Development and Analysis of Analog-Digital Neural Net for Speech Stress Detection

Vasilii G. Arkhangelsky
CIT&S
Moscow, Russia
citis@arkhang.ru

Sergey A. Alyushin, Alexander V. Alyushin
Electronic Measuring Systems, Cybernetics
National Research Nuclear University MEPHi
Moscow, Russia
avalyushin@mail.ru

***Abstract*—Autonomous SSDs (Speech Stress Detector) are widely used at present. Known electronic realizations suffer from low noise immunity, their methods of stress detection require further development and verification. State-of-the-art integrated technologies permit to superpose digital reprogrammable systems with analog neuron like elements and receive new quality of data processing in SSD. Proposed real-time neural network is characterized by rational representation and distribution of main SSD functions in neural net in analog and digital forms. Single layer structure of ADNN (Anolog-Digital Neural Net) supports self-organization process in SSD during speech analysis, new features extraction and classification. Developed by the authors ADNN SSD prototype has shown higher noise immunity level in comparison with conventional realizations, can be treated as promising model for instrumental base development on the one hand and hardware real time realization in the future embeddable systems on the other.**

Keywords—speech stress detector; neural net; pulsed analog-digital neuron

I. INTRODUCTION

Modern express-diagnostics of human psycho-emotional state plays significant role in rapidly developing computerized society. Acoustic level voice measurements are economical and do not require special sophisticated equipment as physiological and phonetic-articulatory levels [1]. Speech synthesis mechanism is rather complicated and autonomous. Affective states are reflected in physiological reactions, somatic nervous system and modulate certain parameters of voice production process (subconscious level). Only distinguished peoples can control this process of speech modulation on the conscious level while trying to simulate different emotional state. During this process, new emotional stress occurs that can be detected as well.

Autonomous speech stress detectors are the most popular ones. Nevertheless, verified information about their parameters especially in noisy environment is insufficient, used methods require further development and verification from validity and reliability points of view.

II. STATE OF THE ART

Basic principles of frequency and time analysis have been developed and realized in the first portable and embeddable SSDs.

A. Frequency Analysis in SSD

In Fullers SSD (Fuller F.H., 1974) stress condition of a speaking person is diagnosed by the speech vibrato variance [3]. Improved Fullers SSD (Fuller F.H., 1974) evaluates affective state by the product of vibrato level and energy in correspondent frequency band [4]. More than that, energy comparison of modulation signals in different frequency bands increases the accuracy of stress level detection (Fuller F.H., 1974) [5]. In Bells SSD (Bell A.D. et al., 1976) speech classification into normal and under stress classes is performed according to 8 ~ 12 Hz micro tremor modulation level [6].

B. Time Analysis in Speech Stress Detectors

Time analysis deals with temporal structure of speech, prosodic features. In Williamsons SSD (Williamson J.D., 1978) ratio of the basic tone periods sum (relating to the same word) to the word duration is analyzed [7]. Time intervals with different basic tone characteristics are allocated. Methods of speech analysis on time base, developed by Silverman (1987, 1992 and 1999), determine stress condition by the form of amplitude time decay of the speech elements [8 - 10].

Allocation and analysis of micro tremor modulations are the basic functions of portable SSD.

C. Structure of Modern SSD

Modern SSDs in addition to frequency and time analysis use statistical data processing for detecting affective condition. They separate acoustic signals of different simultaneously speaking people, classify input speech series into several additional classes beside stress and operate with higher number of speech parameters (up to 200). For example, Deganis SSD (Degani Y., et al., 2009) determines the stress level for each of two speaking persons as the sum of deviations of secondary voice parameters from predicted in normal state [11]. Petrushins SSD (Petrushin V.A., 2007)

978-1-5386-4340-2/18/$31.00©2018 IEEE

estimates maximum value, mean and standard deviation of the fundamental frequency, range of the fundamental frequency and other statistics (for signal energy, first and second formants, and so on), classifies speech features into several classes (normal, happy, angry, sad and afraid), appreciates the confidence of the decision [12]. Generalized structure of modern SSD is shown in Fig. 1, where FEE – Front End Electronics, AFEE – Analog FEE, ADC – Analog to Digital Converter.

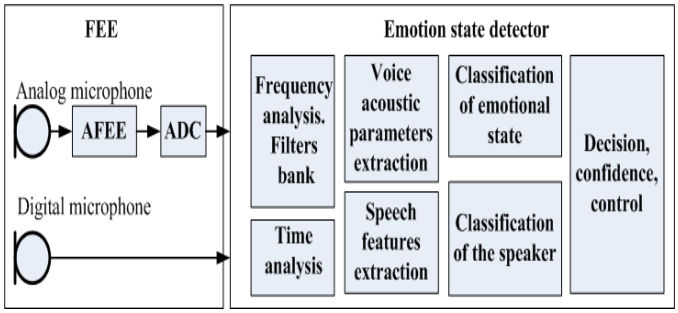

Fig. 1. Generalized structure of modern SSD.

D. Classifier on the Base of a Digital Neural Net

Digital neural network (DNN) processes calculated speech features for correspondent emotional state of a speaker classification into certain class of finite number of possible classes [12]. Trained DNN is used with predetermined states. DNN realization is based on digital hardware microprocessors or software. SSD performance evaluation has shown low level of decision confidence due to speech variability even for a single person in laboratory environment with low acoustic noise level. The aspects of validity and reliability of a used methods have been analyzed not enough detailed [13]. Method validity verification requires reliable electronic real time realizations for stress detection on the fly. Next, we consider electronic realizations of portable SSD.

III. ANALYSIS OF FULLERS TYPE SSD PERFORMANCE

Portable SSDs of Fullers type are characterized by use of AFEE (Fig. 1), limited number of channels for frequency analysis, simple classification and decision schemas [14, 15]. Frequency and time analysis have been performed to determine SSD main parameters, sensitivity and noise immunity (ORCAD v. 10.0).

A. FEE characteristics

Microphone amplifier and input fourth order analog filter are characterized by the following parameters (Fig. 2):

- gain coefficient – 3765,
- frequency band (- 6 dB level), Hz - 340 ~ 800,
- central frequency, Hz - 620.

This frequency range corresponds to the first formant frequency band and is characterized by the location of almost all energy of the speech signal and supposed to have evident level of micro tremor.

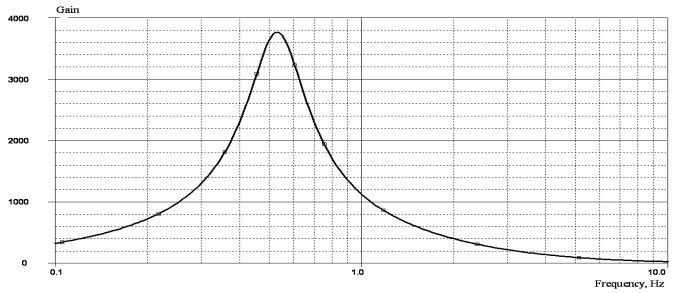

Fig. 2. FEE fourth order filter characteristics.

The envelope of the filtered speech signal is analysed in the frequncy band 0 ~ 120 Hz (look at the freguency response of the envelope third order filter in Fig. 3), which corresponds to fundamental frequency range of an adalt man .

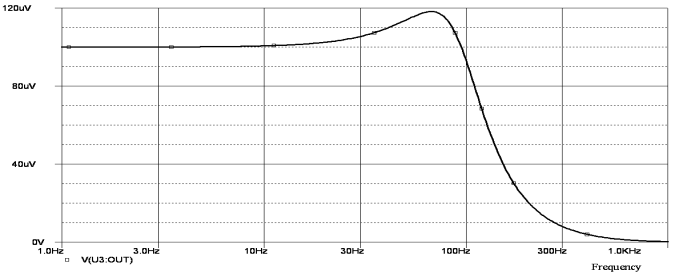

Fig. 3. Third order envelope filter characteristics, where V(U3:OUT) – amplitude of the output signal.

Analog to digital conversion is performed on the base of neuron-like threshold element with Shmitt trigger and monostable multivibrator. Each time the envelope signal exeeds the threshold value, the neuron-like element produces single firing with pulse of constant duration. This sequential digital code is the input signal to emotion state detector.

B. Emotion State Detector

Frequency analysis is performed in two frequency bands A and B. Characteristics of correspondent filters are presented in Fig. 4 and Fig. 5, where V(U6:OUT) and V(R29:2) are the output signals amplitude.

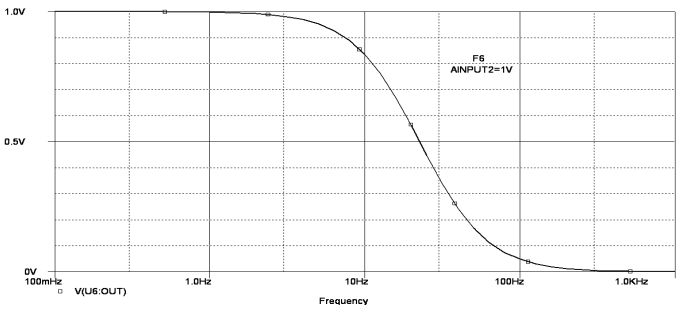

Fig. 4. Second order filter A characteristic.

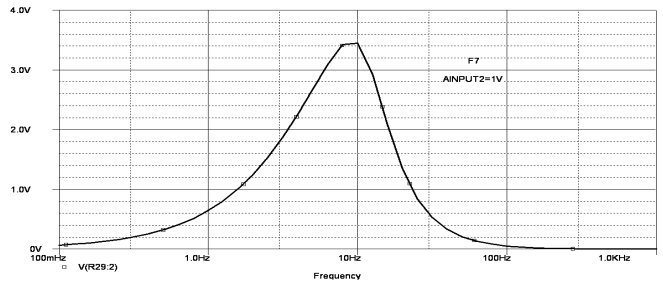

Fig. 5. Second order filter B characteristic.

Main characteristics of these filters are:

- frequency band A, Hz - 0 ~ 25,
- frequency band B, Hz - 3 ~ 20,
- gain coefficient A - 1.0,
- gain coefficient B - 3.42.

Low order of the filters bank provides rapid classification of the stress condition in time interval duration equal to several periods of fundamental frequency. There is no speaker classification. Classification of the emotional state of the speaker is performed by comparison of the mean energy levels in frequency bands A and B. Three classes are used – "Normal", "Normal 50%" and "Stress". High energy level in channel A corresponds to normal state, high energy level in channel B corresponds to stress state. Illustration of the classification process during detecting stress emotional state is presented in Fig. 6. Signal/noise ratio is more than 50 dB.

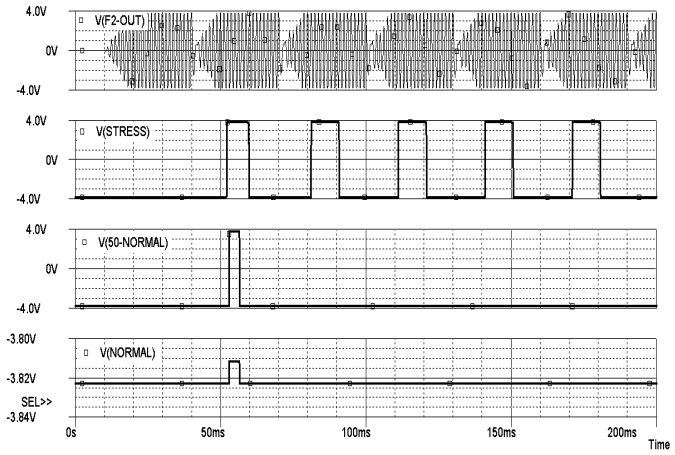

Fig. 6. Stress emotional state detection, where V(F2-OUT) – output signal of the microphone amplifier.

For given value of the threshold voltage and gain coefficient of the FEE, equivalent input signal amplitude can be calculated. In this case, sensitivity of this type of SSD is 0.01 ~ 0.02 mV for the speech signal with 25% ~ 50% micro tremor modulation. This corresponds to near microphone speaker location with low level of environment acoustic noise.

C. Time Analysis of Fulers SSD Performance

Time analysis has been performed in order to determine the SSD behavior in noisy environment. Equivalent electronic circuit for imitation of speech signal in noisy environment is presented in Fig. 7, where V_1 – sinusoidal generator for frequency analysis, V_2 – first formant signal generator, V_3 – micro tremor generator, V_4 – noise generator, V_5 – pulse generator.

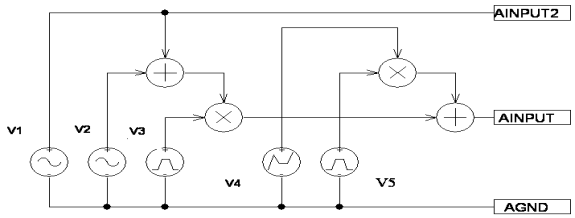

Fig. 7. Equivalent electronic circuit for imitation of the speech signal with micro tremor.

Frequency band of the white noise generator V_4 is 0 ~ 2 KHz. Generator V_5 provides different levels of signal/noise ratio during research. Results of SSD research in noisy environment have shown that reliable detection of normal and stress conditions can be made for signal/noise ratio > 8 ~ 10 dB. In Fig. 8 time diagrams of the output signals for filters A and B, classifier module are presented for different signal/noise ratio.

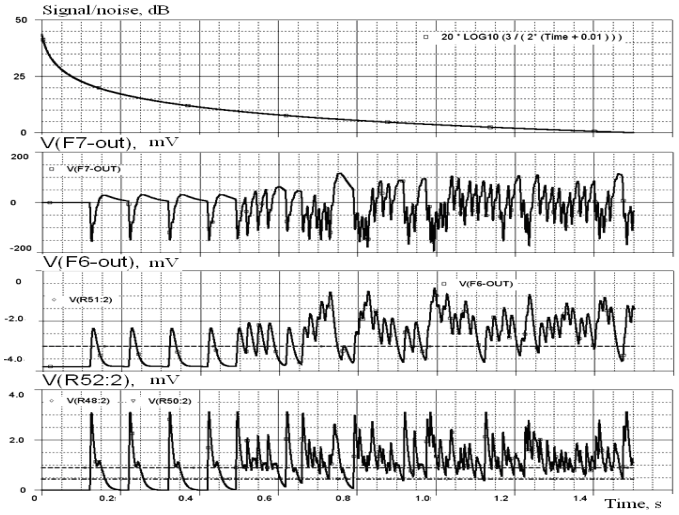

Fig. 8. Output signals of the filters A and B, classifier module for different signal/noise ratio (0 ~ 45 dB), where V(F7-out), V(F6-out) and V(R52:2) – correspondent signals amplitude.

In Fig. 9 and 10 time diagrams of normal and stress conditions detection in noisy environment are presented correspondently, where V(AINPUT) – amplitude of analog input signal; I(D_D17), I(D_D15) and I(D_D14) – SSD output signals, reflecting the states "Stress", "Normal 50%" and "Normal" correspondently.

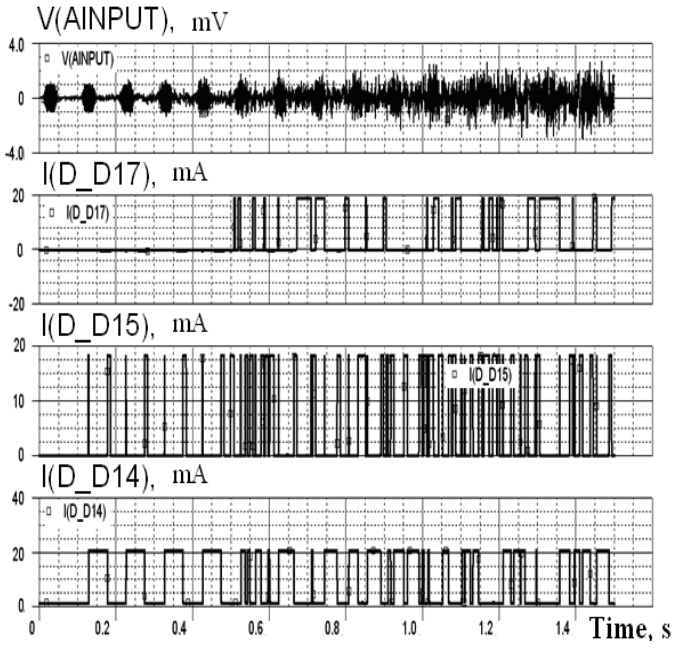

Fig. 9. Time diagrams of normal condition detection in noisy environment.

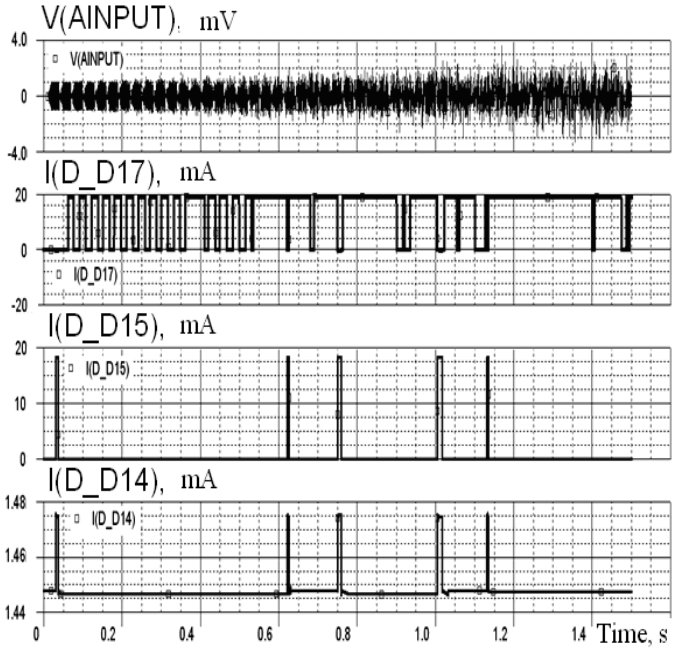

Fig. 10. Time diagrams of stress condition detection in noisy environment.

The most noise sensitive parts of SSD are the neuron-like threshold element, classifier module. Higher SSD noise immunity can be achieved by use of:

- microphone array for beam forming,
- each speaker position monitoring,
- bank of filters with narrow frequency bands,
- neuron type classifier with high noise immunity,
- neurons with frequency selective features.

IV. Emotion State Detection on Base of Analog – Digital Neural Net for Noisy Environment

Further development of affective states detection methods and their hardware realizations requires real-time instrument with rapid reconfiguration, possibility of a new informative parameters allocation on different levels of data processing. State-of-the-art integrated technologies permit to superpose digital reprogrammable systems with and analog neuron like elements and receive new quality of data processing [16]. In this work, we propose to realize all main functions of emotion state detection on the base of analog-digital real-time neural network:

- frequency analysis,
- time analysis,
- acoustic parameters extraction,
- speech features extraction,
- speaker classification,
- library (cluster system) of a speakers formation,
- speaker emotion classification,
- library (cluster system) of the speaker emotion states formation.

Artificial neural net library formation and/or classification are performed simulteneously during a single process of cluster self-organization like Hebbs [17]. This approach permits to extend individual properties of each neuron, use of rational data representation and distribution in SSD in analog and digital forms, perform requared data filtration on each level of data processing with uniform distribution.

Formalisation of the speech criterion adequately reflecting the human affective state is nontrivial task. One of the "intellectual" properties of artificial neural net is its possibility to synthesize appropriate representation of the input data structure by internal self-organization. This direction of research seems to be rather promising in the field of SSD design.

A. Characteristics of Analog – Digital Neuron

Neuron like basic element of an ADNN is charactersed by:

- under threshold data processing,
- integrate or resonance functions,
- time delayed summation, reflecting spatial synapce layout,
- possability to process analog data in wide dynamic range.

Experimental characteristics of designed by authors artificial neurons with resonance properties are illustrated in Fig. 11, where Y_{ib} – output neuron signal, *ib* – neuron index, *b* – step size index, *f* – frequency of the input signal, $A = 1$ – scale coefficient . This figure corresponds to the case with *b* =

1, *n = 26*. In practice, fine spectrum analysis is based on *n*b = 32 ~1024*.

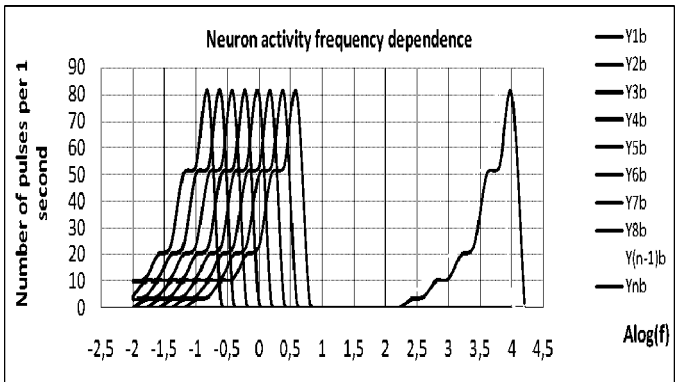

Fig. 11. Neuron activity frequency dependence.

Frequency dependent neuron activity is measured in the pulse number per second metrics. Intencity of neuron firing, resonance frequency are programmed during teaching or self-organisation. According to [13] human muscle vibrato has been detected in the frequency range 1 ~ 100 Hz. Effective speech vibrato frequency band is required to be determined by additional research. Our electronic realization of the artificial neuron supports analysis of input speech in the frequency range 1 ~ 20 000 Hz (Fig. 11, *A = 1*).

B. Structure of ADNN

We have used single layer neural network of analog-digital neurons. In our previous research of a single layer neural net architectural transformations (for example, during self organization or teaching), it has been shown, that this form of a neural net exhibits generalized properties of a whole class of nets with different structures in terms of the number of layers, number of neurons in each layer, connection matrices between layers [18 - 20].

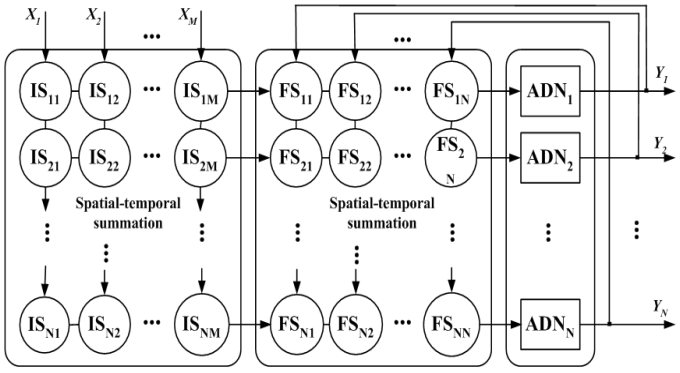

Fig. 12. Structure of a single layer ADNN.

Generalized structure of a single layer ADNN is presented in Fig. 12, where ***X*** and ***Y*** – input and output vectors correspondently, ADN – analog digital neuron, IS – input synapse matrix, FS – feedback synapse matrix. During self-organization, process of grouping of the neurons with similar responses to certain input stimuli characteristics occurs due to interaction between neurons. Illustration of cluster formation in terms of FS matrix is presented in Fig. 13. Self-organized cluster is characterized by two layer network with direct data flow and tree neurons in each layer. In Fig. 14 fundamental frequency parameters detection in correspondent neuron cluster of ADNN is illustrated. Each ADN in this cluster is characterized by different threshold level. Further data processing in the next ADN levels of ADNN highlights precise value of fundamental frequency parameters in wide dynamic range of the input speech signal.

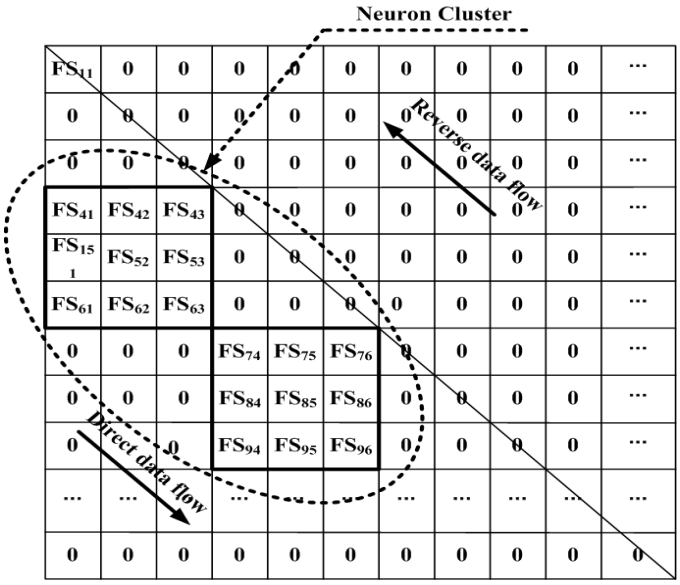

Fig. 13. Illustration of two layer cluster formation with direct data flow.

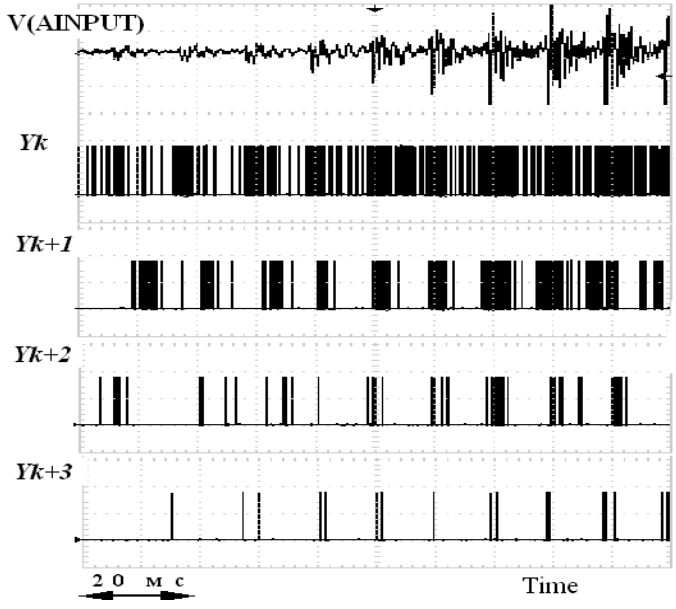

Fig. 14. Fundamental frequency parameters detection in correspondent neuron cluster of ADNN.

Some additional aspects of ADNN implementation are discussed in [21, 22]. Initial teaching and training of the ADNN have been performed on the base of libraries of test speech records. Futher real-time teaching has been organized with the group of several speakers. Experimental research of the trained ANN behavior in noisy environment has shown high noise immunity level. ADNN SSD (laboratory environment, near microphone zone speech sourse location < 1 ~ 5 m, white noise generator). For far microphone zone speech source location microphone array is planned to be used (Fig. 15).

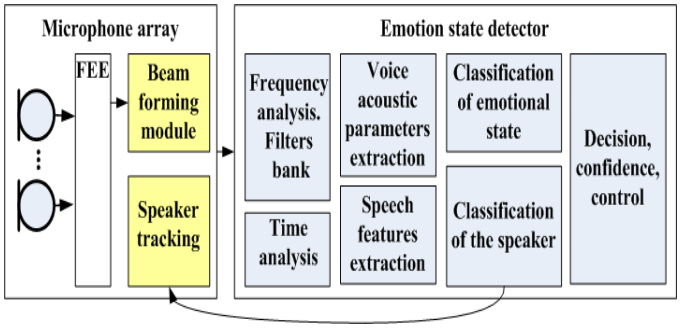

Fig. 15. Generalized SSD structure for high level noise environment.

V. Conclusion

Autonomous SSD are widely used at present. Known electronic realizations suffer from low noise immunity, their methods of stress detection require further development and verification. State-of-the-art integrated technologies permit to superpose digital reprogrammable systems with analog neuron like elements and receive new quality of data processing in SSD. Proposed real-time neural network is characterized by rational representation and distribution of main SSD functions in neural net in analog and digital forms. Single layer structure of ADNN supports self-organization process in SSD during speech analysis, new features extraction and classification. Developed by the authors ADNN SSD prototype has shown higher noise immunity level in comparison with conventional realizations, can be treated as promising model for instrumental base development on the one hand and hardware real time realization in the future embeddable systems on the other.

References

[1] P.N. Juslin, K.R. Scherer, “Speech emotion analysis,” [Online]. Available: http://www.scholarpedia.org/article/Speech_emotion_analysis

[2] P.N. Juslin, P. Laukka, “Communication of emotions in vocal expression and music performance: different channels, same code?” [Online]. Available: http://www.psyk.uu.se/digitalAssets/510/c_510552-l_1-k_juslin_emotion2003.pdf

[3] F.H. Fuller, “Method and apparatus for phonation analysis leading to valid truth/lie decisions by fundamental speech-energy weighted vibrato component assessment,” Patent USA 3855418, G10I 1/04, 17.12.1974.

[4] F.H. Fuller, “Method and apparatus for phonation analysis leading to valid truth/lie decisions by fundamental speech-energy weighted vibrato component assessment,” Patent USA 3855416, G10I 1/04, 17.12.1974.

[5] F.H. Fuller, “Method and apparatus for phonation analysis lending to valid truth/lie decisions by spectral energy region comparison,” Patent № 3855417 USA, G10I 1/04, 17.12.1974.

[6] A.D. Bell et. al., “Physiological response analysis method and apparatus,” Patent № 3971034 USA, G01D 1/04, 20.06.1976.

[7] J.D. Williamson, “Speech analyzer for analyzing pitch or frequency perturbations in individual speech pattern to determine the emotional state of the person,” Patent № 4093821 USA, G10L 1/00, 06.06.1978.

[8] S.E. Silverman, “Method for detecting suicidal predisposition,” Patent № 4675904 USA, G10L 5/00, 23.06.1987.

[9] S.E. Silverman, “Method for detecting suicidal predisposition,” Patent № 5148483 USA, G10L 5/00, 15.09.1992.

[10] S.E. Silverman, “Method for detecting suicidal predisposition,” Patent № 5976081 USA, A61B 5/00, 2.11.1999.

[11] Y. Degani, et. Al., “Method and apparatus for determining emotional arousal by speech analysis,” Patent № 7606701 B2 USA, G10L 11/04, 19/00, 21/00, 20.10.2009.

[12] V.A. Petrushin, “Detecting emotions using voice signal analysis,” Patent USA 7222075, 22.05.2007.

[13] A. Eriksson, F. Lacerda, “Charlatanry in forensic speech science: a problem to be taken seriously,” [Online]. Available: ttps://pdfs.semanticscholar.org/e96e/74efe5606ba1c0ffec30a47486c12a2f5fc4.pdf

[14] C. McNeice, R. Cota, “Build a vocal “Truth” analyzer,” Popular electronics, april 1980, pp. 66–71.

[15] A.V. Alyushin, M.V. Alyushin, S.A. Alyushin, L.V. Kolobashkina, N.A. Korotkova, “The analysis of the vocal stress detector performance in the presence of acoustic noise,” Natural and technical sciences, No. 1, 2010, pp. 283-288.

[16] A.V. Alyushin, M.V. Alyushin, S.A. Alyushin, “The net of pulsed neurons with the delay on the basis of the analog-digital field programmable integrated circuit,” Proceedings of the 5th. Int. conf and Exibition “Digital signal processing and its applications”, Moscow, Russia, March 12-14, 2003, pp. 582-585.

[17] D.O. Hebb, “Organization of behavior,” New York: Science edition, 1949.

[18] A.V. Alyushin, M.V. Alyushin, S.A. Alyushin, “Electronic neural net design methodology,” Proceedings of the 5th. Int. conf and Exibition “Digital signal processing and its applications”, Moscow, Russia, March 12-14, 2003, pp. 585-587.

[19] A.V. Alyushin, S.A. Alyushin, V.G. Arhangelsky, “Scalable processor core for high-speed pattern matching architecture on FPGA,” Proceedings of The Third Int. Conf. on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC2016), Higher Scool of Economics (National Research University), Moscow, Russia, July 06-08, 2016, pp. 148-153.

[20] A.V. Alyushin, S.A. Alyushin, V.G. Arhangelsky, “High-speed pattern matching arhitecture on limited connectivity FPGA,” Proceedings of the 11-th Int. Conf. “Application of information and communication technologies”, AICT2017, Moscow, Russia, 20-22 September, Vol. 1, pp. 57-62.

[21] A.V. Alyushin, S.A. Alyushin, V.G. Arhangelsky, “Electrical activity signal spectrum of the artificial neyral net on the base of pulsed neurons and memristors,” To be published in the Proceedings of the “Conference of Russian Young Researchers in Electrical and Electronic Engineering ElConRus-2018”, Moscow, Russia, 29 january – 1 february 2018.

[22] A.V. Alyushin, S.A. Alyushin, V.G. Arhangelsky, “Bit-vector pattern matching systems on the base of high bandwidth FPGA memory,” To be published in the Proceedings of the “Conference of Russian Young Researchers in Electrical and Electronic Engineering ElConRus-2018”, Moscow, Russia, 29 january – 1 february 2018.