Dynamic Sound Fields Clusterization Using Neuro-Fuzzy Approach

Petia Koprinkova-Hristova and Kiril Alexiev

Inst. of Inf. and Comm. Technologies, Bulgarian Academy of Sciences
Acad. G. Bonchev str. bl.25A,
Sofia 1113, Bulgaria
{pkoprinkova,alexiev}@bas.bg

Abstract. In the presented investigation a recently proposed approach for multidimensional data clustering was applied to create a 3D "sound picture" of the data collected by a microphone array antenna. For this purpose records of acoustic pressure at each point (a microphone in the array) collected for a given period of time were used. Features for classification are extracted using overlapping receptive fields based on the model of direction selective cells in the middle temporal (MT) cortex. Next the clustering procedure using Echo state network and subtractive clustering algorithm is applied to separate these receptive fields into proper number of classes. Obtained for each time step two dimensional "sound pictures" were combined to create a 3D representation of dynamic changes in the sound pressure. We compare our results with the sonograms created by the original software of the producer of microphone array. Although our approach did not account for the distance to the noise source, it allows consideration of dynamically changing sounds.

Keywords: acoustic pressure, Echo state networks, subtractive clustering, receptive field, direction selective cells.

1 Introduction

Localization of sound sources is a task with numerous applications varying from military locators, seismic surveys, medical and machine diagnostic systems etc. For different practical applications there were created many specialized equipments and corresponding mathematical methods for signal processing aimed at accurate noise source localization. An example of such device is acoustic camera that consists of several microphones operating in tandem.

There are developed two basic approaches for processing of acoustic pressure measured by the microphone arrays: acoustic holography and beam forming. The first one reconstructs sound fields near to the camera and has established two realizations: near-field acoustic holography (NAH) [16] and statistically optimal near-field acoustic holography (SONAH) [4]. There are several strong requirements that cumber NAH implementation and limit its application to small sound sources at low frequencies. That is why SONAH was developed to alleviate some of these requirements. The

G. Agre et al. (Eds.): AIMSA 2014, LNAI 8722, pp. 194–205, 2014.

other basic approach – beam forming (BF) – was created for localization of medium and long distance sound sources. In both cases the core of the task to be solved is to divide the area observed by the acoustic camera into sub-areas in dependence on multidimensional measurement data. Hence we decided to apply a clustering approach to solve it.

In spite of numerous developments, clustering of multidimensional data sets is still a challenging task [6]. There are numerous approaches for solving it including intelligent techniques based on fuzzy logic and neural networks. In [8, 11] we proposed a new multidimensional data clustering approach that combines model of direction selective cells in the middle temporal (MT) cortex and recurrent neural networks for features extraction and fuzzy subtractive clustering for blind separation of data into clusters. Variations of the algorithm were successfully applied by now to different static and dynamic data sets: landscape classification using multi-spectral satellite image of a mountain region in Bulgaria [9, 10], clustering of dynamic data taken from an experiment that tests visual discrimination of complex dot motions [11] and classification of accumulated acoustic pressure measured by a microphone array [12].

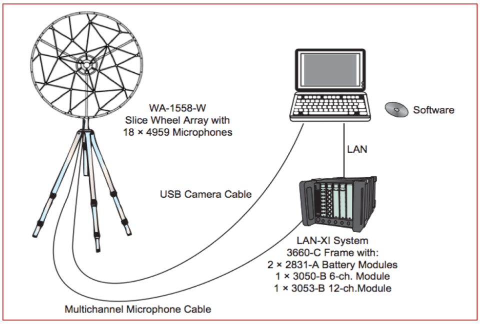
By far the algorithm was used to create 2D picture of clusters of multidimensional data sets. The present investigation extends application of the algorithm to 3D visualization including time course of data as third dimension. The suggested approach is simplified one – it uses sound intensity only and doesn't account for the distance to the sound source. However it will allow detecting not only of static but also of moving sound sources – function that is not included in the original software version supplied by Brüel & Kjær. The obtained results were compared with the 2D sonograms created by original software. Our future intention is to incorporate it in the system we have and to extend it with ability to detect moving sound sources.

The paper is organized as follows: in next section we describe the experimental set-up, the equipment (acoustic antenna) and data collection procedure; section 3 gives short description of our algorithm with accent to its extension to 3D dynamic task; in section 4 results of dynamic data clustering and 3D visualization are presented and discussed in comparison with the sonograms created by the original software of Brüel&Kjær; the paper finishes with conclusions and directions for future work.

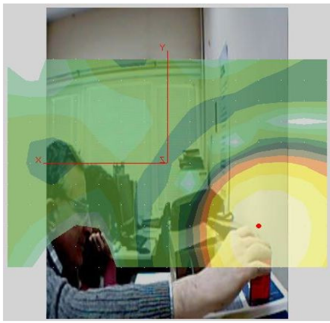
2 Experiment Set-up

2.1 Brüel and Kjær Microphone Array

Multidimensional data for testing of our approach was collected using the system from Brüel & Kjær for sound analysis shown on Figure 1 (a). It consists of 18 microphones array placed randomly in a wheel grid called antenna. At the center of antenna is mounted camera. All microphones are connected to a front-end panel. Both camera and front-end are connected to a computer (via USB and LAN cables correspondingly) with software for sensor information processing. The system measures acoustic pressure and visualizes "sound picture" of the observed by camera area as it is shown on Figure 1 (b).



(a)



(b)

Fig. 1. Brüel & Kjær system for sound analysis (a) and created by it “sound picture” (b)

2.2 Raw Data Collection

Our multidimensional data set consists of raw measurement data from all 18 microphones in antenna array. A piezo beeper WB 3509 (standard Brüel & Kjær equipment – the red box in the right low corner on the picture on Figure 1 (b)) with frequency of 2.43 kHz was used as sound source. After switching on the beeper the system collects acoustic pressure in Pa for 15.9ms – period of time predetermined by the system software – from all 18 microphones. The measurements were taken with time step $1.53*10^{-5}$ s. The collected data are periodic signals with variable amplitude and constant frequency of the noise source (the beeper). The input signal amplitude is different for each microphone due to attenuation of different beeper – microphone path loss.

3 Clusterization Algorithm

3.1 Initial Feature Extraction Procedure

In [11], following the model of human visual perception from [1, 2], we used the receptive fields of MST neurons to pre-process time series of our dot motion data. This model has been widely used to examine the emergence of complex motion pattern properties [1, 2]. The receptive fields are direction selective cells in middle temporal (MT) cortex described by the following equation [1]:

$$f_{il}(t) = \frac{1}{N}\sum_{k=1}^{N} \exp\left(\frac{-(\mu_i - s_k(t))^2}{2\sigma^2}\right) \tag{1}$$

Here $f_{il}(t)$ is the response of i-th MT unit to k-th input stimuli $s_k(t)$ for the l-th receptive field (area of stimuli collection) at time t; μ_i is center and σ is variance of Gaussian curve defining each filter response; N is number of inputs, i.e. stimuli received in the *l*-th field. In present work we divide the area of stimuli (in considered example these are microphone sensors readings) into several overlapping regions, each containing at least one stimulus (sensor) input. In [12] we accumulate receptive fields' outputs at each area and average them over all time period of measurements thus accounting for accumulated acoustic pressure. In present work we use receptive fields' outputs for a given moment in the time in order to account for time changes in the sound picture.

Division of observed by acoustic camera area into 16 overlapping square regions is shown on Figure 2 (each region is surrounded by a dashed line square with rounded edges). The small red dots with numbers represent corresponding microphone position and the big red dot in the center marks camera position. Each region contains at least one microphone (e.g. microphone 5 is the only one in upper right region). Maximal number of microphones in region is four and it is situated at the center of antenna (e.g. region containing microphones 1, 3, 8 and 7).

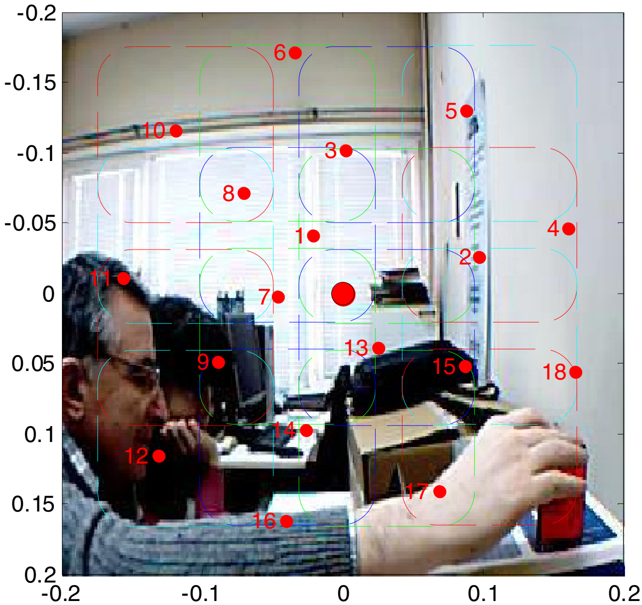


Fig. 2. Regions positions at the antenna area

Next, in order to design our receptive filed units, we divide the dynamic range of raw data (that is from -0.4 to 0.4 Pa) into 11 intervals. For each interval we define a filter with center μ_i at the center of interval and variance σ equal to one third of interval size. Thus our receptive fields overlap covering intervals from -3σ to $+3\sigma$ around their centers.

Equation (1) describes obtained at this first step feature vectors of area number l where $i=1 \div n_f$ and $n_f=11$ is number of filters in our experiment. Thus the obtained for the period of time from 0 to t_f data set of features is:

$$f_{il}(t)\Big|_{t=1\div t_f,\ i=1\div 11,\ l=1\div 16} \tag{2}$$

These features are inputs to the recurrent neural network used in the second step of feature extraction procedure.

3.2 Final Feature Extraction

At this step we exploit the equilibrium states of neurons of a special kind of recurrent neural network – Echo state network (ESN) [5, 13] – as final features extracted from multidimensional data. The basic idea was proposed for the first time in [8]. Here we'll describe it briefly.

The structure of ESN is presented on Figure 3 bellow. It consists of a randomly generated dynamic reservoir of interconnected neurons having also feedback from their own outputs. The reservoir connections weight matrix is denoted by W^{res}. All reservoir neurons receive as input a vector denoted here by u multiplied by input weight matrix denoted by W^{in}. The output of reservoir is a simple sigmoid function (usually hyperbolic tangent) that depends on current input as well as on previous state of the reservoir neurons.

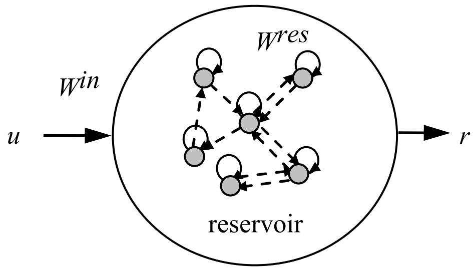


Fig. 3. Echo state network (ESN) structure

Following the proposed in [14, 15] algorithm for initial tuning of reservoir weights and conclusions from [7] that achieved equilibrium states of reservoir neurons after such tuning reflect the structure of training data set, in [8] emerged the idea that the reservoir equilibrium can serve as a feature vector.

Inputs to our second feature extractor – ESN – are initial features extracted by receptive fields, i.e. for l-th area of stimuli collection:

$$u_l(t) = \left[f_{1l}(t) \quad f_{2l}(t) \ \dots f_{11l}(t)\right]_{t=1\div t_f,\ l=1\div 16} \tag{3}$$

Final features used for data clustering are equilibrium states r_e of trained ESN neurons that are calculated by presenting each vector as constant input to the ESN until all neurons outputs settle down, i.e.:

$$r_{el}(t) = \tanh\left(diag(a)W^{res} r_{el}(t) + diag(a)W^{in} u_l(t) + b\right) \tag{4}$$

Here a and b are additional vectors of parameters used to tune the reservoir according to [13, 14].

3.3 Overall Clustering Procedure

Measurement data (stimuli) are collected for given period of time from all sensors in considered area. The clustering algorithm is as follows:

- The collected data are pre-processed using first step feature extraction procedure and data set (2) is generated;
- A random ESN reservoir is generated and tuned to these data;
- The trained reservoir equilibriums are determined according to (4); then they are scaled within interval [-1, +1];
- All possible two dimensional projections between equilibrium states of every two different neurons in the reservoir i and j and for each period of time step t are generated as follows:

$$P_{ij}(t) = \left\lfloor r_{el}^{i}(t), \quad r_{el}^{j}(t) \right\rfloor_{l=1\div 16} \tag{5}$$

- Subtractive clustering procedure [17] is applied to all projections (5) in order to determine number and centers of data clusters. This procedure was chosen since it is reported as one of the best options in the case of unknown number of clusters [3];
- The projections with highest number of clusters are selected;

4 Results and Discussion

It was observed that acoustic pressure data is periodic with period of about 0.412 ms or approximately 28 time steps as it is shown on Figure 4. Hence we decided to investigate time changes of "sound picture" during one period as well as for all the time of measurements with 0.412 ms time step.

At the second step of described above feature extraction algorithm we used ESN reservoirs with different sizes: 10, 30 and 50 neurons. In all cases the number of inputs of ESN was determined by the number of features, i.e. 11 according to the number of receptive fields. For each new generation of reservoir number of obtained by our algorithm clusters varies but in most cases we have mainly 3 or 4 clusters.

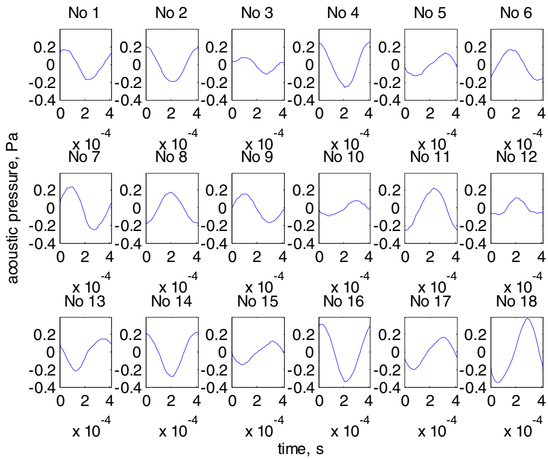


Fig. 4. Microphone signals for the first 28 time steps (approximately one period of signal)

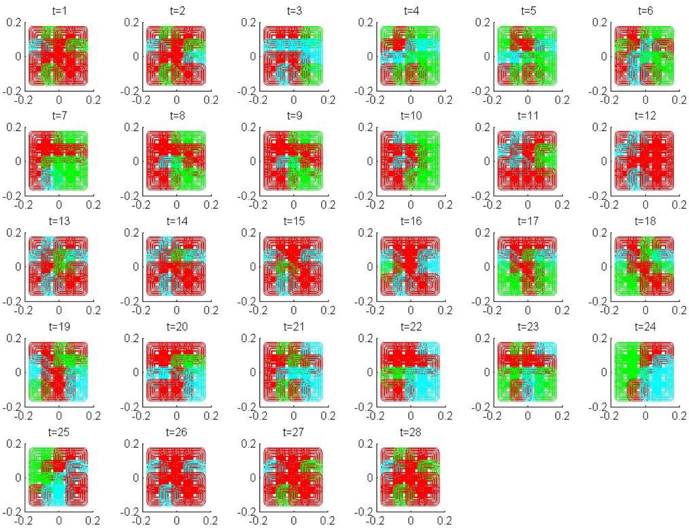
The bigger was number of neurons in ESN reservoir, the bigger is number of possible two dimensional projections with maximal number of clusters. The number of obtained clusters however is smaller in comparison with those obtained in [12] (2D clusters case) where we had about 6 clusters.

We consider each sub-region in antenna area as an area in the picture taken from camera that has to be classified. Each cluster is covered by rectangles with different color. Figures 5, 6 and 7 present classification results obtained by using ESN reservoirs with 10, 30 and 50 neurons respectively. On each figure (a) is "unfolded" 3D picture for the first period of measurements and (b) – for all periods of measurements with time step equal to period duration (0.412 ms).

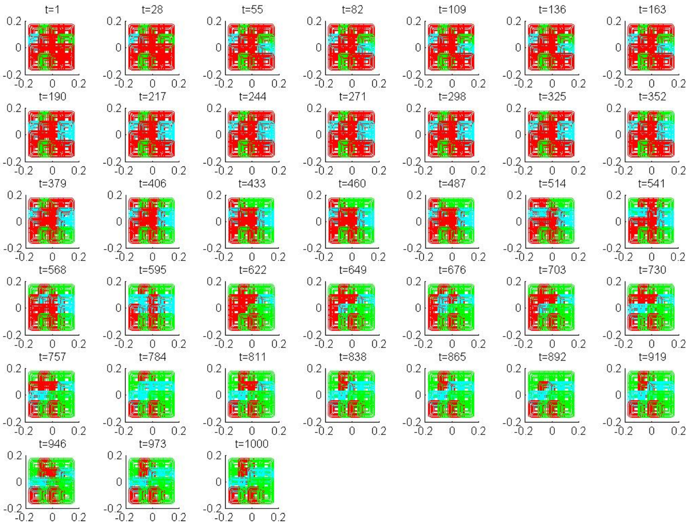
From all figures (a) we can observe "movement" of the sound wave coming from the noise source through receptive fields for the first period of time. The last picture (beginning of new period) is the same as the first one, i.e. our classification is able to reveal periodical characteristics of data. In spite of roughness of our sensing fields, the position of the beeper can be exactly estimated without usage of any information about free-space path loss formula and propagation delay.

On figures (b) we observe time changes of "sound picture during all time of measurements. The "unfolded" 3D picture reveals changes in the acoustic pressure amplitude with time. Although all pictures are from the beginning of current period, they gradually change from the beginning of the measurements to their end. The change in pictures is due to beating frequency of inexact correspondence of sampling frequency and beeper frequency. The pictures visualize 2D beating frequency propagation and can serve as an instrument for spectral analysis of the input signal.

Comparison with Figure 1 (b) is to rough but it is clear that in all pictures the area with noise source (low right corner) is recognized to be different from the other areas in the picture. Having in mind that "sound picture" form original software is accumulative for all time of collecting data and our "sound pictures" are for different time steps, the observed differences reveal dynamic nature of the sound signal.

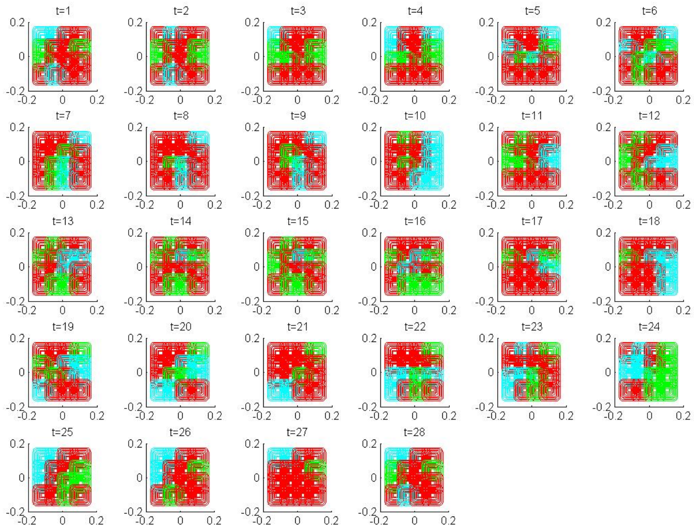


(a)

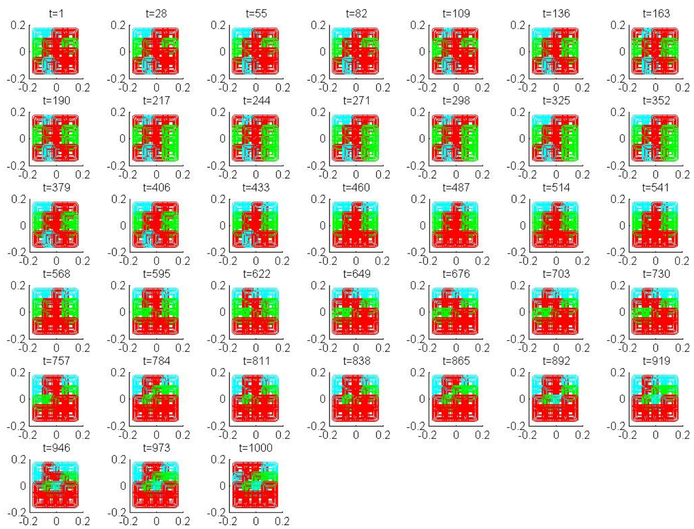


(b)

Fig. 5. Clusters obtained with 10 neurons (a) for the first period and (b) for all the time with step 0.412 ms

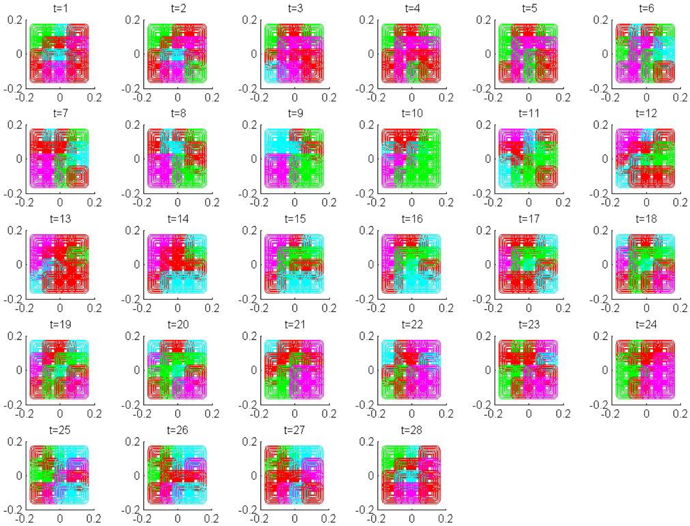


(a)

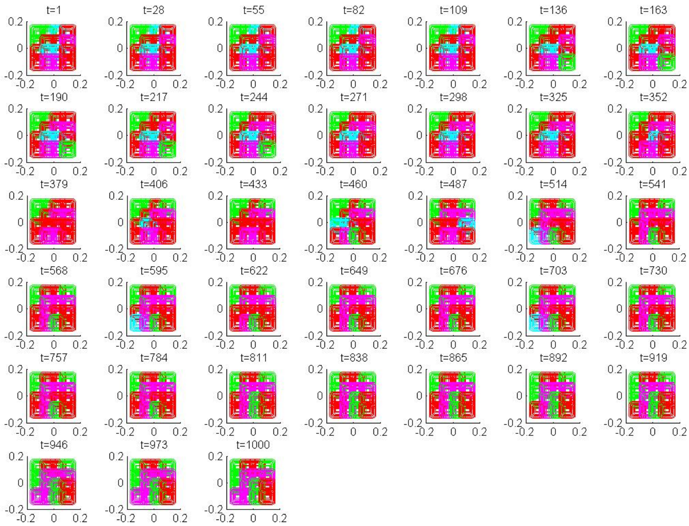


(b)

Fig. 6. Clusters obtained with 30 neurons (a) for the first period and (b) for all the time with step 0.412 ms



(a)



(b)

Fig. 7. Clusters obtained with 50 neurons (a) for the first period and (b) for all the time with step 0.412 ms

5 Conclusions

The presented in the paper application of currently developed algorithm for multidimensional data clustering and its extension to dynamic data representation in 3D showed promising results and pointed out to directions for future developments.

First of all, testing of our approach on yet another multidimensional data set and comparison of the results with a professional signal processing software demonstrated that our approach although not that refined is promising and gives similar results.

Possible refinement of the results could be obtained via fuzzy visualization of clusters. By the moment although we use fuzzy clustering algorithm, we classify each dot on the picture based on only its distance to the cluster centers. However subtractive clustering allows us to have also overlapping clusters with fuzzy membership of each dot. Hence we can obtain pictures with gradual change of colors that will be much closer to that of original software.

Another interesting way for further improvement is the application of described scheme for multisource input signal, localization of each source and spectral identification.

Acknowledgments. The research work reported in the paper is partly supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Program (Research Potential of Convergence Regions).

References

1. Beardsley, S.A., Ward, R.L., Vaina, L.M.: A neural network model of spiral–planar motion tuning in MSTd. Vision Research 43, 577–595 (2003)
2. Grossberg, S., Pilly, P.K.: Temporal dynamics of decision-making during motion perception in the visual cortex. Technical Report BU CAS/CNS TR-2007-001 (February 2008)
3. Hammouda, K.: A comparative study of data clustering Techniques. SYDE 625: Tools of Intelligent Systems Design, Course Project (August 2000)
4. Jacobsen, F., Jaud, V.: Statistically optimized near field acoustic holography using an array of pressure-velocity probes. J. Acoust. Soc. Am. 121(3), 1550–1558 (2007)
5. Jaeger, H.: Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach. GMD Report 159, German National Research Center for Information Technology (2002)
6. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: A review. ACM Computing Surveys 31(3), 264–323 (1999)
7. Koprinkova-Hristova, P., Palm, G.: ESN intrinsic plasticity versus reservoir stability. In: Honkela, T. (ed.) ICANN 2011, Part I. LNCS, vol. 6791, pp. 69–76. Springer, Heidelberg (2011)
8. Koprinkova-Hristova, P., Tontchev, N.: Echo state networks for multi-dimensional data clustering. In: Villa, A.E.P., Duch, W., Érdi, P., Masulli, F., Palm, G. (eds.) ICANN 2012, Part I. LNCS, vol. 7552, pp. 571–578. Springer, Heidelberg (2012)

9. Koprinkova-Hristova, P., Alexiev, K., Borisova, D., Jelev, G., Atanassov, V.: Recurrent neural networks for automatic clustering of multispectral satellite images. In: Bruzzone, L. (ed.) Image and Signal Processing for Remote Sensing XIX, October 17. Proceedings of SPIE, vol. 8892, p. 88920X (2013), doi:10.1117/12, ISSN: 0277-786X, ISBN: 9780819497611
10. Koprinkova-Hristova, P., Angelova, D., Borisova, D., Jelev, G.: Clustering of spectral images using Echo state networks. In: 2013 IEEE International Symposium on Innovations in Intelligent Systems and Applications, IEEE INISTA 2013, Albena, Bulgaria, June 19-21 (2013), doi:10.1109/INISTA.2013.6577633, ISBN: 978-147990661-1
11. Koprinkova-Hristova, P., Alexiev, K.: Echo State Networks in Dynamic Data Clustering. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A.E.P., Appollini, B., Kasabov, N. (eds.) ICANN 2013. LNCS, vol. 8131, pp. 343–350. Springer, Heidelberg (2013)
12. Koprinkova-Hristova, P., Alexiev, K.: Sound fields clusterization via neural networks. In: 2014 IEEE Int. Symposium on Innovations in Intelligent Systems and Applications, Alberobello, Itally, June 23-25 (accepted paper, 2014)
13. Lukosevicius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. Computer Science Review 3, 127–149 (2009)
14. Schrauwen, B., Wandermann, M., Verstraeten, D., Steil, J.J., Stroobandt, D.: Improving reservoirs using intrinsic plasticity. Neurocomputing 71, 1159–1171 (2008)
15. Steil, J.J.: Online reservoir adaptation by intrinsic plasticity for back-propagation-decoleration and echo state learning. Neural Networks 20, 353–364 (2007)
16. Williams, E.G., Maynard, J.D., Skudrzyk, E.J.: Sound source reconstructions using a microphone array. J. Acoust. Soc. Am. 68(1), 340 (1980)
17. Yager, R., Filev, D.: Generation of fuzzy rules by mountain clustering. Journal of Intelligent & Fuzzy Systems 2(3), 209–219 (1994)