



Instytut Telekomunikacji i Cyberbezpieczeństwa

Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

SEKRETARIAT
Rady Dyscypliny AEEITK

Warszawa, 12.01.2026 r.

Wpłynęło dnia 28.01.2026

Zarejestrowano pod nr 510.6.7/25

Podpis Jm

RECENZJA ROZPRAWY DOKTORSKIEJ

mgr inż. **MAGDALENY MARTY RYBICKIEJ**

pt. „Towards discriminative speaker representations for speaker recognition and diarization”, przedłożonej Radzie Naukowej Automatyki, Elektroniki, Elektrotechniki i Technologii Kosmicznych Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie

PRZEDMIOT ROZPRAWY, GŁÓWNE CELE PRACY

Praca Pani mgr inż. **Magdaleny Marty Rybickiej** pt. „Towards discriminative speaker representations for speaker recognition and diarization” (polski tytuł: *Dyskryminatywne reprezentacje mówców dla zadań rozpoznawania mówców i diaryzacji*), dotyczy zagadnień rozpoznawania, diaryzacji i separacji mówców.

Przedstawiona rozprawa jest pracą o charakterze badawczym. Jako cel badawczy Autorka stawia sobie usprawnienie systemów rozpoznawania i diaryzacji mówców. W szczególności Doktorantka badała możliwość usprawnienia uczenia systemów rozpoznawania mówców, poprawę skuteczności rozpoznawania mówców, możliwość zastosowania osadzeń kodera w procesie diaryzacji oraz możliwość łączenia procesów diaryzacji i separacji mówców.

Doktorantka prezentuje w rozprawie następujące tezy:

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.

Rozprawa została napisana w języku angielskim, składa się z ciągu 5 publikacji, poprzedzonych obszernym autoreferatem, składającym się z 4 rozdziałów oraz bibliografii. Rozprawa wraz z załącznikami zajmuje łącznie 131 stron.

Rozdział 1. wprowadza w tematykę rozprawy.

Rozdział 2. prezentuje stan wiedzy w obszarze rozpoznawania, diaryzacji i separacji mówców.

Rozdział 3. prezentuje przeprowadzone badania oraz ich wyniki.

Rozdział 4. podsumowuje całość rozprawy.

MOCNE STRONY ROZPRAWY

Za mocną stroną rozprawy Doktorantki uważam bardzo wysoką jakość zaprezentowanych prac. Doktorantka szczegółowo udokumentowała cały proces badawczy: od wskazania zbiorów danych, używanych metryk, poprzez szczegóły architektury poszczególnych sieci, aż po szczegółową prezentację i dyskusję wyników.

Za ważne osiągnięcie Doktorantki uważam zaproponowanie użycia do diaryzacji mówców metody nieautoregresywnej estymacji atraktorów (ang. *Non-Autoregressive Attractor estimation*, NAA), w którym wszystkie atraktory, służące do określania reprezentacji mówców w danym sygnale, wyznaczone są jednocześnie.

W zaproponowanym podejściu reprezentacje mówców są wyznaczone z wykorzystaniem właściwości osadzeń na poziomie ramek (ang. *frame-based embeddings*). Pozwoliło to na poprawę skuteczności diaryzacji mówców o 50% (!) względem modelu odniesienia dla scenariusza symulowanego oraz 15% dla nagrań rzeczywistych z bazy CALLHOME.

Warte docenienia jest też zaproponowanie przez Doktorantkę struktury sieci, która może zostać wyuczona równocześnie dla zadania diaryzacji i separacji mówców, również z wykorzystaniem zaproponowanej wcześniej metody NAA. Doktorantka wykazała, że zaproponowana metoda działa skutecznie także w przypadku mocno niezrównoważonych proporcji między czasem wypowiedzi poszczególnych mówców. W przedstawionym eksperymencie wykazała ponad 50% (!) poprawę skuteczności diaryzacji i separacji mówców względem stanu wiedzy.

Do ważnych osiągnięć Doktorantki w zakresie rozpoznawania mówców zaliczam również automatyczną adaptację hiperparameterów, która polega na tym, że ich wartości dostosowują się w każdej iteracji uczenia do poprawności odpowiedzi sieci i jej

zbieżności. Metoda ta umożliwiła podniesienie skuteczności oraz przyspieszenie uczenia sieci dla zadania rozpoznawania mówców.

Doktorantka dokonała też udanej modyfikacji struktury sieci ResNet (ang. *residual model*) pod kątem użycia jej w rozpoznawaniu mówców. Autorka zwiększyła rozdzielczość czasową analizy, co zaowocowało poprawą skuteczności rozpoznawania mówcy. Inną innowacją, zaczerpniętą z dziedziny przetwarzania obrazów, było łączne zastosowanie modułów Res2Net oraz Time-Squeeze-and-Excitation (T-SE), a także adaptacja skali permutowanej (ang. *scale-permuted design*), tzw. SpineNet.

Potwierdzeniem osiągnięć Doktorantki jest jej dorobek publikacyjny. Oprócz 5 artykułów stanowiących główny cykl publikacji, z czego 3 były prezentowane na topowej konferencji Interspeech, a 2 zostały opublikowane w wiodących czasopismach z dziedziny przetwarzania mowy (IEEE/ACM TASLP, IEEE Signal Processing Letters), Doktorantka jest jeszcze współautorką 4 innych publikacji prezentowanych na międzynarodowych konferencjach, w tym tak renomowanych jak ICASSP i EUSIPCO.

Uważam, że rozprawa doktorska prezentuje oryginalne rozwiązania w zakresie rozpoznawania i diaryzacji mówców. Doktorantka wykazała się umiejętnością samodzielnego prowadzenia pracy naukowej, sprawnym posługiwaniem warsztatem badawczym, wykazała też biegłą znajomość zagadnień z obszarów na styku elektroniki, automatyki, a także informatyki.

SŁABE STRONY ROZPRAWY

Do słabych stron rozprawy zaliczam pewną niekonsekwencję w strukturze pracy. Doktorantka na wstępie, w Rozdziale 1, starannie wprowadza w tematykę i formułuje wcześniej cytowane cztery tezy pracy. Natomiast pod koniec autoreferatu, w Rozdziale 4 Doktorantka co prawda stwierdza ogólnie, że tezy pracy zostały potwierdzone, ale już ich nie przywołuje ani też bezpośrednio nie uzasadnia, że te tezy zostały spełnione. Uważam, że bardzo przydałoby się przywołanie poszczególnych tez (skoro już zostały sformułowane) i wyraźne wykazanie, że udało się ich dowieźć. Według mnie możliwy byłby też wariant bez formułowania żadnych tez, a jedynie z określeniem obszaru badań i/lub zagadnień badawczych.

Inną pewnego rodzaju niekonsekwencją jest następująca niespójność: Doktorantka informuje na wstępie, że na rozprawę składa się ciąg 5 publikacji. Publikacje są świetne, przeszły na pewno rygorystyczny proces publikacyjny, więc jest to gotowy materiał mogący stanowić podstawę do nadania stopnia naukowego doktora. Czytelnik spodziewa się więc, że właśnie te publikacje będą stanowić trzon rozprawy, który zostanie poprzedzony jedynie niezbędnym wstępem i krótkim autoreferatem.

Tymczasem główną część rozprawy stanowi powtórzenie (i czasami pewne rozwinięcie) treści owych 5 publikacji w Rozdziale 3, a same publikacje znajdują się dopiero w załączniku. Według mnie stanowi to zbędną redundancję. Być może lepsza byłaby wtedy forma klasycznej monografii i wystarczyłoby jedynie odesłanie do opublikowanych prac? Pozwoliłoby to też pominąć procedurę pozyskiwania oświadczeń współautorów.

UWAGI SZCZEGÓŁOWE/POLEMICZNE

- Eksperymenty opisane w Publikacjach I oraz II zostały przeprowadzone na zbiorach VoxCeleb1 oraz VoxCeleb2, które są dość specyficzne: zawierają nagrania celebrytów umieszczone w Internecie. Świetnie, że zastosowana była też augmentacja danych poprzez np. mieszanie z muzyką i emulację różnych warunków akustycznych. A jak zachowałyby się proponowane metody dla nagrań o niższej jakości, np. telefonicznej?
- Doktorantka twierdzi, że zaproponowana w Publikacji IV metoda nieautoregresywnej estymacji atraktorów (NAA) zapewnia większą wyjaśnialność procesu estymacji atraktorów. Dlaczego Doktorantka uważa, że ta metoda jest bardziej wyjaśnialna niż metoda oparta na modelach LSTM?
- W swoich badaniach Doktorantka często wykorzystywała symulowane mieszanie mówców poprzez łączenie składowych sygnałów (np. bazy Libri2Mix i Libri3Mix w Publikacji V czy symulacje w Publikacji III z wykorzystaniem baz NIST SRE). Czy wyniki eksperymentów opisanych w tych publikacjach nie obniżyłyby się znacznie, gdyby analizowany był naturalny sygnał z głosami wielu mówców?
- W Publikacji V jest napisane: „(...) we used single-speaker regions from the train part of CH and simulated 2-speaker mixtures to adapt models for the joint task.” Jak te obszary “single-speaker” były identyfikowane? Manualnie?
- Przedstawione w Rozdziale 1.2 „cele badawcze i hipotezy” (str. 4) to właściwie wyłącznie hipotezy.
- Czy algorytm PESQ rzeczywiście bywa stosowany do oceny jakości separacji mówców? Zwykle stosuje się go do oceny transmisji mowy, jakości kodeków itp. (str. 45).
- W rozprawie brakuje wyraźnego zaznaczenia, które części prac wieloautorских zostały wykonane przez Doktorantkę.
- Występują drobne usterki edycyjne, opisane w następnym rozdziale.

STRONA EDYCYJNA PRACY

Rozprawa doktorska **mgr inż. Magdaleny Marty Rybickiej** jest napisana bardzo poprawnym językiem angielskim. Autorka sprawnie posługuje się stylem naukowym. Edycja rozprawy jest bardzo staranna.

Drobne niedociągnięcia natury językowej i typograficznej to m.in.:

- Przecinek bywa używany jako znak dziesiętny (np. Tabela 3.1 na str. 55), podczas gdy w notacji angielskiej znakiem dziesiętnym jest kropka.
- Na wydruku stron z kolorowymi rysunkami tekst wydaje się być na przemian pisany zwykłą i wytłuszczoną czcionką.
- Po skrócie „e.g.” powinien znajdować się przecinek.
- Błędy językowe w polskim streszczeniu: 1) powinno być „agenty konwersacyjne” (czyli systemy) zamiast „agenci konwersacyjni” (co sugerowałoby... ludzi, takich jak agenci ubezpieczeniowi), na podobnej zasadzie co „piloty” i „piloci”, 2) powinno być „oparte na mierze kątowej” zamiast „oparte o miarę kątową” (str. xiv).
- Bibliografia: [147] naukowiec Alvin Martin na imię ma Alvin, a na nazwisko Martin, nie odwrotnie ;)

Wyżej wymienione drobne uchybienia edycyjne nie umniejszają jednak w żadnym stopniu osiągnięć Doktorantki.

WNIOSKI KOŃCOWE

W podsumowaniu stwierdzam, że cele badawcze postawione w rozprawie **mgr inż. Magdaleny Marty Rybickiej** zostały osiągnięte. Doktorantka opracowała oryginalne rozwiązania w zakresie rozpoznawania, diaryzacji i separacji mówców.

Stwierdzam, że przedstawiona rozprawa doktorska spełnia warunki określone w art. 187 ustawy z dn. 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2024 r. poz. 1571 z późn. zm.), dlatego niniejszym wnioskuje o **dopuszczenie** Doktorantki, Pana **mgr inż. Magdaleny Marty Rybickiej**, do publicznej obrony jej rozprawy doktorskiej. Dodatkowo, ze względu na znaczny dorobek publikacyjny Doktorantki oraz bardzo wysoki poziom przeprowadzonych badań, składam **wniosek o wyróżnienie rozprawy**.

dr hab. inż. Artur Janicki, prof. uczelni
Politechnika Warszawska



PODPIS ZAUFANY
ARTUR JANICKI
27.01.2026 21:04:15 GMT+1
Dokument podpisany elektronicznie
podpisem zaufanym