

Tomasz Orzechowski*, Katarzyna Chmurzyńska*, Piotr Radkowski*

Automatyczna ocena zaburzeń emisji głosu będących wynikiem procesów neurodegeneracyjnych w oparciu o analizę wyizolowanych głosek**

1. Wprowadzenie

Prezentowane w niniejszej publikacji prace są pierwszymi z planowanych kompleksowych, interdyscyplinarnych badań, którym poddani byli pacjenci kliniki CM-UJ Kraków, z chorobą neurodegeneracyjną (uszkodzenie układu pozapiramidowego) ze zmianami w mowie o typie dyzartrii. Przeprowadzono także badania kontrolne u osób zdrowych. Opierając się na wiedzy lingwistycznej z zakresu fonetyki oraz doświadczeniach wynikających z wieloletniej pracy patologa mowy, wybrano elementy, których realizacje badano. Materiał lingwistyczny został celowo dobrany tak, aby wychwycić zmiany głosu charakterystyczne dla pacjentów z dyzartrią. W niniejszym artykule prezentowane będą zjawiska zarejestrowane w izolowanych dźwiękach. W celu przeprowadzenia dalszych analiz w czasie badania rejestrowano także frazy oparte na idiomach polskich.

2. Fizjologia głosu

Wytwarzanie głosu i mowy wymaga ścisłego współdziałania ze sobą wielu narządów, które z foniatrycznego punktu widzenia można by podzielić na narządy [1]:

- wytwarzające niezbędny dla fonacji strumień powietrza wydechowego (płuca, oskrzela, tchawica);
- wzmacniające ton podstawowy (krtań);
- formujące barwę głosu i tworzące głoski mowy (nasada, gardło, jama nosowa, jama ustna).

3. Patologia głosu

Obok typowych zmian wynikających z choroby neurodegeneracyjnej (np. drżenie ciała, kończyn, sztywność mięśni) obserwuje się zmiany w głosie. Badania przedstawione w pra-

* Katedra Automatyki, Akademia Górniczo-Hutnicza w Krakowie

** Pracę wykonano w ramach badań własnych (umowa AGH nr 10.10.120.39)

cy [4] wskazują na poważny problem występowania patologii mowy u aż 75% pacjentów. Można zatem twierdzić, że głos stanowi jeden z ważniejszych komponentów diagnozy neurologicznej, co również zostało przedstawione w pracy [8].

U pacjentów z chorobami neurodegeneracyjnymi (a tacy zostali zbadani przez autorów) występują zmiany w mowie o typie dyzartrii. Dyzartria to zespół zaburzeń opuszkowo-fonacyjno-artykulacyjnych spowodowanych uszkodzeniem ośrodków i dróg unerwiających aparat mówienia, występujących mimo zachowanego planu mowy [4]. Inne definicje charakteryzują dyzartrię jako upośledzenie zdolności wytwarzania artykulowanych dźwięków mowy wskutek zakłócenia nerwowych mechanizmów tworzenia głosu oraz modulacji, natężenia, kształtu i rezonansu [2]. Współcześnie dyzartrię określa się jako grupę motorycznych zaburzeń mowy, będących rezultatem uszkodzenia ogniskowego, wieloogniskowego lub rozlanego w centralnym albo obwodowym układzie nerwowym, albo też w obu tych układach.

Ze względu na dominujący objaw zaburzenia wyróżniono sześć typów dyzartrii [3]. W naszych badaniach pacjenci charakteryzowali się dyzartrią hipokinetyczną i hiperkinetyczną. Chorobie Parkinsona oraz zespołowi parkinsonowskiemu (uszkodzenie układu pozapiramidowego; zaburzenia mowy związane są ze spowolnieniem i sztywnością) towarzyszą zmiany w mowie o typie dyzartrii hipokinetycznej.

Do jej najistotniejszych cech na poziomie głosek w izolacji należą:

- niewyraźna artykulacja powodowana szybkimi, ograniczonymi ruchami języka i warg,
- ograniczenie głośności,
- redukcja realizacji głosek aż do bełkotu.

Zakłócenia w procesie mówienia polegają na nieoczekiwanych przerwach w fonacji. Głos jest monotony, cichy, słaby, nikiący, oraz występuje zanikanie głosu. Drugim typem dyzartrii, który występuje w chorobach neurodegeneracyjnych układu pozapiramidowego jest dyzartria hiperkinetyczna. Fonacja jest zniekształcona przez zmiany głośności, brak głosu. Mogą pojawić się nagłe przerwy w mówieniu. Pojawia się nieprawidłowa artykulacja, a także nieregularne przerwy w artykulacji, przedłużanie głosek, powtarzanie głosek powodowane nieprawidłowym napięciem mięśniowym. Może występować nadmierna nosowość, a utrata powietrza wynikająca z zaburzeń podniebieno-gardłowych powoduje skracanie fraz. Występuje wahanie głośności, głos jest drżący, napięto-zdławiony, słaby, pojawiają się przerwy w emisji.

4. Charakterystyka i klasyfikacja głosek zastosowanych w badaniu

W badaniu wykorzystano samogłoski i spółgłoski. Badani proszeni byli o realizację głosek w izolacji. Grupę samogłosek stanowiły [a], [e] oraz [i]. Wyboru głosek dokonano, kierując się różną wysokością wzniesienia języka oraz różnym stopniem otwarcia jamy ustnej (tab. 1).

Im większe zbliżenie grzbietu języka do podniebienia twardego lub miękkiego, tym mniejszy stopień otwarcia drogi ustnej, im niższe jego położenie, tym otwarcie większe. Położenie języka względem poziomej osi ustnego odcinka kanału głosowego jest podstawą podziału samogłosek na mniej lub bardziej przednie albo tylne [6].

Tabela 1
Podział samogłosek

Głoska	Podział wg stopnia wzniesienia języka	Podział wg stopnia otwarcia jamy ustnej
i	wysokie	przymknięte
e	średniowysokie	półprzymknięte
a	niskie	otwarte

W naszym badaniu wykorzystaliśmy samogłoski przednie. Samogłoska wysoka i przednia [i] charakteryzuje się tym, że środkowa część przesuniętej w górę jamy ustnej masy języka znajduje się bardzo blisko środkowej części podniebienia twardego. Natomiast niska i przednia [a] ma zarówno wybrzuszenie nisko położonej masy języka, jak i miejsce na podniebieniu twardym, ku któremu język się wznosi.

Grupę spółgłosek stanowiły [s], [x], [p], [k] oraz [g]. Również tę kategorię głosek można dzielić na grupy i podgrupy, opierając się na rozmaitych kryteriach artykulacyjnych. Jednym z nich jest sposób artykulacji, ograniczający w różnym stopniu przepływ powietrza przez kanał głosowy, aż do jego całkowitej blokady.

Głoski [s] i [x] są przykładami z grupy głosek szczelinowych. Są to spółgłoski, przy których artykulacji odpowiednie części narządu mowy zbliżają się do siebie, tworząc wąską szczelinę. Szczelina, w której strumień powietrza o odpowiedniej masie i prędkości ulega zakłóceniu, może być utworzona w różnych miejscach ponadkraniowego odcinka kanału głosowego. Spółgłoska [s] jest spółgłoską z grupy przedniojęzykowo-zębowych, natomiast [x] jest tylnojęzykowo-podniebienna.

Głoski [p], [k] i [g] są natomiast przykładem głosek zwarto-wybuchowych. Pierwszą fazę ich trwania stanowi zwarcie, czyli krótkotrwała, całkowita blokada przepływu powietrza przez jamę ustną i nosową. Napływające z płuc powietrze zatrzymuje się przed zaporą, jego ciśnienie wzrasta i w pewnym momencie osiąga wartość krytyczną. Następuje gwałtowne przerwanie blokady, któremu towarzyszy zjawisko akustyczne nazywane wybuchem.

Spółgłoska [p] zaliczana jest do dwuwargowych, natomiast [k] i [g] są spółgłoskami z grupy tylnojęzykowo-podniebiennej.

5. Metoda badania

Grupę osób badanych stanowiło 18 osób chorych w wieku od 20 do 80 lat oraz grupa porównawcza osób zdrowych tej samej liczności w podobnym wieku. Głos badanych pacjentów został zarejestrowany za pomocą profesjonalnego sprzętu cyfrowego, w dźwiękoszczelnym pomieszczeniu w celu eliminacji wszelkich niepożądanych czynników mogących wpłynąć negatywnie na końcowy wynik. Z zarejestrowanego głosu wyizolowano poszczególne głoski, a następnie poddano je obróbce (filtracja i analiza widma). Widmo dźwięku zawiera bardzo wiele szczegółów, stąd dla celów automatycznej klasyfikacji badanego głosu konieczna jest parametryzacja.

W pierwszej kolejności sprawdzono cechy widma głosek istotnych diagnostycznie w badaniu subiektywnym.

5.1. Zmiany w realizacji głosek

Do porównania realizacji głosek osób zdrowych i chorych wykorzystano analizę widma otrzymaną za pomocą dyskretnej transformaty Fouriera, równanie (1) [5, 10]. Analiza widma w skali Barkhausena pozwoliła na lepszą klasyfikację zarejestrowanych danych.

$$X(n) = \sum_{k=0}^{N-1} x(k)e^{-jk2\pi n/N} \quad \text{dla } n = 0..N-1 \quad (1)$$

W aplikacji zastosowano szybką transformatę Fouriera (FFT) możliwą w przypadku, gdy N jest potęgą 2. Transformata ta jest zaimplementowana w większości środowisk do obliczeń naukowych, jak np. MATLAB® [11].

W celu efektywnej klasyfikacji danych wyznaczono macierz energii, a następnie zmniejszono wymiar macierzy przez zsumowanie wartości znajdujących się w kolejnych przedziałach odpowiadających skali Barkhausena $f_{low}(i)$ i $f_{high}(i)$, patrz wzory (2) i (3).

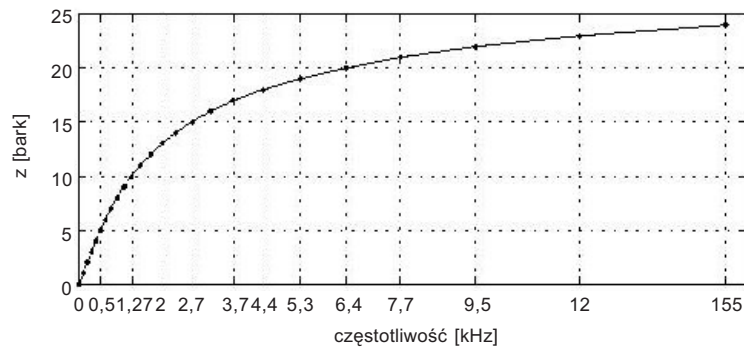
Macierz energii $P(n;t)$, gdzie n jest indeksem częstotliwości, a t czasu, zdefiniowana jest wg wzoru poniżej:

$$P(n,t) = |X_t(n)|^2 \frac{1}{N} \quad \text{dla } n = 0..N/2 \quad (2)$$

$$CB(t,i) = \sum_{n \in I(i)} P(n,t) \quad (3)$$

$$\text{dla } I(i) = \{n : f_{low}(i) < f(n) \leq f_{high}(i)\}$$

Macierz CB zawiera energię w poszczególnych i -tych przedziałach Barkhausena dla odpowiedniego czasu t (rys. 1).



Rys. 1. Charakterystyka skali Barkhausena

U osób chorych, widoczne są zmiany w wymawianych spółgłoskach – dokładnie przejście w inną głoskę w czasie realizacji. Przejście to jest zarówno słyszalne, jak i możliwe jest jego automatyczne wykrycie na podstawie porównania macierzy CB . Zmiany te

obserwowalne były szczególnie dla podanych niżej spółgłosek, dla których zaznaczono także występującą zmianę.

[k] → [a] / [k] → [y]

[g] → [y] / [g] → [e]

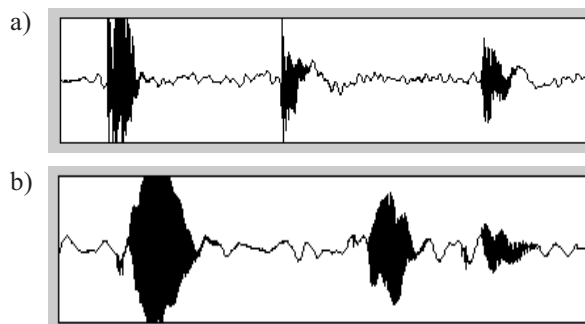
[s] → [y]

[x] → [a]

Wynik takiego zaburzenia spowodowany jest osłabieniem sprężystości mięśni krtani, dlatego spółgłosce towarzyszy samogłoska, która nie wymaga takiego napięcia.

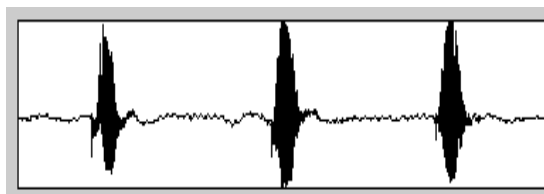
5.2. Natężenie głosek wymawianych wielokrotnie w izolacji

Badani byli poproszeni o powtórzenie cztery razy tej samej głoski zwarto-wybuchowej. Prośba wynikała z wiedzy, iż w czasie realizacji sekwencji tych samych głosek spodziewano się zaburzenia sekwencji dźwięku, pojawienia się mamrotania.



Rys. 2. Amplituda ostatnich trzech głosek osoby chorej: a) osoba „a”; b) osoba „b”

Analizując zarówno przebieg czasowy, jak i amplitudę kolejnych głosek, obserwujemy u większości badanych, że amplituda końcowych głosek była odpowiednio niższa od pierwszej, patrz rysunek 2. Spowodowane jest to sztywnością mięśni, która jest cechą charakterystyczną dla choroby Parkinsona. W grupie kontrolnej u większości badanych nie odnotowano tego zjawiska (rys. 3).



Rys. 3. Amplituda ostatnich trzech głosek osoby zdrowej

5.3. Analiza stabilności dźwięków ciągłych

Badanych poproszono o wypowiedzenie na jednym oddechu jednej z głosek: [a], [e], [i], [s] oraz [x]. Dźwięk dłużej emitowany pozwolił na analizę energii w celu zaobserwowania przeniesienia zmian częstotliwości wynikającej z patologicznego drżenia. Uzyskane wartości porównano z wartościami grupy kontrolnej.

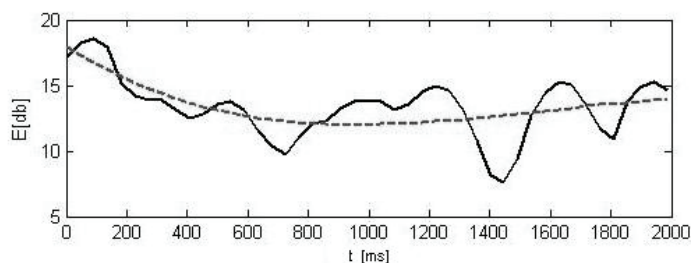
W celu uśrednienia wyników każdy zarejestrowany sygnał głoski został podzielony na ramki 160 ms, z nałożeniem ramek 40 ms. Moc poszczególnych ramek P_2 została wyznaczona przez zsumowanie wartości znajdujących się w zakresie od $x(t)$ do $x(t+m)$, gdzie m jest długością ramki.

$$P_2(t) = \sum_{k=t}^{t+m} |x(k)|^2 \quad (4)$$

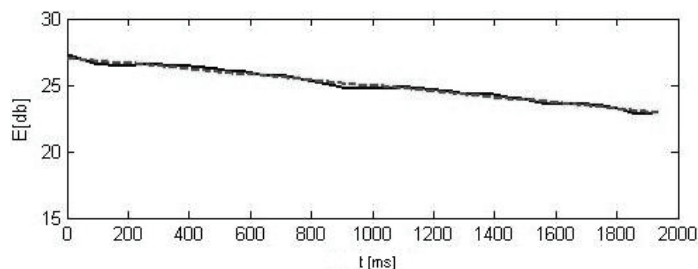
W celu wyznaczenia stabilności realizacji głosek dopasowano do otrzymanego wyżej przebiegu wielomian $p(t)$ stopnia 4, a następnie określono stabilność jako współczynnik *Stab*:

$$Stab = \sum_{k=1}^t |P_2(t) - p(t)| \quad (5)$$

Na rysunkach 4 i 5 przedstawione są przykładowe wyniki badania osoby chorej i zdrowej.



Rys. 4. Głoska [a] z grupy osób chorych, $Stab = 1,43$



Rys. 5. Głoska [a] z grupy osób kontrolnych, $Stab = 0,14$

Współczynnik *Stab* był znacznie większy dla osób chorych, co pozwoliło na wyznaczenie kolejnego, znaczącego parametru klasyfikującego osoby ze zmianami patologicznymi w mowie.

6. Podsumowanie

Prezentowane wyniki stanowią początek badań nad automatyczną klasyfikacją głosu. Nawiązano do zagadnień parametryzacji zarówno przebiegu czasowego, jak i parametryzacji widma dźwięku. Głównym zadaniem postawionym w przyszłych badaniach jest takie zdefiniowanie deskryptorów, które w połączeniu z określonymi algorytmami przeszukiwania pozwolą na prawidłową interpretację zmian w głosie pacjenta.

Prezentujemy także propozycję rejestracji, obróbki i analizy mowy traktowanej jako sygnał cyfrowy.

Planowana jest kontynuacja analizy wyizolowanych głosek, porównywanych w realizacji u osób zdrowych i chorych przy uwzględnieniu wieku i płci oraz badań nad analizą fraz; u chorych rejestrowana będzie dynamika zmian związanych z progresją choroby.

Zarejestrowano też materiał lingwistyczny na poziomie fraz i obecnie przygotowywana jest jego analiza techniczna. W tych badaniach będą analizowane elementy prozodyczne mowy: rytm, tempo, intonacja, akcent, melodia mowy. Wymienione powyżej elementy mowy są dostępne w diagnostyce subiektywnej (określenie typu dyszartrii). Autorom zależy jednak na zbadaniu specyfiki tych zmian w mowie, które są niedostępne w badaniu subiektywnym oraz na stworzeniu kompleksowego modelu automatycznej klasyfikacji. W tym celu wykorzystane będą najnowsze metody analizy rejestracji, obróbki i analizy sygnału.

Literatura

- [1] Duus P.: *Neurologisch-Topische Diagnostik*. Stuttgart, 1983
- [2] Johns D.F.: *Clinical management of neurogenic communicative disorders*. Boston, College Hill, 1985
- [3] Darley F., Aronson A., Bron J.: *Cluster of deviant speech dimension in the dysarthrias*. Journal of Speech and Hearing Research, 12, 1969, 462–496
- [4] Lamig L., Sapir S., Countryman S.: *Intensive voice treatment (LSVT®) for patients with Parkinson's disease: a 2 year follow up*. J Neurol Neurosurg Psychiatry 71, 2001, 439–498
- [5] Tadeusiewicz R.: *Sygnal Mowy*. Warszawa, Wydawnictwa Komunikacji i Łączności 1988
- [6] Wróbel H.: *Gramatyka współczesnego języka polskiego – Fonetyka i fonologia*. Kraków, PAN 1995
- [7] Pruszewicz A.: *Foniatria kliniczna*. Warszawa, PZWL 1992
- [8] Izworski A., Tadeusiewicz R.: *Artificial Intelligence Methods in Diagnostics of the Pathological Speech Signals*. Speech and Language Technology, vol. 6, PPA, Poznań, 2002, 183–197
- [9] Ralston A.: *A first course in numerical analysis*. New York, 1965
- [10] Körner T.: *Fourier Analysis*. Cambridge University Press, 1988
- [11] Hanselman D.C., Littlefield B.L.: *Mastering MATLAB 7*. USA, 2004

