

AGH

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI,
INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

KATEDRA METROLOGII I ELEKTRONIKI

Rozprawa doktorska

*Wykorzystanie sygnału mowy jako źródła informacji diagnostycznej,
kontrolnej i prognostycznej w wybranych problemach medycznych
związanych z otolaryngologią*

*The use of the speech signal as a source of diagnostic, control and
forecasting information in selected medical problems related to
otorhinolaryngology.*

Autor:

Daria Hemmerling

Kierunek studiów:

Biocybernetyka i Inżynieria Biomedyczna

Promotor:

prof. dr hab. inż. Janusz Gajda

Promotor pomocniczy:

dr inż. Andrzej Skalski

Kraków, 2018

Uprzedzona o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystycznego wykonania albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.): „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej «sądem koleżeńskim».”, oświadczam, że niniejszą rozprawę doktorską opracowałam osobiście i samodzielnie i że nie korzystałam(-am) ze źródeł innych niż wymienione w pracy.

Składam serdeczne podziękowania mojemu promotorowi prof. dr hab. inż. Januszowi Gajdzie za wszechstronną pomoc udzielaną mi podczas powstawania tej pracy, a także za życzliwość i miłą atmosferę. Dziękuję również promotowi pomocniczemu, dr inż. Andrzejowi Skalskiemu za nieustanne wsparcie, inspiracje oraz konsultacje naukowe. Serdeczne podziękowania składam również dr n. med. Magdalenie Wójcik-Pędziwiatr za wspaniałą współpracę podczas rejestracji głosu u pacjentów z chorobą Parkinson'a. Chcę podziękować również prof. dr hab. inż. Ryszardowi Tadeusiewicz za zainteresowanie problematyką mojej pracy doktorskiej.

Pragnę podziękować również wszystkim osobom, które wspierały mnie w trakcie prowadzenia badań naukowych: prof. dr hab. inż. Wiesławowi Wszótek, dr eng. Juan Rafael Orozco-Arroyave, prof. dr hab. n. med. Jackowi Składzień, dr hab. inż. Maciejowi Petko, dr n. med. Zygmuntovi Podolec.

Osobne, równie ważne podziękowania składam mojemu mężowi, za wszelką pomoc i motywację. Słowa podziękowania należą się także moim rodzicom za wiarę w moje możliwości.

Spis treści

1. Wprowadzenie	10
1.1. Cel pracy.....	12
1.2. Analiza stanu wiedzy.....	13
1.3. Struktura pracy.....	18
2. Proces generowania sygnału mowy i mowy zdeformowanej	20
2.1. Patologia procesu generowania sygnału akustycznego w ujęciu medycznym.....	22
2.2. Zaburzenia mowy w chorobie Parkinson'a.....	26
2.3. Skale kliniczne opisujące zaawansowanie choroby Parkinsona.....	27
3. Materiał i metoda badawcza	30
3.1. Materiał badawczy.....	30
3.2. Parametryzacja sygnału akustycznego.....	32
3.3. Informacja diagnostyczna w sygnale mowy.....	43
3.3.1. Nieliniowa analiza składowych głównych.....	44
3.3.2. Lasy losowe.....	46
3.3.3. Sieci neuronowe.....	47
3.4. Informacja kontrolna i prognostyczna.....	51
4. Wyniki badań	57
4.1. Algorytmiczna detekcja chorób laryngologicznych.....	57
4.2. Kontrola stanu neurologicznego u pacjentów z chorobą Parkinson'a.....	70
4.3. Predykcja stanu neurologicznego u pacjentów po spożyciu leków.....	73
5. Podsumowanie	83
Bibliografia	86
A. Dodatek A	93
B. Dodatek B	99
C. Dodatek C	106

Opisane w pracy wyniki powstały w ramach realizacji programu *Diamentowy Grant*, nr 0136/DIA/2013/42, AGH 68.68.120.36 (2014-2018) oraz w wyniku współpracy z:

- Katedrą i Kliniką Otolaryngologii Collegium Medicum, Uniwersytet Jagielloński w Krakowie (zgoda Komisji Bioetycznej Uniwersytetu Jagiellońskiego, nr 122.6120.213.2015),
- Krakowskim Szpitalem Specjalistycznym im. Jana Pawła II w Krakowie (zgoda Komisji Bioetycznej przy okręgowej Izbie Lekarskiej w Krakowie, nr 203/KBL/OIL/2016),
- Departamento de Ingeniería Electrónica, Universidad de Medellín w Kolumbii,
- Department of Computer Science, Friedrich-Alexander-University of Erlangen-Nürnberg w Niemczech,
- Dipartimento di Studi Letterari, Linguistici e Comparati, Università degli studi di Napoli L'Orientale we Włoszech.

Streszczenie

W niniejszej rozprawie przedstawiono możliwości wykorzystania metod przetwarzania i analizy sygnału mowy oraz stosowanych narzędzi, metod klasyfikacji oraz regresji w kontekście podejmowania decyzji terapeutycznych związanych z dysfunkcją narządu głosu oraz wybranymi schorzeniami neurologicznymi. Głównym celem pracy było przeprowadzenie badań zmierzających do ustalenia, w jakim stopniu zmiany w sygnale mowy, dające się wykryć i ilościowo określić na podstawie analiz fonetyczno-akustycznych tego sygnału, mogą być źródłem informacji diagnostycznej, prognostycznej i kontrolnej, przy rozwiązywaniu niektórych problemów związanych z wybranymi chorobami w otolaryngologii. Uzyskane wyniki posłużyły do stworzenia systemu, którego celami były: automatyczna detekcja chorób laryngologicznych, prognoza postępu choroby Parkinson'a, estymacja punktacji w skali UPDRS opisującej nasilenie objawów choroby Parkinson'a. Dane niezbędne do zrealizowania wymienionych celów pracy zostały pozyskane z dwóch źródeł: z bazy *Saarbruecken Voice Database* zawierającej nagrania osób, których ojczystym językiem był język niemiecki oraz bazy nagrań pacjentów cierpiących z powodu choroby Parkinson'a zarejestrowanych w ramach współpracy z Krakowskim Szpitalem Specjalistycznym im. Jana Pawła II mówiących w języku polskim. Opis sygnału mowy w postaci wektora parametrów akustycznych oraz wykorzystanie metod uczenia maszynowego pozwoliło na poszukiwanie w sygnale charakterystycznych symptomów diagnostycznych wskazujących na obecność określonej choroby laryngologicznej lub jej braku oraz dodatkowo na monitorowanie przebiegu choroby Parkinson'a w sposób automatyczny. Wyniki analizy akustycznej zostały skorelowane z klinicznym opisem rozpoznania wybranych chorób i ich stopnia zaawansowania. W rezultacie możliwe było stworzenie oprogramowania wspomagającego pracę lekarza-klinicysty w zakresie diagnostyki, monitorowania terapii wraz z ilościową oceną postępu procesu leczenia oraz prognozą skutków tej terapii.

Abstract

This thesis presents the usage possibilities of speech signal processing and its analysis as well as applied tools, classification methods and regression to help therapeutic decisions related to the vocal organ dysfunction and selected neurological disorders. The main objective of this work was to conduct studies to determine the changes in the speech signal, which are detectable and quantitatively determined on the basis of phonetic and acoustic analyzes of this signal, may be the source of diagnostic, prognostic and control information in solving some problems related to selected diseases in otolaryngology. The obtained results were used to create a system whose goals were: automatic detection of laryngeal diseases, the prognosis of Parkinson's disease progression, UPDRS score estimation describing the severity of Parkinson's disease symptoms. The data necessary to achieve mentioned goals was obtained from two sources: from the *Saarbruecken Voice Database* containing the recordings of people whose native language was german and the database of patients suffering from Parkinson's disease registered in cooperation with the John Paul II Hospital in Krakow, all speaking in polish. The speech signal is described by a vector composed of acoustic parameters. The use of machine learning methods allowed the search for characteristic diagnostic symptoms indicating the presence of a specific laryngological disease or its absence, and in addition to monitor the course of Parkinson's disease in an automatic manner. The results of the acoustic analysis were correlated with a clinical description of the selected diseases diagnosis and their severity. As a result, it was possible to create software supporting the physician-clinician work in the field of diagnostics, therapy monitoring along with a quantitative assessment of the treatment progress and prognosis of the effects of this therapy.

SPIS NAJWAŻNIEJSZYCH OZNACZEŃ I SKRÓTÓW

- x - dyskretny, poddany procesowi segmentacji i okienkowania sygnał akustyczny, dB
- F_0 - częstotliwość podstawowa, Hz
- J - współczynnik jitter, %
- S - współczynnik shimmer %
- A_i - amplituda tonu podstawowego w kolejnych fragmentach sygnału, dB
- f_i - częstotliwość tonu podstawowego w kolejnych fragmentach sygnału, Hz
- E - energia sygnału, dB
- M_m - momenty widmowe m -tego rzędu ($m=1,2,3$), dB
- F_m - częstotliwość formantu m -tego rzędu ($m=1,2,3$) [Hz]
- $G(t, f)$ - widmo czasowo-częstotliwościowe sygnału mowy, dB
- HNR - ang. *harmonic to noise ratio*, stosunek części harmonicznej do szumu, dB
- MFCC - ang. *mel - frequency cepstral Coefficients*, współczynniki mel-cepstralne
- FFT ang. *fast Fourier transform*, szybka transformacja Fouriera
- IFFT - ang. *inverse fast Fourier transform*, odwrotna transformacja Fouriera
- DCT - ang. *discrete cosine transform*, dyskretna transformacja kosinusowa
- PCA - ang. *principal component analysis*, analiza składowych głównych
- kPCA - ang. *kernel principal component analysis*, jądrowa analiza składowych głównych
- RF - ang. *random forest*, metoda lasów losowych NLPCA - nieliniowe PCA
- SVR - ang. *support vector regression*, regresja wektorów nośnych
- UPDRS - ang. *Unified Parkinson's Disease Rating Scale*, skala wykorzystywana do oceny postępu choroby Parkinson'a

1. Wprowadzenie

Nieodłącznym elementem funkcjonowania człowieka w społeczeństwie jest proces komunikacji werbalnej. Proces ten opiera się na generowaniu i odbiorze bodźców akustycznych. Narząd mowy jest wyspecjalizowanym układem umożliwiającym generowanie różnych dźwięków poprzez regulację strumienia powietrza wypływającego z płuc i tym samym formowanie dźwięków w postaci zmian chwilowego ciśnienia. Sygnał akustyczny jest odbierany za pomocą narządu słuchu.

Sygnał mowy niesie ze sobą wiele informacji - możliwe jest zrozumienie wypowiedzianej treści, rozróżnienie płci, stanu emocjonalnego i psychicznego. Niekiedy można również określić status społeczny, pochodzenie, wiek. Płód człowieka już w 24 tygodniu od poczęcia potrafi rozpoznać głos matki, reaguje na zmiany tonu i barwy dźwięku. Od tego momentu do końca życia uczymy się jak przetwarzać dźwięki i odpowiednio je interpretować. Prowadzone są prace badawcze nad urządzeniami technicznymi, które mają dorównać funkcjonalnością niezwykle wyspecjalizowanemu narządowi słuchu, który odbiera bodźce i przekazuje je do mózgu.

Sygnał akustyczny był wykorzystywany w diagnostyce medycznej jeszcze przed rozwojem technicznych metod jego analizy. Zmiany głosu towarzyszą rozwojowi chorób laryngologicznych i mogą być spowodowane różnorodnymi czynnikami takimi jak: wysiłek głosowy, infekcje górnych dróg oddechowych, stosowane używki, inne choroby (endokrynologiczne, układu oddechowego i krążenia) i wiele innych. Sygnał mowy stanowi źródło informacji na temat kondycji narządów wewnętrznych człowieka. Uzasadnia to prowadzenie analizy tego sygnału w wybranych celach diagnostycznych. Analizę akustyczną można przeprowadzić wykorzystując: subiektywną ocenę głosu i/lub cyfrową analizę sygnałów akustycznych, spektrografię i sonografię. O kondycji narządu mowy można również wnioskować na podstawie obrazów przedstawiających funkcjonowanie traktu głosowego.

W prezentowanej rozprawie przedstawiono możliwości wykorzystania metod przetwarzania i analizy sygnału mowy oraz stosowanych narzędzi, metod klasyfikacji oraz regresji w kontekście podejmowania decyzji terapeutycznych związanych z dysfunkcją narządu mowy oraz wybranymi schorzeniami neurologicznymi.

W medycynie nie istnieje jeden wzorzec danej jednostki chorobowej, analizie poddaje się zespół objawów występujących u pacjenta i odnosi się to do określonego opisu konkretnej choroby. Rozwój technologiczny pozwala na pomiar coraz większej liczby cech i wykorzystanie uzyskanych wyników do zbudowania pełnego obrazu klinicznego danej jednostki chorobowej. Podejmowanie decyzji klinicznych uwzględniających diagnozę i wybór procedury leczenia jest złożonym i trudnym procesem. Aby ułatwić to zadanie opracowywane są narzędzia do automatycznej analizy otrzymanyh wyników oraz wspomaganie decyzji diagnostycznych i terapeutycznych. Zazwyczaj takie systemy wykorzystują odległość pomiędzy zespołem parametrów występujących u chorego oraz cechami charakterystycznymi dla określonej jednostki chorobowej. Systemy te umożliwiają ocenę stanu chorobowego poprzez porównanie i odniesienie do wartości parametrów przypisanych danej chorobie.

Obecnie zespoły naukowców pracują nad opracowaniem optymalnej metodyki pomiarowej, parametrami wykorzystywanymi do diagnostyki oraz algorytmami umożliwiającymi jej przeprowadzenie. W konsekwencji, prowadzi to do próby odpowiedzenia na pytanie, czy w badanym sygnale głosu występują cechy patologiczne, czy też dany sygnał należy do grupy sygnałów akustycznych prawidłowych [1–4]. Postęp w rozwoju cyfrowych metod przetwarzania sygnałów umożliwia wyznaczenie coraz większej liczby mierzalnych parametrów ludzkiego głosu, a także pozwala na dokładną, obiektywną ich ocenę. Takie parametry niekoniecznie uwiadaczają się w metodzie odsłuchowej stosowanej w ocenie patologii głosu oraz mowy. Systemy wspomaganie diagnostyki mogą pomóc w analizie ogromnej ilości danych i wyników badań umożliwiając tym samym wyciąganie użytecznych wniosków dotyczących danej osoby. Co więcej, parametry sygnału akustycznego wraz z odpowiednimi narzędziami algorytmicznymi mogą posłużyć do prognozy rozwoju choroby u pacjenta.

Prognozowanie w medycynie stanowi niezwykle ważny aspekt w planowaniu terapii pacjenta i w ocenie rokowań. Wybór metody oraz procedury leczenia polega na prognozowaniu jej efektów oraz na porównaniu jej z innymi procedurami leczenia, a następnie wyborze takiej metody, której skutki będą najbardziej pożądane. Podjęcie trafnej decyzji terapeutycznej może skrócić czas terapii i rekonwalescencji pacjenta, jak również pozwolić na szybszą reakcję w przypadku niepożądanych skutków terapii. Poprawne prognozowanie pozwala lepiej zrozumieć pacjentowi jego sytuację, dostrzec możliwe korzyści i efekty planowanej terapii oraz umożliwia na świadome podjęcie ewentualnego ryzyka.

Podstawą każdej prognozy skutków leczenia jest określenie zmiennych prognostycznych lub czynników ryzyka wpływających na wynik końcowy. Dlatego też predykcja zwykle oparta jest na analizie statystycznej zmierzającej do identyfikacji i wyboru takich czynników, które mają

największe znaczenie prognostyczne. W praktyce klinicznej wybór takich zmiennych i czynników odbywa się na podstawie obserwacji, teorii tłumaczącej patogenezę występującego schorzenia oraz na podstawie wyników długotrwałych badań. Wybór następuje w wyniku porównań prowadzonych wśród danych grup chorych w celu wykrycia statystycznie istotnych różnic dotyczących przebiegu choroby między różnymi pacjentami. W rzeczywistości, na ostateczny wynik leczenia wpływa wiele zmiennych i czynników, dlatego też w modelach predykcyjnych uwzględnia się ich wzajemne oddziaływanie. Podstawę predykcji stanowią dane historyczne, tj. informacja o przebiegu w przeszłości tej samej choroby u różnych pacjentów. Obecnie istnieje wiele modeli predykcyjnych, na przykład modele regresyjne liniowe lub nieliniowe. Pozwalają one m.in. wyznaczyć prawdopodobieństwo wystąpienia określonego zdarzenia. W medycynie modele predykcyjne są na ogół skomplikowane i najczęściej charakter zależności pomiędzy zmiennymi wejściowymi i wyjściowymi nie jest znany w sposób jawny. Wówczas wykorzystuje się metody oparte na tzw. uczeniu maszynowym (ang. *machine learning*), które umożliwiają poszukiwanie i analizę zależności oraz badanie i opracowywanie złożonych problemów w dużych bazach danych.

1.1. Cel pracy

Głównym celem pracy było przeprowadzenie badań zmierzających do ustalenia, w jakim stopniu zmiany w sygnale mowy, dające się wykryć i ilościowo określić na podstawie analiz fonetyczno-akustycznych tego sygnału, mogą być wykorzystane jako źródło informacji diagnostycznej, prognostycznej i kontrolnej przy rozwiązywaniu niektórych problemów medycznych związanych z wybranymi chorobami w otolaryngologii. Uwagę skupiono na możliwościach wykorzystania wyników analizy sygnału mowy do stworzenia systemu, którego celami działania są:

- automatyczna detekcja chorób laryngologicznych,
- prognoza postępu choroby Parkinson’a,
- estymacja punktacji w skali UPDRS ¹ opisującej nasilenie objawów choroby Parkinson’a.

Opis sygnału mowy w postaci wektora parametrów akustycznych oraz wykorzystanie metod uczenia maszynowego pozwoliło na poszukiwanie w sygnale charakterystycznych symptomów diagnostycznych wskazujących na obecność określonej choroby laryngologicznej lub jej braku oraz dodatkowo na monitorowanie przebiegu choroby Parkinson’a w sposób automatyczny. Wyniki analizy akustycznej zostały skorelowane z klinicznym opisem rozpoznania

¹*Unified Parkinson’s Disease Rating Scale*, Ujednolicona Skala Oceny Choroby Parkinsona

wybranych chorób i ich stopnia zaawansowania. W rezultacie możliwe było stworzenie oprogramowania wspomagającego pracę lekarza-klinicysty w zakresie diagnostyki, monitorowania terapii wraz z ilościową oceną postępu procesu leczenia oraz prognozą skutków tej terapii.

Dane niezbędne do zrealizowania wymienionych celów pracy zostały pozyskane z różnych źródeł. Pierwszym z nich była baza *Saarbruecken Voice Database* udostępniona przez Instytut Fonetyki w Saarland w Niemczech. Dane te posłużyły do utworzenia systemu do automatycznej detekcji chorób laryngologicznych.

Drugim źródłem danych była baza nagrań pacjentów cierpiących z powodu choroby Parkinson'a zarejestrowanych w ramach współpracy z Krakowskim Szpitalem Specjalistycznym im. Jana Pawła II.

1.2. Analiza stanu wiedzy

Podczas konferencji *Conference on Early Detection of Laryngeal Pathology* w 1973 roku profesor G. Paul Moore podkreślił publicznie potrzebę stworzenia narzędzi do analizy akustycznej sygnału głosu w celu przeprowadzenia oceny klinicznej głosu dla potrzeb diagnostyki laryngologicznej [5]. Zaletą takiego podejścia jest nieinwazyjność, niski koszt urządzeń oraz krótki czas potrzebny na analizę. Przetwarzanie sygnału akustycznego głosu nadal stanowi wyzwanie naukowe ze względu na fakt, że jego brzmienie jest zróżnicowane międzyosobniczo, zależy od języka ojczystego, intonacji wymowy, towarzyszących emocji oraz współistniejących chorób. W ocenie różnicowania zaburzeń głosu można wykorzystać parametry, za pomocą których możliwe jest wykrywanie niestabilności, zawartości szumów oraz stwierdzenie czy występują problemy artykulacyjne. Istotnym elementem naukowym jest również powiązanie wyników analizy z określoną jednostką chorobową i jej nasileniem, a także przewidywanym rozwojem. Najczęściej wykorzystywanymi parametrami jest częstotliwość podstawowa, współczynnik jitter oraz shimmer, stosunek harmonicznym do szumu, współczynniki mel-cepstralne (MFCC) [6]. W ostatnich latach badania dotyczące analizy akustycznej obejmowały głównie tworzenie algorytmów do automatycznej detekcji oraz klasyfikacji wybranych chorób głosu z wykorzystaniem parametrycznych i nieparametrycznych cech, algorytmów do rozpoznawania wzorców (ang. *pattern recognition*) oraz metod statystycznych. Algorytmy do detekcji patologii mowy najczęściej wykorzystują nagrania samogłosek o przedłużonej fonacji ze względu na ich stabilność w czasie oraz łatwość wypowiedziania przez pacjenta [7].

Algorytmy, których celem jest automatyczna detekcja chorób głosu, wymagają wcześniejszego przygotowania reprezentatywnej bazy nagrań pacjentów z określoną jednostką chorobową narządu głosu oraz osób zdrowych. Analiza głosu rozpoczyna się od przygotowania opisu sygnału zarejestrowanego przez mikrofon przy pomocy wybranych parametrów. Taki opis powinien uwzględniać parametry, za pomocą których możliwa jest detekcja wystąpienia lub braku

objawów chorób głosu. Zebranie odpowiednio licznej bazy nagrań głosu wraz z opisem odpowiednich jednostek chorobowych jest zadaniem trudnym i długotrwałym ze względu na ograniczoną dostępność pacjentów. Ponadto wymagana jest współpraca z lekarzami. Z tego powodu większość artykułów w literaturze przedmiotu przedstawia badania z uwzględnieniem ograniczonej liczby nagrań. Autorzy publikacji [8] poddali analizie nagrania głosu 30 pacjentów, wśród których u 24 osób wystąpiła dysfonia głosu. Opis cech uwzględniał widmo długoterminowe LTAS (ang. *long – term average spectrum*), współczynnik zawartości szumów oraz liniową analizę dyskryminacyjną, która następnie była wykorzystana w regresji logistycznej. W wyniku otrzymano 100% dokładność klasyfikacji osób z dysfonią oraz osób zdrowych. Nieco gorszy wynik klasyfikacji, wynoszący 96,1% osób zdrowych i chorych, lecz uwzględniający zdecydowanie bardziej liczną grupę pacjentów, przedstawia praca [9]. Parametrami opisującymi sygnał akustyczny były parametry takie jak średnia częstotliwość podstawowa, jej odchylenie standardowe, współczynniki jitter i shimmer, współczynnik harmonicznych do szumu w dziedzinie czasu i częstotliwości, znormalizowana energia szumu (NNE). Dodatkowo wyprowadzono takie parametry jak: współczynnik szumu turbulentnego (ang. *turbulent noise index*) oraz znormalizowaną energię pierwszej harmonicznej. Klasyfikację wykonano za pomocą algorytmu *k*-najbliższych sąsiadów (ang. *k-nearest neighbours*). Nagrania wykorzystane na poczet tej pracy pochodziły z bazy komercyjnej *Massachusetts Eye and Ear Infirmary* (MEEI) [10] i dotyczyły nagrań samogłoski /a/ 744 pacjentów, z których 638 miało stwierdzoną patologię głosu. Nieco inne podejście zaprezentowano w pracy [11], w której do klasyfikacji wykorzystano sieci neuronowe. Na wejście sieci podano parametry fazowe (ang. *phase-based*) sygnału głosu otrzymując dokładność klasyfikacji na poziomie 95,92% dla 710 pacjentów (657 z chorobami głosu). Badania również wykorzystywały nagrania samogłoski /a/. W innej publikacji [12] baza MEEI została opisana za pomocą 30 parametrów akustycznych, które następnie zostały poddane analizie składowych głównych (ang. *principal component analysis*, PCA). Klasyfikację wykonano wykorzystując algorytm SVM (ang. *support vector machines*). W wyniku 5-krotnego sprawdzianu krzyżowego (krosvalidacji) dokładność klasyfikacji osób zdrowych i osób z chorobami laryngologicznymi oceniono na 98,1%. Nagrania dotyczyły również samogłoski /a/ o przedłużonej fonacji, wypowiedzianej przez 308 osób, z czego 269 stanowiła grupa chorych. Ta sama baza została zaimplementowana w kolejnych badaniach [13]. Wówczas wykorzystano krótkoczasową transformatę Fouriera oraz transformatę falkową. Liniowa analiza dyskryminacyjna oraz analiza SVM zapewniły rozpoznanie nieprawidłowości w głosie ludzkim na poziomie 100%. Algorytm ten był testowany na nagraniach samogłoski /a/ dla 53 zdrowych oraz 67 chorych osób.

Drugą obok bazy MEEI, powszechną bazą nagrań głosu osób zdrowych oraz z chorobami laryngologicznymi jest baza Saarbruecken Voice Database (SVD) [14]. Baza ta została wykorzystana przez autorów publikacji [15], która przedstawia wykorzystanie współczynników

MFCC, stosunku harmonicznego do szumu, znormalizowanej energii szumu oraz współczynnika *glottal – to – noise excitation ratio*. Cechy te były zaimplementowane w modelach mieszanin gaussowskich (ang. *gaussian mixture models*, GMM). Analizę przeprowadzono na podstawie 1970 nagrań samogłoski /a/ wypowiedzianej w normalnej intonacji (1320 osób stanowiły osoby chore, a 650 osoby zdrowe). Analogiczną analizę przeprowadzono dla nagrań 226 osób (173 stanowiły osoby chore) z bazy MEEI. Takie podejście zapewniło dokładność klasyfikacji pacjentów zdrowych i chorych na poziomie 89,2%. Badania te zostały rozwinięte i opublikowane w pracy [16], w której autorzy obliczają te same parametry co w [15], wykorzystując nagrania samogłosek w różnych intonacjach (stabilnej normalnej, wysokiej, niskiej oraz modulowanej: niska-wysoka-niska). Do klasyfikacji wykorzystano algorytm *score level fusion*, w wyniku czego skuteczność wykrycia nieprawidłowości w głosie została oceniona na 94,93% dla osób chorych. Skuteczność wykrycia osób zdrowych przy wykorzystaniu tego samego algorytmu wyniosła 95% [15]. W tym przypadku wykorzystano 1970 nagrań głosu samogłosek /a/, /i/, /u/ w normalnej, wysokiej oraz niskiej intonacji. Nagrania samogłoski /a/ także w trzech intonacjach pochodzących od 1360 pacjentów są przedmiotem analizy w pracy [17]. Autorzy przedstawiają wektor 31 parametrów akustycznych, które następnie zostały poddane analizie składowych głównych w celu redukcji wymiarowości. Klasyfikację oparto na algorytmie *k*-średnich. Do walidacji utworzonego algorytmu wykorzystano 10-krotny sprawdzian krzyżowy. Klasyfikację przeprowadzono osobno dla mężczyzn i kobiet. W efekcie, 100% wszystkich mężczyzn oraz 81,4% kobiet, które wypowiedziały samogłoskę /a/ w normalnej intonacji, zostało poprawnie zaklasyfikowanych odpowiednio do grupy zdrowych lub chorych. Nagrania samogłoski /a/ z trzech różnych baz (SVD, MEEI oraz *Arabic Voice Pathology Database* AVPD) zostały wykorzystane w kolejnej pracy w celu stworzenia algorytmu do automatycznej detekcji osób, u których występuje patologia głosu oraz osób zdrowych z punktu widzenia otolaryngologicznego [18]. Ekstrakcja cech akustycznych została przeprowadzona przy użyciu programu *Multidimensional Voice Program (Kay Pentax CSL, Model 4300)*. Wybór parametrów był oparty na analizie dyskryminacyjnej Fishera, a klasyfikacji wykorzystano algorytm SVM. W wyniku otrzymano dokładność klasyfikacji (stosunek wszystkich przypadków poprawnie zaklasyfikowanych do wszystkich przypadków zaklasyfikowanych) na poziomie 89,5% dla bazy MEEI, 99,68% dla bazy SVD oraz 71,6% dla bazy AVPD. Praca [19] opisuje klasyfikację nagrań pochodzących z 4 baz: MEEI, SVD, AVPD oraz PDA (*Principe de Asturias Database*). Do badań wykorzystano łącznie: 8042 nagrania samogłoski /a/. Dokładność klasyfikacji była na poziomie 73,3% przy wykorzystaniu algorytmu XGBoost oraz 9 parametrów akustycznych. Inne podejście przedstawiają autorzy pracy [20]. Do badań wykorzystano 3 bazy nagrań głosu: MEEI, SVD oraz AVPD. W pracy wykorzystano entropię oraz autokorelację wyznaczoną dla różnych zakresów częstotliwości sygnałów. Dokładność klasyfikacji wyniosła odpowiednio: 99,69%, 92,79%, and 99,79% dla poszczególnych baz.

Większość badań przedstawionych w literaturze opisuje algorytmy, które zostały przetestowane na nierównolicznych grupach pacjentów, niektóre badania nie uwzględniają wystarczającej liczby nagrań warunkujących uzyskanie wiarygodnego wyniku. Opracowanie algorytmu, który potrafiłby w sposób automatyczny wskazać osoby zdrowe oraz posiadające jakąś nieprawidłowość laryngologiczną jest trudnym zadaniem, co potwierdzają wyniki prac, w których żaden nie osiąga 100% dokładności klasyfikacji.

Prowadzone są również prace, których celem jest detekcja konkretnej choroby laryngologicznej. Autorzy publikacji [21] przedstawiają system do wykrywania trzech zaburzeń głosu: napięcie mięśni krtani, refluks żołądkowy, hiperfunkcja. 258 nagrań głosu pochodziło z bazy MEEI. Autorzy zastosowali dyskretną transformację falkową, energię sygnału oraz entropię Shannona w celu wykrycia cech szczególnych dla wybranych jednostek chorobowych. Cechy te zostały poddane optymalizacji przez wieloklasową liniową analizę dyskryminacyjną. Klasyfikacja przeprowadzona była przy pomocy sieci neuronowych. Dokładność klasyfikacji wszystkich 3 jednostek chorobowych wyniosła 96-97%. Praca [22] przedstawia klasyfikację 5 jednostek chorobowych zdiagnozowanych łącznie u 657 osób (baza MEEI): napięcie mięśni krtani, hiperfunkcja, kompresja komór fałdów głosowych, porażenie nerwów krtaniowych oraz refluks żołądkowy z wykorzystaniem ukrytych modeli Markova. Jako cechy wykorzystano 12 współczynników mel cepstralnych. Skuteczność klasyfikacji mieści się w zakresie od 61 do 69% w zależności od wybranej choroby. Autorzy publikacji [23] prezentują algorytm do dyskryminacji 5 chorób: cysty, choroby refleksyjnej przełyku, porażenia nerwów krtaniowych, polipów oraz bruzd. W badaniach wykorzystano dwa pierwsze formanty, a bazę stanowiła grupa 71 pacjentów. Klasyfikacja została przeprowadzona z wykorzystaniem algorytmu kwantyzacji wektorowej zapewniając 67,86% skuteczność dyskryminacji pomiędzy wybranymi chorobami. Praca [24] przedstawia algorytm do klasyfikacji osób z porażeniem nerwów krtaniowych, polipów, guzów oraz osób zdrowych. W tym celu wykorzystano transformatę falkową do obliczenia energii sygnału oraz entropii Shannona. Algorytm SVM wykorzystany do klasyfikacji zapewnił 91% skuteczność wykrycia poszczególnych jednostek chorobowych. Nagrania samogłoski /a/ wykorzystane w pracy zostały pozyskane od łącznie 147 osób.

Analiza akustyczna głosu jest przydatna do poszukiwania chorób nie tylko w dziedzinie otolaryngologii. Innym obszarem jest neurologia. Procesy neurodegeneracyjne w mózgu mogą przyczynić się do powstania zaburzeń emisji głosu [25]. Wyniki pierwszych prac badawczych dotyczących wykorzystania analizy sygnału akustycznego w detekcji chorób neurodegeneracyjnych motywują do kontynuowania badań w tym obszarze oraz w rezultacie do utworzenia systemu nie tylko do detekcji, ale również do monitorowania przebiegu wybranych chorób [26–29]. W niniejszej pracy doktorskiej skupiono się na ocenie choroby Parkinson'a na podstawie sygnału głosu.

Po raz pierwszy w 1817 roku James Parkinson opisał zmiany w głosie spowodowane chorobą Parkinson'a [30]. Choć takie objawy obserwowane były już w odległych czasach, zainteresowanie tą tematyką pojawiło się całkiem niedawno. Przydatność samogłosek do detekcji choroby Parkinson'a, ze względu na ich krótki czas rejestracji oraz analizy, została przedstawiona przez autorów publikacji [31]. 156 nagrań samogłoski /a/ zostało opisanych za pomocą 309 parametrów w celu wykrycia tych najbardziej informacyjnych z diagnostycznego punktu widzenia. Współczynnik shimmer, stosunek mocy sygnału do szumu, 1-3 współczynniki MFCC to parametry niosące największą ilość informacji na temat choroby Parkinson'a [31]. Wykorzystując takie wnioski przeprowadzono klasyfikację za pomocą analizy SVM otrzymując dokładność na poziomie 90%. Inny aspekt związany z chorobą Parkinson'a przedstawiono w pracy [32]. Autorzy opisują występowanie drżenia krtani za pomocą parametrów: częstotliwości podstawowej, współczynników jitter oraz shimmer, stosunku harmonicznego do szumu (HNR) oraz innych cech fonacyjnych. Do testów wykorzystano nagrania samogłoski /a/ powtórzonej trzy razy przez 22 pacjentów z chorobą Parkinson'a, którzy nie byli leczeni lekami dopaminergicznymi oraz 28 zdrowych pacjentów. W wyniku badań stwierdzono wzrost wartości współczynników jitter oraz shimmer u pacjentów z PD, co nie wystąpiło u grupy osób zdrowych. Wartości HNR oraz częstotliwości podstawowej były niższe u pacjentów z PD. Takie same wnioski zostały opublikowane w artykule [33], w którym wykorzystano nagrania 41 pacjentów z PD poddanych leczeniu lekami dopaminergicznymi. Wyniki opublikowane w [34] prezentują badania z wykorzystaniem samogłoski /a/ wypowiedzianej przez 23 pacjentów z PD oraz 8 zdrowych osób. Autorzy skupili się na analizie fonacji w programie Praat obliczając: dwa współczynniki jitter (o wartości absolutnej oraz absolutną średnią różnicę pomiędzy kolejnymi cyklami), iloraz perturbacji amplitudy (ang. *amplitude perturbation quotient*), shimmer, stosunek harmonicznego do szumu, entropię gęstości obecnej w cyklach (ang. *recurrence period density entropy*), beztrendową analizę fluktuacyjną, korelację wymiaru oraz entropię tonu. Na podstawie powyższych parametrów przeprowadzono klasyfikację z wykorzystaniem algorytmu SVM z jądrem Gaussa. Średnia dokładność wykrycia choroby Parkinsona została określona na poziomie 91,4%. Badania zmian w głosie na podstawie takich aspektów jak fonacja, artykulacja oraz prozodia mowy parkinsonowskiej zostały przeprowadzone przez autorów [35] w różnych grupach językowych. Wykorzystano nagrania głosu 46 osób posługujących się językiem czeskim, 23 z PD oraz 23 zdrowych, które zawierały 6 różnych treści: fonacji przedłużonych samogłosek, powtórzeń sylab: /pa-ta-ka/, przeczytanie tekstu o 136 słowach, wygłoszenie monologu, przeczytanie zdań oraz rytmicznego tekstu. W celu analizy samogłosek o przedłużonej fonacji wykorzystano: wariancję częstotliwości podstawowej, dwa współczynniki jitter, shimmer, stosunek harmonicznego do szumu oraz szumu do harmonicznego, częstotliwość 1 oraz 2 formantu. Na podstawie badań stwierdzono, że u 78% badanych wystąpiły zmiany w głosie, głównie w prozodii oraz artykulacji.

Choroba Parkinson'a jest schorzeniem postępującym w czasie. Z tego powodu, w celu wspomagania pracy lekarzy, przydatnym narzędziem jest system za pomocą którego możliwa byłaby obserwacja postępu choroby Parkinson'a w czasie. Realizację takiego zadania podjęli autorzy pracy [35]. Pacjenci wypowiadający samogłoskę /a/ byli nagrywani wiele razy w ciągu sześciu miesięcy. Każdy pacjent został zarejestrowany około 141 razy, w sumie wykonano 5923 nagrania, wykonywane dla każdego pacjenta w 3 okresach czasu: na początku programu, po 3 oraz po 6 miesiącach. Każdy z pacjentów został poddany diagnostyce lekarskiej w celu określenia jego stanu w skali UPDRS (ang. *Unified Parkinson's Disease Rating Scale*). UPDRS jest skalą powszechnie używaną przez lekarzy w celu oceny postępu choroby Parkinson'a. Ideą pracy było mapowanie parametrów akustycznych na skalę UPDRS. Zadaniem pacjenta było wypowiedzenie 6 samogłosek o przedłużonej fonacji. Analizę przeprowadzono osobno dla kobiet i mężczyzn ze względu na różnice w wartościach parametrów akustycznych. Do analizy wykorzystano różne parametry (współczynnik szumu turbulentnego, 6 współczynników MFCC, logarytm energii, częstotliwość podstawową, współczynnik pobudzenia fałdów i inne). Do wyznaczenia punktów UPDRS na podstawie analizy sygnału głosu przetestowano różne techniki regresji: metodę drzew klasyfikacji, regresję (CART) oraz lasy losowe. W wyniku 10-krotnego sprawdzianu krzyżowego stworzonego algorytmu otrzymano różnicę 2 punktów UPDRS względem wyniku lekarskiego. Praca ta wykorzystuje możliwość automatycznego wyznaczenia skali UPDRS na podstawie analizy głosu. Autorzy publikacji [36] przedstawiają wyniki pracy, której celem była estymacja nasilenia choroby Parkinson'a na podstawie sygnału głosu. W badaniach wykorzystano 50 nagrań, a badane cechy to: stosunek części dźwięcznej do bezdźwięcznej sygnału (ang. *voicing ratio*), występowanie cech nieliniowych w sygnale: znormalizowana entropia gęstości prawdopodobieństwa i skalowanie fraktalne (ang. *fractal scaling*). Estymację punktacji UPDRS przeprowadzono za pomocą sieci neuronowych. W wyniku badań stwierdzono, że wykorzystane cechy pozwalają na estymację nasilenia choroby Parkinson'a.

1.3. Struktura pracy

W niniejszej rozprawie przedstawiono metody oraz uzyskane rezultaty badania algorytmów wspomagających pozyskanie informacji diagnostycznej, prognostycznej i kontrolnej różnych jednostek chorobowych w obszarze laryngologii oraz neurologii na podstawie sygnału głosu.

W rozdziale 2 zaprezentowano opis procesu generowania sygnału mowy w ujęciu ogólnym, mowy zdeformowanej z uwzględnieniem wybranych chorób laryngologicznych oraz zaburzenia mowy w chorobie Parkinson'a. Rozdział ten zawiera również opis skal klinicznych wykorzystywanych do oceny stopnia zaawansowania choroby Parkinson'a.

Rozdział 3 przedstawia materiał badawczy wykorzystany w przeprowadzonych analizach. Następnie opisana została procedura parametryzacji sygnału akustycznego mowy, metodologia pozyskiwania informacji diagnostycznej, kontrolnej i prognostycznej z wykorzystaniem sygnału mowy.

W rozdziale 4 zaprezentowano wyniki badań automatycznej detekcji i klasyfikacji chorób laryngologicznych. Badania ta przebiegały dwu-etapowo. W pierwszym, wykorzystując algorytm lasów losowych przeprowadzono automatyczną klasyfikację osób zdrowych i pacjentów z różnymi schorzeniami laryngologicznymi. W drugim etapie, klasyfikację przeprowadzono na podstawie sieci neuronowych, w wyniku której algorytm wskazał jedną z trzech zdefiniowanych wstępnie chorób laryngologicznych lub oznaczył danego pacjenta jako osobę zdrową. Wykorzystując bazę nagrań osób z chorobą Parkinson'a za pomocą algorytmu SVR wykonano najpierw estymację punktacji w skali UPDRS-III w celu przeprowadzenia kontroli zmian stanu pacjenta. Następnie przeprowadzono predykcję stopnia nasilenia choroby Parkinson'a w 180 minucie po spożyciu leków oraz estymację punktacji UPDRS-III, uzyskując obiektywny wynik stanu zdrowia badanych osób.

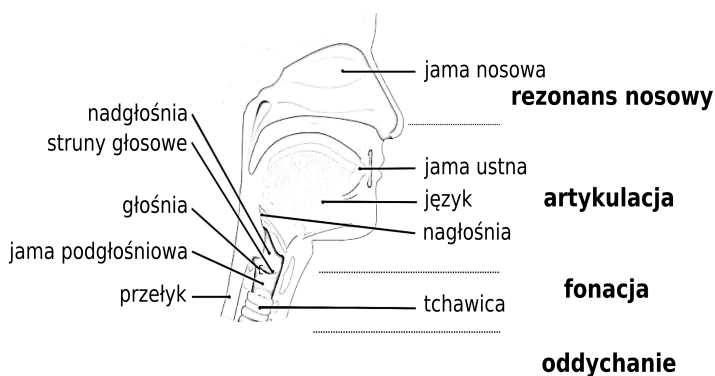
Rozdział 5 przedstawia podsumowanie całej rozprawy.

2. Proces generowania sygnału mowy i mowy zdeformowanej

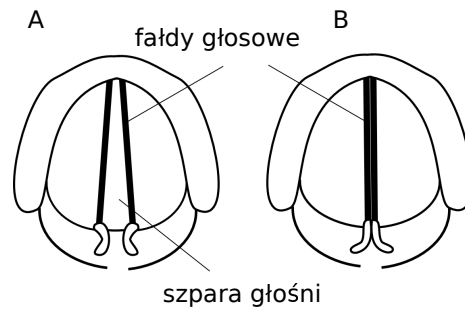
Poznanie i zbadanie mechanizmów generacji sygnału akustycznego mowy człowieka pozwala przedstawić w sposób parametryczny prawidłową i zdeformowaną emisję głosu. Proces tworzenia głosu jest skomplikowanym zjawiskiem akustycznym i mechanicznym, w wyniku którego powstaje dźwięk. Z fizycznego punktu widzenia dźwięk jest drganiem cząstek ośrodka sprężystego w odniesieniu do położenia równowagi. Wibracje cząstek powietrza powodują lokalne zmiany ciśnienia w stosunku do ciśnienia atmosferycznego (tzw. ciśnienie akustyczne) przemieszczające się w postaci fal, na które reaguje ludzkie ucho [37].

Głównym generatorem drgań cząstek powietrza jest krtień. Jest ona umiejscowiona pomiędzy gardłem, a tchawicą (rys. 2.1). Krtień zbudowana jest ze szkieletu chrzęstnego, mięśni i więzadeł, które są połączone stawami, więzadłami oraz mięśniami. W skład mięśni krtani wchodzi: mięśnie zewnętrzne i wewnętrzne. Mięśnie zewnętrzne ustalają położenie krtani, odpowiadają za jej ruch w górę i ku dołowi, przemieszczenie ku przodowi i do tyłu. Mięśnie wewnętrzne krtani, w tym mięśnie głosowe, stanowią fundamentalną rolę w procesie fonacji, czyli w tworzeniu dźwięku [37]. Wewnątrz krtani znajduje się jama krtani, która wyścielona jest błoną śluzową. W części środkowej znajdują się fałdy głosowe. Za generację głosu odpowiadają procesy fizjologiczne i fizyczne, dzięki którym wibracja fałdu głosowego zamieniana jest na mowę. Fałdy głosowe drgają, co w konsekwencji powoduje powstawanie dźwięku (rys. 2.2).

Podczas fonacji, fałdy głosowe mogą być zwarte w różny sposób [38]:



Rys. 2.1. Schemat budowy narządu głosowego.



Rys. 2.2. Schematyczne przedstawienie położenia wiązadeł głosowych i czynności fałdów głosowych, A - faza oddechowa, szpara głośni rozwarta, B- faza fonacyjna - szpara głośni zwarta [37].

- nastawienie miękkie, fałdy głosowe przybliżają się do siebie, nie zwierając się mocno, pomiędzy nimi pozostaje eliptyczna szczelina. Takie nastawienie głosowe jest cechą prawidłową głosu fizjologicznego, brzmienie głosu jest pozbawione napięć i wysiłku (rys. 2.2);
- nastawienie twarde, fałdy głosowe mocno się zwierają, napierają na siebie tak mocno, że utrudnione są ich swobodne drgania. W takim nastawieniu głos tworzony jest siłowo, niezgodnie z fizjologią narządu głosowego, głos jest krzykliwy, wybuchowy. Dodatkowo może występować ból w okolicy krtani. Nastawienie twarde widoczne jest w dysfonii funkcjonalnej i hyperfunkcjonalnej (rys. 2.3);
- nastawienie chuchające, fałdy głosowe przybliżają się do siebie tworząc niepełne zwarcie, a pomiędzy nimi pozostaje szeroka przestrzeń, która może przybierać różne kształty, powstałe przez mięśnie, które uległy porażeniu lub niedowładowi. Nastawienie chuchające może występować przy porażeniu nerwu krtaniowego wstecznego (rys. 2.4).

Główna siła napędowa drgań fałdów głosowych i produkcja głosu zależą od zamiany energii aerodynamicznej na energię akustyczną, podczas gdy są one zamknięte w linii środkowej. Do wytworzenia dźwięku niezbędne jest wytworzenie wibracji powietrza przechodzącego przez szparę głośni. Z zewnątrz na fałdy działa siła pochodząca od ciśnienia atmosferycznego, a od środka działa siła pochodząca od ciśnienia powietrza wydostającego się z płuc. Przepływ powietrza przez fałdy powoduje spadek ciśnienia i w rezultacie samorzutne ich zaciskanie. Zamknięcie fałdów powoduje wyrównanie ciśnienia i ich powrót do położenia pierwotnego. Drgania generowane są w efekcie ruchu fałdów głosowych w kierunku poprzecznym do ich długiej osi [37]. Naprzemienne rozwieranie i zwieranie fałdów głosowych powoduje powstawanie drgań powietrza, a w efekcie dźwięku. Głównym jego parametrem jest częstotliwością podstawową, zwana inaczej tonem krtaniowym lub podstawowym i stanowi odwrotność okresu drgań fałdów. Jego częstotliwość oraz moc zależą od właściwości fałdów głosowych, ich masy, długości, napięcia i elastyczności. Parametry te mogą być modulowane dzięki pracy mięśni krtani.

Ton krtaniowy ma określoną częstotliwość, wysokość i natężenie. W badaniach foniatrycznych nie ma bezpośredniego dostępu do tonu krtaniowego, korzysta się więc z wynikowego sygnału mowy, który powstaje z udziałem całego traktu głosowego (języka, warg, podniebienia, policzków oraz ruchów szczęk). Z punktu widzenia akustycznego trakt głosowy to układ rezonatorów, gdzie kształt i połączenia między nimi zmieniają się w zależności od ułożenia narządów artykulacyjnych (żuchwy, języka, warg, podniebienia). Ponieważ wiele czynników może wpływać na fonację, produkcja głosu jest wysoce zmienna i zależy nie tylko od cech osobniczych.

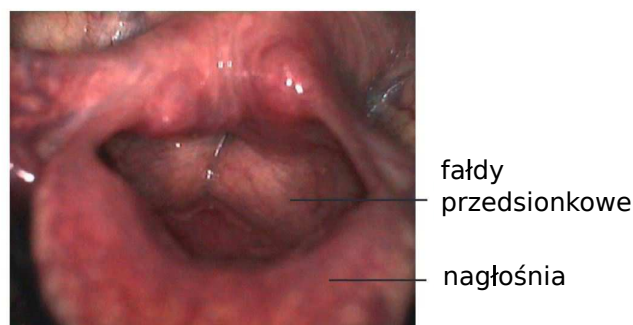
2.1. Patologia procesu generowania sygnału akustycznego w ujęciu medycznym

Do zaburzeń mowy można zaliczyć wszystkie patologiczne zachowania - od pogorszenia jakości głosu, powstania wad wymowy aż do całkowitego jego zaniku i możliwości porozumiewania się. Na zaburzenia mowy wpływ mogą mieć czynniki rozwojowe, funkcjonalne i etiologiczne [39]. Do czynników rozwojowych zalicza się kwestie genetyczne oraz nabyte (powstałe w wyniku zmian chorobotwórczych). Czynniki funkcyjne obejmują zaburzenia artykulacji, fonacji, funkcji językowych oraz płynności mowy. Czynniki etiologiczne uwzględniają uszkodzenie narządów artykulacyjnych, ośrodkowego układu nerwowego, zaburzenia o podłożu psychogennym oraz o niejasnej etiologii [40].

Schorzenia otolaryngologiczne, dla których przeprowadzono badania w ramach tej pracy to:

- dysfonia hyperfunkcjonalna,
- porażenie nerwu krtaniowego wstecznego,
- ostre zapalenie krtani,
- zaburzenia spowodowane chorobą Parkinson'a.

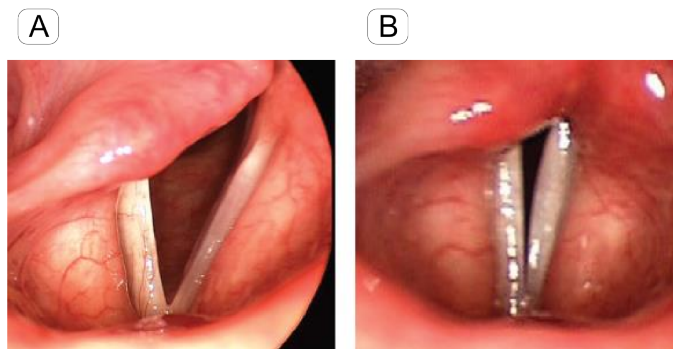
Wśród osób z zaburzeniami mowy i głosu dużą grupę stanowią osoby z tzw. zaburzeniami czynnościowymi, powstałymi w wyniku działania czynników rozwojowych. W początkowym okresie anomalii nie stwierdza się zmian organicznych w krtani (np. guzków głosowych, polipów, obrzęków), zaburzona jest natomiast czynność, tzn. nieprawidłowo pracują mięśnie wewnątrz- i zewnątrz-krtaniowe. Na tym etapie takie zaburzenia (dysfonie) głosu są odwracalne. Nieprawidłowe mechanizmy fonacyjne z zaburzoną koordynacją oddechowo-fonacyjno-artykulacyjną prowadzą do powstania dysfonii czynnościowych [44]. Takie zmiany mogą przyczynić się do powstawania defektów głosu. U osób pracujących głosem zaburzenia czynnościowe wyprzedzają najczęściej pojawienie się zmian organicznych na fałdach głosowych [41]. Jedną z najczęstszych postaci anomalii czynnościowych jest dysfonia, czyli wielopostaciowe



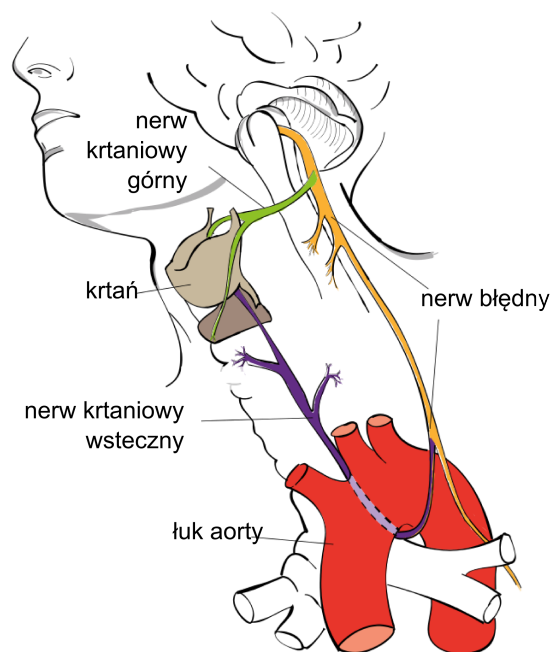
Rys. 2.3. Dysfonia hyperfunkcjonalna, obraz wideolaryngostroboskopowy. Faza fonacyjna – zwanie fałdami przedsionkowymi. Nadmierne zwanie fonacyjne – część nadgłośniowa krtani zasłania fałdy głosowe [46].

zaburzenie głosu, które może występować wskutek: obniżonego napięcia mięśni krtani (dysfonia hypofunkcjonalna) lub zwiększonego napięcia mięśni krtani (dysfonia hyperfunkcjonalna). Obejmuje ona zmiany składowych akustycznych tj. częstotliwości, poziomu głośności, czasu trwania fonacji i barwy głosu. Problem zaburzeń czynnościowych głosu dotyczy najczęściej osób, u których głos stanowi narzędzie pracy.

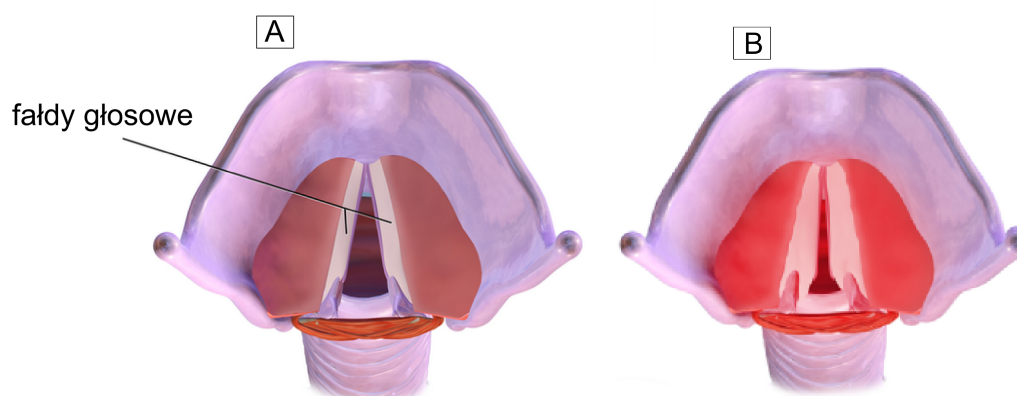
Według literatury, dysfonia głosu dotyka najczęściej nauczycieli [2, 42]. Najczęściej spotykane objawy tej choroby zawodowej to parestezje gardła i krtani, częste chrząkanie, utrudniona modulacja głosu, skrócenie czasu fonacji, chwilowe zaniki głosu, chrypka z bezgłosem [43]. Dysfonie czynnościową można podzielić na trzy typy: hyperfunkcjonalną - najczęstszą (65%), hypofunkcjonalną - 19%, dyfunkcjonalną = 16% [44]. Podczas dużego wysiłku głosowego, Zaburzenia głosu powstałe w wyniku dużego wysiłku głosowego (np. u nauczycieli) mogą prowadzić do powstania dysfoni hyperfunkcjonalnej. Wówczas stwierdza się dyskomfort podczas fonacji pojawiający się w okolicy gardła i krtani, czemu może towarzyszyć nawet silny ból uniemożliwiający generację dźwięku [45]. Podczas badania laryngologicznego można stwierdzić nadmierne napięcie mięśni krtaniowych, podniebienia, gnykowych. Głos ma brzmienie parté, gardłowe lub nosowe, z ograniczonym rezonansem i z nastawieniem twardym [44]. Dodatkowo skraca się czas fonacji, zmniejsza się dźwięczność głosu, powstaje chrypka. Laryngoskopia pośrednia umożliwia obserwację: nadgłośnia jest opadająca, napięte są fałdy głosowe, brzeżnie przekrwione, sfinkterowaty mechanizm fonacji [45]. Inne zmiany, możliwe do detekcji w badaniu stroboskopowym to: zmniejszona amplituda drgań, brzeżne przesunięcie krawędziowe, nieregularne drgania fałdów głosowych [41, 44]. Nieleczenie dysfonii czynnościowej może prowadzić do powstania zmian organicznych fałdów głosowych. Długotrwałe zmiany patologiczne krtani obniżają wydolność narządu głosu oraz mogą utrudniać proces komunikacji [45].



Rys. 2.4. Jednostronne porażenie nerwu krtaniowego wstecznego, A - faza oddechowa, B - faza fonacyjna [51].



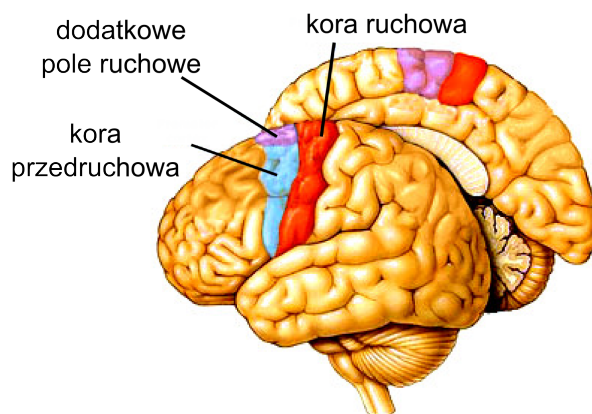
Rys. 2.5. Nerw krtaniowy wsteczny (kolor fioletowy) [52].



Rys. 2.6. Zapalenie krtani, A - krtani zdrowa, B - Stan zapalny krtani [53].

Innym schorzeniem, analizowanym w tej rozprawie, jest porażenie nerwu krtaniowego wstecznego. Jego położenie przedstawia rys. 2.5. Nerw krtaniowy wsteczny unerwia wszystkie wewnętrzne mięśnie krtani, z wyjątkiem mięśnia pierścienno-tarczowego, który unerwiający jest przez nerw krtaniowy górny. Przyczyną porażenia nerwu krtaniowego wstecznego może być długotrwały łagodny przerost tarczycy i ucisk gruczołu na nerw [46]. Do innych przyczyn można zaliczyć nadczynność gruczołu tarczowego (powodującą obrzęk i zrosty pozapalne), a także zapalenie gruczołu tarczowego. Znaczącą liczbę porażań nerwów krtaniowych wstecznych powoduje uraz jatrogenny w czasie operacji tarczycy [47, 48]. Wynika to z bezpośrednich uszkodzeń pnia nerwu w czasie zabiegu (przecięcie, rozerwanie) lub ucisku na jego pień w okresie pooperacyjnym, wywołanego przez krwiak lub obrzęk w otaczających tkankach, a w późniejszym okresie przez zmiany bliznowate [49, 50]. Porażenie tego nerwu może prowadzić do powstania następujących symptomów chorobowych: bezgłosu (przy porażeniu obustronnym) lub chrypki (przy porażeniu jednostronnym), świstu krtaniowego (stridor), zaburzeń połykania (dysfagia), duszności [45].

Zaburzenie w generowaniu sygnału mowy może również być wywoływane przez ostre zapalenie krtani, które jest schorzeniem błony śluzowej krtani (rys. 2.6). Owa dysfunkcja może powstać samoistnie, jako zakażenie zstępujące (ogniskiem pierwotnym jest infekcja nosa, zatok przynosowych, gardła) lub zakażenie będące skutkiem zapalenia oskrzeli [45]. Czynniki predysponującymi do zapalenia krtani są: nadużywanie fałdów głosowych, suche, zapyłone powietrze, choroby współistniejące. Początkowy obraz kliniczny to afonia lub dysfonia, ból, uczucie drapania, suchość w krtani. Pojawia się suchy kaszel oraz coraz większa chrypka dochodząca do bolesnego bezgłosu. W laryngoskopii w ostrym zapaleniu krtani stwierdza się zaczerwienione fałdy głosowe z nastrzykniętymi naczyniami na błonie śluzowej. Fałdy głosowe mogą być obrzęknięte, co powoduje że ich zwarcie przy fonacji jest niepełne [45]. Na fałdach głosowych lub pomiędzy nimi może zalegać wydzielina.



Rys. 2.7. Obszary kory mózgowej zaangażowane w planowanie i wykonywanie ruchu [55].

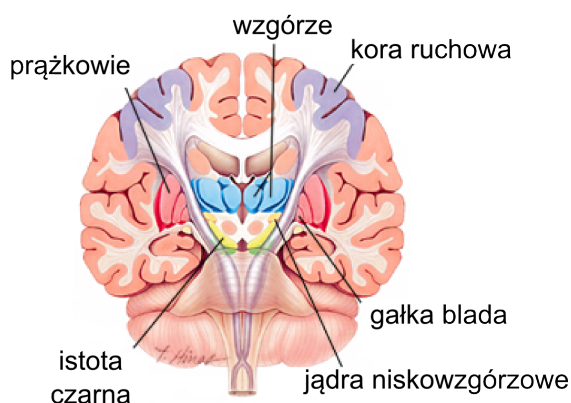
2.2. Zaburzenia mowy w chorobie Parkinson'a

Zaburzenia mowy dotyczą nawet 89% wszystkich pacjentów z chorobą Parkinson'a (PD, ang. *Parkinson's disease*) [54]. Zaburzenia głosu, w tym zmniejszenie głośności mowy, problemy artykulacji i płynności są często jednym z pierwszych objawów. James Parkinson opisał w swojej publikacji [30] powstawanie zaburzeń mowy w skutek PD.

Aby wyjaśnić dlaczego mowa osób z chorobą Parkinsona jest zmieniona, należy odnieść się do wiedzy anatomicznej o mózgu, związanej z motorycznością oraz procesem wytwarzania mowy.

Kora mózgu składa się z kilku części, z których dwie są odpowiedzialne za wykonywanie i planowanie ruchu (kora ruchowa i kora przed-ruchowa). Część obszaru zajmującego się kontrolą mowy należy zarówno do kory ruchowej i przed-ruchowej (rys. 2.7). Większość obszarów mózgu zaangażowanych w funkcje ruchowe można badać poprzez analizę części korowej mózgu, w jądrach podstawnych. Rysunek 2.8 ilustruje taki obszar i obejmuje struktury związane z ruchem. Aktywność mięśni jest inicjowana w korze przed-ruchowej, która przekazuje informację do jąder podstawnych przez prążkowie. Następnie sygnały z jąder przekazywane są z gałki bladej i są przesyłane poprzez wzgórze do kory. Po przetworzeniu informacji kora wysyła sygnały do odpowiednich mięśni.

Zaburzenia mowy u osób z chorobą Parkinsona spowodowane są głównie przez: deficyt czynnościowy krtani, osłabioną pracę mięśni mimicznych, zmniejszoną pojemność życiową płuc i zmniejszony napęd mówienia [57]. Takie zmiany prowadzą do powstania licznych nieprawidłowości w głosie i mowie, w tym: redukcji głośności, obniżenia tonu głosu, ograniczonej



Rys. 2.8. Przekrój głowy w płaszczyźnie szwu wieńcowego: jądra podstawne [56].

modulacji (mowa monotonna), trudności ze zmianami głośności, redukcji napięcia fałdów głosowych, chropawego i ochryplego tonu, a także niewłaściwej artykulacji (mowa staje się niewyraźna) oraz zmiany tempa wypowiedzi [57–59]. Upośledzenia te nazywane są dyzartriami hipokinetycznymi. Dyzartryczna mowa charakteryzuje się dysfunkcjami fonacji, artykulacji oraz prozodii, które powstały w wyniku uszkodzenia ośrodków i dróg nerwowych odpowiadających za unerwienie narządów mowy [39, 40]. Fonacja definiowana jest jako wibracja strun głosowych, w efekcie czego powstaje dźwięk. Artykulacja łączy modyfikację pozycji, napięcia i kształtu struktur oraz języka zaangażowanych w produkcję mowy. Najbardziej widocznymi cechami fonacyjnymi u pacjentów z PD są jednostajność melodyki oraz monotonna mowa. Z klinicznego punktu widzenia, problemy fonacyjne są powiązane z nieprawidłowym ruchem fałdów głosowych oraz niecałkowitym ich zamknięciem [60]. Zmiany w artykulacji są spowodowane zredukowaną amplitudą i prędkością ruchów: warg, szczęki i języka. Prowadzi to do zmniejszonego akcentowania, niedokładnej artykulacji spółgłosek aż do bełkotu. Prozodia to brzmieniowa właściwość mowy uwzględniająca intonację, głośność, akcent i długość trwania fonacji [57, 61]. Nieprawidłowości w prozodii objawiają się mówieniem krótkimi, przyspieszonymi frazami, monotonnością i ograniczoną głośnością mowy, zmianą tempa mowy, pauzami, trudnością w ekspresji emocji, powtarzaniem głosek lub sylab [26].

2.3. Skale kliniczne opisujące zaawansowanie choroby Parkinsona

Rozpoznanie choroby Parkinsona następuje na podstawie całościowego obrazu klinicznego. Typowymi objawami choroby PD są: zaburzenia ruchowe (bradykinezja, hipokinezja, akinezja), sztywność i wzmożone napięcie mięśni, drżenie spoczynkowe (o częstotliwości około 5 Hz), pro- i retropulsja (upadanie do przodu lub do tyłu), mikrografia (pismo mniej wyraźne, drobniejsze), zaburzenia mowy, węchu lub połykania, zaburzenia psychiczne (depresja, zespoły

otępienia). Rozpoznanie choroby Parkinsona odbywa się na podstawie objawów ruchowych i jest niezwykle trudne we wczesnej fazie choroby. Spowolnienie ruchowe może występować we wszystkich postaciach parkinsonizmu, drżenie spoczynkowe może wystąpić w parkinsonizmie polekowym, a niestabilność postawy może wskazywać na rozwój atypowego parkinsonizmu. Z tego powodu, objawy ruchowe dobrze nadają się do oceny postępu choroby w czasie. Do takiej oceny służą między innymi: skala według Hoehn&Yahr [62], skala samodzielności Schwaba i Englanda [63] oraz ujednolicona skala oceny choroby Parkinsona (ang. *Movement Disorders Society - Unified Parkinson's Disease Rating Scale*, MDS-UPDRS) [64].

Skala Hoehn&Yahr ocenia zdolność pacjenta do samodzielnego prowadzenia typowych aktywności życia codziennego, a jej punktacja mieści się w zakresie od I do V. Skala Schwaba i Englanda określa nasilenie trudności z wykonywaniem podstawowych czynności w życiu codziennym oraz stopień zależności od otoczenia i osób trzecich w ich wykonywaniu [65].

Najczęściej stosowaną skalą do oceny nasilenia objawów choroby PD jest skala UPDRS. Stanowi ona dobre narzędzie do monitorowania progresji choroby zarówno w leczeniu objawowym jak i w monitorowaniu objawów pozaruchowych [66]. Skala ta składa się z IV części. Część I uwzględnia stan intelektualny i zaburzenia nastroju (4 zagadnienia), część II ocenia aktywności życia codziennego (13 zagadnień), w części III ocenia się funkcje motoryczne (27 zagadnień), a w części IV oceniane są powikłania leczenia (11 zagadnień) [66]. Każde z zagadnień może otrzymać od 0 (brak objawów) do 4 punktów (znaczące objawy). Całkowita liczba punktów jest sumą z każdej części i może wynosić maksymalnie 220. Wyższy wynik w skali UPDRS wskazuje na bardziej zaawansowane stadium. Wpływ choroby Parkinsona na mowę jest uwzględniony w części III skali UPDRS i najczęściej w badaniach analizujących mowę pacjentów ogranicza się do punktacji uzyskanej tylko z tej części. W takim przypadku liczba punktów może wynosić od 0 do 108 (27 zagadnień x 4 = 108) [67].

W rozdziale przedstawiono opis zjawiska i procesy zachodzące podczas generowania sygnału akustycznego mowy człowieka. Następnie przedstawiono wybrane choroby głosu w ujęciu medycznym. Owe dysfunkcje, zostały wykorzystane do stworzenia systemu klasyfikacji osób zdrowych oraz tych, które posiadały symptomy danej jednostki chorobowej. Podrozdział 2.2 uwzględnia źródło oraz opis zaburzeń głosu i mowy występujące w chorobie Parkinson'a. Do oceny jej nasilenia wykorzystuje się różne skale. Ich opis został przedstawiony w sekcji 2.3. Przedstawiona dysfunkcja pokazuje potencjał metod opartych na analizie w kontekście wykrywania i monitorowania jednostek chorobowych nie tylko laryngologicznych, ale też neurodegeneracyjnych. Analizę należy wzbogacić opisem parametrycznym sygnału głosu w celu kontroli wszystkich zaburzeń jakie mogą występować w trakcie trwania procesu chorobowego. W badaniach tych zaburzeń, wykorzystano samogłoski o przedłużonej fonacji i intonacji. Metodologia zaproponowana w tej pracy została zastosowana do sygnałów zawartych w dwóch

bazach danych zawierających nagrania w językach: polskim oraz niemieckim. Szczegółowe omówienie nagrań sygnału mowy zawartych w każdej bazie danych oraz opis parametryczny tych sygnałów głosu przedstawiono w następnym rozdziale.

3. Materiał i metoda badawcza

Wprowadzenie metod cyfrowego przetwarzania sygnału głosu do diagnostyki laryngologicznej może przyspieszyć i ułatwić lekarzowi postawienie diagnozy oraz skrócić czas badania danego pacjenta. Analiza akustyczna pozwala na opis parametryczny sygnałów generowanych i emitowanych przez narząd głosu człowieka, a jej wyniki dobrze korelują ze stanem foniatrycznym przy prawidłowej i patologicznej emisji głosu [68]. W efekcie możliwe jest określenie stanu i funkcjonalności tego narządu.

Zmiany struktury anatomicznej i ograniczenia czynności krtani wpływają na zmianę parametrów fizycznych źródła tonu krtaniowego i znajdują odzwierciedlenie w postaci zmian struktury akustycznej głosu. Z tego powodu, analiza parametrów akustycznych może wykryć anomalie struktury anatomicznej i wskazać przyczyny ograniczenia czynności ruchowych struktur krtani. Tym samym wyniki takiej analizy mogą wspomagać diagnozę.

W rozdziale tym przedstawiono opis danych wykorzystanych do implementacji metod pozwalających na uzyskanie informacji diagnostycznej, kontrolnej i prognostycznej. Przedstawiono proces parametryzacji sygnału akustycznego omawiając każdy z obliczonych parametrów. W dalszej części przedstawiono opis algorytmów, które zostały zaimplementowane do realizacji postawionych zadań.

3.1. Materiał badawczy

Badania prowadzono na sygnałach dostępnych w bazie *Saarbruecken Voice Database* (SVD) udostępnionej przez Instytut Fonetyki w Saarland w Niemczech [14]. Baza SVD zawiera nagrania pochodzące od ponad 2000 osób, które wypowiadały samogłoski /a/, /i/, /u/ w wysokiej, niskiej, normalnej oraz modulowanej tonacji. Czas trwania nagrań samogłosek wynosił od 1 do 4 sekund. Wszystkie nagrania zostały zarejestrowane z częstotliwością próbkowania 50 kHz i z rozdzielczością 16 bitów. Baza SVD składa się z nagrań mowy pacjentów cierpiących łącznie na 71 różnych chorób narządu głosu. Z bazy usunięto nagrania uszkodzone lub o czasie trwania krótszym niż 1 sekunda. W badaniach wykorzystano nagrania samogłoski /a/, /i/ oraz /u/ dla normalnej, wysokiej oraz niskiej intonacji. Pochodzą one od 900 kobiet, z których 450 były zdrowe, a 450 miało stwierdzoną chorobę narządu głosu oraz od 510 mężczyzn, z których

Tabela 3.1. Liczebność grup pacjentów z bazy SVD, których nagrania zostały poddane analizie.

nazwa choroby	kobiety	mężczyźni
dysfonia hyperfunkcjonalna	166	45
porażenie nerwu krtaniowego wstecznego	138	74
ostre zapalenie krtani	57	82
zdrowi	361	201

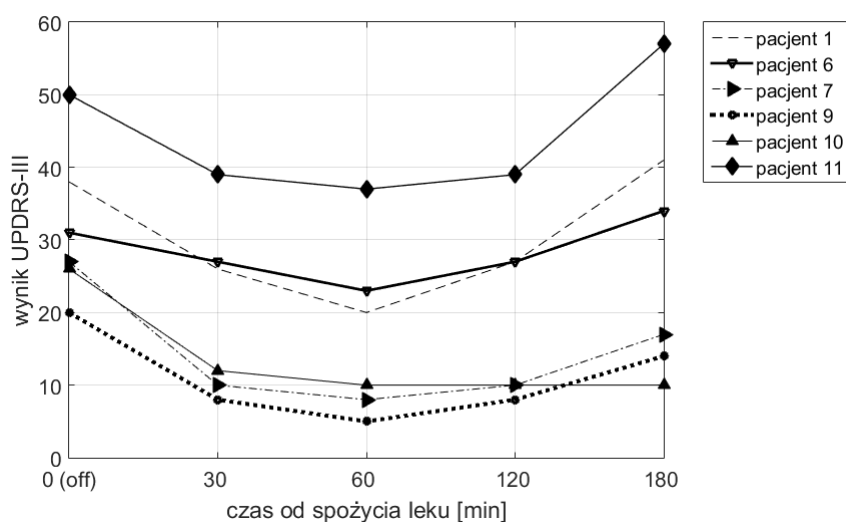
255 było zdrowych, a 255 należało do grupy osób chorych. Analizę akustyczną przeprowadzono osobno dla kobiet i mężczyzn ze względu na istotne różnice w głosie pomiędzy tymi grupami. Tak duża baza danych została wykorzystana w celu zbudowania systemu, który byłby w stanie wykryć istnienie nieprawidłowości lub ich brak w głosie ludzkim. Postawione zadanie polegało na wykryciu konkretnej choroby narządu głosu. Do analizy wykorzystano nagrania pochodzące od osób z dysfonią hyperfunkcjonalną, ostrym zapaleniem krtani, porażeniem nerwu krtaniowego wstecznego oraz od osób zdrowych. Liczebność osób chorych na poszczególne choroby zestawiono się w tabeli 3.1.

Ponadto dla potrzeb badawczych w Krakowskim Szpitalu Specjalistycznym im. Jana Pawła II w Krakowie, na Oddziale Neurologicznym z Poddziałem Udarowym i z Pododdziałem Rehabilitacji Neurologicznej przeprowadzono nagrania głosu pacjentów z chorobą Parkinson'a. Wszystkie nagrania zostały wykonane w tym samym pomieszczeniu, w kontrolowanych warunkach. Próbkę głosu zostały zarejestrowane z częstotliwością próbkowania 44 100 Hz i rozdzielczością 16 bitów, z wykorzystaniem mikrofonu pojemnościowego o charakterystyce kardoidalnej (AKG C1000S) oraz profesjonalnej karty dźwiękowej (M-Audio M-Track MKII).

Pacjenci byli rekrutowani do badań spośród chorych leczonych w Poradni Neurologicznej oraz pacjentów hospitalizowanych w Oddziale Neurologicznym KSS im. Jana Pawła II ¹.

Rejestracja głosu została przeprowadzona w stanie *off* (tj. w momencie zupełnego wysycenia leków) oraz w stanie *on* (tj. w momencie, w którym widoczne jest działanie leku, a czas od spożycia ostatniej dawki leku nie był dłuższy niż 180 minut). W celu oceny progresji choroby badanie głosu zostało powtórzone po upływie 30, 60, 120 i 180 minut od zażycia standardowej dawki preparatu L-dopy (lewodopy). Badanie mowy polegało na zarejestrowaniu głosu pacjentów, których zadaniem było przeczytanie 5 samogłosek: /a/, /i/, /u/, /e/ oraz /o/ wypowiedzianych w intonacji normalnej oraz ze zmienną modulacją. Równocześnie z rejestracją mowy lekarz przeprowadzał ocenę nasilenia objawów parkinsonizmu w skali UPDRS i skali H&Y. Zarejestrowano głos 30 osób, z czego u 27 osób została zdiagnozowana choroba Parkinson'a.

¹Zaproponowana procedura nagrywania głosu uzyskała pozytywną opinię Komisji Bioetycznej przy Okręgowej Izbie Lekarskiej w Krakowie. Ponadto każdy z pacjentów wyraził pisemną zgodę na udział w badaniu.



Rys. 3.1. Wynik skali UPDRS-III w kolejnych punktach pomiarowych dla wybranych pacjentów, wyznaczony przez lekarza.

Wiek pacjentów mieścił się w przedziale od 35-74 lat (średnia 64 ± 10.2). Charakterystyka pacjentów przedstawiona została w tab 3.2. Językiem ojczystym wszystkich osób rekrutowanych do badania był język polski.

3.2. Parametryzacja sygnału akustycznego

Diagnostykę traktu głosowego można przeprowadzić w oparciu o wybrane głoski, alofony, fonemy i samogłoski. Samogłoski stanowią grupę najczęściej wykorzystywaną w diagnostyce medycznej ze względu na stan ustalony występujący podczas ich fonacji, umożliwiający wyznaczenie stabilnych w czasie parametrów sygnału akustycznego. Samogłoski powstają w warunkach swobodnego przepływu powietrza wzdłuż języka. Węzadła głosowe drgają okresowo lub quasi-okresowo, a podniebienie miękkie jest uniesione do góry blokując dostęp powietrza do nosa. Ułożenie środkowej i tylnej części języka względem podniebienia twardego decyduje o tym, jaka samogłoska zostanie wypowiedziana. Częstotliwość tonu krtaniowego może ulegać zmianie w zależności od rodzaju deformacji lub uszkodzenia struktur krtani. Taka zmiana może zostać zidentyfikowana podczas analizy sygnału głosu przy zastosowaniu odpowiednich metod cyfrowego przetwarzania sygnału. Celem analizy było zbadanie stabilności okresu tonu krtaniowego. Ton krtaniowy charakteryzuje się zmiennością swojego przebiegu z okresu na okres. Tylko wybrane samogłoski o przedłużonej fonacji zapewniają stabilność tego parametru [67]. Są to samogłoski: [a, e, i, o, u]. Dlatego takie samogłoski wykorzystywane są do celów diagnostycznych w otolaryngologii.

Tabela 3.2. Ocena pacjentów z chorobą Parkinson’a w skali UPDRS, część III wykonana przez lekarza. UPDRS 30, 60, 120, 180 - wyniki skali UPDRS otrzymane odpowiednio w 30, 60, 120 i 180 minucie od spożycia leków, UPDRS *off* - moment, w którym występuje zupełne wysycenie leków lewodopy.

lp.	wiek	UPDRS <i>off</i>	UPDRS 30	UPDRS 60	UPDRS 120	UPDRS 180	okres choroby [lata]
1	69	38	26	20	27	41	10
2	64	27	21	8	7	8	16
3	74	27	27	22	24	26	9
4	71	46	44	38	39	40	9
5	71	61	23	22	21	49	21
6	65	31	27	23	27	34	8
7	57	27	10	8	10	17	4
8	60	38	27	21	21	27	7
9	62	21	14	10	11	12	14
10	62	26	12	10	10	10	12
11	62	50	39	37	39	57	2
12	67	48	28	22	25	25	6
13	55	24	20	20	13	14	10
14	74	41	16	19	42	24	11
15	65	22	14	9	9	11	7
16	66	21	4	4	5	7	4
17	35	29	18	14	14	14	4
18	57	48	26	26	37	36	9
19	73	20	8	5	8	14	10
20	62	16	8	3	3	5	6
21	70	53	40	42	42	42	10
22	69	22	18	7	7	7	7
23	67	27	8	6	6	14	14
24	67	46	23	18	13	16	10
25	67	46	20	30	30	32	5
26	78	12	9	6	7	8	3
27	50	37	19	13	17	16	9

Aby uzyskać informacje istotne z diagnostycznego punktu widzenia konieczne było wyselekcjonowanie takich cech sygnału i jego widma częstotliwościowego, które determinują sposób jego percepcji przez odbiorcę, pozwalając tym samym na przeprowadzenie procesu diagnostycznego, kontrolnego lub prognostycznego. Wybór diagnostycznych cech opisujących sygnał mowy polegało na znalezieniu takich cech, przy wykorzystaniu których możliwe było różnicowanie przypadków patologicznych i zdrowych. Poszczególne cechy mogą różnić się jednostką miary oraz charakteryzować się zróżnicowanym zakresem zmienności. W przypadku wielu algorytmów klasyfikacyjnych może to mieć znaczący wpływ na wynik końcowy. Dlatego zalecane jest wstępne przetworzenie danych obejmujące ich skalowanie i normalizację, czyli sprowadzenie cech sygnału do postaci bezwymiarowej o unormowanym zakresie zmienności.

W celu ilościowej oceny procesu fonacji konieczny jest jego opis parametryczny. W tej pracy rozważono następujące parametry tego procesu: częstotliwość podstawową, współczynnik jitter i shimmer, pięciopunktowy i jedenastopunktowy współczynnik zaburzeń okresu, pięciopunktowy współczynnik zaburzenia amplitudy, energię, zerowy, pierwszy, drugi i trzeci moment widmowy, kurtozę, współczynnik mocy względnej, amplitudę i częstotliwość 1-, 2-, 3- formantu, wartość maksymalną i minimalną sygnału, 10 współczynników mel-cepstralnych, stosunek harmonicznym do szumu. Parametry te opisują periodyczność, zawartość szumów oraz nieliniowość procesu fonacji.

Periodyczność w mowie obejmuje zdolność do generowania ciągłego przepływu powietrza podczas produkcji samogłosek o przedłużonej fonacji. Stabilność i niezmienność takiego przepływu przez struny głosowe może zostać scharakteryzowana parametrycznie poprzez zmienność amplitudy i/lub częstotliwości.

Odcinek czasu pomiędzy kolejnymi zvarciami fałdów głosowych wyznacza najmniejszą powtarzającą się sekwencję w sygnale mowy. Odcinek ten nazywany jest okresem podstawowym. Odwrotność tego okresu definiuje częstotliwość podstawową (F_0). F_0 jest jednym z ważniejszych parametrów charakteryzujących źródło mowy dźwięcznej. Jego wartość zależy m.in. od płci, wieku, stanu emocjonalnego oraz zdrowotnego mówcy. Dla kobiet wartość F_0 mieści się w przedziale 160-960 Hz, a dla mężczyzn 80-480 Hz. Na podstawie parametru F_0 i badań wyizolowanych głosek dźwięcznych o przedłużonej fonacji możliwe jest: wykrycie anomalii struktury anatomicznej, określenie przyczyny ograniczenia czynności ruchowej fałdów głosowych, które występują w różnych stanach patologicznych, ocena poprawy stanu głosu w czasie leczenia i rehabilitacji, klasyfikacja stopnia depresji pacjenta oraz rozpoznanie emocji mówcy [69]. Do wyznaczenia parametru F_0 można wykorzystać kilka metod, a jedna z nich wykorzystuje analizę cepstralną.

Cepstrum definiowane jest jako logarytm dziesiętny widma częstotliwościowego sygnału (otrzymanego w wyniku szybkiej transformacji Fouriera, FFT) poddanej odwrotnej transformacji Fouriera (IFFT). Położenie maksimum w dziedzinie cepstrum odpowiada częstotliwości

podstawowej F_0 . Algorytm oparty na analizie cepstralnej jest relatywnie niewrażliwy na szum. Wyróżnia się cepstrum zespolone oraz rzeczywiste (cepstrum mocy). Równanie (3.1) opisuje wyznaczenie cepstrum rzeczywistego C [70]:

$$C = IFFT(\log_{10}|FFT(x)|) \quad (3.1)$$

gdzie FFT - dyskretna transformacja Fouriera, $IFFT$ - odwrotna, dyskretna transformacja Fouriera, x - dyskretny, poddany procesowi segmentacji i okienkowania sygnał akustyczny.

Zmienność częstotliwości podstawowej w kolejnych cyklach można opisać za pomocą współczynnika *jitter* (J) oraz *shimmer* (S) [71]. Współczynnik *jitter* pozwala na określenie zmienności częstotliwości podstawowej w kolejnych fragmentach sygnału f_i w stosunku do częstotliwości średniej i jest określony wzorem:

$$J = \frac{\frac{1}{R-1} \sum_{i=1}^{R-1} |f_i - f_{i+1}|}{\frac{1}{R} \sum_{i=1}^R f_i} \quad (3.2)$$

R - liczba ramek sygnału akustycznego.

Do opisu krótkookresowych zmian częstotliwości podstawowej można wykorzystać współczynnik zaburzeń okresu (ang. *period perturbation quotient*, *ppq*). Jest to względna zmiana częstotliwości podstawowej w analizowanym sygnale z wygładzeniem dla zdefiniowanej liczby okresów. W pracy wykorzystano 5 okresów (*ppq5*) oraz 11 okresów (*ppq11*).

Drugi parametr - współczynnik *shimmer* określa zmienność amplitudy tonu podstawowego w kolejnych fragmentach sygnału A_i w stosunku do amplitudy średniej tonu podstawowego i jest wyrażony wzorem:

$$S = \frac{\frac{1}{R-1} \sum_{i=1}^{R-1} |A_i - A_{i+1}|}{\frac{1}{R} \sum_{i=1}^R A_i} \quad (3.3)$$

Krótkoterminowe zmiany amplitudy głosu z wygładzaniem dla kolejnych okresów tonu krtańowego opisuje współczynnik zaburzeń amplitudy (ang. *amplitude perturbation quotient*, *apq*). W pracy wykorzystano trójokresowy (*apq3*) oraz pięciookresowy (*apq5*) współczynnik zaburzeń amplitudy.

Miarą energii niesionej przez sygnał x_i dla $i=1,2,\dots,R$ jest parametr E [72]:

$$E = \sum_{i=1}^R x_i^2 \quad (3.4)$$

Kształt widma sygnału mowy $G(t, f)$ może zostać opisany za pomocą wielu parametrów. W niniejszych badaniach wykorzystano momenty widmowe, gdzie moment m -tego rzędu w j -tej chwili czasowej jest opisany zależnością [73]:

$$M_m(t_j) = \sum_{f_k=f_d}^{f_g} |G(t_j, f_k)| \cdot [f_k]^m \quad (3.5)$$

gdzie:

$G(t_j, f_k)$ stanowi widmo czasowo-częstotliwościowe w chwili czasowej t_j ,

f_k - częstotliwość środkowa k -tego pasma wyróżnionego w analizie częstotliwościowej,

f_d, f_g - częstotliwości ograniczające pasmo wyznaczenia momentu widmowego.

Moment 0-rzędu wykorzystywany jest do normalizacji momentów wyższych rzędów. Momenty unormowane definiowane są jako stosunek momentu m -tego rzędu M_m i momentu 0-rzędu M_0 , co opisuje poniższe równanie:

$$M_{unorm(m)} = \frac{M_m}{M_0} \quad (3.6)$$

Moment normalizujący zerowego rzędu jest wykorzystywany w analizie mowy do wykrywania różnic pomiędzy głoskami dźwięcznymi i bezdźwięcznymi. Wyrażony jest wzorem [74]:

$$M_0(t) = \sum_{f_k=f_d}^{f_g} |G(t, f_k)| \quad (3.7)$$

Unormowany moment pierwszego rzędu zawiera informację o środku ciężkości widma i jest opisany zależnością [74]:

$$M_1(t) = \frac{\sum_{f_k=f_d}^{f_g} |G(t, f_k)| \cdot f_k}{M_0(t)} \quad (3.8)$$

Unormowany moment centralny drugiego rzędu jest kwadratem szerokości widma. Moment ten pozwala różnicować typowe deformacje widma. Opisuje go zależność:

$$M_2(t) = \frac{\sum_{f_k=f_d}^{f_g} |G(t, f_k)| \cdot [f_k - M_1(t)]^2}{M_0(t)} \quad (3.9)$$

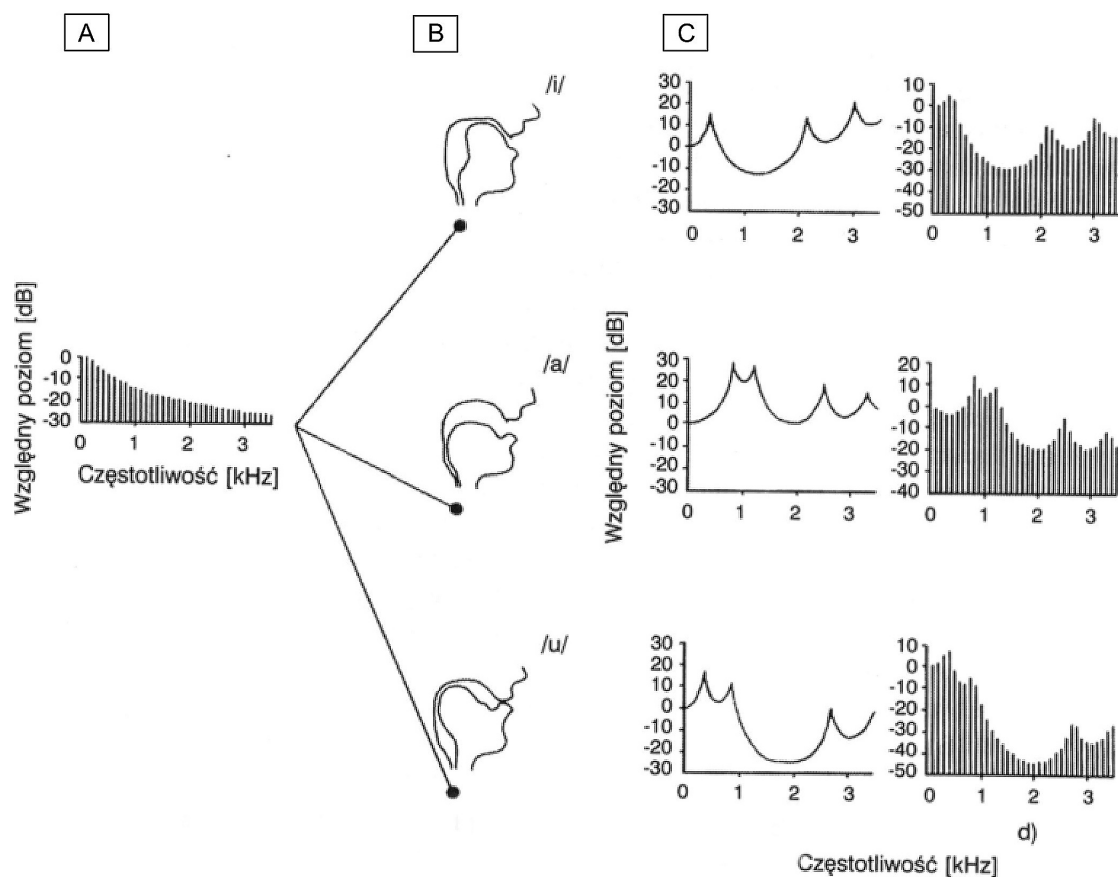
Informację na temat skośności widma opisuje unormowany moment centralny trzeciego rzędu:

$$M_3(t) = \frac{\sum_{f_k=f_d}^{f_g} |G(t, f_k)| \cdot [f_k - M_1(t)]^3}{M_0(t)} \quad (3.10)$$

Na podstawie obliczonych unormowanych momentów centralnych rzędu 2 i 4 wyznaczana jest kurtoza, która stanowi miarę płaskości widma sygnału [72]:

$$kurtosis = \frac{M_4(t)}{M_2(t)^2} \quad (3.11)$$

Kanał głosowy można przedstawić jako układ filtrów (rezonatorów) o określonych wartościach częstotliwości rezonansowej. Rezonatorami w trakcie głosowym są: jama podgłośniowa, tchawica, klatka piersiowa i oskrzela, krtań, jama gardłowa, nosowa i ustna. Sposób generacji dźwięków wybranych samogłosek przedstawiony jest na rys. 3.2. Po przejściu przez układ



Rys. 3.2. Sposób generacji dźwięków wybranych samogłosek. A - widmo dźwięku (tonu krtańowego) wytwarzanego w wyniku drgań fałdów głosowych, B- przekroje poprzeczne kanału głosowego w konfiguracjach odpowiadających różnym samogłoskom, C- charakterystyki przeniesienia kanału głosowego i widma samogłosek po przejściu tonu krtańowego (A) przez charakterystyki filtrów [76].

takich filtrów, widmo tonu krtaniowego charakteryzuje się lokalnymi maksimumami i minimumami. Takie obszary nazywane są formantami i oznaczane są odpowiednio według rosnących częstotliwości: $F1$ - pierwszy formant, $F2$ - drugi formant itd. Ze względu na fizyczne możliwości narządów głosu, a dokładnie na szerokość pasma mowy ludzkiej, która jest ograniczona do 7-8 kHz [75], w mowie występuje maksymalnie 8 formantów. Słuch jest niezwykle czuły na zmiany częstotliwości formantowych, przez co możliwe jest skuteczne rozpoznawanie danych głosek w mowie.

Formanty można opisać za pomocą częstotliwości i amplitudy. Położenie formantów na osi częstotliwości jest ściśle związane z wartościami częstotliwości rezonansowej traktu głosowego i zależy od jego kształtu. Częstotliwość środkowa każdego z formantów jest inna i ściśle związana z kształtem kanału głosowego. Zależy ona zarówno od wypowiedzianej samogłoski jak i od cech indywidualnych mówcy. Na wartość częstotliwości formantu $F1$ ma wpływ ustawienie języka w jamie ustnej. Wartość formantu $F2$ warunkuje ruch języka przed-tył. Formanty $F3$ i wyższe zależą głównie od długości traktu głosowego i ich częstotliwości rezonansowe zmieniają się niewiele podczas wypowiedzania różnych samogłosek. Również poziomy poszczególnych formantów (wyrażone względem formantu F_1) wykazują stałe w przybliżeniu wartości [76].

Metoda wydzielenia częstotliwości formantowych polega na modelowaniu sygnału mowy tak, jakby był on generowany przez jedno źródło i kształtowany przez kilka różnych filtrów. Punktem wyjścia do wyznaczenia parametrów formantowych jest tzw. wygładzone widmo sygnału (obwiednia widmowa). Do jej wyznaczenia w niniejszej pracy wykorzystano metodę liniowego kodowania predykcyjnego (LPC, ang. *linear predictive coding*). Jest to technika analizy sygnału mowy polegająca na interpretacji tego sygnału jako odpowiedzi filtru biegunowego (a konkretnie filtru o nieskończonej odpowiedzi impulsowej, IIR - ang. *infinite impulse response*) na sygnał tonu krtaniowego. Istotnym elementem przy obliczaniu formantów jest określenie współczynników filtru LPC, tak aby minimalizować odległość między sygnałem, a jego predykcją. Do określenia liczby współczynników wybrano metodę *rule of thumb* [77]. Liczba współczynników jest równa wartości częstotliwości próbkowania wyrażonej w kHz powiększonej o 2. Dla przykładu częstotliwość próbkowania równa 10 kHz plus 2 daje 12 współczynników [78].

W pracy wykorzystano również współczynnik mocy względnej, który określa stosunek mocy sygnału w wybranym paśmie częstotliwościowym do mocy sygnału w paśmie (0:12000 Hz) [73]. Moc względna opisana jest równaniem 3.12 [73, 74].

$$W = \frac{\sum_{t=t_{beg}}^{t_{end}} \sum_{f_k=f_d}^{f_g} G(t, f_k)}{\sum_{t=t_{beg}}^{t_{end}} \sum_{f=0}^{12000} G(t, f_k)} \quad (3.12)$$

gdzie:

Tabela 3.3. Zakresy częstotliwości formantów dla mowy niemieckiej, $F1$ - 1 formant, $F2$ - drugi formant, $F3$ - 3 formant, f_g oznacza górną granicę, f_d dolną granicę częstotliwości[79].

samogłoska	kobiety						mężczyźni					
	$F1$ [Hz]		$F2$ [Hz]		$F3$ [Hz]		$F1$ [Hz]		$F2$ [Hz]		$F3$ [Hz]	
	f_g	f_d	f_g	f_d	f_g	f_d	f_g	f_d	f_g	f_d	f_g	f_d
/a/	651	838	1346	1583	2679	2983	529	674	1224	1386	2317	2618
/e/	382	495	1949	2472	2691	3055	328	436	1700	2006	2324	2614
/i/	292	385	2125	2496	2644	3000	266	337	1813	2106	2328	2668

Tabela 3.4. Zakresy częstotliwości formantów dla mowy polskiej, $F1$ - 1 formant, $F2$ - drugi formant, $F3$ - 3 formant, f_g oznacza górną granicę, f_d dolną granicę częstotliwości.

samogłoska	kobiety						mężczyźni					
	$F1$ [Hz]		$F2$ [Hz]		$F3$ [Hz]		$F1$ [Hz]		$F2$ [Hz]		$F3$ [Hz]	
	f_g	f_d	f_g	f_d	f_g	f_d	f_g	f_d	f_g	f_d	f_g	f_d
/a/	650	900	1398	1750	2566	2799	514	984	1020	1475	2149	2712
/e/	451	620	1587	1852	2415	3189	398	666	1653	2124	2154	3007
/i/	175	358	2169	2401	2035	3297	215	317	1938	2183	2002	3154

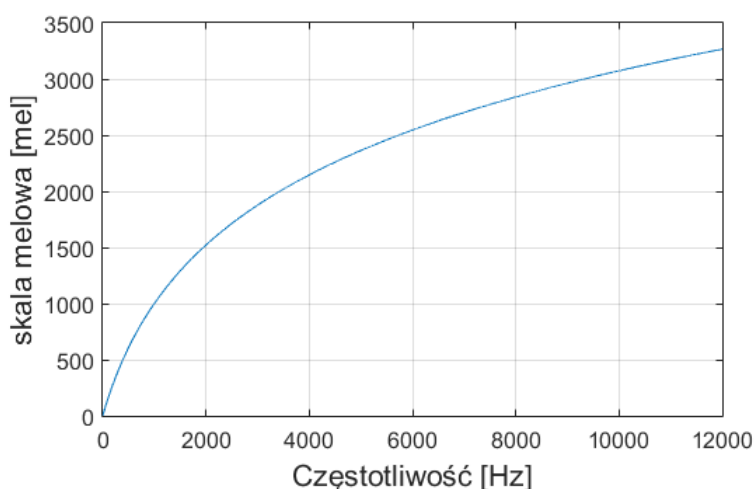
t_{beg}, t_{end} - początek i koniec nagrania sygnału mowy.

Pasma częstotliwościowe do wyznaczenia mocy względnej zostały dobrane zgodnie z występowaniem formantów mowy. Dla języka niemieckiego zakres został dobrany na podstawie artykułu [79], a wyniki zaprezentowane są w tabeli 3.3.

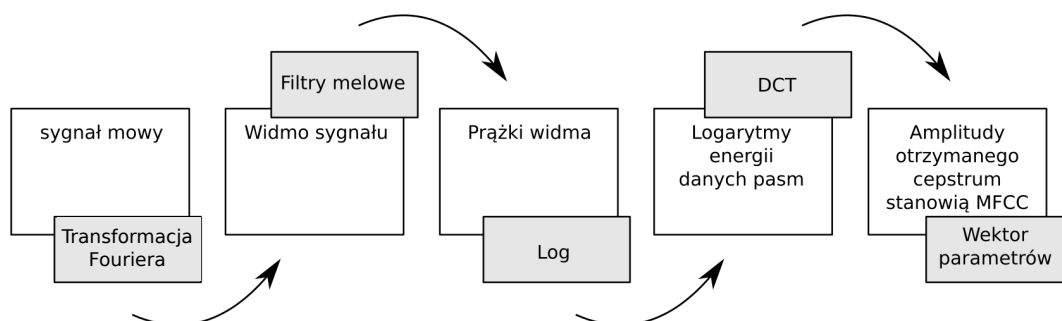
W tabeli 3.4 przedstawiono średnie zakresy częstotliwości trzech pierwszych formantów dla mowy polskiej. Wartości otrzymano na podstawie analizy danych z trzech powtórzeń danej samogłoski wypowiedzianych przez 30 osób zdrowych, dla których język polski był językiem ojczystym.

Mechanizm słyszenia ucha ludzkiego charakteryzuje nieliniowa percepcja częstotliwości sygnałów akustycznych. Analiza sygnału mowy w sposób adekwatny do zjawiska słyszenia, wymaga przekształcenia skali częstotliwości z wykorzystaniem podejścia perceptualnego tzw. cepstrum melowego. Odpowiada ono subiektywnemu wrażeniu wysokości dźwięku. Podstawą tego podejścia jest przekształcenie widma sygnału do melowej skali częstotliwości. Po transformacji do skali melowej, obliczane jest cepstrum. Jako punkt odniesienia przyjmuje się częstotliwość 1 kHz przy poziomie ciśnienia akustycznego o 40 dB ponad progiem słyszenia człowieka i oznacza się go jako 1000 meli.

Wrażenie wysokości dźwięku zależy od głośności, stąd w definicji skali przyjęto poziom natężenia 40 dB. Oznacza to, że skala widma sygnału jest zbliżona do liniowej, ale tylko do częstotliwości 1 kHz, a powyżej tej wartości zależność przechodzi w logarytmiczną, co jest opisane zależnością (3.13) [80]. Charakterystyka ilustrująca związek pomiędzy skalą melową,



Rys. 3.3. Związek między skalą częstotliwości, a skalą melową.



Rys. 3.4. Schemat blokowy wyznaczania współczynników MFCC.

a skalą częstotliwości przedstawiona jest na rys 3.3.

$$F_{mel} = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (3.13)$$

W badaniach prawidłowości sygnału akustycznego stosowane są skale perceptualne do wyznaczenia współczynników mel-cepstralnych MFCC (ang. *mel – frequency cepstral coefficients*).

Proces wyznaczenia współczynników MFCC składa się z poszczególnych etapów:

- preemfaza oraz okienkowanie. Celem preemfazy było wzmocnienie składowych o wysokiej częstotliwości i osłabienie składowych o niskiej częstotliwości. Wzmocnienie wysokich częstotliwości sygnału mowy o ponad 20 dB, skutkuje dużą odpornością na zakłócenia pochodzące z otoczenia. Podczas procedury okienkowania sygnał jest dzielony na odcinki czasu z określoną długością pokrywaniem się kolejnych odcinków (ang.

overlapping). Sygnał akustyczny zachowuje stacjonarny charakter dla odcinków o długości 20-30 ms [81]. Zastosowanie pokrywania kolejnych okien sygnału pozwala na zwiększenie rozdzielczości czasowej wyników analizy oraz zmniejszenie zniekształceń sygnału spowodowanych zastosowaniem procedury okienkowania. Kolejne okna sygnału pokrywają się najczęściej w 1/3-1/2 długości okna [81]. Do obliczeń wykorzystano okno o długości 25 ms z przesunięciem 10 ms. Aby ograniczyć wyciek częstotliwości każda ramka poddawana jest kształtowaniu oknem Hamminga,

- FFT (szybka transformacja Fouriera, ang. *fast Fourier transform*) pozwalająca na przekształcenie sygnału z dziedziny czasu do dziedziny częstotliwości,
- filtracja melowa, czyli wykorzystanie banku filtrów dla kolejnych, zachodzących w połowie na siebie pasm częstotliwości, rozmieszczonych równomiernie na nieliniowej skali perceptualnej. Bank filtrów wykorzystany w obliczeniach przeprowadzonych w pracy składał się z 12 filtrów o trójkątnych charakterystykach równomiernie rozmieszczonych w skali melowej,
- logarytmowanie widma sygnału, co pozwala na rozróżnienie poszczególnych składowych sygnału,
- dyskretna transformacja kosinusowa (DCT, ang. *Discrete Cosine Transform*) obliczona zgodnie ze wzorem [82, 83]:

$$MFCC_n = \sqrt{\frac{2}{L}} \sum_{i=1}^L \log(S_i) \cdot \cos\left[\frac{\pi \cdot n}{L} \cdot (i - 0.5)\right] \quad (3.14)$$

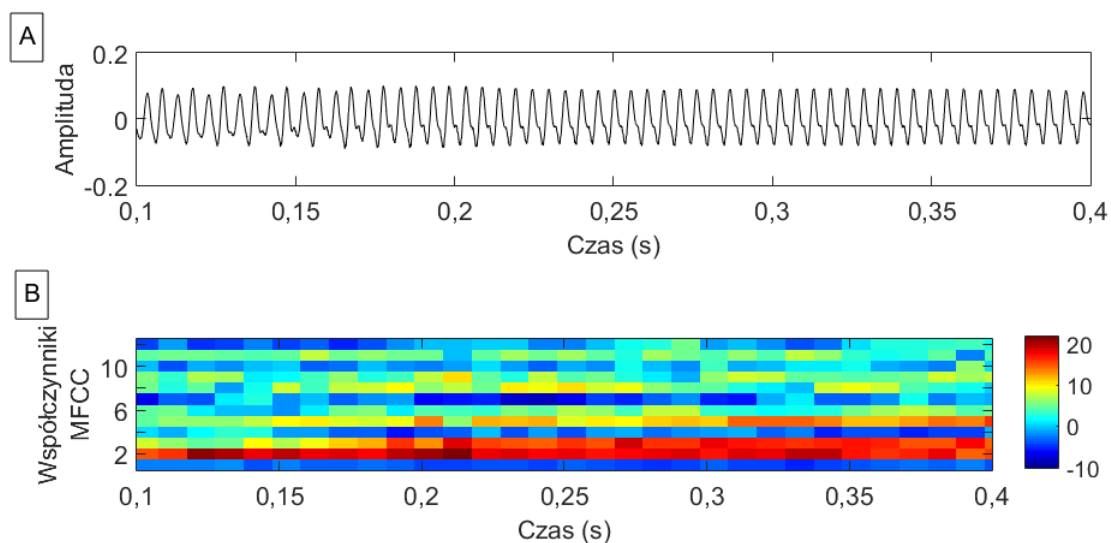
gdzie:

$MFCC_n$ - n -ty współczynnik mel-cepstralny,

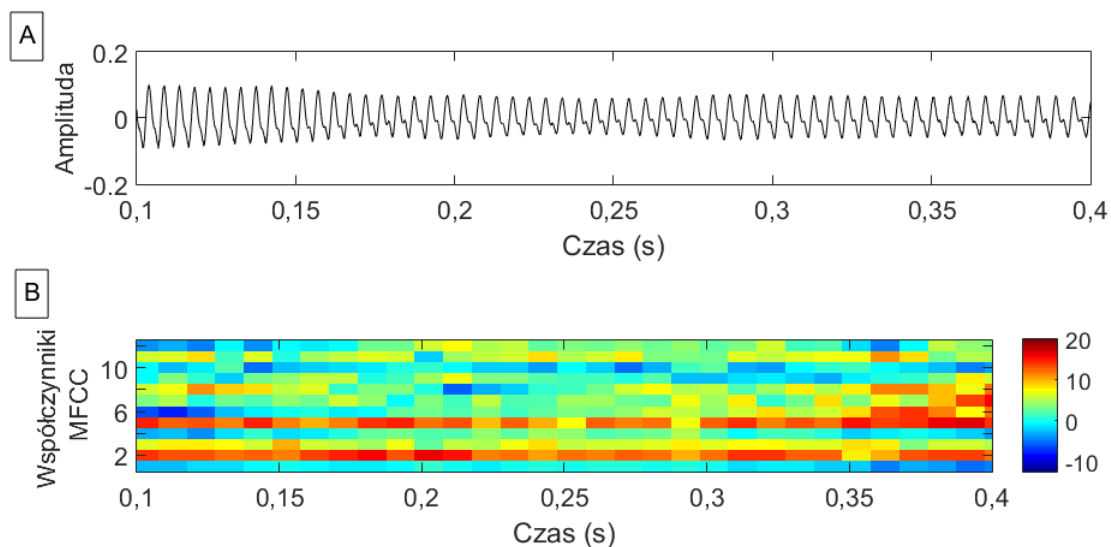
S_i - wartość estymaty gęstości widmowej mocy uśredniona przy użyciu i -tego filtra,

L - liczba filtrów melowych.

Otrzymany wektor MFCC ma długość równą liczbie pasm.



Rys. 3.5. (A) Przebieg czasowy oraz (B) czasowa zmienność współczynników MFCC dla samogłoski /i/, pacjent z chorobą Parkinson'a, lat 71, stan *off*, UPDRS 61.



Rys. 3.6. (A) Przebieg czasowy oraz (B) czasowa zmienność współczynników MFCC dla samogłoski /i/, pacjent z chorobą Parkinson'a, 60 minut od spożycia leku, lat 71, UPDRS 22.

Wyższe (powyżej 12) współczynniki MFCC reprezentują szybkie zmiany w sygnale, co może niekorzystnie wpływać na obiektywny opis sygnału akustycznego [84]. Ogólny schemat przedstawiający algorytm wydzielający wektor parametrów MFCC na podstawie sygnału mowy przedstawiono na rys. 3.4. Na rysunkach odpowiednio 3.5 oraz 3.6 przedstawiono przykładowe przebiegi czasowe sygnału mowy, widmo w skali melowej oraz współczynniki MFCC samogłoski /i/ dla pacjenta cierpiącego z powodu choroby Parkinson'a, którego głos zarejestrowano w stanie *off* (rys. 3.5), oraz 60 minut po zażyciu leku (rys. 3.6). Z porównania wartości współczynników MFCC widoczne jest, że u pacjenta w stanie *off* występują wyższe wartości niż w stanie *on*.

U osób z chorobą Parkinson'a występują niekontrolowane ruchy fałdów głosowych oraz dochodzi do ich niekompletnego zamknięcia. Do opisu tego zjawiska wykorzystano stosunek części harmonicznej do szumu (HNR - *Harmonic to Noise Ratio*). Pod tym pojęciem rozumie się ilość energii przekazywanej w częstotliwości podstawowej i jej harmonicznym (części periodycznej sygnału), podzieloną przez energię w zakresie częstotliwości szumów. Parametr ten związany jest z percepcją szorstkości głosu i chrypki. Zdrowe osoby mają niski poziom szumu i wysoką wartość parametru HNR. W przypadku osób chorych, parametr HNR przyjmuje mniejsze wartości. W patologii mowy występuje zwiększony szum spowodowany turbulentnym przepływem powietrza, wynikający z niepełnego domknięcia fałdów głosowych. Stąd też należy w takim przypadku oczekiwać mniejszej wartości parametru HNR. Parametr ten obliczono w oparciu o prace [85, 86].

3.3. Informacja diagnostyczna w sygnale mowy

Do opisu sygnału akustycznego w kontekście poszukiwania jego nieprawidłowości wykorzystuje się wiele parametrów. Parametry stosowane w tej rozprawie zostały przedstawione w rozdziale 3.2. W efekcie tworzone są wektory opisujące analizowane przypadki. Taki opis jest trudny w interpretacji, przez co konieczne jest użycie metody redukującej wymiar wektora i wskazującej parametry najbardziej informacyjne i użyteczne w procesie klasyfikacji. W tym celu najczęściej wykorzystywane są metody statystyczne (np. analiza składowych głównych). W kolejnym kroku opracowywany jest algorytm klasyfikacyjny, a następnie należy zdefiniować zasadę określenia klasy przynależności analizowanego wektora. Na ogół jest to realizowane na etapie uczenia algorytmu klasyfikacji. W tym celu wykorzystywany jest zbiór danych uczących. W przypadku klasyfikacji binarnej wyróżniane są dwie klasy uwzględniające tylko osoby zdrowe oraz taka, do której przynależą tylko osoby posiadające zdiagnozowaną chorobę laryngologiczną. W pracy opracowano algorytm klasyfikacji binarnej oraz multiklasyfikacji, która wyróżnia 4 klasy. 3 z nich wskazywały na 3 różne patologie głosu, a jedna klasa zawierała

przypadki tylko osób zdrowych. W wyniku takiej klasyfikacji możliwe było nie tylko rozróżnienie pomiędzy pacjentem zdrowym lub chorym, ale również określenie jaką patologię głosu może posiadać. Jako narzędzie klasyfikacyjne wykorzystano metodę lasów losowych oraz sieci neuronowe, które zostaną szczegółowo przedstawione w dalszej części rozdziału.

3.3.1. Nieliniowa analiza składowych głównych

Uszeregowanie parametrów sygnału mowy jest prowadzone pod kątem ich zdolności dyskryminacyjnych. W tym celu można wykorzystać metody liniowe (np. analizę składowych głównych, PCA) lub metody nieliniowe (np. jądrową analizę składowych głównych lub nie-liniowe PCA). Celem PCA jest liniowa transformacja parametrów do nowego układu współrzędnych oraz ich uszeregowanie z punktu widzenia największej wariancji. Te nowe współrzędne nazywamy składowymi głównymi. Metoda ta opiera się na wektorach własnych macierzy kowariancji, wyznaczonej dla wektorów zawierających parametry sygnału mowy. W wyniku analizy PCA otrzymujemy tyle składowych głównych ile wymiarów miał wektor oryginalny, uszeregowanych w kolejności malejącej wariancji. W przypadku, gdy liczba analizowanych parametrów nie może zostać zredukowana w wyniku liniowej transformacji, redukcja wymiaru oryginalnych wektorów wymaga rzutowania ich na krzywą, a nie prostą redukcji [4, 56]. W takim przypadku mówimy o metodach nieliniowych. Przykładem metody realizującej nieliniową transformację danych wielowymiarowych jest jądrowa analiza składowych głównych (ang. *kernel principal component analysis*, kPCA). Metoda ta polega na odwzorowaniu danych na rozszerzoną przestrzeń parametrów poprzez pewne nieliniowe odwzorowanie. Wymiar nowej przestrzeni może być dowolny, mniejszy lub nawet większy niż wymiar oryginalnego wektora parametrów. Aby zrealizować nieliniową transformację parametrów można zastosować tzw. sztuczkę jądra (ang. *kernel trick*) [87]. W pierwszym etapie tego algorytmu należy zapisać dane w postaci macierzy, której kolumny zawierają wartości parametrów wyznaczone dla kolejnego pacjenta. Wówczas zbiór danych ma postać $\{xp_i\}$, gdzie $i=1,2 \dots, M$, M - liczba pacjentów, a każdy xp_i stanowi D -wymiarowy kolumnowy wektor parametrów, gdzie $xp_i=[p_{i1},p_{i2} \dots p_{iD}]$, p - wynik danego parametru. Następnie należy wyznaczyć macierz kowariancji, która w klasycznym PCA ma postać:

$$C = \frac{1}{M} \sum_{i=1}^M (xp_i - \overline{xp})(xp_i - \overline{xp})^T \quad (3.15)$$

gdzie:

xp_i - kolejne wektory parametrów, $i=1, \dots, M$,

\overline{xp} - wartość średnia:

$$\bar{x}_p = \frac{1}{M} \sum_{i=1}^M x_{p_i} \quad (3.16)$$

Kolejnym krokiem jest wyznaczenie wektorów własnych v i wartości własnych λ macierzy kowariancji C spełniających równanie:

$$\lambda v = Cv \quad (3.17)$$

dla wartości własnych $\lambda \geq 0$ oraz $v \in \mathbb{R}^M \setminus \{0\}$.

W wyniku zastosowania nieliniowej transformacji $\phi(x_{p_i})$, gdzie każdy $x_{p_i} = [p_{i1}, p_{i2}, \dots, p_{iD}]$, p - wynik danego parametru, D -wymiar przestrzeni parametrów, otrzymuje się nową przestrzeń parametrów W , gdzie $W > D$. Wówczas każdy punkt x_{p_i} będzie przedstawiony poprzez $\phi(x_{p_i})$.

W pierwszym etapie należy dane znormalizować, tak aby $\sum_{i=1}^M \phi(x_{p_i}) = 0$, a następnie wyznaczyć macierz kowariancji [88]:

$$C_1 = \frac{1}{M} \sum_{i=1}^M \phi(x_{p_i}) \phi(x_{p_i})^T \quad (3.18)$$

W takim przypadku wektory własne oraz wartości własne można przedstawić za pomocą:

$$C_1 v_k = \lambda_k v_k \quad (3.19)$$

gdzie:

$k = 1, 2, \dots, W$. Na podstawie równań 3.18 oraz 3.19 otrzymuje się wyrażenie:

$$\frac{1}{M} \sum_{i=1}^M \phi(x_{p_i}) \{ \phi(x_{p_i})^T v_k \} = \lambda_k v_k \quad (3.20)$$

Z równania 3.20 wynika, że wartości własne są powiązane liniowo z $\phi(x_{p_k})$, co można zapisać wprowadzając współczynnik α_{ki} .

$$v_k = \sum_{i=1}^M \alpha_{ki} \phi(x_{p_i}) \quad (3.21)$$

Połączenie równań 3.20 oraz 3.21 doprowadza do równania:

$$\frac{1}{M} \sum_{i=1}^M \phi(x_{p_i}) \phi(x_{p_i})^T \sum_{j=1}^M \alpha_{kj} \phi(x_{p_j}) = \lambda_k \sum_{i=1}^M \alpha_{ki} \phi(x_{p_i}) \quad (3.22)$$

Najważniejszym punktem obliczeń jest wyznaczenie funkcji jądra $\kappa(x_{p_i}, x_{p_j}) = \phi(x_{p_i})^T \phi(x_{p_j})$

Wykorzystując funkcję jądra równanie (3.22) można przedstawić jako [88]:

$$K^2 \alpha_k = \lambda_k M K \alpha_k \quad (3.23)$$

gdzie: $K_{i,i} = \kappa(xp_i, xp_j)$, a a_k stanowi M -wymiarową kolumnę wektora a_{ki} :

$$a_k = [a_{k1} a_{k2} \dots a_{kM}]^T \quad (3.24)$$

Do wyznaczenia jądrowych składowych głównych (parametrów w nowej przestrzeni W) należy rozwiązać równanie:

$$y_k(xp) = \phi(xp)^T v_k = \sum_{i=1}^M a_{ki} \kappa(xp, xp_i) \quad (3.25)$$

W niniejszej pracy jako jądro przekształcenia wykorzystano funkcję Gaussa.

Podstawowym parametrem jądrowej analizy składowych głównych jest szerokość jądra σ . W przypadku, gdy celem jest rozróżnienie różnych klas w nowej przestrzeni parametrów, pożądane jest aby σ była mniejsza niż odległość pomiędzy klasami. W badaniach przeprowadzonych w ramach pracy wykorzystano wskazówkę zawartą w pracy [89] i opisaną wzorem 3.26. Zapewniana, że σ ma odpowiednią wartość, aby uwzględnić położenie sąsiadujących punktów w danej klasie oraz ograniczyć wpływ klas sąsiadujących.

$$\sigma = qd_i^{MM} \quad (3.26)$$

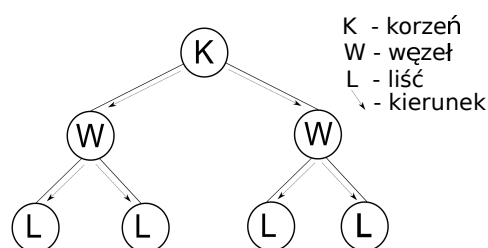
gdzie:

q - wartość dobrana eksperymentalnie,

d_i^{MM} - odległość pomiędzy wektorem xp_i oraz wektorem znajdującym się najbliżej w przestrzeni parametrów (najbliższym sąsiadem).

3.3.2. Lasy losowe

Jedną z najskuteczniejszych i popularnych metod klasyfikacji są drzewa decyzyjne (ang. *decision trees*), zwane inaczej drzewami klasyfikacyjnymi (ang. *classification trees*) [90]. Są one przykładem wielopoziomowego procesu decyzyjnego. Algorytm drzew decyzyjnych opiera się na drzewiastej strukturze. W ich budowie można wyróżnić jeden wierzchołek, tzw. korzeń oraz klasy terminalne, tzw. liście [91]. Każdy wierzchołek drzewa, który nie jest liściem nazywa się węzłem drzewa, a odcinki łączące kolejne węzły gałęziami. W każdym węźle wykonywany jest podział według określonego kryterium dotyczącego danej obserwacji. Na podstawie jego wyniku wybierana jest określona gałąź. Jako kryterium podziału może być stosowana miara informacji wyrażona jako indeks Gini lub entropia wzajemna (ang. *cross – entropy*). W procesie klasyfikacji, od korzenia do liścia może prowadzić tylko jedna droga. Przykładowa struktura drzewa decyzyjnego przedstawiona jest na rysunku 3.7. Zaletą drzew decyzyjnych jako algorytmu klasyfikacji jest niska złożoność obliczeniowa i objętość pamięciowa [92].



Rys. 3.7. Przykładowa struktura drzewa decyzyjnego.

W procesie klasyfikacji pacjentów zdrowych i tych, którzy posiadają chorobę laryngologiczną wykorzystano metodę lasów losowych (ang. *random forest*, RF) polegającą na wykorzystaniu wielu drzew decyzyjnych, przez co bardzo dobrze nadaje się ona do badania problemów opisanych przez wiele parametrów. Drzewa w lasach losowych nie podlegają operacji przycinania (usuwania zbędnych elementów drzewa), powinny osiągać maksymalny możliwy rozmiar. Lasy losowe nie wymagają wiedzy eksperckiej i są niepodatne na nadmierne dopasowanie się do danych (przeuczenie lub przetrenowanie).

W celu ewaluacji algorytmu klasyfikacji zbiór wejściowy jest dzielony na dwa podzbiory: zbiór uczący oraz zbiór OOB (ang. *out-of-bag*). Na podstawie zbioru OOB można określić ile elementów zbioru uczącego zostało błędnie sklasyfikowanych.

Ważnym elementem podczas tworzenia drzew jest określenie kryterium podziału, zwanego inaczej testem, który odbywa się w węzłach. W jego wyniku zostaje określona przynależność testowanego parametru do bezpośredniego potomka rozważanego węzła. W badaniach pracy wykorzystano testy binarne, w rezultacie czego każdy węzeł ma dwóch potomków. Najważniejszym kryterium podziału jest uzyskanie możliwie małego rozrzutu danych w obrębie tych samych klas oraz dużej odległości w kolejnych węzłach pomiędzy wektorami parametrów przynależnymi do różnych klas. W efekcie działania algorytmu lasów losowych możliwe jest wyznaczenie rankingu zmiennych pod względem wartości predykcyjnych. W takim celu oblicza się średnie zmiany indeksu Giniego [93]. Im niższa jest wartość tego indeksu, dana zmienna charakteryzuje się lepszymi właściwościami predykcyjnymi.

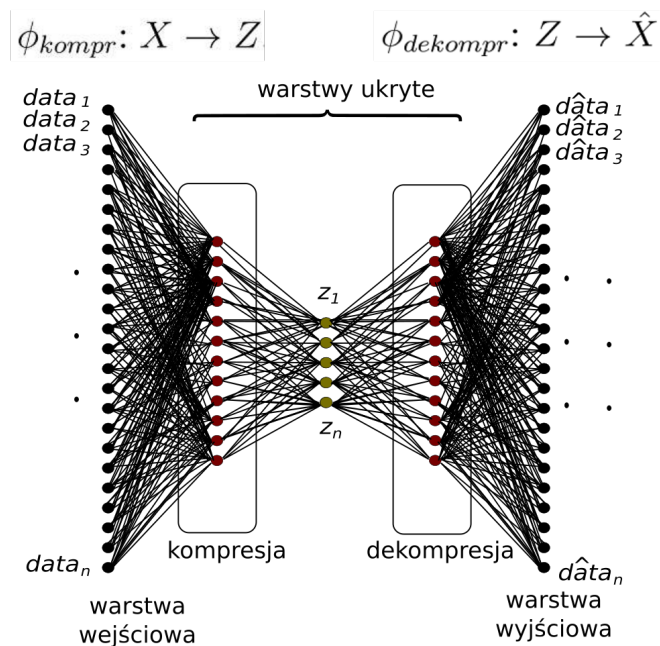
3.3.3. Sieci neuronowe

Sztuczne sieci neuronowe (ANN, ang. *artificial neural networks*) są nowoczesnymi i coraz częściej wykorzystywanymi systemami do analizy i przetwarzania informacji w wielu dziedzinach. Jednak nawet dobrze zaprojektowana sieć nie zastąpi całkowicie procesu diagnostycznego, może natomiast wskazać lekarzowi sugestię diagnozy. Sieci neuronowe są szeroko wykorzystywane do klasyfikacji i rozpoznawania różnych obiektów na podstawie ich charakterystycznych cech i parametrów.

W budowie sieci wyróżnia się warstwy: wejściową, wyjściową i ukryte. Projektowanie sieci składa się z 4 etapów: inicjalizacji, uczenia, walidacji i testowania. Etap inicjalizacji uwzględnia przyporządkowanie wag sygnałom wejściowym. Różnicują one znaczenie poszczególnych wejść, im wyższa wartość wagi tym dany sygnał jest ważniejszy przy obliczaniu odpowiedzi neuronu. Czynniki mającymi wpływ na ustalenie wag oraz dobór prawidłowej struktury sieci są: liczba warstw ukrytych, liczba neuronów w każdej z warstw, funkcja aktywacji sieci. Ostateczne wyniki działania danej sieci neuronowej (algorytmu) otrzymuje się w efekcie procesu uczenia, który polega na ustaleniu wag dla poszczególnych neuronów. Zbiór uczący zawiera przykładowe zadania wraz z ich poprawnymi rozwiązaniami (np. wynikiem diagnostycznym). Dla potrzeb oceny działania sieci neuronowej dane wejściowe dzieli się na trzy zbiory: uczący, walidacyjny i testowy. Zbiór uczący zawiera przykłady do nauczenia sieci, zaś zbiór walidacyjny przykłady do poprawienia wartości parametrów sieci i jej architektury. Zbiór testowy służy do określenia jakości sieci.

Aby klasyfikacja była skuteczna, bardzo ważny jest wybór i przygotowanie danych wejściowych. Duża ilość danych wejściowych utrudnia proces interpretacji i ogranicza możliwość szybkiego podjęcia decyzji. W przypadku wielowymiarowych danych wejściowych w postaci parametrycznej, zależności pomiędzy poszczególnymi parametrami w obrębie różnych klas nie są znane. W takim przypadku celowe jest przedstawienie danych w nowej przestrzeni wielowymiarowej i wzmacnianie tym samym ich zdolności dyskryminacyjnych pomiędzy analizowanymi klasami. Takie zadanie można rozwiązać przy użyciu sieci neuronowej, która zapewnia nieliniową transformację danych. Sieć taka to tzw. nieliniowe PCA (NLPCA). W niniejszej pracy wykorzystano nieliniowe PCA, które bazuje na auto-asocjacyjnych wielowarstwowych sieciach neuronowych [94–96]. Sieć taka składa się z minimum trzech warstw: warstwy wejściowej, wyjściowej oraz ukrytej. Działanie NLPCA opiera się na wielowarstwowym perceptronie (ang. *multi – layer perceptron*, MLP) z auto-asocjacyjną topologią. Taka sieć zwana jest również jako *autoencoder* lub sieć typu "wąskie gardło". Oznacza to, że dane muszą być rzutowane lub kompresowane w reprezentację o niższym wymiarze. Zadanie to realizowane jest poprzez minimalizację kwadratowego błędu rekonstrukcji E (ang. *squared reconstruction error*) pomiędzy wartościami danych w warstwie wejściowej, a wartościami danych w warstwie wyjściowej [94–96].

Działanie sieci auto-asocjacyjnej składa się z dwóch etapów. W pierwszym etapie sieć uczy się jak przedstawić dane wejściowe za pomocą mniejszej liczby sygnałów reprezentowanych przez neurony warstwy ukrytej: $\phi_{kompr}: X \rightarrow Z$. W drugim etapie realizuje proces dekompresji odbywający się pomiędzy warstwą ukrytą i wyjściową: $\phi_{dekompr}: Z \rightarrow \hat{X}$. Przykładowa struktura sieci NLPCA przedstawiona jest na rys. 3.8. Tak przygotowana sieć auto-asocjacyjna, składająca się pierwotnie z trzech warstw, wykonuje transformację danych w warstwie ukrytej do zredukowanego wymiaru, po czym wykonuje kolejną transformację odwrotną w celu uzyskania

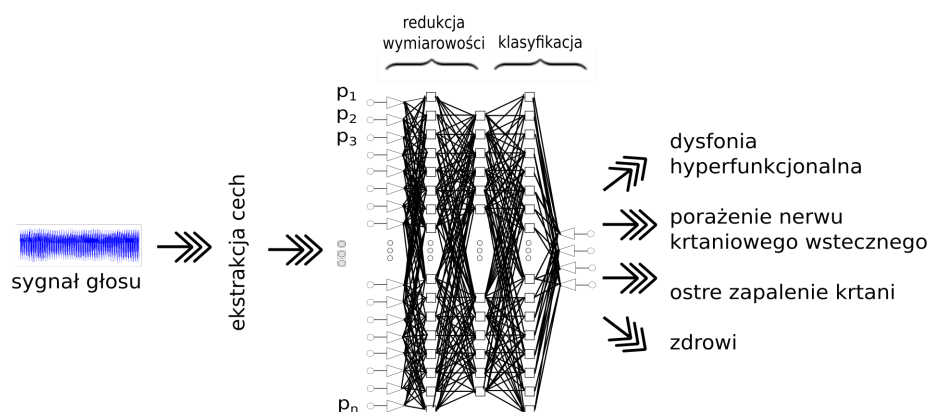


Rys. 3.8. Schemat blokowy auto-asocjacyjnej wielowarstwowej sieci neuronowej.

danych wyjściowych (dekompresja). W przypadku, w którym neurony w warstwie wejściowej i ukrytej mają charakterystyki liniowe, taka sieć odpowiada algorytmowi analizy składowych głównych. Sieć ta pozwala jednak na implementację nieliniowej rekonstrukcji danych poprzez zastosowanie neuronów nieliniowych, które posiadają w swojej strukturze nieliniowy potencjał membranowy (ang. *post – synaptic potential*, PSP). Aby takie zadanie było możliwe do realizacji sieć musi posiadać minimum trzy warstwy ukryte, a całkowita topologia powstałej sieci zawiera 5 warstw. Pierwsza warstwa ukryta realizuje nieliniową transformację danych, a środkowa warstwa ukryta zapewnia redukcję wymiarowości [94–96]. Ostatnia warstwa ukryta odpowiada za dekompresję wcześniej zgrupowanych danych.

W procesie dekompozycji danych na drodze odpowiadającej procedurze PCA w liniowy lub nieliniowy sposób, ważne jest ustalenie pożądanego efektu działań. W wyniku dekompozycji można otrzymać: redukcję wymiarowości danych, identyfikację, dyskryminację i ekstrakcję unikalnych oraz znaczących parametrów. W przypadku redukcji wymiaru parametrów, gdy pożądanym jest usunięcie korelacji pomiędzy parametrami, redukcja szumów i kompresja danych, wymagana jest podprzestrzeń o dużej pojemności opisowej. Jedynym wymaganiem, które wówczas należy spełnić jest, aby opis podprzestrzeni zawierał maksymalną informację w sensie błędu średniokwadratowego.

W pracy wykorzystano algorytm hierarchicznego nieliniowego PCA (*h-NLPCA*), który został zaproponowany przez Scholtz'a oraz Vigario w publikacji [94]. Algorytm ten porządkuje



Rys. 3.9. Schemat przeprowadzonego badania.

parametry względem ich hierarchii i zawartości informacyjnej. W wyniku otrzymuje się nieskorelowane ze sobą składowe główne, gdzie pierwsze n składowych głównych uszeregowane są zgodnie z malejącą wariancją parametrów akustycznych.

W celu znalezienia optymalnych wag podczas uczenia sieci neuronowej wykorzystano metodę nieliniowych gradientów sprzężonych (ang. *nonlinear conjugate gradient descent*) [95, 97].

Całkowita sieć neuronowa składa się z 2 części. Pierwsza odpowiada za redukcję wymiarowości przy pomocy sieci NLPCA, a w drugiej przeprowadzana jest klasyfikacja, której celem jest przydzielenie pacjenta na podstawie jego cech charakterystycznych do określonej jednostki chorobowej lub do klasy osób zdrowych (rys. 3.9). W przypadku klasyfikacji chorób laryngologicznych i grupy zdrowych przeprowadzonej na podstawie wielu wejściowych parametrów akustycznych, nie występują proste reguły klasyfikacyjne. Do rozwiązania takich problemów można wykorzystać ponownie sieci neuronowe [98]. W przypadku zadania klasyfikacji wykorzystano sieć składającą się z wielu wejść i 2 warstw ukrytych, z których druga stanowi warstwę wyjściową składającą się z 4 wyjść. Suma aktywacji wszystkich neuronów warstwy wyjściowej jest równa 1, stąd wynik sieci można interpretować jako prawdopodobieństwo przynależności danego parametru wyjściowego do poszczególnych, wykluczających się wzajemnie klas. Dla potrzeb klasyfikacji, funkcja ta powinna znajdować się w ostatniej warstwie sieci neuronowej.

Do realizacji procesu klasyfikacji wykorzystuje się funkcję błędu opartą na entropii wzajemnej (ang. *cross entropy*). W zależności od zadania klasyfikacji istnieją dwie wersje entropii wzajemnej. Funkcja aktywacji dla sieci z pojedynczym neuronem wyjściowym powinna być typu logistycznego. Ponieważ w pracy sieć wyjściowa składa się z 4 klas, dlatego należało wykorzystać funkcję aktywacji typu *softmax*. Próg akceptacji warstwy wyjściowej, decydujący o zaliczeniu do danej klasy przyjęto równy 0,73.

3.4. Informacja kontrolna i prognostyczna

Stworzenie narzędzi do monitorowania i przewidywania skutków terapii pozwoliłoby lekarzom na lepsze, indywidualne dostosowanie strategii postępowania diagnostyczno-leczniczego. Leczenie pacjentów z chorobą Parkinsona, u których pojawiają się wahania w stanie ruchowym (fluktuacje) i ruchy mimowolne (dyskinezy) za pomocą leczenia doustnego jest niezwykle trudne [99]. Następstwem choroby Parkinsona jest zwyrodnienie neuronów istoty czarnej, co powoduje ograniczenie produkcji dopaminy i powstanie głównie objawów ruchowych. Wybór leczenia odgrywa ważną rolę w przebiegu i rozwoju choroby [100]. Obecnie, do łagodzenia skutków choroby Parkinsona stosowanych jest wiele leków. Złotym standardem i najbardziej skutecznym jak do tej pory lekiem jest lewodopa, która jest prekursorem dopaminy. Lewodopa łagodzi objawy choroby Parkinsona takie jak: drżenie, sztywność i wzmożone napięcie mięśniowe, spowolnienie ruchowe (bradykinezja) i zastygnięcia podczas chodu (ang. *freezing*). Poprawia się również mowa. Lek nie zatrzymuje postępującej choroby, nie zapobiega obumieraniu komórek produkujących dopaminę, łagodzi tylko jej objawy. Lek ten jest najbardziej skuteczny w początkowym stadium rozwoju choroby, w kolejnych stadiach skuteczność maleje, a efekt działania leku jest wprost proporcjonalny do podanej dawki. Postępująca degeneracja neuronów istoty czarnej uniemożliwia uzupełnianie niedoboru dopaminy w mózgu w sposób ciągły. Oznacza to, że w zależności od godziny przyjęcia przez pacjenta i podanej dawki leku objaw działania będzie różny. Z powodu postępującej degeneracji neuronów dopaminergicznych zmienia się metabolizm przemiany lewodopy w dopaminę. Zmieniający się metabolizm powoduje nagromadzenie się dostarczonej i nie wchłoniętej dawki leku. Prowadzi to do powstania zaburzeń sprawności ruchowej zmieniających się w różnych porach doby. Pacjent może nie czuć skutków choroby, a po chwili stan jego się pogarsza, np. pojawia się sztywność, drżenie czy zaburzenia mowy. U niektórych pacjentów obserwuje się nadpobudliwość na działanie leku, powstają nieskoordynowane, mimowolne ruchy kończyn lub całego ciała (dyskinezy). Stosowane dawki nie są stałe i muszą się zmieniać z czasem w miarę postępu choroby i narastania niedoboru dopaminy, w zależności od aktualnych potrzeb chorego [101]. Z postępem choroby, czas działania leku lewodopy skraca się, czego skutkiem jest chwilowe nasilenie się objawów choroby Parkinsona (ang. *wearing off*). Dlatego pacjent powinien regularnie przyjmować leki i mieć ciągły kontakt z lekarzem prowadzącym.

Opracowanie metody kontroli i prognozowania skutków terapii prowadzonej z wykorzystaniem leku lewodopy może pomóc w zapobieganiu wahaniom stanu pacjenta. Pierwszym celem podjętych prac była ocena stanu neurologicznego pacjentów zgodnie ze skalą UPDRS część III na podstawie sygnałów akustycznych. W tym celu wykorzystano metodę regresji wektorów nośnych (SVR, *Support Vector Regression*) [102]. W kolejnym etapie prac, przeprowadzono predykcję stanu pacjenta w 120 i 180 minucie od przyjęcia leku. Badania prowadzono

na podstawie wcześniejszych nagrań głosu pacjentów wykorzystując sieci neuronowe. Wyniki wyrażono w stosowanej przez lekarzy skali UPDRS.

Jedną z najbardziej popularnych metod uczenia maszynowego jest maszyna wektorów nośnych (ang. *support vector machine*, SVM) [103]. Pozwala ona na analizowanie danych i rozpoznawanie wzorców w celu przeprowadzenia klasyfikacji lub regresji. Jednym z głównych etapów realizacji maszyny wektorów nośnych jest znalezienie hiperpłaszczyzny w przestrzeni wielowymiarowej, która optymalnie będzie oddzielać punkty należące do różnych klas z możliwie największym marginesem zaufania. Im większy margines tym odległość do najbliższego punktu danej klasy po każdej stronie hiperpłaszczyzny jest większa. Celem więc jest znalezienie maksymalnego marginesu i osiągnięcie najwyższej skuteczności rozdzielania klas.

Rozszerzenie algorytmu SVM nazywane jest regresją wektorów nośnych (SVR). SVR nie wymaga definiowania startowej funkcji odwzorowania, dzięki czemu możliwa jest analiza danych wielowymiarowych, w których występują relacje o charakterze nieliniowym, bez sprecyzowanej ścisłej postaci funkcyjnej [104]. W modelu ogólnym SVR zastosowano funkcję kosztu, która ignoruje wszystkie dane treningowe odpowiednio bliskie tj. odległe o mniej niż zadana wartość ε (ang. *insensitive loss function-SVR*, ε -SVR) od poprawnej odpowiedzi [102].

Do procesu uczenia niezbędny jest zbiór uczący, wraz z prawidłowymi odpowiedziami. Jeżeli oznaczymy x_{wej} jako wektor wejściowy, a y jako odpowiedź modelu (decyzję), to błąd może zostać określony jako $|y-f(x_{wej})|$. Wynik może znajdować się powyżej lub poniżej oczekiwanej wartości, a zmienna opisująca takie położenie może zostać zdefiniowana jako: $f(x_{wej_i})-y_i > \varepsilon$ i $y_i-f(x_{wej_i}) > \varepsilon$. ε określa próg czułości metody. Aby lepiej wyjaśnić algorytm SVR warto zacząć od omówienia regresji liniowej. Załóżmy, że mamy do czynienia z przypadkiem binarnym. Dane wejściowe opisane są jako $x_{wej_i} \in \mathbb{R}^d$, gdzie parametr d oznacza wymiar, a wyjście (stan, diagnoza pacjenta) oznaczone jest jako $y_i \in [-1, +1]$, $i=1,2,\dots,M$, M - liczba elementów danych wejściowych, np. liczba pacjentów. Zakłada się, że istnieje hiperpłaszczyzna, która jest w stanie oddzielić elementy $y_i = +1$ od elementów $y_i = -1$. Punkt x_{wej_i} znajdujący się w linii hiperpłaszczyzny spełnia wówczas warunek: $\langle w, x_{wej_i} \rangle + b = 0$, gdzie w jest wektorem normalnym do hiperpłaszczyzny, b jest wyrazem wolnym, $\frac{|b|}{\|w\|}$ określa odległość hiperpłaszczyzny od początku układu współrzędnych, a $\|w\|$ stanowi normę euklidesową wektora w (pierwiastek z iloczynu skalarnego tego wektora). Dla danych binarnych, które są całkowicie separowalne spełnione będą warunki [105]:

$$\langle w, x_{wej_i} \rangle + b \geq +1 \quad \text{dla} \quad y_i = +1 \quad (3.27)$$

$$\langle w, x_{wej_i} \rangle + b \leq -1 \quad \text{dla} \quad y_i = -1 \quad (3.28)$$

Równania (3.27) i (3.28) mogą zostać zastąpione jedną nierównością:

$$-y_i(\langle w, x_{wej_i} \rangle + b) + 1 \leq 0 \quad \text{dla } i = 1, 2, \dots, M \quad (3.29)$$

Dla danych $y_i = +1$, odległość hiperpłaszczyzny od początku układu współrzędnych będzie oznaczona jako $\frac{|1-b|}{\|w\|}$, a dla grupy $y_i = -1$ jako $\frac{|-1-b|}{\|w\|}$. Ideą SVR jest znalezienie hiperpłaszczyzn (w przypadku binarnym płaszczyzn H_1 oraz H_2), aby zapewnić maksymalny margines. Zagadnienie to może zostać przedstawione poprzez minimalizację funkcji [105]:

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^M |y_i - f(x_{wej_i})|_\varepsilon \quad (3.30)$$

gdzie C określa poziom dopuszczanego błędu. Wyższa wartość C oznacza wyższą "karę" dla punktów naruszających granice hiperpłaszczyzn. W rzeczywistości rzadko spotykane jest idealne odseparowanie punktów należących do poszczególnych klas, niektóre z punktów mogą odstawać od reszty i znajdować się w obszarze klasy przeciwnej. Biorąc pod uwagę taką sytuację wprowadza się zmienną relaksacyjną ξ_i (tzw. zwis, ang. *slack variable*). Wartość ξ_i powinna być jak najmniejsza. Jeżeli $0 \leq \xi_i \leq 1$, to punkt odpowiadający danym leży wewnątrz strefy marginesu po właściwej stronie, jeżeli $\xi_i > 1$, to punkt znajduje się po niewłaściwej stronie hiperpłaszczyzny.

Przyjmując oznaczenie: ξ gdy występuje $f(x_{wej_i}) - y_i > \varepsilon$ oraz ξ^* gdy $y_i - f(x_{wej_i}) > \varepsilon$, równanie (3.30) będzie miało postać [105]:

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^M (\xi_i + \xi_i^*) \quad (3.31)$$

z zastrzeżeniem:

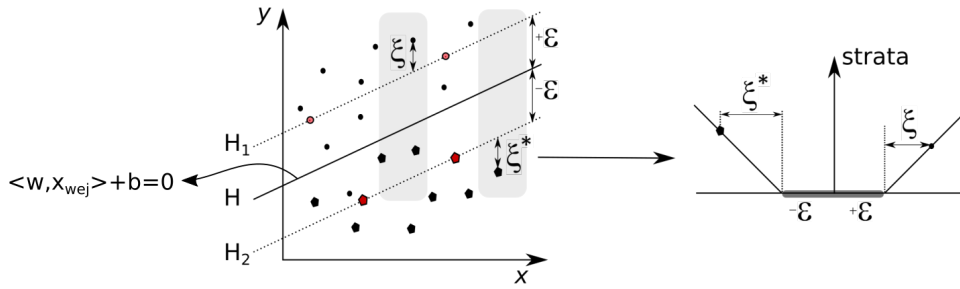
$$\begin{aligned} y_i - \langle w, x_{wej_i} \rangle - b &\leq \varepsilon + \xi_i, \\ -y_i + \langle w, x_{wej_i} \rangle + b &\leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* &\geq 0, i = 1, 2, \dots, M. \end{aligned} \quad (3.32)$$

Powyższe sformułowanie prowadzi do zdefiniowania ε -SVR funkcji straty, która określa błąd dla przewidywanych danych i jest określona jako:

$$|\xi|_\varepsilon =: \begin{cases} 0 & \text{gdy } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{w przeciwnym przypadku} \end{cases} \quad (3.33)$$

ε -SVR funkcja straty została przedstawiona na rysunku 3.10.

Punkty, które znajdują się w linii hiperpłaszczyzn (H_1 , H_2) nazywane są wektorami nośnymi. Na rysunku 3.10 zaznaczone są kolorem czerwonym. W celu spełnienia warunku maksymalizacji marginesu można wykorzystać lagrangian i mnożniki Lagrange'a [105]:



Rys. 3.10. Funkcja straty wykorzystywana w metodzie SVR.

$$\begin{aligned}
 L := & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M (\xi_i + \xi_i^*) - \sum_{i=1}^M (\eta_i \xi_i + \eta_i^* \xi_i^*) \\
 & - \sum_{i=1}^M \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_{weji} \rangle + b) \\
 & - \sum_{i=1}^M \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_{weji} \rangle - b)
 \end{aligned} \tag{3.34}$$

gdzie $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ są mnożnikami Lagrange'a i muszą spełniać nierówność:

$$\{\alpha_i, \alpha_i^*, \eta_i, \eta_i^*\} \geq 0. \tag{3.35}$$

Lagrangian L musi być zminimalizowany ze względu na parametry $\langle w, b \rangle$. Zgodnie z warunkiem koniecznym istnienia ekstremum pochodne cząstkowe po poszukiwanych parametrach w, b, ξ_i, ξ_i^* będą równe zero: $\frac{\partial L}{\partial w} = 0$, $\frac{\partial L}{\partial \xi_i} = 0$, $\frac{\partial L}{\partial \xi_i^*} = 0$ oraz $\frac{\partial L}{\partial b} = 0$, co doprowadza do układu równań [105]:

$$w = \sum_{i=1}^M (\alpha_i^* - \alpha_i) x_{weji} \tag{3.36}$$

$$C = \eta_i + \alpha_i \tag{3.37}$$

$$C = \eta_i^* + \alpha_i^* \tag{3.38}$$

$$\sum_{i=1}^M (\alpha_i - \alpha_i^*) = 0 \tag{3.39}$$

Podstawiając równania (3.36)-(3.39) do (3.34) należy rozwiązać już tylko dualny problem optymalizacyjny:

$$\begin{aligned} \max\{\alpha_i, \alpha_i^*\} = & -\frac{1}{2} \sum_{i,j=1}^M (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_{wej_i}, x_j \rangle \\ & -\varepsilon \sum_{i=1}^M (\alpha_i + \alpha_i^*) + \sum_{i=1}^M y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (3.40)$$

gdzie:

$$\sum_{i=1}^M \alpha_i - \alpha_i^* = 0 \quad \text{oraz} \quad 0 \leq \{\alpha_i, \alpha_i^*\} \leq C \quad (3.41)$$

Dzięki równaniu (3.41) wyeliminowane zostały zmienne η_i i η_i^* . Na podstawie równania $w = \sum_{i=1}^M (\alpha_i^* - \alpha) x_{wej_i}$, otrzymujemy funkcję regresji:

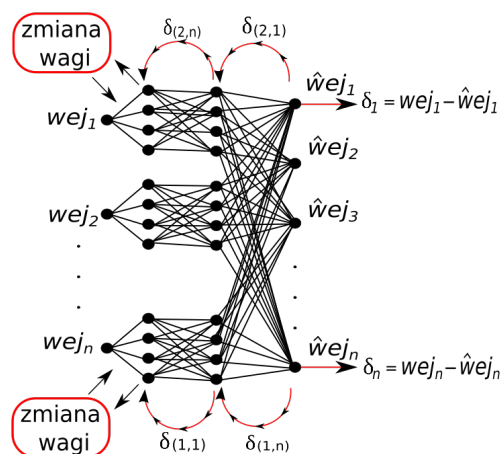
$$f(x) = \sum_{i=1}^M (\alpha_i - \alpha_i^*) \langle x_{wej_i}, x \rangle + b \quad (3.42)$$

Opis wektora w jest liniową kombinacją punktów testowych x_{wej_i} . Opis ten może zostać rozwinięty poprzez wykorzystanie jądra (ang. *kernel-based regression*). Wówczas równania (3.36) i (3.42) będą miały postać:

$$\begin{aligned} w &= \sum_{i=1}^M (\alpha_i - \alpha_i^*) \phi(x_{wej_i}) \\ f(x) &= \sum_{i=1}^M (\alpha_i - \alpha_i^*) k(x_{wej_i}, x) + b \end{aligned} \quad (3.43)$$

gdzie $\phi(x_{wej_i})$ to funkcja bazowa, $k(x_{wej_i}, x)$ to jądro iloczynu skalarnego funkcji bazowych. Jako funkcję bazową w niniejszej pracy wybrano funkcję gausowską. W rezultacie działania algorytmu SVR otrzymano estymację punktacji UPDRS.

W celu przeprowadzenia predykcji zmian stanu pacjenta z chorobą Parkinson'a wykorzystano sieć neuronową ze wsteczną propagacją błędów (ang. *backpropagation*) [108, 109]. Rysunek 3.11 przedstawia przykładowy schemat takiej sieci. W celu nauczania sieci wykorzystano metodę modyfikacji wag według optymalizacji Levenberg-Marquardt (L-M) [110]. Ze względu na niewielką złożoność obliczeniową algorytm ten jest jednym z najczęściej stosowanych do uczenia sieci neuronowych jednokierunkowych. Idea polega na rozwiązaniu nieliniowego problemu najmniejszych kwadratów [111]. Minimalizacja funkcji celu w metodzie L-M polega więc na dobraniu odpowiedniego czynnika regularyzacyjnego oraz wag sieci neuronowej w kolejnych iteracjach algorytmu.



Rys. 3.11. Idea działania sieci neuronowej z metodą wstecznej propagacji błędów.

Zaprezentowane podejście zostało wykorzystane w badaniach do pozyskania informacji diagnostycznej, kontrolnej i prognostycznej w wybranych chorobach laryngologicznych oraz neurodegeneracyjnych. Aby takie badania mogły zostać zrealizowane konieczne było wykorzystanie parametrów opisujących sygnał akustyczny, których opis znajduje się w podsekcji 3.2. Następnie przedstawiono metodologię wykorzystaną w klasyfikacji osób zdrowych i tych, którzy mieli zdiagnozowaną chorobę laryngologiczną. W tym celu wykorzystano algorytm lasów losowych. Klasyfikacja wskazująca na konkretną chorobę laryngologiczną została przeprowadzona w oparciu o sieci neuronowe. Do kontroli stanu postępu nasilenia choroby neurodegeneracyjnej jaką jest choroba Parkinson'a wykorzystano skalę UPDRS-III, wyznaczoną przez lekarza-specjalistę oraz nagrania mowy pozyskane w określonych chwilach czasu. Transformację wektora parametrycznego sygnału mowy do skali UPDRS-III przeprowadzono z wykorzystaniem algorytmu regresji wektorów nośnych. Informację prognostyczną, a więc prognozę wyniku wartości skali UPDRS-III w określonej chwili w przyszłości na podstawie wcześniejszych pomiarów wykonano w oparciu o sieci neuronowe. Sieci te pozwoliły na predykcję wartości wektora parametrów akustycznych, a algorytm regresji wektorów nośnych umożliwił na wyrażenie go za pomocą skali UPDRS-III. W kolejnym rozdziale opisano badania prowadzone na sygnałach mowy (z bazy SVD oraz własnych nagrań) przy wykorzystaniu algorytmów opisanych w tym rozdziale.

4. Wyniki badań

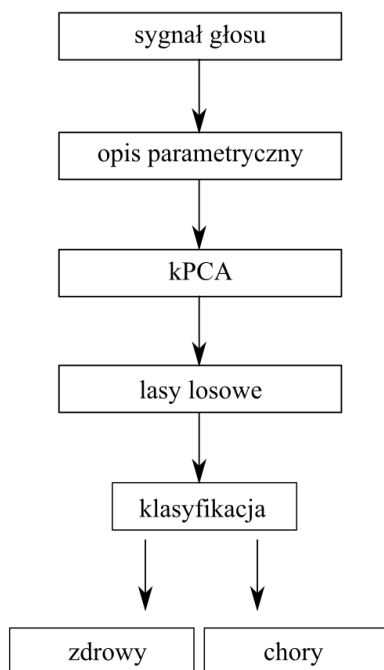
Rozdział ten przedstawia wyniki przeprowadzonych badań eksperymentalnych. Jest on podzielony na trzy główne sekcje. Pierwsza dotyczy algorytmicznej detekcji chorób laryngologicznych. Badania przeprowadzono na sygnałach pochodzących z bazy języka niemieckiego *Saarbruecken Voice Database* [14], opisanej w rozdziale 3, sekcji 3.1. Zadanie to obejmuje klasyfikację binarną, której celem jest wskazanie, czy badany pacjent jest zdrowy, czy też posiada symptomy świadczące o obecności choroby laryngologicznej. Dodatkowo, przeprowadzono tzw. multiklasyfikację, w wyniku której system wskazywał na konkretną, jedną chorobę laryngologiczną, na którą cierpiał dany pacjent.

Sekcja druga 4.2 przedstawia wyniki estymacji punktacji w skali UPDRS-III pacjentów z chorobą Parkinson'a na podstawie sygnału głosu. Proces estymacji został wykonany w oparciu o algorytm ϵ -SVR. Dane do badań zostały zarejestrowane przez autorkę pracy w Krakowskim Szpitalu Specjalistycznym im. Jana Pawła II w Krakowie. Szczegółowy opis bazy nagrań znajduje się w rozdziale 3.1. Sekcja trzecia 4.2 przedstawia rozwiązanie problemu predykcji parametrów akustycznych w trzeciej godzinie od spożycia leków Ldopy oraz estymację na tej podstawie punktacji pacjentów w skali UPDRS-III. W ten sposób rozwiązywany jest problem predykcji stanu pacjenta, wyrażonego w skali UPDRS-III.

4.1. Algorytmiczna detekcja chorób laryngologicznych

Ten etap badań miał na celu stworzenie systemu, za pomocą którego możliwa będzie klasyfikacja osób zdrowych i tych, u których występują różne schorzenia otolaryngologiczne. W pierwszym etapie skupiono się na klasyfikacji binarnej (wykrycie osób zdrowych i chorych), a następnie na możliwościach detekcji konkretnych chorób laryngologicznych. W tym celu wykorzystano bazę *Saarbruecken Voice Database* (SVD) [14], która zawierała nagrania samogłosek /a/, /i/ oraz /u/ wypowiedzianych w trzech intonacjach: wysokiej, niskiej oraz normalnej. Szczegółowy opis bazy SVD znajduje się w rozdziale 3.1.

Do budowy systemu automatycznej diagnostyki chorób laryngologicznych wykorzystano 900 nagrań głosu kobiet oraz 510 nagrań mężczyzn, z czego 50% stanowiła grupa osób zdrowych, a pozostała grupa zawierała osoby z różnymi chorobami głosu. Utworzony został opis



Rys. 4.1. Schemat klasyfikacji chorych laryngologicznie i osób zdrowych.

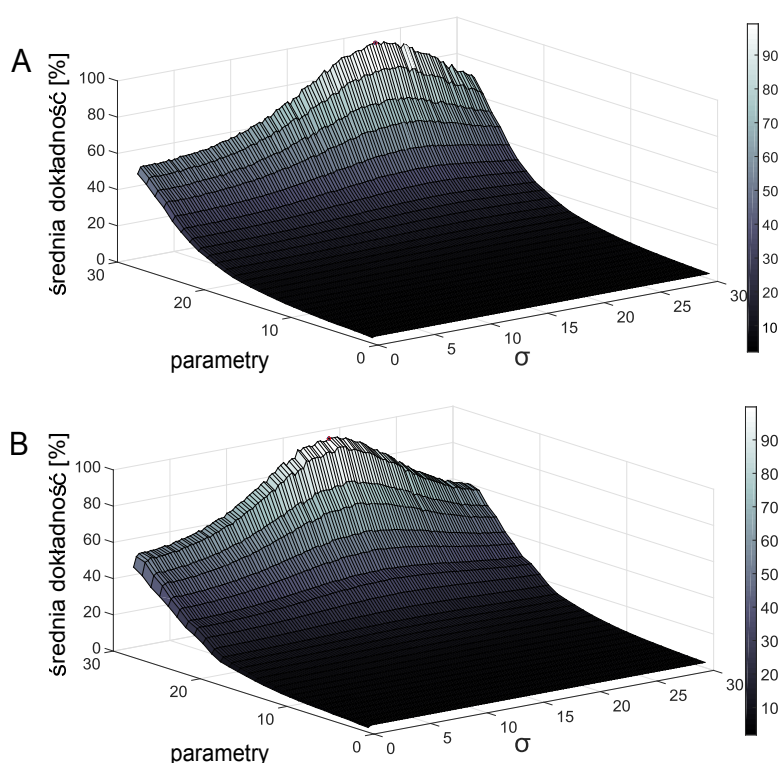
parametryczny nagrań fonetycznych głosu, na podstawie którego możliwa była detekcja zniekształceń sygnału głosu. Analiza została przeprowadzona na podstawie 28 parametrów: częstości podstawowej, współczynników jitter i shimmer, energii, zerowego, pierwszego, drugiego i trzeciego momentu widmowego, kurtozy, współczynnika mocy względnej, amplitudy i częstości 1-, 2-, 3- formantu, wartości maksymalnej i minimalnej sygnału, 10 współczynników mel-cepstralnych. Każdy sygnał akustyczny został więc przedstawiony za pomocą 28-wymiarowego wektora parametrów. Ponieważ parametry różnią się zakresem zmienności oraz wartością i wymiarem fizycznym, przeprowadzono normalizację względem odchylenia od wartości minimalnej określoną wzorem 4.1. Zapewniło to normalizację wartości parametrów do przedziału [0,1]. Schemat badania przedstawiono na rys. 4.1.

$$p_{norm} = \frac{p - p_{min}}{p_{max} - p_{min}} \quad (4.1)$$

gdzie: p - wynik danego parametru,

p_{max} , p_{min} - wartość maksymalna oraz minimalna wektora parametrów.

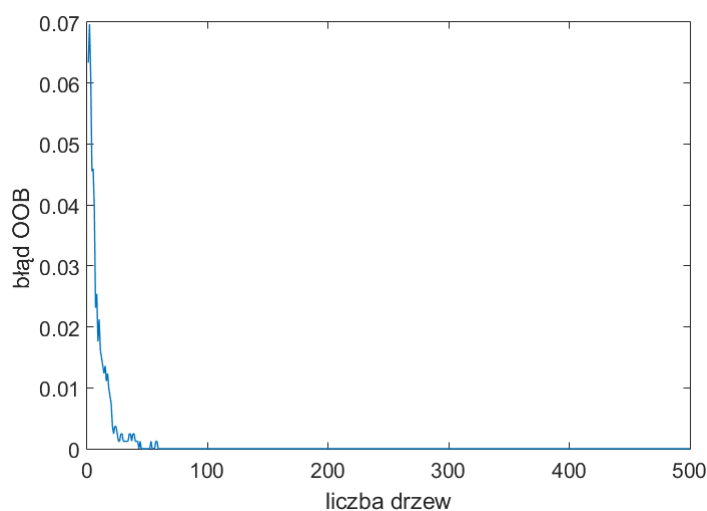
Z punktu widzenia klasyfikacji, ważne było wykorzystanie takich zmiennych, które będą najbardziej informacyjne oraz pozwolą na oddzielenie różnych klas od siebie w przestrzeni parametrów. W tym celu wykorzystano metodę analizy składowych głównych (kPCA). Szczegółowy opis matematyczny tej metody znajduje się w rozdziale 3.3.1. Algorytm implementowany



Rys. 4.2. Wykres przedstawiający wpływ wartości σ i liczby składowych głównych na dokładność klasyfikacji [%] dla samogłoski /a/ wypowiedzianej przez A) kobiety oraz B) mężczyzn w intonacji normalnej [113].

w pracy doktorskiej został sformułowany przez autorów pracy [112]. Do obliczeń wykorzystano jądro Gaussa. W wyniku działania algorytmu kPCA otrzymano tzw. składowe główne (ang. *principal components*, PC). Ważnym parametrem podczas tworzenia algorytmu kPCA jest parametr σ (opisany wzorem (3.26)). W pracy parametr ten został wybrany eksperymentalnie na podstawie analizy każdej samogłoski oraz każdej intonacji u kobiet i mężczyzn osobno. Jako kryterium wyboru wykorzystano średnią dokładność klasyfikacji w zależności od liczby użytych składowych głównych. Dokładność klasyfikacji jest to stosunek przypadków poprawnie zaklasyfikowanych do wszystkich biorących udział w procesie klasyfikacji. Zwiększenie wymiaru wektora składowych głównych nie przyniosło poprawy wyniku klasyfikacji dla żadnej z samogłosek. Wpływ σ na dokładność klasyfikacji dla różnej liczby składowych głównych, dla klasyfikacji prowadzonej na podstawie samogłoski /a/ przedstawiono na rysunku 4.2. Najwyższą dokładność uzyskano dla samogłoski /a/ zarówno u kobiet oraz mężczyzn dla 28 składowych głównych. U kobiet optymalna wartość parametru σ wyniosła 19,5, u mężczyzn $\sigma=17,5$.

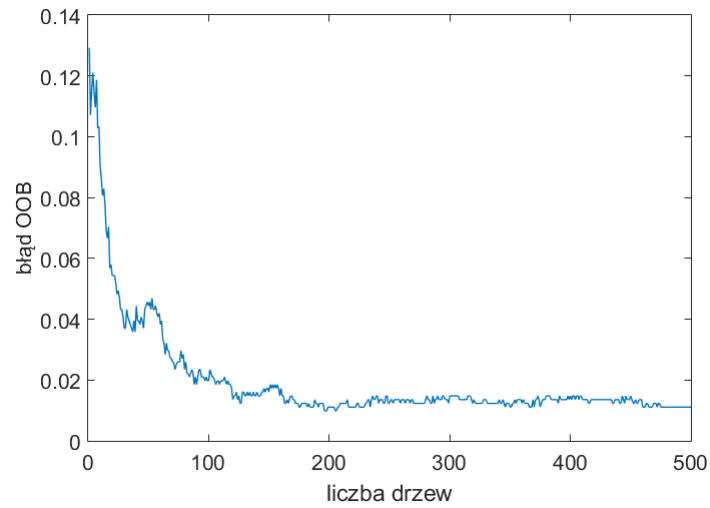
Klasyfikację przeprowadzono metodą lasów losowych (RF, ang. *random forest*), której metodologia opisana jest w rozdziale 3.3.2. Algorytm RF powstał na podstawie pracy [91].



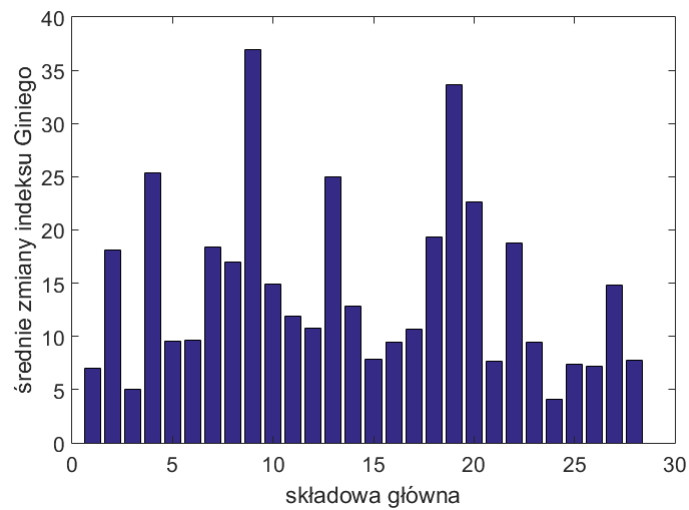
Rys. 4.3. Błąd predykcji (OOB) w zależności od liczby drzew, samogłoska /a/, normalna intonacja, mężczyźni.

Proces klasyfikacji przeprowadzono osobno dla grupy kobiet oraz mężczyzn. Podczas implementacji pierwszym krokiem było wyznaczenie wielkości zbioru treningowego, który stanowił 90% nagrań oraz zbioru testowego (pozostałe 10% nagrań). Dane wejściowe do metody lasów losowych stanowiły wektory składowych głównych otrzymane w wyniku działania algorytmu kPCA. Aby uzyskać obiektywną ocenę funkcjonowania algorytmu, proces ten został poddany walidacji krzyżowej. Na podstawie zbioru testowego obliczono błąd predykcji (OOB) w zależności od liczby drzew. Przykładowe wykresy takich zależności znajdują się na rys. 4.3 oraz 4.4. Poddając analizie wszystkie iteracje algorytmu, uznano liczbę drzew $t=500$ za wystarczającą do uzyskania satysfakcjonujących wyników, co widoczne jest również na przedstawionych wykresach 4.3 i 4.4. Na podstawie wykresu 4.3 wystarczająca liczba drzew wyniosła 59, natomiast w przypadku wykresu 4.4 wystarczająca liczba drzew to 475.

W celu określenia wartości predykcyjnych dla poszczególnych składowych głównych wyznaczono średnie zmiany indeksu Giniego (tab. 4.1). Analizując wartości indeksu Giniego można dostrzec, że średnie wartości oraz ich odchylenia standardowe dla każdej z samogłosek zarówno dla grupy kobiet jak i mężczyzn są do siebie zbliżone. Średnia wartość wynosi od 12,9 do 13,1. Przykładowy wykres prezentujący indeks Giniego wyznaczony dla poszczególnych składowych głównych u mężczyzn, dla samogłoski /a/ w intonacji normalnej przedstawiono na rys. 4.5. Na jego podstawie zauważyć można, że w tym przypadku parametr 8 - energia sygnału posiada najwyższą średnią wartość indeksu Giniego. Kolejnymi parametrami o wysokich wartościach tego indeksu są: parametr 2- współczynnik jitter, 4- jedenastopunktowy współczynnik zaburzeń okresu, 13- kurtoza, 18- częstotliwość 1 formantu, 19- częstotliwość 2 formantu, 20- częstotliwość 3 formantu, 22- 2 współczynnik mel-cepstralny oraz 27- 7 współczynnik mel-cepstralny. Pozostałe parametry mają bardzo zbliżone wartości tego indeksu.



Rys. 4.4. Błąd klasyfikacji w zależności od liczby drzew, samogłoska /a/, wysoka intonacja, kobiety.



Rys. 4.5. Średnie wartości indeksu Ginięgo dla poszczególnych składowych głównych wykorzystanych w klasyfikacji, samogłoska /a/, normalna intonacja, mężczyźni.

Tabela 4.1. Średnie wartości oraz odchylenie standardowe indeksu Giniego dla wszystkich składowych głównych dla każdej z analizowanych samogłosek oraz ich intonacji.

<i>intonacja</i>	<i>kobiety</i>		
	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>
wysoka	13,09±6,26	13,03±5,43	12,94±5,32
niska	12,93±5,41	13,03±5,48	13,03±5,90
normalna	12,88±5,90	12,91±5,76	12,91±5,81
	<i>mężczyźni</i>		
wysoka	12,98±5,65	12,98±5,65	12,92±5,72
niska	12,96±5,65	12,97±5,58	12,98±5,72
normalna	13,05±5,74	12,89±6,18	12,94±5,40

Tabela 4.2. Dokładność klasyfikacji [%] dla osób zdrowych i chorych otrzymana w wyniku algorytmu RF u kobiet i mężczyzn z uwzględnieniem różnych intonacji po zastosowaniu algorytmu kPCA [113].

<i>intonacja</i>	<i>kobiety</i>			<i>mężczyźni</i>		
	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>
wysoka	99,56	99,67	99,56	99,80	99,80	99,61
niska	99,78	99,67	99,67	99,80	99,61	99,80
normalna	99,9	99,78	99,67	100,00	99,61	99,61

W celu uzyskania oceny skuteczności klasyfikacji przeprowadzono sprawdzian krzyżowy oraz obliczono parametry takie jak: dokładność (tab. 4.2), precyzja dla osób zdrowych (tab. 4.3) oraz precyzja dla osób chorych (tab. 4.4). Dokładność zdefiniowana jest jako stosunek liczby przypadków dobrze przypisanych do grupy osób zdrowych i chorych oraz liczby wszystkich sklasyfikowanych przypadków. Precyzja określa stosunek liczby przypadków zaklasyfikowanych do grupy odpowiednio zdrowych/chorych, które w rzeczywistości należały do grupy zdrowych/chorych oraz liczby wszystkich przypadków sklasyfikowanych przez system.

Tabela 4.3. Precyzja dla osób zdrowych [%] otrzymana w wyniku algorytmu RF u kobiet i mężczyzn z uwzględnieniem różnych intonacji po zastosowaniu algorytmu kPCA.

<i>intonacja</i>	<i>kobiety</i>			<i>mężczyźni</i>		
	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>
wysoka	99,56	99,78	99,78	100,00	100,00	99,60
niska	100,00	99,78	99,78	100,00	100,00	100,00
normalna	100,00	100,00	99,78	100,00	99,60	99,62

Tabela 4.4. Precyzja dla osób chorych [%] otrzymana w wyniku algorytmu RF u kobiet i mężczyzn z uwzględnieniem różnych intonacji po zastosowaniu algorytmu kPCA.

<i>intonacja</i>	<i>kobiety</i>			<i>mężczyźni</i>		
	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>	<i>/a/</i>	<i>/i/</i>	<i>/u/</i>
wysoka	99,56	99,56	99,33	99,60	99,62	99,62
niska	99,56	99,56	99,56	99,62	99,22	99,60
normalna	99,78	99,56	99,56	100,00	99,60	99,62

Na podstawie tab. 4.2 widoczne jest, że dokładność klasyfikacji wynosi minimum 99,5%. Wyniki uzyskane dla poszczególnych samogłosek i intonacji są do siebie zbliżone. Wartość bezbłędną 100% poprawnych wyników klasyfikacji otrzymano w przypadku mężczyzn dla samogłoski /a/, dla intonacji normalnej. W grupie kobiet, najwyższą dokładność klasyfikacji otrzymano również dla samogłoski /a/ w intonacji normalnej. Najniższą dokładność otrzymano dla samogłoski /a/ w intonacji wysokiej w przypadku kobiet. Biorąc pod uwagę grupę mężczyzn najniższy wynik uzyskano dla samogłoski /i/ w intonacji niskiej oraz normalnej oraz dla samogłoski /u/ w intonacji wysokiej i normalnej. Spośród wszystkich samogłosek największy stosunek przypadków poprawnie zaklasyfikowanych do wszystkich przypadków poddanych analizie otrzymano dla samogłoski /a/ zarówno w grupie kobiet jak i mężczyzn. Podsumowując, badanie dokładności klasyfikacji, było wyższe w grupie mężczyzn niż w grupie kobiet. Niemniej, średnia różnica pomiędzy wynikami obu grup wynosi 0,03%.

Parametr precyzji wyznaczono osobno w grupie osób zdrowych oraz grupy osób chorych. Wyniki zamieszczono w tabelach 4.3 oraz 4.4. Na ich podstawie można zauważyć, że w przypadku błędnej klasyfikacji danego przypadku, algorytm częściej zalicza osoby chore do zdrowych, niż na odwrót. Dla samogłoski /a/ w intonacji niskiej i normalnej oraz samogłoski /i/ w intonacji normalnej precyzja u osób zdrowych wyniosła 100% w grupie kobiet. W przypadku mężczyzn, samogłoska /a/ wypowiedziana we wszystkich intonacjach, samogłoska /i/ w intonacji wysokiej oraz niskiej, a także samogłoska /u/ w intonacji niskiej zapewniły 100% precyzję dla osób zdrowych. Najniższe wyniki otrzymano dla samogłoski /a/ w intonacji wysokiej u kobiet oraz dla samogłoski /i/ w intonacji normalnej u mężczyzn. Najwyższa precyzja dla osób zdrowych zapewniła samogłoska /a/ i /i/ podczas klasyfikacji kobiet oraz samogłoska /a/ gdy klasyfikację przeprowadzono w grupie mężczyzn.

W grupie osób chorych, najwyższą precyzję otrzymano dla samogłoski /a/ wypowiedzianej w intonacji normalnej w grupie kobiet (99,78%) oraz mężczyzn (100%). Najniższy wynik wystąpił w przypadku, gdy do analizy wykorzystano samogłoskę /u/ w intonacji wysokiej w grupie kobiet oraz samogłoskę /i/ w intonacji niskiej w grupie mężczyzn. Spośród wszystkich samogłosek najwyższa precyzja dla osób chorych otrzymano dla samogłoski /a/ zarówno u kobiet jak i mężczyzn.

Kolejnym etapem prac była diagnostyka wybranych chorób laryngologicznych i wskazanie jednej, która wystąpiła u danego pacjenta. Do badań wybrano 3 najbardziej liczne jednostki chorobowe w bazie SVD: dysfonię hyperfunkcjonalną, porażenie nerwu krtaniowego wstecznego i ostre zapalenie krtani oraz grupę osób zdrowych. Liczby osób chorych na wymienione wyżej jednostki chorobowe znajduje się w tab. 3.1 w rozdziale 3.1. Liczba parametrów akustycznych oraz proces normalizacji danych były analogiczne jak w przypadku zadania binarnej klasyfikacji.

W przypadku klasyfikacji chorób laryngologicznych i grupy zdrowych na podstawie wielu wejściowych parametrów akustycznych, które decydują o przynależności do klasy nie występują proste reguły klasyfikacyjne [6]. Do rozwiązania postawionego zadania wykorzystano sieć neuronową. Całkowita sieć neuronowa składała się z 2 części. Pierwsza dotyczyła redukcji wymiarowości przy pomocy sieci NLPCA (opisanej w rozdziale 3.3.3), a w drugiej przeprowadzona została klasyfikacja, której celem było zdiagnozowanie pacjenta na podstawie jego cech charakterystycznych.

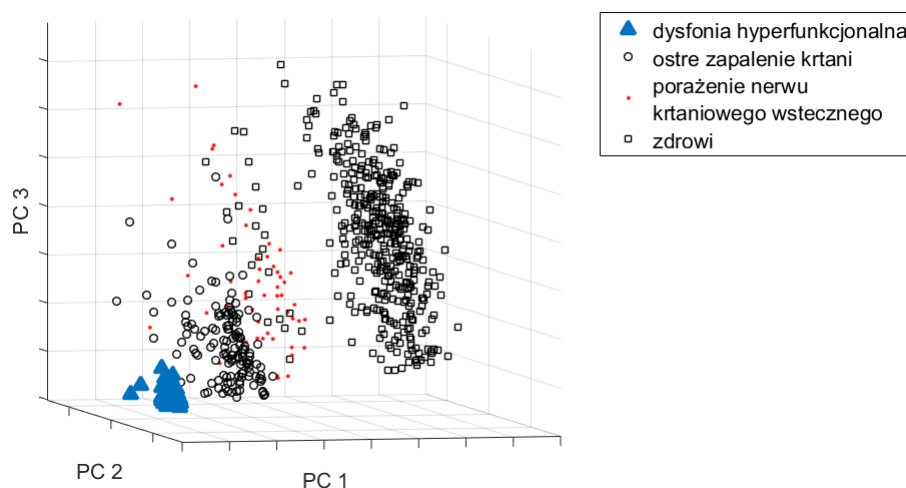
Sieć NLPCA składała się z 5 warstw, warstwy wejściowej, 3 warstw ukrytych i warstwy wyjściowej. Tab. 4.5 przedstawia liczbę neuronów we wszystkich warstwach sieci dla każdej z samogłosek. Do warstwy wejściowej wprowadzono wektory parametrów akustycznych opisujących nagrania określonych dysfunkcji głosu. Pierwsza warstwa ukryta umieszczona pomiędzy warstwą wejściową, a warstwą ukrytą środkową wykonuje wymaganą nieliniową kompresję danych wejściowych. Zadaniem drugiej warstwy ukrytej jest zmniejszenie wymiaru sygnału wejściowego. Strukturą umieszczoną między warstwą ukrytą, a warstwą wyjściową, wykonuje odwrotną transformację (dekompresję) sygnału o wcześniej zmniejszonej wymiarowości.

Podczas uczenia sieci neuronowej, wagi sieci są korygowane w celu zminimalizowania średniego błędu kwadratowego uzyskanego dla każdego wektora parametrów za pomocą algorytmu wstecznej propagacji błędów. Następnie ustalana jest liczba epok. Etap ten zatrzymuje się gdy zostanie osiągnięta minimalna wartość średniego błędu kwadratowego. Proces uczenia zapewnia najlepszą adaptację i modyfikację wag w odpowiedzi na wzorce wejściowe podczas treningu (warstwa wejściowa). W wyniku działania sieci NLPCA otrzymano redukcję wymiarowości wektorów parametrów (tab. 4.5). Liczba parametrów wykorzystanych następnie w procesie klasyfikacji przedstawiona jest w opisie tabeli 4.5 w 2 warstwie ukrytej.

Wykres 4.6 przedstawia rozkład 3 wybranych parametrów powstałych w wyniku działania NLPCA dla samogłoski /a/ u kobiet. Widać na nim przestrzenne rozdzielenie poszczególnych grup pacjentów. Wyniki dla poszczególnych grup skupione są w określonych obszarach, będąc niemalże już rozłączne w przestrzeni trójwymiarowej.

Tabela 4.5. Liczba neuronów w poszczególnych warstwach sieci NLPCA dla każdej z samogłosek, wys- intonacja wysoka, norm - intonacja normalna, nis- intonacja niska, L – warstwa liniowa, N – warstwa o charakterze nieliniowym .

samogłoska [intonacja]	warstwa wejściowa	1 warstwa ukryta	2 warstwa ukryta	3 warstwa ukryta	warstwa wyjściowa
<i>kobiety</i>					
/a/ [wys]	35L	42N	14N	42N	35L
/a/ [norm]	35L	45N	13N	45N	35L
/a/ [nis]	35L	47N	16N	47N	35L
/i/ [wys]	35L	38N	14N	38N	35L
/i/ [norm]	35L	39N	17N	39N	35L
/i/ [nis]	35L	41N	17N	41N	35L
/u/ [wys]	35L	40N	14N	40N	35L
/u/ [norm]	35L	33N	16N	33N	35L
/u/ [nis]	35L	36N	15N	36N	35L
<i>mężczyźni</i>					
/a/ [wys]	35L	36N	12N	36N	35L
/a/ [norm]	35L	45N	13N	45N	35L
/a/ [nis]	35L	37N	15N	37N	35L
/i/ [wys]	35L	38N	15N	38N	35L
/i/ [norm]	35L	38N	16N	38N	35L
/i/ [nis]	35L	39N	17N	39N	35L
/u/ [wys]	35L	41N	17N	41N	35L
/u/ [norm]	35L	42N	19N	42N	35L
/u/ [nis]	35L	41N	19N	41N	35L



Rys. 4.6. Prezentacja wybranych 3 parametrów (PC1, PC2, PC3) powstałych w wyniku działania neuronowej NLPCA dla samogłoski /a/ w intonacji normalnej u kobiet dla 3 chorób głosu oraz grupy kobiet zdrowych [114]).

Klasyfikacja jest realizowana w ostatniej warstwie sieci neuronowej, która ma 4 wyjścia. W neuronach warstwy wyjściowej zastosowano funkcję aktywacji typu *softmax*, która jest znormalizowaną funkcją wykładniczą transformującą wektor wartości rzeczywistych do wektora, którego elementy zawierają się w przedziale wartości [0,1].

Aby ocenić skuteczność klasyfikacji przeprowadzono 8-krotną walidację krzyżową. Zbiór danych został losowo podzielony na 8 różnych podzbiorów, a proces klasyfikacji powtórzono 8 razy. W każdym przypadku do testowania skuteczności używano innego podzbioru. Po zakończeniu sprawdzania krzyżowego wyniki zostały uśredniane we wszystkich powtórzeniach. Do ilościowej oceny skuteczności klasyfikacji wykorzystano macierz pomyłek. Przedstawia ona liczby przypadków poprawnie rozpoznanych przez system oraz tych, które zostały zaklasyfikowane błędnie. Wszystkie macierze pomyłek znajdują się w załączniku A. Przy pomocy macierzy pomyłek wyznaczono 3 miary ewaluacji: dokładność, czułość oraz parametr F1. W przypadku multiklasyfikacji dokładność definiowana jest jako stosunek liczby przypadków poprawnie zaklasyfikowanych do wszystkich przypadków zaklasyfikowanych w obrębie danej grupy. Czułość (ang. *recall*) określa jaka część danych z określonej klasy została zaklasyfikowana poprawnie. Innymi słowy, jest to zdolność algorytmu do prawidłowego rozpoznania choroby. Parametr F1 (ang. *F-score*) stanowi średnią harmoniczną z parametrów precyzji oraz czułości. Precyzja określa liczbę przypadków poprawnie zaklasyfikowanych do danej klasy w stosunku do wszystkich przypadków do niej zaklasyfikowanych. Parametr F1 jest opisany wzorem:

$$F1 = 2 \cdot \frac{p \cdot c}{p + c} \quad (4.2)$$

gdzie:

p - precyzja, c - czułość.

Wyniki oceny dokładności klasyfikacji dla poszczególnych samogłosek wypowiedzianej w różnej intonacji przedstawiono w tab. 4.6 oraz 4.7. Dokładność w grupie kobiet mieści się w zakresie od 79-84% w zależności od wybranej samogłoski. Analizując każdą z samogłosek i intonacji osobno dla grupy kobiet, najlepszy wynik klasyfikacji uzyskano dla samogłoski /u/ o intonacji niskiej (84,13%), a najgorszy dla samogłoski /a/ wypowiedzianej również w intonacji niskiej (79,25%). Jednakże, dokładność klasyfikacji dla samogłosek /a/ - intonacja wysoka i niska, /i/ - intonacja normalna, niska oraz samogłoska /u/ we wszystkich intonacjach jest wyższa od 81%. Zaproponowany klasyfikator w grupie kobiet najlepiej poradził sobie z wykryciem dysfonii hyperfunkcjonalnej (samogłoska /i/, intonacja niska - 85,00%), a najgorzej z porażeniem nerwu krtaniowego wstecznego (samogłoska /i/, intonacja niska - 85,82%). W wyniku klasyfikacji najwyższą dokładność otrzymano dla samogłoski /u/ bez względu na intonację w grupie kobiet.

Detekcja grupy osób zdrowych przebiegła bezbłędnie (dokładność wyniosła 100%) zarówno u kobiet jak i u mężczyzn.

Tabela 4.6. Dokładność klasyfikacji [%] na podstawie sygnału głosu kobiet, u których zdiagnozowano jedną z 3 chorób laryngologicznych oraz kobiet zdrowych, [wys] - intonacja wysoka, [norm] - intonacja normalna, [nis] - intonacja niska.

samogłoska	HF	lar	par	zdrowi
/a/ [wys]	79,38	82,14	65,44	100,00
/a/ [norm]	76,25	73,21	78,68	100,00
/a/ [nis]	76,88	73,21	66,91	100,00
/i/ [wys]	74,38	75,00	68,38	100,00
/i/ [norm]	82,50	75,00	66,91	100,00
/i/ [nis]	85,00	80,36	58,82	100,00
/u/ [wys]	78,75	82,14	67,65	100,00
/u/ [norm]	81,88	75,00	75,74	100,00
/u/ [nis]	81,25	83,93	71,32	100,00

Odmienne niż w grupie kobiet, dokładność klasyfikacji porażenia nerwu krtaniowego wstecznego u mężczyzn była najwyższa spośród wszystkich chorób laryngologicznych i wynosiła 84,72% dla samogłoski /a/ w intonacji wysokiej. Natomiast najniższą dokładność osiągnięto dla ostrego zapalenia krtani (samogłoska /i/ intonacja niska). Biorąc pod uwagę wszystkie choroby oraz grupę osób zdrowych dokładność klasyfikacji mieści się w zakresie od 72 do 80%. Najwyższą dokładność klasyfikacji osiągnięto dla samogłoski /a/ o intonacji normalnej (80,35%), a najniższą dla /a/ o intonacji niskiej (72,36%). Samogłoska /a/ zapewniła najwyższą dokładność klasyfikacji w grupie mężczyzn.

Porównując dokładność, wyższe wyniki osiągnięto w grupie kobiet niż w grupie mężczyzn. W najlepszym przypadku jest to różnica 3,78%. Taka różnica może być spowodowana liczebnością obu grup. Nagrań głosu kobiet było 1,8 raza więcej, niż nagrań mężczyzn, co mogło wpłynąć na jakość uczenia sieci. Algorytm bardzo dobrze poradził sobie z rozpoznaniem ostrego zapalenia krtani u mężczyzn (84,72%), natomiast u kobiet ta jednostka chorobowa była najtrudniejsza do wykrycia (78,68%) spośród wszystkich poddanych analizie.

Na podstawie macierzy pomyłek powstałej w efekcie sprawdzianu krzyżowego obliczono czułość. Wyniki znajdują się w tabelach 4.8 i 4.9. Tak jak w przypadku dokładności czułość u kobiet i mężczyzn jest maksymalna (1,00) podczas detekcji osób zdrowych. Samogłoska /u/ w intonacji niskiej u kobiet wykazuje najwyższy wynik (0,83). Najlepiej wykrywaną chorobą jest paraliż nerwu krtaniowego wstecznego (0,88 dla samogłoski /u/ wysoka intonacja), a najgorzej ostre zapalenie krtani dla samogłoski /a/ w niskiej intonacji. Niski wynik ostrego zapalenia krtani może być spowodowany małą próbą uczącą i testującą. U mężczyzn, najwyższy wynik czułości otrzymano dla ostrego zapalenia krtani (0,86 - samogłoska /a/, intonacja normalna), a najniższy dla dysfonii hyperfunkcjonalnej (0,50 - samogłoska /u/, intonacja normalna). Podobnie jak w przypadku kobiet, grupa osób chorujących na dysfonię hyperfunkcjonalną była

Tabela 4.7. Dokładność klasyfikacji [%] na podstawie sygnału głosu mężczyzn, u których zdiagnozowano jedną z 3 chorób laryngologicznych oraz osób zdrowych, [wys] - intonacja wysoka, [norm] - intonacja normalna, [nis] - intonacja niska.

samogłoska	HF	lar	par	zdrowi
/a/ [wys]	62,50	55,00	84,72	100,00
/a/ [norm]	70,00	75,00	76,39	100,00
/a/ [nis]	55,00	52,50	81,94	100,00
/i/ [wys]	75,00	51,25	80,56	100,00
/i/ [norm]	77,50	50,00	80,56	100,00
/i/ [nis]	67,50	47,50	79,17	100,00
/u/ [wys]	65,00	51,25	75,00	100,00
/u/ [norm]	62,50	50,00	80,56	100,00
/u/ [nis]	67,50	50,00	81,94	100,00

Tabela 4.8. Wartości średnie wraz z odchyleniem standardowym parametru czułość powstałe w wyniku klasyfikacji na podstawie sygnału głosu kobiet, [wys] - intonacja wysoka, [norm] - intonacja normalna, [nis] - intonacja niska.

samogłoska	HF	lar	par	zdrowi
/a/ [wys]	0,78±0,11	0,57±0,11	0,87±0,11	1,00±0,00
/a/ [norm]	0,78±0,08	0,56±0,08	0,88±0,11	1,00±0,00
/a/ [nis]	0,76±0,09	0,53±0,07	0,83±0,08	1,00±0,00
/i/ [wys]	0,76±0,12	0,55±0,05	0,79±0,17	1,00±0,00
/i/ [norm]	0,79±0,12	0,58±0,06	0,84±0,09	1,00±0,00
/i/ [nis]	0,76±0,12	0,60±0,10	0,84±0,11	1,00±0,00
/u/ [wys]	0,78±0,13	0,56±0,05	0,88±0,10	1,00±0,00
/u/ [norm]	0,81±0,10	0,68±0,16	0,84±0,07	1,00±0,00
/u/ [nis]	0,82±0,12	0,64±0,10	0,84±0,11	1,00±0,00

niewielka (łącznie tylko 45 mężczyzn). Samogłoska /a/ w intonacji normalnej wykazuje najwyższą średnią czułość dla wszystkich chorób oraz grupy mężczyzn zdrowych.

Do opisu ilościowego jakości klasyfikacji wykorzystano miarę *F-score*. Wyniki znajdują się w tab. 4.10 i 4.11. W grupie kobiet oraz mężczyzn najwyższy wynik otrzymano w przypadku detekcji grupy zdrowych ($F1=1,00$) oraz dla porażenia nerwu krtaniowego wstecznego - 0,75. Przyglądając się każdej z chorób osobno, w przypadku grupy kobiet najwyższy wynik $F1$ uzyskano dla samogłoski /a/ w intonacji normalnej ($F1=0,83$) podczas klasyfikacji porażenia nerwu krtaniowego wstecznego, a najniższy dla ostrego zapalenia krtani ($F1=0,61$). Najwyższą średnią wartość $F1$ uzyskano dla samogłoski /u/ w intonacji normalnej oraz niskiej ($F1=0,81$).

Tabela 4.9. Wartości średnie wraz z odchyleniem standardowym parametru czułość uzyskane w wyniku klasyfikacji na podstawie sygnału głosu mężczyzn, [wys] - intonacja wysoka, [norm] - intonacja normalna, [nis] - intonacja niska.

samogłoska	HF	lar	par	zdrowi
/a/ [wys]	0,58±0,09	0,78±0,16	0,67±0,07	1,00±0,00
/a/ [norm]	0,56±0,13	0,86±0,06	0,78±0,15	1,00±0,00
/a/ [nis]	0,55±0,13	0,70±0,10	0,66±0,07	1,00±0,00
/i/ [wys]	0,56±0,16	0,73±0,20	0,73±0,12	1,00±0,00
/i/ [norm]	0,59±0,11	0,79±0,17	0,66±0,08	1,00±0,00
/i/ [nis]	0,55±0,16	0,70±0,21	0,65±0,07	1,00±0,00
/u/ [wys]	0,51±0,12	0,73±0,21	0,67±0,10	1,00±0,00
/u/ [norm]	0,50±0,10	0,80±0,10	0,67±0,06	1,00±0,00
/u/ [nis]	0,60±0,18	0,73±0,20	0,70±0,13	1,00±0,00

W grupie mężczyzn, najwyższy wynik F1 osiągnięto dla ostrego zapalenia krtani (F1=0,79, /a/ intonacja normalna), a najniższy dla dysfonii hyperfunkcjonalnej (F1=0,54 /a/ intonacja niska). Według tego parametru najlepsze wyniki klasyfikacji uzyskano dla samogłoski /a/ w intonacji normalnej (F1=0,79).

Na podstawie parametru F1 możliwe jest wyznaczenie samogłoski i intonacji, dla której algorytm najlepiej przeprowadził klasyfikację. U kobiet jest to samogłoska /u/ w intonacji normalnej oraz niskiej, a u mężczyzn samogłoska /a/ wypowiedziana w intonacji normalnej. Uzyskana wysoka wartość parametru F1 wskazuje na możliwości stosowania w praktyce opracowanego algorytmu klasyfikacji.

Tabela 4.10. Wartości średnie wraz z odchyleniem standardowym parametru F1 uzyskane w wyniku klasyfikacji na podstawie sygnału głosu kobiet, [wys] - intonacja wysoka, [norm] - intonacja normalna, [nis] - intonacja niska.

samogłoska	HF	lar	par	zdrowi
/a/ [wys]	0,78±0,06	0,67±0,09	0,74±0,12	1,00±0,00
/a/ [norm]	0,77±0,06	0,63±0,11	0,83±0,10	1,00±0,00
/a/ [nis]	0,76±0,06	0,61±0,05	0,74±0,11	1,00±0,00
/i/ [wys]	0,75±0,10	0,63±0,04	0,73±0,17	1,00±0,00
/i/ [norm]	0,62±0,14	0,58±0,20	0,74±0,09	1,00±0,00
/i/ [nis]	0,80±0,06	0,68±0,06	0,69±0,09	1,00±0,00
/u/ [wys]	0,78±0,09	0,66±0,05	0,76±0,06	1,00±0,00
/u/ [norm]	0,81±0,07	0,71±0,09	0,78±0,10	1,00±0,00
/u/ [nis]	0,81±0,06	0,72±0,08	0,76±0,09	1,00±0,00

Tabela 4.11. Wartości średnie wraz z odchyleniem standardowym parametru F1 uzyskane w wyniku klasyfikacji na podstawie sygnału głosu mężczyzn, [wys] - intonacja wysoka, [norm] - intonacja normalna, [nis] - intonacja niska.

samogłoska	HF	lar	par	zdrowi
/a/ [wys]	0,59±0,13	0,64±0,08	0,75±0,07	1,00±0,00
/a/ [norm]	0,61±0,16	0,79±0,09	0,77±0,09	1,00±0,00
/a/ [nis]	0,54±0,13	0,60±0,05	0,72±0,06	1,00±0,00
/i/ [wys]	0,63±0,15	0,60±0,19	0,78±0,09	1,00±0,00
/i/ [norm]	0,66±0,10	0,60±0,18	0,72±0,07	1,00±0,00
/i/ [nis]	0,59±0,17	0,56±0,17	0,71±0,05	1,00±0,00
/u/ [wys]	0,56±0,13	0,60±0,15	0,70±0,03	1,00±0,00
/u/ [norm]	0,54±0,11	0,60±0,11	0,75±0,06	1,00±0,00
/u/ [nis]	0,61±0,19	0,58±0,14	0,74±0,06	1,00±0,00

4.2. Kontrola stanu neurologicznego u pacjentów z chorobą Parkinson'a

W chorobie Parkinson'a ważne jest śledzenie wszystkich objawów powstających w wyniku postępującej neurodegeneracji mózgu. Dokładne monitorowanie pacjentów z PD (choroba Parkinson'a, ang. *Parkinson's disease*) pozwala na szybkie podejmowanie decyzji dotyczących ich leczenia i terapii. Dodatkowo, jeśli takie monitorowanie odbywa się na podstawie nagrań mowy, wtedy skutki leczenia mogą być śledzone zdalnie, a koszty leczenia mogłyby być zdecydowanie niższe.

W badaniu osób cierpiących z powodu choroby Parkinson'a wzięło udział łącznie 33 pacjentów. 5 pacjentów nie zostało poddanych analizie akustycznej ze względu na ich choroby współtowarzyszące, obecności elektrody do głębokiej stymulacji mózgu czy też konieczność opuszczenia badania przed czasem. Niemalże każdy z pacjentów pojawił się na badaniu z opiekunem. Rejestrację głosu przeprowadzono w wydzielonym pomieszczeniu szpitala. Dla pacjentów, którzy do poruszania się nie wymagali pomocy osób trzecich przygotowano specjalny fotel do ustabilizowania postawy siedzącej i eliminowania dźwięków generowanych przez występujące drżenie ciała. Z otoczenia pacjentów zostały usunięte przedmioty, które mogłyby zaburzyć pomiary (rzeczy prywatne, telefony komórkowe) poprzez wytwarzanie dodatkowego dźwięku. W ocenie pacjentów, postawione zadanie wypowiedzenia samogłosek o stabilnej intonacji było łatwe. Trudnością jedynie było utrzymanie przedłużonej fonacji, bez przerywania, jednakże po kilku próbach zadanie to było wykonywane bezbłędnie. Podczas wywiadu lekarskiego przeprowadzanego przed przystąpieniem do badania, trudnością dla pacjentów było przedstawienie obiektywnego obrazu klinicznego choroby, przedstawienie wszystkich symptomów choroby. Bardzo często w trakcie trwania badania przez autora pracy lub lekarza obserwowane były dodatkowe objawy choroby, których pacjenci wcześniej nie zgłaszali lub nie zauważali w tym

nic niepokojącego. Dotyczyły one: pojawienia się niewyraźnej mowy, jąkania się, trudności w rozpoczęciu mówienia, pojawienia się mowy monotonnej (bez zmian intonacji), obecności dodatkowych ruchów płasawicznych kończyn dolnych, drżenia pojedynczych palców kończyn górnych.

Do oceny stanu neurologicznego pacjentów z chorobą Parkinsona wykorzystano w pracy nagrania samogłosek /a/, /e/, /i/, /o/ oraz /u/ o przedłużonej fonacji. Równocześnie z procesem nagrywania głosu pacjentów lekarz przeprowadzał ocenę stanu pacjenta w skali UPDRS. Ta ocena stanowiła referencję pozwalającą ocenić skuteczność klasyfikacji pacjentów na podstawie sygnału mowy. Dla potrzeb implementacji algorytmu, dane wejściowe oraz odpowiadające im referencyjne wyniki wyrażone w skali UPDRS zostały przedstawione w postaci numerycznej za pomocą wektorów zawierających parametry akustyczne podzielonych na 3 grupy:

- zawartość szumów: stosunek harmonicznym do szumu,
- periodyczność: częstotliwość podstawowa, współczynnik jitter, pięciopunktowy i jedena-stopunktowy współczynnik zaburzeń okresu (*ppq5*, *ppq11*), współczynnik shimmer, trój-punktowy współczynnik zaburzeń amplitudy (*apq3*), pięciopunktowy współczynnik zaburzeń amplitudy (*apq5*), energia, zerowy, pierwszy, drugi i trzeci moment widmowy, kurtoza, współczynnik mocy względnej,
- nieliniowość: amplituda i częstotliwość 1-, 2-, 3- formantu, 12 współczynników mel-cepstralnych.

Szczegółowy opis poszczególnych parametrów przedstawiono w rozdziale 3.2.

Końcowy wybór parametrów został poprzedzony przeprowadzeniem testu *Pearson'a* (r) oraz testu *Spearman'a* (ρ). Na podstawie parametru r otrzymanego w wyniku testu *Pearson'a* określono siłę liniowego związku pomiędzy wyróżnionymi grupami parametrów akustycznych i skalą UPDRS-III. Aby wykryć zależności nieliniowe wykorzystano test *Spearman'a*. W jego wyniku obliczono parametr ρ , który opisuje siłę związku nieliniowego pomiędzy parametrami akustycznymi i skalą UPDRS-III. Wyniki obu testów dla każdej z samogłosek oraz wyróżnionych grup parametrów znajdują się w tabeli 4.12.

Najwyższe korelacje (*Spearman'a* oraz *Pearson'a*) obliczone dla każdej z samogłosek oraz analizowanych grup parametrów akustycznych (tabela 4.12) otrzymano dla wszystkich wybranych parametrów akustycznych łącznie. Ponadto, test *Spearman'a* praktycznie dla każdej samogłoski daje wyższe wyniki niż test *Pearson'a*, co świadczy o większym udziale zależności nieliniowych pomiędzy parametrami i skalą UPDRS-III.

Do badań wykorzystano nagrania samogłosek o przedłużonej fonacji pacjentów cierpiących z powodu choroby Parkinson'a. Wszyscy pacjenci byli polsko-języczni. Grupy uczące i testujące zostały utworzone losowo, przy czym w żadnym przypadku określony pacjent nie został przypisany równocześnie do grupy testowej i uczącej. Do wyznaczenia punktacji w skali

Tabela 4.12. Wyniki testów korelacji *Pearson'a* (r) oraz *Spearman'a* (ρ) dla analizowanych grup parametrów akustycznych i punktacji w skali UPDRS-III.

grupy parametrów	zawartość szumów		periodyczność		nieliniowość		wszystkie parametry	
	ρ	r	ρ	r	ρ	r	ρ	r
/a/	0,29	0,27	0,42	0,41	0,50	0,42	0,70	0,63
/e/	0,39	0,38	0,43	0,52	0,58	0,52	0,68	0,66
/i/	0,49	0,41	0,54	0,40	0,50	0,46	0,74	0,62
/o/	0,32	0,37	0,43	0,48	0,63	0,58	0,74	0,62
/u/	0,45	0,42	0,46	0,42	0,58	0,52	0,69	0,57

Tabela 4.13. Porównanie wyników estymacji punktacji [pkt] w skali UPDRS-III z uwzględnieniem parametrów opisujących zawartość szumów, periodyczność, nieliniowość oraz wszystkich parametrów łącznie z oceną lekarską, STD - odchylenie standardowe.

samogłoska	/a/	/e/	/i/	/o/	/u/
<i>zawartość szumów</i>					
średnia modułu różnicy [pkt]	13,52	16,48	12,25	20,41	6,22
STD różnicy [pkt] 9,45	10,22	8,37	10,71	10,70	
<i>periodyczność</i>					
średnia modułu różnicy [pkt]	12,12	14,95	11,87	15,60	15,47
STD różnicy [pkt]	10,26	10,95	9,43	6,72	8,82
<i>nieliniowość</i>					
średnia modułu różnicy [pkt]	9,84	12,42	10,29	11,69	10,42
STD różnicy [pkt]	8,70	8,31	8,68	9,27	8,72
<i>wszystkie parametry</i>					
średnia modułu różnica [pkt]	8,84	8,36	8,82	8,85	9,50
STD różnicy [pkt]	7,05	7,63	7,23	7,12	7,26

UPDRS wykorzystano regresję ϵ -SVR z użyciem jądra Gaussa. Ocena jakości estymacji punktacji w skali UPDRS-III została wykonana przy użyciu k -krotnego sprawdzianu krzyżowego, gdzie $k=9$.

Na podstawie wektorów parametrów akustycznych algorytm ϵ -SVR przeprowadzał estymację punktacji w skali UPDRS opisując stan neurologiczny pacjentów. Estymację wykonano w 4 konfiguracjach parametrów akustycznych: uwzględniając tylko zawartość szumów, parametry opisujące periodyczność, nieliniowość oraz wszystkie parametry łącznie. Wyniki przedstawiono w tabeli 4.13. Na ich podstawie, można zauważyć, że wektor wejściowy uwzględniający grupę parametrów określających zawartość szumów i periodyczność prezentuje średnie różnice w punktacji wyjściowej w porównaniu z oceną lekarską na poziomie 6-16 punktów, z odchyleniem standardowym mieszczącym się w zakresie 6-11 punktów. Nieco lepsze wyniki otrzymano na podstawie wektorów uwzględniających parametry opisujące nieliniowość. Średnia różnica

punktów dla każdej z samogłosek wynosi 8-12 punktów, STD 8-9 punktów. Wykorzystując wszystkie parametry razem otrzymano najmniejszy błąd wyrażony różnicą punktów pomiędzy oceną uzyskaną na podstawie badanego algorytmu, a oceną lekarską i wynosi on 8-9 punktów, a odchylenie standardowe mieści się w zakresie 7,05-7,63 punktów. Średnia różnica pomiędzy wartościami otrzymanymi w wyniku algorytmu ϵ -SVR oraz punktacją skali UPDRS-III wyznaczoną przez lekarza nie przekracza 10 punktów dla każdej z samogłosek. Minimalna wartość średniego błędu jest na poziomie 7,74% (8,36 pkt), a maksymalna - 8,8% (9,50 pkt). Odchylenie standardowe różnicy punktów również jest zbliżone dla wszystkich samogłosek oraz parametrów i wynosi 6,5%. Podsumowując, najlepsze wyniki estymacji punktacji UPDRS-III uzyskano, gdy wykorzystano łącznie wszystkie parametry akustyczne oraz nagrania samogłoski /a/, jednak wyniki uzyskane dla pozostałych samogłosek są bardzo zbliżone.

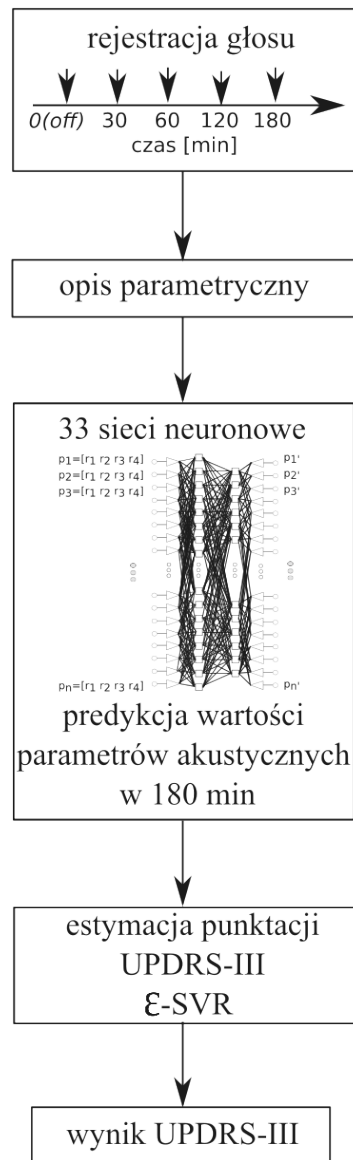
4.3. Predykcja stanu neurologicznego u pacjentów po spożyciu leków

Pod pojęciem predykcji rozumie się proces przewidywania jaką wartość przyjmie zmienna zależna przy ustalonych wartościach zmiennej niezależnej w przyszłości [115]. Wynikiem procesu predykcji jest prognoza wyrażona wynikiem liczbowym. W pracy predykcja polegała na wyznaczeniu punktacji UPDRS-III w 180 minucie po spożyciu leku na podstawie opisu parametrycznego sygnału głosu pacjentów z chorobą Parkinson'a zarejestrowanego w chwili 0 oraz 60 i 120 minut po podaniu leku.

W tym celu przeanalizowano możliwość prognozy nasilenia objawów choroby Parkinson'a z wykorzystaniem tylko sygnału głosu. Zadanie to polega na predykcji wartości parametrów akustycznych opisujących sygnał mowy oraz estymacji na tej podstawie stanu pacjenta w skali UPDRS-III. Ponieważ skala UPDRS tylko w jednym z podpunktów uwzględnia ocenę mowy pacjenta, to zadanie estymacji przyszłego stanu pacjenta jest trudne. Powyższa idea opiera się na założeniu, że istnieje dostatecznie silna korelacja pomiędzy upośledzeniem mowy pacjentów wyrażonym w skali UPDRS, a oceną jego stanu.

Zadanie predykcji oceny stanu pacjenta wyrażonej w skali UPDRS-III zostało podzielone na dwie części. Pierwsza dotyczy predykcji wektora parametrów akustycznych w 3 godzinie od spożycia leków, natomiast druga uwzględnia przypisanie wyznaczonemu wektorowi wartości skali UPDRS-III. Podstawą predykcji były wyniki analizy sygnału mowy zarejestrowanej w 4 określonych odstępach czasowych. Schemat przeprowadzonego procesu predykcji przedstawiono na rys. 4.7.

Predykcja parametrów akustycznych została przeprowadzona z wykorzystaniem sieci neuronowych. Dane wejściowe stanowił opis parametryczny 4 nagrań głosu pacjentów z PD. Pomiar polegał na rejestracji samogłosek o przedłużonej fonacji /a/, /e/, /i/ /o/ oraz /u/. Pierwszą rejestrację przeprowadzono, kiedy pacjent znajdował się w stanie *off*, kolejną po 30 minutach



Rys. 4.7. Schemat procesu predykcji stanu pacjenta w 180 minucie.

od spożycia leków podanych przez lekarza, następnie po 60 i 120 minutach od pierwszego pomiaru. Na etapie uczenia dane wyjściowe sieci neuronowej stanowiły wyniki analizy akustycznej wykonanej w oparciu o nagrania głosu wykonane w 180 minucie od zażycia leków. Sieci neuronowe wykorzystane w tym celu to sieci typu perceptron wielowarstwowy z uczeniem nadzorowanym (z nauczycielem). Metoda uczenia sieci to algorytm wstecznej propagacji błędów (BP, ang. *backpropagation neural network*). Dokładny opis grupy badawczej znajduje się w rozdziale 3.1.

W celu przystosowania sieci do potrzeb postawionego problemu predykcji stanu pacjenta dane pomiarowe zostały podzielone na 3 zbiory w proporcjach: 80% stanowił zbiór uczący, 10% zbiór walidacyjny, 10% zbiór testowy. Dane ze zbioru uczącego były wykorzystane do obliczania gradientu i aktualizacji wag. Zbiór walidacyjny służył do określenia momentu, w którym sieć zostaje przeuczona. Do określenia takiego momentu posłużył błąd średniokwadratowy (ang. *mean squared error*, MSE). Jest to współczynnik błędu prognozy *ex-post*, zdefiniowany wzorem:

$$MSE = \frac{1}{M} \sum_{i=1}^M (w\tilde{e}j_n - wyj_n)^2 \quad (4.3)$$

gdzie:

$w\tilde{e}j_n$ - wartość zwrócona przez model (wyjście sieci),

wyj_n - rzeczywista wartość danego parametru,

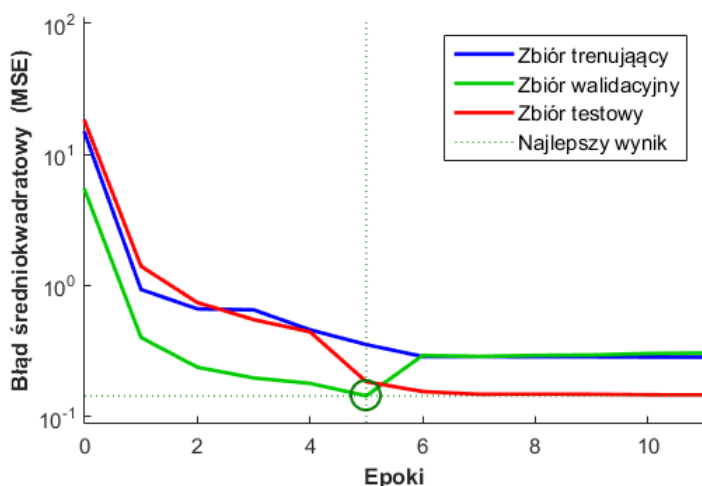
M - liczba pacjentów.

Im mniejszy błąd MSE, tym model jest lepszy. Objawem przeuczenia jest wzrost błędu MSE zbioru walidacyjnego oraz osiągnięcie ustalonej wartości maksymalnej lub kilkukrotny (cykliczny) wzrost błędu. Przykładowy proces generalizacji dla parametru 1 (predykcja częstotliwości podstawowej w 180 minucie) przedstawiony jest na rysunku 4.8. Uczenie dla tego przypadku zostało zakończone już po 5 epokach.

Uczenie zostaje zatrzymane w miejscu, w którym błąd dla zbioru walidacyjnego jest minimalny. Zbiór testowy służył do wyznaczenia miary jakości sieci.

Dla każdego z parametrów akustycznych opisanego w rozdziale 4.2 została utworzona osobna sieć neuronowa o indywidualnej budowie i parametrach. W badaniach przeprowadzono analizę dla różnych kombinacji ustawień sieci i wybrano te, dla których wynik predykcji był najbardziej zbliżony do wartości rzeczywistej. Testy zostały przeprowadzone dla różnych konfiguracji sieci uwzględniając liczbę neuronów w warstwie ukrytej, rodzaj algorytmu uczącego sieć, rodzaj użytej funkcji aktywacji w neuronach ukrytych oraz wyjściowym. Ustawienia sieci dla każdego z parametrów znajdują się w dodatku B w tab. B.1, B.2, B.3, B.4, B.5, B.6.

Najlepsze wyniki osiągnięto dla dwóch metod trenowania sieci jednokierunkowych: metody Lavenberg'a (*trainlm*) oraz metody gradientowej Broydena-Fletcher-Goldfarba-Shano



Rys. 4.8. Proces generalizacji sieci neuronowej dla predykcji częstotliwości podstawowej (parametr 1) w 180 minucie, samogłoska /a/.

(*trainbfg*). *Trainlm* jest najszybszą metodą uczenia wszystkich sieci jednokierunkowych. Oparta jest na optymalizacyjnym algorytmie Levenberga–Marquardta [116]. Podczas testów symulacyjnych przeprowadzonych w ramach badań, algorytm *trainlm* pozwalał na uzyskanie poprawnych wyników po znacznie mniejszej liczbie epok procesu uczenia w porównaniu do innych metod.

Do oceny wyniku działania sieci wykorzystano liniowe modele regresyjne wyznaczone dla każdego ze zbiorów: uczącego, walidacyjnego i testowego wiążące ten wynik z wartością rzeczywistą. W idealnym przypadku, wyniki powinny układać się wzdłuż prostej nachylonej pod kątem 45 stopni [117]. Oznacza to, że otrzymane wyniki działania sieci są identyczne jak rzeczywiste. Przykładowe modele dla dwóch parametrów: częstotliwości podstawowej i współczynnika shimmer dla 3 losowo wybranych pacjentów ze zbioru testowego w wyniku sprawdzianu krzyżowego znajdują się na rys. 4.9. Jako kryterium ilościowe pozwalające ocenić jakość sieci przyjęto współczynnik korelacji R, którego wartość dla wszystkich analizowanych zbiorów musiała być równa lub większa od wartości 0,70. W innym przypadku, sieć była trenowana ponownie z innymi ustawieniami aż do spełnienia przyjętego kryterium.

Po zakończeniu uczenia, walidacji oraz testów, w każdej 8-krotnej próbie sprawdzianu krzyżowego wyznaczany jest błąd średniokwadratowy MSE dla każdego zbioru testowego. Uzyskane wartości błędów średniokwadratowych MSE znajdują się w dodatku C w tab. C.1, C.2, C.3, C.4, C.5, C.6. Najmniejszy błąd MSE, który wynosił $0,02 \pm 0,02$ otrzymano dla zerowego momentu widmowego (współczynnik korelacji $R = 0,99 \pm 0,01$), samogłoski /e/ oraz dla częstotliwości 3-formantu, samogłoski /u/ (współczynnik korelacji $R = 0,98 \pm 0,03$). Największy błąd MSE dla zbioru testowego wykazywał pięciopunktowy współczynnik zaburzeń amplitudy dla samogłoski /o/ i wynosił $0,18 \pm 0,27$, współczynnik $R = 0,97 \pm 0,05$ oraz dla amplitudy 3-ciego

Tabela 4.14. Średnia [%] oraz odchylenie standardowe (STD) modułu różnicy pomiędzy wartością rzeczywistą i predykcją stanu pacjenta w skali UPDRS-III dla wszystkich parametrów sygnału mowy poddanych analizie dla każdej z samogłosek.

samogłoska	/a/	/e/	/i/	/o/	/u/
średni błąd prognozy [%]	3,27	3,69	3,35	3,52	3,26
STD błędu prognozy [%]	3,30	3,40	3,18	3,64	3,15

Tabela 4.15. Średnia [pkt] oraz odchylenie standardowe (STD) modułu różnicy [pkt] pomiędzy punktacją rzeczywistą i wynikiem prognozy stanu pacjenta w skali UPDRS-III uzyskanej z wykorzystaniem sieci neuronowej oraz algorytmu ϵ -SVR dla każdej z analizowanych samogłosek.

samogłoska	/a/	/e/	/i/	/o/	/u/
średnia różnica [pkt]	3,93	4,41	4,58	3,64	3,93
średnia różnica [%]	3,63	4,08	4,24	3,37	3,64
STD średniej różnicy [pkt]	5,34	5,33	3,94	4,60	4,56
STD średniej różnicy [%]	4,94	4,94	3,64	4,26	4,22

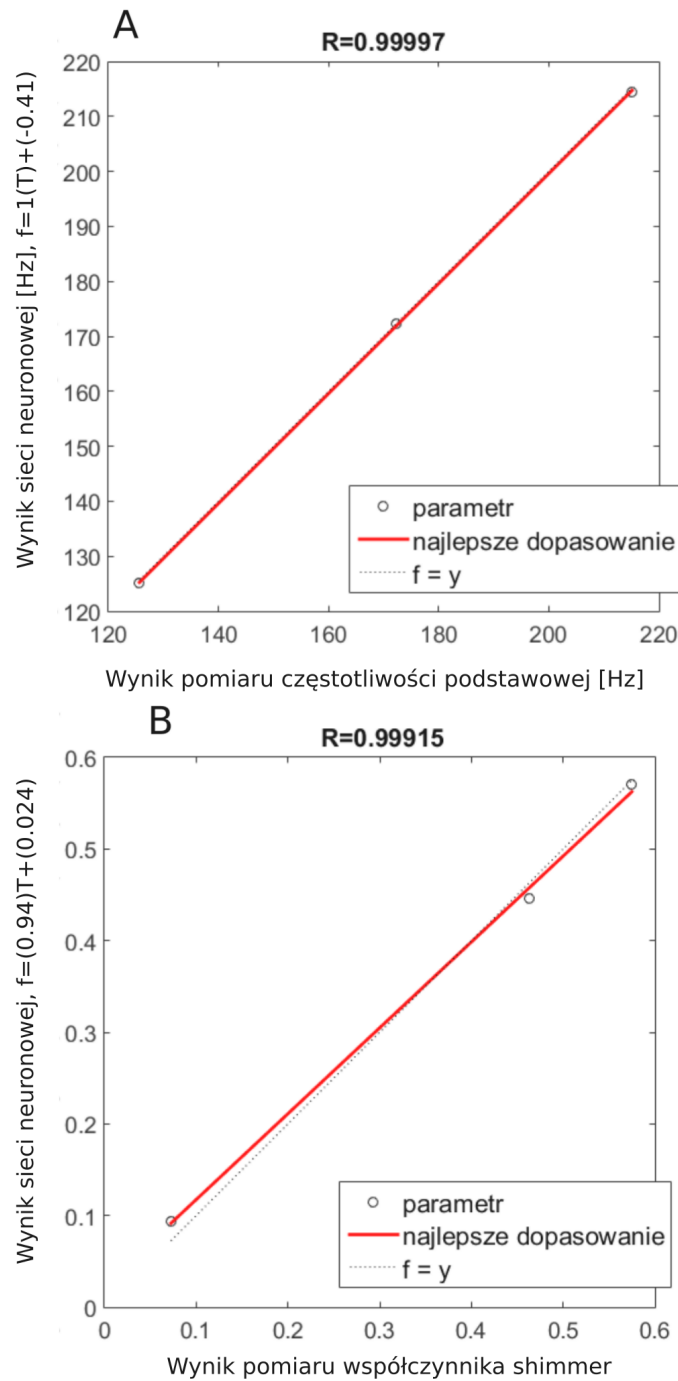
formantu samogłoski /e/ i wynosił on $0,18 \pm 0,23$, $R=0,83 \pm 0,37$. Wyznaczone modele regresyjne były podstawą predykcji wartości danych parametrów akustycznych w 180 minucie od spożycia leków.

Biorąc pod uwagę każdą z samogłosek osobno, wyniki predykcji najbardziej zbliżone do wartości rzeczywistych uzyskano dla parametrów:

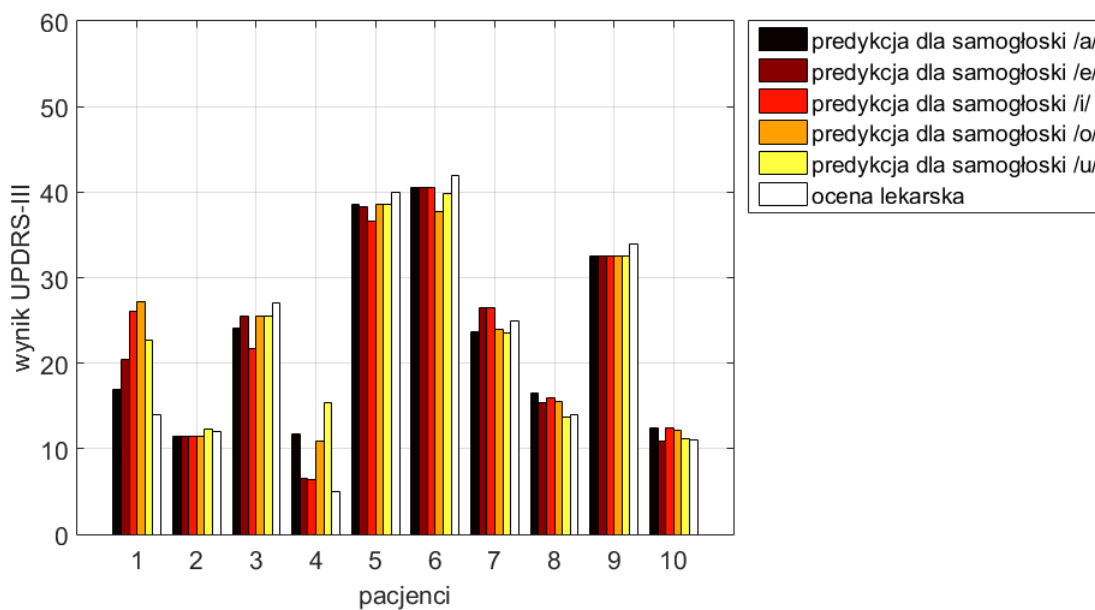
- /a/: amplituda 1 formantu, 7 współczynnik mel-cepstralny ($MSE_{test} = 0,04$),
- /e/: zerowy moment widmowy ($MSE_{test} = 0,02$),
- /i/: częstotliwość 3 formantu ($MSE_{test} = 0,05$),
- /o/: 10 i 11 współczynnik mel-cepstralny ($MSE_{test} = 0,05$),
- /u/: 3 formant ($MSE_{test} = 0,02$).

Wyniki średniego błędu prognozy dla każdej z analizowanych samogłosek, biorąc pod uwagę wszystkie parametry akustyczne, zamieszczono w tabeli 4.14. Wartość błędu prognozy oraz odchylenia standardowego otrzymane dla zdefiniowanych samogłosek są porównywalne i mieszczą się w przedziale 3,3-3,7%. Najmniejszy średni błąd prognozy otrzymano dla samogłoski /a/.

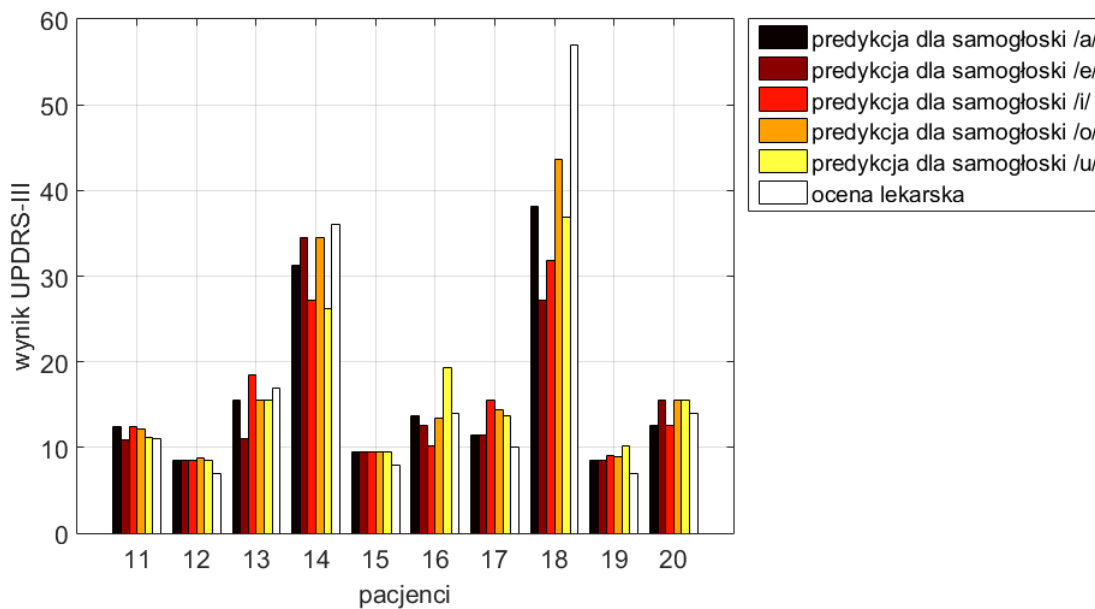
Ostatnim etapem prac było przypisanie punktacji w skali UPDRS-III wyznaczonym prognozom wektora parametrów sygnału mowy. W tym celu wykorzystano ponownie algorytm ϵ -SVR, którego opis znajduje się w rozdziale 3.4. Wykresy zaprezentowane na rysunkach 4.10, 4.11 oraz 4.12 przedstawiają wyniki punktacji dla każdego z pacjentów dla każdej z samogłosek oraz wartość referencyjną (wynik ustalony przez lekarzy).



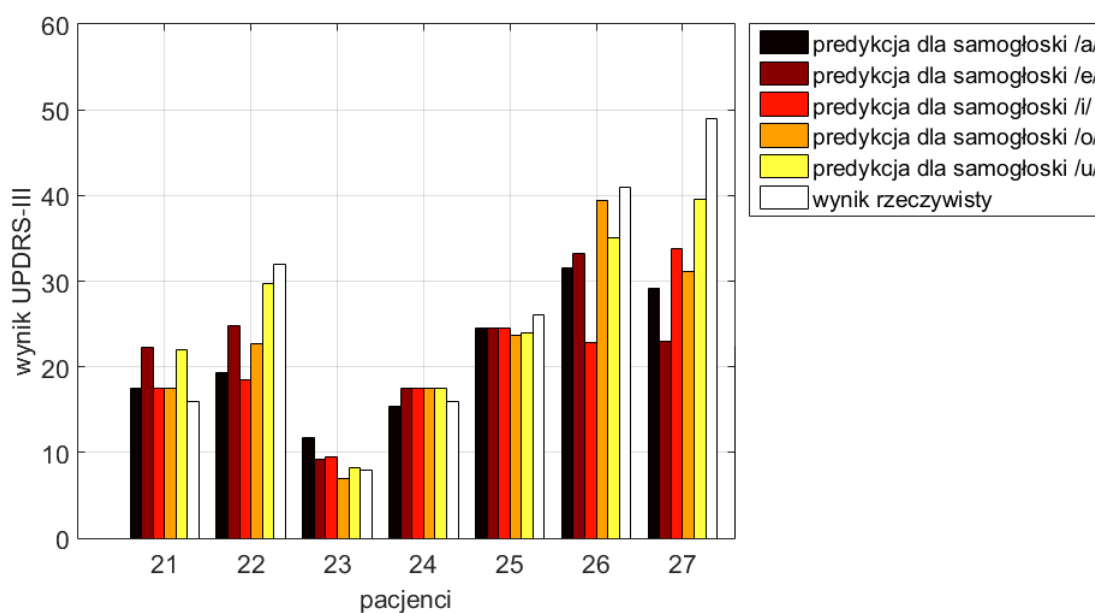
Rys. 4.9. Modele regresyjne dla 3 pacjentów z grupy testowej dla A) częstotliwości podstawowej, samogłoska /a/, B) współczynnika shimmer, samogłoska /i/, T =rzeczywista wartość danego parametru.



Rys. 4.10. Predykcja punktacji UPDRS-III w 180 minucie na podstawie nagrań samogłosek o przedłużonej fonacji dla 10 pacjentów.



Rys. 4.11. Predykcja punktacji UPDRS-III w 180 minucie na podstawie nagrań samogłosek o przedłużonej fonacji dla kolejnych 10 pacjentów.

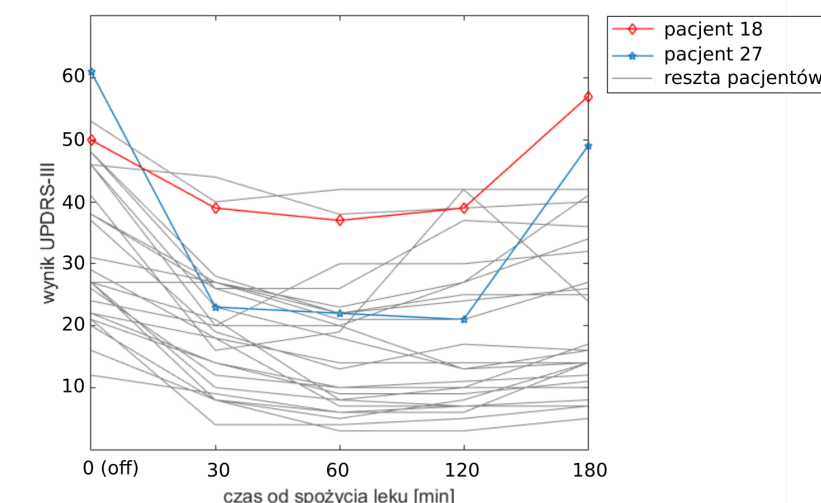


Rys. 4.12. Predykcja punktacji UPDRS-III w 180 minucie na podstawie nagrań samogłosek o przedłużonej fonacji dla pozostałych 8 pacjentów.

Jakość prognozy skali UPDRS wyrażona jako wartość średnia modułu różnicy pomiędzy punktacją rzeczywistą, a prognozowaną znajduje się w tabeli 4.15. Na podstawie wyznaczonej prognozy stanu pacjenta wyrażonego w punktacji UPDRS (rys. 4.10, 4.11, 4.12, tabela 4.15) widoczne jest, że najbliższy wartości rzeczywistej wynik został osiągnięty dla samogłoski /o/ (średnia różnica = 3,64%). Biorąc pod uwagę również odchylenie standardowe wskazujące na stabilność otrzymanej wartości, warto przyjrzeć się samogłosce /i/, którą charakteryzuje najniższe STD. Należy tutaj zauważyć, że średnia różnica punktacji dla tej samogłoski w porównaniu do samogłoski /o/ jest wyższa o około 1 punkt UPDRS.

Na rys. 4.10, 4.11 oraz 4.12 widoczna jest znacząca różnica pomiędzy prognozą, a wartością rzeczywistą dla dwóch pacjentów (rys. 4.11 - pacjent 18, 4.12 - pacjent 27). U obydwu pacjentów występowały fluktuacje ruchowe spowodowane skróconym czasem działania leków, dystonia końca dawki (np. bolesne kurcze mięśni z wykręcaniem kończyny) oraz nadmiar ruchów płasawicznych. Podczas badania nie stwierdzono słyszalnych zmian mowy. Wyniki punktacji w skali UPDRS-III były najwyższe u tych pacjentów spośród całej grupy badawczej w 180 minucie badania (rys. 4.13). Pomimo, że w obu przypadkach występował spadek punktacji UPDRS wraz z podaniem leków, a następnie wzrost w 180 minucie od spożycia lewo-dopy, algorytm nie poradził sobie z tym zadaniem. Przyczyną takiego odstępstwa była prawdopodobnie mało liczna grupa ucząca zawierająca takie szczególne przypadki.

Badania przeprowadzone na grupie 27 pacjentów wykazały skuteczność zaproponowanego podejścia. Średni błąd predykcji oceny chorych na PD w skali UPDRS wynosi od 3,3% do



Rys. 4.13. Graficzna prezentacja zmian punktacji.

3,7% w zależności od analizowanej samogłoski. Najniższy błąd otrzymano dla samogłoski /a/ ($\sim 3,27\%$). Najwyższy natomiast wykazała samogłoska /e/ ($\sim 3,67\%$).

Średni błąd estymacji punktacji UPDRS-III obliczony na podstawie wartości parametrów prognozowanych wynosi od 3,9-4,6 punktów, z czego najmniejszą różnicę pomiędzy punktacją rzeczywistą (wykonaną przez lekarza) i wynikiem predykcji wykazały samogłoska /a/ oraz /u/ ($\sim 3,93$ punktów). Największą różnicę punktacji otrzymano dla samogłoski /i/ ($\sim 4,58$ punktów).

Skala UPDRS pozwala na ocenę nasilenia objawów oraz stopnia zaawansowania choroby Parkinson'a. W pracy skupiono się na ocenie pacjentów w okresie trwającym łącznie około 4 godzin, uwzględniając stan *off* oraz okres do 3 godzin od spożycia leków łagodzących skutki choroby. Na podstawie pojawiających się zmian stanu pacjentów, możliwa była walidacja i ewentualna modyfikacja dawek leków i dostosowanie do aktualnego stopnia zaawansowania choroby i nasilenia objawów. Nie jest możliwe, aby lekarz spędził okres około 4 godzin z pacjentem sam na sam, aby móc ocenić dokładnie jego stan w danym przedziale czasowym. Taki okres pozwala jednak na obiektywną ocenę nasilenia choroby i pojawiających się objawów. Narzędzie wspomagające lekarza w ocenie punktacji UPDRS-III pacjenta w danej chwili czasowej może ułatwić i udokumentować szczegółowo proces monitorowania chorych. Dodatkowo, możliwość predykcji stanu pacjenta w przyszłości skraca czas monitorowania. Na podstawie przedstawionej metodologii badawczej możliwa jest automatyzacja oceny stanu pacjenta w skali UPDRS-III oraz skrócenie czasu badania. W wyniku realizacji zaproponowanej procedury otrzymano różnicę punktów na poziomie 3-4%. Taka mała wartość błędów nie ma znaczenia diagnostycznego.

Uzyskane wyniki należy uznać za obiecujące. Narzędzie, które nie będzie wymagać interwencji lekarza, pozwalające na krótko-czasową (3 godzinną) analizę zmian stanu neurologicznego pacjenta i jego reakcję na leki pozwoliłoby na spersonalizowaną diagnozę, utworzoną w zdecydowanie krótszym czasie. W przyszłości planuje się przeprowadzenie identycznego badania w celu porównania wyników analizy akustycznej w kontekście nasilenia choroby Parkinson'a w dłuższym okresie czasu.

5. Podsumowanie

Celem rozprawy było przeprowadzenie badań zmierzających do ustalenia w jakim stopniu zmiany w sygnale mowy mogą zostać wykorzystane jako źródło informacji diagnostycznej, prognostycznej i kontrolnej przy rozwiązywaniu niektórych problemów medycznych związanych z wybranymi chorobami laryngologicznymi. W wyniku badań stworzono system, którego zadaniami było: automatyczna detekcja zaburzeń głosu, prognoza postępu choroby Parkinson'a (PD), estymacja punktacji w skali UPDRS opisującej nasilenie objawów choroby Parkinson'a. Przyjęty cel rozprawy został osiągnięty. Informacje będące wynikiem działania systemu mogą zostać wykorzystane w diagnostyce wybranych schorzeń laryngologicznych. Mogą również wesprzeć lekarza w przewidywaniu skutków terapii oraz pomóc kontrolować zmieniający się stan pacjentów z chorobą Parkinson'a. Choroby głosu oraz choroba neurodegeneracyjna Parkinson'a dotyczą milionów ludzi na całym świecie, a koszty leczenia wzrastają każdego dnia. Rozwój narzędzi obliczeniowych, które wspomagają ocenę i monitorowanie chorób, może przyczynić się do zmniejszenia obciążenia ekonomicznego szpitali, przychodni oraz gabinetów lekarskich, a także pomóc pacjentom w poprawie jakości ich życia.

Z przeprowadzonych badań wynika, że samogłoski wypowiedziane w przedłużonej fonacji i w określonej intonacji pozwalają na uzyskanie informacji o stanie kanału głosowego. Wypowiedzenie przez pacjenta jedynie samogłoski o danej intonacji jest łatwe i szybkie, co jest atutem przedstawionej metody w kontekście czasu trwania badania.

W ramach rozprawy zaproponowano i przebadano zestawy parametrów, które umożliwiły przedstawienie sygnału głosu w formie parametrycznej. Taki opis sygnału mowy oraz poddanie go działaniu wybranych algorytmów klasyfikacji, uczenia maszynowego oraz sieci neuronowych pozwoliło na wykrycie nie tylko, czy pacjent posiada schorzenie laryngologiczne czy jest zdrowy, ale też wskazać rodzaj choroby narządu głosu. W tym celu wykorzystano 1410 nagrań samogłosek /a/, /i/, /u/ o przedłużonej fonacji pochodzących z bazy *Saarbruecken Voice Database* (SVD). Klasyfikację binarną, w wyniku której system wskazywał, czy osoba jest zdrowa czy nie, przeprowadzono implementując algorytm jądrowej analizy składowych głównych oraz lasów losowych. Dokładność klasyfikacji była najwyższa dla samogłoski /a/ wypowiedzianej w intonacji normalnej dla grupy kobiet i wyniosła ona 99,89%, a dla mężczyzn 100%. Zaproponowana metodologia pozwala ocenić stan zdrowia pacjenta, bez względu na

postęp choroby narządu głosu. Taki system może być niezwykle użyteczny podczas testów przesiewowych.

Kolejnym etapem prac była detekcja wybranych chorób laryngologicznych i wskazanie jednej, która została zdiagnozowana u danego pacjenta. Do badań wybrano 3 najbardziej liczne jednostki chorobowe występujące w bazie SVD: dysfonię hyperfunkcjonalną, porażenie nerwu krtaniowego wstecznego i ostre zapalenie krtani. Czwartą grupę stanowiły osoby zdrowe. Do realizacji tego zadania wykorzystano sieć neuronową NLPCA. Dokładność diagnozy wskazującej na konkretną patologię głosu była najwyższa (84,13%) dla samogłoski /u/ w intonacji niskiej. System bezbłędnie wykrywał osoby zdrowe. Automatyczne obiektywne i nieinwazyjne narzędzie służące do wykrywania różnych schorzeń narządu głosu na podstawie analizy sygnałów akustycznych może odgrywać ważną rolę we wczesnej diagnostyce, śledzeniu postępu, a nawet poprawie skuteczności leczenia danej jednostki chorobowej. W wyniku choroby Parkinson'a upośledzone zostają czynności ruchowe osoby tj. pojawia się drżenie, sztywność mięśni. Zmiany takie opisywane są przez neurologów w skali UPDRS, części III (UPDRS-III). Skala ta jest uważana za globalny standard oceny pacjentów z PD i została przyjęta w tej pracy jako odniesienie dla walidacji proponowanych metod. Jednakże, skala UPDRS-III składa się z 27 pytań, z czego tylko jedno dotyczy zagadnienia mowy. W związku z tym, w pracy założono, że istnieje silna korelacja pomiędzy skalą UPDRS, a zaburzeniami głosu i mowy obserwowanymi u pacjentów z PD.

Autorka pracy wykonała nagrania głosu u osób z chorobą Parkinson'a w Krakowskim Szpitalu Specjalistycznym im. Jana Pawła II. Nagrania obejmowały wypowiedzenie samogłosek /a/, /e/, /i/, /o/ oraz /u/ w przedłużonej intonacji przez 27 pacjentów. Pierwszy pomiar występował wtedy, kiedy u pacjenta zdiagnozowano zupełne wysycenie leków łagodzących objawy choroby. Następnie lekarz podawał lek lewodopy. Kolejne pomiary wykonano po 30, 60, 120 i 180 minutach od spożycia leku. Przed każdym pomiarem, lekarz neurolog wykonywał badanie i opisywał stan pacjenta w skali UPDRS.

Przy pomocy opisu parametrycznego, skali UPDRS-III, sieci neuronowej z metodą wstecznej propagacji błędów oraz metody regresji wektorów nośnych przeprowadzono prognozę postępu nasilenia choroby w okresie do 3 godzin od spożycia leku lewodopy. Na podstawie zrealizowanych badań stwierdzono, że istnieje korelacja pomiędzy zestawem wybranych parametrów akustycznych, a skalą UPDRS-III. Estymacja punktacji skali UPDRS-III została przeprowadzona przy użyciu metody regresji wektorów nośnych. Następnie porównano otrzymane wyniki ze skalą lekarską. Na ich podstawie wskazano, że wykorzystanie łącznie wszystkich wytypowanych parametrów sygnału mowy dla samogłoski /e/ zapewnia najmniejszą wartość średnią modułu różnicy pomiędzy wartościami wskazanymi przez lekarza i algorytm, wyniosła ona 8,36 pkt (7,74%).

Prognoza polegała na wyznaczeniu punktacji UPDRS-III w 180 minut po spożyciu leku na podstawie opisu parametrycznego sygnału głosu pacjentów z chorobą Parkinson'a zarejestrowanego w chwili pełnego wysycenia lekami oraz 60 i 120 minut po podaniu leku. Zadaniem było przewidzenie wartości składowych wektora akustycznego opisującego sygnał mowy oraz na tej podstawie estymacja oceny stanu pacjenta w skali UPDRS-III. Do tego celu wykorzystano sieć neuronową. Wartości błędów prognozy i jego odchylenia standardowego otrzymane dla zdefiniowanych samogłosek są porównywalne i mieszczą się w przedziale 3,3-3,7%. Ostatnim etapem prac było przypisanie wyznaczonym wektorom parametrów sygnału mowy punktacji w skali UPDRS-III. Średni błąd estymacji punktacji UPDRS-III obliczony na podstawie wartości parametrów prognozowanych wynosi od 3,9-4,6 punktów, z czego najmniejszą różnicę pomiędzy punktacją rzeczywistą i wynikiem prognozy wykazały samogłoski /a/ oraz /u/ (~3,93 punktów).

Uzyskane wyniki należy uznać za obiecujące. Na podstawie przedstawionej metodologii badawczej możliwa jest automatyzacja oceny stanu pacjenta w skali UPDRS-III oraz skrócenie czasu badania. W wyniku realizacji zaproponowanej procedury otrzymano różnice w ocenie stanu pacjenta na poziomie 3-4%. Jest to niewielki błąd w porównaniu z korzyściami, jakie dostarcza przedstawiona metodologia. Narzędzie, które nie będzie wymagać interwencji lekarza, pozwalające na krótko-czasową (3 godzinną) analizę zmian stanu neurologicznego pacjenta i jego reakcję na leki pozwala na utworzenie spersonalizowanej diagnozy wykonanej w zdecydowanie krótszym czasie. Takie podejście pozwala kontrolować proces terapii i przewidzieć jej skutki wykorzystując jedynie sygnał głosu.

Zebrań dodatkowej bazy nagrań sygnałów mowy zdeformowanej pozwoli na wykorzystanie większej liczby przypadków uczących i udoskonalenie wykorzystanych metod zaimplementowanych w rozprawie. Ponadto, przeprowadzenie badań mowy (uwzględniając przeczytanie określonego tekstu) pozwoli na pozyskanie dodatkowych informacji o stanie narządów głosotwórczych. Ważnym aspektem w dalszych badaniach jest wybór urządzenia oraz miejsca, w którym powinna odbywać się rejestracja głosu. Pożądana jest możliwość przeprowadzenia nagrań przy pomocy np. smartphona lub innego małego, przenośnego urządzenia w warunkach na przykład domowych. Dodatkowo, należy mieć na uwadze, aby system do rejestracji sygnału głosu był odporny na ewentualne niepożądane dźwięki z otoczenia pacjenta.

Wykorzystanie analizy głosu może pozwolić na zdalne rejestrowanie sygnału i ocenę stanu zdrowia pacjentów. Ponadto, połączenie analizy głosu z innymi metodami pomiarowymi (np. analiza pisma) mogłoby pomóc w monitorowaniu nie tylko pacjentów z chorobą Parkinson'a, ale również osób starszych.

Wyniki badań związane z pracą zostały opublikowane w pracach [4, 17, 118–121] oraz były prezentowane podczas wielu wystąpień na konferencjach międzynarodowych.

Bibliografia

- [1] Dimitar D Deliyski, Robert E Hillman i Daryush D Mehta. „Laryngeal high-speed videoendoscopy: Rationale and recommendation for accurate and consistent terminology”. W: *Journal of Speech, Language, and Hearing Research* 58.5 (2015), s. 1488–1492.
- [2] Ewa Niebudek-Bogusz i in. „The application of short-term efficiency analysis in diagnosing occupational voice disorders.” W: *Medycyna pracy* (2015).
- [3] J Potepa and Szaleniec i in. „Analysis of Voice Modifications for Persons After Tonsillectomy”. W: *Acta Physica Polonica, A*. 125 (2014).
- [4] Daria Panek. „Ocena parametrów analizy akustycznej w detekcji patologii mowy”. W: *Przegląd Elektrotechniczny* 90.5 (2014), s. 126–129.
- [5] Steven B Davis. „Acoustic characteristics of normal and pathological voices”. W: *Speech and language: advances in basic research and practice* 1.27 (1979), s. 133–164.
- [6] Mirjana Petrović-Lazić i in. „Acoustic voice analysis of patients with vocal fold polyp”. W: *Journal of Voice* 25.1 (2011), s. 94–97.
- [7] Maria Claudia Franca. „Acoustic comparison of vowel sounds among adult females”. W: *Journal of Voice* 26.5 (2012), 671–e9.
- [8] Tanya L Eadie i Philip C Doyle. „Classification of dysphonic voice: acoustic and auditory-perceptual measures”. W: *Journal of Voice* 19.1 (2005), s. 1–14.
- [9] Stefan Hadjitodorov i Petar Mitev. „A computer system for acoustic analysis of pathological voices and laryngeal diseases screening”. W: *Medical engineering & physics* 24.6 (2002), s. 419–429.
- [10] *MEEI Voice Database*. <http://www.masseyeandear.org/specialties/otolaryngology/voice-laboratory/>. [Online; dostęp 05-Lipiec-2017].
- [11] Thomas Drugman, Thomas Dubuisson i Thierry Dutoit. „Phase-based information for voice pathology detection”. W: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 2011, s. 4612–4615.
- [12] Ce Peng i in. „Pathological voice classification based on a single Vowel’s acoustic features”. W: *Computer and Information Technology, 2007. CIT 2007. 7th IEEE International Conference on*. IEEE. 2007, s. 1106–1110.
- [13] Meisam Khalil Arjmandi i Mohammad Pooyan. „An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine”. W: *Biomedical Signal Processing and Control* 7.1 (2012), s. 3–19.

- [14] Saarbruecken Voice Database. <http://www.stimmdatenbank.coli.uni-saarland.de/>. [Online; dostęp 05-Lipiec-2017].
- [15] David Martínez i in. „Score level versus audio level fusion for voice pathology detection on the Saarbrücken Voice Database”. W: *Advances in Speech and Language Technologies for Iberian Languages* (2012), s. 110–120.
- [16] David Martínez i in. „Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit”. W: *Advances in Speech and Language Technologies for Iberian Languages*. Springer, 2012, s. 99–109.
- [17] Daria Panek, Andrzej Skalski i Janusz Gajda. „Quantification of linear and non-linear acoustic analysis applied to voice pathology detection”. W: *Information Technologies in Biomedicine 4.2014* (2014), s. 355–364.
- [18] Ahmed Al-nasheri i in. „Voice pathology detection with MDVP parameters using Arabic voice pathology database”. W: *Information Technology: Towards New Smart World (NSITNSW), 2015 5th National Symposium on*. IEEE. 2015, s. 1–5.
- [19] Pavol Harar i in. „Towards robust voice pathology detection”. W: *Neural Computing and Applications* (2018), s. 1–11.
- [20] Ahmed Al-Nasheri i in. „Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions”. W: *IEEE Access* 6 (2018), s. 6961–6974.
- [21] Ali Akbari i Meisam Khalil Arjmandi. „An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features”. W: *Biomedical Signal Processing and Control* 10 (2014), s. 209–223.
- [22] Alireza A Dibazar, Theodore W Berger i Shrikanth S Narayanan. „Pathological voice assessment”. W: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE. 2006, s. 1669–1673.
- [23] Ghulam Muhammad i in. „Automatic voice disorder classification using vowel formants”. W: *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE. 2011, s. 1–6.
- [24] Hossein Khadivi Heris, Babak Seyed Aghazadeh i Mansour Nikkhah-Bahrami. „Optimal feature selection for the assessment of vocal fold disorders”. W: *Computers in Biology and Medicine* 39.10 (2009), s. 860–868.
- [25] Pedro Gómez-Vilda i in. „Characterizing neurological disease from voice quality biomechanical analysis”. W: *Cognitive Computation* 5.4 (2013), s. 399–425.
- [26] J Ruz i in. „Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease”. W: *The Journal of the Acoustical Society of America* 129.1 (2011), s. 350–367.
- [27] Kamalini G Ranasinghe i in. „Abnormal vocal behavior predicts executive and memory deficits in Alzheimer’s disease”. W: *Neurobiology of Aging* 52 (2017), s. 71–80.
- [28] CJ Pérez i in. „Diagnosis and Tracking of Parkinson’s Disease by using Automatically Extracted Acoustic Features”. W: *J Alzheimers Dis Parkinsonism* 6.260 (2016), s. 2161–0460.
- [29] Juan Rafael Orozco-Arroyave i in. „New Spanish speech corpus database for the analysis of people suffering from Parkinson’s disease.” W: *LREC*. 2014, s. 342–347.

- [30] James Parkinson. „An essay on the shaking palsy”. W: *The Journal of Neuropsychiatry and Clinical Neurosciences* 14.2 (2002), s. 223–236.
- [31] Athanasios Tsanas i in. „Objective automatic assessment of rehabilitative speech treatment in Parkinson’s disease”. W: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.1 (2014), s. 181–190.
- [32] Félix Javier Jiménez-Jiménez i in. „Acoustic voice analysis in untreated patients with Parkinson’s disease”. W: *Parkinsonism & Related Disorders* 3.2 (1997), s. 111–116.
- [33] Javier Gamboa i in. „Acoustic voice analysis in patients with Parkinson’s disease treated with dopaminergic drugs”. W: *Journal of Voice* 11.3 (1997), s. 314–320.
- [34] Athanasios Tsanas i in. „Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson’s disease symptom severity”. W: *Journal of the Royal Society Interface* 8.59 (2011), s. 842–855.
- [35] J Ruz i in. „Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease”. W: *The Journal of the Acoustical Society of America* 129.1 (2011), s. 350–367.
- [36] Dávid Sztahó i Klára Vicsi. „Estimating the severity of Parkinson’s disease using voiced ratio and nonlinear parameters”. W: *International Conference on Statistical Language and Speech Processing*. Springer. 2016, s. 96–107.
- [37] *Anatomia i fizjologia narządu głosu (krtani)*. http://old.imp.lodz.pl/emisja_glosu/powstglos/index.html. [Online; dostęp 05-Lipiec-2017].
- [38] Wanda Kostecka. „Emisja głosu jako ważny przedmiot w kształceniu nauczycieli”. W: *Zeszyty Naukowe WSSP* 16 (2013).
- [39] Wioletta Pawlukowska, Krystyna Honczarenko i Monika Gołąb-Janowska. „Charakter zaburzeń mowy w chorobie Parkinsona”. W: *Neurologia i Neurochirurgia Polska* 47.3 (2013), s. 263–270.
- [40] Frederic L Darley, Arnold E Aronson i Joe R Brown. „Differential diagnostic patterns of dysarthria”. W: *Journal of Speech, Language, and Hearing Research* 12.2 (1969), s. 246–269.
- [41] M Śliwińska-Kowalska i in. „Rehabilitacja zawodowych zaburzeń głosu. Poradnik dla nauczycieli”. W: *Red. M. Śliwińska-Kowalska, E. Niebudek-Bogusz. Łódź, Instytut Medycyny Pracy* (2009).
- [42] Ewelina Woźnicka i in. „Ocena wartości manualnej terapii krtani w rehabilitacji głosu u pacjentów z dysfonią.” W: *Otorynolaryngologia* 15.1 (2016).
- [43] Ewa Niebudek-Bogusz i in. „Zastosowanie wskaźnika niepełnosprawności głosowej Voice Handical Index-VHI w ocenie efektywności terapii głosu nauczycieli”. W: *Medycyna Pracy* 58.6 (2007), s. 1–9.
- [44] *Choroby zawodowe narządu głosu - patofizjologia i obraz kliniczny*. <http://www.programyzdrowotne.pl/programypowrotudopracy/glos/Programrehabilitacjiglosu/Chorobyzawodowenarzdugosu/Default.aspx/>. [Online; dostęp 17-Marzec-2017].
- [45] Grzegorz Janczewski i Ewa Osuch Wójcikiewicz. *Ostry dyżur Otolaryngologia*. Alfa Medica Press, 2003.
- [46] Nikolay Sapundzhiev i in. „Surgery of adult bilateral vocal fold paralysis in adduction: history and trends”. W: *European Archives of Oto-Rhino-Laryngology* 265.12 (2008), s. 1501.
- [47] J Shvero i in. „Laser posterior ventriculocordectomy with partial arytenoidectomy for the treatment of bilateral vocal fold immobility”. W: *The Journal of Laryngology & Otology* 117.07 (2003), s. 540–543.

- [48] G Hope. „Laryngeal obstruction: removal of a portion of the vocal cord”. W: *Internationales Centrallblatt Laryngologie Rhinologie verwandte Wissenschaften* 12 (1896), s. 292.
- [49] JG Bizakis i in. „The combined endoscopic CO2 laser posterior cordectomy and total arytenoidectomy for treatment of bilateral vocal cord paralysis”. W: *Clinical Otolaryngology & Allied Sciences* 29.1 (2004), s. 51–54.
- [50] E Ferri i Purriños FJ García. „Diode laser surgery in the endoscopic treatment of laryngeal paralysis”. W: *Acta Otorrinolaringologica Espanola* 57.6 (2005), s. 270–274.
- [51] *Porażenie nerwu krtaniowego wstecznego*. <https://entokey.com/introduction-to-videostroboscopy/>. [Online; dostęp 05-Lipiec-2017].
- [52] *Nerw krtaniowy wsteczny*. http://www.wikiwand.com/pl/Nerw_krtaniowy_wsteczny/. [Online; dostęp 17-Marzec-2017].
- [53] *Ostre zapalenie krtani*. <http://www.wikiwand.com/en/Laryngitis/>. [Online; dostęp 17-Marzec-2017].
- [54] Lorraine O Ramig, Cynthia Fox i Shimon Sapir. „Speech treatment for Parkinson’s disease”. W: *Expert Review of Neurotherapeutics* 8.2 (2008), s. 297–309.
- [55] *Obszary kory mózgowej zaangażowane w planowanie i wykonywanie ruchu*. http://thebrain.mcgill.ca/flash/d/d_06/d_06_cr/d_06_cr_mou/d_06_cr_mou.html. [Online; dostęp 05-Lipiec-2017].
- [56] Jian Yang i Jing-yu Yang. „Why can LDA be performed in PCA transformed space?” W: *Pattern Recognition* 36.2 (2003), s. 563–566.
- [57] Tatiana Lewicka. „Diagnoza i terapia zaburzeń mowy u pacjentów z chorobą Parkinsona”. W: *Logopedia* 1.7 (2009), s. 116–25.
- [58] Lorraine Olson Ramig, Cynthia Fox i Shimon Sapir. „Speech disorders in Parkinson’s disease and the effects of pharmacological, surgical and speech treatment with emphasis on Lee Silverman voice treatment (LSVT®)”. W: *Handbook of Clinical Neurology* 83 (2007), s. 385–399.
- [59] Jeri A Logemann i in. „Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients”. W: *Journal of Speech and Hearing Disorders* 43.1 (1978), s. 47–57.
- [60] Kathe S Perez i in. „The Parkinson larynx: tremor and videostroboscopic findings”. W: *Journal of Voice* 10.4 (1996), s. 354–361.
- [61] Sabine Skodda, Wenke Visser i Uwe Schlegel. „Vowel articulation in Parkinson’s disease”. W: *Journal of Voice* 25.4 (2011), s. 467–472.
- [62] Margaret M Hoehn i Melvin D Yahr. „Parkinsonism onset, progression, and mortality”. W: *Neurology* 17.5 (1967), s. 427–427.
- [63] Monika Biercewicz i in. „Ocena stanu chorego z dysfunkcją układu pozapiramidowego—wybrane narzędzia pomiarowe w neurogeriatrii”. W: *Pielęgniarstwo Neurologiczne i Neurochirurgiczne* 3.2 (2014).
- [64] S Fahn i RL Elton. „Unified rating scale for Parkinson’s disease”. W: *Recent developments in Parkinson’s disease*. Florham Park, New York: Macmillan (1987), s. 153–163.
- [65] SteveEmilia Jadwiga Sitek Young. „Neuropsychologiczna ocena samoświadomości objawów w chorobach Huntingtona, Parkinsona i w dystonii sztywniej”. W: *Rozprawa Doktorska* (2010).
- [66] Christopher G Goetz i in. „Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results”. W: *Movement Disorders* 23.15 (2008), s. 2129–2170.

- [67] IR Titze. „Principles of Voice Production (National Center for Voice and Speech, Iowa City, IA)”. W: *Chap* 6 (2000), s. 149–184.
- [68] P Świdziński. „Obiektywna metoda akustycznego różnicowania zaburzeń głosu”. W: *Nowa Medycyna* 3.99 (2001), s. 24–27.
- [69] B Wilk. „Wyznaczanie wartości chwilowej częstotliwości podstawowej tonu krtaniowego za pomocą analizy falkowej sygnału mowy”. W: *Przegląd Elektrotechniczny* 91 (2015), s. 305–308.
- [70] Tukey J Bogert B. P. Healy M. J.R. „The quefrency analysis of time series for echoes: Cepstrum, pseudo-auto-covariance, cross-cepstrum and saphe cracking”. W: *Proceedings of the Symposium on Time Series Analysis* 15 (1963), 209—243.
- [71] Meike Brockmann i in. „Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task”. W: *Journal of Voice* 25.1 (2011), s. 44–53.
- [72] Tomasz Piotr Zieliński. *Cyfrowe przetwarzanie sygnałów: od teorii do zastosowań*. Wydawnictwa Komunikacji i Łączności, 2007, s. 6.
- [73] Zbigniew Witold Engel, Maciej Kłaczyński i Wiesław Wszolek. „A vibroacoustic model of selected human larynx diseases”. W: *International Journal of Occupational Safety and Ergonomics* 13.4 (2007), s. 367–379.
- [74] Maciej Kłaczyński. „Zjawiska wibroakustyczne w kanale głosowym człowieka”. W: *Rozprawa Doktorska* (2007).
- [75] Dennis H Klatt. „Software for a cascade/parallel formant synthesizer”. W: *The Journal of the Acoustical Society of America* 67.3 (1980), s. 971–995.
- [76] BCJ Moore. *Wprowadzenie do psychologii słyszenia*. 1999.
- [77] Jacob Benesty, M Mohan Sondhi i Yiteng Huang. *Springer Handbook of Speech Processing*. Springer Science & Business Media, 2007.
- [78] Dieter Maurer. *Acoustics of the Vowel-Preliminaries*. Peter Lang LTD International Academic Publishers, 2016.
- [79] Matthias Patzold i Adrian P Simpson. „Acoustic analysis of German vowels in the Kiel Corpus of Read Speech”. W: *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel* 32 (1997), s. 215–247.
- [80] Steve Young. „A review of large-vocabulary continuous-speech”. W: *IEEE Signal Processing Magazine* 13.5 (1996), s. 45.
- [81] Veena S Chakravarthi, Yasha Jyothi M Shirur i P Rekha. *Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking (VCASAN-2013)*. T. 258. Springer Science & Business Media, 2013.
- [82] Steven Davis i Paul Mermelstein. „Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”. W: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28.4 (1980), s. 357–366.
- [83] Romuald Mazurek. „Przestrzenno-czasowe rozkłady pola akustycznego zespołów źródeł szerokopasmowych i ich wpływ na zniekształcenia sygnałów mowy”. W: *Rozprawa Doktorska* (2014).

- [84] Fayek Haytham. *Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between*. <http://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>. [Online; dostęp 17-Marzec-2017].
- [85] Yannis Stylianou, Marcos Faundez-Zanuy i Anna Eposito. *Progress in nonlinear speech processing*. T. 4391. Springer, 2007, s. 23–28.
- [86] Dawn Black. *Emotion Detection, Queen Mary, University of London*. <https://code.soundsoftware.ac.uk/users/87>. [Online; dostęp 01-Listopad-2015].
- [87] Bernhard Scholkopf. „The kernel trick for distances”. W: *Advances in Neural Information Processing Systems* (2001), s. 301–307.
- [88] Bernhard Schölkopf, Alexander Smola i Klaus-Robert Müller. „Nonlinear component analysis as a kernel eigenvalue problem”. W: *Neural Computation* 10.5 (1998), s. 1299–1319.
- [89] Quan Wang. „Kernel principal component analysis and its applications in face recognition and active shape models”. W: *arXiv preprint arXiv:1207.3538* (2012).
- [90] Wei-Yin Loh i Yu-Shan Shih. „Split selection methods for classification trees”. W: *Statistica Sinica* (1997), s. 815–840.
- [91] *Randomforest-matlab: RandomForest(Regression, Classification and Clustering) Implementation for Matlab(and Standalone)*. <https://code.google.com/archive/p/randomforest-matlab/>. [Online; dostęp 2014].
- [92] Pall Oskar Gislason, Jon Atli Benediktsson i Johannes R Sveinsson. „Random forests for land cover classification”. W: *Pattern Recognition Letters* 27.4 (2006), s. 294–300.
- [93] Leo Breiman i in. *Classification and Regression Trees*. CRC press, 1984.
- [94] Matthias Scholz i Ricardo Vigário. „Nonlinear PCA: a new hierarchical approach.” W: *ESANN*. 2002, s. 439–444.
- [95] Matthias Scholz i in. „Non-linear PCA: a missing data approach”. W: *Bioinformatics* 21.20 (2005), s. 3887–3895.
- [96] Matthias Scholz, Martin Fraunholz i Joachim Selbig. „Nonlinear principal component analysis: neural network models and applications”. W: *Principal manifolds for data visualization and dimension reduction*. Springer, 2008, s. 44–67.
- [97] Magnus Rudolph Hestenes i Eduard Stiefel. *Methods of conjugate gradients for solving linear systems*. T. 49. 1. NBS, 1952.
- [98] Christopher M Bishop. „Pattern recognition”. W: *Machine Learning* 128 (2006), s. 1–58.
- [99] Jarosław Sławek i Andrzej Bogucki. „Dojelitowa postać lewodopy w leczeniu zaawansowanej choroby Parkinsona”. W: *Neurologia i Neurochirurgia Polska* 44.4 (2010), s. 396–403.
- [100] Jarosław Sławek. „Lewodopa w leczeniu choroby Parkinsona–wczoraj i dziś”. W: *Neurologia i Neurochirurgia Polska* 46.1 (2012), s. 63–75.
- [101] Jarosław Sławek. „Choroba Parkinsona—jak właściwie rozpoznawać, skutecznie i bezpiecznie leczyć?” W: *Forum Medycyny Rodzinnej*. T. 8. 6. 2014, s. 281–291.
- [102] Marcin Wolter. *Metody analizy wielu zmiennych w fizyce wysokich energii*. Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego Polskiej Akademii Nauk, 2012.

- [103] Marti A. Hearst i in. „Support vector machines”. W: *IEEE Intelligent Systems and Their Applications* 13.4 (1998), s. 18–28.
- [104] Janusz Rusek. „Zastosowanie metody Support Vector Machine w analizie stanu technicznego zabudowy terenu górniczego LGOM”. W: *Eksploatacja i Niezawodność* 19.1 (2017), s. 54.
- [105] Alex J Smola i Bernhard Scholkopf. „A tutorial on support vector regression”. W: *Statistics and computing* 14.3 (2004), s. 199–222.
- [106] William Karush. „Minima of functions of several variables with inequalities as side conditions”. W: *Master thesis, University of Chicago* (1939).
- [107] Harold W Kuhn, Albert W Tucker i in. „Nonlinear programming”. W: *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. The Regents of the University of California. 1951.
- [108] Paul John Werbos. „Beyond regression: New tools for prediction and analysis in the behavioral sciences”. W: *Doctoral Dissertation, Applied Mathematics, Harvard University, MA* (1974).
- [109] David E Rumelhart, Geoffrey E Hinton i Ronald J Williams. *Learning internal representations by error propagation*. Spraw. tech. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [110] Martin T Hagan i Mohammad B Menhaj. „Training feedforward networks with the Marquardt algorithm”. W: *IEEE transactions on Neural Networks* 5.6 (1994), s. 989–993.
- [111] Donald W Marquardt. „An algorithm for least-squares estimation of nonlinear parameters”. W: *Journal of the Society for Industrial and Applied Mathematics* 11.2 (1963), s. 431–441.
- [112] Bernhard Scholkopf, Alexander Smola i Klaus-Robert Muller. „Kernel principal component analysis”. W: *International Conference on Artificial Neural Networks*. Springer. 1997, s. 583–588.
- [113] Daria Hemmerling, Andrzej Skalski i Janusz Gajda. „Voice data mining for laryngeal pathology assesment”. W: *Computers in Biology and Medicine* 69 (2016), s. 270–276.
- [114] Daria Hemmerling. „Voice pathology distinction using autoassociative neural networks”. W: *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE. 2017, s. 1844–1847.
- [115] *Przydycja, Encyklopedia PWN*. <https://encyklopedia.pwn.pl/haslo/przydycja;3962016.html>. [Online; dostęp 27-Luty-2018].
- [116] Tadeusiewicz Ryszard i in. *Sieci neuronowe w inżynierii biomedycznej*. Akademicka Oficyna Wydawnicza EXIT, 2013.
- [117] *Sieci Neuronowe, Matlab*. <https://www.mathworks.com/help/nnet/gs/fit-data-with-a-neural-network.html>. [Online; dostęp 13-Maj-2018].
- [118] Daria Panek i in. „Acoustic analysis assessment in speech pathology detection”. W: *International Journal of Applied Mathematics and Computer Science* 25.3 (2015), s. 631–643.
- [119] Daria PANEK, Andrzej SKALSKI i Janusz GAJDA. „Wpływ długości fonacji na ilość informacji zawartej w sygnale głosu ludzkiego”. W: *Przegląd Elektrotechniczny* 91.5 (2015), s. 57–59.
- [120] Daria Panek, Andrzej Skalski i Janusz Gajda. „Voice pathology detection by fuzzy logic”. W: *Instrumentation and Measurement Technology Conference (I2MTC), 2015 IEEE International*. IEEE. 2015, s. 289–293.
- [121] Daria Hemmerling i in. „Automatic Detection of Parkinson’s Disease Based on Modulated Vowels.” W: *InterSpeech*. 2016, s. 1190–1194.

A. Dodatek A

Tabela A.1. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /a/, intonacja wysoka, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	79,38	13,75	6,88	0,00
	lar	12,50	82,14	5,36	0,00
	par	23,53	11,03	65,44	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.2. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /a/, intonacja normalna, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	76,25	17,50	6,25	0,00
	lar	17,86	73,21	8,93	0,00
	par	18,38	2,94	78,68	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.3. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /a/, intonacja niska, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	76,88	15,00	8,13	0,00
	lar	17,86	73,21	8,93	0,00
	par	22,79	10,29	66,91	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.4. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /a/, intonacja wysoka, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	62,50	25,00	12,50	0,00
	lar	13,75	55,00	31,25	0,00
	par	9,72	5,56	84,72	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.5. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /a/, intonacja normalna, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	70,00	7,50	22,50	0,00
	lar	15,00	75,00	10,00	0,00
	par	13,89	9,72	76,39	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.6. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /a/, intonacja niska, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	55,00	35,00	10,00	0,00
	lar	13,75	52,50	33,75	0,00
	par	11,11	6,94	81,94	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.7. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /i/, intonacja wysoka, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	74,38	13,75	11,88	0,00
	lar	17,86	75,00	7,14	0,00
	par	22,06	9,56	68,38	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.8. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /i/, intonacja normalna, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	82,50	9,38	8,13	0,00
	lar	17,86	75,00	7,14	0,00
	par	21,32	11,76	66,91	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.9. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /i/, intonacja niska, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	85,00	8,75	6,25	0,00
	lar	12,50	80,36	7,14	0,00
	par	28,68	12,50	58,82	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.10. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /i/, intonacja wysoka, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	75,00	17,50	7,50	0,00
	lar	26,25	51,25	22,50	0,00
	par	11,11	8,33	80,56	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.11. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /i/, intonacja normalna, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	77,50	10,00	12,50	0,00
	lar	18,75	50,00	31,25	0,00
	par	11,11	8,33	80,56	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.12. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /i/, intonacja niska, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	67,50	20,00	15,00	0,00
	lar	21,25	47,50	31,25	0,00
	par	9,72	11,11	79,17	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.13. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /u/, intonacja wysoka, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	78,75	13,75	7,50	0,00
	lar	14,29	82,14	3,57	0,00
	par	21,32	11,03	67,65	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.14. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /u/, intonacja normalna, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	78,75	13,75	7,50	0,00
	lar	14,29	82,14	3,57	0,00
	par	21,32	11,03	67,65	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.15. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /u/, intonacja niska, grupa kobiet.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	81,88	7,50	10,63	0,00
	lar	17,86	75,00	7,14	0,00
	par	16,91	7,35	75,74	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.16. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /u/, intonacja wysoka, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	65,00	15,00	20,00	0,00
	lar	23,75	51,25	25,00	0,00
	par	11,11	13,89	75,00	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.17. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /u/, intonacja normalna, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	62,50	22,50	15,00	0,00
	lar	21,25	50,00	28,75	0,00
	par	15,28	4,17	80,56	0,00
	zdrowi	0,00	0,00	0,00	100,00

Tabela A.18. Macierz pomyłek klasyfikacji wybranych chorób laryngologicznych [%], samogłoska /u/, intonacja niska, grupa mężczyzn.

		Klasyfikacja			
		HF	lar	par	zdrowi
Diagnoza	HF	67,50	25,00	7,50	0,00
	lar	20,00	50,00	30,00	0,00
	par	8,33	9,72	81,94	0,00
	zdrowi	0,00	0,00	0,00	100,00

B. Dodatek B

Tabela B.1. Parametry sieci neuronowej dla poszczególnych parametrów: 1 - częstotliwość podstawowa, 2 - współczynnik jitter, 3 - ppq5, 4 - ddp, 5 - współczynnik Shimmer, 6 - apq3.

parametr	1					2				
	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
parametr										
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów warstw ukrytych	13	6	14	6	16	9	10	7	8	8
Liczba neuronów 2 warstwy ukrytej	18	25	20	33	29	30	32	23	16	25
funkcja aktywacji 1 warstwy	logsig	tansig	logsig	tansig	logsig	tansig	logsig	logsig	logsig	purelm
funkcja aktywacji 2 warstwy	logsig	logsig	logsig	purelm	logsig	purelm	logsig	logsig	logsig	tansig
funkcja aktywacji 3 warstwy	logsig	logsig	logsig	logsig	tansig	logsig	logsig	logsig	logsig	logsig
metoda uczenia	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm
parametr										
parametr										
samogłoska	3					4				
parametr										
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów 1 warstwy ukrytej	16	8	8	11	13	18	11	14	14	12
Liczba neuronów 2 warstwy ukrytej	30	30	33	14	15	36	34	21	29	20
funkcja aktywacji 1 warstwy	tansig	tansig	tansig	purelm	purelm	tansig	logsig	logsig	logsig	logsig
funkcja aktywacji 2 warstwy	logsig	purelm	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig
funkcja aktywacji 3 warstwy	logsig	logsig	logsig	logsig	tansig	logsig	purelm	tansig	logsig	logsig
metoda uczenia	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm
parametr										
parametr										
samogłoska	5					6				
parametr										
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów 1 warstwy ukrytej	16	16	10	15	13	8	6	14	16	10
Liczba neuronów 2 warstwy ukrytej	22	18	35	31	34	16	17	26	29	34
funkcja aktywacji 1 warstwy	purelm	logsig	logsig	logsig	tansig	logsig	logsig	logsig	purelm	logsig
funkcja aktywacji 2 warstwy	logsig	logsig	logsig	logsig	logsig	logsig	purelm	logsig	logsig	logsig
funkcja aktywacji 3 warstwy	tansig	purelm	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig
metoda uczenia	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm

Tabela B.2. Parametry sieci neuronowej dla poszczególnych parametrów: 7 - apq5, 8 - HNR, 9 - energia, 10 - 0-moment widmowy, 11 - 1-wszy moment widmowy, 12 - 2-gi moment widmowy.

parametr	7					8				
	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
parametr	7					8				
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów 1 warstwy ukrytej	10	15	10	8	11	10	7	6	8	14
Liczba neuronów 2 warstwy ukrytej	22	33	18	11	16	17	20	21	21	33
funkcja aktywacji 1 warstwy	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig
funkcja aktywacji 2 warstwy	logsig	logsig	logsig	purelin	logsig	logsig	logsig	logsig	logsig	logsig
funkcja aktywacji 3 warstwy	logsig	purelin	logsig	logsig	logsig	logsig	logsig	logsig	purelin	logsig
metoda uczenia	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg
parametr	9					10				
parametr	9					10				
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów 1 warstwy ukrytej	12	10	13	12	11	7	16	8	14	12
Liczba neuronów 2 warstwy ukrytej	21	16	16	25	14	12	29	20	31	18
funkcja aktywacji 1 warstwy	logsig	logsig	logsig	purelin	logsig	logsig	purelin	logsig	logsig	logsig
funkcja aktywacji 2 warstwy	purelin	logsig	logsig	logsig	logsig	logsig	purelin	purelin	logsig	logsig
funkcja aktywacji 3 warstwy	logsig	logsig	logsig	logsig	logsig	purelin	logsig	logsig	logsig	logsig
metoda uczenia	trainbfg	trainbfg	trainbfg	trainbfg	trainbfg	trainlm	trainlm	trainlm	trainlm	trainlm
parametr	11					12				
parametr	11					12				
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów 1 warstwy ukrytej	12	7	6	17	15	15	10	12	18	10
Liczba neuronów 2 warstwy ukrytej	25	25	10	25	26	29	18	18	21	26
funkcja aktywacji 1 warstwy	logsig	purelin	logsig	purelin	logsig	logsig	logsig	logsig	purelin	logsig
funkcja aktywacji 2 warstwy	purelin	logsig	logsig	logsig	logsig	purelin	purelin	logsig	logsig	logsig
funkcja aktywacji 3 warstwy	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig	logsig
metoda uczenia	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm

Tabela B.6. Parametry sieci neuronowej dla poszczególnych parametrów: 30-33 - od 10 do 12 współczynnika MFCC.

parametr	31					32				
	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
Liczba neuronów 1 warstwy ukrytej	8	7	9	10	14	10	15	8	14	18
Liczba neuronów 2 warstwy ukrytej	12	10	33	18	32	12	32	25	25	27
funkcja aktywacji 1 warstwy	purelin	logsig	logsig	logsig	logsig	logsig	logsig	logsig	tansig	logsig
funkcja aktywacji 2 warstwy	logsig	logsig	logsig	logsig	logsig	logsig	tansig	logsig	purelin	logsig
funkcja aktywacji 3 warstwy	logsig	purelin	tansig	logsig	logsig	logsig	logsig	purelin	logsig	logsig
metoda uczenia	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm
parametr	33									
parametr	/a/	/e/	/i/	/o/	/u/					
Liczba neuronów 1 warstwy ukrytej	10	16	13	10	18					
Liczba neuronów 2 warstwy ukrytej	15	29	29	24	27					
funkcja aktywacji 1 warstwy	logsig	logsig	logsig	logsig	logsig					
funkcja aktywacji 2 warstwy	logsig	logsig	logsig	logsig	logsig					
funkcja aktywacji 3 warstwy	logsig	purelin	logsig	logsig	logsig					
metoda uczenia	trainlm	trainbfg	trainlm	trainbfg	trainbfg					

C. Dodatek C

Tabela C.1. Zestawienie wyników predykcji dla parametrów od 1- częstotliwość podstawowa, 2- współczynnik jitter, 3- pięciopunktowy współczynnik zaburzeń okresu, 4- jedenastopunktowy współczynnik zaburzeń okresu, 5- współczynnik shimmer, 6- trójpunktowy współczynnik zaburzeń amplitudy, z wykorzystaniem sieci neuronowej dla każdego z analizowanych parametrów, MSE_{tr} - błąd średniokwadratowy dla zbioru uczącego, MSE_{val} - błąd średniokwadratowy dla zbioru walidacyjnego, MSE_{test} - błąd średniokwadratowy dla zbioru testowego, R - współczynnik korelacji dla wszystkich zbiorów.

parametr	1					2				
	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
MSE _{tr}	0,05±0,06	0,06±0,02	0,04±0,02	0,07±0,04	0,07±0,08	0,04±0,05	0,02±0,11	0,03±0,06	0,06±0,05	0,03±0,01
MSE _{val}	0,08±0,07	0,06±0,01	0,05±0,07	0,08±0,07	0,09±0,07	0,07±0,06	0,08±0,03	0,06±0,06	0,06±0,06	0,05±0,04
MSE _{test}	0,08±0,07	0,09±0,01	0,07±0,06	0,09±0,07	0,12±0,10	0,10±0,17	0,09±0,09	0,08±0,09	0,07±0,07	0,05±0,04
R	0,99±0,02	0,98±0,04	0,96±0,08	0,91±0,25	0,96±0,08	0,86±0,31	0,98±0,03	0,86±0,21	0,89±0,30	0,92±0,20
parametr	3					4				
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
MSE _{tr}	0,08±0,05	0,02±0,06	0,05±0,04	0,06±0,01	0,02±0,03	0,07±0,04	0,06±0,04	0,07±0,06	0,05±0,06	0,04±0,09
MSE _{val}	0,09±0,09	0,09±0,02	0,07±0,06	0,06±0,08	0,07±0,08	0,08±0,08	0,06±0,08	0,07±0,06	0,05±0,07	0,05±0,06
MSE _{test}	0,09±0,09	0,09±0,15	0,08±0,10	0,08±0,09	0,08±0,06	0,14±0,16	0,14±0,17	0,15±0,12	0,17±0,26	0,10±0,11
R	0,95±0,08	0,97±0,06	0,98±0,04	0,81±0,23	0,72±0,26	0,96±0,05	0,97±0,04	0,98±0,03	0,96±0,10	0,98±0,03
parametr	5					6				
samogłoska	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
MSE _{tr}	0,06±0,03	0,05±0,05	0,03±0,15	0,06±0,30	0,04±0,03	0,04±0,06	0,05±0,07	0,06±0,18	0,07±0,11	0,05±0,15
MSE _{val}	0,06±0,10	0,05±0,10	0,06±0,08	0,06±0,07	0,05±0,07	0,05±0,08	0,07±0,05	0,06±0,06	0,08±0,08	0,07±0,07
MSE _{test}	0,11±0,12	0,1±0,10	0,11±0,10	0,24±0,26	0,07±0,08	0,11±0,10	0,13±0,11	0,11±0,15	0,09±0,07	0,12±0,09
R	0,98±0,03	0,99±0,01	0,96±0,07	0,92±0,14	0,99±0,01	0,99±0,02	0,97±0,06	0,99±0,01	0,99±0,01	0,99±0,02

Tabela C.3. Zestawienie wyników predykcji dla parametrów 13- kurtoza, 14- współczynnik mocy względnej, 15 - amplituda 1-formantu, 16- amplituda 2-formantu, 17- amplituda 3-formantu, 18- częstotliwość 1-formantu, MSE_{tr} - błąd średniokwadratowy dla zbioru uczącego, MSE_{val} - błąd średniokwadratowy dla zbioru walidacyjnego, MSE_{test} - błąd średniokwadratowy dla zbioru testowego, R - współczynnik korelacji dla wszystkich zbiorów.

parametr	13						14					
	/a/	/e/	/i/	/o/	/u/	/w/	/a/	/e/	/i/	/o/	/u/	/w/
MSE _{tr}	0,04±0,06	0,03±0,06	0,02±0,06	0,04±0,07	0,04±0,06	0,04±0,06	0,05±0,04	0,03±0,06	0,05±0,01	0,05±0,05	0,03±0,04	0,03±0,04
MSE _{val}	0,06±0,03	0,10±0,09	0,03±0,05	0,07±0,04	0,08±0,10	0,08±0,10	0,05±0,04	0,05±0,06	0,05±0,05	0,05±0,06	0,04±0,40	0,04±0,40
MSE _{test}	0,07±0,07	0,12±0,15	0,03±0,07	0,08±0,07	0,12±0,2	0,12±0,2	0,08±0,14	0,06±0,08	0,09±0,01	0,06±0,05	0,07±0,07	0,07±0,07
R	0,79±0,60	0,91±0,27	0,89±0,23	0,87±0,24	0,70±0,66	0,70±0,66	0,92±0,08	0,98±0,03	0,84±0,23	0,85±0,31	0,94±0,13	0,94±0,13
parametr	15						16					
samogłoska	/a/	/e/	/i/	/o/	/u/	/w/	/a/	/e/	/i/	/o/	/u/	/w/
MSE _{tr}	0,06±0,04	0,03±0,05	0,04±0,05	0,06±0,08	0,04±0,05	0,04±0,05	0,02±0,05	0,08±0,07	0,09±0,09	0,06±0,06	0,05±0,06	0,05±0,06
MSE _{val}	0,05±0,06	0,06±0,03	0,07±0,04	0,08±0,05	0,04±0,01	0,04±0,01	0,03±0,02	0,13±0,12	0,14±0,07	0,08±0,06	0,06±0,04	0,06±0,04
MSE _{test}	0,08±0,09	0,10±0,11	0,08±0,08	0,07±0,08	0,05±0,05	0,05±0,05	0,04±0,06	0,13±0,10	0,16±0,21	0,13±0,17	0,07±0,06	0,07±0,06
R	0,97±0,06	0,96±0,06	0,98±0,01	0,78±0,60	0,98±0,01	0,98±0,01	0,98±0,04	0,82±0,23	0,97±0,04	0,98±0,03	0,91±0,21	0,91±0,21
parametr	17						18					
samogłoska	/a/	/e/	/i/	/o/	/u/	/w/	/a/	/e/	/i/	/o/	/u/	/w/
MSE _{tr}	0,06±0,08	0,06±0,06	0,04±0,07	0,04±0,06	0,05±0,09	0,05±0,09	0,01±0,03	0,1±0,10	0,05±0,06	0,05±0,07	0,04±0,10	0,04±0,10
MSE _{val}	0,08±0,10	0,13±0,10	0,06±0,03	0,07±0,03	0,13±0,11	0,13±0,11	0,13±0,16	0,11±0,09	0,08±0,08	0,11±0,09	0,02±0,03	0,02±0,03
MSE _{test}	0,15±0,16	0,18±0,23	0,06±0,08	0,10±0,10	0,11±0,11	0,11±0,11	0,10±0,10	0,14±0,11	0,13±0,17	0,07±0,08	0,05±0,10	0,05±0,10
R	0,97±0,08	0,83±0,37	0,78±0,76	0,68±0,66	0,96±0,06	0,96±0,06	0,78±0,67	0,93±0,15	0,91±0,11	0,97±0,06	0,99±0,01	0,99±0,01

Tabela C.4. Zestawienie wyników predykcji dla parametrów 19- częstość 2-formantu, 20- częstość 3-formantu [21-24] - 1:4 współczynnik mel-cepstralny; MSE_{tr} - błąd średniokwadratowy dla zbioru uczącego, MSE_{val} - błąd średniokwadratowy dla zbioru walidacyjnego, MSE_{test} - błąd średniokwadratowy dla zbioru testowego, R - współczynnik korelacji dla wszystkich zbiorów.

parametr	19						20					
	/a/	/e/	/i/	/o/	/u/	/y/	/a/	/e/	/i/	/o/	/u/	
MSE _{tr}	0,10±0,09	0,07±0,04	0,07±0,03	0,13±0,11	0,07±0,02	0,11±0,09	0,08±0,08	0,04±0,02	0,12±0,14	0,07±0,11		
MSE _{val}	0,04±0,08	0,04±0,07	0,06±0,07	0,04±0,05	0,06±0,08	0,05±0,05	0,04±0,05	0,03±0,03	0,04±0,05	0,01±0,02		
MSE _{test}	0,13±0,14	0,09±0,10	0,10±0,09	0,15±0,02	0,09±0,06	0,12±0,16	0,08±0,06	0,05±0,04	0,17±0,23	0,02±0,02		
R	0,99±0,04	0,92±0,15	0,98±0,02	0,96±0,08	0,97±0,04	0,96±0,06	0,97±0,04	0,96±0,07	0,77±0,38	0,98±0,03		
parametr	21						22					
samogłoska	/a/	/e/	/i/	/o/	/u/	/y/	/a/	/e/	/i/	/o/	/u/	
MSE _{tr}	0,05±0,05	0,05±0,06	0,04±0,05	0,03±0,05	0,01±0,01	0,04±0,01	0,07±0,02	0,07±0,09	0,05±0,03	0,07±0,05		
MSE _{val}	0,07±0,07	0,09±0,10	0,06±0,06	0,11±0,14	0,07±0,09	0,05±0,07	0,09±0,07	0,04±0,04	0,10±0,12	0,09±0,08		
MSE _{test}	0,04±0,06	0,13±0,10	0,16±0,21	0,13±0,17	0,07±0,06	0,05±0,07	0,09±0,07	0,04±0,04	0,10±0,12	0,09±0,08		
R	0,91±0,23	0,96±0,08	0,98±0,04	0,98±0,02	0,99±0,02	0,99±0,02	0,80±0,34	0,97±0,05	0,93±0,11	0,92±0,09		
parametr	23						24					
samogłoska	/a/	/e/	/i/	/o/	/u/	/y/	/a/	/e/	/i/	/o/	/u/	
MSE _{tr}	0,06±0,08	0,12±0,22	0,05±0,05	0,05±0,06	0,04±0,08	0,03±0,05	0,06±0,06	0,01±0,01	0,05±0,07	0,05±0,06		
MSE _{val}	0,05±0,01	0,13±0,17	0,06±0,02	0,07±0,04	0,10±0,09	0,11±0,13	0,05±0,02	0,08±0,14	0,09±0,09	0,05±0,01		
MSE _{test}	0,07±0,06	0,11±0,10	0,11±0,11	0,05±0,05	0,09±0,11	0,09±0,11	0,08±0,09	0,09±0,14	0,17±0,21	0,07±0,07		
R	0,80±0,41	0,93±0,1	0,72±0,61	0,91±0,19	0,53±0,79	0,93±0,15	0,95±0,10	0,92±0,13	0,90±0,19	0,84±0,44		

Tabela C.5. Zestawienie wyników predykcji dla parametrów [25-30] - 5:10 współczynnik mel-cepstralny, MSE_{tr} - błąd średniokwadratowy dla zbioru uczącego, MSE_{val} - błąd średniokwadratowy dla zbioru walidacyjnego, MSE_{test} - błąd średniokwadratowy dla zbioru testowego, R - współczynnik korelacji dla wszystkich zbiorów.

parametr	25						26					
	/a/	/e/	/i/	/o/	/u/		/a/	/e/	/i/	/o/	/u/	
MSE_{tr}	0,06±0,06	0,04±0,04	0,02±0,02	0,05±0,06	0,04±0,05		0,02±0,02	0,05±0,07	0,04±0,05	0,04±0,07	0,02±0,03	
MSE_{val}	0,11±0,08	0,10±0,07	0,03±0,03	0,05±0,02	0,05±0,02		0,13±0,18	0,13±0,13	0,08±0,05	0,05±0,02	0,06±0,06	
MSE_{test}	0,16±0,16	0,13±0,13	0,06±0,07	0,06±0,07	0,06±0,07		0,13±0,13	0,06±0,04	0,07±0,06	0,07±0,09	0,07±0,08	
R	0,98±0,03	0,98±0,02	0,99±0,02	0,93±0,12	0,92±0,09		0,71±0,60	0,92±0,10	0,87±0,22	0,76±0,66	0,82±0,18	
parametr	27						28					
samogłoska	/a/	/e/	/i/	/o/	/u/		/a/	/e/	/i/	/o/	/u/	
MSE_{tr}	0,05±0,05	0,05±0,06	0,04±0,05	0,03±0,05	0,01±0,01		0,04±0,01	0,07±0,02	0,07±0,09	0,05±0,03	0,07±0,05	
MSE_{val}	0,07±0,07	0,09±0,1	0,06±0,06	0,11±0,14	0,07±0,09		0,05±0,07	0,09±0,07	0,04±0,04	0,10±0,12	0,09±0,08	
MSE_{test}	0,04±0,06	0,13±0,01	0,16±0,21	0,13±0,17	0,07±0,06							
R	0,91±0,23	0,96±0,08	0,98±0,04	0,98±0,02	0,99±0,02		0,99±0,02	0,80±0,34	0,97±0,05	0,93±0,11	0,92±0,09	
parametr	29						30					
samogłoska	/a/	/e/	/i/	/o/	/u/		/a/	/e/	/i/	/o/	/u/	
MSE_{tr}	0,09±0,08	0,07±0,06	0,03±0,08	0,03±0,02	0,06±0,07		0,03±0,06	0,04±0,07	0,04±0,07	0,03±0,03	0,02±0,05	
MSE_{val}	0,12±0,07	0,07±0,02	0,15±0,15	0,08±0,05	0,07±0,05		0,05±0,03	0,07±0,04	0,05±0,03	0,06±0,04	0,08±0,09	
MSE_{test}	0,09±0,08	0,10±0,08	0,15±0,19	0,10±0,08	0,14±0,12		0,07±0,09	0,08±0,07	0,06±0,07	0,05±0,03	0,08±0,09	
R	0,96±0,04	0,92±0,12	0,96±0,06	0,87±0,2	0,88±0,24		0,78±0,58	0,97±0,04	0,99±0,03	0,96±0,06	0,86±0,28	

Tabela C.6. Zestawienie wyników predykcji dla parametrów [31-32] - 11:12 współczynnik mel-cepstralny, 33- stosunek harmonicznych do szumu, MSE t_{tr} - błąd średniokwadratowy dla zbioru uczącego, MSE v_{val} - błąd średniokwadratowy dla zbioru walidacyjnego, MSE t_{test} - błąd średniokwadratowy dla zbioru testowego, R - współczynnik korelacji dla wszystkich zbiorów.

parametr	31					32				
	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
parametr samogłoska										
MSE t_{tr}	0,03±0,04	0,06±0,08	0,07±0,11	0,03±0,03	0,05±0,05	0,05±0,08	0,01±0,02	0,08±0,11	0,06±0,08	0,04±0,07
MSE v_{val}	0,12±0,13	0,05±0,03	0,12±0,13	0,04±0,01	0,06±0,04	0,09±0,03	0,04±0,07	0,06±0,03	0,13±0,07	0,07±0,06
MSE t_{test}	0,13±0,21	0,09±0,11	0,09±0,11	0,05±0,05	0,12±0,12	0,11±0,15	0,03±0,03	0,09±0,09	0,10±0,09	0,06±0,08
R	0,94±0,09	0,84±0,37	0,84±0,29	0,97±0,05	0,81±0,11	0,89±0,13	0,99±0,02	0,92±0,20	0,85±0,36	0,76±0,46
parametr samogłoska	33									
MSE t_{tr}	0,07±0,08	0,08±0,11	0,04±0,07	0,07±0,08	0,08±0,10					
MSE v_{val}	0,08±0,04	0,12±0,1	0,07±0,04	0,10±0,08	0,15±0,10					
MSE t_{test}	0,12±0,11	0,14±0,12	0,09±0,14	0,15±0,18	0,20±0,20					
R	0,72±0,41	0,80±0,31	0,83±0,37	0,79±0,42	0,83±0,26					

Oświadczenie

Wykonałam pracę samodzielnie, nie korzystałam z innych źródeł niż te wymienione w niniejszej rozprawie.