JOURNAL OF MEDICAL INFORMATICS & TECHNOLOGIES Vol. 7/2004, ISSN 1642-6037

neural network training, gait analysis, gait disturbances, hemiparesis diagnosis, Parkinson's disease diagnosis

Sławomir CHANDZLIK*, Jan PIECHA*,**

THE INTERFERENCE SPECTRUM EXTRACTION OF A GAIT CHARACTERISTICS DATA RECORD

The paper shows several aspects of the gait data record analysis describing neurological diseases. The diagnosis of the gait abnormalities concerns interferences level of the patient physiological records. The disease source and level can be classified by the relevant interference functions. These functions were used for artificial records creation to multiply the necessary set of data needed for neural network training.

1. INTRODUCTION

Effectiveness of medical diagnosis depends on equipment quality that a doctor has for his disposal. Various computer systems provide the operator precise measures not only classifying the disease but (what is more important) describing the abnormality level. Numbers of works done by the paper authors discuss the gait characteristics of patients suffering from neurological diseases [1,2,3,5]. These works were based on Parotec System for Windows (PSW) applications [4,5,8,9]. The PSW is used to register the pressure distribution on a foot describing a foot shape for orthopaedic purposes [4,8,9]. These elementary options were widely described in several earlier works [2,4,5,6,8,9]. The PSW system was also provided with extensions for neurological diseases classification, discussed already as well [1,2,5,6].

The present paper shows our recent investigations that allow avoiding very difficult condition for neural network training process – the developer needs large number well classified records used for running the artificial conclusion making system.

Majority of experiments we have done for two groups: left- and right- lateral hemiparesis and Parkinson disease. In the control group several regularities were observed then extracted. These regularities were defined by math-formulas (the disturbance gait functions) and used as an interference spectrum of the physiological gait record. The defined formulas allow producing virtual data records on the basis of data records determined as physiological.

The virtual data record generator produced above 40 thousand of virtual data records on a basis of 92 clinical measurements. The concatenation of these functions also permits to obtain new distribution of the gait.

* University of Silesia, Institute of Informatics, Dept. of Computer Systems, Sosnowiec, Poland
** Silesian University of Technology, Faculty of Transport, Department of Transport Informatics, Katowice, Poland

Also the Parotec System for Windows (PSW) was elaborated for orthopaedic diseases classification and many valuable data can be extracted from its record.

2. THE INTERFERENCES DEFINITIONS

An aim of neural network training is finding a global minimum of the cost function. This process is realized by determining adequate M-dimension vector of weights W on the basis a number of parameters of the neural network.

The minimum of the function is found as faster as the training is better controlled (is not running chaotically) [6,7]. It also depends on complexity of the task of the neural network, that has to classify, and from the network topology. Relatively small values of teaching factor η are being applied during teaching process for that reason. The N–dimension training sequence has to be given on the inputs of the neural network because the weights vector W was determined correctly at so low value of the teaching factor η. Number of N, how is resulting from experience and examples given in literature [2,5,6], is approximately from a few thousands to a few millions.

Various experiments with the neural network selection were carried out on the basis of four sets of records: the control group of patients, the group with left-lateral hemiparesis, the right-lateral hemiparesis and records describing as Parkinson's disease.

The almost one hundred cases for putting the diagnosis has been assigned. These clinical records were divided into a following groups: 25 cases of the control group, 29 records classified by medical experts as the left-lateral hemiparesis, 28 cases for the right-lateral hemiparesis and 10 records concerning the Parkinson's disease.

It is obvious that this number of records is not sufficient for carrying any optimisation of neural network training procedures [1,6,7]. Cycling repetition of these records within the process of training causes the neural network "learning by heart". That is why for increasing the size N of the training data large number of virtual records has to be somehow produced.

a) b)

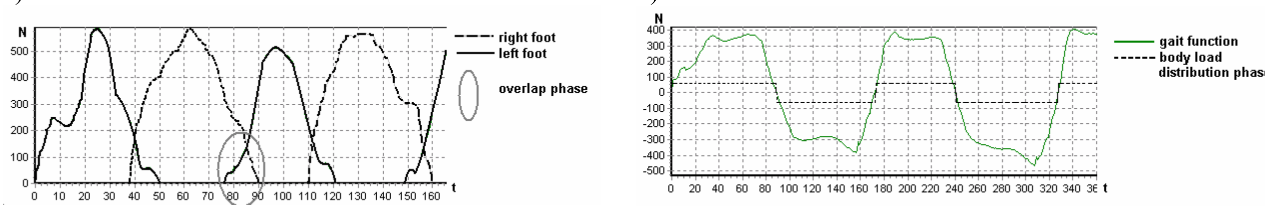

Fig. 1. The forces time distribution: a) the functions of forces F, b) the gait function W.

Below one can find several formalities explaining how the virtual records are produced.

Definition 1.

A step has to be understood as a period of time while a patient's foot (left or right) touches a floor.

Definition 2.

A steps-cycle is a time between a left step (or right) beginning and a right step (or left) ends if the left step (or right) is finished. It defines the observation cycle for only one left-foot step and only one right-foot step within the single steps-cycle.

Definition 3.

The overlap phase is a time period of parallel floor contact of both feet in the dynamic part (in a walking time) of the data. It is a time when the body weight is totally moved from one of the foot to the other one.

Let us assume that the value of α defines an active foot:

- for left foot $\alpha = \mathrm{l}$
- for right foot $\alpha = \mathrm{r}$.

Let us also determine the force measured on an α-foot in a current *i* steps-cycle as:

$$F(\alpha,t) = \sum_{i=1}^{n_\alpha} F_i(t) = \sum_{i=1}^{n_\alpha} P_i(t) \cdot S_i \qquad (1)$$

where: n_α – is a number of sensors installed on an insole (of the α-foot),
$F_i(t)$ – determines the force recorded in a time t on each sensor i,
$P_i(t)$ – determines a pressure value in a time t on each sensor i,
S_i – describes the hydrocell surface of every sensor i.

The time distribution of these forces (on a footprint) has been presented by functions *F* in Fig. 1a. Similarly the gait function *W* (Fig. 1b) can be defined as:

$$W(t) = F(l,t) - F(r,t) \qquad (2)$$

The positive values of the gait function *W* determine a gravity centre of the patient's body movement into the left side of the body. The negative values of *W* determine the overload on a right foot where the gravity centre moved into the right side of the body.

A dynamic part of the data record contains samples of pressure registered during the gait time. For this data part the functions F_i (1) registered on the sensors were defined by a spline interpolation method that produces continuous functions F_i – widely described in the paper [3]. Thanks to this approximation, the functions F_i have been continuous and they have been realizing a representation:

$$F_i : [0, T_D] \rightarrow \Re^+ \cup \{0\} \qquad (3)$$

where: T_D is a time period of a dynamic part of the measurement.

The virtual data records can be produced in the case the interferences of the pathological records are recognised. Then multiplying the clinical cases into well-defined classes can cover the needs of large number of training data set. The function *F* representing distribution of forces at a patient’s foot was determined on the basis of control group.

Let *A* denotes the gait function of the virtual record, as:

$$A(t) = W(t) + E(t) \tag{4}$$

where: W – is a gait function given by formula (2) for a record of the control group, E – represents a gait disturbances function defined as:

$$E(t) = W_1(t) - W_2(t) \tag{5}$$

where: W_i – defines the gait functions obtained from formula (2) from two clinical data record R_1 and R_2.

In the case the numbers of steps-cycles m_1, m_2 are different then they have to be reduced into the same size – into smaller number of steps-cycles $m = \min\{m_1, m_2\}$. Moreover functions W_1 and W_2 operate on these sets:

$$\begin{aligned} W_1 &: [0, T_{D1}] \to \Re \\ W_2 &: [0, T_{D2}] \to \Re \end{aligned} \tag{6}$$

where: T_{D1}, T_{D2} concern the time markers of the dynamic units for R_1 and R_2 data records respectively for m number of steps-cycles.

The values of T_{D1} and T_{D2} should be additionally equal. This condition enable that the $W_i(t)$ functions are defined for every time unit $t \in T_{DD}$ and the T_{DD} set is defined as:

$$T_{DD} = [0, T_{D1}] \tag{7}$$

Then the time T_{D2} has to be redefined by k_T factor, as:

$$k_T = \frac{T_{D1}}{T_{D2}} \tag{8}$$

and an adaptation time T_{D2} period is defined by formula:

$$T_{D2}^{'} = k_T T_{D2} \textit{ for m steps-cycles.} \tag{9}$$

After this operations $T_{D1} = T_{D2}^{'}$ are equal and both functions W_1 and W_2 are defined for a whole T_{DD} time period. What is more the values of disturbances functions E respond to the real data of the data record.

3. THE INTERFERENCE SPECTRUM EXTRACTION

Let us compare two data records R_w and R_p. Let us also assume that R_w record represents clinical case defined as physiological (the control group record) and R_p record is a pathological case (outside the control group).

For these cases the gait functions W_w and W_p are determined in accordance with the formulas presented above (illustrated in Fig. 2 a, b). The interferences of the gait physiology are obtained as:

$$E(t) = W_p(t) - W_w(t) \tag{10}$$

The above E function shows the character of the gait data spectrum interferences. This function of interferences is then used for multiplying the records with the extracted character of interferences; here with the left-lateral hemiparesis.

The example of the E expression for a left-lateral hemiparesis is presented in Fig. 2c.

a) b) c)

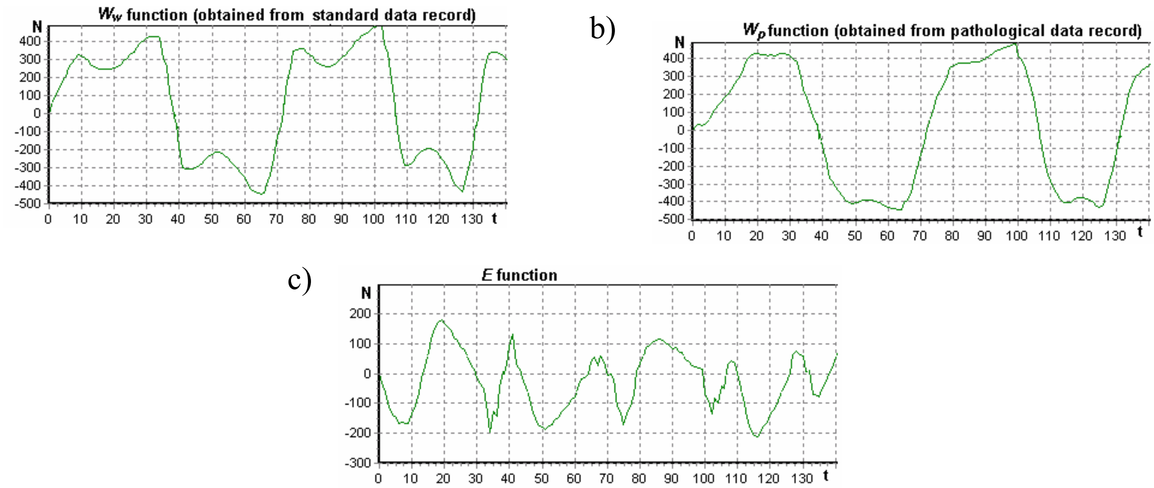

Fig. 2. Interferences of physiological records: a) the gait function W_w of a physiological record, b) the gait function W_p of a pathological record, c) interferences E.

4. THE EXPERIMENTS DISCUSSION

The gait interferences defined by function E is determined for each sensor of the insole. The same operation is done for every data record. In table 1 four classes of a gait abnormality have been determined – defined as *.efz files, with virtual products of the data record.

Table 1. Number of virtual *.efz files

	The clinical records			
	of control group	of hemiparesis		of Parkinson’s disease
		left-lateral	right-lateral	
	25	29	28	10
The virtual data products; *.efz files for four disease classes	625	725	700	250

The function E of the gait interferences are used to produce the *.efz files, however the gait functions W are determined from any clinical data record of the control group.

4.1. AN OVERLAP PHASE RECOGNITION

An overlap phase is one of the characteristic features of the patient's gait that has to be determined in the set of virtual records (def. 3).

The overlap phase number p is defined by formula:

$$p = 2m - 1 \tag{11}$$

where: m – is a number of steps-cycles.

There is possible to obtain approximations of distribution of the force function F_O, that appears at a patient's foot during the overlap phase in the virtual data record, directly from the gait function A given by formula (4). The following formula determines discussed approximation of force distribution that has to be assigned to every of overlap phases i:

$$F_{Oi}(\alpha, t) = A(t) - \beta_\alpha(t) \tag{12}$$

where: A – is a gait function generated for the virtual data record,
β – is an approximation factor defined by:

$$\begin{aligned} \beta_l(t) &= m_o - m_o\left(1 - f_2(t)\right) f_1(t) \\ \beta_r(t) &= M_o - M_o\left(1 - f_2(t)\right)\left(1 - f_1(t)\right) \end{aligned} \tag{13}$$

where: $m_o = \min\limits_{t=[0;t_{Oi}]}(A(t))$ – is a minimal value of the function A of i overlap phase,
$M_o = \max\limits_{t=[0;t_{Oi}]}(A(t))$ – is a maximum value of the function A of i of the overlap phase,
f_1 – is a linear function determining monotonic character of the function E, given by formula:

$$f_1(t) = \begin{cases} g_1(t), & if\ A(t_{SOi}) < A(t_{SOi} + t_{Oi}) \\ 1 - g_1(t), & if\ A(t_{SOi}) > A(t_{SOi} + t_{Oi}) \end{cases} \tag{14}$$

where: $g_1(t) = \dfrac{t}{t_{Oi}}$
t_{Oi} – is a time duration of overlap phase i,
t_{SOi} – is a time overlap phase i marker, where the overlap phase begin,
f_2 – is a function of correction defined by formula:
$f_2(t) = c_G \sin\left(\Pi g_1(t)\right)$ where: c_G – is a scaling constant.

The F_{Oi} functions are assigned at ranges of $[t_{SOi}; t_{SOi} + t_{Oi}]$, i.e. only for time markers where the overlap phase exists. The example forces distribution on feet in time period of the overlap phase and its approximations by F_{Oi} functions are presented in Fig. 3a.

4.2. THE OVERLAP APPROXIMATION ERROR ANALYSIS

For the virtual overlap estimation the error analysis has been carried out first. This makes the optimisation of the function β coefficient possible. For all experiments the relative error has been defined:

$$R_E(t) = \frac{\left|F_O(l,t) - F_O(r,t) - \left(F(l,t) + F(r,t)\right)\right|}{F(l,t) + F(r,t)} \tag{15}$$

where: F_O – concerns estimated values of approximated forces on a foot distribution,
F – concerns real values of the forces on a foot.

With this error definition its measure can be defined, as:

$$R_i = \int_{t_{SOi}}^{t_{SOi}+t_{Oi}} R_E(t)dt \tag{16}$$

where: t_{Oi}, t_{SOi} – are described as in the above formula (14).

The 50 clinical records have been discussed carefully, where the estimation error was the analysis subject. The smallest value of the measure R was obtained for coefficient c_G = 0.22. With the average value of the R measure for this coefficient c_G was $\overline{R} = 1.7704$ (compare with the Fig. 3b).

a) b)

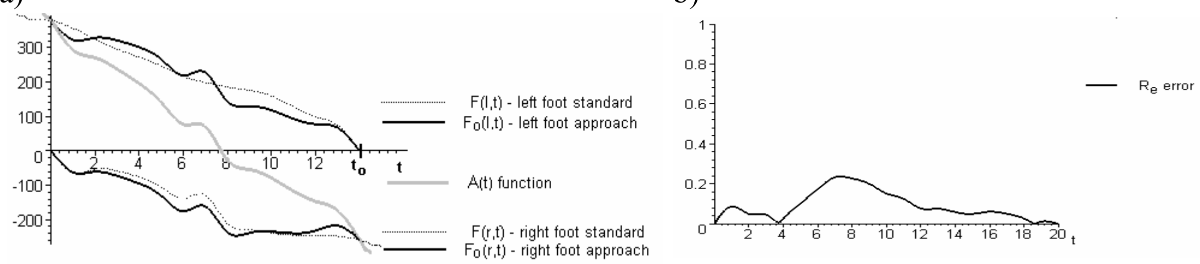

Fig. 3. a) The overlap functions A, b) an error R_E

4.3. THE DISTRIBUTION OF FORCES IN VIRTUAL RECORDS

An algorithm describes the forces distribution for the virtual record describing the patient's foot load in the dynamic part of the data:

- the gait function W extraction from a pattern record as in Fig. 2a,
- normalisation of a time duration T_{Dw} of the dynamic part of the pattern record using the scaling coefficient k_T,
- the interference function E (Fig. 2c) reading from a *.efz file,
- the gait function A extraction from the virtual record as a product of functions concatenation E z W (equation 4 and Fig. 1b),
- extraction of the forces F_O occur on an overlap of the feet using the gait function A,
- absolute values of forces, representing pressure recorded on a right foot.

For virtual product definition one file *.efz and one pattern file is needed. This way a large number of addition records can be obtained (table 2).

Table 2. The virtual record database structure

Number of clinical records			
control group	hemiparesis		Parkinson's disease
	left-lateral	right-lateral	
25	29	28	10
Number of virtual records			
control group	hemiparesis		Parkinson's disease
	left-lateral	right-lateral	
15 625	18 125	17 500	6 250

5. CONCLUSIONS

The discussed method of forces distribution was evaluated for the dynamic part of the record. Anyhow the same methodology can be used for static part of the record, although several modifications are needed. For example the gait functions are defined for left and right foot separately, similarly as interference functions. Also overlap phases do not exists for static units of the data.

The interferences concern duration of the gait time, the amplitude of interference functions E and the amplitude of the gait function W for a pattern clinical record. The only limits for virtual records production concern the range of the disease classes. These virtual records (first properly classified) have been used for the conclusion-making unit training [2].

In the conclusion making unit examination of the automatic classification of current records have been done for: left- and right- lateral hemiparesis, Parkinson's disease and control group.

BIOBLIOGRAPHY

[1] CHANDZLIK S., KOPICERA K.: Experiments with neural network parameters – selection for Foot abnormalities Recognition, Journal of Medical Informatics & Technologies. Vol. 5, pp: CS-71 – CS-78. ISBN 83-909517-2-7, 2000.

[2] CHANDZLIK S., PIECHA J.: The body balance measures for neurological disease estimation and classification. Journal of Medical Informatics & Technologies. Vol. 6, pp: IT-87 – IT-94. ISSN 1642-6037, 2003.

[3] CHANDZLIK S., PIECHA J.: A patient walk-data-record modelling using a spline interpolation method. Journal of Medical Informatics & Technologies. Vol. 3, pp: MIT-153 – MIT-160. ISSN 1642-6037, 2002.

[4] KOPICERA K., PIECHA J.: The fuzzy estimation unit of foot-print abnormality recognition. Journal of Medical Informatics & Technologies. Vol. 2, pp: MI183 – MI188. ISSN 1642-6037, 2001.

[5] KOPICERA K., PIECHA J., ZYGUŁA J.: The neural networks in diagnostics support for PSW system. Proc. of Int. Conference ASIS'99, pp. 113-118, Krnov, 1999.

[6] PIECHA J.: The neural network conclusion-making system for foot abnormality recognition. Proceedings of IMACS World Congress, Lausanne, Switzerland, August 2000.

[7] PIECHA J.: The neutral network selection for a medical diagnostic system using an artificial data set. Journal of Computing and Information Technology CIT, Vol.9, pp. 123-132, ISSN: 1330-1136, 2001.

[8] PIECHA J., ZYGUŁA J., ŁYCZAK J., GAŹDZIK T., PROKSA J.: The advanced measuring system for orthopaedic pathologies diagnostics using a static and dynamic footprints, Chirurgia narządów ruchu i ortopedia polska vol. LXI 1996, suplement 3B, pp.119-124. (in Polish)

[9] ZBROJKIEWICZ J., PIECHA J,. JARZĄBEK-STĘPNIAK A.: The gait pattern detection in PSW records of the acute sciatic neuralgia. Proc. on KOSYR'01, pp. 29-36, ISBN 83-911675-2-6, 2001.

The work is supported by Committee for Scientific Research funds, KBN Grant No. 3 T11E 055 26 in 2004/2005.