

Poznań, 7.07.2025 r.

Prof. dr hab. inż. Adam Dąbrowski
Politechnika Poznańska
Wydział Automatyki, Robotyki i Elektrotechniki
Instytut Automatyki i Robotyki
Zakład Układów Elektronicznych
i Przetwarzania Sygnałów

OCENA

rozprawy doktorskiej pt.: „*System automatycznego rozpoznawania mówcy wsparty behawioralnymi cechami sygnału mowy*”
Doktorant por. mgr inż. Dominik Mały

1 Ocena trafności doboru tematyki pracy, określenia problemu naukowego oraz celu i zakresu przeprowadzonych badań

Tematem ocenianej rozprawy doktorskiej, napisanej w języku polskim, jest opracowanie systemu automatycznego rozpoznawania mówcy (ang. automatic speaker recognition, ASR) zawierającego, oprócz klasycznych biometrycznych cech fizycznych mowy, dobrze dobrany zbiór cech behawioralnych.

Opracowana i przetestowana przez Doktoranta technika umożliwia zwiększenie poprawności rozpoznawania mówcy, zwłaszcza w trudnych warunkach akustycznych.

Podjęte i przedstawione w ocenianej rozprawie prace badawcze, programistyczne i eksperymentalne zaliczam do bardzo ważnej i w pełni aktualnej tematyki, która ma coraz większe znaczenie w związku z rozwojem zastosowań biometrii i ogólnie — sztucznej inteligencji.

Tematyka rozprawy została trafnie przypisana do dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne.

Na stronie 8 rozprawy w streszczeniu Doktorant podał następujący cel badawczy:

„Celem niniejszej rozprawy było opracowanie i implementacja w istniejącym systemie ASR zbioru cech behawioralnych umożliwiających zwiększenie liczby poprawnych identyfikacji tożsamości mówców, w szczególności w utrudnionych warunkach rejestracji.”

Doktorant sformułował także następującą tezę naukową (także na str. 8 oraz ponownie na str. 11 i str. 45):

„możliwe jest zdefiniowanie i wyekstrahowanie z sygnału mowy takich cech behawioralnych, które w połączeniu z cechami fizycznymi zaimplementowanymi w istniejącym rozwiązaniu znacząco poprawią jego skuteczność w obecności zewnętrznych zakłóceń.”

Uważam, że zarówno cel pracy jak i teza naukowa są zgodne z podjętym tematem i zakresem przeprowadzonych badań. Sformułowane są jasno i precyzyjnie, chociaż teza jest niepotrzebnie powtarzana w różnych miejscach tekstu rozprawy.

Tytuł rozprawy trafnie określa zakres podjętych i przeprowadzonych badań, które dotyczą, jak już to wcześniej zaznaczyłem, ważnego i w pełni aktualnego a nawet w dużym stopniu nadal nowatorskiego zagadnienia naukowego o dużych walorach aplikacyjnych. Doktorant przeprowadził skrupulatnie analizę zastanego stanu wiedzy oraz opisał własne pomysły, prace programistyczne i eksperymenty wykonane w różnorodnych warunkach.

2 Ocena treści, tekstu rozprawy, analizy źródeł i dorobku publikacyjnego Doktoranta

Recenzowana rozprawa, jak już wspomniałem, została napisana w języku polskim. Zawiera 125 stron tekstu głównego oraz 55 osobno numerowanych stron z załącznikami, jest więc dość obszerna.

Na stronach 3–5 tekstu głównego jest spis treści. Indeks skrótów mieści się na str. 6 i 7. Powinien on sąsiadować z listą używanych symboli, ale tej w pracy niestety brakuje. Na stronie 8 tekstu rozprawy jest streszczenie w języku polskim a na stronie 9 — streszczenie w języku angielskim. Na stronach 10 i 11 Doktorant zawarł przedmowę do rozprawy i na str. 12 zamieścił podziękowania.

Właściwa część tekstu rozprawy rozpoczyna się od rozdziału 1, który jest wprowadzeniem do zagadnień rozważanych w rozprawie, zawierającym opis modelowania procesów generowania sygnału mowy a także ogólniej — ludzkiego głosu. W tym rozdziale Doktorant zaproponował także klasyfikację systemów automatycznego rozpoznawania mowy oraz przeprowadził obszerny przegląd literatury.

W rozdziale 2 pt.: „Proces przetwarzania sygnału mowy w aspekcie automatycznego rozpoznawania mowy” Doktorant omówił cechy fizyczne i behawioralne zawarte w sygnale mowy oraz podstawowe procesy przetwarzania sygnału mowy, tj. procesy rejestracji i segmentacji. Opisał także wykorzystaną przez Niego bazę nagrań głosów SLR-12.

Rozdział 3 pt.: „System odniesienia oparty na cechach fizycznych” zawiera omówienie standardowego rozwiązania rozpoznawania mowy, rozpoznawanie tzn. wyłącznie za pomocą analizy cech fizycznych głosu mowy.

Natomiast rozdział 4 pt.: „System oparty na cechach behawioralnych” został już poświęcony opisaniu pomysłu Doktoranta, polegającego na wprowadzeniu cech behawioralnych do systemu ASR.

Połączenie obu typów cech w jednym systemie ASR (tj. cech fizycznych i behawioralnych) zostało przedstawione w kolejnym rozdziale 5 pt.: „Fuzja systemów”.

Rozdział 6 pt.: „Badania eksploatacyjne systemu ASR” zawiera opis wykorzystanych baz danych oraz testy opracowanych rozwiązań w warunkach bezszumowych, w zróżnicowanych warunkach szumowych oraz w zróżnicowanych warunkach akustycznych przy zróżnicowanych jakościach nagrań.

W ostatnim rozdziale 7 pt.: „Konkluzja” Doktorant opisał i scharakteryzował wyniki otrzymane za pomocą opracowanego przez siebie systemu SARM-WAT-B oraz porównał je z wynikami uzyskiwanymi za pomocą systemu ARMiA, wcześniej opracowanego w zespole Promotora Doktoranta Pana Profesora Andrzeja Dobrowolskiego przez Promotora pomocniczego Pana mjr. dr. inż. Kamila Kamińskiego.

Główny tekst rozprawy można podzielić na dwie wzajemnie powiązane części.

W pierwszej z nich, na którą składają się: rozdział 1, punkty 2.1 i 2.2 rozdziału 2 i rozdział 3, Doktorant przedstawił standardowe rozwiązanie systemów ASR oparte na badaniu cech fizycznych sygnału mowy, uwzględniając system ARMiA oparty na analizie cepstralnej sygnału mowy i modelach mieszanin gaussowskich. Ta część jest wprowadzeniem do tematyki rozprawy i pełni funkcję pomocniczą.

Część drugą, stanowiącą zasadniczą, oryginalną część rozprawy, w której Doktorant zawarł swoje wyniki badawcze, polegające zarówno na wykorzystaniu cech fizycznych głosu jak na uwzględnieniu zaproponowanych przez Doktoranta cech behawioralnych, tworzą: punkty 2.2.3, 2.2.4, 2.3 i 2.5 rozdziału 2 oraz rozdziały 4, 5, 6 i 7. Pan por. mgr inż. Dominik Mały zaproponował ponadto sposób fuzji obu typów danych, a następnie zaprezentował wyniki testów opracowanego przez Niego systemu w warunkach istnienia zewnętrznych zakłóceń oraz zastosowania różnych torów akustycznych służących do rejestracji i transmisji sygnału mowy.

Spis literatury zawierający łącznie 158 pozycji (w tym 10 adresów stron internetowych), które dobrze i wyczerpująco przedstawiają aktualny stan wiedzy i badań w zakresie objętym rozprawą, zajmuje str. 114–125 tekstu rozprawy.

Spis ten zawiera tylko jedną publikację Doktoranta (pozycja [1] w rozprawie i pierwsza pozycja w poniższym zestawieniu), choć Pan por. mgr inż. Dominik Mały jest współautorem siedmiu następujących publikacji:

- w czasopismach z listy A

1. D. Mały, A. Dobrowolski, Behavioral features of the speech signal as part of improving the effectiveness of the automatic speaker recognition system, *Inżynieria Bezpieczeństwa Obiektów Antropogenicznych*, vol. 4, 2023, pp. 26–34, DOI: 10.37105/iboa.187, (140 pkt.)

- w czasopismach z listy B

1. D. Mały, A. Dobrowolski, Detekcja faz oddechu w rejestracjach stetoskopowych, *Elektronika: konstrukcje, technologie, zastosowania*, nr 6/2019, ss. 4–7
2. D. Mały, A. Dobrowolski, Ekstrakcja danych numerycznych z wykresów słuchowych potencjałów wywołanych, *Elektronika: konstrukcje, technologie, zastosowania*, nr 3/2020, s. 4–7

- w materiałach konferencji międzynarodowych

1. D. Mały, A. Dobrowolski, Extraction of numerical data from auditory evoked potentials charts, The Students' International Conference "CERC 2020", Bucharest, 06–07.11.2020

- w materiałach konferencji krajowych

1. D. Mały, A. Dobrowolski, Detekcja faz oddechu w rejestracjach stetoskopowych, XXXV Konferencja Elektroniki, Telekomunikacji i Energetyki Studentów i Młodych Pracowników Nauki (SECON'19), Warszawa, 25–26.04.2019



2. D. Mały, A. Dobrowolski, Behawioralne cechy sygnału mowy jako element poprawy skuteczności systemu automatycznego rozpoznawania mowy, XXXIV Konferencja Naukowo-Techniczna Inżynieria Bezpieczeństwa — Ochrona przed Skutkami Nadzwyczajnych Zagrożeń „Ekomilitaris 2023”, Zakopane, 19–22.09.2023
3. A. Dobrowolski, E. Majda-Zdancewicz, D. Mały, K. Kamiński, Ocena przydatności fizycznych i behawioralnych cechy mowy do badań fonoskopijnych, Krajowa Konferencja Metrologii (KKM 2025), Waplewo, 26–29.05.2025 r.

Biorąc pod uwagę powyższe zestawienie prac, dorobek publikacyjny Doktoranta oceniam jako wartościowy i dużym stopniu wspierający ocenianą rozprawę.

Dalszy tekst rozprawy stanowią załączniki. Załącznik Z.1 zawiera zbiór definicji wszystkich cech behawioralnych zastosowanych przez Doktoranta w opracowanym przez Niego systemie ASR. Załącznik Z.2 jest dokumentacją ewolucji opracowanego systemu ASR. W załączniku Z.3 są podane wyniki przeprowadzonych eksperymentów w zróżnicowanych warunkach i środowiskach akustycznych (np. w przypadku ulewnego deszczu czy przy obecności odgłosów tłumu). Rozprawę zamyka załącznik Z.4, który jest opisem opracowanego oprogramowania SARM-WAT-B (System Automatycznego Rozpoznawania Mówcy Wsparty ATrybutami Behawioralnymi).

Kompozycja pracy jest prawidłowa. Tekst jest jasny, rzeczowy i treściwy. Doktorant najpierw przeprowadził skrupulatnie analizę zastanego stanu wiedzy a następnie opisał własne pomysły, badania, eksperymenty i ich wyniki.

3 Ocena uzyskanych wyników i przeprowadzonych eksperymentów oraz poprawności ich przedstawienia

Pan por. mgr inż. Dominik Mały zaproponował i przetestował wykorzystanie zestawu licznych, odpowiednio dobranych cech behawioralnych głosu do automatycznego rozpoznawania mowy.

Do głównych zadań podjętych i wykonanych przez Doktoranta oraz do najważniejszych uzyskanych przez Niego wyników należy, moim zdaniem, zaliczyć:

- określenie głównego kierunku badań w zakresie tworzenia właściwego zbioru cech behawioralnych na podstawie analizy działania systemu ASR ARMiA autorstwa Promotora pomocniczego rozprawy Pana mjr. dr. inż. Kamila Kamińskiego
- utworzenie (selekcja za pomocą algorytmu genetycznego) zbioru 27 cech behawioralnych głosu, stanowiących podstawę działania systemu ASR opracowanego przez Doktoranta
- wybór efektywnego zestawu najbardziej dystynktywnych cech behawioralnych spośród wszystkich zdefiniowanych wstępnie 71 cech za pomocą metod ewolucyjnych
- opracowanie metody ekstrakcji cech opartej na przetwarzaniu nakładających się ramek sygnału mowy
- propozycję metody integracji danych wykorzystującej ich fuzję zarówno na poziomie decyzji jak i na poziomie cech
- eksperymentalną weryfikację poprawności opracowanych rozwiązań ASR w zróżnicowanych warunkach akustycznych i zróżnicowanych jakościach rejestracji sygnałów.

Wymienione powyżej osiągnięcia Doktoranta oceniam wysoko. Świadczą one nie tylko o Jego pomysłowości i dużej wiedzy w zakresie automatycznego rozpoznawania mówcy, ale także o systematyczności, skrupulatności, pracowitości i skuteczności osiągnięcia założonego celu badawczego.

Język pracy jest poprawny a edycja tekstu — staranna. Występują tylko nieliczne, mało znaczące błędy gramatyczne, np. na str. 19 „próbka głosu znajduje się” zamiast „próbka głosu znajduje się” i „występuje układ decyzyjny” zamiast „występuje układ decyzyjny”.

Można mieć jednak pewne zastrzeżenia wobec stosowanego nazewnictwa, występującego np. w indeksie skrótów na str. 6 i 7. Nazywanie „maszyny wektorów nośnych” (ang. support vector machine, SVM) „siecią neuronową wektorów nośnych” jest kontrowersyjne. Działanie tzw. „maszyny wektorów nośnych” jest oparte na ściśle zdefiniowanym algorytmie uczenia maszynowego, przydatnym do rozwiązywania zadań klasyfikacji liniowej lub nieliniowej, regresji, a nawet wykrywania danych błędnych (tzw. „odstających” od pozostałych). W odróżnieniu od SVM działanie sieci neuronowych jest oparte na heurystyce. Liczba parametrów SVM rośnie liniowo wraz ze wzrostem rozmiaru problemu. Natomiast wielkość struktury sieci neuronowej nie jest bezpośrednio związana z rozmiarem problemu, ponieważ sieć neuronowa może mieć dowolną liczbę warstw.

Określenia są stosowane bez należytej konsekwencji. Na przykład skrót DFT (discrete Fourier transform) jest tłumaczony jako „dyskretna transformacja Fouriera a skrót FT (Fourier transform) — jako „transformata Fouriera”.

Ponadto tłumaczenie na język polski określenia HMM (hidden Markov model) jako „ukryty model Markowa” jest niezręczne a nawet niepoprawne, ponieważ to nie sam model jest ukryty. Chodzi bowiem o jawny a nie ukryty model systemu, w którym obserwuje się przechodzenie między szeregiem ukrytych stanów, przy jednoczesnej obserwowalności wyników. Zatem poprawna nazwa to „model Markowa o ukrytych stanach”.

Wyrażenia matematyczne są na ogół poprawne, choć nie wszystkie. Poza tym nazewnictwo i w przypadku wyrażen matematycznych bywa błędne. Na przykład wyrażenia (2.1) na str. 35 są wzorami a Doktorant nazywa je równaniami. Pierwsze z nich

$$x_i = \frac{x_i - \bar{x}}{\sqrt{s}}$$

można by nawet potraktować jako równanie z niewiadomą x_i i wtedy, jako takie, byłoby poprawne, ale tak nie jest. Naprawdę to nie jest jednak równanie a wzór na znormalizowaną i pozbawioną składowej średniej i -tą próbkę sygnału, która została błędnie oznaczona tak samo jak próbka nieznormalizowana po prawej stronie znaku równości. Taki zapis jest stosowany w językach programistycznych w znaczeniu „podstaw jako”, ale jest niedopuszczalny w wyrażeniach matematycznych.

Podobne błędy występują także w innych miejscach tekstu rozprawy, np. w wyrażeniach (3.1) i (3.2) na str. 49.

Ponadto nietypowe oznaczenia: $^{-2}\sqrt{2}$, $^3\sqrt{2}$, $^6\sqrt{2}$ i $^{-6}\sqrt{2}$ zastosowane we wzorach (4.3), (4.5) i w tekście na str. 67 można by zastąpić znacznie bardziej czytelnymi, odpowiednio: $2^{-1/2}$, $2^{1/3}$, $2^{1/6}$ i $2^{-1/6}$.

Podsumowując podkreślam, że uzyskane przez Doktoranta wyniki zasługują na pozytywną ocenę. Mają nie tylko wartość poznawczą, ale przede wszystkim duże walory aplikacyjne. Doktorant rozwiązał postawione problemy, dowodząc, że posiadał umiejętności związane z metodyką i metodologią prowadzenia badań naukowych.

4 Konkluzja

Podsumowując powyższą charakterystykę recenzowanej rozprawy doktorskiej stwierdzam, że Pan por. mgr inż. Dominik Mały zamieścił w niej wartościowe i oryginalne wyniki prac przeprowadzonych przez Niego pod kierunkiem Promotora rozprawy Pana Profesora dr. hab. inż. Andrzeja Dobrowolskiego i Promotora pomocniczego Pana mjr. dr. inż. Kamila Kamińskiego.

Oceniam, że Doktorant osiągnął zakładany cel badawczy i wykazał prawdziwość postawionej tezy naukowej. Stwierdzam więc, że przedłożona praca bardzo dobrze spełnia wymagania stawiane przez stosowne przepisy rozprawom doktorskim i uważam, że Pan por. mgr inż. Dominik Mały nie tylko powinien być dopuszczony do dalszych etapów procedury doktorskiej, w tym do publicznej obrony ocenionej przeze mnie rozprawy, ale także, ze względu na oryginalność i nowatorski charakter uzyskanych przez Niego wyników oraz Jego wartościowy dorobek publikacyjny, należy poddać głosowaniu wyróżnienie tej rozprawy.

