

Ocena zmienności HNR w czasie maksymalnie przedłużonej fonacji samogłoski /a/ na przykładzie osób zdrowych i pacjentów z chorobą Parkinsona

Streszczenie. Parametr HNR (ang. Harmonic to Noise Ratio) jest miarą zawartości szumu w sygnale reprezentującym mowę dźwięczną. W pracy przedstawiono metodę i wyniki oceny zmienności HNR w czasie maksymalnie przedłużonej fonacji samogłoski /a/. Uwzględnienie początkowej i końcowej fazy fonacji można wykorzystać do obiektywnej oceny funkcji fonacyjnej głośni u pacjenta z chorobą Parkinsona.

Abstract. HNR (Harmonic to Noise Ratio) is a measure of the noise content in the signal representing voiced speech. This paper presents the method and results of the assessment of HNR variability during the maximally prolonged phonation of the vowel /a/. Taking into account the initial and final phases of phonation can be used to objectively assess the phonation function of the glottis in patient with Parkinson's disease. (Assessment of HNR variability during maximally prolonged phonation of the vowel /a/ on the example of healthy subjects and patients with Parkinson's disease).

Słowa kluczowe: przedłużona fonacja samogłoski, HNR, funkcja autokorelacji, mowa dźwięczna w chorobie Parkinsona.

Keywords: prolonged phonation of vowel, HNR, autocorrelation function, voiced speech in Parkinson disease.

Wstęp

Choroba Parkinsona (ang. Parkinson Disease, PD) to przewlekła, postępująca choroba zwyrodnieniowa układu nerwowego, którą cechują typowe objawy ruchowe [1]:

- drżenie spoczynkowe (dotyczy 70% chorych),
- sztywność mięśni,
- bradykinezja (tj. spowolnienie ruchowe),
- zaburzenia postawy i chodu.

W chorobie Parkinsona występują także zaburzenia mowy w zakresie fonacji, artykulacji oraz prozodii (tzw. dyszartria hipokinetyczna), które nasilają się wraz z rozwojem tej choroby [2].

Zmniejszona objętość wydychanego powietrza w wyniku dysfunkcji mięśni krtani, upośledzenie procesu oddychania (tj. spływanie wdechu i skrócenie fazy wydechowej) oraz brak koordynacji cyklu oddechowego i fonacji przyczynia się do osłabienia dynamiki głosu chorego, tzn. głos staje się cichy i monotony w trakcie wypowiedzi. Chory podczas dłuższej wypowiedzi oraz śpiewania odczuwa brak powietrza i zmęczenie.

Ponadto wzmożone napięcie mięśniowe i bradykinezja mięśni artykulacyjnych powodują nieprecyzyjną realizację głosek. Mowa pacjenta wraz z rozwojem choroby Parkinsona staje się stopniowo coraz mniej zrozumiała dla otoczenia i prowadzi do utraty zdolności komunikacji z pacjentem.

W chorobie Parkinsona zazwyczaj zaburzenia głosu poprzedzają zaburzenia artykulacyjne.

W wyniku niedomykalności głośni spowodowanych przez sztywność mięśni napinających fałdy głosowe oraz nieregularności drgań fałdów głosowych w sygnale mowy dźwięcznej (którego widmo cechuje struktura harmoniczna) występuje addytywny szum. Obiektywną ocenę zawartości tego szumu w sygnale mowy dźwięcznej umożliwia parametr HNR (ang. Harmonic to Noise Ratio) lub jego odwrotność, tj. NHR (ang. Noise to Harmonic Ratio).

W niniejszym artykule opisano metodę przyjętą do wyznaczenia parametru HNR w czasie maksymalnie przedłużonej fonacji wyizolowanej samogłoski /a/. Zaproponowano nowy sposób analizy zmienności HNR ze szczególnym uwzględnieniem początkowej i końcowej fazy fonacji. Przedstawiono wyniki oceny zmienności wartości parametru HNR, które uzyskano dla mówców zdrowych

oraz pacjentów w różnym stanie zaawansowania choroby Parkinsona.

Wykazano, że obiektywna ocena zmęczenia głosu pacjenta z chorobą Parkinsona, która została wywołana przez niewielki wysiłek głosowy, tj. maksymalnie przedłużoną fonację głośli /a/, może być przeprowadzona na podstawie analizy zmienności parametru HNR w początkowej i końcowej fazie fonacji.

Mowa dźwięczna – pobudzenie krtaniowe

Przyjmuje się, że quasi-stacjonarny segment sygnału reprezentujący mowę dźwięczną jest wynikiem spłotu pobudzenia krtaniowego oraz odpowiedzi impulsowej filtru liniowego, który modeluje trakt głosowy.

Pobudzenie krtaniowe jest wytwarzane przez drgania fałdów głosowych, które cyklicznie modulują strumień powietrza przepływającego z płuc do kanału głosowego. Warunkiem prawidłowej fonacji są m.in. symetryczne drgania fałdów głosowych z pełną wydolnością fonacyjną głośni. Podczas fonacji fałdy głosowe powinny zbliżyć się do siebie tak, aby wąska, eliptyczna szczelina pomiędzy nimi umożliwiała im swobodne drgania.

Czas trwania jednego cyklu pracy fałdów głosowych odpowiada okresowi tonu krtaniowego (T_0), a jego odwrotność jest nazywana częstotliwością podstawową tonu krtaniowego (F_0).

O wartości częstotliwości F_0 decyduje długość i masa fałdów głosowych, ich napięcie i sztywność oraz wartość ciśnienia podgłośniowego. Wartość częstotliwości F_0 zależy od płci, stanu emocjonalnego mówcy i od intonacji wypowiedzi. Zmienia się także wraz z wiekiem mówcy.

Widmo sygnału reprezentującego mowę dźwięczną bez zaburzeń cechuje struktura znamienna dla sygnału quasi-okresowego, który zawiera harmoniczne do częstotliwości 4 kHz oraz niski poziom szumu.

W przypadku niepełnego zamknięcia głośni podczas cyklu pracy fałdów głosowych przepływ powietrza z płuc do kanału głosowego jest turbulentny i generuje addytywny szum.

Jak już wspomniano, w chorobie Parkinsona główną przyczyną niepełnego zwarcia fonacyjnego są sztywność mięśni napinających fałdy głosowe oraz niesymetryczne drgania tych fałdów.

Definicja i metody wyznaczania *HNR*

Do oceny zawartości szumu występującego w sygnale mowy dźwięcznej wykorzystuje się tzw. parametr *HNR* (ang. Harmonic to Noise Ratio), który definiuje się jako wartość stosunku energii przypadającej na częstotliwość podstawową F_0 oraz jej harmoniczne do energii reprezentowanej przez szum. Wartość *HNR* określa się dla różnych pasm częstotliwości [3].

Osoby zdrowe cechuje niski poziom szumu, tj. duża wartość parametru *HNR*. Parametr *HNR* często przyjmuje się do opisu szorstkości głosu oraz oceny chrypki, a także wykorzystuje się do określenia wpływu wysiłku głosowego na funkcję fonacyjną krtani [4]. Na podstawie *HNR* można obiektywnie ocenić stopień zmęczenia całego organizmu człowieka, który jest wynikiem zaburzeń snu.

Do wyznaczania wartości *HNR* opracowano kilka metod, które szczegółowo omówiono m. in. w pracach [4-8].

Zazwyczaj wartość *HNR* oblicza się na podstawie quasi-stacjonarnego segmentu sygnału mowy dźwięcznej, który obejmuje tylko kilka cykli pracy fałdów głosowych. Segmenty te są wydzielane za pomocą ruchomego okna czasowego przesuwanego z zakładką lub bez zakładki.

Sygnał mowy $s(n)$ można opisać za pomocą zależności:

$$(1) \quad s(n) = (p(n) + a(n)) * v(n)$$

gdzie: $p(n)$ i $a(n)$ – składowe pobudzenia krtaniowego odpowiednio quasi-okresowa i aperiodyczna (tj. addytywny szum), $v(n)$ – odpowiedź impulsowa traktu głosowego, * – operacja splotu.

Jedną z metod stosowaną do obliczania *HNR* (tzw. algorytm de Krom'a) polega na usunięciu harmonicznych z Cepstrum i wyznaczeniu za pomocą transformaty Fouriera widma sygnału mowy, który nie zawiera składowej okresowej [4, 6]. Różnica pomiędzy takim widmem a widmem oryginalnego segmentu sygnału mowy jest definiowana jako *HNR*.

Natomiast podstawą tzw. algorytmu d' Alessandro jest przyjęcie dla badanego segmentu sygnału mowy modelu parametrycznego, który umożliwia uzyskanie sygnału pobudzenia krtaniowego [4]. W celu wyznaczenia składowej aperiodycznej przeprowadza się złożoną analizę widma amplitudowego pobudzenia krtaniowego dzieląc je na „aperiodyczne i periodyczne regiony” na podstawie informacji o pierwszej harmonicznej występującej w Cepstrum pobudzenia krtaniowego. Odtworzenie składowej aperiodycznej wymaga wykonania kilku iteracji algorytmu. Składową periodyczną otrzymuje się w wyniku operacji odejmowania składowej aperiodycznej od sygnału pobudzenia krtaniowego.

W pracy [8] zaproponowano wykorzystanie metody EMD (ang. Empirical Mode Decomposition) do dekompozycji widma amplitudowego wyznaczanego dla segmentu sygnału mowy dźwięcznej objętego przez ruchome okno czasowe (o długości 0,2 s). Metoda ta umożliwia wydzielenie trzech składowych widma, tj. jego obwiedni, harmonicznych oraz szumu.

Najczęściej do wyznaczania wartości *HNR* stosowana jest metoda bazująca na funkcji autokorelacji (tzw. algorytm Boersma) [4, 5, 7].

Funkcja autokorelacji jest miarą podobieństwa sygnału i jego kopii przesuniętej w czasie:

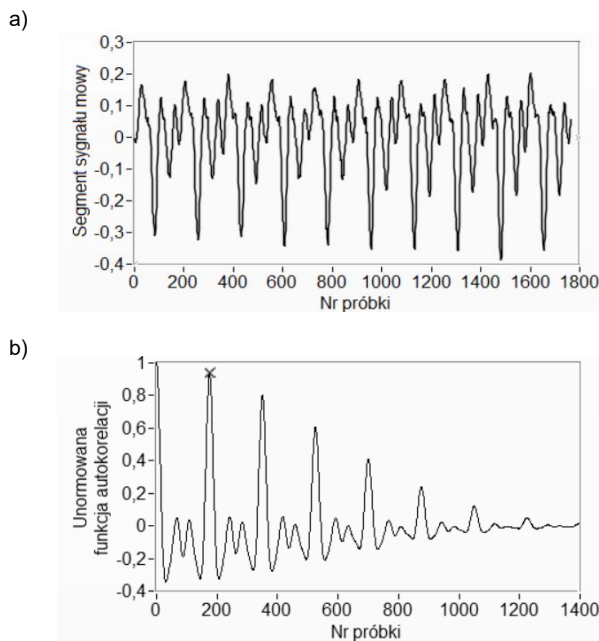
$$(2) \quad R_{xx}(k) = \sum_{n=0}^{N-1-k} s(n)s(n+k), \quad k = 0, 1, 2, \dots, N-1$$

gdzie: $s(n)$ – segment sygnału mowy dźwięcznej, który obejmuje N próbek.

Funkcja autokorelacji przyjmuje największą wartość dla przesunięcia τ równego zero. Zakłada się, że ta wartość funkcji autokorelacji jest równa energii sygnału pobudzenia krtaniowego, tzn. uwzględnia zarówno składową okresową, jak i składową aperiodyczną.

Na podstawie lokalizacji czasowej pierwszego lokalnego maksimum funkcji autokorelacji (dla przesunięcia τ większego od zera) oblicza się okres tonu krtaniowego T_0 . Przyjmuje się, że amplituda tego maksimum odpowiada energii składowej quasi-okresowej pobudzenia krtaniowego.

Na rysunku 1 przedstawiono quasi-stacjonarny segment sygnału mowy dźwięcznej (głoski /a/) oraz przebieg jego funkcji autokorelacji. W celu ułatwienia detekcji pierwszego lokalnego maksimum amplituda funkcji autokorelacji została unormowana, tzn. wartości funkcji autokorelacji podzielono przez jej wartość maksymalną.



Rys. 1. (a) Quasi-okresowy segment sygnału mowy dźwięcznej bez zaburzeń, (b) unormowana funkcja autokorelacji tego sygnału oraz wynik detekcji pierwszego lokalnego maksimum (częstotliwość próbkowania f_s wynosi 44100Hz)

Wartość *HNR* wyrażoną w dB wyznacza się wg wzoru:

$$(3) \quad HNR = 10 \log_{10} \frac{R_{xx}(T_0)}{R_{xx}(0) - R_{xx}(T_0)}$$

gdzie: $R_{xx}(0)$ oraz $R_{xx}(T_0)$ – to odpowiednio wartość funkcji autokorelacji sygnału mowy dźwięcznej dla przesunięcia $\tau=0$ oraz $\tau=T_0$. W przypadku funkcji autokorelacji o unormowanej amplitudzie $R_{xx}(0) = 1$.

Stanowisko pomiarowe

Sygnał mowy dźwięcznej, tj. głoskę /a/ o maksymalnie przedłużonej fonacji wypowiedzianą kilkakrotnie przez badanego mówcę zarejestrowano za pomocą cyfrowego rejestratora dźwięku SR-Q Saramonic. Cyfrowa rejestracja dźwięku bazuje na 16-bitowym przetworniku A/C. Częstotliwość próbkowania wynosiła 44100 Hz.

Wspomniany rejestrator umożliwia archiwizację danych w postaci plików w formacie *.wav. W czasie nagrywania rejestrator leżał na stole w odległości ok. 40 cm przed badaną osobą, która znajdowała się w pozycji siedzącej.

W badaniach wzięło udział 30 osób zdrowych (grupa Control) oraz 30 pacjentów w różnym stadium zaawansowania choroby Parkinsona (grupa PD).

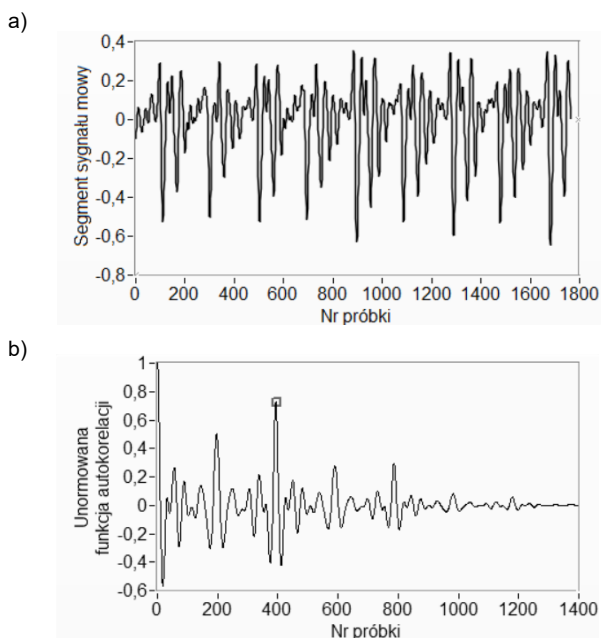
Opis metody

W ramach przetwarzania wstępnego sygnał akustyczny zarejestrowany za pomocą cyfrowego rejestratora dźwięku został poddany filtracji w celu ograniczenia pasma częstotliwości, dla którego będzie wyznaczany parametr *HNR*. Największy udział harmonicznym w strukturze widma sygnału mowy dźwięcznej jest do częstotliwości 1000 Hz, dlatego do oceny zawartości addytywnego szumu w sygnale mowy osób z chorobą Parkinsona wybrano pasmo częstotliwości od 40 Hz do 1000 Hz.

Następnie z sygnału po filtracji wydzielane są tylko te segmenty, które reprezentują samogłoskę /a/. Detekcja początku i końca fonacji pojedynczej samogłoski odbywa się metodą progowania obwiedni amplitudowej, która wyznaczana jest za pomocą transformaty Hilberta na podstawie zarejestrowanego sygnału akustycznego po operacji filtracji. Uzyskanie dostatecznie gładkiej obwiedni amplitudowej zapewnia filtr średniej ruchomej.

Założono, że czas trwania tzw. stanu nieustalonego na początku i na końcu fonacji samogłoski /a/ nie powinien przekroczyć 0,1 s. Do dalszej analizy wybierany jest automatycznie segment sygnału mowy dźwięcznej z pominięciem tych fragmentów, które są związane ze stanami nieustalonymi. Następnie wykonywana jest operacja segmentacji sygnału reprezentującego przedłużoną fonację głoski /a/ za pomocą ruchomego okna czasowego (tj. okna Hamminga) o długości 40 ms i przesuwanego bez zakładki. Dla każdego wydzielonego przez okno segmentu sygnału mowy dźwięcznej wyznaczana jest funkcja autokorelacji i określana jest wartość *HNR* wg wzoru (3).

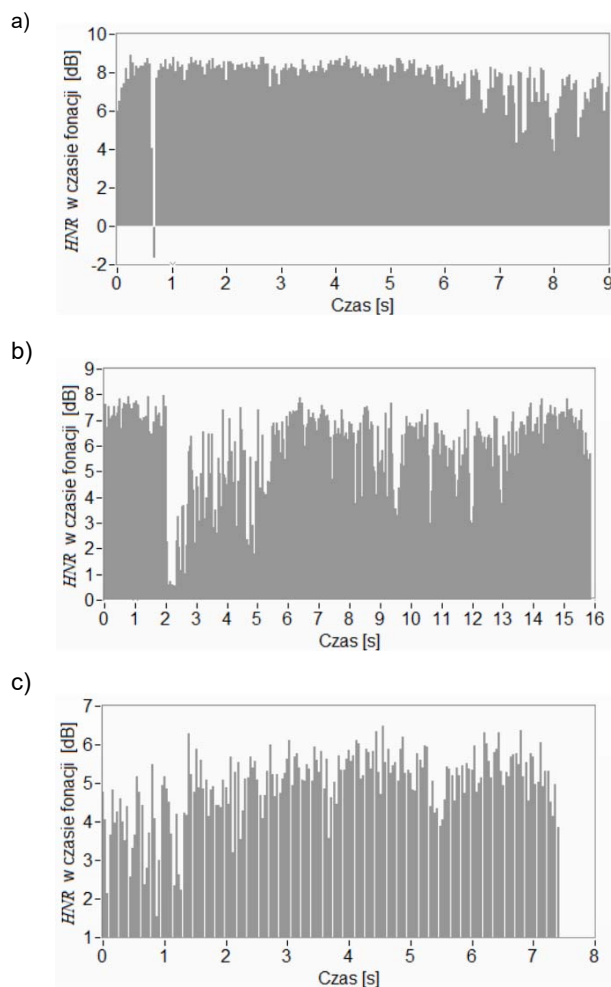
Poprawną detekcję pierwszego lokalnego maksimum odpowiadającego okresowi tonu krtańowego T_0 utrudnia obecność tzw. subharmonicznej o częstotliwości $F_0/2$. Problem ten dobrze ilustruje rysunek 2. W takim przypadku wartość *HNR* jest bardzo mała (rys. 3a, b).



Rys. 2. (a) Quasi-okresowy segment sygnału mowy dźwięcznej osoby z chorobą Parkinsona w początkowym stadium, (b) unormowana funkcja autokorelacji tego sygnału oraz wynik detekcji pierwszego lokalnego maksimum (częstotliwość próbkowania f_s wynosi 44100Hz)

W wyniku końcowym uzyskuje się serię wartości *HNR*, które są znamienne dla poszczególnych lokalizacji ruchomego okna.

Na rysunku 3 zaprezentowano przykładowe wyniki uzyskane dla maksymalnie przedłużonej fonacji głoski /a/ wypowiedzianej przez osobę zdrową i pacjentów w różnym stopniu zaawansowania choroby Parkinsona.



Rys. 3. Wartości *HNR* wyznaczone na podstawie maksymalnie przedłużonej fonacji samogłoski /a/ dla osoby zdrowej (a) oraz dla pacjentów w początkowym (b) i w zaawansowanym stadium choroby Parkinsona (c)

Na podstawie serii wartości *HNR* oblicza się tradycyjnie wartość średnią i odchylenie standardowe, które cechują całą przedłużoną fonację.

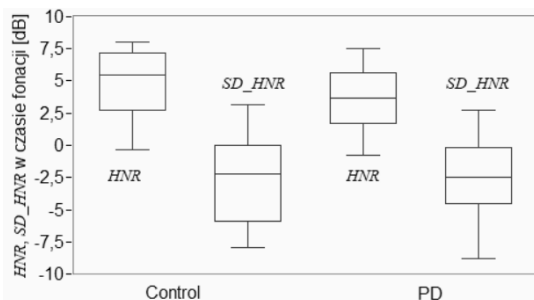
Na podstawie przeglądu literatury przedmiotowej można stwierdzić, że dotychczas nie badano zmienności wartości *HNR* w początkowej i końcowej fazie fonacji.

Przyjęto założenie, że dla badanych pacjentów maksymalnie przedłużona fonacja samogłoski /a/ może stanowić znaczny wysiłek głosowy. Aby to wykazać, zaproponowano porównanie zmienności *HNR* w 1-sekundowych interwałach reprezentujących początek i koniec fonacji (co oznaczono jako blok danych A i B). Do oceny zmienności *HNR* przyjęto parametry statystyczne.

Wyniki

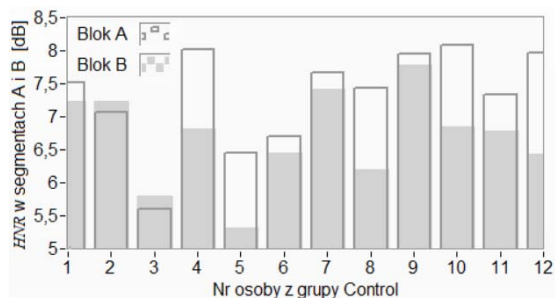
Na rysunku 4 przedstawiono wyniki uzyskane dla osób zdrowych (z grupy Control) i pacjentów z chorobą Parkinsona (PD). Wykresy pudełkowe ilustrują rozrzut wartości średniej *HNR* oraz odchylenia standardowego SD_{HNR} , które zostały wyznaczone dla każdej z badanych osób na podstawie analizy całej maksymalnie przedłużonej fonacji samogłoski /a/.

Pacjentów PD cechuje mniejsza wartość mediany *HNR* niż osoby zdrowe (oznaczone jako Control).

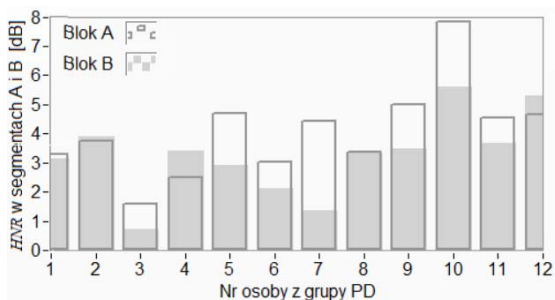


Rys. 4. Rozrzut wartości średniej *HNR* i odchylenia standardowego *SD_HNR* na podstawie maksymalnie przedłużonej fonacji głoski /a/ (Mediana, kwartyle, zakres Max – Min)

Natomiast na rysunkach 5 i 6 zaprezentowano wartość średnią *HNR* w początkowej oraz końcowej fazie fonacji (tj. odpowiednio w blokach A oraz B) dla grupy 12 osób zdrowych i 12 pacjentów w różnym stadium zaawansowania choroby Parkinsona. Dla większości mówców wartość średnia *HNR* w początkowej fazie fonacji (blok A) jest większa niż w końcowej fazie fonacji (blok B). Krótkookresowy wysiłek głosowy związany z wypowiedzianiem głoski /a/ o maksymalnie przedłużonej fonacji powoduje zwiększenie poziomu szumu także u osób zdrowych. W przypadku pacjentów z zaawansowaną chorobą Parkinsona wartości średnie parametru *HNR* są znacznie mniejsze niż dla osób zdrowych.



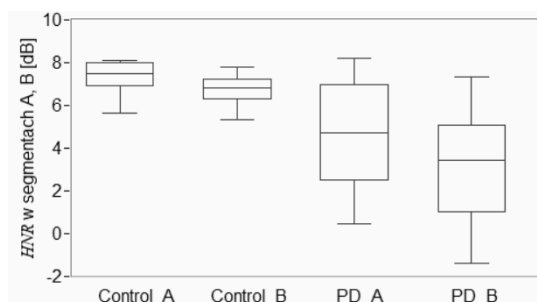
Rys. 5. Wartość średnia *HNR* obliczona dla bloku danych A oraz B – grupa osób zdrowych



Rys. 6. Wartość średnia *HNR* obliczona dla bloku danych A oraz B – grupa osób z chorobą Parkinsona

Wykresy pudełkowe przedstawione na rys. 7 dobrze ilustrują wpływ wysiłku głosowego na rozrzut średniej wartości *HNR* w początkowej i końcowej fazie maksymalnie przedłużonej fonacji głoski /a/. Na podstawie tych wykresów można stwierdzić, że średnia *HNR* osiąga większe wartości i znacznie mniejszy rozrzut dla grupy osób zdrowych. Pacjentów PD charakteryzuje dużo większy spadek wartości mediany *HNR* w końcowej fazie fonacji w porównaniu do fazy początkowej. Przyczyną dużego

rozrzutu wartości średniej *HNR* w blokach A i B w grupie pacjentów PD jest zróżnicowany stan zaawansowania choroby Parkinsona.



Rys. 7. Rozrzut wartości średniej *HNR* na podstawie początkowej (blok A) i końcowej (blok B) fazy maksymalnie przedłużonej fonacji głoski /a/ (Mediana, kwartyle, zakres Max – Min)

Podsumowanie

W artykule opisano metodę wyznaczania wartości parametru *HNR*, który umożliwia ocenę zawartości szumu w sygnale reprezentującym głoskę /a/ o maksymalnie przedłużonej fonacji. Parametr *HNR* określany jest na podstawie unormowanej funkcji autokorelacji dla poszczególnych segmentów sygnału mowy dźwięcznej, które są wydzielane przez ruchome okno czasowe (40 ms). Na podstawie uzyskanej serii wartości *HNR* obliczono wartość średnią odpowiadającą całkowitej fonacji oraz wartości średnie znamienne dla początkowej i końcowej fazy fonacji. Wyniki analizy zmienności *HNR* z uwzględnieniem początkowej i końcowej fazy maksymalnie przedłużonej fonacji głoski /a/ (tj. w pierwszej sekundzie od rozpoczęcia fonacji i w ostatniej sekundzie trwania fonacji) wykazały, że można obiektywnie ocenić krótkotrwały wysiłek głosowy zarówno u osób zdrowych, jak i u pacjentów z chorobą Parkinsona, co warto wykorzystać w diagnostyce medycznej do monitorowania stanu pacjenta z chorobą Parkinsona.

Autor: dr inż. Barbara Wilk, Politechnika Rzeszowska, Katedra Metrologii i Systemów Diagnostycznych, ul. W. Pola A, 35-959 Rzeszów, E-mail: bmwilki@prz.edu.pl

LITERATURA

- [1] Friedman A. (red. naukowa), Choroba Parkinsona od mechanizmów do leczenia, PZWL Wydawnictwa Lekarskie, Warszawa (2019)
- [2] Moro-Velazquez L. et al., Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects, *Biomedical Signal Processing and Control*, 66 (2021) 102418
- [3] Gelzinis A. et al., Automated speech analysis applied to laryngeal disease categorization, *Comput Methods Programs Biomed.* 91 (2008) No. 7, 36-47
- [4] Severin F. et al., HNR extraction in voiced speech, oriented towards voice quality analysis, *13th European Signal Processing Conference, Antalya, Turkey, (2005)*, 1-4
- [5] Fernandes, J. F. T. et al., Determination of Harmonic Parameters in Pathological Voices - Efficient Algorithm, *Appl. Sci.* 13 (2023), No. 4, 2333
- [6] de Krom G., A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals, *Journal of Speech and Hearing Research*, 36 (1993), No. 4, 254-265
- [7] Boersma P., Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sample sound, *IFA Proceedings*, 17 (1993), 97-110
- [8] Kacha A. et al., On the harmonic-to-noise ratio as an acoustic cue of vocal timbre of Parkinson speakers, *Biomedical Signal Processing and Control*, 37 (2017), 32-38