

Prof. dr hab. inż. **Krzysztof Ślot**  
Instytut Informatyki Stosowanej  
Politechnika Łódzka

## Recenzja rozprawy doktorskiej

Pana magistra inż.  
**Kacpra Radzikowskiego**  
pt.

### **A study on speech recognition and correction for non-native English speakers**

#### **1. Tematyka i cele rozprawy**

Przedstawiona do recenzji rozprawa doktorska dotyczy problematyki automatycznego rozpoznawania mowy (ang. *Automatic Speech Recognition*, ASR), przy czym szczególnym wątkiem badawczym tej niezwykle obszernej dziedziny, którego dotyczą prace badawcze Autora, jest rozpoznawanie mowy wypowiedzianej przez osoby, dla których język wypowiedzi nie jest językiem ojczystym. Zagadnienie podjęte przez Doktoranta jest zarówno **istotne** jak i niezwykle **aktualne**. Postępująca globalizacja i związane z tym procesy społeczne, kulturowe i ekonomiczne wymagają umiejętności sprawnego komunikowania się z osobami z innych kręgów językowych, co w ostatnich dziesięcioleciach doprowadziło do powszechnej akceptacji jako narzędzi komunikacji jedynie kilku wybranych języków (przede wszystkim, języka angielskiego). Ten stan rzeczy okazał się poważnym wyzwaniem dla metod automatycznego rozpoznawania mowy, które stanowią istotny element wspierający postęp techniczny i cywilizacyjny, a które okazują się być niewystarczająco skuteczne w konfrontacji z mową pozyskiwaną od mówców ‘nienatywnych’.

Celem badań Doktoranta była próba opracowania metod rozpoznawania mowy pozwalających na skuteczne poradzenie sobie z przedstawionym, trudnym problemem. Skala złożoności zadania, wynikająca z wielu czynników: ogromnej różnorodności sposobów artykulacji i konstruowania wypowiedzi, często znacznie odbiegających od istniejącego dla danego języka kanonu, naleciałości i nawyków fonetycznych ukształtowanych przez inny język ojczysty, ubóstwa leksykalnego i gramatycznego, jest na tyle duża, że jak dotąd nie zostały opracowane żadne skuteczne metody jego rozwiązania. Dlatego też zadanie, którego rozwiązania podjął się Doktorant jest nietrywialne, nakreślony cel pracy – ambitny, a uzyskanie znaczących efektów prac może stanowić osiągnięcie o randze spełniającej kryteria merytoryczne stawiane przed rozprawą doktorską i stanowić

zauważalny wkład do dyscypliny naukowej **informatyka techniczna i telekomunikacja**, w której bez wątpienia lokuje się obszar badań podjętych przez Doktoranta.

## **2. Struktura i tezy rozprawy**

Metodyka postępowania, przyjęta przez Doktoranta dla osiągnięcia postawionych celów rozprawy nie budzi żadnych zastrzeżeń. We wstępnej części rozprawy, Doktorant kompetentnie podsumowuje stan wiedzy w zakresie istniejących propozycji metod poprawy skuteczności rozpoznawania mowy nienatywnej oraz wskazuje podstawowe przyczyny utrudniające właściwe działanie standardowych narzędzi rozpoznawania mowy. Z uwagi na oczywistą trudność precyzyjnej identyfikacji czynników decydujących o odmienności mowy wypowiedzianej przez osoby nienatywne, Doktorant zwraca się w kierunku wykorzystania uczenia maszynowego, jako jedynej sensownej metodyki adekwatnej dla możliwości rozwiązania problemu. Jednocześnie, dokonuje trafnej diagnozy jednego z głównych powodów trudności wdrożenia konwencjonalnych metod uczenia maszynowego, jakim jest brak wystarczająco obszernego, etykietowanego materiału eksperymentalnego, niezbędnego dla budowy skutecznych modeli rozpoznawania. W konsekwencji, uwaga Doktoranta skupia się na jedynym realnym alternatywnym źródle informacji - bazach mowy pozbawionych etykiet, których zgromadzenie w odpowiednio dużych wolumenach nie nastrocza problemu, ale których skuteczne wykorzystanie wymaga opracowania zaawansowanych algorytmów analizy.

Prace nad poszukiwaniem metod rozpoznawania mowy dla języka nie będącego językiem ojczystym mówcy, zostały przez Doktoranta przeprowadzone w trzech różnych kierunkach, odpowiadających trzem zasadniczym celom naukowym, które wytyczył dla swoich badań, i które można uznać za tezy przedstawionej rozprawy. Pierwszy z nich dotyczy opracowania metody pozwalającej na budowę modeli algorytmicznych ukierunkowanych na analizę i rozpoznawanie wypowiedzi mówców nienatywnych. Drugi cel to próba sformułowania metodyki modyfikacji zarejestrowanej mowy, dokonywanej na poziomie fonetycznym i zmierzającej do dostosowania struktury oraz brzmienia wypowiedzi do kanonów rozważanego języka, pozwalająca na wykorzystanie do rozpoznawania algorytmów opracowanych dla mówców natywnych. Wreszcie, ostatni cel badawczy sformułowany w rozprawie to próba weryfikacji możliwości zbudowania algorytmu dokonującego korekty tekstu będącego efektem analizy wypowiedzi, dostosowującej go do wymogów składni i gramatyki języka docelowego, pozwalającej na dodatkowe zwiększenie poprawności rozpoznawania. Celem prac w każdym z wymienionych wątków badawczych było uzyskanie rozwiązań o możliwie najbardziej ogólnym charakterze, pozwalającym na realizację algorytmów w odniesieniu do dowolnych języków wypowiedzi.

## **3. Merytoryczna ocena pracy**

Koncepcje zaproponowane przez Doktoranta w ramach realizacji prac są, w odniesieniu do dwóch pierwszych wymienionych wątków badawczych, czyli budowy systemu ASR wyspecjalizowanego w analizie mowy nienatywnej i opracowania metody fonetycznej transformacji wypowiedzi, bardzo ciekawe, oryginalne i merytorycznie znaczące. Podejście zastosowane w odniesieniu do realizacji trzeciego postawionego celu – opracowania algorytmu korekcji wyników rozpoznawania modułu ASR, mają w mojej opinii, raczej szablonowy charakter i dlatego w dalszej części recenzji zostanie mu poświęcone znacznie mniej uwagi. Punktem wyjścia dla sformułowanych przez Doktoranta

propozycji jest bardzo dobra identyfikacja właściwych, zaawansowanych i stosunkowo jeszcze świeżych koncepcji uczenia maszynowego, stanowiących bazę dla przedstawionych przez Niego, bardzo pomysłowych metod. Na uznanie zasługuje intuicja badawcza, której efektem jest formułowanie adekwatnych kryteriów sterujących przebiegiem uczenia rozważanych algorytmów oraz rzetelność i rozległość procedur eksperymentalnej weryfikacji hipotez, obejmująca trafny wybór materiału doświadczalnego, a także stosowanie właściwych miar i kryteriów oceny uzyskiwanych wyników, decydujące o wiarygodności procedury testowania i istotności otrzymanych rezultatów. Przedstawiona w dalszej części ocena dokonań Doktoranta została podzielona na dwie części, odzwierciedlające układ treści przedstawianych w pracy.

### **3.1 Metoda budowy modeli ASR dostosowanych do analizy mowy nienatywnej**

Podstawowym ograniczeniem dla możliwości utworzenia modelu ASR dla mowy wypowiedzianej w języku innym niż język ojczysty mówcy jest brak dostępności odpowiednio obszernego, etykietowanego materiału eksperymentalnego. Zapewnienie odpowiedniej złożoności modelu rozpoznawania mowy, niezbędnego z punktu widzenia złożoności rozwiązywanego problemu, wymaga dostarczenia podczas procesu uczenia wystarczającej wiedzy o problemie, co stoi w sprzeczności z istniejącym ograniczeniem. Dlatego, jedynym sposobem osiągnięcia zakładanego celu jest opracowanie procedury korzystającej z nieetykietowanych baz danych, co jest zadaniem nietrywialnym, wymagającym kreatywności i doskonałej orientacji w zaawansowanych koncepcjach uczenia.

Zaproponowane przez Doktoranta rozwiązanie – metoda budowy modeli ASR dla analizy mowy nienatywnej, stanowi twórcze rozwinięcie koncepcji Podwójnego Uczenia Nadzorowanego (Double Supervised Learning - DSL), adaptowane w sposób umożliwiający osiągnięcie założonego celu z wykorzystaniem wyłącznie dwóch nieetykietowanych źródeł danych: pozbawionych transkrypcji nagrań mowy oraz pozbawionych akustycznej ilustracji tekstów. Przedstawiony przez Doktoranta, niezwykle pomysłowy sposób rozwiązania zadania to wykorzystanie dwóch komplementarnych modeli transformacji danych: tekstowych w akustyczne (który będę określać dalej jako *Text-to-Speech*: T2S) i akustycznych w tekstowe (*Speech-to-Text*: S2T), pozwalające na realizację procedury uczenia nadzorowanego w konwencji autoenkodowania, korzystającego z danych generowanych przez dwa dodatkowe modele: generacji tekstu (dokonywanej przez wstępnie wytrenowany model języka) i generacji wypowiedzi (dokonywanej przez wstępnie wytrenowany model akustyczny).

Obydwa modele generacyjne: językowy i akustyczny są budowane na bazie nienadzorowanych, łatwo dostępnych (a więc, obszernych) zbiorów danych, przy czym ich dodatkową funkcjonalnością jest możliwość oceny poprawności podawanych na ich wejścia próbek tekstu i mowy. Dla budowy modelu języka Doktorant wykorzystuje korpus współczesnego języka angielskiego (COCA), stanowiący dostatecznie obszerną bazę informacji niezbędnej do zapewnienia poprawnej reprezentacji wiedzy, zaś dla budowy modelu akustycznego, równie bogaty, nieetykietowany zbiór nagrań języka angielskiego wypowiedzianego przez mówców polsko- lub japońsko-języcznych. Procedura uczenia systemu ASR dla mowy nienatywnej, który jest tożsamy z modułem S2T przedstawionej komplementarnej pary, obejmuje również uczenie modelu T2S i jest dokonywana w pętli obejmującej: naprzemienną generację próbek albo przez generator języka, albo przez generator

akustyczny, przetworzenie tych próbek przez obydwa uczone modele i dokonanie odpowiednich korekt na podstawie odpowiednio zdefiniowanych funkcji straty. W dwuetapowym przekształceniu danych dokonywane są dwie oceny poprawności ich realizacji – cząstkowa (w jakim stopniu wygenerowany tekst / wypowiedź odpowiada modelowi języka / modelowi akustycznemu) i końcowa (zgodność zrekonstruowanej próbki z oczekiwaną, czyli wygenerowaną), co skłania Autora do przyjęcia interpretacji procesu treningu jako metody uczenia ze wzmocnieniem.

Dopełnieniem pomysłowej koncepcji treningu modelu rozpoznawania mowy jest nie budzący zastrzeżeń wybór implementacji jej komponentów składowych, przy czym oprócz narzucających się dla realizacji zadań cząstkowych neuronowych architektur rekurencyjnych (podstawowej RNN i LSTM), Doktorant dodatkowo wprowadza metodę n-gramów jako możliwy, alternatywny względem sieci neuronowych, sposób implementacji modułu modelowania języka. Jednocześnie, ponieważ moduł T2S stanowi w zaproponowanym algorytmie komponent o charakterze pomocniczym, Doktorant buduje go w oparciu o sprawdzoną architekturę Wavenet.

Weryfikacja eksperymentalna zaproponowanej metody jest również przeprowadzona w sposób nie budzący żadnych zastrzeżeń. W celu umożliwienia oceny przydatności swojej koncepcji Doktorant określa metodę referencyjną klasyfikacji, którą jest rekurencyjna sieć neuronowa trenowana w schemacie nadzorowanym. Celem eksperymentów jest nie tylko określenie podstawowych parametrów służących ocenie zaproponowanego algorytmu klasyfikacji, ale również weryfikacja dodatkowego pomysłu poprawy skuteczności jego działania poprzez wprowadzenie dwóch wariantów wstępnego przygotowania (pre-treningu) modułów T2S i S2T. Pierwszy z nich, określony przez Doktoranta jako ‘miękki start’ to inicjalizacja parametrów algorytmu dokonywana na podstawie niewielkich (a więc, dostępnych) etykietowanych baz danych. W drugim scenariuszu, inicjalizacja jest dokonywana tylko dla modułu klasyfikatora (S2T), co Doktorant określa terminem startu ‘pół-miękkiego’.

Uzyskane przez Doktoranta wyniki potwierdzają nie tylko słuszność hipotezy o możliwości zbudowania systemu rozpoznawania mowy wypowiedzianej w języku, który nie jest językiem ojczystym osób mówiących, ale pokazują, że osiągnięta poprawność klasyfikacji przewyższa wyniki możliwe do uzyskania dla metody referencyjnej. Ten wynik, w świetle zupełnie innych uwarunkowań dla budowy modelu ASR dla obydwu przypadków, gdzie proponowana przez Doktoranta metoda jest tworzona na bazie zbiorów nieetykietowanych, podczas gdy druga metoda wymaga takiego zbioru, ma w mojej opinii niezmiernie istotne znaczenie i stanowi dowód na możliwość skutecznego rozwiązania postawionego problemu.

Na dodatkowe uznanie zasługuje podjęta przez Doktoranta próba uzupełnienia opracowanej metody o dodatkową procedurę zmierzającą do uzyskania dalszej poprawy skuteczności rozpoznawania mowy, stanowiąca realizację trzeciego z zapowiedzianych celów badawczych rozprawy. Istotą zaproponowanej koncepcji jest wprowadzanie do wyników rozpoznawania otrzymanych za pomocą omówionego wcześniej algorytmu, korekt wynikających z wiedzy o różnicach między gramatykami i konwencjami języka wypowiedzi i natywnego języka mówcy, nabytej w drodze zastosowania procedur uczenia maszynowego. Jako metodę realizacji korekt Doktorant wykorzystuje dwustopniową strukturę transformera (koder-dekoder), trenowaną na zbiorze par tekstów zawierających wyrażenia niepoprawne i skorygowane. Efektem wprowadzenia zaproponowanej

transformacji jest uzyskanie dodatkowej poprawy rozpoznawania, czyli dalsza poprawa właściwości opracowanego przez Doktoranta modelu ASR.

Podsumowując ocenę dokonań Doktoranta w pierwszym wątku przedstawionych przez Niego prac, chcę stwierdzić, że są one zdecydowanie **znaczące** i w mojej opinii, stanowią **zauważalny wkład** do nauki w skali światowej.

### **3.2 Metoda korygowania fonetycznych błędów artykulacji wypowiedzi**

O ile pierwszym obszarem prac Doktoranta było zbudowanie kompleksowego algorytmu automatycznego rozpoznawania mowy, o tyle drugi zasadniczy wątek prac przedstawiony w rozprawie dotyczy badań prowadzonych nad modyfikacją postaci danych wejściowych – mowy wypowiedzianej przez osobę nienatywną, dopasowującej sposób artykulacji wypowiedzi do kanonu obowiązującego dla rozważanego języka. Zadanie, określone przez Doktoranta jako modyfikacja akcentu wypowiedzi, jest realizowane w dziedzinie spektrogramu, a istotą realizacji zadania jest opracowanie algorytmu dokonującego przekształcenia spektrogramu według ‘wyuczonego’ dla danej grupy etnicznej mówców schematu. Narzędziem realizacji zadania jest albo koncepcja Autoenkodera albo metoda ‘transferu stylu’, a podstawą dla treningu metod jest zbiór danych złożony z par nagrań, zawierających te same zdania, wypowiedziane przez mówców natywnych i mówców z rozważanej, nienatywnej grupy etnicznej. Należy w tym miejscu zaznaczyć, że utworzenie takiego zbioru danych nie jest procesem kosztownym, a więc zgromadzenie odpowiedniej ilości materiału eksperymentalnego jest jak najbardziej realne.

Pierwsza z zaproponowanych strategii wprowadzania korekt ujawnia doskonałą intuicję Doktoranta, który dostrzega podobieństwo podjętego zadania z ‘generycznym’ dla koncepcji Autoenkodera problemem ‘odszumiania’ danych wejściowych. W rozważanym przez Doktoranta zagadnieniu, dane wejściowe to wypowiedzi zaburzone niewłaściwym sposobem akcentowania, a więc efektem redukcji lub eliminacji ‘szumu’ powinna być generacja wypowiedzi z poprawną artykulacją (próbka referencyjna mowy wypowiedzianej przez natywnego mówcę), usuwającą zniekształcenia obecne w próbce wejściowej (wypowiedzianej przez osobę nienatywną).

Druga z rozważanych przez Autora metod jest bardziej złożona, a jej istotą jest próba modyfikacji sposobu artykulacji mowy, bazująca na założeniu możliwości dokonania separacji między komponentami ‘treści’ i ‘stylu’ wypowiedzi, wykorzystywanej w koncepcji tzw. transferu stylu. Podstawę dla możliwości dokonania takiej separacji stanowi założenie, że informacja o treści danych jest wydzielana na etapie filtracji dokonywanej w początkowych warstwach konwolucyjnych sieci. Dlatego też, wyniki oceny przetwarzania danych przedstawiających tą samą treść ale w różniącej się formie, uzyskiwane na wyjściach tych warstw powinny być identyczne. Ta obserwacja jest wyrażona ilościowo jako pierwszy składnik funkcji straty stosowanej do treningu Autoenkodera (równanie 10). Zabiegiem stosowanym dla identyfikacji komponentu stylu zawartego w danych jest przyjęcie, że miarą indywidualnego sposobu prezentacji treści jest statystyczna (pozbawiona informacji o przestrzennym rozkładzie) różnica wyników przetwarzania danych wejściowych przez dalsze warstwy konwolucyjne (równanie 11). Bazując na założeniu możliwości separacji stylu i treści, Doktorant proponuje procedurę, która korzystając z par wypowiedzi: referencyjnej (osoby natywnej) i badanej (osoby nienatywnej) dokonuje ekstrakcji odpowiednich

komponentów, a następnie przekształca, korzystając z modułu Autoenkodera, spektrogram wypowiedzi osoby nienatywnej w sposób redukujący obydwie komponenty funkcji straty (a więc, dąży do zachowania tej samej treści w wypowiedziach referencyjnej i analizowanej oraz upodabnia styl obydwu wypowiedzi).

Wypowiedzi skorygowane z użyciem zaproponowanej przez Doktoranta procedury są następnie przedmiotem rozpoznawania w systemie ASR, wytrenowanym w sposób 'konwencjonalny', a więc na podstawie danych zawierających wypowiedzi natywnych mówców języka. Doktorant rozważył w swoich pracach dwa warianty procedury rozpoznawania: pierwszy bazujący na analizie zmodyfikowanego spektrogramu i wykorzystujący odpowiednio zaprojektowany klasyfikator konwolucyjny, oraz drugi, bazujący na przekształceniu zmodyfikowanego spektrogramu do postaci czasowej i użycie klasyfikatora rekurencyjnego. O ile pierwszy ze wspomnianych sposobów został w pracy opisany w sposób jasny, o tyle takiej jasności nie mam w odniesieniu do drugiego z nich – nie potrafię powiedzieć, jak Doktorant przekształcił spektrogram w sygnał mowy (odwrotna transformacja Fouriera wymaga wiedzy o fazach prążków widma, a te nie są przekształcane).

Podsumowanie eksperymentów ewaluacji zaproponowanych metod korekcji stylu wypowiedzi, dokonane z użyciem właściwych do tego celu miar ilościowych wskazuje na bardzo znaczącą poprawę dokładności rozpoznawania, uzyskiwaną dzięki zastosowaniu obydwu metod. Jednocześnie otrzymane wyniki jednoznacznie identyfikują transfer stylu, jako strategię modyfikacji sposobu artykulacji wypowiedzi dającą lepsze efekty. Bardzo interesującą właściwością metody transferu stylu, na którą zwraca uwagę Doktorant, jest możliwość konfrontacji w trakcie procedury treningowej próbek różnych wypowiedzi, co znacznie wzbogaca liczbę przykładów używanych do treningu. W konsekwencji może to oznaczać, że lepsze wyniki uzyskiwane dla metody transferu stylu są albo wynikiem zastosowania znacznie bardziej ogólnego podejścia do analizy nagrań ('styl' i 'zawartość' są pewnymi pośrednio analizowanymi komponentami), ale też, może wynikać ze wskazanego zwiększenia rozmiarów zbioru danych treningowych (co, w obliczu szczupłości baz możliwych do zgromadzenia dla większości problemów praktycznych, i tak świadczy na korzyść metody transferu stylu).

Podsumowując ocenę prac Doktoranta w obszarze dotyczącym modyfikacji stylu artykulacji, chcę jednoznacznie stwierdzić, że podobnie jak w odniesieniu do poprzedniego wątku, są one **oryginalne, wartościowe** i stanowią **interesujący** i w moim przekonaniu, **zauważalny wkład** do nauki.

#### **4. Mankamenty tekstu**

Rozprawa jest napisana w sposób bardzo jasny, ma poprawny i logiczny układ i zawiera jedynie nieliczne usterki edycyjne lub fragmenty, które wymagają dodatkowego doprecyzowania lub komentarza i które zostaną wskazane w dalszej części niniejszego rozdziału.

Zanim przejdę do ich prezentacji, chciałbym jednak zwrócić uwagę na jedyną dostrzeżoną przeze mnie słabość przedstawionego tekstu, która dotyczy prezentacji kontekstu realizowanych prac. Przedstawiony przegląd stanu wiedzy jest obszerny. Autor przyjmuje w nim konwencję zwięzłej prezentacji głównych treści wybranych przez siebie pozycji, przy czym kryterium wyboru jest

prawdopodobnie istotność, według oceny Autora, wpływu prezentowanych prac na rozwój dziedziny. Przyjęcie takiej konstrukcji jest dla Autora dość wygodne, bo dzięki temu Doktorant unika konieczności dokonania trudnej systematyzacji podejść i metod a więc wskazywania, które wątki stanowią rozwinięcia wcześniejszych pomysłów, a które są zupełnie nowymi propozycjami. Mankamentem przyjętego podejścia jest oczywiście subiektywność wyboru – oprócz pozycji o niepodważalnie fundamentalnym charakterze, Doktorant w przytoczonych opisach omawia sporo prac o znaczeniu, w mojej opinii, niewielkim (co zresztą też jest poglądem subiektywnym). Jednak moje główne zastrzeżenie w odniesieniu do przeglądu stanu wiedzy w obszarze automatycznego rozpoznawania mowy dotyczy pominięcia rozwiązań zaproponowanych w ostatnich latach. Przedstawione przez Autora pozycje są datowane najpóźniej na 2014 rok, co uwzględniając długość cyklu wydawniczego oznacza, że Autor praktycznie wcale nie odnosi się do stanu dziedziny jaki ma miejsce w chwili obecnej. Postępy w tym obszarze, które za sprawą pojawienia się głębokich sieci neuronowych dokonały się w ostatnich ośmiu latach, nie objętych przeglądem są kolosalne (tak jak i w innych obszarach analizy danych). Chcę jednocześnie podkreślić, że wspomniana uwaga dotyczy wyłącznie edytorskiego mankamentu przedstawionego tekstu, a nie kompetencji Doktoranta, który w swoich pracach korzysta z najbardziej aktualnych i zaawansowanych koncepcji istniejących w obszarze tworzenia algorytmów ASR.

Dostrzeżone przez mnie w tekście, nieliczne fragmenty i zdania, których zrozumienie jest trudne lub niemożliwe są następujące:

Str. 29: The skilled speakers with advances semantics and syntax, and phonology related to their language, despite the diversity that is always present in native speech. - nie wiem, co Autor chciał tu powiedzieć.

str. 30: „As the vectors of phonemes existing for every new language that is new, may be listed based on current linguistic information or the extension of the corpora with its phonemic representation, not existing in languages defined before inside the system, could be defined. „ - niestety, tego zdania też nie rozumiem – co jest podmiotem, który ‘mógłby być zdefiniowany’?

Str 69: „As the loss network model for automatic speech recognition tasks, properties of convolutional and recurrent layers were combined, where the former layers become, in fact, employed as feature extractors”. Zdanie jest trudne do zrozumienia: wynika z niego, że modelem są właściwości ..

Prezentacja wyników metody uczenia modelu ASR na rys. 5 jest w mojej opinii niefortunna – wykresy ilustrujące przebieg procesu uczenia modelu powinny być przedstawione dla przypadku, który daje najlepsze efekty (te pokazane prawdopodobnie odpowiadają pierwszemu lub trzeciemu scenariuszowi treningu - są gorsze niż wyniki uzyskiwane dla sieci referencyjnej).

Przedstawione na rys.9 odniesienia do fragmentów pracy dotyczących komponentów przedstawianej metody są niewłaściwe, przy czym wydaje się, że błąd polega między innymi na użyciu niewłaściwej pierwszej cyfry identyfikowanych podrozdziałów (nie 2 tylko 4).

## 5. Wniosek końcowy

Podsumowując przedstawioną recenzję, chcę jednoznacznie stwierdzić, że Doktorant osiągnął wszystkie postawione przed sobą, ambitne cele badawcze i opracował funkcjonalne metody

pozwalające na skuteczną realizację zadania rozpoznawania mowy wypowiedzianej przez osoby nie będące natywnymi mówcami danego języka. Opracowane przez Doktoranta algorytmy są oryginalne, pomysłowe i stanowią cenny wkład naukowy do dziedziny automatycznego rozpoznawania mowy, o istotnym znaczeniu w skali światowej i, jestem przekonany, dużym potencjale oddziaływania.

Konkludując recenzję, chciałbym stwierdzić, że przedłożona do recenzji rozprawa doktorska Pana magistra inżyniera Kacpra Radzikowskiego pt. „A study on speech recognition and correction for non-native English speakers” **spełnia** w moim przekonaniu z dużym nadmiarem wymagania określone w odnośnej ustawie o stopniach i tytule naukowym i tym samym **wnioskuję o dopuszczenie Autora rozprawy do publicznej obrony**.

Dodatkowo, uwzględniając bardzo wysoki poziom merytoryczny zaproponowanych przez Doktoranta rozwiązań chciałbym zwrócić się z wnioskiem o wyróżnienie rozprawy.