Speech Signal Processing using Pulse Coupled Neural Network

M. Jurečka; B. Šarkan
Department of Technical Cybernetics - Faculty of Management Science and Informatics, Department of Road and Urban Transport - Faculty of Operation and Economics of Transport and Communications
University of Žilina
Žilina, Slovakia
matus.jurecka@fri.uniza.sk, branislav.sarkan@fpedas.uniza.sk

Abstract—**This paper presents results obtained from the analysis of Pulse Coupled Neural Network (PCNN) behavior in speech signal processing. Authentic PCNN model inspired by neurobiology was initially used for the image recognition [3] problems such as automatic target recognition in military systems. This original PCNN model was modified for 1--dimensional audio signals processing represented by sequence of PCM samples. At first we attempted to express the system of difference equation describing PCNN behavior in more general closed form. The closed form expression of difference equation is often advisable in the analysis of its behavior and can bring to light such properties of the system as periodicity and stability. Unfortunately, the non-linearity in the system of difference equations describing PCNN behavior prevents us from closed form expression formulation. Then PCNN behavior was analyzed using simulation methods exploring PCNN generated features changes depending on different input signals. It has been shown that PCNN generated features appear to be a very good speech signal representation for speaker identification applications.**

Keywords-component: Pulse Coupled Neural Network; Speech Signal Processing;Speaker Identification;

I. INTRODUCTION *(HEADING 1)*

All manuscripts must be in English. These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. In general, speech signal processing includes wide set of methods such as filtration, boosting, feature extraction...etc. Speech acquisition begins with a person speaking into a microphone. Speech signal is converted onto digital form using pulse coded modulation (PCM). This means of speech signal representation is rather bit rate demanding and hence it is not so suitable for pattern recognition. However, it can be represented by a limited set of features. There are several methods available for features extraction and dimension reduction. The goal of dimension reduction is to obtain significant features for the unique pattern representation. The dimension reduction is a transformation of an input signal space into a feature space with a lower dimension. Classical methods of features extraction in digital signal processing for speech recognition include coefficients of Fourier transform, Linear Predictive Coefficients (LPC) or derived Cepstral Coefficients. [5]. The method that is scoring a growing interest for a dimension reduction & feature extraction in a field of image processing is Pulse Coupled Neural Network (PCNN) [1].

II. THE STRUCTURE OF PCNN

The structure of standard PCNN comes out from the structure of input pattern, which will be processed. Let us consider that the input pattern is a matrix of values for an input isolated word. PCNN is a single layered, two-dimensional, laterally connected neural network of pulse coupled neurons connecting each other with values of input matrix. The standard PCNN model is described as iteration by the following equations:

$$
\begin{aligned}
F_{ij}(n) &= S_{ij} + F_{ij}(n-1)\cdot e^{-\alpha F} + V_F\cdot(M * Y(n-1))_{ij}\\
L_{ij}(n) &= L_{ij}(n-1)\cdot e^{-\alpha L} + V_L\cdot(W * Y(n-1))_{ij}\\
U_{ij}(n) &= F_{ij}(n)\cdot(1+\beta\cdot L_{ij}(n))\\
\Theta_{ij}(n) &= \Theta_{ij}(n-1)\cdot e^{-\alpha\Theta} + V_\Theta\cdot Y(n-1)\\
Y_{ij}(n) &= \begin{cases} 1 & \text{if } U_{ij} > \Theta_{ij} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{1}
$$

Where F_{ij} is the feeding input, L_{ij} is the linking input, n is an iteration step, S_{ij} is a value at i,j coordinates in the input matrix. W and M are the weight matrices, $*$ is the convolution operator, Y_{ij} is the output of the neuron at i,j coordinates, V_L and V_F are potentials, α_L and α_F are decayed constants.

Single signals of the linking input are biased and then multiplied together. Next, input values F_{ij} , L_{ij} are modulated in linking part of neuron. We also obtain internal activity of neuron U_{ij} in the specific iteration step. If internal activity is greater than dynamic threshold Θ_{ij}, then neuron generates output pulse. Otherwise output equals to zero. The neuron output Y_{ij} does not necessarily need to be binary. It is possible to use a sigmoid pulse generator, where the neuron takes the analogue value from 0 to 1. The input matrix is through PCNN transformed into a sequence of temporary binary matrixes. Each of these binary matrixes has the same dimension as input matrix. The sum of all activities in specific iteration step gives one value, which representing one feature for the classification. If we have N iteration steps, we obtain N features. The one-dimensional time signal, generated from the values of output matrix $Y_{ij}(n)$ in every iteration step n can be defined as follows:

$$G(n) = \sum_{i,j} Y_{ij}(n) \qquad (2)$$

Significant advantage of PCNN which is useful mainly in image recognition is the invariance of generated time signal to rotation, dilatation or translation of images [1]. Therefore PCNN is advisable for the feature generation and pattern recognition in the classification tasks using conventional neural networks or other methods. It is evident that PCNN is not the neural network in the term of classification. It is only a mean of feature extraction for pattern classification using conventional neural network models, like that of multi-layer perceptron. Several models of PCNN have been developed. The most used PCNN models are, for example, PCNN with modified feeding input [2], fast-linking PCNN [4],[6]or feedback PCNN [3]. PCNN model is also advisable in dealing with image processing problems such as vehicle classification. [11] The above-mentioned PCNN model was designed to process grayscale image data which are naturally 2-dimensional with unipolar (positive) pixel values.

For the needs of 1--dimensional bipolar PCM speech signal processing PCNN was reduced to 1--dimensional structure in which the amount of PCNN neurons is equal to the amount of PCM samples. The weight matrices W and M were reduced to 1--dimensional structures, too. This also helped to decrease convolution computation complexity. The weight vectors were formed using Gaussian function as follows:

$$w(n) = m(n) = \frac{e^{-\frac{(n-\mu)^2}{2\sigma^2}}}{\omega\sqrt{2\pi}} \qquad (3)$$

where $\sigma = 0.9$, $\mu = r_0/2$ and length of the vectors $r_0=7$

PCNN parameters were set as follows: $\alpha_F = 17.3$, $\alpha_L = 0.03$, $\alpha_\theta = 0.03$, $V_F = 3$, $V_L = 0.1$, $V_\theta = 0.3$, $\beta = 0.4$, $r_0 = 7$, $N = 200$.

This modified PCNN network was subject to the experiments described below.

III. Analytical Means of Exploration

Modified PCNN can be described by the following system of difference equations:

$$F(k,n) = S(k) + F(k,n-1)\cdot e^{-\alpha F} + V_F(M * Y(k,n-1)$$

$$L(k,n) = L(k,n-1)\cdot e^{-\alpha L} + V_L(M * Y(k,n-1)$$

$$U(k,n) = F(k,n)\cdot(1+\beta\cdot L(k,n))$$

$$\Theta(k,n) = \Theta(k,n-1)\cdot e^{-\alpha\Theta} + V_\Theta\cdot Y(k,n-1)$$

$$Y(k,n) = \begin{cases} 1 & \text{if } U(k,n) > \Theta(k,n) \\ 0 & \text{otherwise} \end{cases}$$

(4)

where n is the time dimension and k is describing position of PCM sample. For the needs of detailed behavior analysis of such difference equation system it would be ideal to express it in closed form, i.e. to get rid of recursions and to express output $Y(k,n)$ as a function of time n, space k and input signal $s(k)$.

$$Y(n,k) = f(n,k,S(k)) \qquad (5)$$

Generated features can be computed be the summation of function f along space dimension k.

Unfortunately, the non-linearity in the system of difference equations (4) does not enable to use closed form expression formulation as it can be shown on logistic

difference equation which can often act chaotically [10].

IV. Simulation Means of Exploration

In this section the PCNN output $G(n)$ is being analyzed for various input signals. PCNN parameters were set as follows: $\alpha_F = 17.3$, $\alpha_L = 0.03$, $\alpha_\theta = 0.03$, $V_F = 3$, $V_L = 0.1$, $V_\theta = 0.3$, $\beta = 0.4$, $r_0 = 7$, $N = 200$. Weight vectors were formed using Gaussian functions as described in equation (3). At first PCNN input signals were harmonic signals with different amount of periods (Fig. 1), different sampling frequency (Fig. 2) and different phase (Fig. 3).

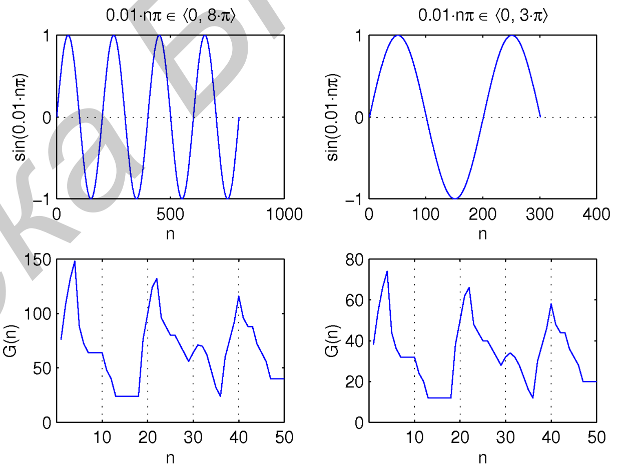

Figure 1. PCNN output *G(n)* dependence on the number of input signal periods

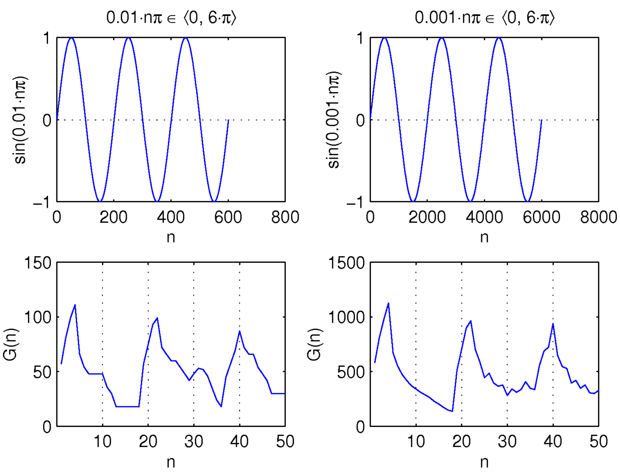

Figure 2. PCNN output *G(n)* dependence on sampling frequency of input signal

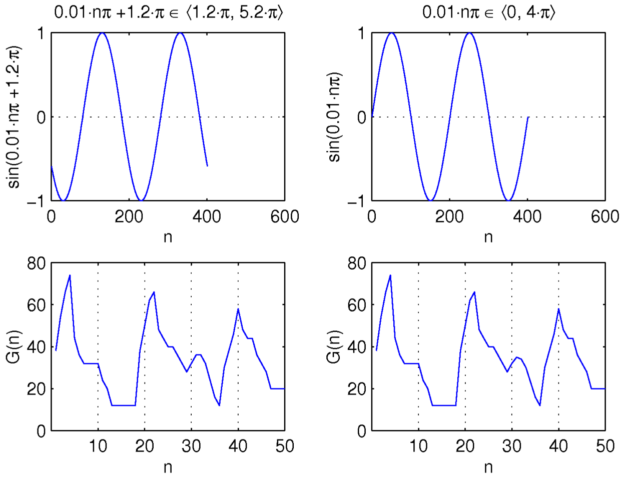

Figure 3. PCNN output *G(n)* dependence on the phase of input signal

While the length of input signal has influence on the DC shift of *G(n)* (due to different sizes of binary vectors Y and hence different amount of ”ones” in these vectors) it is quite reasonable to normalize of *G(n)* to range < 0, 1 >.

Further experiments were performed on signals obtained as an output from LPC model of speech. PCNN output was analyzed for different pitch periods of input signal (Fig. 4) and different coefficients of the LPC filter (Fig. 5).

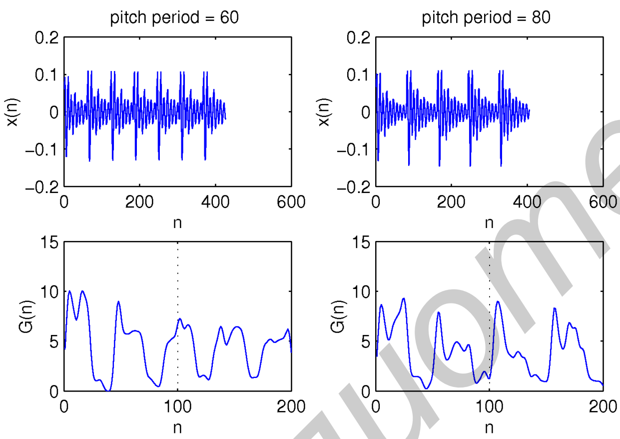

Figure 4. PCNN output *G(n)* dependence on different input signal pitch period $T_0=60$ and $T_0=80$

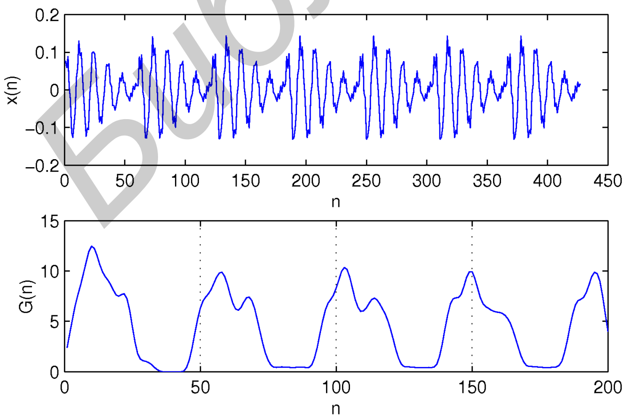

Figure 5. PCNN output *G(n)* dependence on changes in LPC of input signal for the pitch period of $T_0=60$

The above mentioned figures clearly show that the amount of periods and different phases does not change normalized output *G(n)*. On the other hand, sampling frequency does, but the "period" of *G(n)* is maintained. Periodicity of PCNN output was presented in details in [7].

Experiments with LPC model show that normalized PCNN output *G(n)* is not maintained. *G(n)* "period" is maintained only if the pitch period is not changed. It seems that growing pitch period causes growing *G(n)* "period".

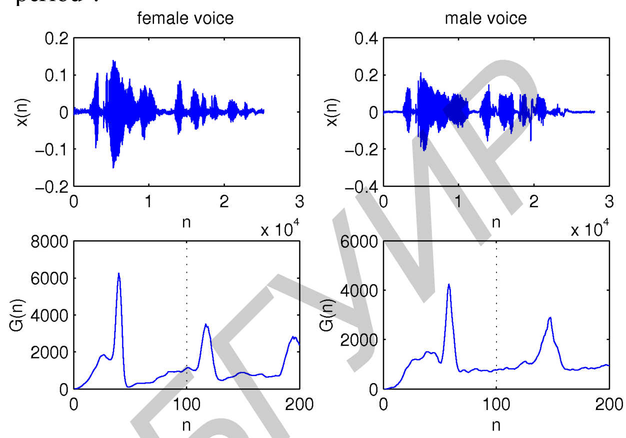

Figure 6. Impact of speaker's gender on PCNN output *G(n)*

This hypothesis was tested in the experiment with male and female voices. In general, male voice has a greater pitch period than female voice. Influence of the speaker's gender on *G(n)* is shown in Fig. 6. This Figure shows that *G(n)* "period" for female voice is smaller than *G(n)* "period" for the male voice.

Tab.1 summarizes influence of different PCNN input signals on its output.

TABLE I. INFLUENCE OF DIFFERENT PCNN INPUT SIGNALS ON ITS OUTPUT *G(N)*

input signal change	**normalized G(n) identical**	**preserved G(n) period**
phase	yes	yes
number of periods	yes	yes
sampling frequency	no	yes
pitch period	no	no
LPC	no	yes
spealers gender	no	no

Fig.(7) shows normalized *G(n)* for 25 male and 18 female speech samples. The speech sample provided by every speaker consisted of one line from Slovak poem.

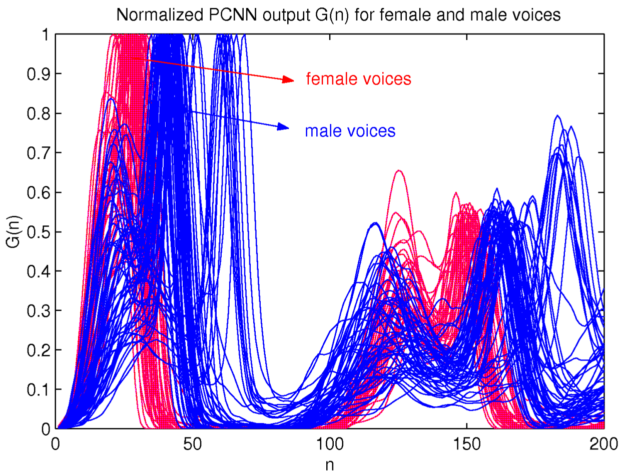

Figure 7. Normalized PCNN output *G(n)* for 25 male and 18 female speech samples

V. CONCLUSION

It seems that PCNN output is sensitive to the change in excitation (drive) signal of the LPC speech production model. In real speech signals this excitation signal change could be represented for example by pitch period change. In general, pitch period T_0 or fundamental frequency f_0 is one of the most important parameters in the systems for speaker identification followed by other formant frequencies f_1, f_2, f_3...[9] Although pitch period is in general different in single voiced phonemes, average pitch period depends mainly on voice color of the speaker. This could be the explanation why speaker recognition systems described in [8] are giving so promising recognition results.

ACKNOWLEDGMENT

This work has been supported by FRI Institutional Grant 2012.

REFERENCES

[1] R. Forgáč, I. Mokriš, "Umelé neurónové siete na redukciu dimenzie priestoru príznakov a klasifikáciu," Univerzita Mateja Bela, Banská Bystrica 2002. ISBN 80-8055-743-8J.

[2] G. Frank, G. Hartman, "An artificial Neural Network Accelerator for Pulse-Coupled Model Neurons, Proc. of International Conference on Neural Networks, Vol. 4, Perth 1995.

[3] J.L. Johnson, M.L. Padget, "PCNN Models and Applications," IEEE Transaction on Neural Networks, Vol. 2, No.3, 1999, pp. 480–498

[4] J.M. Kiser, J.L. Johnson, "Implementation of Pulse-Coupled Neural Networks in the CNAPS Enviroment," IEEE Transactions on Neural Networks, Vol. 10, No. 3, 1999, pp. 584–590

[5] J.G. Proakis, D.G. Manolakis, "Digital Signal Processing," MPC, New York 1992..

[6] P.Ševčík, "Implementation Design of Poulse Coupled Neural Network Neuron Into Field Programmable Gate Array Device," Proc. of International Conference on Applied Electronics, Plzeň 2006.

[7] M. Jurečka, "Pulse Couple Neural Network Parameter Settings Recomendations," Journal of Information, Control and Management Systems Vol.9, No.3, Žilina 2010.

[8] M. Jurečka, "Speaker Identification Using Pulse Couple Neural Network," Proc. of International Conference MEMICS, Znojmo 2008, pp 86–90.

[9] J.M. Gutiéerez-Arriola, "Analysis of Parameter Importance in Speaker Identity," Madrid: Universidad Politecnica de Madrid, Spain, 2003, [online] http://www-gth.die.upm.es/˜cordoba/Papers/AI-82Ana-03.doc

[10] R. M. May,"Simple mathematical models with very complicated dynamics," 1976, Nature. 261: 459-467

[11] O. Karpiš, "System for Vehicles Classification and Emergency Vehicles Detection" 11th IFAC Workshop on Programmable Devices and Embedded Systems PDeS 2012, 23-25.5.2012, pp.155-159, Brno