BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **15**, No 7
Special Issue on Information Fusion

Sofia • 2015 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.1515/cait-2015-0088

3D Visualization of Sound Fields Perceived by an Acoustic Camera

Nevena Popova, Georgi Shishkov, Petia Koprinkova-Hristova, Kiril Alexiev

Institute of Information and Communication Technologies, BAS, 1113 Sofia, Bulgaria
Emails: veni_93@abv.bg boxich@gmail.com pkoprinkova@bas.bg alexiev@bas.bg

Abstract: *The paper summarizes the application results of a recently proposed neuro-fuzzy algorithm for multi-dimensional data clustering to 3-Dimensional* (*3D*) *visualization of dynamically perceived sound waves recorded by an acoustic camera. The main focus in the present work is on the developed signal processing algorithm adapted to the specificity of multidimensional data set recorded by the acoustic camera, as well as on the created software package for real-time visualization of the "observed" sound waves propagation.*

Keywords: *Acoustic camera, multi-dimensional data, feature extraction, direction selective cells* (*MT neurons*), *Echo state network, fuzzy clustering.*

1. Introduction

An acoustic camera is an imaging device used to locate sound sources and to characterize them [1]. It consists of a group of microphones (microphone array) and optionally may be accompanied by an optical camera. Since the sound propagates through different media with known speed, each sound source in the "observed" area is perceived by each microphone in the array at a different time instant and with different sound intensity in dependence on both the sound source location and the particular microphone location. There are two basic approaches [4, 13] for localization of the sound sources – beamforming for far distances and acoustic holography for near distances. However, in both of the cases the signal processing techniques used needs a lot of memory and powerful hardware in order to cope with

the computational complexity. That is why it is usually performed off-line after accumulation of data thus allowing detecting only static sound sources while the development of on-line implementations increases dramatically the price and size of the acoustic cameras. This motivated the need to work towards less complex memory-consuming and faster implementations.

The approach presented here is an attempt towards simplification of acoustic camera signal processing via application of intelligent techniques based on the following motivations:

- the task of sound source localization consists of separation of the "observed" area into sub-areas in dependence on the measured characteristics of the sound coming from them, i.e., it could be considered as a clustering task;
- since the final outcome of signal processing is a "picture" of the observed sounds, a parallel between the sound field perception and visual information processing by human brain can be implied.

Following the above motives, a new approach for sound sources localization was proposed in [6, 7]. Since the acoustic camera is composed of multiple microphones, the data perceived by it is multidimensional. So the algorithm for multi-dimensional data clustering [8] combining neural networks and fuzzy logic was adopted. The need to consider the perceived sensory data as a "sound picture" that has to be mapped onto observed by optical camera scene led us to the idea to upgrade the algorithm from [8] with a layer for pre-processing of the raw signals mimicking the perception of visual information in the human brain. In [6] the approach was tested first to create two static dimensional pictures from accumulated sensory data of acoustic camera. Next in [7] the approach was extended to dynamic visualization of the perceived by the acoustic camera sound field, thus creating a three dimensional picture of the sound waves propagation through time. Next in [10, 12] investigations of the proposed algorithm aimed at its refinement and tuning of its key parameters continued. Experiments with two sound sources, as well as with a moving sound source were carried out too.

The present paper summarizes the proposed algorithm and results from different experiments carried out focusing on the created software package for real-time visualization of propagation of the "observed" sound waves by the camera. It is organized as follows: Section 2 describes briefly the basic steps of the developed hierarchical structure for signal processing, the features extraction and clustering of multidimensional data recorded by the acoustic camera; Section 3 describes the corresponding software modules implementing the algorithm; Section 4 demonstrates software package application results; the paper concludes with discussion and directions for future work.

2. Algorithm description

The overall algorithm developed through [6, 7, 10, 12] has hierarchical structure shown in Fig. 1. Since all the experiments were carried out with Brüel & Kjær acoustic camera, the description of the raw data pre-processing is based on its specific characteristics (e.g., the microphones positions and their correspondence to

the observed by the acoustic camera “screen”) as described in our previous works [6, 7, 10, 12]. The acoustic camera we have consists of 18 microphones array placed randomly in a wheel grid with mounted optical camera at its centre. The applications to different shapes of the microphone arrays can be done in a similar way and following the basic principles of the algorithm are described here.

The basic steps of our algorithm are as follows:

- raw data signal pre-processing and extraction of initial features;
- mapping of the initial features onto a bigger feature space;
- creation of two-dimensional projections from the extracted multiple features and choice of proper projection(s);
- clustering of the chosen projections and obtaining of a “sound picture” for each time instant of the period of observations;
- gathering of all sound pictures to produce a 3D picture of sound waves propagation.

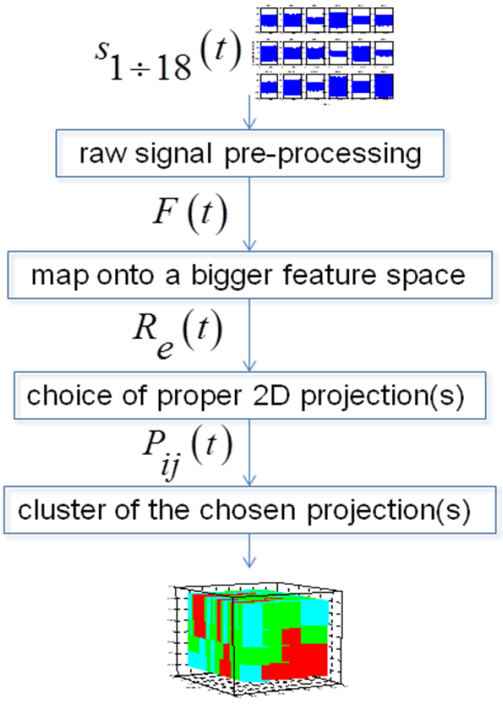

Fig. 1. Hierarchical structure of the proposed algorithm

The blocks composing the hierarchical structure of our algorithm are described further.

2.1. Initial features extraction

The role of this step is to mimic the perception of visual information in human brain, and particularly the way humans discriminate among complex visual motion

patterns. It exploits a part of the model of human visual perception proposed in [2] – the direction selective cells in the middle temporal cortex called MT neurons. For this purpose we consider the plane of acoustic camera microphones as a "screen" observed by the camera. The signal perceived by each sensor (microphone) is considered as a "stimulus". The screen is divided into sub-areas called "receptive fields", each one receiving at least one stimulus. In [6, 7] it was decided to divide the observed area into 16 overlapping square regions, each receiving at least one stimulus as shown in Fig. 2. MT neurons serve as filters that discriminate the stimuli in each receptive field (that is the acoustic pressure measured by the microphones there) by their average amplitude. For this purpose, the working range of the stimuli was divided into several overlapping intervals. For each of these intervals a corresponding MT filter was defined as follows:

$$\mathrm{MT}_i\left(s_k\left(t\right)\right) = \exp\left(\frac{-\left(\mu_i - s_k\left(t\right)\right)^2}{2\sigma^2}\right), \tag{1}$$

where μ_i is the centre of the i-th MT filter; and σ is the variance (common parameter for all filters); $s_k(t)$ is a stimulus (the acoustic pressure measured by the k-th microphone) at time step t. Then the average received stimulus in the l-th receptive field is

$$f_{il}\left(t\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{MT}_i'\left(s_k\left(t\right)\right), \tag{2}$$

where n is the number of stimuli in the l-th receptive field.

The key parameter that has to be chosen is the number of MT filters n_f. Then the parameters of each MT filter were defined as follows: the dynamic range of the stimuli (the acoustic pressure data) was divided into n_f intervals with equal width; their centres define the parameters μ_i; the variance σ was chosen to be equal to one third of the interval width so that the filters overlap [6, 7]. Thus the matrix $F\left(t\right) = \left[f_{il}\left(t\right)\right]$, i=1, …, n_f, l=1, …, 16, contains initial features extracted from the instantly observed "sound picture".

2.2. Mapping onto bigger feature space

The idea of mapping from the initial features space to a bigger features space using a special kind of recurrent neural network – Echo State Network (ESN) [5, 9] – was proposed for the first time in [8].

The basic structure of ESN is shown in Fig. 3. It consists of a randomly generated dynamic reservoir of interconnected neurons. The external input u is fed into all neurons in the reservoir. Each neuron has feedback connection from itself as well as from other randomly chosen neurons. The non-linear transformation of the weighted input (a sigmoid, preferably hyperbolic tangent) produces the state of the reservoir r.

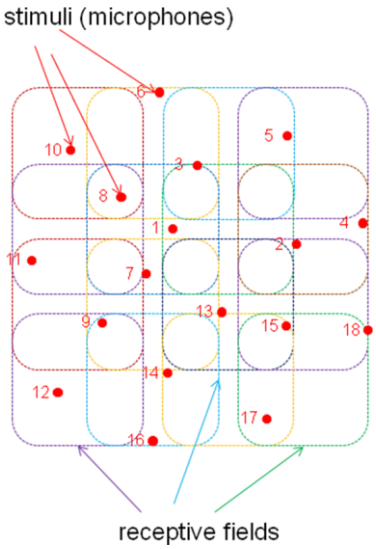

Fig. 2. Observed by acoustic camera area ("screen") and its sub-areas ("receptive fields"). The numbered dots denote the microphones positions and the receptive fields are surrounded by different colour squares

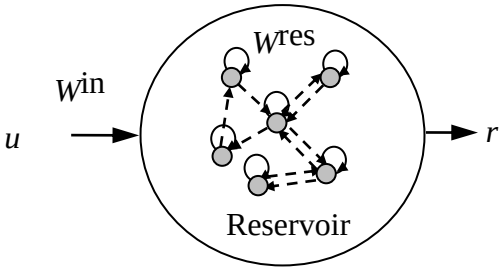

Fig. 3. Echo state network: *W** denotes the vector (*=*in*) and matrix (*=*res*) of the corresponding input and reservoir connection weights

The core of feature extraction is explained in details in [8]. Briefly, the parameters of the ESN reservoir (vectors *a* and *b* in Equation (3) below) are tuned to reflect the input data structure using the algorithm from [11] and then mapping onto the state of equilibrium states of reservoir neurons is as follows:

$$r_e = \tanh\left(\mathrm{diag}(a)W^{res}r_e + \mathrm{diag}(a)W^{in}u + b\right), \quad u = \mathrm{const}. \tag{3}$$

for a given set of constant inputs calculated iteratively. The set of inputs containing initial features obtained for each receptive field l=1, …, 16 is $u_l = F_l(t) = \left[f_{il}(t) \right]$. Thus the extracted features form the matrix $R_e(t) = \left[r_{el}^i \left(u_l(t) \right) \right]$, $i = 1,...,n_r$, $l = 1,...,16$, contain equilibrium states for all n_r neurons in the ESN for all receptive fields and all time instants *t* of the period of observations.

2.3. Choice of 2D projection(s)

All possible 2D projections contain all possible combinations between the features (equilibrium states of ESN) extracted at the previous stage:

$$P_{ij}(t) = \left[r_{el}^{i}(t), \quad r_{el}^{j}(t) \right], \quad l = 1,...,16. \tag{4}$$

Initially in [6, 7] the projections revealing the maximum number of clusters were selected. For this aim the subtractive clustering was applied to all possible 2D projections. However, since the subtractive clustering was the "bottleneck" of the algorithm, while increasing the reservoir sizes of the number of projections, the computational burden increased dramatically. Thus we returned to the originally proposed in [8] idea to choose the projection(s) between neurons having the biggest number of maxima of the probability density distribution of their equilibrium states. Although even in this case there is a chance to obtain more than one projection, it appeared to be the best option among the other tested approaches [10, 12].

2.4. Clustering of chosen projection(s) and 3D visualization

For clustering purposes two fuzzy algorithms were tested – subtractive clustering [14] and fuzzy C-means clustering [3]. The first is optimization procedure that reveals a proper number of clusters. The second separates the data into pre-determined number of clusters and is faster. During preliminary experiments with test sound sources with constant frequency [6, 7, 10, 12], it appeared that 3-4 clusters were enough to obtain a clear "sound picture". Hence, for creation of 3D visualization here we will present the results with only three clusters for simplicity.

The 3D picture of sound waves propagation is obtained by gathering "slices" of 2D pictures. These 2D pictures visualize each receptive field on the observed "screen" with different colour in dependence on the cluster it belongs. The correspondence between the clusters and receptive fields was done by calculating the minimum distance to the clusters centres determined by a fussy C-means procedure.

3. Software package

The overall description of the created software package in Matlab environment is given in Fig. 4. The first three modules are intended for off-line adjusting of the individual elements in the "sound picture" visualization system, i.e., tuning of ESN to the data perceived by the acoustic camera, choice of 2D projection and calculation of the cluster centres. For this purpose data accumulated for some period of measurements (called further training data set) were used. The last module is intended to work in real time, receiving raw signals from acoustic camera at each time instant, extracting their features and creation of 2D slice by mapping each receptive field to a cluster using already tuned at off-line stage elements (ESN, chosen projection and cluster centres). Then 2D slice is added to a 3D visualization "cube".

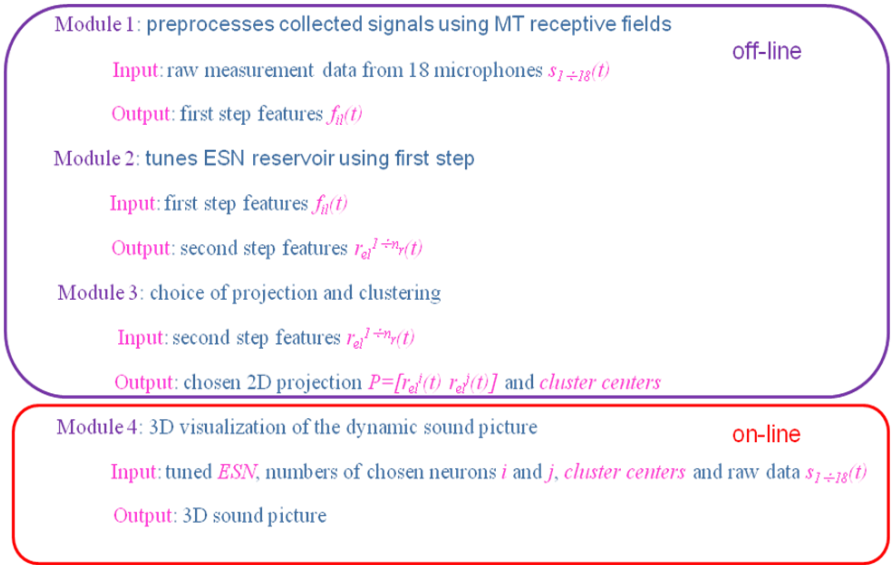

Fig. 4. Software package modules

The functions of the basic modules from Fig. 4 are as follows:

- Module 1 calculates the initial features matrix for accumulated training data set by a number of filters set by the user using Equations (1) and (2). If the results obtained at the end of the off-line procedure are not satisfactory, the user can return to repeat calculations with different number of MT filters.
- Module 2 uses ESN toolbox [5] with additional IP training function [11]. It generates a random ESN reservoir with a chosen number of neurons and tunes its parameters (the vectors *a* and *b* from Equation (3)), using the features matrix generated by the first module as input data. Next the module calculates the equilibrium states of reservoir neurons (as in Equation (3)) that are the features extracted at the second step of the algorithm. The number of the reservoir neurons can be defined by the user. If at the next step (performed by Module 3) the results are non-satisfactory, the user can return to this step or can choose also to return to module 1 changing also the number of MT filters.
- Module 3 calculates the density distributions of neurons equilibrium states and determines those with the biggest maximal number. Then it performs clustering of the chosen 2D projections into user defined number of clusters allowing the user to select one if there are several options. This module can visualize also 3D cube gathering 2D slices as chosen by the user period of time or can show in parallel the several 2D pictures for selected time instants to support the decision. When the results are approved by the user, the tuned ESN parameters, the numbers of the chosen neurons for 2D projection and the calculated cluster centres are kept into a file to be used further in on-line mode by the last module.
- Module 4 performs at each time step all procedures performed by Modules 1, and 2 obtaining the initial features and then the second step features. Then it maps each receptive field to a cluster using known cluster centres for the chosen 2D

projection, visualizes a 2D slice and adds it to the 3D cube containing all consecutive trough time slices. This module generates a 3D figure in Matlab format that can be rotated. The collection of 2D slices can be kept to be used for animated projection. An option to visualize the different clusters propagation through time can also be included depending on the users need.

4. Results

Here the work of the software package is demonstrated on the test experiment with a moving sound source.

In the experiment carried out a piezo beeper WB 3509 with constant frequency of 2.43 kHz was used as a sound source. The acoustic pressure in Pa measured by all microphones is collected in five data buffers for totally 2.49 s with a sampling time of 1.53×10^{-5} s. Fig. 5 shows the experimental set-up (left) and the produced by an acoustic camera software "sound picture" in case of a static sound source (right). During the measurement the beeper was manually moved from left to right in an attempt to test the moving sound source detection ability of our software. The raw data taken is shown in Fig. 6. The collected data are periodic signals with variable amplitude and constant frequency of the noise source (the beeper).

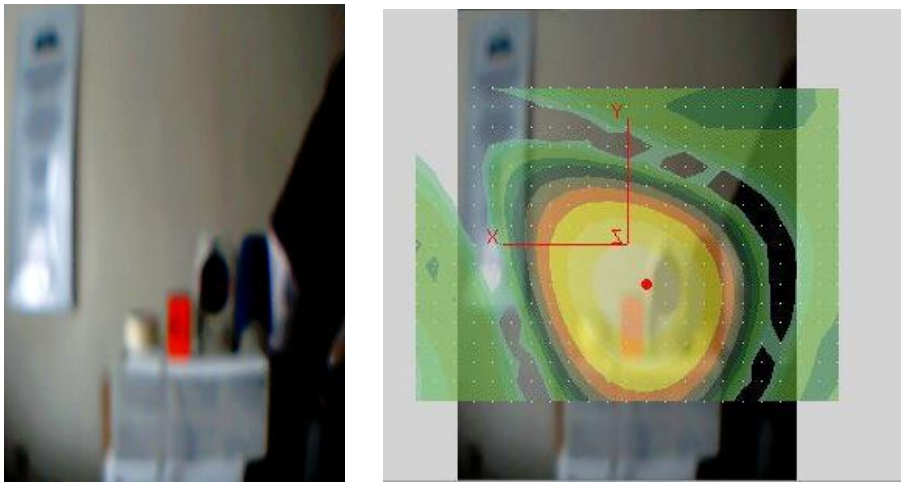

Fig. 5. Picture form the optical camera (left) and the sound picture produced by original software (right). The sound source is the red box

The investigations for proper parameters of our algorithm [10, 12] were carried out using a static beeper as a sound source, as well as two static beepers. Based on the results it was decided that 11 MT filters, 30 ENS neurons and 3 clusters suit well to demonstrate the performance of our algorithm for this particular sound source. These parameters were used during the experiment with the moving sound source above described. The obtained results are demonstrated bellow.

The data from the first buffer was used as a training data set. Fig. 7 demonstrates the esults from the first step of our signal processing procedure – outputs of several MT filters for several time steps. The black dots mark the centres of the 16 receptive fields (as in Fig. 2).

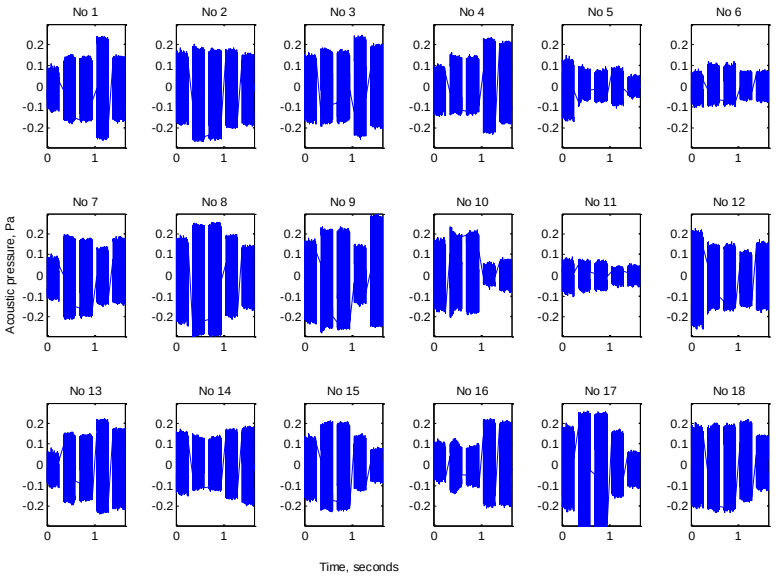

Fig. 6. Stimuli (acoustic pressure) recorded by 18 sensors (microphones) through all periods of measurements separated into five buffers

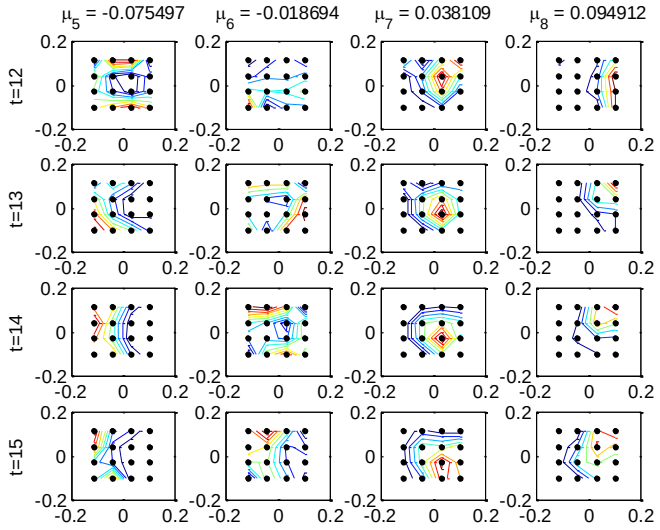

Fig. 7. Example of receptive fields for several MT filters and several consecutive time instants

Figs 8 and 10 demonstrate the obtained in 2D slices “sound pictures” produced by our software for the first 28 time steps (since this is approximately one period of our beeper) for the training data set (the first buffer) and for a part of the rest of data or testing data set (here the 4th buffer was chosen). Figs 9 and 11 demonstrate 3D views obtained by gathering the 2D slices from the corresponding to them Figs 8 and 10.

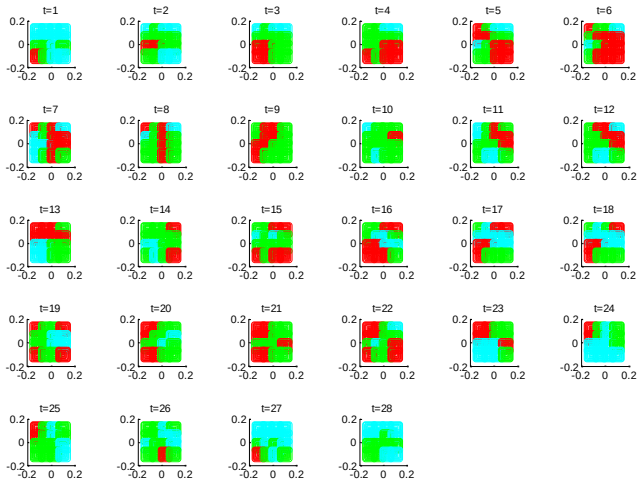

Fig. 8. Sound waves propagation through the first 28 time steps (buffer 1)

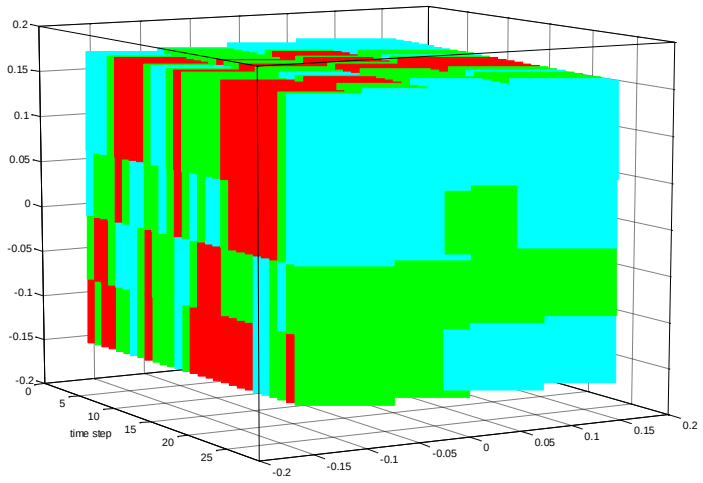

Fig. 9. 3D view of sound waves propagation through the first 28 time steps (buffer 1)

From these figures the following conclusions are made:

- The consecutive 2D sound pictures reveal the periodic nature of the sound waves recorded during the experiment since the first and the last slices are almost identical for both buffers.
- The changed position of the sound source was observed too. Knowing its initial and final position we can say that the beeper is marked by the red cluster in all pictures. Hence, its movement from left to right can be detected by comparing the corresponding 2D sound pictures from the first and fourth data buffer.

- The produced by our software 3D sound pictures visualize the propagation of sound waves arriving to the plane of the microphone array through time. Since reflection from different walls and barriers in the room is inevitable, its influence on the recorded through time acoustic pressure is also detected. Hence, these visualizations can be used to reveal the specific acoustic characteristics of the room where the experiment was carried out.

5. Conclusion

The initial tests of our algorithm and the developed Matlab software package demonstrated ability to visualize sound waves propagation through time and to detect the static and dynamic sound sources in the observed by acoustic camera area.

The future work intended to improve the visualization quality will be focused towards the following two directions:

- The obtained by far results use a very rough grid of only 16 receptive fields. Increasing their number and hence, their density will allow obtaining of more detailed sound pictures.
- Exploitation of the fuzzy membership to a cluster of each receptive field instead of crisp mapping used in the present study will allow including much more grades between the different clusters that will enrich the visualization quality too.

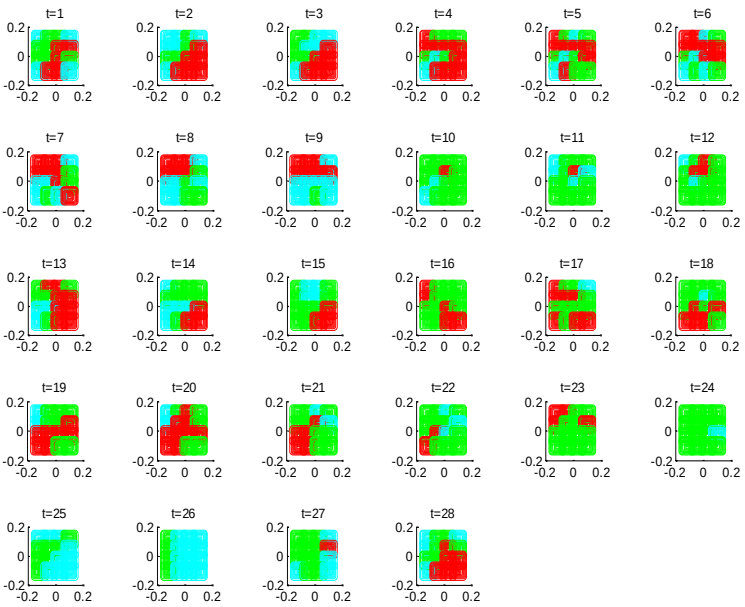

Fig. 10. Sound waves propagation through the first 28 time steps after movement of the sound source (buffer 4)

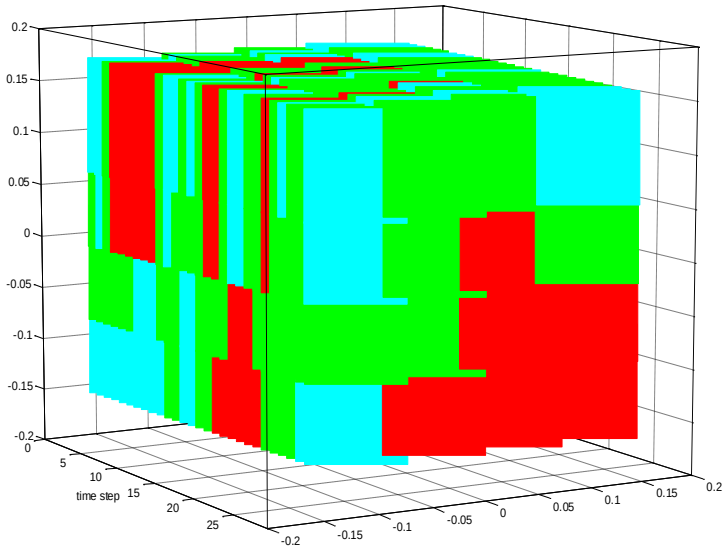

Fig. 11. 3D view of sound waves propagation through the first 28 time steps after movement of the sound source (buffer 4)

***Acknowledgements*:** This work is partially supported by Project AComIn, "Advanced Computing for Innovation", Grant 316087, funded by FP7 Capacity Program (Research Potential of Convergence Regions) Grant 316087, funded by FP7 Capacity Program and Project No DO1-192.

References

1. Bai, M. R., J.-G. Benesty, J. Ih. Acoustic Array Systems: Theory, Implementation, and Application. 1st Edition. Wiley-IEEE Press, 2013.
2. Beardsley, S. A., R. L. Ward, L. M. Vaina. A Neural Network Model of Spiral-Planar Motion Tuning in MSTd. – Vision Research, Vol. **43**, 2003, pp. 577-595.
3. Bezdek, J. C., R. Ehrlich, W. Full. FCM: The Fuzzy c-Means Clustering Algorithm. – Computers & Geosciences, Vol. **10**, 1984, No 2-3, pp. 191-203.
4. Jacobsen, F., V. Jaud. Statistically Optimized Near Field Acoustic Holography Using an Array of Pressure-Velocity Probes. – Journal of the Acoustical Society of America, Vol. **121**, 2007, No 3, pp. 1550-1558.
5. Jaeger, H. The "Echo State" Approach to Analysing and Training Recurrent Neural Networks with an Erratum Note. GMD Report 148: German National Research Center for Information Technology, 2001.
6. Koprinkova-Hristova, P., K. Alexiev. Sound Fields Clusterization via Neural Networks. – In: 2014 IEEE International Symposium on Innovations in Intelligent Systems and Applications, 23-25 June 2014, Alberobello, Italy. DOI 10.1109/INISTA.2014.6873646.
7. Koprinkova-Hristova, P., K. Alexiev. Dynamic Sound Fields Clusterization Using Neuro-Fuzzy Approach. – In: Lecture Notes in Computer Science. Vol. **8722**. 2014, pp. 194-205.
8. Koprinkova-Hristova, P., N. Tontchev. Echo State Networks for Multi-Dimensional Data Clustering. – Lecture Notes in Computer Science. Vol. **7552**. Part 1. 2012, pp. 571-578.

9. L u k o s e v i c i u s, M., H. J a e g e r. Reservoir Computing Approaches to Recurrent Neural Network Training. – Computer Science Review, Vol. **3**, 2009, pp. 127-149.
10. P o p o v a, N., G. S h i s h k o v, P. K o p r i n k o v a - H r i s t o v a. Dynamic Sound Fields Clustering via Neuro-Fuzzy Approach. – In: Proc. of XIII Nat. Scientific-Practical Conference for Young Researchers, 27-28 April. Sofia. ISSN: 1314-8931. pp. 89-94 (in Bulgarian).
11. S c h r a u w e n, B., M. W a n d e r m a n n, D. V e r s t r a e t e n, J. J. S t e i l, D. S t r o o b a n d t. Improving Reservoirs Using Intrinsic Plasticity. – Neurocomputing, Vol. **71**, 2008, pp. 1159-1171.
12. S h i s h k o v, G., N. P o p o v a, K. A l e x i e v, P. K o p r i n k o v a - H r i s t o v a. Investigation of Some Parameters of a Neuro-Fuzzy Approach for Dynamic Sound Fileds Visualization. – In: 2015 IEEE International Symposium on Innovations in Intelligent Systems and Applications, 2-4 September 2015, Madrid, Spain, DOI: 10.1109/INISTA.2015.7276769.
13. W i l l i a m s, E. G., J. D. M a y n a r d, E. J. S k u d r z y k. Sound Source Reconstructions Using a Microphone Array. – Journal of the Acoustical Society of America, Vol. **68**, 1980, No 1, pp. 340-344.
14. Y a g e r, R., D. F i l e v. Generation of Fuzzy Rules by Mountain Clustering. – Journal of Intelligent & Fuzzy Systems**,** Vol. **2**, 1994, No 3, pp. 209-219.