

19 sierpnia 2019 r.

prof. dr hab. inż. Paweł Strumiłło  
Instytut Elektroniki  
Politechniki Łódzkiej

RECENZJA ROZPRAWY DOKTORSKIEJ  
DLA WYDZIAŁU INŻYNIERII BIOMEDYCZNEJ  
POLITECHNIKI ŚLĄSKIEJ

*Tytuł rozprawy: Ocena normatywności cech artykulacyjnych głosek dentalizowanych z wykorzystaniem wielowymiarowej analizy sygnału mowy*

*Autor rozprawy: mgr inż. Zuzanna Miodońska*

*Promotor: dr hab. inż. Paweł Badura*

*Promotor pomocniczy: dr inż. Joanna Czajkowska*

**I. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autora?**

Przedłożona do recenzji rozprawa dotyczy zastosowania komputerowych metod analizy akustycznej sygnału mowy do wspomagania diagnostyki logopedycznej. Autorka skupiła się w pracy na jednej z najczęściej występującej wadzie wymowy jaką jest seplenienie zwane też sygmatyzmem. Wada ta wynika z nieprawidłowej artykulacji tzw. głosek dentalizowanych, tj. takich, których wymowa wymaga zbliżenia do siebie górnych i dolnych zębów oraz odpowiedniego ułożenia języka i warg. Takie ułożenie narządów artykulacyjnych często sprawia trudność dzieciom w wieku przedszkolnym. Szacuje się, że wady te stanowią ok. 30% wad wymowy u dzieci. Co prawda komputerowe metody analizy wymowy, a szczególnie rozpoznawania mowy są szeroko rozwijane to brak jest metod pomiarowych i obliczeniowych przeznaczonych specjalnie do diagnozy patologii wymowy, w szczególności patologii wymowy występujących u dzieci.

Autorka sformułowała i wykazała w pracy następującą tezę:

*Wykorzystanie cech sygnału mowy oraz statystycznych modeli czasowych głosek dentalizowanych /f/ i /ʒ/ pozwala na przeprowadzenie detekcji wybranych nienormatywnych cech artykulacji tych głosek u dzieci w wieku przedszkolnym.*

W pierwszym rozdziale pracy problem badawczy określony w tezie rozprawy został wystarczająco jasno scharakteryzowany. Autorka nakreśliła zagadnienia związane z wymową zarówno w ujęciu fizjologicznym jak i technicznym oraz opisała szczegółowo normatywne i patologiczne artykulacje głosek dentalizowanych.

## II. Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?

Podjęty przez Doktorantkę problem badawczy jest ważny i w dużym stopniu pionierski. Odpowiednio wczesna i prawidłowa diagnoza wad wymowy u dzieci jest podstawą skutecznej terapii logopedycznej. Jest ona szczególnie ważna w okresie rozwoju dziecka gdyż nie podjęta odpowiednio wcześnie może prowadzić do utrwalenia wad wymowy, a także może mieć konsekwencje psychologiczne. Badacze niezwiązani z naukami technicznymi, m.in. lekarze w coraz większym stopniu przekonują się o skuteczności komputerowych metod analizy danych i sygnałów. Również komputerowe metody wspomagające diagnostykę i terapię medyczną w tym logopedyczną zyskują na znaczeniu.

## III. Czy autor rozwiązał postawiony problem i czy użył to do tego właściwych metod?

Zgodnie z postawioną tezą pracy Autorka sformułowała następujące szczegółowe cele badawcze pracy:

- C1. Przygotowanie bazy mowy zawierającej zapisy mowy normatywnej oraz mowy nienormatywnej (z patologiami artykulacyjnymi) wraz z anotacjami logopedycznymi.
- C2. Opracowanie metodyki przetwarzania wstępnego sygnału mowy oraz wydzielania parametrów, w pasmie szumu, charakteryzujących głoski dentalizowane (sybilanty).
- C3. Wyznaczenie powiązań pomiędzy parametrami sygnału mowy a artykulacjami patologicznymi.
- C4. Opracowanie statystycznego modelu artykulacji głosek z wykorzystaniem ukrytych modeli Markowa.
- C5. Opracowanie komputerowych metod oceny prawidłowych i nieprawidłowych artykulacji.

Ad. C1. Przgotowanie bazy mowy zawierającej zapisy mowy normatywnej oraz mowy nienormatywnej. Przygotowanie odpowiednio opracowanej bazy danych z rejestracjami reprezentatywnych przykładów mowy prawidłowej i patologicznej jest podstawą do przeprowadzenia wiarygodnych badań wad wymowy. Autorka rozprawy podjęła w tym zakresie współpracę z logopedami i językoznawcami. Wykonała zapisy mowy u 79 dzieci w wieku 5 i 6 lat, które uzyskały zgodę rodziców na badanie, nie miały infekcji dróg oddechowych i miały kompletne uzębienie. Rejestracje cyfrowe sygnału z częstotliwością próbkowania 44100 Hz i 16 bitowym kodowaniem próbki wykonano z wykorzystaniem specjalnie zaprojektowanej maski akustycznej zapewniającej powtarzalne warunki akustyczne nagrań. Wśród dzieci wydzielono cztery grupy: z wymową normatywną (25), addentalną (7), interdentalną (10) oraz dentalną (11). W nawisach podano liczebność każdej z grup. Badane wady wymowy zostały w pracy odpowiednio objaśnione, m.in. na rys. 3.1. Z grupy 79 dzieci wybrano rejestracje od 53 dzieci, dla których występuje pojedyncza patologia wymowy. Wykonane nagrania zawierały izolowane słowa zawierające głoski detalizowane występujące na początku, w środku i na końcu wypowiedzianych słów. Każde dziecko bezpośrednio po rejestracji nagrań przechodziło badanie logopedyczne. Należy stwierdzić, że baza mowy wraz z anotacjami eksperckimi została przygotowana

z należytą starannością i stanowi wartość samą w sobie. Może być podstawą dla szeregu innych prac badawczych z dziedziny logopedii.

- Ad. C2. Opracowanie metodyki przetwarzania wstępnego sygnału mowy oraz jego parametryzacji. W zakresie tego celu badawczego Autorka zastosowała klasyczne metody parametryzacji sygnału mowy. Przeprowadziła segmentację sygnału mowy na głoski za pomocą autorskiej aplikacji, w której podstawą segmentacji były przebiegi czasowe sygnału oraz spektrogramy. Dla segmentowanych fragmentów sygnału wyznaczono 78 parametrów, do których należały: współczynniki melcepstralne (MFCC), współczynniki cepstralne (CC) i (NCC), momenty statystyczne wyznaczone na podstawie widma sygnału, parametry powiązane z formantami oraz energie pasma szumu (NE). Jest to bogaty zestaw parametrów uwzględniający w szczególności pasmo szumu związane z artykulacją głosek dentalizowanych, tj. pasmo pomiędzy częstotliwościami występowania pierwszego i czwartego tzw. formantu szumowego.
- Ad. C3. Wyznaczenie powiązań pomiędzy parametrami sygnału mowy a artykulacjami patologicznymi – analiza akustyczno-artykulacyjna. Ten wątek badawczy realizowano w dwóch etapach. W pierwszym przeprowadzono analizę rozkładów statystycznych dla parametrów opisujących sygnał mowy. Stwierdzono (na podstawie testu Shapiro-Wilka), że zdecydowana większość parametrów nie ma rozkładu normalnego w grupach. Fakt ten ma istotne konsekwencje w dalszych badaniach związanych z testowaniem hipotez statystycznych odnośnie badanych parametrów ich opisu statystycznego. Do oceny zdolności dyskryminacyjnej różnych typów artykulacji na podstawie parametrów sygnału mowy zastosowano nieparametryczny test Kruskala-Wallisa. Oceniono różnice median i średnich między grupami. Przeprowadzono również analizę związków korelacyjnych pomiędzy wartościami parametrów a przynależnością do grup artykulacji. Badania statystyczne podsumowano obszerną dyskusją, w której stwierdzono m.in. że perceptualna skala częstotliwościowa (skala melowa) nie wniosła znaczącej przewagi nad skalą liniową. Pokazano też, że największy związek korelacyjny pomiędzy położeniem języka a opisem parametrycznym sygnału stwierdzono dla wybranych współczynników cepstralnych. Jednak zasadniczym wnioskiem z badań statystycznych jest stwierdzenie, że różnorodny i wieloparametryczny opis sygnału może stanowić podstawę do dokładniejszej dyskryminacji różnych typów artykulacji (normatywnych i nienormatywnych). Trzeba stwierdzić, że badania te przeprowadzono bardzo skrupulatnie z zachowaniem wszelkich reguł analizy statystycznej sygnałów.
- Ad. C4. Opracowanie statystycznego modelu artykulacji głosek z wykorzystaniem ukrytych modeli Markowa (ang. *Hidden Markov Model – HMM*). Niniejszy cel badawczy polegał na zastosowaniu podejścia modelowego do wykrywania nienormatywnych artykulacji. Zastosowaną strukturą modelu jest model HMM. Jest to uzasadniony i zweryfikowany w zastosowaniach analizy sygnałów mowy dwuwarstwowy model statystyczny. Niezależne modele cząstkowe opracowano dla mowy normatywnej oraz oddzielnie dla każdej z analizowanej patologii. Modele podlegały uczeniu, polegającemu na poszukiwaniu takiego zestawu parametrów modelu, dla którego jest

osiągane najlepsze dopasowanie modelu dla poszczególnych artykulacji. Dobór parametrów modelu przeprowadzono wg iteracyjnego algorytmu Bauma-Welcha. Uzasadniono wybór tego algorytmu. Opracowane modele dla sygnałów odpowiadających różnym artykulacjom zastosowano jako jedno z podejść do klasyfikacji mowy normatywnej i różnych typów mowy nienormatywnej.

Ad. C5. Opracowanie metodyki komputerowej klasyfikacji prawidłowych i nieprawidłowych artykulacji oraz analiza uzyskanych wyników. Osiągnięte wcześniej cele badawcze były etapami przygotowawczymi do przeprowadzenia klasyfikacji artykulacji normatywnych i nienormatywnych. Zastosowano kilka metod klasyfikacji danych (z modelem sygnału i bez modelu) oraz typów klasyfikacji, tj. dwuklasowej i wieloklasowej. Spośród wielu klasyfikatorów zastosowano klasyfikator wektorów podtrzymujących (ang. *Support Vector Machine*) oraz ten klasyfikator w autorskim układzie hybrydowym z modelem HMM. Klasyfikator SVM jest uznawany za jeden z najbardziej skutecznych klasyfikatorów w zadaniach nieliniowej klasyfikacji danych, w szczególności danych o dużej liczbie wymiarów. Dodatkowo również zbadano skuteczność klasyfikacji dla klasyfikatorów wybranych jako referencyjne, tj.  $k$ -najbliższych sąsiadów ( $k$ -NN) i klasyfikatora wieloklasowego SVM implementowanego z wykorzystaniem modelu ECOC (ang. *Error Correcting Output Codes*). W tym obszarze badań Autorka dokładnie zbadała działanie klasyfikatorów dla różnych ich parametrów. Dla SVM badała działanie klasyfikatora dla różnych wartości parametru regularyzacyjnego oraz „szerokości” funkcji jądra, a dla modelu HMM optymalizowała liczbę stanów modelu. Przebadła również skuteczność klasyfikatorów dla różnego zestawu wyznaczonych parametrów sygnału mowy. Zastosowane metryki skuteczności klasyfikacji to: czułość, swoistość i dokładność. Przedstawiła bogaty zestaw wyników dla klasyfikacji dwu- i wielo-klasowych oraz dla różnych badanych klasyfikatorów, m.in. w postaci tzw. macierzy pomyłek. Wyniki te są bardzo wartościowe z punktu widzenia diagnostyki logopedycznej. Zostały one skrupulatnie przedyskutowane, m.in. stwierdzono, że lepsze wyniki klasyfikacji uzyskano dla głoski /f/ niż dla głoski /ʃ/, dla pełnego zestawu parametrów sygnału mowy oraz dla klasyfikacji dwuklasowych. Największą dokładność klasyfikacji patologii wymowy wyniosła 86%. Wynik ten oraz szereg obszernych analiz innych wyników potwierdza osiągnięcie celu badawczego postawionego w pracy.

#### **IV. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora? Jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?**

Tematyka automatycznego rozpoznawania mowy jest dziedziną intensywnie i z sukcesami od lat rozwijaną. Autorka rozprawy podjęła jednak szczegółowy, słabo rozpoznany w literaturze światowej wątek badawczy związany z rozpoznawaniem wad wymowy sybilantów. Autorka rozprawy powołuje się na pozycję literatury [112] z roku 2012, która również dotyczyła rozpoznawania wad wymowy głosek dentalizowanych, ale badania te były przeprowadzone dla wad wymowy u osób dorosłych. Koncentracja kierunku badań na wadach wymowy u dzieci

jest nowym podejściem badawczym i stanowiącym oryginalny wkład Autorki do komputerowych metod wspomagających diagnostykę logopedyczną. W szczególności za najbardziej wartościowe osiągnięcia Autorki rozprawy uważam:

- 1) wyznaczenie zależności ilościowych pomiędzy parametrami sygnału mowy, a prawidłowo i nieprawidłowo wypowiedzianymi głoskami dentalizowanymi,
- 2) opracowanie autorskich, odpowiednio dostrojonych statystycznych modeli cząstkowych HMM odwzorowujących sygnały mowy prawidłowej i mowy z wadami artykulacyjnymi dla głosek dentalizowanych,
- 3) opracowanie metody hybrydowej wykorzystującej połączenie klasyfikatora SVM oraz modeli HMM do dyskryminacji wad artykulacji głosek dentalizowanych.

#### **V. Czy rozprawa świadczy o dostatecznej wiedzy autora i znajomości współczesnej literatury z dyscypliny naukowej, której dotyczy?**

Autorka ma bardzo dobre rozeznanie w literaturze krajowej i światowej dotyczącej tematyki rozprawy. Zamieszczony w pracy wykaz literatury liczy 130 pozycji. W szczególności w sekcji pracy 1.4 *Komputerowa analiza wymowy w literaturze* w sposób metodyczny i syntetyczny dokonano przeglądu najaktualniejszych prac w dynamicznie i szeroko rozwijanej dziedzinie analizy mowy w tym w obszarze diagnostyki i terapii logopedycznej. W podanym wykazie literatury zamieszczono 6 współautorskich publikacji Autorki rozprawy w tym dwie prace opublikowane w czasopismach naukowych z listy JCR.

#### **VI. Jakie są słabe strony rozprawy i jej główne wady?**

Strona formalna rozprawy jest bardzo dobra. Praca jest przygotowana starannie pod względem edycyjnym i graficznym oraz wydana w wygodnym dla czytelnika formacie książkowym.

Punkty dyskusyjne (niepodważające mojej pozytywnej oceny całości pracy), na które chciałbym zwrócić uwagę są następujące:

- w wielu miejscach pracy, również w tezie jest mowa o „cechach sygnału”, „cecha” nie jest wielkością mierzalną charakteryzującą sygnał, powinna tu być raczej mowa o parametrach sygnału,
- proszę o komentarz odnośnie opisu parametrycznego sygnału; wybranych 78 parametrów jest licznym ale ograniczonym zbiorem parametrów, dlatego dla przykładu nie uwzględniono parametrów statystycznych wyznaczanych bezpośrednio z próbek sygnału, np. momenty statystyczne.

Praca jest napisana dobrym stylem. Zauważono jednak następujące błędy językowe i edycyjne:

- w bardzo wielu miejscach pracy Autorska stosuje określenia „wyższe” lub „niższe” m.in. w odniesieniu do dokładności klasyfikacji, podczas gdy określenie to raczej dotyczy rozmiaru fizycznego związanego z wysokością; dla przykładu na str. 86 napisano: „...najwyższym wynikiem...”,
- inne niepoprawne stylistycznie sformułowania to:
  - str. 33: „...pik o niższej amplitudzie...” (powinno być: „maksimum o mniejszej amplitudzie”).

- str. 43: „...*ekstrakcji formantów szumowych ...*” (powinno być: „*wydzielaniu formantów szumowych*”),
- str. 49: „... *detekowanych szczytów*” (powinno być: „*wykrywanych maksimów*”),
- str. 65 „... *zwracane przez SVM*” (powinno być: „*wyznaczane przez SVM*”), ,
- str. 93: „... *jakikolwiek wnioskowanie nie ma szans na bycie przetestowanym w praktyce.*” (powinno być: „*jakikolwiek wnioskowanie nie jest możliwe w praktyce*”),
- w tekście pracy jest wiele odwołań do nr wierszy w tabelach 7.1 i 7.4, podczas gdy wiersze w tych tabelach nie są numerowane.
- w pracy jest używanych bardzo wiele skrótów, przydatny dla czytelnika byłby wykaz stosowanych skrótów i oznaczeń.
- zwyczajowo w pracach doktorskich zamieszcza się streszczenia pracy w języku polskim i angielskim, których w pracy zabrakło.

## VII. Wniosek końcowy

Przedłożona do oceny rozprawa doktorska jest wartościowym opracowaniem wnoszącym oryginalny wkład do dyscypliny *Inżynieria Biomedyczna* w zakresie analizy sygnałów mowy i komputerowych metod wspomagania diagnostyki logopedycznej. Autorka rozprawy zaproponowała oryginalną, autorską metodykę analizy sygnałów mowy u dzieci. Skoncentrowała się na wadach wymowy głosek detalizowanych stanowiących najczęściej występującą wadę wymowy u dzieci w wieku przedszkolnym. Postawioną tezę pracy dowiodła na drodze eksperymentalnej, w której na podstawie opracowanej bazy danych mowy poprawnej i patologicznej wykazała skuteczność opracowanych metod dyskryminacji podstawowych typów artykulacji nienormatywnych.

Biorąc pod uwagę oryginalne osiągnięcia badawcze Autorki rozprawy, bardzo dobry dorobek publikacyjny (dwa współautorskie artykuły w czasopismach naukowych z listy JCR oraz inne publikacje) **zglaszam rozprawę doktorską mgr inż. Zuzanny Miodońskiej do wyróżnienia.**

Przedłożona do recenzji rozprawa doktorska spełnia wymagania Ustawy o stopniach naukowych i tytule naukowym z 14 marca 2003 roku, Dziennik Ustaw Nr 65, poz. 595 z późn. zm. oraz Rozporządzeniem Ministra Nauki i Szkolnictwa Wyższego z dnia 19 stycznia 2018 roku w sprawie szczegółowego trybu i warunków przeprowadzania czynności w przewodach doktorskich, postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora. Wnioskuję o dopuszczenie rozprawy doktorskiej mgr inż. Zuzanny Miodońskiej do publicznej obrony.

*Paweł Stankiewicz*