

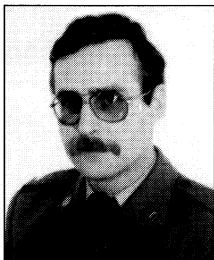
Jacek JAKUBOWSKI, Krzysztof KWIATOS

WOJSKOWA AKADEMIA TECHNICZNA,
INSTYTUT PODSTAW ELEKTRONIKI

Zastosowanie statystyk wyższych rzędów w procesie różnicowania sygnałów pochodzących od drżeń kończyn człowieka

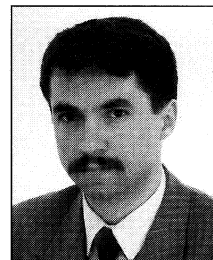
Mgr inż. Krzysztof KWIATOS

– absolwent Wydziału Elektroniki Wojskowej Akademii Technicznej z 1986 r. Zatrudniony w Zakładzie Teorii Obwodów i Sygnałów Elektrycznych Instytutu Podstaw Elektroniki WAT na stanowisku asystenta naukowo-dydaktycznego. Obszar zainteresowań: metrologia biomedyczna.



Mgr inż. Jacek JAKUBOWSKI

– absolwent Wydziału Elektroniki Wojskowej Akademii Technicznej z 1993 r. Od 1995 r. pracuje jako asystent naukowo-dydaktyczny w Zakładzie Miernictwa Elektronicznego Instytutu Podstaw Elektroniki WAT. Obszar jego zainteresowań to wykorzystanie statystyk wyższych rzędów w przetwarzaniu sygnałów pomiarowych.



Streszczenie

Referat przedstawia próbę rozróżnienia najczęściej występujących trzech rodzajów drżeń rąk człowieka polegającą na wykonaniu pomiarów akcelerometrycznych i wyliczeniu dla uzyskanych w ten sposób szeregów czasowych zbioru cech je charakteryzujących. Cechy te zostały wygenerowane za pomocą różnych opisów matematycznych bazujących na statystyce zarówno II jak i wyższych rzędów. Wykorzystanie prostego klasyfikatora neuronowego umożliwiło uzyskanie błędę klasyfikacji drżeń na poziomie 3.6%.

Abstract

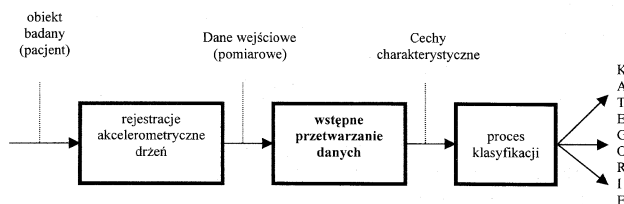
The paper presents an attempt to separate the most often observed three types of human tremor with objective techniques by measuring the acceleration of the hand tremor and by calculating some characteristic features of these time series. Different mathematical descriptions have been adopted in order to generate the features. They have come from both second and higher-order statistics. A simple neural classifier has enabled the separation of these kinds of tremor with 3.6% error rate.

Wprowadzenie

Drżenie rąk jest zjawiskiem występującym zarówno w populacjach ludzi zdrowych jak i w populacjach wykazujących zmiany neurologiczne. Wśród przypadków patologicznych stosunkowo najczęściej spotykanymi schorzeniami są choroba Parkinsona i drżenie samoistne. W praktyce klinicznej postawienie diagnozy w jak najwcześniejszym stadium choroby jest niezwykle cenne, gdyż umożliwia przejście na odpowiedni sposób leczenia warunkujący skuteczność podawanych środków farmakologicznych. Z reguły diagnoza bazuje na bezpośrednim obrazie klinicznym i wywiadzie chorobowym. Niestety – brak pełnej obiektywności oceny drżenia spowodowany podobieństwem cech klinicznych, stanowią dużą trudność dla badacza próbującego wyodrębnić istotne dla niego informacje.

Współczesna neurologia zgłasza silne zapotrzebowanie na narzędzie pozwalające jednoznacznie ocenić lub też wesprzeć proces różnicowania występujących patologii. Jedną z wykorzystywanych powszechnie metod obiektywnego zbierania informacji o drżeniach jest ich rejestracja akcelerometryczna [1]. Wówczas diagnozowanie (lub raczej jego wspomaganie) sprowadzi się do klasyfikacji szeregów czasowych drżeń tj. do podjęcia decyzji dotyczącej przypisania zbioru pewnych ich cech do jednej z kilku określonych kategorii. Wobec zmienności drżeń w obrębie tej samej kategorii

(tej samej jednostki chorobowej) konieczne jest wyodrębnienie spośród analizowanych danych zbioru cech charakterystycznych dla rozpatrywanych chorób i znacznie mniej nadmiarowych niż dane pierwotnie zebrane w procesie akwizycji. Operacji tej dokonuje się w ramach tzw. wstępnego przetwarzania danych wejściowych które, po zastosowaniu pewnej reguły klasyfikacyjnej, stanowi podstawę decyzji o przypisaniu drżenia do określonej kategorii – rys.1.



Rys. 1. Struktura blokowa procesu klasyfikacji drżeń

Czynione w ostatnich latach próby obiektywnego rozróżnienia występujących jednostek chorobowych polegające na zastosowaniu do opisu zarejestrowanych szeregów czasowych drżeń widmowej gęstości mocy nie zapewniają zadowalających rezultatów [2]. Okazuje się, iż o ile w przypadku konieczności rozróżnienia pomiędzy silną patologią a drżeniem fizjologicznym dostatecznym i wygodnym kryterium jest poziom samego drżenia, o tyle dyskryminacja pomiędzy różnymi rodzajami drżeń patologicznych oraz pomiędzy słabą patologią a drżeniem fizjologicznym napotyka na duże trudności. Ani częstotliwość drżenia, ani jego poziom, określane na podstawie estymowanej widmowej gęstości mocy nie są wskaźnikami wiarygodnymi pozwalającymi jednoznacznie stwierdzić jaki rodzaj drżenia występuje w danym konkretnym przypadku [1-3]. Powstają więc tutaj zasadnicze pytania: czy nie byłoby celowe rozwiązanie parametrów opisujących widmową gęstość mocy innych niż dwa powyżej wspomniane i czy w ogóle widmowa gęstość mocy jest narzędziem wystarczającym do tego, aby opisywać i różnicować sygnały pochodzące od tak złożonego obiektu biologicznego, jakim jest człowiek. Widmowa gęstość mocy pozwala wykryć występujące w szeregach czasowych okresowości, ale istnieją również inne opisy mogące posłużyć ich efektywnemu różnicowaniu.

Dyskusja problemu adekwatnego opisu szeregów czasowych drzeń

Mimo szerokiego zakresu zastosowań widmowa gęstość mocy charakteryzuje się istotnym ograniczeniem. Dowolny proces stochastyczny $x(n)$ jest w pełni określony przez podanie jego łącznej wielowymiarowej funkcji gęstości prawdopodobieństwa. Postać tej funkcji jest jednak jednoznacznie zdeterminowana za pomocą wartości średniej oraz funkcji autokorelacji lub za pomocą funkcji autokowariancji jedynie dla procesów normalnych. Wykorzystanie widmowej gęstości mocy do opisu tego rodzaju sygnałów dzięki tw. Wienera-Chinczyna jest więc w zupełności wystarczające. W przypadku rozróżniania sygnałów innych niż normalne, tzw. sygnałów niegaussowskich, widmowa gęstość mocy może okazać się narzędziem zbyt słabym.

Przy rozpatrywaniu szeregów czasowych pochodzących od drzeń kończyn człowieka założenie o ich gaussowskim rozkładzie nie jest niestety niczym usprawiedliwione. Mogą one być:

- 1) efektem przepuszczania procesów o nieznannej łącznej wielowymiarowej funkcji gęstości prawdopodobieństwa przez układy nieliniowe o nieznanym typie nieliniowości,
- 2) odpowiedzią układów liniowych na nieznanne pobudzenie, przy czym istotne dla procesu klasyfikacji są charakterystyki fazowe tych układów,
- 3) sygnałami niegaussowskimi zawierającymi addytywne zakłócenia gaussowskie o charakterze białym lub kolorowym lub
- 4) sygnałami gaussowskimi w ramach jednej kategorii drzeń i niegaussowskimi w ramach innej.

Widmowa gęstość mocy na mocy tw. Wolda [4] o dekompozycji sygnału, wg którego dla dowolnego procesu stochastycznego $x(n)$ można zawsze znaleźć takie nieskorelowane wymuszenie $u(n)$ oraz filtr $h(n)$ aby spełnione było:

$$x(n) = \sum_{k=0}^{\infty} h(k)u(n-k) \quad (1)$$

nie może opisywać procesów powstałych w układach nieliniowych. Odpowiednikiem (1) w dziedzinie częstotliwości jest znana zależność:

$$S_{2x}(f) = \sigma_u^2 |H(f)|^2 \quad (2)$$

wg której widmowa gęstość mocy $S_{2x}(f)$ jest jednoznacznie określona modułem hipotetycznej, ale możliwej do określenia liniowej transmitancji $H(z)$ (będącej transformatą Fouriera $h(n)$). Tym samym zastosowanie $S_{2x}(f)$ do przypadku 1) nie zapewnia pełnego, tzn. z uwzględnieniem istniejącej nieliniowości, opisu procesu. Brak przenoszenia informacji fazowej przez widmową gęstość mocy wyklucza również w sposób oczywisty, przy braku dostępu do wymuszenia, jej zastosowanie do przypadku 2). W przypadkach 3) i 4) widmowa gęstość mocy dostarczając informacji o rozkładzie intensywności procesu wzdłuż osi częstotliwości jednakowo traktuje procesy gaussowskie jak i niegaussowskie, wobec czego jej wykorzystanie jest również mało przydatne. Można sobie wprost wyobrazić dwa różne procesy (tj. charakteryzujące się odmiennymi postaciami wielowymiarowych łącznych funkcji gęstości prawdopodobieństwa) białe lub pasmowe, które będą posiadały jednakowe widmowe gęstości mocy.

Dokonując pełnego scharakteryzowania widmowej gęstości mocy należy stwierdzić, iż jest ona transformatą Fouriera momentu drugiego rzędu (inaczej funkcji autokorelacji) procesu stochastycznego. W przypadku opisu sygnałów niegaussowskich naturalne jest więc sięgnięcie do momentów wyższych rzędów a zwłaszcza do pewnych ich nieliniowych kombinacji zwanych kumulantami. Ich estymatory oraz estymatory ich transformat Fouriera określa się wspólnym mianem statystyk wyższych rzędów [5].

W teorii procesów stochastycznych rozważa się istnienie pierwszej funkcji charakterystycznej, będącej transformatą Fouriera łącznej wielowymiarowej funkcji gęstości prawdopodobieństwa

procesu oraz drugiej powstałej przez zlogarytmowanie pierwszej. Kolejne współczynniki rozwinięcia tych funkcji w szeregi Taylora tworzą momenty oraz kumulanty odpowiednich rzędów. Wobec unikalnej własności zerowania się współczynników rozwinięcia drugiej funkcji charakterystycznej o numerach porządkowych większych od dwu dla procesów normalnych oraz własności liniowości tej funkcji [6] bardziej pożądane jest wykorzystanie kumulantów oraz ich częstotliwościowych odpowiedników zwanych polispektrami. Używając notacji wartości oczekiwanej kumulanty stacjonarnego procesu rzeczywistego o zerowej wartości średniej można zapisać w następujący sposób:

$$c_{2x}(k) = E[x(n)x(n+k)] \quad \text{– kumulanta II rzędu (funkcja autokorelacji),} \quad (3)$$

$$c_{3x}(k, l) = E[x(n)x(n+k)x(n+l)] \quad \text{– kumulanta III rzędu,} \quad (4)$$

$$c_{4x}(k, l, m) = E[x(n)x(n+k)x(n+l)x(n+m)] \quad \text{– kumulanta IV rzędu,} \quad (5)$$

$$c_{2x}(k)c_{2x}(l-m) - c_{2x}(l)c_{2x}(k-m) - c_{2x}(m)c_{2x}(m-l)$$

a ich transformaty Fouriera:

$$S_{2x}(f) = E[X(f)X^*(f)] = DFT\{c_{2x}(k)\} \quad \text{– widmowa gęstość mocy,} \quad (6)$$

$$S_{2x}(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1+f_2)] = DFT^2\{c_{3x}(k)\} \quad \text{– bispektrum,} \quad (7)$$

$$S_{4x}(f_1, f_2, f_3) = E[X(f_1)X(f_2)X(f_3)X^*(f_1+f_2+f_3)] = DFT^3\{c_{4x}(k)\} \quad \text{– trispektrum.} \quad (8)$$

Motywacją do wykorzystania kumulantów oraz polispektrowych rzędu $k > 2$ w przetwarzaniu sygnałów jest następująca:

- wobec wspomnianych własności liniowości drugiej funkcji charakterystycznej, jeżeli $z(n) = x(n) + y(n)$, gdzie $x(n)$ i $y(n)$ są procesami niezależnymi, to:

$$c_{kz}(m_1, m_2, \dots, m_{k-1}) = c_{kx}(m_1, m_2, \dots, m_{k-1}) + c_{ky}(m_1, m_2, \dots, m_{k-1}) \quad (9)$$

- wobec wspomnianej pierwszej własności drugiej funkcji charakterystycznej, jeżeli $x(n)$ jest procesem normalnym to:

$$c_{kx}(m_1, m_2, \dots, m_{k-1}) = 0 \quad (10)$$

- stąd, jeżeli $z(n) = x(n) + w(n)$, gdzie $w(n)$ jest szumem normalnym i niezależnym od $x(n)$, to $c_{kz}(m_1, m_2, \dots, m_{k-1}) = c_{kx}(m_1, m_2, \dots, m_{k-1})$. A zatem możliwe jest wyznaczenie kumulanty rzędu $k > 2$ procesu niegaussowskiego nawet w obecności szumu białego lub kolorowego. Możliwe jest więc również odróżnienie procesu gaussowskiego od niegaussowskiego (p. punkty 3) i 4) niniejszego rozdziału);
- jeżeli proces $x(n)$ jest rzeczywiście procesem liniowym (a więc nie tylko w sensie dekompozycji Wolda) czyli,

$$x(n) = \sum_{k=0}^{\infty} h(k)u(n-k), \quad \text{gdzie } u(n) \text{ jest nieskorelowanym}$$

wymuszeniem wyższego rzędu (tzw. proces i.i.d.), to można wyказаć [7], że:

$$S_{2x}(f) = c_{2x}(0) |H(f)|^2, \quad (11)$$

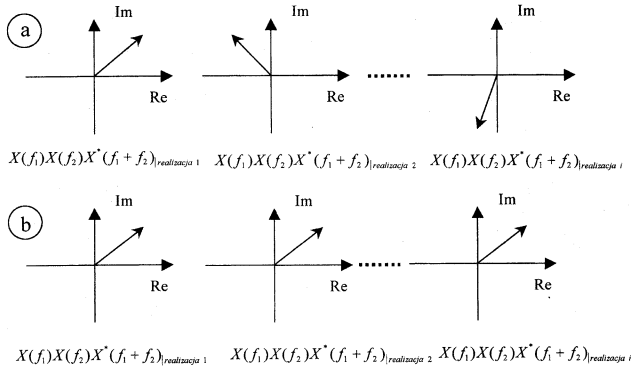
$$S_{3x}(f_1, f_2) = c_{3x}(0,0) H(f_1) H(f_2) H^*(f_1 + f_2), \quad (12)$$

$$S_{4x}(f_1, f_2, f_3) = c_{4x}(0,0,0) H(f_1) H(f_2) H(f_3) H^*(f_1 + f_2 + f_3). \quad (13)$$

Widmowa gęstość mocy $S_{2x}(f)$ nie przenosi informacji o fazie. W przeciwieństwie do niej, jeżeli tylko $u(n)$ jest procesem niegaus-

sowskim, widma wyższych rzędów taką informację zapewniają. Równania (12) i (13) stały się w ciągu ostatniej dekady podstawą szeregu algorytmów parametrycznej i nieparametrycznej estymacji fazy $H(f)$ (p. więc punkt 2));

- polispektra są wyrażone wartościami zespolonymi, tak więc przykładowo: wartość bispektrum w każdym punkcie (f_1, f_2) na płaszczyźnie zmiennej zespolonej może być przedstawiona za pomocą wektora. Wykorzystując taką interpretację oraz znajomość sposobu estymacji bispektrum np. poprzez uśrednianie kolejnych periodogramów wyższego rzędu [8] (p. też (7)) – rys. 2,



Rys. 2. Graficzna ilustracja detekcji sprzężeń fazowych pomiędzy częstotliwościami f_1 i f_2 sygnału za pomocą uśredniania periodogramów III rzędu (bispektrum): a) niekoherentne uśrednianie – brak sprzężenia, b) uśrednianie koherentne – występuje sprzężenie

można na podstawie bispektrum wnioskować o istnieniu bądź nie tzw. sprzężeń fazowych pomiędzy częstotliwościami f_1 i f_2 sygnału [4,5]. Sprzężenia tego typu występują właśnie na wyjściu układów nieliniowych pobudzonych procesem harmonicznym (p. więc punkt 1)).

Warunkiem wykorzystania kumulantów oraz polispektrow w odniesieniu do drżeń jest jednak zgodnie z (10) niegaussowskość ich rozkładu. Badania tej własności jak również przynależności do procesów liniowych zostały wykonane przez autorów za pomocą testu Hinicha [9] i przedstawione w [10]. Wyniki wskazują, iż rejestrowane przebiegi czasowe drżeń parkinsonowskich oraz samoistnych są realizacjami procesów niegaussowskich, podczas gdy przebiegi drżeń fizjologicznych pochodzą od procesów o symetrycznej funkcji gęstości prawdopodobieństwa. Wyniki testu liniowości nie dały jednoznacznej odpowiedzi co do charakteru mechanizmu powstawania drżeń (tzn. liniowego lub nieliniowego).

Procedura rejestracji i materiał

Dane do badań (szeregi czasowe drżeń) zostały zebrane w prostym systemie akwizycji złożonym z umieszczonej w komputerze karty przetwornika A/C oraz czujników przyspieszeń (akcelerometrów) wraz ze wzmacniaczami ładunkowymi. Czujniki były umieszczane na rękach pacjenta, po jednym na każdej. Do badań wykorzystywano sygnały pochodzące z tzw. rąk dominujących poszczególnych pacjentów, tzn. sygnały, które charakteryzowały się wyższą wartością skuteczną. W trakcie pomiarów pacjenci przyjmowali pozycję siedzącą, z przedramionami opartymi o poręczę fotela. Dłonie pozostawały lekko uniesione ku górze. Procedura rejestracyjna pierwotnie umożliwiała nieprzerwaną akwizycję 80. sekundowych bloków danych zbieranych z częstotliwością próbkowania 100Hz. Ponieważ dane literaturowe (np. [11]) oraz wyniki naszych poprzednich badań [12] wykazują istnienie fluktuacji częstotliwości chwilowej oraz amplitudy drżenia z przedziałami stacjonarności trwają-

cymi od kilku do kilkudziesięciu sekund, z pierwotnych wektorów danych złożonych z 8192 próbek wyodrębniono krótsze, 10. sekundowe rekordy charakteryzujące się stosunkowo mało zmienną w czasie widmową gęstością mocy. Selekcji odcinków tak rozumianego stabilnego drżenia dokonano za pomocą analizy czasowo-częstotliwościowej [15]. Aby uniknąć wpływu różnych wartości amplitudy drżeń na wyniki, rekordy te zostały znormalizowane, tzn. ustalono zerową wartość średnią i jednostkową wariancję.

Zebrany materiał objął 139 rekordów sygnałów pochodzących od drżenia parkinsonowskiego, 98 rekordów pochodzących od drżenia samoistnego, i 339 pochodzących od osób uznanych za zdrowe. W przypadku drżeń patologicznych dysponowano rejestracjami zebranymi przed zacyciem leku.

Klasyfikator

W ramach wspomnianego wstępnego przetwarzania danych wejściowych dozwolona jest w zasadzie dowolna operacja matematyczna generująca liczbową cechę sygnału. Na podstawie tak uzyskanego zbioru cech dokonuje się przypisania sygnału do jednej z kilku kategorii stosując odpowiedni klasyfikator. Klasyfikator może być klasyczny czyli liniowy (bazuje wtedy na pewnej mierze odległości w przestrzeni cech) lub nieliniowy, możliwy do realizacji przez sieć neuronową wykorzystującą nieliniowe funkcje aktywacji neuronów. W przypadku wykorzystania tylko jednej cechy wyniki klasyfikatora liniowego i nieliniowego są jednakowe. W przypadku wielowymiarowych wektorów cech stosowanie klasyfikatorów nieliniowych zapewnia jednak o wiele lepsze rezultaty.

Zagadnieniem o podstawowym znaczeniu jest jednak dobór cech dobrze różnicujących rozpatrywane sygnały, czyli cech reprezentatywnych. W celu oceny cechy c – kandydata do wektora różnicującego zastosowano wskaźnik jakości fom_{A-B} zdefiniowany dla dwu kategorii A i B w sposób następujący:

$$fom_{A-B} = \frac{|c_A - c_B|}{\sigma_A^2 + \sigma_B^2} \quad (14)$$

gdzie:

\bar{c}_A, \bar{c}_B – średnie arytmetyczne cechy w zbiorze A i B ,

σ_A^2, σ_B^2 – wariancje cechy w zbiorze A i B .

Pożądane wysokie wartości tego wskaźnika (znacznie różniące się wartości średnie i małe wariancje) wskazują na dobre odseparowanie kategorii, czyli wysoką reprezentatywność cechy c .

W celu sprawdzenia efektywności różnicowania sygnałów pochodzących od drżeń parkinsonowskich, fizjologicznych i samoistnych za pomocą cech generowanych przez różne opisy, zastosowano klasyfikator neuronowy bazujący na perceptronie wielowarstwowym wykorzystujący w procesie uczenia metodę zmiennej metryki w połączeniu z minimalizacją kierunkową [13]. Wektory d wartości pożądanych na wyjściu sieci w procesie uczenia przyjęto następujące:

drżenie parkinsonowskie $\rightarrow d = [1 \ 0 \ 0]$,

drżenie fizjologiczne $\rightarrow d = [0 \ 1 \ 0]$,

drżenie samoistne $\rightarrow d = [0 \ 0 \ 1]$,

tzn. tak, aby wartości na poszczególnych pozycjach odpowiedzi sieci na dane testujące, które nie były eksponowane w procesie uczenia stały się miarą przynależności do każdej z wymienionych kategorii drżeń (od 0 do 1). Jako błąd uczenia lub testowania potraktowano każdy przypadek w którym odpowiedź sieci y różniła się od wektora d o więcej niż 0.5 na dowolnej pozycji, czyli:

$$\text{błąd} \Leftrightarrow |y_i - d_i| > 0.5 \text{ dla dowolnego } i=1, 2, 3.$$

W celu określenia efektywności procesu różnicowania posłużono się liczbą błędów testowania sieci odniesioną do liczności zbioru danych testujących, przy czym w celu poprawnej i uczciwej oceny przypadków uzyskiwania małych błędów testowania przy jednocześnie wysokich błędach uczenia, te elementy zbioru uczącego, których sieć nie była w stanie się „nauczyć” włączono do zbioru danych testujących. Błąd różnicowania został więc określony przez:

$$err = \frac{lbu + lbt}{lbu + lt} \cdot 100\% \quad (15)$$

gdzie:

lbu , lbt – liczba błędów uczenia i testowania,
 lt – liczba danych testujących,

a efektywność różnicowania:

$$eff = \left(1 - \frac{lbu + lbt}{lbu + lt}\right) \cdot 100\% \quad (16)$$

Zbiór cech

Wobec niejednoznaczności wyników badań w zakresie określenia charakteru mechanizmu powstawania drżeń, opisy sygnałów, na podstawie których dokonano ekstrakcji cech zostały wybrane w oparciu o następujące założenia:

- mechanizm powstawania drżeń jest liniowy,
- mechanizm powstawania drżeń jest nieliniowy.

Przy przyjęciu założenia a) wykorzystano następujące opisy:

- nieparametryczną postać modułu funkcji przenoszenia układu liniowego $|H(f)|$ wyznaczoną na podstawie statystyki II rzędu (wykorzystano klasyczny estymator PSD wg procedury Welch) – PSD_{II},
- nieparametryczną postać modułu funkcji przenoszenia układu liniowego $|H(f)|$ wyznaczoną na podstawie statystyki IV rzędu (wykorzystano transformatę Fouriera diagonalnej estymaty kumulanty IV rzędu) [4] – PSD_{IV},
- parametryczną postać funkcji przenoszenia układu liniowego $H(f)$ wyznaczoną na podstawie estymacji parametrów modelu ARMA (wykorzystano tzw. algorytm „q-slice” [14] bazujący na kumulantach wyższych rzędów).

Przy przyjęciu założenia b) wykorzystano natomiast opisy następujące:

- nieparametryczną postać modułu diagonalnej bispektrum - $\text{diag}(|\text{bispl}|)$, której maksima można traktować jako miary sprzężeń fazowych pomiędzy pierwszą a drugą harmoniczną [5,8],
- nieparametryczną postać modułu diagonalnej trispektrum - $\text{diag}(|\text{trisp}|)$, której maksima można traktować jako miary sprzężeń pomiędzy pierwszą a trzecią harmoniczną.

Dla wszystkich opisów nieparametrycznych określono następujące cechy:

- c1** – częstotliwość wystąpienia maksimum,
c2 – wartość maksimum,
c3+c8 – tzw. momenty spektralne, przy których wyznaczaniu dany opis widmowy $P(k)$ traktuje się jak funkcję gęstości prawdopodobieństwa zmiennej losowej (nie mają one nic wspólnego z rozpatrywanymi statystykami wyższych rzędów i posłużyły jedynie do liczbowego scharakteryzowania opisu $P(k)$):

$$(17-22)$$

$$c3 = m1 = \sum_{k=1}^N kP^1(k), \quad c4 = m2 = \sum_{k=1}^N (k-1)^2 P^1(k),$$

$$c5 = m3 = \sum_{k=1}^N (k-1)^3 P^1(k), \quad c6 = m4 = \sum_{k=1}^N (k-1)^4 P^1(k),$$

$$c7 = m3_u = \left(\frac{1}{m2}\right)^{3/2} \cdot m3, \quad c8 = m4_u = \left(\frac{1}{m2}\right)^2 \cdot m4,$$

$$\text{gdzie: } P^1(k) = \frac{P(k)}{\sum_{k=1}^N P(k)}, \text{ czyli } \sum_{k=1}^N P^1(k) = 1.$$

- c9_a** – liczba próbek opisu widmowego unormowanego $P''(k) = P(k)/\max[P(k)]$ o wartościach większych od parametru a wyrażonego w procentach maksimum $P''(k)$ (dla jednego opisu można więc uzyskać więcej niż jedną cechę typu c9),
c10 – suma wartości wszystkich próbek opisu widmowego nie unormowanego,
c11 – suma wartości wszystkich próbek podzielona przez wartość maksymalną.

Przy opisie parametrycznym jako cechy przyjęto wartości poszczególnych parametrów ARMA.

Tab. Wartości fom dla wybranych cech

Lp.	cecha	fom		
		park-fizjo	park-samo	samo-fizjo
1	$c\ 2(\text{PSD}_{II})$	2,62	1,40	1,23
2	$c\ 1(\text{PSD}_{II})$	1,78	1,72	1,06
3	$c\ 3(\text{PSD}_{II})$	2,81	2,47	1,45
4	$c\ 4(\text{PSD}_{II})$	2,23	2,13	1,13
5	$c\ 6(\text{PSD}_{II})$	1,73	1,23	1,27
6	$c\ 9_{0,72}(\text{PSD}_{II})$	3,77	1,63	2,95
7	$c\ 9_{2,42}(\text{PSD}_{II})$	2,50	1,77	1,70
8	$c\ 9_{95,3}(\text{PSD}_{II})$	1,43	1,21	0,91
9	$c\ 7(\text{PSD}_{II})$	2,26	2,06	0,73
10	$c\ 10[\text{diag}(\text{bispl})]$	0,96	1,37	1,44
11	$c10(\text{bispl})$	2,08	0,78	1,77
12	$c\ 10\{\lg[\text{diag}(\text{bispl})]\}$	2,33	1,62	1,75
13	$c10[\lg(\text{bispl})]$	2,34	1,63	1,75
14	$c\ 3[\text{diag}(\text{bispl})]$	2,35	2,45	1,21
15	$c\ 4[\text{diag}(\text{bispl})]$	2,22	2,28	1,04
16	$c\ 3\{\lg[\text{diag}(\text{bispl})]\}$	1,92	1,59	1,09
17	$c\ 2[\text{diag}(\text{trisp})]$	1,46	1,02	1,01
18	$c\ 11[\text{diag}(\text{trisp})]$	2,10	1,61	0,74
19	$c\ 3[\text{diag}(\text{trisp})]$	1,89	2,08	0,92
20	$c\ 4[\text{diag}(\text{trisp})]$	1,26	0,17	1,33
21	$c\ 5[\text{diag}(\text{trisp})]$	1,10	1,46	0,60
22	$c\ 9_{0,29}[\text{diag}(\text{bispl})]$	3,84	2,63	0,77
23	$c\ 9_{4,3}[\text{diag}(\text{bispl})]$	1,73	1,93	1,00
24	$c\ 9_{0,15}[\text{diag}(\text{trisp})]$	2,18	0,38	3,29
25	$c\ 9_{5,7e-6}[\text{diag}(\text{trisp})]$	1,46	1,27	1,27
26	$c\ 7[\text{diag}(\text{bispl})]$	1,84	1,75	0,97
27	$c\ 1(\text{PSD}_{IV})$	2,29	1,56	0,91
28	$c\ 3(\text{PSD}_{IV})$	2,34	1,58	1,47
29	$c\ 5(\text{PSD}_{IV})$	2,26	1,47	1,24
30	$c\ 9_{0,43}(\text{PSD}_{IV})$	3,67	1,31	1,33

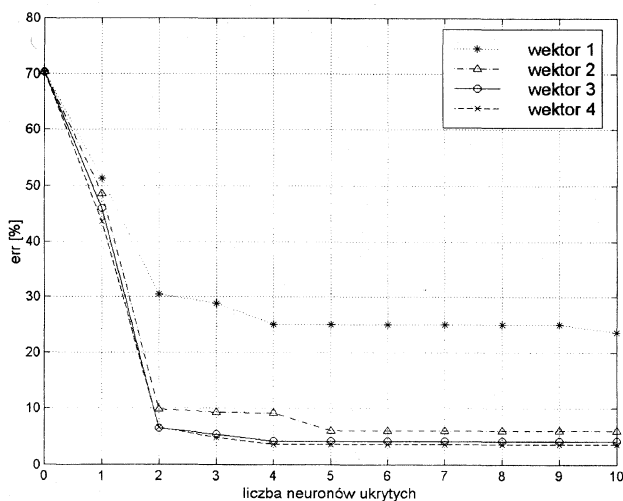
Reprezentatywność tak wygenerowanych cech została zweryfikowana wprowadzonym wskaźnikiem jakości (14). Do tworzenia wektorów cech, na podstawie których oparł się proces klasyfikacji neuronowej wytypowano tylko te cechy, które charakteryzowały się wartością $fom > 1$ dla przynajmniej dwu z trzech możliwych rozróżnień (drżenie parkinsonowskie – fizjologiczne, fizjologiczne – samoistne i parkinsonowskie – samoistne). Uzyskany w ten sposób zbiór 30 cech przedstawiony został wraz z wartościami fom w tabeli.

Wyniki klasyfikacji i wnioski końcowe

Na bazie zaprezentowanego zbioru 30 cech charakteryzujących rozpatrywane drżenia wyznaczone zostały cztery wektory różnicujące:

- wektor 1 – zawierający wspomniane dwa parametry widmowej gęstości mocy, tzn. wartość maksymalną oraz częstotliwość jej wstąpienia (cechy 1+2 z tabeli),
- wektor 2 – zawierający cechy wektora 1 oraz 7 innych cech opisujących widmową gęstość mocy (łącznie cechy 1+9 z tabeli),
- wektor 3 – zawierający 21 cech uzyskanych przy opisie z wykorzystaniem statystyk wyższych rzędów (cechy 10+30 z tabeli),
- wektor 4 – zawierający wszystkie cechy z wektorów 2 i 3 (łącznie 30 cech).

Posiadany zbiór danych został podzielony na zbiór uczący (ok. 70% ogółu danych) i zbiór testujący (ok. 30% ogółu danych), które utworzyły w ten sposób materiał do oceny efektywności różnicowania drżeń za pomocą wektorów 1-4. W celu zapewnienia jak najwyższej zdolności uogólniania, w strukturze klasyfikatora neuronowego przyjęto prosty schemat z jedną warstwą ukrytą. W warstwie wejściowej zastosowano liniową funkcję aktywacji neuronów, w dwu pozostałych funkcję sigmoidalną. Wyniki badania efektywności klasyfikacji przedstawione są na rys. 3. w postaci błędu klasyfikacji (15) wyznaczonego w funkcji złożoności klasyfikatora (czyli liczby neuronów ukrytych).



Rys. 3. Błąd klasyfikacji w funkcji złożoności klasyfikatora dla czterech wektorów różnicujących

Badania przeprowadzone w zakresie od 0 do 10 neuronów ukrytych wykazują, iż:

- wektor 1, potraktowany jako wektor odniesienia, zawierający cechy najczęściej spotykane w literaturze umożliwił uzyskanie błędu klasyfikacji na poziomie ok. 23-25%,

- uwzględnienie w wektorze 2 innych dodatkowych cech opisujących widmową gęstość mocy (m.in. momenty spektralne) zdecydowanie poprawia efektywność klasyfikacji – uzyskano redukcję błędu do poziomu ok. 6%, przy czym nie udało się uzyskać całkowitego nauczenia sieci,
- wykorzystanie wektora 3 (wszystkie cechy pochodzące od statystyk wyższych rzędów) umożliwiło całkowite nauczenie sieci neuronowej przy 3 neuronach ukrytych oraz dalszą redukcję błędu do poziomu ok. 4% przy 4 neuronach ukrytych,
- zastosowanie wektora 4, powstałego przez dodanie do wektora 3 cech uzyskanych na bazie widmowej gęstości mocy umożliwiło nauczenie sieci już przy 2 neuronach ukrytych oraz dalsze zmniejszenie błędu klasyfikacji do poziomu 3.6% przy 4 neuronach ukrytych.

Przedstawione wyniki klasyfikacji z wykorzystaniem cech pochodzących od statystyk wyższych rzędów w porównaniu z klasycznym (wektor 1) a nawet rozszerzonym (wektor 2) wektorem cech uzyskanym na bazie statystyki II rzędu wydają się zadowalające. Wydaje się jednak również, iż konieczne jest jednak przeprowadzenie dodatkowych eksperymentów zmierzających do zmniejszenia wymiarowości wektora 4 (kombinowany zestaw cech) przy zachowaniu tego samego poziomu błędu. Być może możliwe jest jeszcze również dodatkowe poprawienie zdolności generalizacji innymi metodami, np. metodą wtrącenia szumu, sieci Pao itp., lub przez wykorzystanie innego rodzaju sieci [13].

Literatura

- [1] J. TIMMER, C. GANTERT, G. DUSCH, J. HONERKAMP: Characteristics of hand tremor time series; Biological Cybernetics, vol. 70, 1993, str. 75-80.
- [2] R. J. ELBLE, W. C. KOLLER: Tremor; John Hopkins University Press, Baltimore, 1990.
- [3] C. GANTERT, J. HONERKAMP, J. TIMMER: Analyzing the dynamics of hand tremor time series; Biological Cybernetics, vol. 66, 1992, str. 479-484.
- [4] A. SWAMI, J. M. MENDEL: Higher-Order Spectral Analysis Toolbox. For Use with Matlab; The Mathworks Partner Series, Natick, 1995.
- [5] J. M. MENDEL: Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications; Proc. IEEE, vol. 79, 1991, str. 278-305.
- [6] A. PAPOULIS: Prawdopodobieństwo, zmienne losowe i procesy stochastyczne; WNT, Warszawa, 1972.
- [7] G. B. GIANNAKIS: Cumulants: A Powerful Tool in Signal Processing; Proc. IEEE, vol. 75, 1987, str. 1333-1334.
- [8] C. NIKIAS, J. M. MENDEL: Signal Processing with Higher-Order Spectra; IEEE Technology Update Series, Nowy Jork, 1995.
- [9] M.J. HINICH: Testing for Gaussianity and Linearity of Stationary Time Series; J. of Time Series Analysis, vol. 3, 1982, str. 169-176.
- [10] R. BOBROWICZ i inni: Badanie procesów powstawania drżeń kończyn górnych człowieka za pomocą testu Hinicha, I Sympozjum Modelowanie i Pomiar w Medycynie „MPM'99”, Krynica Górská, kwiecień 1999, materiały konferencyjne str. 256-265.
- [11] P. O'SUILLEBHAIN: Fluctuations of instantaneous tremor frequency, Tremor Symposium, Kiel, 1997.
- [12] R. BOBROWICZ i inni: Analiza czasowo-częstotliwościowa drżeń patologicznych; XXX Międzyuczelniana Konferencja Metrologów, Międzyzdroje, 1998, str. 205-210.
- [13] S. OSOWSKI: Sieci neuronowe w ujęciu algorytmicznym; WNT, Warszawa, 1996.
- [14] A. SWAMI, J. M. MENDEL: ARMA Parameter Estimation Using Only Output Cumulants; IEEE Trans. ASSP, vol. 38, 1990, str. 1257-1264.
- [15] J. JAKUBOWSKI: Wykorzystanie reprezentacji czasowo-częstotliwościowej w analizie sygnałów niestacjonarnych, Biuletyn Wojskowej Akademii Technicznej, nr 2, 1999, str. 65-88.