

Recenzja rozprawy doktorskiej por. mgr. inż. Dominika Małego pt. „System automatycznego rozpoznawania mowy wsparty behawioralnymi cechami sygnału mowy”

promotor rozprawy: prof. dr hab. inż. Andrzej P. Dobrowolski

promotor pomocniczy: mjr dr inż. Kamil Kamiński

zlecona pismem z dnia 22.04.2025r. przez Przewodniczącego Rady Dyscypliny Naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne, prof. dr hab. inż. Jana K. Jabczyńskiego.

Przedłożona do recenzji rozprawa doktorska przedstawia wyniki badań mgra Dominika Małego nad opracowaniem i zastosowaniem zestawu cech behawioralnych w problemie rozpoznawania mowy. Przedstawiony problem badawczy i opracowane metody dotyczą zadań analizy sygnałów akustycznych, wstępnego przetwarzania i podziału na sekwencje pomiarowe, ekstrakcji cech, i klasyfikacji. Badania te, pod względem zastosowanych podejść i metod, lokują się w zakresie dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne, jak również są powiązane z informatyką, inżynierią biomedyczną oraz medycyną. W recenzji kolejno charakteryzuję rozprawę doktorską, przedstawiam jej ocenę, wnoszę uwagi krytyczne i stawiam wniosek w postępowaniu o stopień doktora nauk technicznych.

Charakterystyka zawartości pracy.

Recenzowana rozprawa składa się ze spisu treści, indeksu skrótów, streszczenia, przedmowy, sześciu rozdziałów, konkluzji, wykazu literatury oraz czterech załączników i liczy 125 + 55 (załączniki) stron. Bibliografia zawiera 148 pozycji literaturowych (w tym artykułów naukowych i monografii) oraz 10 odniesień internetowych (baz danych). Zamieszczone jest streszczenie pracy w języku polskim i angielskim.

W przedmowie autor zawarł krótki wstęp do zagadnienia badawczego uzasadniając podjęcie tematu rozpoznawania mowy, sformułował tezę rozprawy mówiącą o celowości włączenia cech behawioralnych do systemów automatycznego rozpoznawania mowy opartych na cechach fizycznych. W tym miejscu przedstawiona została zawartość rozdziałów rozprawy doktorskiej grupując je do trzech części głównych. Po przedmowie pojawia się podziękowanie, nie ujęte w spisie treści.

Pierwszy rozdział rozprawy, zatytułowany „Wprowadzenie”, najpierw określa czym są cechy behawioralne i jakie przyjmuje się kryteria oceny biometryk. Następnie zamieszczony jest anatomiczny model generacji sygnału mowy. Wracając do zagadnień technicznych, autor rozprawy przywołuje

podział systemów automatycznego rozpoznawania mowy (ASR), różne procesy identyfikacji oraz zestawy cech ekstrahowanych z sygnału mowy. Rozdział kończy ponad dziewięciostronicowy krytyczny przegląd literatury wraz z podsumowaniem.

W rozdziale 2 przedstawione są podstawy rejestracji sygnału mowy, przetwarzania analogowo-cyfrowego, standaryzacji, wycinania ciszy i segmentacji. Wymienione i skomentowane są przykłady cech fizycznych oraz behawioralnych. Zamieszczono w tym miejscu ocenę zagrożeń poprawności działania systemów ASR. W tym rozdziale przedstawiono i uzasadniono wybór do badań bazy danych sygnałów mowy oraz sformułowanej tezy rozprawy.

Rozdział 3 opisuje opracowane na Wydziale Elektroniki Wojskowej Akademii Technicznej systemy do automatycznego rozpoznawania mowy. Jak podano, pierwszym z nich był system „identyfikator mowy” autorstwa Pani Doktor Eweliny Majdy-Zdancewicz (2013 rok), a kolejnym system ARMIA autorstwa Doktora Kamila Kamińskiego (2018 rok). Ogólnie, prace w tym zakresie są prowadzone na Wydziale Elektroniki od 2010 roku. Zawarty w tym rozdziale opis przedstawia zarówno metody wstępnego przetwarzania sygnału, jak również parametryzację, analizę i klasyfikację. Należy podkreślić, iż istniejące systemy były oparte na cechach fizycznych sygnału mowy.

W rozdziale 4 Autor rozprawy przedstawia opisy opracowanego systemu opartego na cechach behawioralnych. Następuje tutaj standaryzacja parametrów przetwarzanego sygnału, zdefiniowanie znanych oraz zaproponowanych nowych cech behawioralnych, zastosowanie metod ewolucyjnych do selekcji cech najbardziej użytecznych w procesie klasyfikacji oraz samej klasyfikacji. W kolejnym rozdziale zaproponowana jest fuzja systemów z szerokim studium różnych podejść do tego zagadnienia. Przedstawione zostały wyniki klasyfikacji dla różnych realizacji fuzji.

Rozdział 6 zawiera opisy szeregu badań wykonanych przez Autora rozprawy w ramach oceny użytkowej opracowanego systemu ASR. Brane pod uwagę są różne bazy głosów, testy systemu w warunkach bezszumowych oraz ze zróżnicowanym poziomem i typem szumu oraz zakłóceń. Rozpatrywany jest również wpływ jakości nagrań oraz toru akustycznego na skuteczność systemu ASR.

Podstawową treść rozprawy zamyka rozdział 7 zawierający konkluzje autorskich badań. Szczegółowe definicje cech behawioralnych, testy eksperymentalne, obszerne wyniki symulacji i krótki opis interfejsu systemu zawarto w czterech załącznikach.

Ocena rozprawy doktorskiej.

Autor pracy jednoznacznie formułuje oraz realizuje przyjęte cele badawcze zachowując różne możliwości wykorzystania opracowanych metryk w zależności od otrzymanych wyników pośrednich symulacji. Wybór i parametryzacja elementów składowych przetwarzania sygnału jest udokumentowana częściowo w podstawowej treści rozprawy oraz znacznie rozszerzona w załącznikach Z.2 oraz Z.3.

Materiał badań początkowo opierał się na jednej bazie danych, natomiast na etapie badań eksploatacyjnych został rozszerzony do sześciu zbiorów głosów. Jest to z pewnością pewne ograniczenie przeprowadzonych eksperymentów.

Zamieszczony w pracy wykaz literatury, jak również znaczna część pierwszego rozdziału zawierająca analizę dotąd opublikowanych wyników badań w tej tematyce dowodzi znajomości

aktualnego stanu wiedzy. W rozprawie wyczerpująco odniesiono się do innych istniejących rozwiązań oraz wyników eksperymentów, wskazując na ich możliwości i ograniczenia.

Znaczna część metod zastosowanych na etapie przetwarzania wstępnego stanowi już pewien standard. Wkładem Autora rozprawy są propozycje nowych cech behawioralnych, ocena poziomu informacji dystynktywnej w nich zawartej oraz fuzja z aktualnie stosowanymi systemami. Bardzo ważnym aspektem prowadzonych badań jest ocena wpływów różnych czynników na dokładność rozpoznania mówcy. Metodyka zastosowane w badaniach jest adekwatna do problemu, a dobór metod składowych jest uzasadniony typem danych oraz kierunkiem badań.

Jako oryginalny dorobek i główny wkład w rozwój dyscypliny naukowej Autora wymienić można:

- Zaproponowanie nowych cech dystynktywnych;
- Ocenę możliwości rozpoznawania mówcy na podstawie wyłącznie cech behawioralnych,
- Autorskie podejście do ekstrakcji cech,
- Badania nad różnymi technikami fuzji prowadzące do poprawy skuteczności istniejących systemów ASR,
- Eksperymenty oceniające wpływ zakłóceń na dokładność rozpoznania mówcy.

Uwagi merytoryczne do tekstu rozprawy.

Do uwag ogólnych zaliczam:

1. Pewne części rozprawy stanowią opis istniejących rozwiązań i nie są wkładem własnym Autora rozprawy w rozwój dyscypliny naukowej.
2. Analizy wyników ograniczają się do podania procentowej wartości sprawności systemu rozpoznawania mówcy. Wyraźnie brakuje szerszych analiz statystycznych.
3. Na podstawie tekstu rozprawy bardzo trudno jest orzec, które cechy behawioralne są propozycjami autorskimi. Przykładowo, w podpunkcie 4.2.6 jest stwierdzenie „Jednym z autorskich parametrów ...”, podczas gdy przy innych cechach Autor rozprawy nie wyróżnia tej kwestii.
4. W związku z poprzednią uwagą, na podstawie tekstu rozprawy nie jestem w stanie stwierdzić, jakie cechy behawioralne były wskazywane w procesie ich selekcji.
5. W rozprawie w różnych rozdziałach wymieniane są zastosowania systemów ASR. Brakuje jednak zebrania w całość potencjalnych zastosowań systemu i oceny liczebności punktów oraz pomiarów, które mogłyby być nim wykonywane.
6. Na etapie klasyfikacji zastosowane metody nie są zbyt zaawansowane. Co było powodem takiego ograniczenia?

W mojej opinii praca jest bardzo starannie przygotowana pod względem edytorskim i nie wnoszę do niej żadnych tego typu uwag z wyjątkiem bliskości barw użytych na niektórych wykresach.

Wymienione w recenzji braki natury ogólnej nie przekreślają istotnego wkładu Autora pracy w uprawianą tematykę badawczą. Recenzowana rozprawa doktorska spełnia wymagania na stopień doktora.

Konkludując, stwierdzam, iż recenzowana rozprawa doktorska por. mgra inż. Dominika Małego pt. **„System automatycznego rozpoznawania mowy wsparty behawioralnymi cechami sygnału mowy”** w dyscyplinie Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne spełnia wymagania wynikające z art. 179 ust.2 Ustawy z dnia 3 lipca 2018 roku Przepisy wprowadzające ustawę – Prawo o szkolnictwie wyższym i nauce (Dz. U. z dnia 30 sierpnia 2018 r., poz. 1669) w związku z art. 13 ust. 1 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z 2003 r. nr 65, poz. 595 z późn. zm.) oraz Rozporządzenia 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, i na tej podstawie wnoszę o dopuszczenie do publicznej obrony.

Z poważaniem



Tomasz Markiewicz