

dr hab. inż. Józef Kotus, prof. PG
Katedra Systemów Multimedialnych
Wydział Elektroniki Telekomunikacji i Informatyki
Politechnika Gdańska
Ul. Narutowicza 11/12
jozkotus@pg.edu.pl

Gdańsk 17.01.2026

SEKRETARIAT
Rady Dyscypliny AEEITK

Wpłynęło dnia 22.01.2026

Zarejestrowano pod nr 510.6.6/25

Podpis

Recenzja rozprawy doktorskiej
mgr inż. Magdaleny Marty Rybickiej

**„Towards discriminative speaker representations
for speaker recognition and diarization”**

Podstawa wykonania recenzji

Niniejsza recenzja dotyczy rozprawy doktorskiej mgr inż. Magdaleny Marty Rybickiej, zatytułowanej: „Towards discriminative speaker representations for speaker recognition and diarization”, tytuł rozprawy w języku polskim brzmi: „Dyskryminatywne reprezentacje mówców dla zadań rozpoznawania mówców i diaryzacji”. Praca została zgłoszona w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne. Promotorem przedłożonej do oceny rozprawy jest dr hab. inż. Konrad Kowalczyk, profesor uczelni Akademii Górniczo-Hutniczej im. Stanisława Staszica. Funkcję promotora pomocniczego pełnił dr Jesús Villalba, Assistant Research Professor na Johns Hopkins University.

Podstawą do przygotowania niniejszej recenzji jest Uchwała numer 163/2025 Rady Dyscypliny Automatyka, Elektronika i Elektrotechnika i Technologie Kosmiczne Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie z dnia 6 listopada 2025 r.

Zgodnie z przywołaną wyżej uchwałą, postępowanie o nadanie stopnia naukowego doktora mgr inż. Magdaleny Marty Rybickiej jest procedowane w oparciu o przepisy określone w Ustawie z dnia 20 lipca 2018 roku Prawo o szkolnictwie wyższym i nauce. W związku z powyższym czynności wykonane w ramach niniejszej recenzji koncentrowały się na sprawdzeniu, czy przedłożona rozprawa spełnia wymagania ujęte w artykule 187 w/w Ustawy.

Kotus

Organizacja rozprawy i sformułowane hipotezy

Rozprawę doktorską stanowi praca pisemna, zredagowana w formie zbioru 5 opublikowanych i powiązanych tematycznie artykułów naukowych, poprzedzonych streszczeniami w języku angielskim i polskim. Praca zawiera również rozszerzone streszczenie w języku polskim. Zasadnicza część pracy liczy 89 stron. Została zredagowana w języku angielskim i obejmuje 4 rozdziały oraz bibliografię zawierającą 210 pozycji. Do pracy dołączono również pełne teksty opublikowanych prac składających się na cykl publikacji. Pracę otwiera wykaz ważniejszych oznaczeń i akronimów, po którym następuje rozdział 1, w którym doktorantka zawarła: motywację i znaczenie podjętych badań (podrozdział 1.1), sformułowała problem naukowy oraz cztery cele badawcze i hipotezy (podrozdział 1.2), które poniżej przywołuję w oryginalnej formie:

- 1. Odpowiednie wartości hiperparametrów funkcji celu opartej na mierze kątowej (ang. angular-based loss function) poprawiają wydajność oraz zbieżność (ang. convergence) procesu uczenia systemu rozpoznawania mówców;*
- 2. Zastosowanie cech wieloskalowych (ang. multi-scale features), zwiększenie rozdzielczości czasowej oraz uwzględnienie zależności częstotliwościowych przyczyniają się do poprawy skuteczności modeli rozpoznawania mówców;*
- 3. Informacje zakodowane w wektorach osadzeń kodera na poziomie ramek (ang. frame-level encoder embeddings) modelu diaryzacji niosą względne informacje o mówcach i umożliwiają ich rozróżnianie w obrębie jednego nagrania;*
- 4. Informacje o mówcach, wyekstrahowane przy użyciu metod zaproponowanych dla diaryzacji, mogą być wykorzystane w modelu jednoczesnej diaryzacji i separacji mówców, co pozwala na poprawę działania dla obu zadań.*

Cele i hipotezy zostały sformułowane jasno i poprawnie. W podrozdziale 1.3 doktorantka przedstawiła listę pięciu publikacji naukowych, składających się na tematycznie powiązany cykl publikacji, są to:

1. M. Rybicka and K. Kowalczyk, "On Parameter Adaptation in Softmax-Based Cross-Entropy Loss for Improved Convergence Speed and Accuracy in DNN-Based Speaker Recognition", Interspeech, Shanghai, China, 2020.
2. M. Rybicka, J. Villalba, P. Zelasko, N. Dehak, K. Kowalczyk, "Spine2Net: SpineNet with Res2Net and Time-Squeeze-and-Excitation Blocks for Speaker Recognition", Interspeech, Brno, Czech Republic, 2021.
3. M. Rybicka, J. Villalba, N. Dehak and K. Kowalczyk, "End-to-End Neural Speaker Diarization with an Iterative Refinement of Non-Autoregressive Attention-based Attractors", Interspeech, Incheon, South Korea, 2022.
4. M. Rybicka, J. Villalba, T. Thebaud, N. Dehak and K. Kowalczyk, "End-to-End Neural Speaker Diarization with Non-Autoregressive Attractors", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024.
5. M. Rybicka, K. Kowalczyk, T. Thebaud, N. Dehak, J. Villalba, "Joint Diarization and Separation Using SepFormer with Non-Autoregressive Attractors", IEEE Signal Processing Letters, 2025.

Następnie w syntetyczny sposób doktorantka wymienia wkład wniesiony w stan wiedzy.

Podrozdział 1.4 zawiera pozostały dorobek doktorantki, w tym zestawienie pozostałych referatów konferencyjnych (4), raporty techniczne i wyzwania badawcze (4), prezentacje konferencyjne (5), udział w projektach (4) oraz zestawienie zagranicznych wyjazdów badawczych (3).

Treść tego podrozdziału wyraźnie dowodzi dużej aktywności naukowej doktorantki. Jej zaangażowanie w liczne konferencje, projekty, starze zagraniczne, formułowanie wyzwań naukowych, a przede wszystkim cykl publikacji, opublikowanych na trzech edycjach prestiżowej konferencji Interspeech oraz w czasopismach z listy JCR, w których doktorantka jest wiodącym autorem, świadczą o rozległej wiedzy doktorantki oraz umiejętności samodzielnego prowadzenia badań naukowych, jak również umiejętności pracy w zespole.

Podrozdział 1.5 stanowi swoisty przewodnik po zawartości dalszych rozdziałów pracy.

Rozdział 2 zawiera opisy podstawowych pojęć stosowanych w publikacjach. Doktorantka precyzyjnie wyjaśnia, na czym polegały jej badania, wspierając te opisy przejrzystymi ilustracjami. Rozdział ten zawiera cztery podrozdziały w których poruszono kluczowe zagadnienia rozważane w zaprezentowanym cyklu publikacji. Są to: rozpoznawanie mówcy, diaryzacja oraz separacja mowy. Uzupełnieniem opisu tych pojęć są również zestawienia wykorzystywanych metryk, dobranych do poszczególnych zagadnień, tj.: weryfikacji mówcy, diaryzacji oraz separacji mowy.

Ostatni podrozdział rozdziału drugiego to przedstawienie zbiorów danych wykorzystanych w toku przeprowadzonych badań. Dla każdego rozważanego zagadnienia doktorantka wykorzystwała odrębne zbiory nagrań. Na potrzeby badań nad zagadnieniem weryfikacji mówcy były to: NIST SRE datasets, SITW, VoxCeleb. Na potrzeby badań nad diaryzacją wykorzystano zbiory: CALLHOME, DIHARD, AMI, NIST SRE, Switchboard simulated mixtures and conversations. Zagadnienie separacji i diaryzacji doktorantka zgłębiała przy wykorzystaniu zbiorów: LibriMix, SparseLibriMix, LibriheavyMix oraz LibriCSS. Na uwagę zasługuje fakt, iż są to zbiory publicznie dostępne, stanowiące swoisty punkt odniesienia w przedmiotowych badaniach, umożliwiające łatwe porównanie skuteczności opracowanych metod z wynikami osiągniętymi przez innych badaczy.

Rozdział ten jest wartościową częścią pracy, dowodzi że doktorantka posiada rozległą, ogólną wiedzę teoretyczną we wskazanej dyscyplinie naukowej.

W rozdziale 3 doktorantka przedstawiła przegląd wykonanych własnych oryginalnych badań i przedstawia wynikający z nich, autorski wkład w stan wiedzy. Rozdział ten zawiera najważniejsze treści, z punktu widzenia dowodów na oryginalne rozwiązanie problemu naukowego.

Rozdział 4 zawiera treści podsumowujące udokumentowane osiągnięcie naukowe oraz zawiera wskazanie obszarów przyszłych prac badawczych.

Problem badawczy objęty zakresem rozprawy

Doktorantka, w swojej rozprawie doktorskiej podejmuje ważne i istotne problemy badawcze jakim są: rozpoznawanie mówcy, separacja mowy oraz diaryzacja mówcy. Zagadnienia te, w świecie, w którym jesteśmy świadkami dynamicznie rozwijających się technologii informacyjnych, są szczególnie istotne, a systemy bazujące na wykorzystaniu przetwarzania języka naturalnego, w tym ludzkiego głosu, odgrywają coraz większą rolę.

Badania Doktorantki mają znaczenie nie tylko w interakcjach człowiek – człowiek, lecz również w relacji człowiek – maszyna oraz w szczególnych przypadkach maszyna – maszyna. Doktorantka w swoich pracach proponuje nowatorskie podejścia, które umożliwiają osiągnięcie wyników przekraczających znacząco stan techniki.

Warto przy tym podkreślić szeroki zakres podjętej pracy, zróżnicowany pod względem zbiorów danych, metryk i sposobów rozwiązywania poszczególnych problemów. Doktorantka podjęła się zadania o dużym stopniu złożoności i udowadnia, że rozważane problemy mają pewną cechę wspólną. Bazując na tym spostrzeżeniu przedstawia systematycznie i konsekwentnie rozwijane metody.

W związku z powyższym uważam podjętą pracę za aktualną, ważną i wartościową.

Analiza źródeł (w tym literatury światowej i stanu techniki)

Doktorantka dokonała szczegółowego i starannego doboru literatury. Spis cytowany prac liczy imponujące 210 pozycji. Pozycje te odnoszą się do książek, artykułów naukowych, referatów konferencyjnych oraz zbiorów danych. Wymienione opracowania dotyczą podjętej tematyki i są efektem starannie i gruntownie wykonanego przeglądu stanu wiedzy. Autorka umiejętnie i skutecznie odnajduje obszary nauki w których proponuje własne, autorskie rozwiązania. Czego dowodem jest jej obszerny i różnorodny dorobek naukowy. Pod tym względem praca zasługuje na szczególne uznanie.

Poprawność pracy

Praca jest poprawna pod względem układu poszczególnych rozdziałów oraz zawartych w nich treści. Autorka systematycznie i skrupulatnie opisuje podjęte zagadnienie. Przyjęta metodyka badawcza jest szczegółowo wyjaśniona, uzasadniona i zilustrowana licznymi rysunkami. Doktorantka biegle posługuje się zarówno adekwatnym do podjętych zagadnień aparatem matematycznym, jak również zróżnicowanym zestawem programów komputerowych i różnorodnymi bazami danych. Dowodzi tym samym swojej dojrzałości w zakresie prowadzenia samodzielnych badań naukowych.

Oryginalne wyniki rozprawy i ich znaczenie dla dyscypliny naukowej

Oryginalny wkład doktorantki do dziedziny wiedzy został ujęty w poszczególnych pracach, składających się na cykl publikacji. Dodatkowo, główne osiągnięcia zostały przedstawione w formie listy wypunktowanej na stronach: 5 i 6 oraz w formie syntetycznego opisu na stronach 71 i 72 rozprawy doktorskiej.

W odniesieniu do problemu rozpoznawania mówców, realizowanego w zagadnieniu weryfikacji mówcy wkład doktorantki to:

- Propozycja dwóch funkcji celu opartych na mierze kątowej, opisanych szczegółowo w publikacji 1. Pierwsza metoda, opisana w rozdziale 2.3 publikacji 1 polega na wprowadzeniu dodatkowego parametru (MAda), który usprawnia proces treningu sieci. Druga metoda, opisana w rozdziale 2.4 publikacji 1 polega na wprowadzeniu dodatkowego parametru (ParAda) w funkcji kosztu opartej na funkcji softmax.
- Adaptacja i modyfikacja architektury ResNet poprzez zwiększenie rozdzielczości czasowej modelu w celu poprawy wydajności rozpoznawania mówców;
- Przeprowadzenie badań dla zaproponowanych modyfikacji na publicznie dostępnych zbiorach danych oraz uzyskanie wyników wyrażonych wskaźnikiem EER znacząco przewyższającym stan techniki, ujętych w tabelach 1 i 2 publikacji 1.

Wyniki te dowodzą słuszności tezy 1 rozprawy.

- Zastosowanie modelu ze skalą permutowaną (SpineNet) na potrzeby poprawy skuteczności rozpoznawania mówcy. Szczegółowy opis wprowadzonych modyfikacji oraz przeprowadzonych badań zaprezentowano w publikacji 2. Kluczowe wyniki zostały zamieszczone w tabelach 2 i 3. Testy zaproponowanej, zmodyfikowanej architektury sieci neuronowej przeprowadzono na trzech różnych zbiorach danych: VoxCeleb1-E (Extended), VoxCeleb1-H (Hard), VoxCeleb1-O (Original). Najmniejszą wartość wskaźnika EER uzyskano dla sieci oznaczonej jako T-SE-Spine2Net-49, dla wszystkich zbiorów testowych.

Wyniki te dowodzą słuszności tezy 2 rozprawy.

W odniesieniu do problemu diaryzacji mówców wkład doktorantki to:

- Zaproponowanie metody nieautoregresywnej estymacji atraktorów (ang. Non-Autoregressive Attractor estimation) – czyli sposobu określania reprezentacji mówców występujących w danym nagraniu, odmiennego od stosowanego w literaturze podejścia, bazującego na metodzie autoregresywnej, EDA (Encoder-Decoder-based Attractors). Szczegółowy opis metody został przedstawiony w rozdziale 3 publikacji 3.
- Zaproponowanie podejścia dla zadania diaryzacji polegającego na zastąpieniu atraktora opartego na sieci LSTM nowym, nieautoregresywnym podejściem opartym na architekturze sieci typu transformer. Testy metody przeprowadzono dla zbiorów nagrań głosów dwóch mówców w próbie testowej (Sim2Spk). Wykorzystano zbiory: Switchboard-2, Switchboard Cellular, NIST Speaker Recognition Evaluation. Wykorzystano również rzeczywiste nagrania zawierające głosy dwóch osób, zaczerpnięte z bazy CALLHOME. Kluczowe wyniki przedstawiono w tabeli 2 w publikacji 3. Zaproponowana metoda umożliwiła uzyskanie lepszych wyników, wyrażonych za pomocą miary DER, w porównaniu z metodą referencyjną.
- Rozszerzenie metody nieautoregresywnej estymacji atraktorów na potrzeby diaryzacji mówców dla przypadków znanej i nieznannej liczby mówców, metoda

została oznaczona jako (EEND-NAA) i została szczegółowo opisana w publikacji 4 w rozdziałach IV i V.

- Przeprowadzenie badań i opracowanie wyników dla różnych wariantów zaproponowanej metody, w badaniach wykorzystano publicznie dostępne korpusy: na potrzeby treningu były to: Switchboard-2, Switchboard Cellular, NIST Speaker Recognition Evaluation, na potrzeby testowania wykorzystano ponadto próbki mowy z bazy CALLHOME oraz bazy DIHARD II. W trakcie testów wykorzystano różne kombinacje liczby mówców oraz współczynnika nakładania mowy. Wyniki zaprezentowano w formie licznych tabel. Doktorantka, wraz z zespołem współautorów, wykazała wyższość zaproponowanej metody w porównaniu z metodą referencyjną (EEND-EDA), uzyskując niższe wartości metryki DER we wszystkich rozważanych scenariuszach testowych. Na uwagę zasługuje fakt, że wartość metryki DER wyraźnie rosła wraz ze wzrostem liczby różnych głosów w analizowanym nagraniu.

W ten sposób doktorantka dowiodła słuszności tezy 3 rozprawy.

W odniesieniu do połączonych problemów diaryzacji i separacji mówców wkład doktorantki obejmuje:

- Rozszerzenie modelu separacji do ujednoliconej struktury dla wspólnej diaryzacji mówcy i separacji mowy, umożliwiającej efektywne i skuteczne wykonywanie tych dwóch czynności.
- Dostosowanie metody nieautoregresywnej estymacji atraktorów dla reaziacji czynności diaryzacji i separacji mowy, umożliwiającej przetwarzanie nagrań w których mowa ma charakter rozmowy (mały współczynnik nakładania) jak i nagrań, w których osoby mówią jednocześnie. Metoda została szczegółowo opisana w publikacji 5 w rozdziale III. Testowano dwa warianty zaproponowanej metody: Cluster-Based Attractor Estimation (SepDiarCA) – dla niewielkich wartości współczynnika mowy jednoczesnej oraz Diarization-Based Attractor Estimation (SepDiarDA) – dla dużych wartości współczynnika mowy jednoczesnej.
- Opublikowanie wyników badań polegających na testowaniu skuteczności łącznej realizacji dwóch zadań: diaryzacji i separacji mowy, z uwzględnieniem różnej wartości współczynnika nakładania się głosów poszczególnych mówców. Na potrzeby oceny skuteczności separacji mowy wykorzystano metrykę SI-SDR_i, a do oceny skuteczności diaryzacji wykorzystano metrykę DER. W trakcie badań wykorzystano różnorodne zbiory nagrań mowy, w tym: SparseLibri2Mix, CALLHOME, Libri2Mix, Libri3Mix LibriheavyMix. Kluczowe wyniki zostały ujawnione w tabelach I, II i III w publikacji 5. Zaproponowane metody uzyskały najlepsze wartości poszczególnych metryk we wszystkich rozważanych scenariuszach testowych.

W ten sposób doktorantka dowiodła słuszności tezy 4 rozprawy.

Zaprezentowane w niniejszej rozprawie doktorskiej różnorodne, pomysłowe sposoby przetwarzania mowy oraz cenne wyniki wielu badań mają istotne znaczenie dla rozważanej dziedziny nauki i wytyczają nowe ścieżki badań dla społeczności naukowców na całym świecie.

Pytania dotyczące zaproponowanych metod i wykonanych badań

Zaprezentowane w poszczególnych publikacjach wyniki nie budzą wątpliwości. Dokładna analiza treści rozprawy oraz poszczególnych publikacji skłania jednak recenzenta do zadania różnorodnych pytań. Zostały one pogrupowane pod względem rozważanego zagadnienia.

Pytania do publikacji dotyczących weryfikacji mówcy:

- jakie były wartości SNR dla próbek głosu w zbiorze testowym?
- czy proponowana architektura i metodyka pracy dają nadzieję na praktyczne wdrożenie opracowanej metody weryfikacji mówcy w systemach kontroli dostępu?
- czy rozważane były nagrania próbek głosu w różnych odległościach od mikrofonu?
- czy opracowana technika może być użyteczna do weryfikacji autentyczności głosu na potrzeby wykrywania nagrań typu deepfake?

Pytania do publikacji dotyczących separacji i diaryzacji mówcy:

- jaki był procentowy udział poszczególnych głosów w nagraniach przygotowanych na potrzeby testowania metody diaryzacji, czy czas trwania poszczególnych głosów był porównywalny, czy też któryś z mówców był dominujący,
- jaka była energia mowy poszczególnych mówców w nagraniach testowych w odniesieniu do problemu diaryzacji oraz łącznego problemu diaryzacji i separacji mowy,
- czy określono minimalny czas aktywności głosowej mówcy w nagraniu wielogłosowym, który zapewnia skuteczną separację i diaryzację? Czy doktorantka może w przybliżeniu określić ten czas?
- badania wykonano na referencyjnych zbiorach danych, czy doktorantka testowała opracowane architektury sieci na własnych nagraniach wielogłosowych?
- czy opracowane metody mogą działać w czasie rzeczywistym?

Poziom edycyjny rozprawy

Praca jest napisana bardzo starannie. Zdania są sformułowane poprawnie pod względem stylistycznym. Czyta się ją bardzo przyjemnie. Treści poszczególnych rozdziałów są ze sobą połączone, tworzą spójną całość. Numeracja wzorów, rysunków i tabel jest poprawna. Odwołania do rysunków i tabel znajdują się w tekście pracy. Każdy tego rodzaju obiekt zamieszczony w pracy ma swoje uzasadnienie. Wkład doktorantki w rozwój dziedziny wiedzy jest wyraźnie zaznaczony i wyjaśniony.

Końcowe wnioski recenzji

Podsumowując uwagi i opinie przedstawione w niniejszej recenzji stwierdzam, że przygotowana przez panią Magdalenę Martę Rybicką rozprawa doktorska spełnia wszystkie ustawowe wymagania określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce ze znacznym nadmiarem. W związku z powyższym wnioskuję do Rady Dyscypliny Automatyka, Elektronika i Elektrotechnika i Technologie Kosmiczne Akademii Górniczo-Hutniczej o dopuszczenie recenzowanej rozprawy do publicznej obrony.

Uwzględniając ponadto obszerny dorobek doktorantki, w szczególności prace składające się na cykl publikacji, tj.: dwa artykuły opublikowane w prestiżowych czasopismach, a także 3 referaty opublikowane w materiałach renomowanej konferencji o zasięgu światowym, w których to pracach doktorantka jest pierwszym autorem, stwierdzam, że zostały spełnione przesłanki ujęte w punkcie 2, podpunkt d, podpunkt i, zawarte w uchwale nr 83/2025 Rady Dyscypliny Automatyka, Elektronika i Elektrotechnika i Technologie Kosmiczne Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie z dnia 6 lutego 2025 r. Na tej podstawie stawiam wniosek o wyróżnienie niniejszej rozprawy doktorskiej.

JÓZEF KOTUS