

Dr hab. inż. Mirosław Świercz, prof. PB
Katedra Automatyki i Robotyki
Wydział Elektryczny
Politechnika Białostocka

Białystok, 2 lipca 2025 r.

RECENZJA ROZPRAWY DOKTORSKIEJ

Pana
por. mgr. inż. Dominika Małego
pt. „**System automatycznego rozpoznawania mowy wsparty behawioralnymi cechami
sygnału mowy**”

(zlecenie Przewodniczącego Rady Dyscypliny Naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego, prof. dr hab. inż. Jana K. Jabczyńskiego, na podstawie Uchwały Nr 53/RDN AEEITK/2025 z dnia 16 kwietnia 2025 r.)

1. Ogólna charakterystyka, przedmiot, teza i cel rozprawy

Rozprawa doktorska por. mgr. inż. Dominika Małego została wykonana pod kierunkiem promotora – prof. dr hab. inż. Andrzeja P. Dobrowolskiego oraz promotora pomocniczego mjr. dr hab. inż. Kamila Kamińskiego. Rozprawa zawiera łącznie 180 stron, z czego 125 stron stanowi właściwa treść rozprawy, zaś pozostałe 55 stron zajmują załączniki Z.1-Z.4. Treść rozprawy została podzielona na następujące części:

- indeks skrótów użytych w rozprawie;
- streszczenie (w języku polskim i angielskim);
- przedmowę, w której Autor przedstawia w skrócie motywację do podjęcia prac badawczych, których rezultaty zostały opisane w rozprawie, formułuje tezę rozprawy oraz charakteryzuje zawartość poszczególnych rozdziałów pracy;
- podziękowania złożone promotorom pracy oraz rodzicom Doktoranta;
- wprowadzenie, w którym Doktorant charakteryzuje układ generacji sygnału mowy, opisuje podział systemów automatycznego rozpoznawania mowy (ASR) oraz dokonuje przeglądu literatury światowej, poświęconej modelowaniu głosów oraz rozpoznawaniu mowy;
- pięć rozdziałów merytorycznych, prezentujących autorską koncepcję rozbudowy i udoskonalenia systemu automatycznego rozpoznania mowy, istniejącego na Wydziale Elektroniki Wojskowej Akademii Technicznej oraz wyniki obszernych badań efektywności rozwiązań, zaproponowanych przez Doktoranta (w tym weryfikację jakości rozpoznawania w warunkach środowiskowych, zbliżonych do rzeczywistych);
- podsumowanie uzyskanych wyników oraz charakterystykę osiągnięć własnych Autora rozprawy;
- spis literatury, zawierający 158 pozycji, w tym jedną pozycję, opublikowaną wspólnie przez Doktoranta i promotora rozprawy;
- cztery załączniki, w których Doktorant prezentuje m.in. zbiór definicji wszystkich cech behawioralnych występujących w opracowanym przez niego systemie ASR oraz pełne wyniki badań i testów, przeprowadzonych w trakcie przygotowywania recenzowanej rozprawy doktorskiej.

h. Świercz

W rozdziale „Przedmowa” na stronie 11 Autor formuluje tezę rozprawy: „możliwe jest zdefiniowanie i wyekstrahowanie z sygnału mowy takich cech behawioralnych, które w połączeniu z cechami fizycznymi zaimplementowanymi w istniejącym rozwiązaniu znacząco poprawią jego [tj. systemu wspomagającego rozpoznawanie mówcy – M. Ś.] skuteczność w obecności zewnętrznych zakłóceń”. Jak dalej wyjaśnia Doktorant, zbiór cech behawioralnych i oparte na nich algorytmy stanowiąc będą wsparcie dla opracowanego wcześniej w macierzystej uczelni systemu automatycznego rozpoznawania mówcy, opartego na cechach fizycznych. Problem badawczy, którego rozwiązania podjął się Doktorant, polega więc na udoskonaleniu istniejącego rozwiązania technicznego poprzez wprowadzenie dodatkowego zbioru cech sygnału, powstałych w rezultacie oryginalnych koncepcji ekstrakcji cech oraz istotnej modyfikacji algorytmów klasyfikacji i rozpoznawania wzorców.

Do sformułowania tezy rozprawy Autor powraca na str. 45, na koniec podsumowania Rozdziału 2, gdzie określa zadania szczegółowe, mające na celu udowodnienie powyższej tezy rozprawy, tj. [cyt. – M. Ś.]:

- opis systemu odniesienia opartego na cechach fizycznych;
- propozycję samodzielnego systemu ASR opartego wyłącznie na cechach behawioralnych (w celu przebadania szerokiego zbioru cech i dokonania ich wstępnej selekcji);
- rozważenie różnych rodzajów fuzji obu systemów;
- badania eksploatacyjne oraz wskazanie najlepszego wariantu pracy systemu ASR i eksperymentalne potwierdzenie słuszności postawionej tezy.

Zdaniem recenzenta, takie sformułowanie zadań badawczych oraz opisana w początkowych rozdziałach tekstu rozprawy motywacja do podjęcia tematyki badawczej i poszukiwania rozwiązania problemu są jak najbardziej prawidłowe i wystarczające do wykazania słuszności tezy.

Udowodnienie tezy rozprawy oraz osiągnięcie zakładanych celów szczegółowych było zadaniem ambitnym, które wymagało zastosowania zaawansowanych algorytmów przetwarzania sygnałów oraz klasyfikacji ich cech, wzbogaconych o oryginalne, autorskie rozwiązania teoretyczne. Stopień złożoności zadania, którego rozwiązania podjął się Doktorant, wynika przede wszystkim z charakteru sygnału mowy ludzkiej – „trudnego” do opisu matematycznego, zmieniającego się w wyniku działania wielu indywidualnych czynników, o zmiennej naturze i zmiennym, niestacjonarnym wpływie na ten sygnał. Zdaniem recenzenta, cel rozprawy i zadania badawcze zostały sformułowane poprawnie, a podjęta w pracy próba ulepszenia działania istniejącego systemu identyfikacji i rozpoznawania mówcy jest niewątpliwie zadaniem naukowym odpowiadającym wymaganiom stawianym rozprawie doktorskiej. Dodatkowo, tematyka rozprawy wynika z potrzeby rozwiązania realnego problemu technicznego, który jak dotychczas (pomimo wielu dość skutecznych prób, raportowanych w literaturze) nie doczekał się kompleksowego rozwiązania – można więc z pewnością stwierdzić, że podjęta i rozwinięta w rozprawie problematyka badawcza ma znaczenie zarówno z teoretycznego, jak i z praktycznego punktu widzenia.

2. Ocena rozprawy – najważniejsze osiągnięcia Autora

Problem badawczy, którego rozwiązanie zostało przedstawione w rozprawie, wynika z praktycznej potrzeby wykorzystania wiedzy teoretycznej oraz algorytmów obliczeniowych w wielu obszarach współczesnej techniki, bankowości, kryminalistyki, jak również obronności – często w aplikacjach wykorzystujących interakcję człowiek-komputer, działających w czasie rzeczywistym. Podjęty przez Doktoranta temat identyfikacji tożsamości mówcy, w szczególności w utrudnionych warunkach rejestracji sygnału mowy, nie doczekał się, jak dotychczas, jednoznacznego i kompleksowego rozwiązania – przede wszystkim ze względu na złożoność charakterystyki mowy ludzkiej. Świadczą o tym wyniki prac nad systemami automatycznego rozpoznania mówcy (ang. *Automatic Speaker Recognition*,

ASR), prowadzonych w renomowanych ośrodkach badawczych, przytoczone przez Doktoranta w rozdziale „1.3. Krytyczny przegląd literatury”. W tym miejscu należy podkreślić, że Doktorant wykazał się bardzo dobrą znajomością aktualnego stanu wiedzy w zakresie, którego dotyczy tematyka ocenianej rozprawy doktorskiej.

Autor rozprawy podjął się opracowania systemu wspomagającego rozpoznawanie mowy na podstawie klasyfikacji behawioralnych cech dystynktywnych głosu ludzkiego, które zostały wyodrębnione z sygnału mowy za pomocą metod analizy czasowej, częstotliwościowej, czasowo-częstotliwościowej i cepstralnej sygnału. Autorskie klasyfikatory opracowane w rozprawie miały za zadanie rozszerzać możliwości systemów opracowanych przez naukowców z Wojskowej Akademii Technicznej, m.in. systemu „Automatyczne Rozpoznawanie Mówcy i Autoryzacja (ARMIa)”. Doktorant skoncentrował się na dostosowaniu własnego rozwiązania do rzeczywistych warunków powstawania i rejestracji sygnału, a także uodpornieniu systemu rozpoznania na wpływy zniekształceń sygnału generowanego przez mówcę (zróżnicowanych torów akustycznych, maseczek, przebiegów itp.).

W opinii recenzenta, przyjęte założenie o wbudowaniu oryginalnych koncepcji i algorytmów opracowanych przez Doktoranta w istniejący system uzasadnia potrzebę dość precyzyjnego opisu środowiska, w którym te pomysły są implementowane. Jednakże, zdaniem recenzenta, proporcja pomiędzy ilością informacji, która nie jest opisem rezultatów prac Doktoranta (tj. prezentacją stanu wiedzy oraz systemów ASR opracowanych w WAT), a własnymi dokonaniem została w rozprawie zaburzona – prezentacja dorobku Autora rozpoczyna się w zasadzie od połowy tekstu rozprawy, tj. od rozdziału „4. System oparty na cechach behawioralnych”. Trudno również w rozprawie znaleźć spójną prezentację architektury systemu ASR, opracowanego przez Doktoranta, wraz z charakterystyką tej koncepcji – rozproszone w drugiej części tekstu schematy blokowe (np. rys. 5.8. „Proces fuzji danych na poziomie decyzji” na str. 88, czy rys. 5.9. „Proces fuzji danych na poziomie cech” na str. 91) nie spełniają w pełni tego wymagania.

Recenzent podziela opinię Doktoranta, który jako swe główne osiągnięcia wskazuje (str. 113):

- opracowanie autorskiej metody ekstrakcji cech charakteryzujących mówcę, która wykorzystuje przetwarzanie nakładających się ramek mowy;
- wybór efektywnego zestawu cech behawioralnych, poprzez selekcję, dokonaną za pomocą algorytmów ewolucyjnych, dających najlepszą efektywność klasyfikacji;
- opracowanie autorskich metod integracji danych wykorzystujące fuzję na poziomie decyzji oraz na poziomie cech;
- eksperymentalną ocenę dokładności i efektywności działania systemu w różnorodnych warunkach, zbliżonych do warunków rzeczywistych.

Zdaniem recenzenta, poziom złożoności tematyki, której dotyczy praca, odpowiada wymaganiom stawianych tematów rozpraw doktorskich. Doktorant prawidłowo sformułował cel i plan badań oraz dobrał zaawansowane metody rozwiązania sformułowanego problemu. Słuszność tezy rozprawy Doktorant potwierdził w sposób eksperymentalny, wykorzystując metody i algorytmy przetwarzania sygnałów oraz algorytmy ekstrakcji zbiorów cech sygnału mowy, które zapewniają skuteczne działanie klasyfikatorów. Treść i poziom rozprawy świadczy o dobrej znajomości przez Doktoranta współczesnego stanu wiedzy w obszarze bezpośrednio związanym z tematem rozprawy, jak również w obszarach zbliżonych tematycznie do głównego nurtu rozprawy. Ogólna ocena przedstawionej do recenzji rozprawy doktorskiej jest zatem pozytywna.

3. Uwagi krytyczne, dyskusyjne i komentarze do rozprawy

Oprócz niewątpliwie pozytywnej ogólnej opinii na temat jakości rozprawy doktorskiej, podczas jej lektury nasuwa się również kilka uwag krytycznych. Recenzent proponuje, aby podczas publicznej obrony rozprawy Doktorant ustosunkował się do niektórych uwag dyskusyjnych, zamieszczonych poniżej.

3.1. Opisy eksperymentów, testujących poprawność i odporność proponowanego rozwiązania.

Doktorant zrealizował budzącą uznanie liczbę eksperymentów symulacyjnych – zarówno podczas konstruowania i wyboru najlepszego modelu automatycznego rozpoznania mówcy, jak i na etapie ilościowej oceny poprawności działania tych modeli w różnych warunkach środowiskowych. Jednakże, zdaniem recenzenta, słabszą stroną rozprawy jest ilość i jakość komentarzy Autora do rezultatów badań eksperymentalnych.

Dla przykładu – na str. 5.6 i 5.7 Doktorant przytacza zaczerpnięte z literatury schematy blokowe systemów fuzji danych na poziomie decyzji (*nota bene* nie jest dobrą praktyką nadawanie tego samego tytułu dwóm rysunkom prezentującym zupełnie inne treści). Nie jest jasne, czy i w jakiej wersji Doktorant użył tych schematów swoim systemie ASR, z jakimi technicznymi realizacjami klasyfikatorów, z jakimi parametrami liczbowymi, itp.

Podobnie – tabele 5.1 i 5.2 na stronach 89 i 90 przedstawiają sprawność systemu ASR opartego na cechach fizycznych z preselektorem behawioralnym dla różnych czasów trwania sygnału mowy, różnej liczby najbliższych sąsiadów i zastosowanych miar odległości. Z komentarza do tych rysunków nie wynika jednak wniosek, których wartości w/w parametrów użył Doktorant, jaka jest jego własna ocena uzyskanych wyników, czy na przykład uważa liczbę 600 najbliższych sąsiadów za racjonalną, itp.

Analogiczne zastrzeżenia wyraża recenzent wobec opisu tabelarycznych i graficznych wyników eksperymentów, zamieszczonych w rozdziale „6. Badania eksploatacyjne systemu ASR”, a także w dodatkach (na przykład brak jest komentarza do bardzo małych wartości sprawności systemów, zamieszczonych w tabeli Z.2.9 na str. Z.2.8).

3.2. Testy opracowanych rozwiązań w warunkach bezszumowych.

Na rys. 6.2 (str. 99) Doktorant przedstawia główny rezultat badania skuteczności identyfikacji mówcy w różnych rozwiązaniach ASR dla zróżnicowanej długości segmentu treningowego. Analiza wzrokowa wyników uzyskanych dla trzech systemów: istniejącego systemu ARMiA, wariantu modyfikacji tego systemu z fuzją na poziomie cech oraz dla „Preselekcji 1000 NN” wskazuje na daleko idące podobieństwo skuteczności identyfikacji mówcy. Przeczy temu komentarz Doktoranta: „Bazując na powyższych wynikach widać przewagę podejść fuzji na poziomie cech oraz decyzji nad innymi wariantami systemów ASR (dla każdego wariantu długości segmentu treningowego).” Czy Doktorant wyznaczał statystyczną istotność różnic wartości skuteczności identyfikacji mówcy, przedstawionych na tych rysunkach? Podobne pytanie odnosi się do wyników przedstawionych na rys. 6.6-6.8.

3.3. Spadek efektywności działania proponowanego rozwiązania w obecności szumów i zakłóceń.

W kolejnych podrozdziałach rozdziału 6 Doktorant opisał badania zachowania systemu ASR w warunkach istnienia addytywnych zakłóceń nałożonych na sygnał mowy: szumu białego o różnej wariancji (SNR wynosił odpowiednio 20, 15 oraz 10 dB) oraz w obecności zakłóceń w postaci nagrania tłumu i nagrania ulewnego deszczu, o różnej intensywności (zakłócenia „słyszalne”, „uciążliwe” i „przeważające”, jakkolwiek Doktorant nie podaje liczbowych wartości parametrów, charakteryzujących te poziomy zakłóceń). W niektórych przypadkach wartości

sprawności systemów ASR (systemu ARMiA oraz jego modyfikacji, zaproponowanych przez Doktoranta) spadają poniżej 50%, czyli poniżej poziomu losowego wyboru. Doktorant słabo skomentował w treści rozprawy te wyniki, nie podjął się dokonać analizy przyczyn tak znacznej degradacji sprawności systemu (ponad 95% w warunkach bezszumowych, rys. 6.2), ani nie zaproponował kierunków przyszłych modyfikacji systemu, pozwalających na poprawę skuteczności jego działania.

3.4. Zbiór cech behawioralnych występujących w systemie ASR.

W dodatku „Z1. Zbiór definicji wszystkich cech behawioralnych występujących w systemie ASR” Doktorant definiuje zbiór 71 cech, używanych (w różnej liczebności, w różnych podzbiorach tych cech) do klasyfikacji/rozpoznania mówcy. Niektóre z tych cech są silnie skorelowane, a nawet zależne – na przykład cechy c_2 i c_4 , cechy c_{18} i c_{22} . Niestety, w treści rozprawy trudno jest znaleźć informację, czy te zależne cechy nie tworzą w eksperymentach numerycznych przestrzeni wejść klasyfikatorów, co z punktu widzenia działania algorytmów optymalizacyjnych nie byłoby właściwym podejściem. Zdaniem Recenzenta problem budowy przestrzeni wejść klasyfikatorów z wektorów zależnych liniowo i nieliniowo oraz ewentualnego wpływu takiej konstrukcji na rozwiązanie problemu powinien zostać przedyskutowany podczas obrony rozprawy.

3.5. Niekonsekwencje w używaniu symboli we wzorach i tekście rozprawy.

W wielu miejscach rozprawy Doktorant używa tych samych symboli do oznaczania różnych wielkości fizycznych – często w tym samym podrozdziale tekstu. Przykładem może być tu użycie symbolu ‘N’ do oznaczania długości ciągu, liczby ramek, itp., co niekiedy zmniejsza czytelność tekstu, utrudniając śledzenie myśli Autora (przykład z rozdziału 2.2 został opisany w dalszej części recenzji).

Jeszcze większą wadą jest użycie tego samego symbolu do oznaczania dwóch zmiennych – oryginalnej i transformowanej. Dla przykładu – wzór (3.1) na str. 49 potraktowany jako równanie matematyczne prowadzi do wniosku, że suma N próbek sygnału $x(n)$ jest zerowa. Z kolejnego wzoru potraktowanego jako równanie (wzór (3.2)) wynika jedynie, że maksymalna wartość modułów próbek ciągu czasowego $x(n)$ jest jednostkowa. Można przypuszczać, że w/w wzory w tekście rozprawy są mechanicznym przeniesieniem instrukcji kodu programu, jakkolwiek taki sposób kodowania również nie jest dobrą praktyką, nawet jeśli jest stosowany w ciele funkcji.

Ponadto dla formalnej poprawności wzoru (3.2) brakuje wyjaśnienia, w jaki sposób Doktorant skaluje segment „kompletnej” ciszy, tzn. ciąg próbek, w którym $\max_{0 \leq n \leq N-1} (|x(n)|) = 0$.

3.6. Kryteria detekcji występowania sygnału mowy.

Na str. 49 Doktorant formułuje kryterium detekcji aktywności głosowej mówcy: „Jako ramki zawierające sygnał mowy przyjmowane są ramki spełniające zależność $E_i > p_r \bar{E}$ ”, gdzie symbole ‘E’ odnoszą się do energii sygnału. Tymczasem na str. 51, w celu detekcji występowania mowy, Doktorant formułuje kryterium oparte na mocy sygnału: „Drugim etapem selekcji ramek jest ponowna detekcja występowania mowy w sygnale głosowym. ... W tym przypadku autor wykorzystał nierówność (3.6) z empirycznie wyznaczonym progiem dla kryterium mocowego $P_i > p_p P_s$ ”.

Z jakich powodów Doktorant stosuje dwa kryteria (wprawdzie związane ze sobą) do detekcji aktywności mówcy?

1. Sielga

3.7. Niewystarczający komentarz dotyczący wyznaczania parametru sygnału mowy.

Na rys. 41 (str. 65) został przedstawiony przebieg tempa przekroczeń zera trzysekundowego fragmentu sygnału mowy, wyznaczonego według wzoru (4.1). Krzywa na rysunku jest ciągła, co sugeruje, że wartość tego parametru została wyznaczona dla każdej chwili próbkowania sygnału mowy. Jednakże nie jest jasne, czy (jeśli powyższe spostrzeżenie o ciągłości krzywej jest prawdziwe), w jaki sposób Doktorant określa długość 'N' ciągu próbek – czy 'N' oznacza liczbę próbek od początku obserwacji (a więc wzrasta z każdą chwilą próbkowania), czy jest stałą (a więc w ramce przesuwanej w każdej chwili czasu zawarta jest stała liczba próbek sygnału)?

Ponadto prawa strona równości (4.1) nie jest poprawna – dla wartości $n = 0$ następuje odwołanie do próbki sygnału $x(-1)$, która nie jest zdefiniowana.

3.8. Proces ekstrakcji cech – wyznaczanie „najbardziej znaczących maksimów”.

Na str. 65-66 Doktorant omawia proces ekstrakcji cech, ilustrując ten proces schematem blokowym przedstawionym na rys. 4.3. Istotnym elementem decyzyjnym w procesie ekstrakcji jest porównanie wartości ekstremalnych (wyznaczonych z sygnału) z wartościami progowymi (zapewne określonymi w jakiś – mniej lub bardziej – arbitralny sposób). Wydaje się również, że wartości progowe mogą być parametrami identyfikującymi mówcę, a przynajmniej zależnymi od fizycznych charakterystyk jego mowy. Niestety, w treści rozdziału 4.2.6 Doktorant nie podaje sposobu określenia wartości progowych.

3.9. „Niedouczenie” i „przeuczenie” modelu.

Na str. 84-85 Doktorant prowadzi rozważania na temat ogólnych parametrów, które charakteryzują jakość modeli uzyskiwanych metodami uczenia z danych: błędu systematycznego oraz wariancji. Ilustrując swoje rozważania rysunkami 5.3 i 5.4, Doktorant używa terminów: „model niedouczony” i „model przeuczony”, które sugerują, że właściwości aproksymacyjne modeli, ukazane na w/w rysunkach, mają coś wspólnego z czasem uczenia (liczbą iteracji wykonanych przez algorytm uczenia). Tymczasem przebieg linii prostej na rys. 5.3 (reprezentującej liniowy model zależności wejściowo-wyjściowej) nie ma nic wspólnego z długością procesu uczenia – najprawdopodobniej współczynniki modelu liniowego, którego nie da się bardziej „nauczyć”, zostały wyznaczone za pomocą metody najmniejszych kwadratów, nie wymagającej iteracji. Podobnie – krzywa pokazana na rys. 5.4 ilustruje zachowanie aproksymacyjne modelu zbyt wysokiego rzędu (tzw. *overfitting*), a efekt nadmiernego dopasowania do danych, który rzeczywiście ujawnia się najczęściej z postępu procesy treningu i jest nazywany „przeuczeniem”, jest zjawiskiem wtórnym i powinien być sygnałem do zmiany rzędu modelu. Temat ten, potraktowany przez Doktoranta w zbyt uproszczony sposób, jest rozważany w wielu pozycjach literatury, m.in.: S. Haykin: „Neural Networks and Learning Machines. Third Edition”, Pearson Prentice Hall, 2009.

Autor nie ustrzegł się również drobnych usterek (w tym usterek redakcyjnych), skrótów myślowych i uproszczeń, wymienionych poniżej:

- Str. 7 – angielski termin „*Vector Quantization*” w polskiej literaturze występuje na ogół jako „kwantyzacja wektorowa”.
- Str. 15 – w tytułach kolumn Tabeli 1.1 Doktorant używa nazw rodzajów biometryk („Uniwersalność”, „Unikatowość”, „Mierzalność” itp.) nie wyjaśniając, co oznaczają te nazwy. Ponadto oceny tych biometryk („Niska”, „Średnia”, „Wysoka”) sugerują, że na przykład „Unikatowość” można ocenić w sposób ilościowy.

HSi42

- Str. 21 – trudno jest zrozumieć znaczenie zamieszczonego w nawiasie fragmentu zdania: „Często charakterystyki te nazywa się prozodycznymi bądź prozodyjnymi (w nawiązaniu do brzmieniowych właściwości cech).”.
- Str. 26 – rysunek 1.10 został dość pobieżnie skomentowany w tekście rozprawy; w szczególności nie jest jasne, jakie wnioski można wyciągnąć na podstawie przebiegu residuów predykcji liniowej sygnału oraz filtracji wszechpasmowej (o której brak wzmianki w tekście).
- Str. 26 – w jednym z bloków funkcjonalnych przedstawionych na rys. 1.11 Doktorant używa terminu „Okienkowanie zsynchronizowane ze skokiem”, nie omawiając znaczenia tego sformułowania.
- Str. 29 – nie jest jasne, dlaczego oś czasu na lewym wykresie rys. 1.15 została wyskalowana w sekundach, a na prawym wykresie – w milisekundach.
- Str. 32 – Doktorant niepotrzebnie poświęca uwagę wyjaśnianiu i omawianiu elementarnych pojęć dotyczących techniki rejestracji sygnału mowy („Urządzenie takie nosi nazwę mikrofonu ...”), podczas gdy w tekście rozprawy często występują terminy niezdefiniowane.
- Str. 34 – kolejność czterech wykresów składających się na rys. 2.3 nie odpowiada kolejności opisu poszczególnych etapów przetwarzania sygnału, zamieszczonego w akapicie powyżej tego rysunku (na str. 33).
- Str. 34 – w ostatnim zdaniu akapitu tekstu poniżej rys. 2.3 (jak również w kolejnym akapicie) Doktorant używa terminu „szybkość próbkowania”, zamiast „częstotliwość próbkowania”.
- Str. 34 – co oznacza sformułowanie „Wartości sygnału reprezentowane są na 16 bitach w zapisie stałoprzecinkowym.”?
- Str. 35 – we wzorze (2.2) nie wyjaśniono znaczenia indeksu i .
- Str. 36 – nie jest jasne, czy znaczenie symbolu N (górny zakres sumowania we wzorze (2.3)) jest takie samo, jak znaczenie tego symbolu we wzorze (2.2). Czy „środek ciężkości widma” (symbol SC) jest tym samym, co częstotliwość środkowa tego widma?
- Str. 36 – w akapicie rozpoczynającym się od sformułowania „Otrzymane wartości parametrów porównywane są z odpowiednimi progami E_i oraz SS_i ” brak jest wyjaśnienia, według jakich kryteriów i w jaki sposób są ustalane te progi.
- Str. 36 – w linii poniżej wzoru (2.5) Doktorant błędnie wyjaśnia: „gdzie T_p to częstotliwość próbkowania”, zamiast „okres próbkowania”, a ponadto fizyczne znaczenie parametrów stanowiących lewe strony równań (2.5) nie zostało wyjaśnione.
- Str. 37-38 – Doktorant nie wyjaśnia pojęć „listek główny” i „listki boczne” okien czasowych, co wydaje się pewną niekonsekwencją wobec wcześniejszych wyjaśnień terminu „mikrofon”.
- Str. 39 – w pierwszym akapicie rozdziału 2.2.2 powinno być „kilku przykładowych”.
- Str. 40 – w przedostatnim akapicie rozdziału 2.2.2 Doktorant używa symbolu N jako liczby zastosowanych filtrów, podczas gdy ten symbol był już kilkakrotnie wykorzystywany w treści rozprawy do oznaczenia innych wielkości.
- Str. 42 – Doktorant używa sformułowania „wartość logarytmu z ilorazu liczby ramek przypadających na konkretne słowo” – czy chodzi po prostu o liczbę ramek przypadających na słowo?
- Str. 42 – w pierwszym zdaniu rozdziału 2.3 powinno być „coraz”.
- Str. 43 – w tabeli 2.1 powinno być: „mała ilość” (zamiast „niska ilość”), „mały wpływ” (zamiast „niski wpływ”), jak również, w odpowiednich sformułowaniach, „duży” zamiast „wysoki”.
- Str. 50 – błędna numeracja wzorów, mająca konsekwencje w całej dalszej części tekstu rozprawy – wzór (3.4) wystąpił już w ostatniej linii tekstu na str. 49.
- Str. 51 – we wzorze (3.7) występują niezdefiniowane symbole F_{0c} i F_{0ac} . Ponadto, wobec braku tych objaśnień nie jest oczywiste, że w każdym przypadku wyrażenie po prawej stronie nierówności (3.7) jest nieujemne – w przeciwnym przypadku nierówność (3.7) byłaby sprzeczna.

- Str. 59 – nieznane jest znaczenie indeksu j w liczniku ułamka po prawej stronie równania (3.24), nie jest również jasne, dlaczego ten indeks nie występuje po lewej stronie równości.
- Str. 64 – nie została określona liczba „setek” milisekund w sformułowaniu: „dokonywany jest również podział na krótsze segmenty o czasie trwania rzędu setek milisekund”, ani też, według jakich kryteriów ta liczba jest ustalana.
- Str. 70 – każdy z rysunków 4.6 i 4.7 przedstawia zbiór 6 punktów we współrzędnych czas-częstotliwość środkowa widma; punktów tych nie powinno się łączyć linią ciągłą (tworząc łamaną), gdyż takie połączenie sugeruje, że na przykład wartość częstotliwości środkowej widma dla przykładowego czasu 0,75 [s] jest średnią odpowiednich wartości dla czasów 0,5 [s] i 1,0 [s], co nie jest prawdą.
- Str. 75 – w ostatnim zdaniu rozdziału 4.3 Doktorant podaje zbiór cech sygnału, wykorzystanych do dalszych badań, używając jedynie symboli (C_j), które bez wyjaśnienia ich fizycznego znaczenia nie dają czytelnikowi rozprawy praktycznie żadnej informacji.
- Str. 75 – z konstrukcji wzoru (4.20) wynika, że Doktorant operuje wierszową postacią wektorów – informacja o takiej konwencji nie pojawiła się dotychczas w treści rozprawy.
- Str. 76 – nie jest jasna myśl, zawarta w zdaniu: „W odróżnieniu od najpopularniejszej metryki euklidesowej, odległość typu Manhattan minimalizuje wpływ pojedynczych dużych różnic, poprzez brak ich potęgowania [1]”.
- Str. 81 – w zdaniu „Model związany z abstrakcją danych wejściowych i/lub wyjściowych można przedstawić za pomocą rozszerzonego 5 stopniowego schematu Dasarathy’ego.” brak jest odwołania do odpowiedniej pozycji bibliograficznej.
- Str. 81 – sformułowanie „w celu uzyskania bardziej precyzyjnego pomiaru średniej temperatury” jest niepoprawne, gdyż nie można zmierzyć wartości średniej serii (ciągu) pomiarów – wartość tę uzyskuje się w wyniku obliczeń.
- Str. 85 – nie jest jasne, co Autor rozumie pod pojęciem klasyfikatora „tego samego typu”.
- Str. 86 – podobnie nie jest jasne, co Autor rozumie pod pojęciem „mniejszych klasyfikatorów”.
- Str. 93 – w jaki sposób uzyskano wartości liczbowe elementów macierzy i jaki cel ma pokazywanie tych liczb?
- Str. 113 – nie jest jasna myśl wyrażona w zdaniu: „Opracowano dwie autorskie metody integracji danych, tj. na poziomie cech oraz na poziomie decyzji.”.
- Str. Z.1.3 – symbol t_N występujący w mianowniku wyrażenia (Z.1.5) nie jest zdefiniowany.

Ponadto Doktorant używa w rozprawie słowa „wyekstrahowanie”, odmieniając je przez wszystkie przypadki, co niekiedy brzmi dość sztucznie i niezgrabnie. Czy polski termin „wyodrębnienie” nie oddaje w pełni istoty działań, które prowadziły do uzyskania cech sygnału mowy, używanych w zadaniach detekcji/klasyfikacji?

4. Podsumowanie oceny rozprawy doktorskiej

Treść rozprawy dotyczy bardzo złożonych zagadnień klasyfikacji sygnału mowy, które mają istotne znaczenie praktyczne w systemach automatycznego rozpoznawania mówcy (ASR). Zdaniem recenzenta, Doktorant, wykorzystując oryginalne, autorskie koncepcje przetwarzania sygnałów w celu ekstrakcji ich tzw. cech dystynktywnych, algorytmy uczenia maszynowego oraz techniki klasyfikacji wielowymiarowych zbiorów danych, za pomocą ogromnej liczby eksperymentów symulacyjnych potwierdził słuszność tezy sformułowanej na str. 11 rozprawy. Zadania badawcze, których rozwiązania podjął się Doktorant, charakteryzują się poziomem merytorycznym w pełni adekwatnym do wymagań stawianych rozprawom doktorskim i zostały sformułowane i rozwiązane w poprawny sposób. Doktorant w sposób wyczerpujący, na ogół jasny i precyzyjny przedstawił w treści rozprawy oryginalne algorytmy dekompozycji sygnału mowy, ekstrakcji i klasyfikacji jego cech. Pewne zastrzeżenia recenzenta,

sformułowane w poprzedniej sekcji recenzji, budzi niekiedy szczupłość komentarzy na temat wyników eksperymentów numerycznych oraz jakość analizy tych wyników i wniosków, które wynikają z wykonanych eksperymentów.

W treści rozprawy można w sposób jednoznaczny wyodrębnić oryginalny i samodzielny dorobek Autora. Rozprawa zawiera również zestawienie i podsumowanie osiągnięć własnych Doktoranta, które nie wzbudza żadnych zastrzeżeń ze strony recenzenta. Można również uznać, że wybrane rezultaty osiągnięte przez Autora, opublikowane w czasopiśmie „Inżynieria Bezpieczeństwa Obiektów Antropogenicznych”, zostały pozytywnie zweryfikowane przez specjalistów, zajmujących się tematyką odpowiadającą zakresowi rozprawy doktorskiej.

W opinii recenzenta, rozprawa stanowi samodzielne i oryginalne rozwiązanie problemu badawczego, potwierdzając tym samym odpowiednie przygotowanie Doktoranta do prowadzenia działalności badawczej. Poziom merytoryczny rozprawy świadczy o dużej wiedzy teoretycznej i umiejętnościach Doktoranta w zakresie podstawowej dyscypliny naukowej, do której należy tematyka rozprawy, a także o odpowiednio wysokim poziomie znajomości zagadnień z pokrewnych obszarów nauki, związanych z treścią rozprawy. Redakcja pracy jest na ogół staranna (usterki i błędy redakcyjne, zauważone przez recenzenta, zostały wymienione powyżej), zaś układ rozprawy można uznać za ogólnie poprawny.

Uwagi polemiczne oraz krytyczne komentarze, zamieszczone w poprzednim rozdziale recenzji, wynikają przede wszystkim ze stopnia złożoności interdyscyplinarnego problemu naukowego, którego rozwiązywania podjął się Doktorant. Recenzent stwierdza jednoznacznie, że rozprawa wnosi interesujący przyczynek do metodologii budowy inteligentnych narzędzi, przeznaczonych do automatycznego rozpoznawania mówcy, pogłębiając dotychczasową wiedzę teoretyczną w tym zakresie oraz wzbogacając zbiór rozwiązań praktycznych. Mocną stroną rozprawy jest weryfikacja uzyskanego rozwiązania w różnorodnych warunkach, jak najbardziej zbliżonych do rzeczywistych warunków, w jakich powinny pracować systemy ASR – pomimo pewnych uwag krytycznych, jakie zostały sformułowane powyżej. Recenzent wyraża również nadzieję, że przedstawione w recenzji uwagi polemiczne są warte rozważenia i ewentualnego uwzględnienia w dalszych pracach Doktoranta – nie wpływają one jednak w istotny sposób na pozytywną ocenę opiniowanej rozprawy doktorskiej.

Oceniając wagę pomysłów doktoranta na modyfikację i poprawę działania systemu ASR (ARMiA) w kontekście złożoności problemu i jakości proponowanego rozwiązania, jak również staranność zaplanowania i przeprowadzenia przez Doktoranta eksperymentów numerycznych, rozprawa zasługiwałaby na wyróżnienie. Jednak szereg uwag krytycznych, zamieszczonych w poprzedniej sekcji recenzji, powstrzymuje recenzenta przed jednoznacznym sformułowaniem takiego wniosku. Recenzent zastrzega sobie jednak możliwość sformułowania wniosku o wyróżnienie rozprawy w zależności od przebiegu dyskusji podczas publicznej obrony rozprawy doktorskiej.

5. Wniosek końcowy

Zgodnie z zaprezentowaną powyżej oceną merytoryczną rozprawy Pana por. mgr. inż. Dominika Małego stwierdzam, że opiniowana praca spełnia warunki i wymagania stawiane rozprawom doktorskim, określone w ustawie z dnia 18 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (tj. Dz. U. z 2024 r., poz. 1751, z późn. zm.).

W związku z Art. 186 ust. 1 oraz Art. 187 ust. 1 i 2 w/w ustawy stawiam zatem wniosek o dopuszczenie rozprawy doktorskiej por. mgr. inż. Dominika Małego do publicznej obrony, zgodnie z procedurą określoną przez Radę Dyscypliny Naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Wojskowej Akademii Technicznej.