

Warszawa, dnia 16.08.2023 r.

**prof. dr hab. inż. Ryszard Romaniuk**  
**Politechnika Warszawska**  
**Wydział Elektroniki i Technik Informatycznych**  
**Instytut Systemów Elektronicznych**

**KWESTIONARIUSZ – RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY DYSZYPLINY  
NAUKOWEJ AUTOMATYKA, ELEKTRONIKA, ELEKTROTECHNIKA I TECHNOLOGIE  
KOSMICZNE POLITECHNIKI WROCŁAWSKIEJ**

**Tytuł rozprawy:** Automatyczna klasyfikacja wybranych zniekształceń sygnałów muzycznych z wykorzystaniem konwolucyjno-rekurencyjnych sieci neuronowych bez porównania do sygnału referencyjnego

**Autor rozprawy:** *mgr inż. Kamila Organiściak*

Podstawą recenzji jest uchwała nr 703/30/RDND02/2021-2024 Rady Dyscypliny naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Politechniki Wrocławskiej z dnia 10 lipca 2023 r. w sprawie wyznaczenia recenzentów w postępowaniu w sprawie nadania stopnia doktora w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne Pani mgr inż. Kamili Organiściak. Uchwała jest podpisana przez Przewodniczącego RDN AEEiTK Politechniki Wrocławskiej prof. dr hab. inż. Andrzeja Dziedzica. Pismo informacyjne w tej sprawie skierowane do mnie od Zastępcy Przewodniczącego RDN AEEiTK Politechniki Wrocławskiej dr hab. inż. Ireneusza Jabłońskiego, profesora uczelni z dnia 31.07.2023 r.

**1. Jaka jest obecność i pozycja doktorantki w środowisku zawodowym i naukowym, zapisy bibliometryczne, publikacje w szczególności związane z pracą doktorską, zespół i otoczenie laboratoryjne doktorantki?**

Pani mgr inż. Kamila Organiściak jest obecna w bazie danych DONA Politechniki Wrocławskiej jako współautorka, na pierwszym miejscu, artykułu Single-ended quality measurement of a music content via convolutional recurrent neural networks, opublikowanym w kwartalniku PAN Metrology and Measurement Systems. 2020, vol. 27, nr 4, s. 721-733. Współautorem jest promotor profesor Józef S.Borkowski. Publikacja ta jest z bazy DONA linkowana do Scopusa oraz do Web of Science, ze względu na indeksowanie kwartalnika MMS w tych bazach bibliograficznych. W związku z tym doktorantka posiada osobiste zapisy w tych kluczowych bazach. Wymieniona publikacja jest także obecna w bibliotece PAN journals.pan.pl, oraz jest indeksowana w bazie międzynarodowej EBSCO.

W kraju doktorantka jest obecna w bazie BazTech prowadzonej przez ICM UW. W zawodowej społecznościowej bazie LinkedIn doktorantka posiada skromny zapis osobowy z informacją o odbytych studiach na Politechnice Wrocławskiej. Podobnie jest w bazie ResearchGate. Tam także znajduje się zapis osobowy doktorantki z informacjami o związku z Politechniką Wrocławską. Wymienione są specjalności – telekomunikacja, elektronika, akustyka, oraz zainteresowania naukowe i umiejętności jak DSP, przetwarzanie mowy, analiza częstotliwościowo-czasowa sygnałów.

RDN AEEiTK/140/2023

Publikacja indeksowana w wymienionych powyżej bazach, opublikowana w r. 2020, dotyczy testów zastosowania konwolucyjnej dwukierunkowej rekurencyjnej sieci neuronowej, określanej w literaturze skrótem CBRNN (convolutional bidirectional recurrent neural network), do obiektywnego pomiaru jakości treści muzycznej. Takich prac w literaturze, dotyczących zastosowania sztucznych sieci neuronalnych ANN, także w wersji CBR i innych wersjach sieci neuronalnych, jest wiele. ANN stanowią efektywne środowisko analizy akustycznej, np.: cytowane w artykule poz. [23], [26], ale także np. 10.21437/Interspeech.2020-2585 – Attentive convolutional recurrent neural network using phoneme-level acoustic representation for rare sound event detection. Ponadto istnieją liczni komercyjni dostawcy rozwiniętych środowisk ANN/CNN do klasyfikacji audio, np.: <https://www.atmosera.com/blog/audio-classification-with-cnns/>. Oprócz tego istnieją zaawansowane, zawodowe i quasi-zawodowe, grupy społecznościowe bardzo aktywne w tematyce estymacji jakości złożonych sygnałów audio. W grupach tych powstaje adekwatne zaawansowane oprogramowanie, najczęściej dostępne publicznie dla uczestników grup.

Autorzy twierdzą, że proponowaną w pracy nowością jest efektywny model CBRNN w którym detekcja anomalii akustycznych nie wymaga zapewnienia czystego sygnału odniesienia. Sygnał muzyczny jest rozważany jako znacznie bardziej złożony od czystej mowy, śpiewu, pojedynczego instrumentu, a także sygnałów generowanych sztucznie. Sygnał muzyczny zawiera często wszystkie takie składniki jednocześnie, multiplikowane w różnych, bardzo zmiennych proporcjach w domenach czasu, częstotliwości i przestrzeni. Ogólnie można twierdzić, że dla takich sygnałów nie ma łatwego dostępnego odniesienia i stosownych wzorców. Z drugiej jednak strony, można powiedzieć, że przy obecnych możliwościach systemów DSP oraz technik Big Data, budowa bardzo dużych baz danych sygnałów odniesienia nie stanowi problemu.

Biorąc zapewne to pod uwagę autorzy twierdzą, że praca jest jedynie rodzajem sprawdzenia koncepcji modelu, proof-of-concept. Choć, w artykule stwierdzają wprost, że zgodnie z ich najlepszą wiedzą takie bazy nie istnieją. Badane artefakty obejmują: błędy kwantyzacji, brakujący dźwięk, zniekształcenie charakterystyki wzmocnienia, dodatkowy dźwięk szumowy. Autorzy twierdzą, że zaproponowany model może być rozszerzony w celu detekcji innych zniekształceń złożonego sygnału muzycznego. Także, może być zastosowany jako wstępnie wytrenowany model w innych procesach transferu uczenia się. Najważniejszą częścią tej pracy są testy wydajności opracowanego modelu przy pomocy różnych eksperymentów praktycznych. Surowe próbki audio były transformowane do spektrogramów melowych. Spektrogramy stosowano jako wejście dla modelu początkowo niezależnie a następnie uzupełnione o dodatkowe cechy jak kontrast spektralny SC, częstość przecięcia zera ZCR. Uzyskano znaczny wzrost dokładności działania modelu w przypadku rozszerzonej informacji wejściowej. Analizowano wpływ warstw rekurencyjnych na sprawność klasyfikacji artefaktów. Celem pierwotnym pracy jest efektywna klasyfikacja czy sygnał audio zawiera, czy nie zawiera artefaktu. Dalszym celem jest klasyfikacja rodzaju artefaktu i określenie jakie artefakty są najtrudniejsze do detekcji dla opracowanego modelu. Celem planowanym jest rozszerzenie tych badań na różne rodzaje muzyki.

Artykuł zawiera materiał, który jest następnie w rozszerzonej i odpowiednio przeredagowanej wersji umieszczony w pracy doktorskiej. W pewnym sensie, opracowany tam model jest rdzeniem części modelowej rozprawy. Ponieważ jest to jedyny opublikowany artykuł doktorantki, dla recenzenta jest istotny charakter tej pracy i jej powiązanie z rozprawą doktorską. Artykuł ma charakter mieszany przeglądowo - przyczynkowy, podczas gdy oceniana praca ma typową strukturę rozprawy doktorskiej o objętości 94 stron. Artykuł jest dość długi, ma 13 stron i zawiera sporo materiału przeglądowego z dość liczną listą odnośników literaturowych - aż 41 pozycji. Inaczej mówiąc, artykuł jest prezentacją

własnego, relatywnie wąskiego, specjalistycznego materiału przyczynkowego na dość szerokim tle dużego i interesującego obszaru badawczego jakości sygnałów audio. Struktura artykułu jest następująca: wstęp, detekcja artefaktów, opis modelu i danych eksperymentalnych, wyniki, wnioski i podsumowanie.

We wstępie autorzy dzielą testy jakości audio na subiektywne i obiektywne – inwazyjne i nieinwazyjne. Testy subiektywne, kosztowne, manualne i długotrwałe, są trudne do obiektywizacji, ale są standaryzowane przez ITU-RBS.1284-2. Testy obiektywne są trudne do modelowania, ale jest ich wiele. Autorzy wymieniają predyktory jakości audio wymagające odniesienia, i powołują ich źródła: PEAQ/PESQ, POLQA, PEMO-Q, STOI, VISQOL, pocziwy klasyczny współczynnik szumu SNR. Wymienione są rozwiązania bezreferencyjne mierzące np. jakość mowy: ANIQUE, HASQI, POSQE, SRMR ale także syntetycznego audio, obrazów i wideo. Metryki inwazyjne wymagają porównania zdegradowanego sygnału testowego z odniesieniem w postaci oryginalnego sygnału nieprzetworzonego. Konieczny jest dostęp do pliku z sygnałem pierwotnym. W układach rzeczywistych złożonych, wieloelementowych, często takiego dostępu nie ma. Stosuje się wówczas korekcję kanałową, jednak niewystarczającą w systemach immersyjnych i interakcyjnych, ze względu na ograniczoną kombinację kanałów. W takich systemach wprowadzono zróżnicowane, wysoce efektywne, formaty obiektowe. Ocena jakości, mimo dostępu do pliku wejściowego, wymaga wówczas dodatkowo zastosowania adekwatnego zestawu metadanych lub wpierw renderowania zawartości. Opcją jest zastosowanie metody nieinwazyjnej, i metryk nie wymagających sygnału odniesienia, estymujących jakość audio na podstawie sygnału użytkownika. Możliwy jest ciągły monitoring jakości sygnału dostarczanego do użytkownika, lub testy regresyjne dla określonego węzła systemu rozgłoszeniowego. Motywacją do badań autorów w tym obszarze (2020 r.) jest stwierdzenie, że zgodnie z najlepszą wiedzą nie istnieją rekomendowane bezreferencyjne metody estymacji jakości rzeczywistego złożonego sygnału muzycznego.

W artykule badany jest relatywnie prostym model bazujący na CBRNN. Model jest porównany z klasycznymi rozwiązaniami estymatorów jakości na CNN. Autorzy twierdzą, że model ma mniejsze wymagania pamięciowe i zwiększoną efektywność statystyczną. Model składa się z dwóch bloków funkcjonalnych, przetwarzania wstępnego i klasyfikacji. Na wejściu standardowy plik audio 44.1 kHz, trwający 3 sek. jest konwertowany do spektrogramów melowych o 256 pasmach i standaryzowany. W bloku klasyfikacyjnym następuje ciąg operacji typowych dla CBRNN umożliwiając w sposób potokowy głębokie uczenie na danych sekwencyjnych. Model trenowano na spektrogramach melowych ekstrahowanych z surowych próbek audio. Model optymalizowano wydajnościowo w funkcji liczby cech i technik ekstrakcji.

W części eksperymentalnej autorzy korzystają z bazy MUSDB18 i uzasadniają ten wybór. Przy okazji omawiają niektóre właściwości innych baz. Tworzą pięć w pełni etykietowanych podzbiorów, jeden oryginalny i cztery z różnymi kategoriami artefaktów – zniekształcenie wzmocnienia, brakujący dźwięk, defekt kwantyzacji oraz dodatkowy dźwięk szumowy. Charakteryzują wymienione kategorie. Przedstawiają spektrogramy melowe badanych sygnałów audio, oryginalnego i zniekształconych, dodatkowo z trzaskami. Zbiory danych autorzy podzielili na niepokrywające się podzbiory - trenujący, testowy i walidacyjny. W artykule nie dokonano dodatkowego podziału zbiorów danych w zależności od rodzaju muzyki, anonsując taki podział do dalszych prac i określenia efektywności modelu dla różnych typów muzyki. Autorzy opisują kolejno sposób wstępnego przetwarzania danych, architekturę i metodologię zastosowania modelu, oraz uzyskane wyniki. Wstępnego porównania efektywności klasyfikacji dokonano między architekturami CNN i CBRNN z zastosowaniem metryk specyficzności, precyzji, zapamiętania, testu F1 i ogólnej dokładności

oraz spektrogramów melowych skalowanych do 256 pasm. Przedyskutowano różnice uzyskując przewagę efektywności w modelu CBRNN we wszystkich badanych kategoriach. Wykonano szereg eksperymentów porównawczych i testujących model CBRNN dla różnej liczby cech wejściowych i ich zawartości. Zastosowano technikę transferu uczenia w celu zbadania wpływu dodatkowych cech i ich zmienności. Istniejący model CBRNN zastosowano jako zintegrowany ekstraktor cech w dwóch innych i różniących się od siebie modelach – odpowiednio z ekstraktorem cechy ZCR i SC. Porównano charakterystyki warstw obu modeli. Wyniki prac w artykule mają charakter typowo akademicki. Są pokazaniem zasady a nie opracowaniem gotowego zamkniętego produktu programistycznego. Autorzy w zakończeniu wymieniają brakujące elementy zarówno do poprawy efektywności oprogramowania jak i do produktu, zdając sobie sprawę, że są swoim opracowaniu odlegli od niego. Wśród niezbędnych uzupełnień wymieniają: ekstrakcję dodatkowych cech, dodanie klasyfikacji wielo-etykietowej, rozszerzenie bazy danych o więcej próbek muzyki, w szczególności z zawartością treści elektronicznych i syntetycznych.

Podsumowując, twórcza obecność naukowa doktorantki w publicznym środowisku zawodowym i naukowym jest dość skromna. Jednak z formalnego punktu widzenia jest wystarczająca do ubiegania się o stopień naukowy doktora nauk inżynierjno-technicznych w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne. Zakładając 50% udział doktorantki w analizowanym powyżej artykule, recenzent uznaje, że merytoryczna zawartość naukowo-techniczna jest wystarczająca do spełnienia formalnych warunków dorobku w procesie doktoryzowania.

**2. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autorkę? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?**

Obszarem badawczym pracy doktorskiej są studia nad zastosowaniem konwolucyjno-rekurencyjnych, sztucznych sieci neuronowych do analizy jakości sygnału audio w wersji monofonicznej. To pewne, a nawet znaczne, uproszczenie. Obszar badań jakości złożonych sygnałów audio budzi cały czas duże zainteresowanie w świecie akademickim i naukowym, ale jeszcze bardziej w świecie biznesowym i przemysłowym. Stosowanych jest wiele różnych metod badawczych. Szczególną rolę zawierają metody sztucznych sieci neuronowych w różnych rozwiązaniach. Ostatnio stosowane są zaawansowane metody sztucznej inteligencji. Praca doktorska ogranicza to wielkie pole do metody CBR-ANN (CBRNN) i do złożonych sygnałów muzycznych, a także do bazy danych MUSDB18 i rekomendacji ITU BS.1284-2 na podstawie których opracowano własny zbiór sygnałów testowych.

Zagadnieniem naukowym rozpatrzonym w pracy jest realizacja, z wykorzystaniem CBRNN, bezreferencyjnej, automatycznej klasyfikacji wybranych zniekształceń w złożonych sygnałach muzycznych. Wynikiem realizacji pracy doktorskiej jest adekwatne oprogramowanie klasyfikujące bazujące na autorskim modelu prototypowym. Model ten był w pracy testowany wieloparametrowo pod względem efektywności klasyfikacji na własnym zbiorze sygnałowym. Tezą pracy jest pokazanie możliwości realizacji skutecznego, w porównaniu z klasycznymi rozwiązaniami stosującymi ANN, klasyfikatora bezreferencyjnego bazującego na CBRNN. Rozwiązywane w rozprawie zagadnienie naukowe jest odpowiednio sformułowane na tle problematyki badawczej oceny jakości złożonych sygnałów audio.

Podsumowując, recenzent stwierdza, że zagadnieniem naukowym rozpatrzonym w pracy jest samodzielne opracowanie oryginalnego modelu, korzystając z CBRNN, przygotowanie oprogramowania klasyfikującego złożone sygnały audio, oraz środowiska testowego weryfikującego charakterystyki i efektywność klasyfikatora. Wymienione zagadnienie

naukowe zostało odpowiednio jasno sformułowane przez autorkę. Praca ma charakter, zarówno teoretyczny w warstwie tworzenia modelu, oraz doświadczalny, w warstwie testowania modelu. Zagadnienie naukowe posiada pod względem merytorycznym wagę w obu tych warstwach odpowiednią dla prac doktorskich w dziedzinie nauk inżyniersko-technicznych i dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne.

**3. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań / świadczący o dostatecznej wiedzy autorki. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?**

Praca zawiera znaczną liczbę odwołań literaturowych, 163 pozycje. Powołania rozpoczynają się od rozdziału 1 dotyczącego ogólnej charakterystyki oceny jakości sygnału audio. Są to pozycje bardzo ogólne, łagodnie wprowadzające w tematykę różnych powodów trudności oceny jakości sygnału audio, od ograniczeń pasma i technicznych, charakterystyki dźwięku, po fizjologiczne, percepcyjne, kodeki percepcyjne, pasma krytyczne, rodzaje maskowania, obszar słyszalności, psychoakustyczne, itp. Przetwarzanie cyfrowe i transmisja strumieniowa audio ma wpływ na jakość dźwięku. Istotna jest budowa całego łańcucha przetwarzania i transmisji od generacji dźwięku po odtwarzanie. Istotny jest potencjalnie dostęp do wszystkich rodzajów plików dźwiękowych na tej drodze. W tym łańcuchu autorka omawia właściwości kodowania cyfrowego i charakterystyki kodeków stratnych i bezstratnych, także powołując ogólną literaturę. Wymienia podstawowe zniekształcenia kodeków jak błędy kwantyzacji, aliasing, błędy algorytmów.

Rozwój konfiguracji systemów audio w kierunku surround i obecnie immersive wymaga sprawnego przetwarzania sygnałów w górę i w dół z odpowiednią liczbą obsługiwanych kanałów. Techniki są określane terminami upmix i downmix. Autorka wymienia kontrolery zakresu dynamiki i ich właściwości – bramkowanie, limityery, kompresory i ekspandery, także standaryzatory poziomu sygnału wg. normy EBU R128. Kolejnym komponentem systemu jest transmisja sygnału audio. Inne możliwości interakcji z sygnałem dźwiękowym polegają np. na jego znakowaniu wodnym co ma znaczenie w ochronie praw autorskich. Wymienione powyżej zagadnienia obejmują cytowania w zakresie pozycji [1-17].

Rozdział 2 pt. manualne subiektywne metody oceny jakości audio dotyczy problematyki nie będącej przedmiotem rozprawy, ale autorka dodaje go aby zarysować całość problematyki i skontrastować z metodami obiektywnymi maszynowymi. Literatura związana z rozdziałem 2 oceny subiektywnej obejmuje takie zagadnienia jak: standardy ITU, dobór grupy słuchaczy wyszkolonych, procedury i metody testowe, sposób oceny jakości, błędy testów odsłuchowych. Powołana literatura obejmuje adekwatne i liczne standardy ITU w tym zakresie i pozycje [18-19].

Rozdział 3 pt. automatyczne metody oceny jakości audio powołuje literaturę w zakresie problematyki zasad testów automatycznych, podział metod na trzy kategorie – referencyjne, bezreferencyjne i parametryczne, charakterystyki poszczególnych automatycznych kategorii testowych, wrażliwości metod w domenie czasu na przesunięcia fazowe, metody referencyjne PEAQ, PEMO-Q, ViSQOLAUDIO, metody bezreferencyjne ANIQUE, HASQI, POSQE, SRMR, WaveNets, Quality-Net, zastosowanie ANN do przetwarzania muzyki. Wymieniona tutaj tematyka ogólna wprowadzająca w pracę doktorską jest łączona z pozycjami literaturowymi [20-40], oraz w zestawieniu tabelarycznym zastosowań ANN w pozycjach [41-73] i tabelach podsumowujących metody referencyjne [74-103] i bezreferencyjne [104-119].

Rozdział 4 opisujący prace własne nad metodą bezreferencyjną odwołuje się do rekomendacji ITU BS1284-2 [18] i pozycji literaturowych [120-122] dotyczących ogólnych właściwości sieci ANN z warstwami konwolucyjnymi i rekurencyjnymi. Rozdział 5 dotyczący

autorskiej bazy danych audio powołuje pozycje literaturowe [123-143]. Odwołania dotyczą problematyki pre-emfazy, SNR, operacji DFT, STFT, skali melowej, parametrów ZCR, OBSC, LKFS, parametrów głośności, ważenia K, true-peak i rekomendacji ITU-R BS1770-4, skali dBTP.

Rozdział 6 dotyczący projektu modelu powołuje pozycje literaturowe [144-156] i szereg poprzednich. Problematyka odwołań dotyczy następujących zagadnień: zasad modelowania w sieciach ANN i CNN oraz RNN, RNN typu LSTM, BRNN, schematy modeli, parametryzacja kontroli procesu uczenia się - hiperparametry, główne warstwy sieci i ich operacje, konwolucyjne, rekurencyjne, aktywacyjne, klasyfikacja wieloetykieta, funkcje charakterystyczne – kosztów, optymalizująca, aktywacji (ReLU, softmax), mapy atrybutów, filtry konwolucyjne, podpróbkiowanie, operacje grupowania, rozwijania, itp.

Rozdział 7 dotyczący ewaluacji modelu powołuje pozycje literaturowe [157-160] poruszające problematykę implementacji modelu, modyfikacji i transferu uczenia, wymagań sprzętowych. Rozdziały 8-10 dotyczące modeli sieciowych odpowiednio konwolucyjnych, konwolucyjno-rekurencyjnych, i konwolucyjno-rekurencyjnych z zastosowaniem dodatkowych parametrów wejściowych powołuje literaturę związaną z siecią konwolucyjną dwukierunkową LSTM, pozycje [161-163].

Powołania literaturowe są równomiernie rozłożone w pracy i adekwatne do prowadzonego wywodu merytorycznego. Analiza tych powołań wskazuje na fakt zdobycia przez autorkę znacznej wiedzy w bezpośrednim obszarze tematyki pracy i obszarze szerzej otaczającym tą tematykę. Podsumowując, recenzent stwierdza, że w rozprawie przeprowadzono w sposób właściwy analizę źródeł, w tym literatury światowej, stanu wiedzy i zastosowań, świadczący o dobrej wiedzy autorki. Wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący.

#### **4. Czy autorka rozwiązała postawione zagadnienia, czy użyła właściwej do tego metody i czy przyjęte założenia są uzasadnione?**

Praca jest napisana w języku polskim. Zawiera 94 strony, i rozpoczyna się streszczeniami po polsku i angielsku, wykazem wybranych używanych skrótów anglojęzycznych i oznaczeń, oraz nienumerowanym wprowadzeniem. Struktura rozprawy doktorskiej jest następująca. Wprowadzenie zawiera zbiór podstawowych pojęć związanych z badaniem jakości sygnałów audio, np. PCM, PEAQ, listę metod inwazyjnych – referencyjnych i nieinwazyjnych – bezreferencyjnych, odniesienie do standardu-rekomendacji ITU BS.1284-2, oraz krótkie omówienie zawartości poszczególnych rozdziałów pracy. Po wprowadzeniu autorka przedstawia explicite tezę, cel i zakres pracy, co nie jest ściśle formalnym, ale bardzo często zwyczajowym i dość silnym nieformalnym wymogiem.

Nawet bez sformułowania tych kategorii – teza, cel, zakres, w sposób odrębny we wprowadzeniu są one wyraźnie widoczne w kilku odpowiednich częściach tekstu pracy, zarówno w opisie motywacji, usytuowaniu prac własnych na tle szerszego obszaru badawczego, potem w części dotyczącej prac własnych, jak i w podsumowaniu. A więc mamy tezę, cel i zakres powtórzony w pracy wielokrotnie, a tutaj we wprowadzeniu, niejako wyciągnięte, być może dla wygody recenzenta?, przed nawias. Jak już wspomniano powyżej, teza, cel i zakres pracy są sformułowane jasno i jak najprościej to możliwe. W szczególności w liście zadań do zrobienia, wymienionych w zakresie pracy autorka ogranicza się głównie do wymienienia autorskich, oryginalnych działań własnych.

Rozdziały od 1 do 3.3 są przyjaznym dla czytelnika ogólnym, podstawowym wstępem podręcznikowym i jednocześnie przeglądem literatury dotyczącym szerokiego tła naukowo-technicznego pracy doktorskiej. Rozdziały te nie tyle dodają pracy doktorskiej wartość merytoryczną ile, po pierwsze znacznie ułatwiają jej czytanie, po drugie sytuują dokładnie prace własne na odpowiednim tle, ale co najważniejsze pokazują że autorka przestudiowała znaczną

liczbę odpowiednich pozycji literatury i posiadała odpowiednią wiedzę do zrealizowania doktoratu. I to jest najważniejszy fakt dla recenzenta dotyczący tej części pracy. Dodatkowo, autorka potrafiła przekazać tę wiedzę w odpowiednio syntetyczny sposób dopasowany do własnych treści naukowych.

Rozdział 3.4. jest dalszą częścią ogólnego wstępu, ale ponieważ zawiera już wyłącznie treści związane z metodami bezreferencyjnymi stosowanymi w pracy, zawiera również w tabeli 3.4. porównanie z pracą własną. Autorka pisze tam odważnie, że porównanie dotyczy ostatniej dostępnej publikacji [119]. Nie kwestionując tego stwierdzenia, sygnowanego dodatkowo datą, należy stwierdzić, że prowadzone są bardzo liczne i bardzo zaawansowane prace w tym zakresie albo niepublikowane w obszarze komercyjnym, oferowane jako środowiska własnościowe, ale także prowadzone są zaawansowane działania standaryzacyjne i społecznościowo-przemysłowe w globalnych grupach roboczych koordynowanych przez przemysł ale i przez adekwatne zawodowe organizacje międzynarodowe np. IEEE. Do wielu z tych grup wymagane jest logowanie i uzyskanie członkostwa wymaga wsparcia afiliacyjnego instytucji zatrudniającej. Należy podkreślić, że rola takich silnych grup społecznościowo-standaryzacyjnych jest w badaniach naukowych bardzo ważna, nie tylko w badaniach podstawowych ale i w inżynierii-technicznych.

W dzisiejszych czasach źródła naukowe nie ograniczają się do publikacji w artykułach w czasopismach archiwalnych. W szczególności w odpowiednio bogato punktowanych przez nasze administracyjne władze naukowe. Bardzo ciekawe materiały można znaleźć w arXiv, a znakomite oprogramowanie w repozytoriach GitHub, itp. Aktywność w tym zakresie prowadzi ASA [acousticalsociety.org]. Autorka o tym nie wspomina. Czy te zasoby przeszukiwała? Czy zagłębiała do zasobów kluczowych standaryzacyjnych grup roboczych?

Rozdziały 4-10 już wyłącznie są skoncentrowane na pracy własnej, ale także zawierają sporo odnośników do odpowiedniej literatury specjalistycznej. Krótki rozdział 4 wywodzi zaproponowaną własną metodę automatycznej bezreferencyjnej klasyfikacji sygnału audio w wersji mono z rekomendacji ITU BS.1284-2. Autorka zaznacza, że ogranicza liczbę docelowych kategorii do wykrycia do czterech pomijając inne występujące w układach wielokanałowych i stereo, jak korelacje, przesłuchy, cechy przestrzenności, itp. Wybrane kategorie to defekty kwantyzacji związane z ograniczonej rozdzielczości bitowej dla sygnału audio, zniekształcenia charakterystyki wzmocnienia, dodatkowy dźwięk np. szum, brakujący dźwięk wskutek utraty pewnych składowych. Detekcję zniekształceń autorka traktuje jako wielomianowy problem klasyfikacji. Funkcja klasyfikatora jest parametryzowana za pomocą modelu ANN. Danymi wejściowymi są kategoryzowane spektrogramy.

Detekcja zniekształceń polega na znalezieniu modelu  $h$ , który dla każdej instancji spektrogramu  $x$  przypisze wartość prawdopodobieństwa warunkowego  $P$  przynależności do danej klasy  $y$ :  $h(x) = \arg \max P(x|y)$ ,  $y$  należy do  $C$ . Warstwy konwolucyjna i rekurencyjna w ANN zwiększają efektywność modelu, zwiększają skuteczność przetwarzania sygnału audio i detekcji zdarzeń akustycznych, zmniejszają złożoność obliczeniową, pozwalają na analizę zdarzeń sekwencyjnych. Szczegółowe role tych warstw dla poszczególnych rozwiązań modelu autorka prezentuje odrębnie odpowiednio w rozdziałach 6-10. Rozdział 4 autorka kończy listą etapów implementacji modelu: przygotowanie bazy sygnałów audio, projekt modelu, wstępne testy, ulepszenia i modyfikacje oraz uczenie, model optymalny, walidacja. Autorka testowała model najpierw dla każdego sygnału muzycznego a potem sprawdzała jaki wpływ na skuteczność klasyfikacji mają dodatkowe parametry ekstrahowane z sygnałów wejściowych.

Rozdział 5 opisuje sposób utworzenia bazy danych audio i wymagane do testów parametry obiektów w tej bazie. Wymienia po kolei standardowo wymagane w takich przypadkach cechy jak czasowe, widmowe, cepstralne, percepcyjne, obrazowo-spektrograficzne. Następnie wymienia te cechy które zostały wybrane do badań i opisuje je szerzej. Utworzona przez autorkę baza danych ma raczej charakter wygodny, testowy, podręczny, ograniczony dla swoich modeli, niż ogólny. Natomiast, kilkukrotnie w pracy powtarza, że według najlepszej wiedzy, nie jest

znana oficjalna...., tak jakby się bała, że w innym przypadku wartość zrealizowanej pracy byłaby mniejsza. W pracy doktorskiej nie chodzi o bicie oficjalnych rekordów świata, a rzetelne przeprowadzenie wywodu naukowo-technicznego na dobrym poziomie akademickim. Z takim przypadkiem rzetelnego wywodu mamy tutaj do czynienia. Dodatkowo, jeśli już, to rekordy świata w pracach doktorskich bije się niestety w języku angielskim.

Autorka korzysta z bazy MUSDB18 i opracowuje własny odpowiedni wybór parametrów wymaganych do testowania w swoich warunkach. Dalej, w rozdziale 5 opisuje kolejne etapy swoich działań czyli: po implementacji programu dokonuje wstępnego przetwarzania sygnałów wejściowych do formatu możliwego do analizy przez model konwolucyjno-rekurencyjny ANN. Porusza problem konieczności stosowania pre-emfazy, poprawy SNR, obliczania dla ramki okienkowej DFT w wersji STFT, konwersji do skali melowej, zastosowania parametrów ZCR, OBSC, L/ML, True-Peak, i odwiecznego kompromisu między możliwą redukcją złożoności obliczeniowej, czasu przetwarzania i wystarczającej ilości danych do uczenia implementowanego modelu.

Rozdział 6 opisuje projekt modelu klasyfikacji zniekształceń dla sygnałów muzycznych. Autorka opisuje założenia i ogólny schemat działania modelu, parametry stosowane do kontrolowania procesu uczenia modelu nazywane hiperparametrami, główne warstwy sieci neuronowej i operacje z nimi związane – konwolucyjne, rekurencyjne i funkcje aktywacji. Schemat działania modelu i wybór hiperparametrów jest typowy dla zastosowanego rodzaju sieci neuronowych i nie wymaga specjalnego komentarza. Dalszy tekst opisowy obu rodzajów sieci i operacji podpróbkowania, grupowania, rozwijania, obliczania wag, dwukierunkowości, wybranych funkcji aktywacji ReLu i softmax, itp. jest także typowy, nie wnosi wkładu własnego, ale pokazuje odpowiednie zaznajomienie autorki z tematem co jest warunkiem koniecznym prawidłowego wykonania projektu CBRNN.

Rozdział 7 dotyczy niezbędnego etapu budowy modelu czyli jego ewaluacji i, jak autorka twierdzi, podsumowuje trzy etapy badań: implementację, modyfikację i aplikację do transferu uczenia. Autorka wymienia typowe, stosowane w takich przypadkach, miary klasyfikacji, przypominając ich definicje, jak dokładność, specyficzność, precyzja, czułość, współczynnik F1. Obliczenia wykonuje autorka w serwisie AWS na instancji AECC2 inaczej Amazon EC2. Opis zadeklarowanych wymagań sprzętowych i parametrów obliczeń zamieszcza autorka w rozdziale 7.3. Wirtualną maszynę obsługuje serwis ASSS, inaczej Amazon S3. Autorka przedstawia graficznie środowisko przygotowania ewaluacji modelu ANN na rys. 7.1.

Rozdziały 8, 9 i 10 posiadają podobną strukturę i opisują kolejno modele sieciowe konwolucyjne, konwolucyjno-rekurencyjne, oraz konwolucyjno-rekurencyjne z zastosowaniem dodatkowych parametrów wejściowych. Rozdziały posiadają wprowadzenie, opis architektury modelu kolejno CNN, CBRNN/LSTM, oraz CBRNN z parametrami ZCR, OBSC i L/ML, TruePeak, tabelaryczne zestawienie architektury zaimplementowanego modelu z charakterystyką warstw, tabelaryczne zestawienie wyników ewaluacji modelu. Każdy z tych rozdziałów "sieciowych" posiada odrębne podsumowanie z osiągniętymi wartościami wymienionych poprzednio metryk. Rozdział 10 zawiera dodatkowo osobne omówienie zastosowania poszczególnych parametrów dodatkowych i uzyskanych wyników w efekcie ich zastosowania. Ogólnym wnioskiem z rozdziałów 8, 9 i 10 jest możliwość wyboru optymalnej architektury modelu. Autorka dokonuje wyboru prototypu w postaci rozwiązania CBRNN z parametrem OBSC i uzasadnia ten wybór wynikami testów.

Rozdział 11 jest podsumowaniem całego wysiłku autorki włożonego w realizację rozprawy doktorskiej. Jeszcze raz pisze co było tematem, celem i zakresem pracy. Ponownie różnicuje metody referencyjne i bezreferencyjne, jakby dodatkowo silniej uzasadniając swój wybór metody bezreferencyjnej. Odnosi się ponownie do kilkakrotnie wymienianej tutaj rekomendacji ITU. Ponownie wspomina o podziale pracy na trzy części w postaci rozdziałów 1-3 wstęp teoretyczny, 4-7 proponowana metoda detekcji i klasyfikacji zniekształceń w rzeczywistych sygnałach muzycznych pozwalająca na analizę automatyczną bezreferencyjną, i 8-10 zbiór



wyników badań zaprojektowanej i zaimplementowanej metody. Recenzent dokonał nieco innego podziału tej pracy, co zostało opisane powyżej.

Rozdział 11 zawiera tabelaryczne porównanie ogólnej skuteczności klasyfikacji badanych w pracy sieci ANN i ponowne omówienie występujących różnic. Dalej w podsumowaniu, autorka pisze, że w ten sposób potwierdziła postawioną na początku tezę, wymienia największe zalety opracowanego modelu prototypowego, wskazuje na kierunki możliwych dalszych badań, oraz ponownie wymienia wszystkie najważniejsze elementy wkładu autorskiego.

Podsumowując, recenzent stwierdza, że autorka rozwiązała bardzo dobrze postawione zagadnienia, użyła właściwej do tego metody i przyjęte założenia były uzasadnione.

#### **5. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autorki, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?**

W powyższych częściach tekstu niniejszej opinii recenzent wymienia w sposób rozproszony oryginalne, własne elementy autorskie. Autorka ułatwia dodatkowo pracę recenzentów wymieniając w kilku miejscach rozprawy najważniejsze elementy opracowań własnych. Takie listy wkładu własnego znajdują się zarówno we wstępie jak i podsumowaniach rozdziałów 8, 9, 10 oraz w rozdziale 11 podsumowującym. Recenzent nie będzie powtarzał tej listy, potwierdzając tutaj pełną zgodność z deklaracją autorki. Praca zawiera wiele elementów autorskich o charakterze nowatorskiego opracowania własnego. Najważniejsze z tych wszystkich elementów było, że złożyły się one w działający efektywny model klasyfikacji opracowany w środowisku CBRNN. Sformułowanie rozprawy doktorskiej poprzedził opublikowany artykuł w kwartalniku PAN MMS.

Pozycja rozprawy w stosunku do stanu wiedzy i poziomu techniki reprezentowanych przez literaturę światową jest bardzo dobra. Należy jednak zwrócić uwagę, że jest to praca o charakterze akademickim, w którą możliwe było zaangażowanie bardzo skromnych środków materialnych i ograniczonego wysiłku osobowego, głównie doktorantki, zapewne z życzliwą naukową poradą i pomocą promotora. W warunkach przemysłowych i komercyjnych w takie prace angażowane są zupełnie inne środki i wykonuje się ją w inny sposób.

Podsumowując, recenzent stwierdza że waga naukowo-techniczna w obszarach teoretycznym i eksperymentalnym przedstawionych w rozprawie oryginalnych wyników działań doktorantki jest formalnie i zwyczajowo wystarczająca z nadmiarem do przyznania stopnia doktora nauk inżynierjno-technicznych w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne. Praca w naturalny sposób zawiera także znaczny wkład w dyscyplinę informatyki technicznej i akustyki technicznej.

#### **6. Czy autorka wykazała umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników /zwięzłość, jasność, poprawność redakcyjna rozprawy/?**

Statystycznie rzecz biorąc z doświadczenia recenzenta wynika, że panie doktorantki piszą rozprawy doktorskie bardzo porządnie. Tutaj mamy taki wzorowy przypadek. Praca jest przygotowana formalnie, merytorycznie i edycyjnie bardzo poprawnie. Wyniki badań są omówione bardzo przekonująco. Praca jest w odpowiedni sposób zwięzła. Części opisowe prezentujące tło badań są dostatecznie powściągliwe, tak aby nie nadużywać cierpliwości specjalistów. Jednocześnie części zawierające opracowania autorskie są prezentowane na odpowiednim poziomie szczegółowości. Ten poziom szczegółowości pokazuje krok po kroku wysiłek intelektualny autorki prowadzący do weryfikacji postawionej tezy.

Podsumowując, recenzent stwierdza że autorka wykazała w ocenianej rozprawie umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników. Zwięzłość, jasność, oraz poprawność redakcyjna rozprawy nie budzą zastrzeżeń.

## 7. Jakie są słabe strony rozprawy i jej główne wady?

Rozprawa ma skromny przyczynkowy charakter akademicki. Została zrealizowana skromnymi środkami jako test wybranej metody. Nie była realizowana w ramach żadnego, dodatkowo finansowanego, szerszego projektu badawczego. Przynajmniej nic na ten temat recenzentowi nie wiadomo. Gdyby była realizowana w ramach projektu jej układ i zakres byłby prawdopodobnie inny. Tematyka prezentowana w pracy bez wątpienia zasługuje na ujęcie i zgłoszenie programowe. Jeśli powyższe uwagi można byłoby potraktować jako potencjalnie słabe strony to oczywistym jest, że nie jest to w żadnym wypadku wina ani zasługa doktorantki.

Czy wadą jest że projekt i testowa wersja oprogramowania nie zostały doprowadzone do postaci zamkniętej, eksploatacyjnej? Autorka nazywa wersję końcową swojego opracowania wersją prototypową. Trochę chyba na wyrost, w odniesieniu do formalnej nomenklatury TRL – poziomów gotowości technologicznej. Ilość pracy wymagana do domknięcia jest tak znaczna że przekraczała bez wątpienia możliwości doktorantki. Doktorantka wymienia co trzeba byłoby zrobić, aby projekt uzupełnić. Ale to oczywiście nie wszystkie potencjalnie wymagane składniki. Projekty zamknięte realizuje się w większych zespołach badawczych i to najczęściej, jeśli są realizowane w warunkach akademickich, na zlecenie dużych kontrahentów naukowych. Tutaj nie mamy takiego przypadku. Na marginesie warto zapytać dlaczego? Tematyka jest odpowiednia i warta takiego wysiłku.

## 8. Jaka jest przydatność rozprawy dla nauk technicznych?

Praca ma wąski, czysto akademicki charakter przyczynkowy. Realizacja wielu takich prac, rozwiązujących szczegółowe zagadnienia, takie jak w niniejszej pracy, stanowi pewien front badawczy, który zaczyna w pewnych przypadkach budzić zainteresowanie przemysłu i komercji. Musi zostać przekroczony pewien próg krytyczny sumy wagi takich niewielkich kontrybucji. Chyba, że waga indywidualnego wkładu jest rzeczywiście ciężka.

Obszar akustyki muzycznej jest obszarem bardzo specjalnym, opanowanym w dużej mierze przez komercję i to komercję bardzo dużych rozmiarów. Podjęcie takiej tematyki już stanowi w takich warunkach pewną odwagę. Ale taka jest rola działań akademickich. Bez wątpliwości, przydatność rozprawy dla nauk inżynierjno-technicznych jest znaczna.

## 9. Podsumowanie, do której z następujących kategorii recenzent zalicza rozprawę:

- a/ nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy
- b/ wymagająca wprowadzenia poprawek i ponownego recenzowania
- c/ spełniająca wymagania
- d/ spełniająca wymagania z wyraźnym nadmiarem**
- e/ wybitnie dobra, zasługująca na wyróżnienie

Praca jest zrealizowana bardzo poprawnie pod względem merytorycznym naukowo-technicznym i pod względem sposobu prezentacji wyników. Gdyby autorka była szerzej obecna w środowisku zawodowym z większym osobistym dorobkiem naukowym, to recenzent byłby skłonny wystąpić z wnioskiem o wyróżnienie, oczywiście po wysłuchaniu doktorantki w czasie obrony i po sposobie przebiegu dyskusji nad rozprawą. Podsumowując, recenzent wnioskuje o dopuszczenie doktorantki do dalszych etapów procesu doktryzowania.

/R.Romaniuk/

