

Parametryzacja sygnału mowy reprezentującego samogłoskę o przedłużonej fonacji w celu oceny zmęczenia głosu

Streszczenie. W artykule opisano metodę przetwarzania sygnału mowy, który reprezentuje wyizolowaną samogłoskę /a/ o maksymalnie przedłużonej fonacji. Do oceny zmęczenia głosu zaproponowano parametry opisujące zmienność amplitudy sygnału mowy (tj. *Shimmer*, *APQ*) oraz zmienność okresu tonu krtańowego (tj. *Jitta*, *Jitter*, *RAP*, *PPQ*, a także *SD1* i *SD2* wyznaczone na podstawie wykresu Poincaré) z uwzględnieniem każdego cyklu pracy fałd głosowych.

Abstract. In this article, the parameterization method of speech signal representing the isolated maximally sustained vowel /a/ is described. For assessing vocal fatigue the parameters describing the variability of speech signal amplitude and pitch period with respect to each vocal folds cycle were proposed (i.e. *Shimmer*, *APQ*, *Jitta*, *Jitter*, *RAP*, *PPQ*, *SD1* and *SD2* determined from the Poincaré plot). (**Parameterization of speech signal representing vowel during sustained phonation for assessing vocal fatigue**).

Słowa kluczowe: mowa dźwięczna, przedłużona fonacja, zmienność tonu krtańowego, wykres Poincaré.

Keywords: voiced speech, sustained phonation, pitch variability, Poincaré plot.

Wstęp

Wytwarzanie dźwięków mowy wymaga koordynacji kilku procesów: psychologicznego (konceptualizacja informacji), neurologicznego (wygenerowanie pobudzenia w ośrodkowym układzie nerwowym i przesyłanie bodźców do poszczególnych elementów układu artykulacyjnego), artykulacji oraz aerodynamicznego (związanego z przepływem powietrza w torze głosowym i wzbudzeniem drgań akustycznych o złożonej strukturze widmowo-czasowej).

Przyjmuje się, że quasi-stacjonarny segment sygnału mowy dźwięcznej jest wynikiem splotu tzw. pobudzenia krtańowego oraz odpowiedzi impulsowej filtra liniowego, który reprezentuje charakterystykę amplitudowo-częstotliwościową traktu głosowego. Pobudzenie krtańowe powstaje w wyniku drgań fałdów głosowych, które modulują strumień powietrza przepływającego z płuc do kanału głosowego. Drgania fałdów głosowych zależą od masy fałdu głosowego, jego napięcia, wartości ciśnienia podgłośnia, przepływu powietrza przez głośnie oraz wzajemnego stosunku faz zamknięcia i otwarcia głośni. Warunkiem prawidłowej fonacji są m.in. symetryczne drgania fałdów z pełną wydolnością fonacyjną głośni oraz miękkie nastawienie głosowe (tzn. podczas fonacji fałdy głosowe zbliżają się do siebie, a pomiędzy nimi pozostaje wąska, eliptyczna szczelina, która pozwala im na swobodne drgania). Czas trwania jednego cyklu pracy fałdów głosowych nazywany jest okresem tonu krtańowego (T_0), a jego odwrotność częstotliwością podstawową tonu krtańowego (F_0).

Zmęczenie głosowe (ang. vocal fatigue) występuje po nadmiernym wysiłku głosowym najczęściej u osób, których głos jest podstawowym narzędziem pracy. Kandydaci do takich zawodów powinni mieć odpowiednie warunki głosowe (tj. prawidłową budowę anatomiczną krtań i narządów artykulacyjnych, sprawny układ oddechowy, pełną wydolność oddechowo-fonacyjną) oraz znać i stosować techniki poprawnej emisji głosu [1, 2, 3].

W stanie określanym jako zmęczenie głosowe m.in. dochodzi do znacznego nagromadzenia mleczanów w mięśniach krtań oraz wydzielania się więcej energii cieplnej podczas drgań fałdów głosowych z powodu zwiększenia lepkości śluzu produkowanego przez gruczoły wydzielnicze błony śluzowej głośni. Śluz o odpowiedniej lepkości zapobiega nadmiernemu „tarceniu” i „przeegrzewaniu” się krawędzi fałdów głosowych w trakcie ich drgań. Suchość błony śluzowej krtań powoduje także znaczny wzrost

wartości ciśnienia podgłośnia, co ma znaczny wpływ na jakość oraz wydolność głosu. Krtań z objawami zmęczenia głosowego cechuje zazwyczaj obrzęk, przekrwienie fałdów głosowych oraz niedomykalność głośni spowodowaną osłabieniem mięśni napinających fałdy głosowe. Zmęczeniu głosowemu po dłuższym i głośnym mówieniu towarzyszą subiektywnie odczuwalne dolegliwości np. suchość w gardle, okresowa chrypka, kaszel, spadek dynamiki głosu, zmiany barwy głosu, załamywanie się głosu, uczucie napięcia oraz sztywności mięśni krtań, gardła i szyi [2, 3]. Po odpoczynku dolegliwości te powinny ustąpić. Konsekwencją częstego przeciążania głosu mogą być stany patologiczne (np. guzki śpiewacze, zmiany przerostowe lub niedowłady fałdów głosowych), które utrudniają pełne zwarcie fonacyjne fałdów głosowych i wymagają interwencji chirurgicznej. W przypadku patologii narządu głosu wysiłek głosowy zazwyczaj nasila dysfunkcję (tzw. dysfonię).

Obiektywna ocena wpływu wysiłku głosowego na funkcję fonacyjną krtań, czyli ocena stopnia wydolności narządu głosu może być przeprowadzona na podstawie analizy sygnału mowy dźwięcznej, który reprezentuje wyizolowaną samogłoskę o przedłużonej fonacji [1, 2, 3]. Najczęściej rejestrowana jest samogłoska /a/ przed i po próbie obciążeniowej, która polega np. na głośnym czytaniu w czasie 1 h.

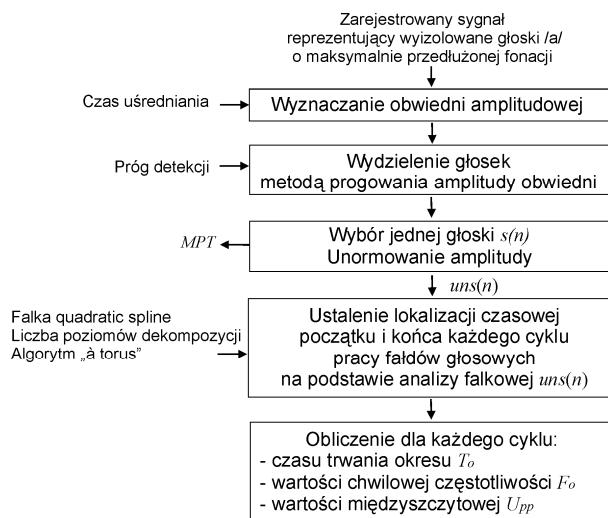
Pierwszymi oznakami nieprawidłowej adaptacji krtań do wysiłku głosowego są niestabilność częstotliwości podstawowej tonu krtańowego oraz natężenia głosu podczas przedłużonej fonacji. Do oceny zmienności częstotliwości F_0 lub okresu T_0 przyjmuje się parametry z tzw. grupy *Jitter*, zaś do oceny zmienności amplitudy sygnału mowy parametry z grupy *Shimmer* [2, 3]. Wspomniane parametry są zazwyczaj wyznaczone na podstawie analizy segmentów sygnału mowy, które obejmują co najmniej kilka cykli pracy fałdów głosowych.

W niniejszym artykule zaproponowano wykorzystanie do oceny zmęczenia głosowego różne parametry, które opisują zmienność okresu tonu krtańowego oraz amplitudy sygnału mowy z uwzględnieniem każdego cyklu pracy fałdów głosowych występujących podczas maksymalnie przedłużonej fonacji głoski /a/.

Opis metody

Na rysunku 1 przedstawiono algorytm opracowany do przetwarzania sygnału mowy, który reprezentuje wyizolowane głoski dźwięczne o maksymalnie przedłużonej

fonacji. Przyjęto założenie, że głoska /a/ o przedłużonej fonacji będzie wypowiediana przez badanego mówcę kilkakrotnie z zachowaniem odstępów czasu pomiędzy kolejnymi powtórzeniami. Algorytm umożliwia wyznaczenie czasu trwania okresu tonu krtańowego T_0 , a także wartości międzyszczytowej sygnału mowy znamiennych dla każdego cyklu pracy fałdów głosowych.



Rys. 1. Algorytm przetwarzania sygnału mowy

W ramach przetwarzania wstępnego z sygnału akustycznego zarejestrowanego mikrofonem wydzielane są tylko te segmenty, które reprezentują głoskę. Do realizacji tego zadania zastosowano metodę progowania obwiedni amplitudowej sygnału. Obwiednia amplitudowa jest wyznaczana za pomocą transformaty Hilberta (opisanej w pracy [4]), a następnie wygładzana metodą średniej ruchomej. Metoda progowania obwiedni amplitudowej sygnału umożliwia także automatyczne wyznaczenie czasu trwania fonacji głoski (MPT).

Natężenie głosu podczas wypowiedzania wyizolowanej głoski zależy od mówcy, dlatego sygnał mowy $s(n)$ reprezentujący tylko jedną, wybraną głoskę poddawany jest najpierw operacji unormowania amplitudy zgodnie ze wzorem:

$$(1) \quad uns(n) = \frac{s(n)}{\max\{|s(n)|\}}$$

gdzie $uns(n)$ oznacza sygnał o unormowanej amplitudzie, która przyjmuje wartości tylko z przedziału $[-1, 1]$.

W celu ustalenia lokalizacji czasowej początku i końca każdego cyklu pracy fałdów głosowych zastosowano metodę szczegółowo opisaną w pracy [5], która bazuje na analizie falkowej sygnału mowy dźwięcznej. Na podstawie wydzielonego segmentu sygnału mowy, który odpowiada tylko jednemu cyklowi pracy fałdów głosowych określa się wartość okresu tonu krtańowego (T_0), wartość chwilową częstotliwości (F_0) oraz wartość międzyszczytową sygnału mowy (U_{pp}).

Parametry przyjęte do oceny zmęczenia głosowego

Parametr aerodynamiczny MPT , definiowany jako maksymalny czas fonacji wyizolowanej samogłoski podczas pełnego wydechu, jest jednym z najlepszych parametrów stosowanych do oceny wydolności głosu. Prawidłowy czas fonacji powinien wynosić co najmniej 20 s [3]. Skrócony czas fonacji (tzn. wartość MPT poniżej 20 s) świadczy o mniejszej wydolności narządu głosu.

Do oceny zmienności okresu tonu krtańowego w czasie fonacji głoski /a/ przyjęto parametry $Jitta$ i $Jitter$, które są czułe na zmiany wartości T_0 zachodzące pomiędzy sąsiednimi cyklami pracy fałdów głosowych [6]:

$$(2) \quad Jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_0(i) - T_0(i+1)|$$

$$(3) \quad Jitter = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_0(i) - T_0(i+1)|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \cdot 100$$

gdzie N oznacza liczbę wszystkich cykli pracy fałdów głosowych występujących podczas przedłużonej fonacji.

Natomiast parametry RAP , $PPQ5$, $PPQ11$ określają względne zmiany wartości T_0 z różnym współczynnikiem wygładzania [6]:

$$(4) \quad RAP = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| T_0(i) - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_0(n) \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \cdot 100$$

$$(5) \quad PPQ5 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| T_0(i) - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} T_0(n) \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \cdot 100$$

$$(6) \quad PPQ11 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| T_0(i) - \left(\frac{1}{11} \sum_{n=i-5}^{i+5} T_0(n) \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \cdot 100$$

Ocenę zmienności okresu tonu krtańowego można także przeprowadzić na podstawie wykresu Poincaré, który tworzą punkty o współrzędnych $(T_0(i), T_0(i+1))$. Metoda ta jest często wykorzystywana w diagnostyce medycznej do oceny zmienności rytmu serca [7]. Na wykresie Poincaré wykreśla się elipsę, której półoś mała jest prostopadła do prostej $y=x$, a półoś wielka jest równoległa do prostej $y=x$. Długość półosi małej jest równa odchyleniu standardowemu $SD1$, które jest miarą rozproszenia odległości punktów wykresu od prostej $y=x$. Natomiast odchylenie standardowe $SD2$ będące miarą rozrzutu odległości punktów wykresu od prostej $y=-x+2T_0sr$ opowiada wartości półosi wielkiej.

W celu określenia względnych zmian amplitudy sygnału mowy (tj. natężenia głosu) przyjęto parametry, których wzory definicyjne podano poniżej [6]:

$$(7) \quad Shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |U_{pp}(i) - U_{pp}(i+1)|}{\frac{1}{N} \sum_{i=1}^N U_{pp}(i)} \cdot 100$$

$$(8) \quad APQ3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| U_{pp}(i) - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} U_{pp}(n) \right) \right|}{\frac{1}{N} \sum_{i=1}^N U_{pp}(i)} \cdot 100$$

$$(9) \quad APQ5 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| U_{pp}(i) - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} U_{pp}(n) \right) \right|}{\frac{1}{N} \sum_{i=1}^N U_{pp}(i)} \cdot 100$$

$$(10) \quad APQ11 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| U_{pp}(i) - \left(\frac{1}{11} \sum_{n=i-5}^{i+5} U_{pp}(n) \right) \right|}{\frac{1}{N} \sum_{i=1}^N U_{pp}(i)} \cdot 100$$

gdzie $U_{pp}(i)$ oznacza wartość międzyszczytową wyznaczoną dla i – tego cyklu pracy fałdów głosowych na podstawie unormowanej amplitudy sygnału mowy.

Przykładowe wyniki

Za pomocą mikrofonu i karty dźwiękowej (z 16-bitowym przetwornikiem A/C) zarejestrowano wyizolowaną głoskę /a/ o maksymalnie przedłużonej fonacji dla kilku nauczycieli akademickich przed i po zajęciach dydaktycznych. Mikrofon był umieszczony w odległości ok. 10 cm od ust mówcy. Podczas rejestracji każdy mówca co najmniej trzykrotnie wypowiadał głoskę /a/. Częstotliwość próbkowania sygnału mowy wynosiła 22 kHz.

W tabeli 1 podano wartości parametrów przyjętych do oceny zmęczenia głosu, które zostały wyznaczone dla kobiety w średnim wieku przed i po 8 godzinach prowadzenia zajęć laboratoryjnych z grupą studentów pierwszego roku w laboratorium z podstaw metrologii. Kierowanie pracą licznej grupy studentów, którzy nie posiadają żadnego doświadczenia w zakresie obsługi podstawowych przyrządów pomiarowych wymaga częstego udzielania wyjaśnień oraz głośnego mówienia, co powoduje znaczne obciążenie głosu.

Tabela 1. Wybrane parametry mówcy przed i po wysiłku głosowym

Parametr	Przed zajęciami	Po zajęciach
MPT [s]	18,8	10,2
$F_0 \dot{s}r \pm SD$ [Hz]	215 \pm 5	157 \pm 4
$T_0 \dot{s}r \pm SD$ [ms]	4,65 \pm 0,11	6,36 \pm 0,15
$Max T_0$ [ms]	5,16	6,84
$Min T_0$ [ms]	4,06	5,47
$Jita$ [μ s]	12	53
$Jitter$ [%]	0,26	0,83
RAP [%]	0,14	0,50
$PPQ5$ [%]	0,16	0,47
$PPQ11$ [%]	0,18	0,49
$SD1$ [ms]	0,008	0,033
$SD2$ [ms]	0,114	0,159
$SD2/SD1$	14,2	4,8
$U_{pp} \dot{s}r \pm SD$	0,41 \pm 0,13	0,45 \pm 0,17
$Shimmer$ [%]	1,95	4,45
$APQ3$ [%]	1,95	2,42
$APQ5$ [%]	1,98	2,66
$APQ11$ [%]	2,19	3,38

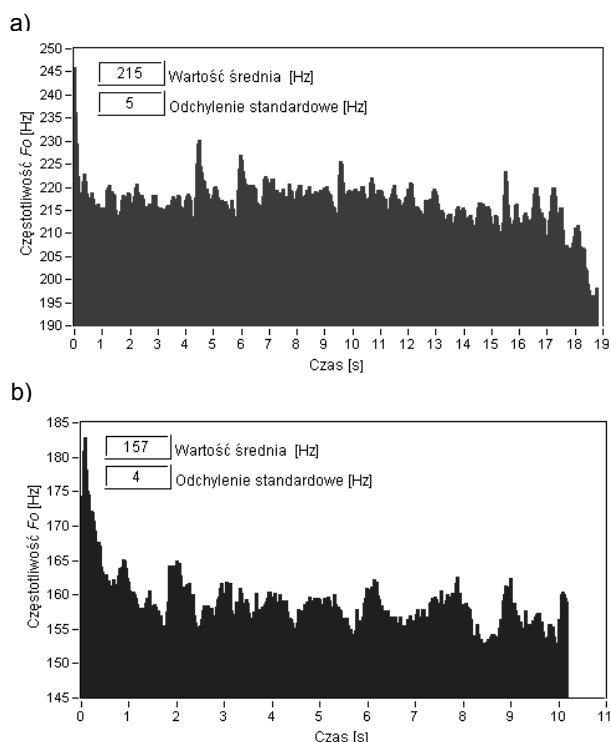
Oznaczenia: SD – odchylenie standardowe

Skrócony czas fonacji MPT po wysiłku głosowym (ponad 40%) świadczy o istotnym spadku wydolności głosu badanego mówcy.

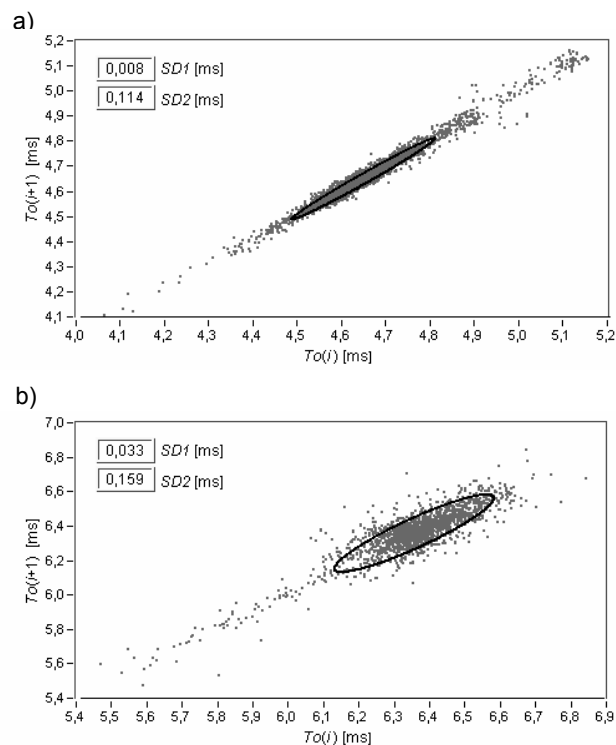
Wysiłek głosowy spowodował także wydłużenie czasu trwania cyklu pracy fałdów głosowych (tj. zwiększenie wartości $T_0 \dot{s}r$ oraz $Max T_0$ i $Min T_0$, które określają zakres zmian T_0). Wartości chwilowe częstotliwości podstawowej tonu kraniowego F_0 w czasie fonacji głoski /a/ przed i po zajęciach dydaktycznych zaprezentowano na rysunku 2. Wartości średnie ($F_0 \dot{s}r$) znacznie się różnią, zaś różnica

między odchyleniami standardowymi (SD obliczonymi dla częstotliwości F_0) nie przekracza 1 Hz.

O zwiększonych fluktuacjach wartości T_0 zachodzących w kolejnych cyklach pracy fałdów głosowych pod wpływem zmęczenia głosu świadczą parametry $Jita$ i $Jitter$ oraz parametry z grup $PPQn$ opisujące zmienność T_0 w sąsiednich n cyklach.



Rys. 2. Częstotliwość podstawowa tonu kraniowego podczas przedłużonej fonacji głoski /a/ zarejestrowanej: (a) przed i (b) po zakończeniu zajęć dydaktycznych



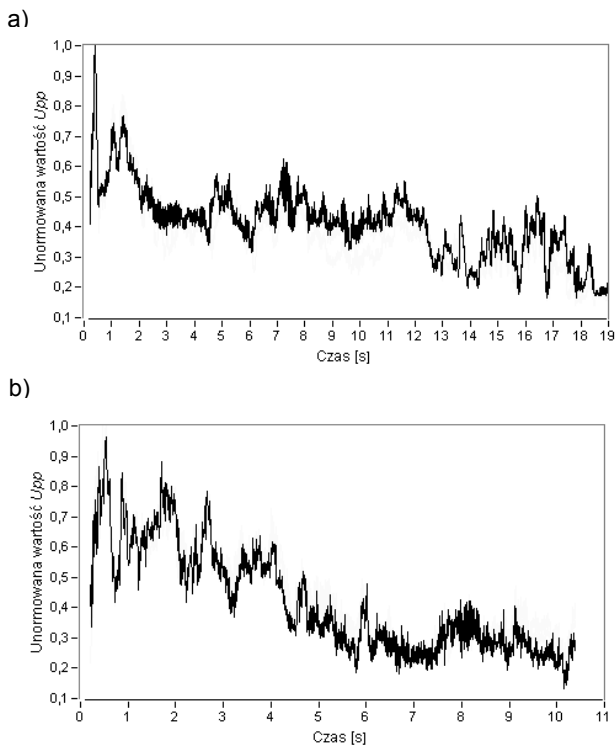
Rys. 3. Wykres Poincaré reprezentujący zmienność czasu trwania cyklu pracy fałdów głosowych podczas przedłużonej fonacji głoski /a/: (a) przed i (b) po zakończeniu zajęć dydaktycznych

Wykresy Poincaré przedstawione na rysunku 3 dobrze ilustrują zmienność okresu tonu krtniowego T_0 podczas przedłużonej fonacji. Parametry elipsy $SD1$ i $SD2$ opisują odpowiednio krótkoczasową i długoczasową zmienność okresu T_0 . „Kolisty kształt” elipsy na rysunku 3b świadczy o większej zmienności krótkoczasowej okresu T_0 w warunkach zmęczenia głosowego, tj. o większych fluktuacjach wartości T_0 w sąsiednich cyklach pracy fałdów głosowych niż przed wysiłkiem głosowym.

Parametry elipsy $SD1$ oraz $SD2$ warto wykorzystać do ilościowej oceny funkcji fonacyjnej krtni, bowiem dostarczają one więcej informacji o zmienności okresu tonu krtniowego niż typowe parametry statystyczne (np. wartość średnia i odchylenie standardowe).

Na rysunku 4 zaprezentowano wykresy wartości międzyszczytowej sygnału mowy wyznaczone dla każdego cyklu pracy fałdów głosowych, które uzyskano na podstawie unormowanej amplitudy sygnału reprezentującego badaną głoskę. Łatwo zauważyć, że w stanie zmęczenia głosowego mówca ma duże trudności z utrzymaniem stałego natężenia głosu podczas fonacji głoski.

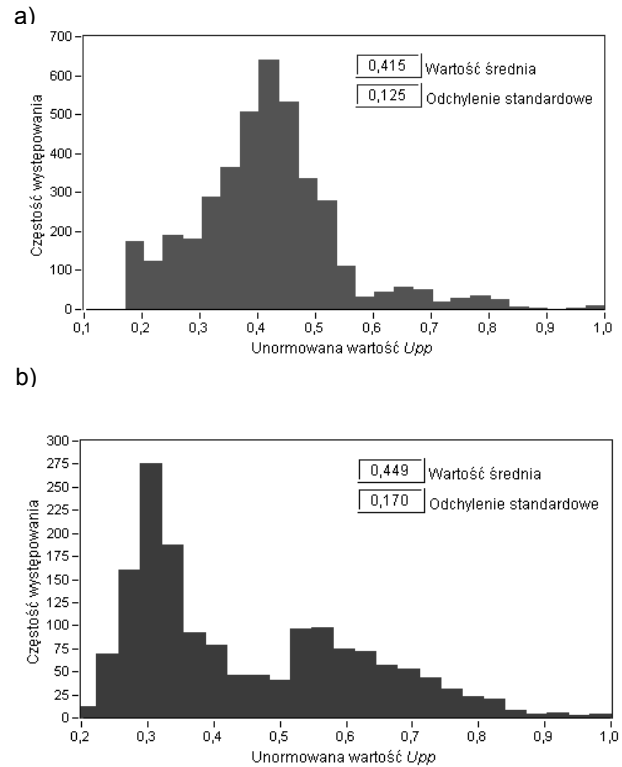
O znacznych fluktuacjach amplitudy sygnału mowy w czasie przedłużonej fonacji świadczą także duże wartości parametrów *Shimmer* i *APQ* podane w tabeli 1.



Rys. 4. Unormowana wartość międzyszczytowa wyznaczona dla każdego cyklu pracy fałdów głosowych podczas przedłużonej fonacji głoski /a/: (a) przed i (b) po zakończeniu zajęć dydaktycznych

Wstępne badania wykazały, że na podstawie kształtu histogramu reprezentującego częstość występowania unormowanej wartości międzyszczytowej sygnału mowy można wykryć stan zmęczenia głosowego.

Histogram znamieny dla stanu zmęczenia głosowego (rys.5b) cechują dwa „maksima”, które odpowiadają maksymalnej częstości występowania małych oraz dużych amplitud. Na podstawie histogramów znamienych dla stanu przed i po wysiłku głosowym można określić, w jakim stopniu zwiększyła się częstość występowania małych amplitud.



Rys. 5. Histogram reprezentujący zmienność unormowanej wartości międzyszczytowej wyznaczonej dla każdego cyklu pracy fałdów głosowych podczas przedłużonej fonacji głoski /a/: (a) przed i (b) po zakończeniu zajęć dydaktycznych

Podsumowanie

W artykule opisano metodę przetwarzania sygnału mowy, który reprezentuje wyizolowaną samogłoskę /a/ o maksymalnie przedłużonej fonacji. Metoda ta umożliwia wyznaczenie dla każdego cyklu pracy fałdów głosowych wartości chwilowych okresu i częstotliwości podstawowej tonu krtniowego, a także amplitudy sygnału mowy.

Przeprowadzono także ocenę zmienności wymienionych parametrów w warunkach przed i po obciążeniu głosowym za pomocą różnych metod. W dalszych badaniach warto wykorzystać parametry $SD1$ i $SD2$ obliczone na podstawie wykresu Poincaré, bowiem dobrze opisują krótkoczasową oraz długookresową zmienność okresu tonu krtniowego.

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

LITERATURA

- [1] Guzmán M. et al, Acoustic Changes of the Voice as Signs of Vocal Fatigue in Radio Broadcasters: Preliminary Findings, *Acta Otorrinolaringol Esp.*, 64 (2013), n.3, 176-183
- [2] Fujiki R. B., Sivasankar M. P., A Review of Vocal Loading Tasks in the Voice Literature, *Journal of Voice*, 31 (2017) n.3, 33-39
- [3] Niebudek-Bogusz E. i in., Zastosowanie krótkookresowej analizy wydolnościowej w diagnozowaniu zaburzeń głosu o podłożu zawodowym, *Medycyna Pracy*, 66 (2015), 225-234
- [4] Zieliński T. P., Cyfrowe przetwarzanie sygnałów – od teorii do zastosowań, WKiŁ, Warszawa, 2005
- [5] Wilk B., Wyznaczanie wartości chwilowej częstotliwości podstawowej tonu krtniowego za pomocą analizy falkowej sygnału mowy, *Przegląd Elektrotechniczny*, 91 (2015), nr.11, 305-308
- [6] Teixeira J.P., Fernandes P.O., Acoustic analysis of vocal dysphonia, *Procedia Computer Science* 64 (2015), 466-473
- [7] Acharya U.R. et al., Heart rate variability: a review, *Med. Bio. Eng. Comput.* 44 (2006), 1031-1051