

Ocena przydatności wybranych cech sygnału mowy wyróżniających osoby ze zmianami neurodegradacyjnymi

Streszczenie. W artykule przedstawiono problematykę oceny aparatu głosu u osób z chorobami neurodegradacyjnymi. W ramach badań dokonano opisu sygnału akustycznego w oparciu o parametry wyodrębnione przy użyciu powszechnie wykorzystywanych metod analizy akustycznej. Następnie przeprowadzono wstępną ocenę przydatności wyekstrahowanych cech z zastosowaniem wybranych miar statystycznych, w kontekście możliwości ich wykorzystania w systemach ukierunkowanych na wczesne wykrycie tzw. stanów otępiennych.

Abstract. The paper presents the problem related to the evaluation of speech organ in the context of persons with neurodegenerative changes. As part of the research, the acoustic signal was described based on the parameters extracted using commonly used methods of acoustic analysis. Next, a preliminary assessment of the usefulness of the extracted features with the use of selected statistical measures was carried out, in the context of the possibility of their use in systems aimed at early detection of the so-called dementia. The problem related to the evaluation of speech organ in the context of persons with neurodegenerative changes

Słowa kluczowe: analiza akustyczna głosu, choroby neurodegradacyjne, selekcja cech.

Keywords: acoustic analysis, neurodegradative diseases, feature selection.

Wprowadzenie

Choroby neurodegeneracyjne związane są z postępującymi uszkodzeniami komórek tworzących struktury układu nerwowego. Schorzenia te są groźne głównie z tego względu, że dojrzałe komórki układu nerwowego (z bardzo nielicznymi wyjątkami) nie wykazują tendencji do regeneracji. Uwaga badaczy koncentruje się więc na chorobach neurodegeneracyjnych z uwagi na fakt, że medycyna wciąż nie dysponuje zadowalającymi metodami ich leczenia. Choroby neurodegradacyjne cechują się postępującym przebiegiem, prowadzącym z czasem do znacznych ograniczeń w życiu codziennym chorych. Poszukiwane są zatem nowe metody ukierunkowane na wczesne wykrycie i precyzyjną diagnostykę tych zaburzeń. Szansą i dobrym wskaźnikiem wczesnej diagnostyki stanów otępiennych mogą być zmiany w głosie osób chorych [1-7]. Typowymi reprezentantami chorób neurodegradacyjnych są *choroba Alzheimera* oraz *choroba Parkinsona*, ale do grupy tych schorzeń zalicza się również: *otępienie z ciałami Lewy'ego*, *schorzenia prionowe* (inaczej *encefalopatie gąbczaste*, np. *choroba Creutzfeldta-Jakoba* czy *śmiertelna bezsenność rodzinna*), *otępienie czołowo-skroniowe*, *stwardnienie zanikowe boczne* i wiele innych [8].

Zaburzenia mowy w chorobach Parkinsona

W ramach prezentowanych w artykule badań autorzy skupili się na chorobie Parkinsona. Jest to heterogenny, przewlekły zespół hipokinetyczny o nieznanym etiologii, występujący u 3% osób powyżej 65 roku życia, dwa razy częściej u mężczyzn niż u kobiet. W Polsce zarejestrowano ponad 70 tysięcy chorych, a rocznie na chorobę Parkinsona zapada około 8 tysięcy osób [9]. Główną przyczyną patologii jest śmierć dużej liczby neuronów dopaminergicznych w istocie czarnej znajdującej się w śródmózgowiu. Objawy kliniczne występują z reguły, gdy ulegnie uszkodzeniu ok. 80% komórek odpowiedzialnych za wytwarzanie dopaminy. Przyczyny choroby Parkinsona są dotychczas nieznanymi poza przypadkami uwarunkowanymi genetycznie. Uważa się, że są one następstwem uszkodzeń toksycznych organizmu [3].

Jednym z pierwszych objawów świadczących o chorobie Parkinsona mogą być zmiany w głosie. Niestety lekarz najczęściej ich nie zauważa i ponieważ nie przebywa na co dzień z osobą badaną. W konsekwencji, ze względu na brak odniesienia, głos chorego nie wydaje mu

się w żaden sposób zmieniony. W chorobie Parkinsona w zakresie zmian aparatu głosu dochodzi do osłabienia mięśni krtani, gardła, podniebienia miękkiego, języka i ust. Mowę chorego charakteryzują zaburzenia oddechowo-artykulacyjno-fonacyjne, powstałe w wyniku uszkodzenia jąder podkorowych, przede wszystkim ciała prążkowanego i gałki bladej. Do zaburzeń tych należą m.in. *dyzartria*, *aprozodia* i *dysprozodia* (mowa monotonna, brak modulacji, zaburzenia rytmu i akcentowania), w tym *upośledzenie artykulacji głosek*, *zanikanie głosu*, *palilalia* (mimowolne powtarzanie słów lub sylab), *parafazje* (używanie fraz o podobnym brzmieniu zamiast właściwych), *dźwięczność dysząca* i *szorstka, różnice w akcencie kontrastowym*, *krótsze frazy*, *zwiększona długość pauz i hezycji* (tzw. *jęków namysłu*) oraz *zmienne tempo mowy*, w tym *bradyalalia* (spowolnienie mowy) w przypadku uszkodzenia gałki bladej i *tachylalia* (zwiększenie tempa mowy), jeśli uszkodzeniu uległo ciało prążkowane [10].

Baza danych

W pracy wykorzystano nagrania pochodzące z [11]. Zasadnicza baza danych to zbiór nagrań samogłosek *a*, *o* i *u* o czasie trwania od 1 do 4 s. Ciągłe samogłoski, pozbawione artefaktów językowych są wygodnym materiałem do analizy akustycznej głosu. Najczęściej wybiera się głoskę *a*, ponieważ według wielu badaczy najbardziej nadaje się ona do analizy tonu krtaniowego [12]. Wszystkie nagrania zostały zarejestrowane z szybkością próbkowania 50 kSa/s i rozdzielczością 16 bitów. W niniejszej pracy do analizy wykorzystano nagrania samogłoski *a* oraz *o* dla 20 osób. Bazę danych podzielono na 2 podgrupy: osoby zdrowe oraz osoby ze stwierdzoną chorobą Parkinsona.

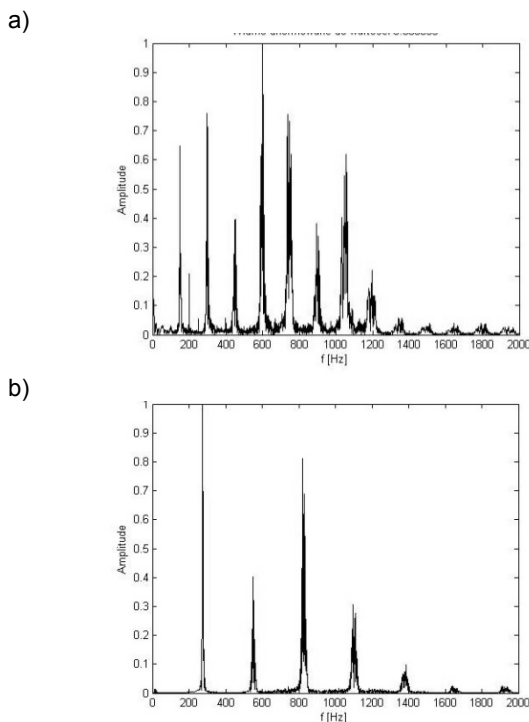
Obiektywna analiza akustyczna głosu

Obiektywna analiza akustyczna głosu ma na celu ocenę aparatu mowy i automatyczne rozpoznanie jego dysfunkcji. W odniesieniu do klasycznych metod diagnozowania choroby Parkinsona analiza tego typu pozwoli na obiektywną, dokładną ocenę różnych cech głosu, ocenianych dotąd jedynie subiektywnie w metodzie osłuchowej, a także na wykrywanie dyskretnych, niezauważalnych percepcyjnie zaburzeń głosu [13].

Analiza prowadzona w dziedzinie czasu służy głównie do oceny prozodycznych właściwości głosu, takich jak *akcent*, *iloczas*, *tempo* i *siła głosu*. Analiza w dziedzinie

czasu jest punktem wyjścia do dalszej analizy akustycznej mowy, ponieważ operuje na pierwotnej postaci zapisu sygnału mowy. W przypadku oceny pacjenta przy pomocy niezależnie wypowiedzianego przez niego tekstu można dodatkowo ocenić płynność mowy pacjenta korzystając z takich parametrów jak: *tempo mówienia*, *całkowity czas wypowiedzi* czy *czas trwania pauz pomiędzy wyrazami*.

Sygnał mowy w postaci czasowej jest kompletny, ale zbyt nadmiarowy, dlatego w celu parametryzacji poszukujemy skutecznych transformacji do innych dziedzin. Inspiracją funkcjonowaniem ludzkiego aparatu mowy oraz zmysłu słuchu prowadzi do metod w dziedzinie częstotliwości, bowiem w procesie artykulacji kształtowana jest głównie obwiednia amplitudowo-częstotliwościowa sygnału (poprzez odpowiednio formowaną strukturę rezonansową traktu głosowego), zaś w procesie percepcji, przed etapem analizy sygnału w sieciach neuronowych mózgu, następuje etap wydzielenia składowych o poszczególnych częstotliwościach przez wyspecjalizowane struktury ucha wewnętrznego. Ze względu na to, że analizę prowadzimy bezpośrednio na cyfrowej reprezentacji sygnału mowy, do analizy widmowej wykorzystujemy szybką transformację Fouriera. Przykładowe widma głoski *a* wypowiedziane przez zdrową i chorą osobę przedstawiono na rys. 1.

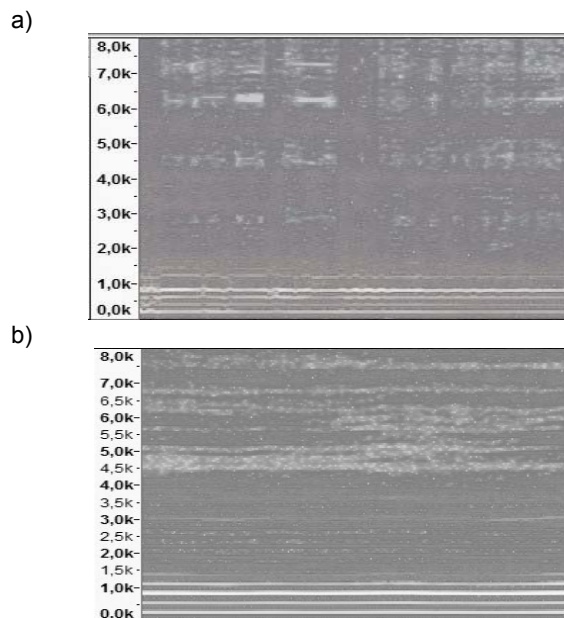


Rys. 1. Widmo pojedynczej głoski „a”: a) osoba chora, b) osoba zdrowa

Wadą klasycznej analizy widmowej jest brak jawnych informacji o czasie (z wyjątkiem informacji fazowych) w widmie sygnału. W analizie sygnału mowy, który można traktować jako ciąg pewnych zdarzeń (zmian częstotliwości, amplitudy, następstwa fonemów, słów itd.) informacja o czasie i kolejności tych zdarzeń jest bardzo istotna. Aby śledzić zmiany widma w funkcji czasu można zastosować *krótkoczasową transformację Fouriera*. Umożliwia ona wydobycie z sygnału informacji o tym, jak zmienia się jego widmo w czasie, czyli jednoczesną obserwację jego właściwości w obu dziedzinach.

Na rys. 2 przedstawiono przykładowe spektrogramy głosek wypowiedzianych przez kobietę zdrową oraz chorą.

Duża liczba szczegółów, jaką niesie ze sobą analiza spektralna utrudnia interpretację i rozpoznanie istotnych informacji zawartych w sygnale akustycznym mowy, dlatego podobnie jak w dziedzinie czasu, wyznacza się z przebiegu czasowo-częstotliwościowego te cechy, które są przydatne w analizowanym problemie. Na uzyskanych spektrogramach widać, że choroba Parkinsona wpływa na cały zakres częstotliwości, zaburzając ciągłość linii widmowych. Świadczy to o różnicach w poprawnym zamykaniu fałd głosowych. Im lepsza ciągłość linii widmowych tym lepsza praca fałd głosowych [14,15].



Rys. 2. Spektrogram pojedynczej głoski o długości 0.7 sekundy: a) osoba chora, b)- osoba zdrowa,

Ostatnim rodzajem analizy, wykorzystanym podczas prezentowanych badań jest analiza cepstralna, w której dzięki zastosowaniu operacji logarytmowania widma, multiplikatywne składowe ludzkiego głosu zmieniają relacje na addytywne, co umożliwia ich separację i odrębną analizę. Analiza cepstralna budzi coraz większe nadzieje na wykorzystanie kliniczne. W dużej metaanalizie badającej przydatność metod akustycznych w diagnozowaniu dysfonii, obejmującej 25 publikacji oraz 87 parametrów akustycznych międzynarodowy zespół autorów ocenił najwyżej współczynniki cepstralne, jako najlepiej korelujące ze stopniem dysfonii [16]. Odmianą tej metody jest metoda melcepstralna, wykorzystująca nieliniowość postrzegania dźwięku przez człowieka. Tzw. *skala perceptualna* wykorzystywana w tym algorytmie charakteryzuje się w dużym przybliżeniu liniowym odwzorowaniem częstotliwości niskich oraz logarytmicznym częstotliwości wysokich, co wynika z subiektywnego związku pomiędzy częstotliwościami czystych tonów harmonicznymi i częstotliwościami postrzeganymi przez człowieka.

Ekstrakcja cech

W oparciu o wyżej wymienione techniki przetwarzania sygnału mowy autorzy zdefiniowali zestaw 45 parametrów różnicujących przypadki prawidłowe i patologiczne. Ekstrakcji parametrów dokonano przy użyciu oprogramowania *Praat* oraz *Matlab*. Listę wszystkich 45 wyekstrahowanych parametrów zamieszczono w tabeli 1. Szerszy opis parametrów zawarty jest w [17-21].

Oceny jakości wyekstrahowanych parametrów (cech) można dokonać na wiele różnych sposobów. W ramach

badani zdecydowano się wykorzystać dwie podstawowe metody. Pierwszą z nich była ocena w oparciu o metodę Fishera, która bazuje na obliczeniu tzw. *miary Fishera* zdefiniowanej zależnością

$$(1) \quad S_{ij}(f) = \frac{|c_i - c_j|}{\sigma_i + \sigma_j}$$

gdzie c_i i c_j są wartościami średnimi rozpatrywanej cechy w grupie przypadków patologicznych i prawidłowych, natomiast σ_i oraz σ_j odpowiednimi odchyleniami standardowymi. Drugą metodą zastosowaną przez autorów była ocena parametrów przy użyciu transformacji PCA. Metoda ta miała na celu graficzną wizualizację i ocenę proponowanych cech.

Tab. 1. Lista wyekstrahowanych cech sygnału mowy

Cecha	Grupa cech
1. Odchylenie standardowe częstotliwości podstawowej dźwięku 2. Energia sygnału 3. Moc sygnału	Cechy częstotliwościowe (c_1-c_3)
4. Jitter 5. Jitter – RAP 6. Jitter PPQ	Cechy oceniające względną zmianę częstotliwości: (c_4-c_6)
7. Shimmer 8. Shimmer APQ3 9. Shimmer APQ5 10. Shimmer APQ11	Cechy oceniające względną zmianę amplitudy (c_7-c_{10})
11. Autokorelacja sygnału 12. Wartość skuteczna sygnału 13. Natężenie dźwięku	Cechy czasowe ($c_{11}-c_{13}$)
14. NHR (noise to harmonic rate), 15. HRN (harmonic to noise)	Cechy względnych pomiarów szumu i zakłóceń ($c_{14}-c_{15}$)
16-45. Cechy wyekstrahowane w oparciu o 30 banków filtrów	Cechy melcepstralne ($c_{16}-c_{45}$)

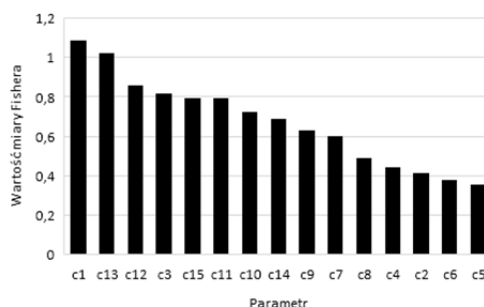
Badania przeprowadzono w dwóch rozłącznych zbiorach. Jedną grupę stanowiły parametry (cechy) wyekstrahowane w oparciu o analizę czasową i częstotliwościową (15 parametrów), drugą zaś parametry melcepstralne (30 parametrów). Na rys. 3 przedstawiono wartości miar Fishera w porządku malejącym osobno dla obydwu grup parametrów.

W grupie cech czasowych i częstotliwościowych największą zdolność dyskryminacyjną w analizowanym zbiorze posiada cecha c_1 będąca odchyleniem standardowym częstotliwości podstawowej dźwięku. Niestabilność częstotliwości podstawowej tonu to pierwsza oznaka nieprawidłowej adaptacji krtani do wysiłku głosowego u osób chorych. W konsekwencji osoby chore nie potrafią stabilnie wymawiać pojedynczych głosek, co uwidacznia się w analizie częstotliwościowej sygnałów.

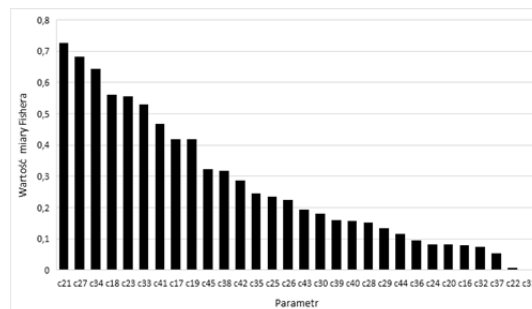
Parametry *Jitter*, *Shimmer* oraz częstotliwość podstawowa są bezpośrednio związane z drganiami fałdów głosowych. Są to najczęściej używane parametry w analizie aparatu głosu. Do oceny zmienności częstotliwości podstawowej przyjmuje się parametry z tzw. grupy *Jitter*, zaś do oceny zmienności amplitudy sygnału mowy parametry z grupy *Shimmer* [21]. Z przeprowadzonych badań wynika, iż parametry z grupy *Jitter* nie wnoszą zbyt dużo informacji dystynktywnej w analizowanym zbiorze. Lepszą jakością charakteryzują się parametry z grupy *Shimmer*. Oznacza to, że amplituda głosu u osób z chorobą Parkinsona nie jest stabilna, co spowodowane jest najczęściej dyszartrią będącą jedną z głównych zmian w mowie osób chorujących na to schorzenie. Dość dużą wartością miary Fishera charakteryzują się parametry

czasowe. Z punktu widzenia analizy sygnału głosowego parametry te determinują objawy dysfunkcji fałdów głosowych u osób chorych. Wybrane głosy charakteryzowały się chrypką, co bezpośrednio przełożyło się na uwydatnienie wartości cech w dziedzinie czasu. Również poziom głośności nagrań osób chorych oraz zdrowych zasadniczo się różnił, o czym świadczy duża wartość różnicująca w przypadku parametru c_{13} . W przypadku parametrów melcepstralnych występuje dość duże zróżnicowanie obliczonych miar Fishera dla wyekstrahowanych parametrów. Współczynniki melcepstralne niższych rzędów uwzględniają prace strun głosowych człowieka. Te wyższych rzędów związane są ze zmianami w samym kanale głosowym człowieka (geometrycznymi, masowymi, postaciowymi). Interesujący jest również fakt, iż zmiany zachodzące w widmie akustycznym sygnału w paśmie od 2125 Hz do 4375 Hz umożliwiają dostrzeżenie istotnych różnic pozwalających ocenić stan krtani chorego [22].

a)



b)



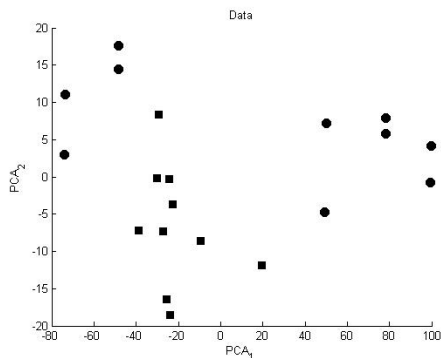
Rys. 3. Wykres sumarycznego współczynnika istotności Fishera dla cech wyekstrahowanych w poszczególnych dziedzinach

Końcowym etapem badań była wizualizacja danych otrzymanych w wyniku ekstrakcji i selekcji cech przy użyciu miary Fishera. Analizę przeprowadzono śledząc zmiany położenia poszczególnych wektorów określających mówcę na płaszczyźnie PCA_1/PCA_2 . Metodę tę zastosowano z uwagi na duży wymiar pełnego wektora cech. Na rys. 4 przedstawiono rozkład analizowanych przypadków z wykorzystaniem cech czasowych i częstotliwościowych (c_1-c_{15}). Na wykresach tych każdy z punktów reprezentuje jedną próbkę głosu. Punkty oznaczone kółkami reprezentują osoby zdrowe, natomiast oznaczone kwadratami osoby ze zdiagnozowaną chorobą Parkinsona.

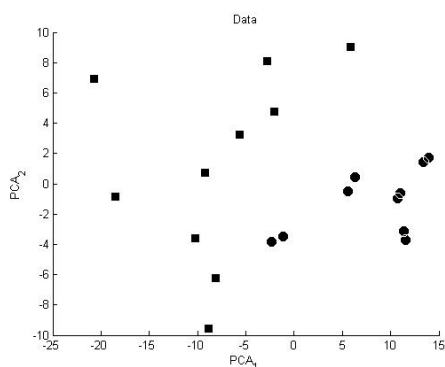
Z przeprowadzonych badań wynika, że ograniczenie liczby cech do tych posiadających dobrą zdolność dyskryminacyjną poprawia separację danych. Najbardziej widoczne jest to w przypadku parametrów melcepstralnych. Ponadto w przypadku parametrów czasowych oraz częstotliwościowych obserwuje się lepsze skupienie dla klasy przypadków prawidłowych. Klasa przypadków patologicznych (punkty w postaci kwadratów) jest względnie rozproszona, co świadczy może o zmienności parametrów

w tej klasie spowodowanych najprawdopodobniej różnym stopniem zaawansowania choroby Parkinsona.

a)



b)



Rys. 4. Rozkład danych zrzuconych na 2 najważniejsze składniki główne; a) pełny zbiór cech z grupy cech czasowych i częstotliwościowych, b) zbiór 10-ciu wyselekcjonowanych cech

Wnioski

Sygnał mowy niesie ze sobą nie tylko treść słowną oraz informacje o mówcy i emocjach towarzyszących danej wypowiedzi, ale również informacje o stanie drogi głosowej całego organizmu. Badania wykazały, że blisko 90% ludzi cierpiących na chorobę Parkinsona wykazuje pewne zmiany w mowie, będące jednocześnie niezauważalnymi przez bliskich objawami wskazującymi na rozwój tej choroby. W ramach badań wyekstrahowano szereg parametrów sygnału mowy oraz dokonano ich wstępnej oceny przy użyciu wybranych miar statystycznych. W tym celu posłużono się rankingową metodą Fishera oraz analizą składników głównych. Uzyskane wyniki potwierdzają skuteczność zarówno analizy czasowej, częstotliwościowej, jak i melcepstralnej w diagnostyce osób z chorobą Parkinsona. Ocena skuteczności wybranych cech będzie przeprowadzona podczas realizacji kolejnego etapu badań, z uwzględnieniem zaprojektowanego klasyfikatora i po poszerzeniu bazy sygnałów głosowych.

Praca jest efektem finansowania projektu „Badania podstawowe w obszarze technologii sensorowej z wykorzystaniem innowacyjnych metod przetwarzania danych” w programie Grant Badawczy MON GBMON/13-996/2018.

Autorzy: dr inż. Ewelina Majda-Zdancewicz, Wojskowa Akademia Techniczna, Wydział Elektroniki, ul. Gen. S. Kaliskiego 2, 00-908 Warszawa, E-mail: ewelina.majda@wat.edu.pl; prof. dr hab. inż. Andrzej P. Dobrowolski, Wojskowa Akademia Techniczna, Wydział Elektroniki, ul. Gen. S. Kaliskiego 2, 00-908 Warszawa, E-mail: andrzej.dobrowolski@wat.edu.pl.

LITERATURA

- [1] World Health Organization. Basic documents, Dementia - a public health priority. *Geneva: World Health Organization*, 2012.
- [2] Braaten AJ, Parsons TD, McCue R, Sellers A, Burns WJ., Neurocognitive differential diagnosis of dementing diseases: Alzheimer's dementia, vascular dementia, frontotemporal dementia, and major depressive disorder, *International Journal of Neuroscience*, 116(11), s.1271–1293, 2006;
- [3] Daria Biechowska, Edyta Orłowska, Neuropsychologiczna charakterystyka wybranych zespołów otępiennych, *Polski Przegląd Neurologiczny*, nr 8 (2), s. 66-75, 2012
- [4] Szepietowska EM, Daniluk B. Zaburzenia językowe w demencji w ujęciu neuropsychologii klinicznej, *Audiofonologia*, XVI, 2000, 117-134.
- [5] C. Stewart, et al., Speech dysfunction in early Parkinson's disease, *Movement Disorders*, nr. 10 (5), s. 562–565, 1995
- [6] T. Wasilewski, Speech impediments in alzheimer's disease, *Neuropsychiatry i Neuropsychologia*, 11(3), s. 93-98, 2016
- [7] B. Klimova, K. Kuc, Speech and language impairments in dementia, *Journal of Applied Biomedicine*, nr 14, 2, s. 97-103, 2016
- [8] World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. *Geneva: World Health Organization*; 1992
- [9] M. Rudzińska, I. Gatkowska, E. Mirek, A. Szczudlik, Choroba Parkinsona. Poradnik dla pacjentów i ich rodzin, *Wydawnictwo Uniwersytetu Jagiellońskiego*, s. 51Kraków 2009,
- [10] A. Krysiak, Zaburzenia języka, mowy i komunikacji w chorobie Parkinsona, *Neuropsychiatry i Neuropsychologia*; 6, 1: 36–42, 2011
- [11] Little Max A., McSharry P. E., Roberts S. J., Costello D. A. E., Moroz I. M, Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection, *BioMedical Engineering OnLine*, 2007.
- [12] E. Niebudek-Bogusz, Ocena parametrów analizy akustycznej głosu u kobiet zdrowych, *Otolaryngologia*, nr 3(1), s. 33-39, 2004
- [13] J. Mekyska i wsp., Robust and complex approach of pathological speech signal analysis, *Neurocomputing*, vol. 167, 1, pp. 94-111, 2015
- [14] P. Lieberman, Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges, *The Journal of the Acoustical Society of America*, nr 35(3), s. 344-353, 1963
- [15] S. Hadjitodorov, P. Mitev, A computer system for acoustic analysis of pathological voices and laryngeal diseases screening, *Medical Engineering & Physics*, nr 24, 6, s. 419-429, 2002
- [16] Y. Maryn i wsp., Acoustic measurement of overall voice quality: a meta-analysis, *The Journal of the Acoustical Society of America*, nr 126(5), s. 2619-34, 2009
- [17] Praat Shport Tutorial. *University of Toronto*, 2003
- [18] J. P. Teixeira, Accuracy of Jitter and Shimmer Measurements, *Procedia Technology*, nr 16, s. 1190 – 1199, 2014
- [19] S. Hadjitodorov, P. Mitev, A computer system for acoustic analysis of pathological voices and laryngeal diseases screening, *Medical Engineering & Physics*, nr 24, 6, s. 419-429, 2002
- [20] J. Mekyska i wsp., Robust and complex approach of pathological speech signal analysis, *Neurocomputing*, vol. 167, 1, pp. 94-111, 2015
- [21] J. Rafael et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, *Digital Signal Processing: A Review Journal*, 2017
- [22] M. Kłaczyński, Zjawiska wibroakustyczne w kanale głosowym człowieka, *Rozprawa doktorska*, Akademia Górniczo-Hutnicza w Krakowie, Kraków, 2007