion detector, two situations were used for the simulation, one in which an object is moving in a horizontal line in front of the sensor and the other in which the object is moving toward and backward in front of the sensor.

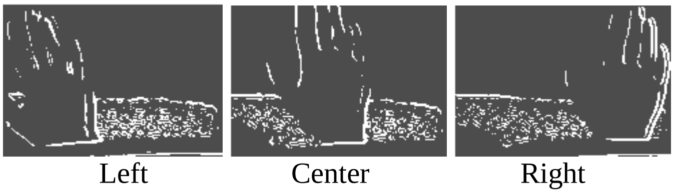

Fig. 8. Positions of the object in horizontal plane during the simulation.

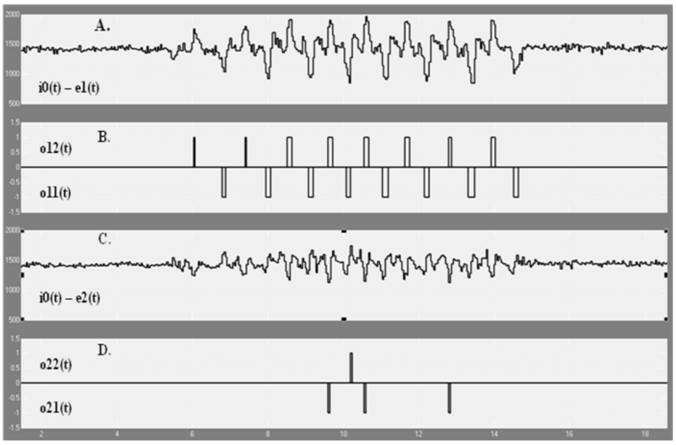

Fig. 9. Horizontal motion detection simulation results.

A. Channel-1 inhibition-excitation differential signal $i_0(t)$-$e_1(t)$
B. Channel-1 outputs : $o_{11}(t)$, $o_{12}(t)$
C. Channel-2 inhibition-excitation differential signal $i_0(t)$-$e_2(t)$
D. Channel-2 outputs : $o_{21}(t)$, $o_{22}(t)$

In the first situation created, as shown in Fig.8, the object (the hand) is moving in front of the sensor in horizontal line from left to right and from right to left with a rest time between successive movements.

When the inhibition-excitation differential signal $i_0(t)$-$e_1(t)$ will exceed a threshold, the output of the comparators from channel-1 will generate a pulse which is positive when is a right-left movement and a negative one when is a left-right movement.

Because of the noise and the correlation process, the inhibition-excitation differential signal $i_0(t)$-$e_2(t)$ from channel-2, will exceed for short periods of time the setup threshold, so will generate a few spikes.

Therefore, if an object is moving in the horizontal line in front of the detector, then channel-1 will generate a pulse train.

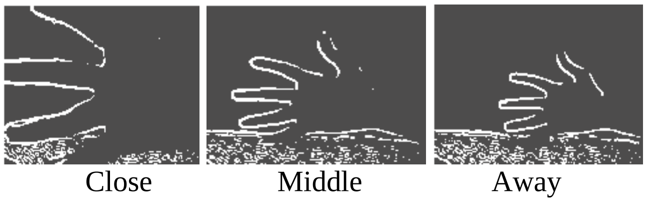

Fig. 10. Positions of the object in front of the sensor during the simulation.

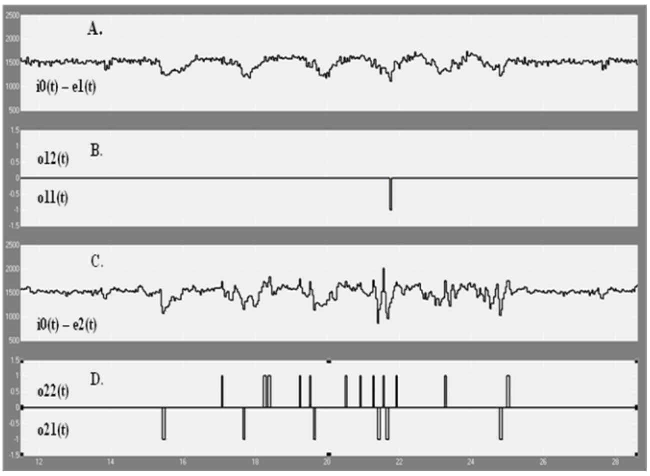

Fig. 11. Frontal motion detection simulation results.

A. Channel-1 inhibition-excitation differential signal $i_0(t)$-$e_1(t)$
B. Channel-1 outputs : $o_{11}(t)$, $o_{12}(t)$
C. Channel-2 inhibition-excitation differential signal $i_0(t)$-$e_2(t)$
D. Channel-2 outputs : $o_{21}(t)$, $o_{22}(t)$

In the same way, in the second case when the object is moving backward and forward in front of the sensor, channel-2 will generate a pulse train and the opposite channel will generate some spikes because of the correlation process.

When the hand is approaching to the sensor (Fig. 10), the inhibition-excitation differential signal $i_0(t)$-$e_2(t)$ (Fig. 11) will exceed an upper threshold so the comparator will generate a positive pulse and when will go under a lower threshold it will generate a negative pulse.

To put these proprieties in evidence and for the simplicity of the experiments we used for the simulation a controlled environment with an appropriate signal-to-noise ratio, knowing the fact that Reichardt correlators have some problems when they are used in a wide range of luminance levels and contrasts.

It is important to mention that for the simulation is used a low resolution sensor of only 160x120 pixels, because this aspect matters if we want a system to compute in real time.

3 AVR used as a MMI

The man-machine interface developed in the present research exploits the remarkable abilities of the human hearing system in identifying sound source positions in 3D space. The proposed solution relies on the AVR concept, which can be considered as a substitute for the lost sight of blind and visually impaired individuals.

According to the AVR concept, the presence of obstacles in the surrounding environment and the path to the target will be signalized to the subject by burst of sounds, whose virtual source position suggests the position of the real obstacles and the direction of movement, respectively.

Generation of sounds that suggest virtual sources whose positions can be placed in any point within 3D space it is not a simple task. When sound waves are propagated from a vibrating source to a listener, the pressure waveform is altered by diffraction caused by the torso, shoulders, head and pinnae. In engineering terms, these propagation effects can be expressed by two transfer functions, one for the left and another for the right ear, that specify the relation between the sound pressure of the source and the sound pressures at the left and right ear drums of the listener [14]. As a result, there is a pair of filters for every position of a sound source in the 3D space [15]. These, so-called Head Related Transfer Functions (HRTFs) are acoustic filters which not only vary both with frequency and with the heading, elevation and range to the source [16], but also vary significantly from person to person [17], [18].

Inter-subject variations may result in significant localization errors (front-back confusions, elevation errors), when one person hears the source through another person’s HRTFs [18]. Thus, individualized HRTFs are needed to obtain a faithful perception of spatial location.

If a monaural sound signal representing the source is passed through these filters and heard through headphones, the listener will hear a sound that seems to come from a particular location (direction) in space. Appropriate variation of the filter characteristics will cause the sound to appear to come from any desired spatial location [19], [20].

The practical implementation of the AVR concept encounters some difficulties due to the HRTF, which should be known for each individual and for a limited number of points in the 3D space.

These functions can be determined using a quite complex procedure, which requires many experimental measurements [14].

The proposed solutions in our research avoid these difficulties by generating the HRTF's values using Artificial Neural Networks (ANNs). Two ANNs are necessary in order to generate the HRTFs corresponding to each ear. The ANNs have been trained using the Listen HRTF Database, which is a public HRIR database, available for the whole scientific community. The development of this database took place in the frame of the Listen project. AKG and Ircam, the two partners of the project, have both performed the HRIR/HRTF and morphological measurement sessions. The database includes measurement data of 49 test subjects. For each subject there are 187 pairs of HRIRs (these HRIRs are Fourier pairs of the above mentioned HRTFs), each of these HRIRs corresponding to a particular point (particular azimuth-elevation pair) in the 3D space.

3.1 HRTF for one particular person

In our previously research [19], [20], an ANN capable to generate the values of the HRTF for a single individual and for every point in the 3D space has been developed. The proposed solution is presented here, for better understanding of the whole system, designed to guide the visually impaired persons.

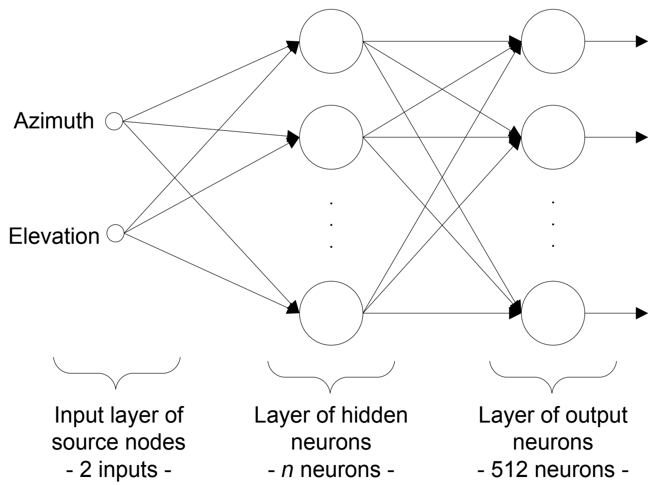

Fig. 12. The architecture of the proposed ANN.

The proposed ANN is a multilayer perceptron feed-forward backpropagation network. The network structure is shown in Fig. 7. The network consists of three parts: an input layer of source nodes (2 inputs), a hidden layer (50 neurons) and an output layer (512 neurons). There are two inputs necessary, which will be the azimuth and the elevation of the desired virtual sound source (network inputs). As a result, each of the two ANNs gives us a set of 512 values corresponding to the Head Related Impulse Responses (HRIRs) for the desired virtual sound source.

The Listen HRTF Database contains, as mentioned before, a total of 49 test subject. From all these subjects one subject was chosen based on the anthropometric similarities with our test person (ZH). The chosen subject from the Listen HRTF Database was the subject known as IRC_1031. The compared anthropometric measurements were presented in Table 2. The training of the ANNs was conducted using the data set corresponding to the selected subject.

Extent	Measurement	IRC_1031		Subject	
		L	R	L	R
x_1	Head width	153		162	
x_3	Head depth	216		225	
x_{12}	Shoulder width	530		550	
d_1	Cavum concha height	20	20	18	19.3
d_3	Cavum concha width	17	17	15.4	15.4
d_4	Fossa height	22	23	21.2	16.7
d_5	Pinna height	62	66	66.5	64.7
d_6	Pinna width	28	31	26.9	28.8

Table 2. Selected anthropometric measurements in [24]. Specified values are expressed in millimeters.

The conducted experimental measurements, presented in [20], will be shortly described as follows. Their purpose is to evaluate the performances of the proposed network structure through localization experiments. The experimental setup, presented in the following, proposes to determine if the practically obtained sound really offers the appropriate perception of the desired virtual sound source in the auditory space of an arbitrary subject. In other words, the paper intends to establish if the obtained sound, acquired by filtering a monaural sound with the obtained HRTFs, really seems to come from, or at least near, the azimuth-elevation couple specified at the inputs of the ANNs.

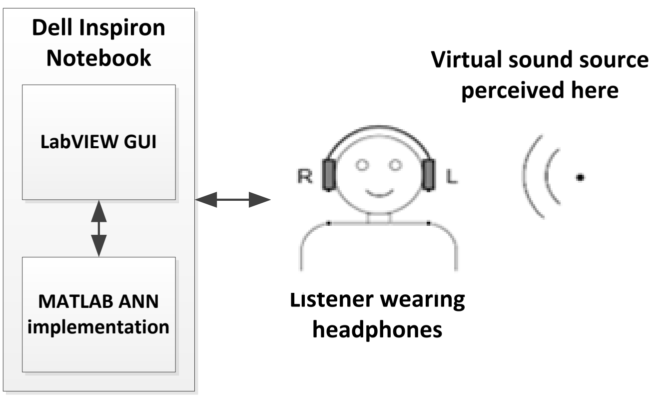

Fig. 13. The experimental setup.

The experimental setup used in our experiments is presented in Fig. 13. A short description of the setup follows. Our experiments were conducted on a Dell Inspiron 1520 notebook running Windows 7 Ultimate 32-bit operating system with the following hardware configuration: Intel Core 2 Duo 2.00 GHz CPU, 2GB DDR2 RAM, 160 GB HDD. Sennheiser HD 435 headphones were connected to the headphone output of the notebook and used for headphone listening of the generated acoustic signals. As software environment, we used an improved version of the graphical user interface (GUI), presented in [25]. The GUI was developed using National Instruments' LabVIEW, which is a graphical programming environment widely used by engineers and scientists.

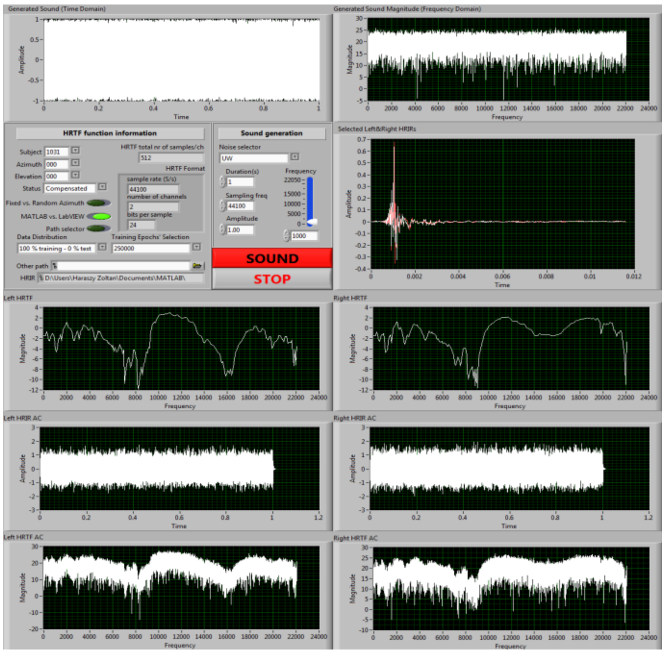

Fig. 14. A snapshot of the running GUI [25].

In the following the GUI will be presented briefly. Fig. 14 shows a snapshot of the running GUI. The basic idea was to implement a basic version of the AVR concept. One can obtain a simple AVR, which means obtaining the left and right acoustic signals intended for headphone listening, whose virtual sound source is a certain point (specified via an azimuth-elevation pair) in the 3D space. The acoustic signal for one of the ears is obtained by convolving a locally generated sound (for example: sine or white noise) with the corresponding HRIR values for the desired point in space.

The GUI also offers the possibility to choose between different neural network states after training the ANNs for a specified number of epochs (50000, 100000, 150000, 200000, 250000 or 300000) and different training/test data distributions (100% training – 0% test or 75% training – 25% test). The possibility to choose between states is possible because the proposed ANNs were trained prior to the localization experiments and their states were saved in .mat format on the hard disc.

In the following we will focus only on the proposed ANN based method. Our method was implemented using the Neural Network Toolbox, included in MATLAB, which is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation. Neural Network Toolbox extends MATLAB with tools for designing, implementing, visualizing, and simulating neural networks.

Our localization experiments were conducted in our Bioinspired Systems' Laboratory from the Department of Applied Electronics, in Politehnica University of Timisoara. The test subject (ZH) was placed on a rotating chair, with a field compass placed horizontally between its legs for simple angle measurement. A loudspeaker was placed straight ahead of the test subject at the same height with the subject's head. Expressed in azimuth-elevation coordinates the loudspeaker position is at (0˚;0˚). The distance between the loudspeaker and the subject's head was fixed at 1,5 meters. The loudspeaker served as a reference sound source in this localization experiment. The reference signal coming from the loudspeaker was a burst of uniform white noise of 1 s with 0,2 s silence and was repeated throughout the localization experiment.

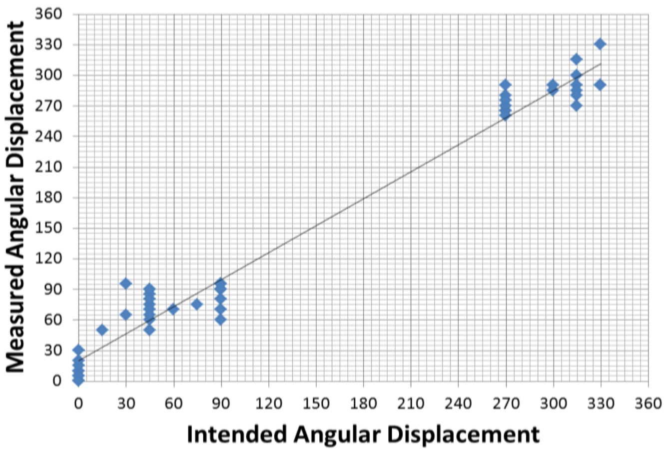

Fig. 15. Obtained localization results.

Acoustic signals were played through the headphones to the subject. The stimulus signals were generated using the GUI from LabVIEW. The subject was told to rotate the chair without rotating its head and modifying the relative height from the ground until the virtual sound source of the acoustic signals (stimulus signals) coincides with the physically present sound source position (reference signal). When the two sources coincide, the angular displacement (azimuth) is read from the field compass. The chosen acoustic signal was a burst of uniform white noise of 1 s with 0,3 s silence and was played repeatedly until the test subject gave an estimation of the virtual sound source azimuth in its auditory space by remaining in the same position with the rotating chair. In the next phase, the virtual sound source position and the actually used azimuth-elevation pair is compared (by calculating the absolute difference between the used and measured azimuth values) to verify the accuracy of the sound localization, this way the performance of the whole system.

The obtained results of the conducted localization experiments are represented in Fig. 15. In frame of this experiment, the stimulus signals were only generated in the frontal plane and at 0˚ elevation. For azimuth, this means angles between – 90˚ and 90˚. The authors would also like to underline that in the conducted localization experiments the following ANN states were used:

- Data distribution – 75% training and 25% test;
- Numbers of training epochs used – 300000.

Table 3 gives us an overview of the obtained absolute error. The first row represents the considered absolute azimuth error intervals. The second row shows the number of measured azimuth values, which fall into the corresponding error interval. Finally, the third row gives us the percentage of one error interval with respect to all of the measured azimuth values.

Absolute azimuth error intervals	**Number of judged azimuth values from the specified azimuth error interval**	**Share of the given error interval**
[0,15)	32	59,26 %
[15,30)	10	18,52 %
[30,45)	11	20,37 %
[45,60)	0	0 %
[60,∞)	1	1,85 %

Table 3. Overview of the obtained azimuth errors.

After comparing the results from Table 3 with the ones from H. Hu et al. [24], we can state that the localization performances presented in the current paper are better than in [24], if we compare them with the results obtained using non-individual HRTFs. But, when we consider the results obtained using individual HRTFs, the localization using our method is less accurate.

Results show that these ANNs have an appropriate behavior and can be used in practical applications.

3.2 HRTFs for any person

In our recent research [21] a more complex and more performance ANN have been developed. This ANN is capable to generate the HRTF values for any subject and for every point in the 3D space. It requires to its inputs the following data:

- Certain anthropometric measurements that define a certain subject, listed in Table 4,
- The azimuth and the elevation, which define the position of a certain point in the 3D space.

As a result, the ANNs outputs will provide the values of the corresponding HRTF.

The multi-subject – multi-point ANN is under development but the already obtained results are very promising [21]. A short presentation of our recent research follows.

Variable	Measurement
x_1	Head width
x_2	Head height
x_3	Head depth
x_4	Pinna offset down
x_5	Pinna offset back
x_6	Neck width
x_7	Neck height
x_8	Neck depth
x_9	Torso top width
x_{10}	Torso top height
x_{11}	Torso top depth
x_{12}	Shoulder width
$\cancel{x_{13}}$	~~Head offset forward~~
$\cancel{x_{14}}$	~~Height~~
$\cancel{x_{15}}$	~~Seated height~~
x_{16}	Head circumference
x_{17}	Shoulder circumference
d_1	Cavum concha height
d_2	Cymba concha height
d_3	Cavum concha width
d_4	Fossa height
d_5	Pinna height
d_6	Pinna width
d_7	Integral incisure width
$\cancel{d_8}$	~~Cavum concha depth~~
$\cancel{\theta_1}$	~~Pinna rotation angle~~
$\cancel{\theta_2}$	~~Pinna flare angle~~

Table 4. Anthropometric measurements available for every subject used in training or testing phase of the proposed ANNs (data written in strikethrough format were not available for every subject).

The structure of the proposed multi-subject – multi-point ANN is presented in Fig. 16. The network includes three parts: an input layer of source nodes (23 inputs), a hidden layer (50 neurons) and an output layer (512 neurons). The necessary input data set consists of the available 21 anthropometric measurements (applied to the inputs in the same order as enumerated in Table 4) and the position (the azimuth and the elevation) in the 3D space of the desired virtual sound source. As a result, each of the two ANNs gives us a set of 512 values representing the Head Related Impulse Responses (HRIRs) for the desired virtual sound. These HRIRs are Fourier pairs of the above mentioned HRTFs.

The proposed ANN is a multilayer perceptron feed-forward backpropagation network and was implemented using MATLAB's Neural Network Toolbox.

After careful consideration, the selected public database was divided in two smaller databases, as follows. The existing data sets for subjects labeled as IRC_10XX, where XX takes values between 08 and 49, totaling a number of 35 subjects, were selected for the training phase of the ANNs. The remaining data sets, for subjects labeled as IRC_10XX, where XX takes values between 50 and 59, were used as test data for the same networks (subject IRC_10XX, where XX takes values between 02 and 07, were not used because of some missing anthropometric measurement values). Using this approach, we were able to evaluate the performances of the proposed ANNs using the previously obtained test data. In the conducted experiments, pre-processing of the input data or post-processing of the output data was not used.

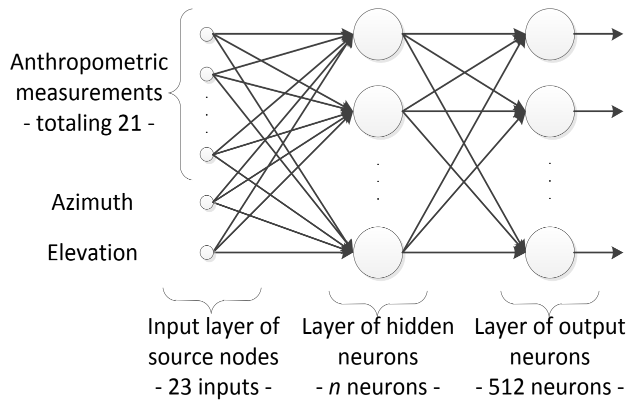

Fig. 16. The architecture of the proposed ANN solution.

The training process was carried out until the accepted error margin was reached. The error margin was considered to be enough for our current experiment in the range of 3-5%. This error was obtained for the test data set.

After the training phase, the next goal was to evaluate the performances of the two networks. As a performance criterion, we have chosen the mean squared error (MSE), which was calculated between the resulted values on each network output and the corresponding test data values. To give a measure to the extent of this error, we compared the resulting MSE to the absolute maximum value of the network outputs. Depending on the initial bias and weight values there are more or less training epochs necessary to obtain the same values of the error. The maximum number of epochs used for training was 150, but usually a number of 100 epochs are more than sufficient.

We would like to present shortly some results obtained for 10 neural networks, which all have the same, proposed structure. The only differences are the initial weight and bias values before the training of these neural networks.

Fig. 17 presents the resulting MSE for 10 ANNs corresponding to the right ear of one particular subject. On the horizontal axis, we represented in each of the 10 columns the obtained errors for one of the 10 evaluated right side ANNs. The training process for each of these ANNs was stopped after 10, 20, 30, 50 and 100 epochs and the corresponding MSE was calculated using the test data set. The obtained MSE for different number of epochs (evaluation phase) is represented by different kinds of signs (point sign – 10 epochs, plus sign – 20 epochs, multi-cross sign – 30 epochs, circle sign – 50 epochs, single-cross sign – 100 epochs). On the vertical axis, the obtained MSE was represented for each evaluation phase.

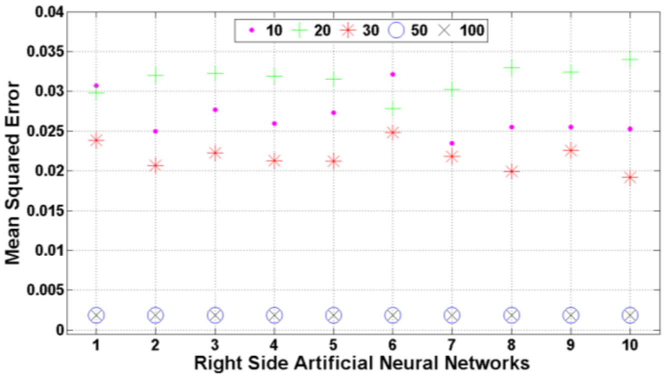

Fig. 17. Obtained performances after the testing procedure of 10 ANNs with the proposed structure.

It can be observed that the MSE for the investigated ANNs vary due to the different initial weight and bias values.

One can observe that the MSE calculated between the resulted output values on each network and the corresponding test data values decreases as the number of training epochs increases. The obtained values confirm our previous statement that a number of 100 epochs are sufficient for the training of the ANNs, if the accepted error margin is 3-5%.

4 Conclusion

Many efforts have been invested in the last years, based on ingenious devices and information technology, in order to develop ETA equipments as a substitute for the lost sight of blind and visually impaired individuals. As a result, we can conclude now that the AVR based MMI is an appropriate solution for ETA equipments. However, research efforts are still necessary in order to optimize the procedure of HRTF generation using ANN. The AVR concept can be then software implemented on a microcontroller system.

In a recent research, our team developed ANNs capable to generate values of the HRTFs for every point in the 3D space [19], [20] and for more than one subject [21]. The proposed method will speed up the implementation of the AVR concept after the ANN training has been completed.

In spite of these results [26], there are still some open questions, which have to be investigated in order to find appropriate solutions for them.

The ODS, usually equipped with ultrasonic transducers, have the advantage of simplicity and low cost, but in the same time these types of detectors have a limited resolution and, in certain situations, encounters errors do to reflections. The laser-based 3D sensors are a valuable alternative solution, but they are still in the development stage [22], [23].

More promising seems to be the bio-inspired solutions, proposed in the present research. Inspired from the ODS model of flies or locusts, our research team is developing, with promising results, a collision detection sensor with applicability in the field of ETA systems.

Acknowledgements

This work was partially supported by the strategic grant POSDRU/88/1.5/S/50783, Project ID50783 (2009), co-financed by the European Social Fund – Investing in People, within the Sectoral Operational Programme Human Resources Development 2007-2013.

References:

[1] V. Tiponut, A. Gacsadi, L. Tepelea, C. Lar, I. Gavrilut, Integrated Environment for Assisted Movement of Visually Impaired, *Proceedings of the 15th International Workshop on Robotics in Alpe-Adria-Danube Region, (RAAD 2006),* ISBN: 9637154 48 5, Balatonfured, Hungary, 2006, pp. 234-239.

[2] J. Reijniers, H. Peremans, Biometric Sonar System Performing Spectrum-based Localization, *IEEE Trans. on Robotics*, vol. 23, no. 6, 2007, pp. 1151-1159.

[3] R. Z. Shi, T. K. Horiuchi, A Neuromorphic VLSI Model of Bat Interaural Level Difference Processing for Azimuthal Echolocation, *IEEE Trans. Circuits and Systems*, vol. 54, 2007, pp. 74-88.

[4] N. Franceschini, J. M. Pichon, C. Blanes, From Insect Vision to Robot Vision, *Philosophical Transactions: Biological Sciences*, Volume 337, Issue 1281, 1992, pp. 283-294.

[5] R. R. Harrison, C. Koch, A silicon implementation of the fly‟s optomotor control system, *Neural Comput.*, vol. 12, 2000, pp. 2291–2304.

[6] Zaghloul K. A., Boahen K., A silicon retina that reproduces signals in the optic nerve, *Journal of Neural Engineering*, 3/2006, pp. 257-267.

[7] M. O'Shea, J. L. D. Williams, The anatomy and output connection of a locust visual interneurone; the lobular giant movement detector (LGMD) neurone" *J. Comparative Physiology*, vol. 91, 1974, pp. 257-266.

[8] F. C. Rind and P. J. Simmons, "Orthopteran DCMD neuron: a reevaluation of responses to moving objects. I. Selective responses to approaching objects," *J. Neurophys.*, vol. 68, 1992, pp. 1654–1666.

[9] F. C. Rind, D. I. Bramwell, Neural network based on the input organization of an identified neuron signaling impending collision, *Journal of Neurophysiology*, Vol 75, Issue 3, 1996, pp. 967-985.

[10] G. Linan-Cembrano, L. Carranza, C. Rind, A. Zarandy, M. Soininen, A. Rodriguez-Vazquez, Insect-vision inspired collision warning vision processor for automobiles, *Circuits and Systems Magazine, IEEE*, Volume 8, Issue 2, 2008, pp. 6-24.

[11] D. Ianchis, Z. Haraszy, V. Tiponut. Collision Detection Inspired by Locust Neural System, *Sesiunea de Comunicari Stiintifice Doctor Etc 2009*, September 24-25, Timisoara, 2009, pp 45-48.

[12] Sergi Bermudez i Badia, Pawel Pyk, Paul F.M.J. Verschure, A fly-locust based neuronal control system applied to an unmanned aerial vehicle: the invertebrate neuronal principles for course stabilization, altitude control and collision avoidance, *International Journal of Robotics Research*, vol. 26, 2007, pp. 759-772.

[13] B. Hassenstein, W. Reichardt, Structure of a mechanism of perception of optical movement, *Proceedings of the 1st International Conference on Cybernetics*, 1956, pp. 797–801.

[14] R. O. Duda, Modeling head related transfer functions, in *Proc. 27th Ann. Asilomar Conf. Signals, Systems and Computers*, 1993.

[15] J. Blauert, *Spatial Hearing*, MIT Press, 1997.

[16] C. P. Brown, R. O. Duda, A Structural Model for Binaural Sound Synthesis, *IEEE Transactions on Speech and Audio Processing*, Vol. 6, No. 5, 1998, pp. 476 – 488.

[17] D. J. Kistler, F. L. Wightman, A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction, *Journal of the Acoustical Society of America*, Vol. 91, 1992, pp. 1637–1647.

[18] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, Localization using nonindividualized head-related transfer functions, *Journal of the Acoustical Society of America*, Vol. 94, 1993, pp. 111–123.

[19] Z. Haraszy, D. Ianchis, V. Tiponut, Generation of the Head related Transfer Functions Using Artificial Neural Networks, *Proceedings of the 13th WSEAS International Conference on CIRCUITS*, ISBN: 978-960-474-096-3, ISSN: 1790-5117, 2009, pp. 114-118.

[20] Z. Haraszy, D. G. Cristea, V. Tiponut, T. Slavici, Improved Head Related Transfer Function Generation and Testing for Acoustic Virtual Reality Development, *Proceedings of the 14th WSEAS International Conference on SYSTEMS*, ISBN: 978-960-474-199-1, ISSN: 1792-4235, 2010, pp. 411-416.

[21] Z. Haraszy, S. Micut, V. Tiponut, T. Slavici, Multi-Subject Head Related Transfer Function Generation using Artificial Neural Networks *Proceedings of the 14th WSEAS International Conference on SYSTEMS*, ISBN: 978-960-474-199-1, ISSN: 1792-4235, 2010, pp. 399-404.

[22] D. Stoppa, L. Pancheri, M. Schandiuzzo, A CMOS 3-D Imager Based on Single Photon Avalanche Diode, *IEEE Trans. Circuits and Systems*, vol. 54, 2007, pp. 4-12.

[23] * * * www.canesta.com.

[24] H. Hu, L. Zhou, H. Ma, Z. Wu, HRTF personalization based on artificial neural network in individual virtual auditory space, *Applied Acoustics*, 2008; 69(2), pp.163–172.

[25] Z. Haraszy, D. Ianchis, V. Tiponut, Acoustic Virtual Reality Generation with Head Related Transfer Function Visualization, *Proceedings of Scientific Communication Session "Doctor ETC 2009"*, ISSN: 2066-883X, 2009, pp. 40-44.

[26] V. Tiponut, D. Ianchis, M. Basch, Z. Haraszy, Work Directions and New Results in Electronic Travel Aids for Blind and Visually Impaired People, *Proceedings of the 14th WSEAS International Conference on SYSTEMS*, ISBN: 978-960-474-199-1, ISSN: 1792-4235, 2010, pp. 347-352.