

dr hab. Marek Pluta, prof. AGH
Akademia Górniczo-Hutnicza w Krakowie
Wydział Inżynierii Mechanicznej i Robotyki
Katedra Mechaniki i Wibroakustyki
al. Mickiewicza 30, 30-059 Kraków
tel. 12-617-34-16
e-mail: pluta@agh.edu.pl

Kraków, 5.01.2025

**Recenzja rozprawy doktorskiej
mgr inż. Mariusza Klecia**

pod tytułem

Efektywna reprezentacja danych w systemach przetwarzania sygnałów dźwiękowych
przygotowanej pod opieką naukową promotor dr hab. Alicji Wieczorkowskiej, prof. PJATK, oraz
promotora pomocniczego dr hab. inż. Krzysztofa Szklanego, prof. PJATK, na Wydziale
Informatyki Polsko-Japońskiej Akademii Technik Komputerowych w Warszawie
wykonana na podstawie decyzji Rady Naukowej Dyscypliny Informatyka Techniczna
i Telekomunikacja Polsko-Japońskiej Akademii Technik Komputerowych w Warszawie
(pismo Przewodniczącej Rady Naukowej Dyscypliny prof. dr hab. Elżbiety Orłowskiej)

Informacje o rozprawie

Recenzowana rozprawa doktorska mgr inż. Mariusza Klecia pod tytułem „Efektywna reprezentacja danych w systemach przetwarzania sygnałów dźwiękowych” została wykonana na Wydziale Informatyki Polsko-Japońskiej Akademii Technik Komputerowych w Warszawie pod opieką naukową głównej promotor dr hab. Alicji Wieczorkowskiej, prof. PJATK, oraz promotora pomocniczego dr hab. inż. Krzysztofa Szklanego, prof. PJATK.

Rozprawa ma formę spójnego tematycznie cyklu pięciu publikacji:

1. Kleć, M., Szklanny, K. & Wieczorkowska, A. (2024). **Developing a Corpus for Polish Speech Enhancement by Reducing Noise, Reverberation, and Disruptions**. In B. Marcinkowski, A. Przybyłek, A. Jarzębowski, N. Iivari, E. Insfran, M. Lang, H. Linger, & C. Schneider (Eds.), *Harnessing Opportunities: Reshaping ISD in the post-COVID-19 and Generative AI Era (ISD2024 Proceedings)*. Gdańsk, Poland: University of Gdańsk. ISBN: 978-83-972632-0-8,
2. Kleć, M. & Korzinek, D. (2015). **Pre-trained deep neural network using sparse autoencoders and scattering wavelet transform for musical genre recognition**. *Computer Science*, 16 (2), 133–144,
3. Kleć, M. (2018). **Early Detection of Heart Symptoms with Convolutional Neural Network and Scattering Wavelet Transformation**. In: Ceci, M., Japkowicz, N., Liu, J., Papadopoulos, G., Raś, Z. (eds) *Foundations of Intelligent Systems. ISMIS 2018. Lecture Notes in Computer Science*, vol. 11177. Springer, Cham,
4. Kleć, M., Wieczorkowska, A., Szklanny, K. & Strus, W.: **Beyond the Big Five personality traits for music recommendation systems**. *J. Audio Speech Music Proc.* 2023, 4 (2023),
5. Kleć, M., Wieczorkowska, A. (2021). **Music Recommendation Systems: A Survey**. In: Ras, Z.W., Wieczorkowska, A., Tsumoto, S. (eds) *Recommender Systems for Medicine and Music. Studies in Computational Intelligence*, vol. 946. Springer, Cham,

obejmujących dwa artykuły w materiałach konferencyjnych (pozycje 1 i 3), dwa artykuły w czasopiśmie naukowym znajdujących się w wykazie czasopism punktowanych Ministra Nauki z dn. 5.01.2024 przypisanych do dyscypliny Informatyka Techniczna i Telekomunikacja (pozycje 2 i 4) oraz rozdział monografii (pozycja 5). Mgr inż. Mariusz Kleć jest pierwszym (pozycje 1, 2, 4 i 5) bądź samodzielnym (pozycja 3) autorem wszystkich wymienionych publikacji.

Praca obejmuje 94 strony i można ją podzielić na dwie części. Zasadniczą część pracy stanowi ostatni, szósty rozdział, zawierający wymienione wyżej publikacje, zamieszczone w całości, w oryginalnej formie. Nie są one uporządkowane chronologicznie, lecz tematycznie. Pozostałe rozdziały stanowią część wstępną, będącą komentarzem do zamieszczonych publikacji. Ich rolą jest wykazanie spójności tematycznej załączonego cyklu pięciu publikacji oraz przedstawienie hipotez badawczych. Na część wstępną składa się pięć pierwszych rozdziałów rozprawy obejmujących „Wprowadzenie” (2 strony), „Przegląd literatury” (5 stron), „Przegląd opublikowanych artykułów” (8 stron), „Podsumowanie” (2 strony) i „Bibliografię” (4 strony, 46 pozycji). Każdy z zamieszczonych w rozdziale 6 artykułów zawiera własną bibliografię, w której można znaleźć część z pozycji wymienionych w rozdziale 5. Całość materiału graficznego, obejmującego 15 rysunków, zamieszczona jest wyłącznie w publikacjach składających się na rozdział 6. Spośród 10 zamieszczonych w pracy tabel 9 znajduje się, podobnie jak rysunki, w publikacjach zamieszczonych w rozdziale 6, natomiast jedna znajduje się w rozdziale 3.

Rozdział 1 wprowadza w problematykę rozprawy. Jest nią wpływ doboru i reprezentacji szeroko rozumianych danych dźwiękowych na efekt uzyskany dzięki ich przetwarzaniu w systemach klasyfikacji i rekomendacji. Przedstawiona jest motywacja do podjęcia przedmiotowej tematyki. Wskazana jest konieczność poprawy jakości narzędzi przetwarzających duże ilości danych dźwiękowych poprzez przyjęcie odpowiedniej ich reprezentacji. Celem tej poprawy ma być doprowadzenie do odkrywania nowych cech przetwarzanych danych, a w konsekwencji zwiększenie możliwości dyskryminacyjnych i szeroko rozumianej skuteczności operujących na nich narzędzi. Wymienione są w tym kontekście systemy przetwarzania i transkrypcji mowy, narzędzia wspierające diagnostykę medyczną, a także systemy rekomendacji muzycznej. Na końcu rozdziału przedstawione są trzy tezy, odnoszące się do usprawnień we wspomnianych trzech grupach systemów.

W rozdziale 2 omówiona jest aktualna literatura, podzielona na osiem obszarów. Są to w przeważającej części wyniki najnowszych prac, bo blisko 60% (27 z 46) spośród cytowanych pozycji pochodzi z lat 2020–2024. Punktem wyjścia jest stwierdzenie, że do celów klasyfikacji wykorzystuje się zwykle dwuwymiarową reprezentację sygnału akustycznego uzyskaną na drodze krótkookresowej transformacji Fouriera, albo transformacji falkowej oryginalnego sygnału. Przytaczana literatura dotyczy przede wszystkim przetwarzania dźwięku przez sieci neuronowe. Wymienione są prace dotyczące ekstrakcji cech dźwięku za pomocą spłotowych sieci neuronowych (rozdział 2.1) oraz badania nad zastosowaniem sieci rekurencyjnych (rozdział 2.2) i transformerów (rozdział 2.3) do analizy zależności czasowych. Wspomniane zostają również prace poruszające problematykę nie nadzorowanej ekstrakcji cech dźwięku przy pomocy autoenkoderów (rozdział 2.4). W kontraście do wymienionych podejść zaprezentowane są również próby automatycznej parametryzacji oryginalnego sygnału, bez jego wstępnego przekształcenia do postaci dwuwymiarowej (rozdział 2.5). Kolejne prace koncentrują się już na problematyce przetwarzania i klasyfikacji w ściśle określonych zastosowaniach, czyli w rekomendacji muzycznej (rozdział 2.6) oraz w analizie nagrań fizjologicznych (rozdział 2.7). Na końcu omówione są prace dotyczące dostosowywania zbiorów danych do określonych zadań klasyfikacyjnych (rozdział 2.8).

Rozdział 3 jest centralnym rozdziałem części wstępnej, omawiającym i komentującym cykl włączonych do rozprawy publikacji. Rozpoczyna się spisem danych bibliograficznych omawianych prac, po którym następuje syntetyczna prezentacja każdej z nich. Publikacje omówione są w kontekście głównego tematu rozprawy, co ma wykazać ich spójność tematyczną. Dlatego, jak wspomniano

wcześniej, ich kolejność nie jest chronologiczna, lecz odzwierciedla logikę przyjętego podejścia do rozwiązania postawionego w temacie rozprawy problemu.

Pierwsza praca (rozdział 3.1) dotyczy utworzenia korpusu dla systemów poprawiających zrozumiałość mowy w języku polskim. Zaproponowane jest w niej generowanie korpusu poprzez syntezę dźwięku z szeregu warstw, obejmujących naturalne nagrania mowy, warstwę tła, dźwięki dodatkowych zdarzeń nie będących mową, a także sygnał pogłosowy nałożony na mowę i dźwięki zdarzeń techniką splotu. Zaproponowane podejście zostało zbadane w połączeniu z kilkoma modelami separacji mówców opartymi na sieciach splotowych i transformerach. W tym miejscu omówione są także wyniki rozszerzonych badań, które nie znalazły się w zamieszczonej publikacji. Druga praca (rozdział 3.2) przedstawia metodę rozpoznawania gatunków muzycznych za pomocą sieci neuronowej, w której wagi są inicjalizowane wagami autoenkoderów wstępnie wytrenowanych z wykorzystaniem danych dźwiękowych pozyskanych z nagrań poddanych działaniu rozproszonej transformacji falkowej. Rozwinięcie tego tematu, a więc podobną reprezentację danych, zastosowaną w połączeniu z siecią splotową do wykrywania symptomów choroby na podstawie nagrań bicia serca, omawia praca trzecia (rozdział 3.3). Czwarta z omawianych publikacji (rozdział 3.4) przedstawia wyniki prac nad hybrydowym systemem rekomendacji muzycznej. System ten działa w oparciu o zbiór informacji dotyczących cech osobowości słuchacza oraz o wysokopoziomowe cechy dźwięku pozyskane na drodze analiz nagrań przeprowadzonych narzędziami MIR (ang. *Music Information Retrieval*). Ostatnia z publikacji (rozdział 3.5) jest przeglądem systemów rekomendacji muzycznej, uwzględniającym podejścia klasyczne, oparte na statystyce, oraz nowe, wykorzystujące głębokie sieci neuronowe.

Podsumowanie rozprawy, zamieszczone w rozdziale 4, prezentuje wnioski ogólne odnoszące się do kwestii udowodnienia postawionych w pracy hipotez badawczych. Zawiera również komentarz dotyczący możliwości wykorzystania uzyskanych wyników oraz upublicznionych zasobów do dalszych badań, które przyczynią się do rozwoju omówionych w pracy metod, narzędzi i systemów.

Ocena rozprawy

Praca jest poświęcona próbom wskazania takich sposobów doboru, reprezentacji, bądź przygotowania danych dźwiękowych, które poprawią efektywność ich przetwarzania w systemach operujących na dużych zbiorach danych. Są to systemy przetwarzania, klasyfikacji i rekomendacji operujące na sygnałach mowy i muzyki oraz na sygnałach o charakterze medycznym, np. na nagraniach bicia serca. W zaprezentowanym ujęciu przetwarzanie odnosi się przede wszystkim do różnego rodzaju sieci neuronowych, może jednak również dotyczyć metod statystycznych, jak w niektórych systemach rekomendacji muzycznej.

Poruszone w pracy zagadnienie jest aktualne i ważne pod kilkoma względami. Po pierwsze, przetwarzanie dużych ilości danych w wymienionych obszarach wiąże się z długim czasem pracy jednostek obliczeniowych. Każde usprawnienie tego procesu może przynieść wymierną oszczędność nie tylko czasu, ale i energii zużytej podczas obliczeń, co może być szczególnie widoczne w przypadku zastosowania głębokich sieci. Po drugie, stosowanie tych samych kombinacji metod przetwarzania i sposobów reprezentacji danych prowadzi do stagnacji nawet pomimo pozyskiwania nowych danych wejściowych. Dalsza poprawa skuteczności wymaga zastosowania nowej metody przetwarzania lub nowego podejścia do danych. Po trzecie, skutkiem poszukiwania lepszej reprezentacji danych może być umożliwienie zastosowania nowych lub zmodyfikowanych metod, co w konsekwencji pozwoli na uzyskanie nowych efektów.

Pewne zastrzeżenia można mieć do sformułowania tematu rozprawy. Pojęcie przetwarzania sygnałów dźwiękowych wydaje się z jednej strony zbyt ogólne, a z drugiej strony nie obejmuje wszystkich omawianych w pracy sytuacji. Jest zbyt ogólne z uwagi na to, że w tych obszarach gdzie praca dotyczy przetwarzania sygnałów dźwiękowych mowa jest o przetwarzaniu za pomocą

technik uczenia maszynowego, przy wykorzystaniu transformacji Fouriera i transformacji falkowej. Jest to jedynie niewielki wycinek spośród dużego zbioru technik przetwarzania dźwięku. Z drugiej strony, publikacja nr 4 (pt. *Beyond the Big Five Personality Traits for Music Recommendation Systems*) omawia działania na danych dotyczących osobowości słuchaczy i wiąże je z preferencjami uczestników testów. Dane te są zdecydowanie nadrzędne w stosunku do danych pochodzących z przetwarzania samego dźwięku, uzyskanych z wykorzystaniem narzędzi MIR.

Drugie zastrzeżenie można mieć do pojęcia efektywności. Nie jest ono w tekście pracy zdefiniowane, czy nawet szerzej omówione. Z tytułu pracy nie wynika, czy dotyczy ono szybkości przetwarzania danych, ich objętości, czy też możliwości uzyskania lepszego wyniku, a w tym ostatnim wypadku, jaka byłaby miara jakości tego wyniku. Skoro nie są znane kryteria obliczania efektywności, trudno jest ocenić czy rzeczywiście i jakim stopniu efektywność uległa poprawie na skutek zastosowania omawianych w pracy sposobów reprezentacji danych. Zapoznając się z poszczególnymi publikacjami wchodzącymi w skład rozprawy można dojść do wniosku, że w każdej z nich pojęcie to dotyczy czegoś innego.

O ile niekiedy trudno jest znaleźć trafne sformułowanie, jeżeli nie jest to możliwe, wskazane byłoby zamieścić w treści odpowiednie wyjaśnienia. Pewną wskazówką mogą w tym zakresie stanowić trzy postawione w rozprawie hipotezy badawcze.

1. „Włączenie zdarzeń dźwiękowych do korpusów mowy, wygenerowanych razem z pogłosem i dźwiękami tła, poprawia odporność modeli przetwarzających mowę na występowanie tych zakłóceń.”
2. „Wykorzystanie różnej długości ramek czasowych rozproszonej transformaty falkowej oraz splotowych sieci neuronowych przyczynia się do poprawy rozpoznawania wczesnych objawów chorób serca.”
3. „Wykorzystanie modelu stanowiącego rozszerzenie standardowego modelu tzw. Wielkiej Piątki do reprezentacji osobowości użytkownika przyczynia się do redukcji błędów popełnionych przez systemy rekomendacji muzycznej, w porównaniu z modelem standardowym.”

Tak sformułowane tezy obejmują zakres zagadnień całego cyklu publikacji zamieszczonych w rozprawie, ukazując je jako elementy większej całości i spajając tematycznie.

W publikacji pt. *Developing a Corpus for Polish Speech Enhancement by Reducing Noise, Reverberation, and Disruptions* (rozdział 6.1) zaproponowane jest rozwiązanie problemu niedostatecznej ilości danych dźwiękowych do treningu narzędzi poprawiających zrozumiałość mowy w języku polskim. Rozwiązanie, które można zaklasyfikować jako technikę syntezy dźwięku, jest oparte na miksie siedmiu warstw. Cztery z nich są fragmentami naturalnych nagrań pochodzących z istniejących baz: dwóch mówców, dźwięków zdarzeń zakłócających mowę oraz dźwięku otoczenia. Pozostałe trzy warstwy są tworzone przez sygnał pogłosu wygenerowany z nagrań mowy oraz zdarzeń zakłócających w oparciu o odpowiedzi impulsowe. Pogłos ma wspólne dla wszystkich trzech warstw losowe parametry, mieszczące się w zakresie typowym dla danego rodzaju dźwięku otoczenia. Oprócz miksu korpus zawiera zapis poszczególnych warstw, co umożliwia utworzenie sygnału składającego się z dowolnych spośród dostępnych elementów. Dzięki możliwości wyłączania poszczególnych warstw oraz przez zastosowanie różnych parametrów pogłosu możliwe jest generowanie praktycznie nieograniczonej liczby różniących się od siebie przykładów na bazie ograniczonego zbioru nagrań składowych. Co godne podkreślenia, skrypty służące do generowania korpusu zostały udostępnione w formie kodu źródłowego dla środowiska Matlab w otwartym repozytorium sieciowym. Praca przedstawia wyniki badań, w których wygenerowany korpus wykorzystano do treningu narzędzi poprawy zrozumiałości mowy opartych na kilku różnych modelach. Wyniki te potwierdzają pierwszą z postawionych w rozprawie hipotez badawczych.

Druga z publikacji, a pierwsza chronologicznie, pt. *Pre-trained deep neural network using sparse autoencoders and scattering wavelet transform for musical genre recognition* (rozdział 6.2) prezentuje technikę rozpoznawania gatunków muzycznych opartą na głębokiej sieci neuronowej. Sieć jest inicjalizowana wagami uzyskanymi z autoenkoderów, podlegających wstępnemu treningowi. Do treningu wykorzystywane są nagrania przekształcone za pomocą rozproszonej transformacji falkowej. Badania mają wstępny charakter, a zaproponowane narzędzie wykorzystuje jedynie informację spektralną, nie uwzględniając struktury czasowej nagrań, istotnej dla klasyfikacji gatunków muzycznych, co jest podkreślone w podsumowaniu publikacji.

Trzecia publikacja, pt. *Early Detection of Heart Symptoms with Convolutional Neural Network and Scattering Wavelet Transformation* (rozdział 6.3) kontynuuje temat reprezentacji danych za pomocą rozproszonej transformacji falkowej. Tym razem jednak zadanie klasyfikacji realizuje sieć spłotowa, a metoda jest zastosowana do medycznych badań przesiewowych i służy analizie nagrań bicia serca celem wykrycia symptomów choroby. Uzyskane wyniki potwierdzają drugą z postawionych w rozprawie hipotez badawczych.

O ile nie budzi wątpliwości włączenie w cykl publikacji trzech pierwszych prac, o tyle dwie ostatnie są do pewnego stopnia odrębne, a powiązanie z poprzednimi ma mniej wyraźny charakter. Czwarta z cyklu publikacji pt. *Beyond the Big Five personality traits for music recommendation systems* (rozdział 6.4) omawia wyniki zakrojonych na szeroką skalę (279 uczestników, 745 utworów) badań nad możliwościami systemu rekomendacji muzycznej opartego na hybrydowym podejściu, w którym brane są pod uwagę szczegółowe dane charakteryzujące osobowość słuchacza oraz cechy wysokiego poziomu pozyskane na drodze analizy nagrań utworów narzędziami MIR. Dużą wartość ma zawężenie zbioru przebadanych aspektów osobowości do czterech, które pozwalają dokonać rekomendacji z najmniejszym błędem, a tym samym opracować uproszczone kwestionariusze dla użytkowników, po wypełnieniu których możliwe jest uzyskanie wiarygodnych rekomendacji muzycznych. Należy podkreślić, że otrzymane dane eksperymentalne zostały udostępnione w publicznym repozytorium i mogą stanowić podstawę dalszych analiz, również w innych zespołach badawczych. Otrzymane rezultaty potwierdzają ostatnią z postawionych w rozprawie hipotez badawczych.

Piąta i zarazem ostatnia publikacja pt. *Music Recommendation Systems: A Survey* (rozdział 6.5) również dotyczy systemów rekomendacji muzycznej. Ma ona jednak charakter przeglądowy, podsumowujący, nie zawiera natomiast nowych, własnych wyników, przez co nie wnosi dodatkowej wartości merytorycznej do recenzowanej rozprawy.

W pracy brakuje nieco głębszego i zarazem syntetycznego podsumowania całości osiągnięcia, bez rozbicia na poszczególne publikacje, wykraczającego poza komentarze dotyczące udowodnienia postawionych tez. Brakuje również całościowego spojrzenia na ograniczenia osiągniętych efektów oraz wskazania w jednolitym ujęciu najważniejszych kierunków dalszych badań. Elementy te wstępują wyłącznie w odniesieniu do poszczególnych publikacji.

Pozytywnie natomiast należy ocenić rozdziały 2 i 3, które relatywnie obszernie omawiają całość prezentowanego w rozprawie zagadnienia. Uzupełniają one również informacje które nie znalazły się w zamieszczonych publikacjach.

Podsumowując ogólną ocenę rozprawy należy zauważyć, że postawione w pracy tezy zostały udowodnione, co zostało wyraźnie podkreślone w rozdziale 4. Wykazano, że zbudowanie korpusu w który włączone są elementy utrudniające rozpoznanie mowy, takie jak dźwięki otoczenia, dźwięki dodatkowych zdarzeń oraz sygnał pogłosowy, poprawia odporność modeli przetwarzających mowę na tego typu zakłócenia. Wykazano również, że można uzyskać lepsze wyniki rozpoznawania wczesnych objawów chorób serca, jeżeli dźwięk reprezentowany przez różnej długości ramki czasowe rozproszonej transformaty falkowej jest analizowany przez spłotową sieć neuronową. W końcu wykazano, że można uzyskać bardziej wiarygodne wskazania systemu rekomendacji muzycznej dzięki zastosowaniu rozszerzonego modelu opisującego osobowość słuchacza. Tezy te wiążą przedstawione publikacje w spójny tematycznie cykl, mieszczący się w zakresie tematycznym dyscypliny Informa-

tyka Techniczna i Telekomunikacja. Tym samym rozprawa spełnia wymagania stawiane rozprawom doktorskim.

Komentarze

Proszę Doktoranta o odniesienie się do uwag zamieszczonych poniżej.

1. W pracy nie wyszczególniono wkładu Doktoranta w powstanie publikacji nr 1, 2, 4 i 5. Co prawda Doktorant jest ich pierwszym autorem, a dołączone *curriculum vitae* pozwala z wysokim prawdopodobieństwem przewidzieć za które elementy mógł być odpowiedzialny, miałbym jednak prośbę o napisanie, w odniesieniu do każdej z tych czterech publikacji, kilku zdań ze wskazaniem własnego wkładu w ich powstanie.
2. W artykule nr 1 (*Developing a Corpus for Polish Speech Enhancement by Reducing Noise, Reverberation, and Disruptions*) podany jest odnośnik do kodu źródłowego, co pozwala dotrzeć do tej informacji, proszę jednak o jej podanie wprost: w jaki sposób ustalane są relatywne poziomy sygnału poszczególnych warstw i jakie są między nimi różnice. Czy te różnice można kontrolować?
3. W odniesieniu do artykułu nr 2 (*Pre-trained deep neural network using sparse autoencoders and scattering wavelet transform for musical genre recognition*): dlaczego zdecydowano się zastosować sieci które nie uwzględniały struktury czasowej nagrań? Przecież struktura ta zawiera bardzo wiele informacji różnicujących gatunki. Czy były późniejsze próby wykorzystania tej informacji? A jeżeli nie, to czy na podstawie znanych Autorowi prac można oszacować poprawę skuteczności klasyfikacji gatunku dzięki wykorzystaniu tej informacji?
4. W artykule nr 3 (*Early Detection of Heart Symptoms with Convolutional Neural Network and Scattering Wavelet Transformation*) brakuje informacji dotyczących analizowanego sygnału. Jakie były jego parametry i cechy charakterystyczne?
5. W tym samym artykule nr 3: jakie były kryteria doboru struktury sieci? Czy tę strukturę wyznaczono na podstawie eksperymentów (i czy były to eksperymenty systematyczne, czy jedynie wyrwykowe), czy jakichś wcześniejszych założeń?
6. Z opisu przedstawionego w artykule nr 4 (*Beyond the Big Five personality traits for music recommendation systems*) wynika, że wysokopoziomowe cechy dźwięku pozyskiwano uśredniając je dla całych utworów. Czy to jest właściwe podejście, skoro charakter wielu utworów ulega w ich trakcie dużym przemianom? W jakich gatunkach muzycznych lub w jakiego rodzaju utworach może to prowadzić do nieprawidłowej analizy? Z jakimi niskopoziomowymi parametrami będzie to związane? Czy ten problem był badany, a może uniknięto go w inny sposób?

Uwagi redakcyjne

Rozdział 6, obejmujący największą część rozprawy, zawiera opublikowane wcześniej prace zamieszczone w oryginalnej formie. Z tej przyczyny uwagi redakcyjne mają sens jedynie w odniesieniu do rozdziałów 1–5.

Od strony redakcyjnej praca jest pozbawiona większych usterek. Można zauważyć nieliczne błędy literowe oraz pewną niestaranność w danych bibliograficznych zamieszczonych na str. 9, gdzie część danych dotyczących publikacji nr 2 została przypisana do publikacji nr 3, widoczna jest również niekonsekwencja w stosowaniu przecinka oddzielającego autorów prac oraz kończącego opis danej pozycji. W rozdziale 2 (str. 3) mowa jest o krótkookresowej transformacji Fouriera (ang.

Short-Term Fourier Transform), której nazwa została omyłkowo przetłumaczona jako „dyskretna transformata Fouriera”. Są to wyłącznie drobne niedociągnięcia.

Wnioski końcowe

Oceniana rozprawa doktorska przedstawia nowe wyniki badań naukowych, mające wysoką wartość teoretyczną i praktyczną. Rozprawa wskazuje, że Doktorant posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka Techniczna i Telekomunikacja, poprzez umiętność budowania i badania złożonych systemów oraz interpretacji skomplikowanych wyników. Nieliczne uwagi krytyczne nie wpływają na wysoką ocenę merytoryczną przedłożonej do recenzji dysertacji.

Na podstawie przedstawionej mi do recenzji rozprawy doktorskiej Pana mgr inż. Mariusza Klecia pod tytułem „Efektywna reprezentacja danych w systemach przetwarzania sygnałów dźwiękowych” stwierdzam, że spełnia ona wymagania ustawy z dnia 20 lipca 2018 roku, Prawo o szkolnictwie wyższym i nauce, Dz. U. 2023, poz. 742.

Po wnikliwym zapoznaniu się z pracą stawiam wniosek o przyjęcie przedstawionej rozprawy doktorskiej oraz o dopuszczenie jej do publicznej obrony.

