



**Politechnika Łódźka**

Instytut Informatyki

PW WETI Kancelaria  
wpłynęło dnia 10.02.2022  
numer .....

Łódź, 24 stycznia 2022 roku

dr hab. inż. Adam Wojciechowski, prof. uczelni  
Instytut Informatyki  
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej  
Politechnika Łódźka  
Al. Politechniki 8, 93-590 Łódź

### RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: **Artificial Intelligence Solutions for Artistic Multimedia Musical Content Creation Support**

Autor rozprawy: mgr inż. **Mateusz Modrzejewski**

Promotor rozprawy: prof. dr hab. inż. **Przemysław Rokita**

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

Pan mgr inż. Mateusz Modrzejewski w ramach rozprawy doktorskiej przedstawił wyniki badań związanych z zastosowaniami sztucznej inteligencji w procesie wspierania wytwarzania artystycznych treści muzycznych oraz analizy ich treści. Badania miały na celu wypełnienie wybranych luk w rozważanych obszarach tworzenia i analizy utworów muzycznych, ze szczególnym uwzględnieniem wybranych typów sieci neuronowych, w tym sieci głębokich, rekurencyjnych, czy modeli generatywnych. W ramach pracy podjęto trzy szczegółowe problemy: przenoszenie stylu brzmieniowego (gitary i pianina) pomiędzy utworami muzycznymi, autonomiczne komponowanie treści utworów z nastawieniem na tworzenie, nowych oryginalnych brzmień oraz klasyfikacja gatunków utworów muzycznych. W poszukiwaniu efektywnych rozwiązań Doktorant sięgał do różnych form reprezentacji treści muzycznych, w tym różnych reprezentacji graficznych i reprezentacji MIDI oraz eksperymentował zarówno z modelami generatywnymi (wytwarzanie nowych fraz muzycznych, transfer stylu muzycznego) jak i splotowymi modelami rekurencyjnymi.

Problemy badawcze, postawione w rozprawie doktorskiej, zostały sformułowane w sposób jasny i zrozumiały. Bardzo pomocne w trafnym i przepełnionym użyciu sformułowaniu pytań badawczych było doświadczenie muzyczne Doktoranta, który może poszczycić się bogatym zbiorem osiągnięć na polu artystycznym (dodatek B). Tym samym, chociaż rozprawa nosi znamiona pracy interdyscyplinarnej, to sposób ujęcia problemów i zaproponowanych rozwiązań czyni ją pełnoprawnym osiągnięciem o charakterze doświadczalnym w Dziedzinie nauk inżynierijno-technicznych, w Dyscyplinie Informatyka Techniczna i Telekomunikacja.

**Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej i stanu zagadnień w przemyśle) świadczącej o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?**

Rozprawa doktorska została podzielona na trzy odrębne zagadnienia, które mieszczą się w obszarze wytwarzania i analizowania treści muzycznych. Każdemu z zagadnień poświęcono odrębny rozdział (sekcje nr 3, 4 i 5), ale analizę źródeł w ramach poszczególnych zagadnień poprzedza osobna sekcja (nr 2) omawiająca fundamentalne pojęcia i zagadnienia muzyczne, które są niezbędne do zrozumienia kontekstu muzycznego poruszanej problematyki.

Zakres i forma przeglądu potwierdza, że przegląd źródeł literaturowych został dokonany w sposób właściwy i obszerny. Niektóre aspekty charakterystyk utworów muzycznych mogłyby wprawdzie zostać uzupełnione o wybrane miary techniczne (inżynierskie) służące do monitorowania rytmu, melodii, harmonii czy struktury utworów muzycznych, które pomogłyby czytelnikowi o mniejszym doświadczeniu muzycznym w ilościowej ocenie cech utworów. Z drugiej jednak strony aspekty te można z powodzeniem znaleźć w literaturze dziedzinowej.

Przegląd literatury dotyczący trzech odrębnych zagadnień: przenoszenia stylu brzmieniowego pomiędzy utworami muzycznymi, generatywnego komponowania treści muzycznych oraz klasyfikacji gatunków utworów muzycznych, obejmuje trzy dedykowane i obszerne przeglądy literatury obejmujące poszczególne obszary. Nie pozostawia on wątpliwości, że wiedza Doktoranta w rzeczonych obszarach jest obszerna i wyczerpująca. Przegląd przytacza wiele rozwiązań referencyjnych i dyskutuje szczególnie wybranych metod literaturowych. Przegląd ma charakter głównie jakościowy jednak warto w tym miejscu zauważyć, że cennym uzupełnieniem przeglądu byłaby również krytyczna analiza ilościowa rozwiązań referencyjnych, która dostarczyłaby wyraźnego tła dla oczekiwanej efektywności rozwiązań zaproponowanych przez Doktoranta. Przykładowo takim obszarem jest wielkość modeli sieci neuronowych, których analiza była celem pracy Doktoranta, podczas gdy wielokrotnie pojawiające się kwantyfikatory: „mały”, „duży” w odniesieniu do modelu sieci mogłyby być z powodzeniem zastąpione ich złożonością pamięciową lub liczbą parametrów modeli.

Nie zmienia to faktu, że za każdym razem wnioski z przeglądu literatury sformułowano w sposób jasny i przekonujący. Wskazywały one na istotne luki w dotychczasowych rozwiązaniach literaturowych i dawały wyraźne tło dla poszukiwań i badań Autora rozprawy.

#### **Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?**

Tematyka rozprawy jest aktualna i ważna. Jest to szczególnie widoczne gdy śledzimy dynamiczny postęp badań w obszarze sztucznej inteligencji na potrzeby przetwarzania i analizy obrazów oraz tekstu. Analiza utworów muzycznych i przetwarzanie sygnałów dźwiękowych zdają się zatem być nie mniej ważnymi i relatywnie słabo rozpracowanymi zagadnieniami. Najczęściej rozwiązania opracowane dla obrazów, czy tekstu są bezrefleksyjnie przenoszone na obszar analizy dźwięków, ale brakuje rozwiązań, które uwzględniają specyfikę tej dziedziny badań i potrafią wykorzystać wiedzę dziedzinową w tworzeniu nowych, lepszych rozwiązań. Tę lukę wypełniają właśnie badania przedstawione w rozprawie doktorskiej Pana Mateusza Modrzejewskiego, który łącząc swoje doświadczenie artystyczne (muzyczne) z kompetencjami informatycznymi zaproponował dedykowane, nowe modele sztucznych sieci neuronowych i przeprowadził cenne badania, które miały istotny wkład w rozwój przedmiotowej dziedziny.

#### **Czy Autor rozwiązał postawione zagadnienie, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?**

W ogólności, wyuczenie modelu sieci neuronowej i jego zdolność do przenoszenia stylu pomiędzy utworami muzycznymi jest rozważane na trzech poziomach ogólności. Na niskim poziomie reprezentacji dźwięku możemy mówić o stylu barwowym i związanym z nim modulowaniem barwy dźwięku, przy określonym stylu wykonania utworu. Poziom średni charakterystyki stylu utworu muzycznego odnosi się do stylu wykonania utworu przy zachowaniu partytury utworu. Z kolei wysokopoziomową cechą stylu utworu jest styl kompozycyjny odnoszący się do reprezentacji partytury przy określonym konturze melodii i charakterystykach harmonii utworu.

W rozprawie Doktorant zaproponował oryginalną metodę transferowania stylu brzmieniowego utworu muzycznego pomiędzy wybranymi instrumentami (gitarą i pianinem), stosującą rekurencyjny

model sieci neuronowej typu autoenkoder. Deklarowanymi cechami metody jest zdolność do zachowania treści utworu i modyfikacji jego brzmienia (tembru) oraz wydajność metody przewyższająca rozwiązania literaturowe - przykładowo wydajność modelu WaveNet. Faktycznie spostrzeżenie Doktoranta, który zwrócił szczególną uwagę na możliwość uwzględnienia następstwa czasowego sekwencji dźwięków dzięki sieciom typu LSTM, jako mechanizm zapewniający odpowiedni transfer stylu brzmieniowego, daje metodzie niekwestionowaną wartość dodaną. Metoda została przetestowana na podzbiorach ogólnie dostępnych zbiorów sekwencji MIDI, co zapewnia odpowiednią otwartość i powtarzalność eksperymentu. Metodologia eksperymentu jest zrozumiała i prawidłowa. Otrzymane wyniki dowodzą, że z każdym nowym wariantem metody, cechuje się ona mniejszym błędem średniokwadratowym dopasowania utworu muzycznego względem wzorca i jednocześnie mniejszym odchyleniem standardowym.

Pomimo niekwestionowanych zalet zaproponowanej metody, niektóre kwestie nie są jednakże jasno sprecyzowane. O ile sposób przetwarzania wstępnego danych został przedstawiony zrozumiale to nie jest jednoznaczne jak wygląda dokładnie wektor wejściowy modelu i sam proces nauki. Schematy modelu pozwalają uchwycić istotne różnice ideowe pomiędzy kolejnymi wariantami, ale wykorzystane w nich symbole i indeksy nie zostały objaśnione. Być może dobrym rozwiązaniem byłoby pokazanie różnic pomiędzy kolejnymi wariantami metody na poziomie szczegółowości z rysunku 3.6? Pomimo odwołania się w treści rozprawy do kryteriów jakościowych takich jak szum, czy jakość dźwięku, praca mogłaby zawierać analizę wybranych miar pozwalającą czytelnikowi samodzielnie ocenić ilościowo i jakościowo różnice pomiędzy sygnałem wyjściowym modelu a sygnałem odniesienia. Swoją drogą jak miara błędu dopasowania MSE koreluje ze słyszalną jakością utworu muzycznego? Nie jest do końca jasne dlaczego nie porównano wyników jakościowych zaproponowanej przez Doktoranta metody z metodami referencyjnymi, skoro zostały one przytoczone i omówione w części przeglądowej?

Drugim obszarem badań Doktoranta było tworzenie treści muzycznych. Ten obszar badań jest obecnie bardzo modny i eksploatowany ze względu na popularność technik sztucznej inteligencji we wspomaganie pracy artystów, np. kompozytora. Literatura przedmiotu analizuje zagadnienie kompozycji utworów w różnych aspektach, w tym jako problem generowania sekwencji akordów, melodii, określonego gatunku muzyki, czy nawet muzyki polifonicznej lub monofonicznej. Autor wpasowuje się jednak w inny nurt badań, który wynika z niskiej oceny komponowanych przez sztuczną inteligencję utworów. Doszukuje się wartości w oryginalności komponowanych treści muzycznych, np. tworzenia nowych instrumentów muzycznych lub nowych brzmień z udziałem sztucznej inteligencji. To bardzo ciekawy i oryginalny nurt komputerowej kreatywności czy sztuki. W tym kontekście na uwagę zasługuje dodatek C, który uzupełnia dysertację o przegląd i dyskusję aspektów prawnych towarzyszących zagadnieniu komponowania muzyki.

Doktorant zaproponował autorski model sieci neuronowej oparty o głęboką architekturę splotową generatywnej sieci adwersaryjnej (GAN - ang. *Generative Adversarial Network*). Celem sieci było głównie skomponowanie utworów nowych, wcześniej nie słyszanych aniżeli komponowanie utworów przynależnych do konkretnego gatunku muzycznego. W tym celu Autor wykorzystał graficzną reprezentację utworu muzycznego, kodującą rytm, melodię i harmonię za pomocą pozycji oraz kolorów pikseli obrazu, przypominającą taśmę perforowaną do odtwarzania muzyki (tzw. *piano roll*). W tym celu utwór jest początkowo kwantyzowany (próbki 30 ms) i przycinany do 64 z pełnej, 88 klawiszowej klawiatury pianina. Model sieci został przedstawiony w postaci obrazowych schematów na pewnym poziomie ogólności – nie jest do końca jasne jak w obrazie o rozmiarze 64x64 pikseli (np. rys. 4.1) zapisano 20 sekund utworu.

Wyniki działania metody zostały przedstawione w sposób opisowy. Szkoda, że spostrzeżeniom Autora nie towarzyszą zestawienia ilościowe lub odpowiednie metryki potwierdzające przytoczone konkluzje.

Przykładowo niejednoznaczne jest stwierdzenie dotyczące wizualnego podobieństwa obrazów (ang. „*visual similarity*”, „*images ... are indistinguishable for the human eye*”), które przy dość abstrakcyjnych znaczeniowo reprezentacjach graficznych treści muzycznych nie są semantycznie jednoznaczne. Dla spostrzeżeń typu „*certain rhythmic patterns*”, czy „*chaotic structure*” przydałyby się obrazujące to przykłady. Przeprowadzona analiza harmoniczna polega na zliczaniu wybranych określeń muzycznych (akordy, kadencje, frazy, in.) dla nowo powstałego utworu i porównywaniu ich spójności z danymi wejściowymi (treningowymi). Chociaż stwierdzono podobieństwo w zakresie harmonii z przykładami treningowymi to nie zostało to jednak poparte zamieszczonymi w pracy wynikami – wyniki ilościowe przytoczone są jedynie dla 10 wybranych przykładów wyjściowych.

Analiza rytmu wyjściowych utworów wskazuje na problemy z jego odtworzeniem, przynajmniej dla niektórych zbiorów danych uczących. Jednocześnie niezbitą wartością skomponowanej muzyki jest obecność długich, różnorodnych i interesujących motywów muzycznych, których obecność jest trudna w obiektywnym zmierzeniu, ale średnio wprawny słuchacz może je usłyszeć analizując dostarczone przez Doktoranta utwory muzyczne. Właśnie w tym kontekście wartym odnotowania osiągnięciem jest przygotowany, z użyciem opracowanego modelu, album utworów muzycznych o nazwie DROP (<https://bit.ly/3G4rkXj>), który powstał w ścisłej współpracy ze znanym muzykiem, Panem Michałem Milczarkiem i stanowi bardzo ciekawe, interdyscyplinarne uzupełnienie badań Doktoranta.

Trzecim nurtem badań Doktoranta jest klasyfikacja gatunków utworów muzycznych, który wpisuje się w szerszy obszar pozyskiwania informacji z utworów muzycznych (ang. *music information retrieval*). Badania były przeprowadzane na różnych zbiorach danych, w tym na stosunkowo obszernym i współczesnym zbiorze danych FMA. Doktorant porównał trzy architektury sieci spłotowych – dwie podstawowe architektury sieci spłotowych (CNN – ang. *convolutional neural network*) o różnej liczbie warstw, odpowiednio 2 i 4, oraz równoległą rekurencyjną sieć spłotową (PRCNN – ang. *parallel recurrent convolutional neural network*) dla dwóch form reprezentacji graficznej: spektrogramów i chromogramów. Chociaż architektura PRCNN została już zaproponowana w 2017 roku (Feng et al., Choi et al.) to Doktorant zaproponował jej modyfikację polegającą na dodaniu jednej warstwy spłotowej, celem uchwycenia lokalnych charakterystyk utworów muzycznych.

Wszystkie trzy architektury zostały przetestowane na różnych zbiorach danych (FMA, GTZAN) dla dwóch różnych reprezentacji graficznych: spektrogramów i chromogramów. Metodologia badań była prawidłowa i została zobrazowana klasycznymi miarami stosowanymi w zadaniu klasyfikacji: precyzją, czułością i miarą F1. Otrzymane wyniki generalnie potwierdzają skuteczność sieci spłotowych w klasyfikacji gatunków utworów muzycznych na literaturowym poziomie. Nie zamieszczono w pracy wprawdzie porównania do rozwiązań referencyjnych, ale Doktorant nadmienia, że otrzymane wyniki plasują zaproponowane przez niego rozwiązanie w grupie czołowych metod klasyfikacji gatunków muzycznych.

Ciekawa dyskusja została przeprowadzona w kontekście macierzy konfuzji w zadaniu klasyfikacji poszczególnych gatunków muzycznych. Doktorant analizował błędy klasyfikacji w świetle cech charakterystycznych (struktury, harmonii, częstotliwości, linii basów i in.) gatunków utworów muzycznych. Bardzo ciekawym rozszerzeniem byłoby dokonanie sprzężenia zwrotnego, w którym poczynione spostrzeżenia zostałyby zamienione na konkretne modyfikacje modelu sieci neuronowej w zadaniu klasyfikacji gatunku muzycznego, co w rezultacie zweryfikowałoby zasadność poczynionych spostrzeżeń. Pomimo naturalnego niedosytu wynikającego z ciekawej i stymulującej dyskusji oraz wielu cennych sugestii, dyskusję oceniam bardzo wysoko i uważam, że poprawia ona zrozumienie procesu analizowania cech utworów muzycznych. Świadczy również o doskonałym zrozumieniu dziedziny badań.

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

Oryginalność rozprawy doktorskiej leży w opracowaniu nowych, oryginalnych modeli sieci neuronowych, które uwzględniają cechy utworów muzycznych i dzięki temu charakteryzują się lepszą efektywnością przetwarzania oraz analizy sygnałów niż rozwiązania bezrefleksyjnie zaczerpnięte z pokrewnych obszarów badań (analiza obrazów, tekstów). W szczególności najważniejszymi, moim zdaniem, osiągnięciami są:

- Nowa metoda transferowania stylu brzmieniowego utworów muzycznych. Pomimo przetestowania metody dla pianina i gitary, metoda nosi znamiona generalizacji i powinna dać się adaptować dla innych instrumentów muzycznych;
- Nowa generatywna metoda tworzenia treści muzycznych, która redukuje ograniczenia wcześniejszych rozwiązań i dostarcza konstruktywnych inspiracji kompozytorom oraz twórcom muzyki. Metoda wpisuje się w bardzo ciekawy wątek autonomicznej twórczości artystycznej, która jest zasługą wyłącznie algorytmów komputerowych lub dostarcza twórcom muzyki nowych wzorców i inspiracji;
- Analiza problemu klasyfikacji gatunków muzycznych w świetle własności modeli sztucznych sieci neuronowych, formatów reprezentacji graficznej utworów muzycznych i cech szczególnych samych gatunków muzycznych;

Wymienione osiągnięcia stanowią bezsprzecznie oryginalny wkład Doktoranta w rozwój Dyscypliny Informatyka Techniczna i Telekomunikacja. Wartością dodaną są również interdyscyplinarne rozważania i wyniki prac, w tym mini album muzyki skomponowanej z użyciem autorskich algorytmów, czy dyskusja na temat aspektów prawnych generatywnie komponowanej muzyki.

**Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięźłość, jasność, poprawność redakcyjna rozprawy)?**

Wszystkie przeprowadzone przez Doktoranta badania zostały zrealizowane zgodnie z przyjętą w dziedzinie metodologią badawczą, wyniki zostały opracowane w syntetyczny i klarowny sposób. Własności zaproponowanych rozwiązań, choć na pewnym poziomie ogólności, zostały przedstawione również w sposób czytelny i zrozumiały. Badania zostały wkomponowane w strukturę rozprawy doktorskiej, czyniąc ją systematycznym i logicznie opracowanym dokumentem.

Praca została zredagowana bardzo starannie, zawiera odpowiednie spisy, odwołania i odsyłacze do literatury. Chociaż nie udało się uniknąć nielicznych błędów edytorskich i reakcyjnych, np. „and and” (str. 8), „in a ... languages” (str. 9), „for of all” (str. 26), „for for” (str. 62) to nie obniża to mojej wysokiej oceny pracy pod kątem redakcyjnym.

Wartym podkreślenia jest fakt publikacji treści muzycznych w otwartej formule (linki do źródeł internetowych), ale również w postaci usystematyzowanego albumu DROP skomponowanego przez opracowane algorytmy i jedynie wyselekcjonowanego przez artystę muzyka. Wartym odnotowania elementem rozprawy są dodatki, a szczególnie dodatek C poświęcony aspektom prawnym komponowania treści muzycznych przez algorytmy. Stanowi to bardzo cenny głos uzupełniający formalny kontekst podjętej problematyki badawczej.

Potwierdzeniem bardzo dobrej umiejętności poprawnego i przekonującego przedstawiania uzyskanych przez siebie wyników jest bogata lista publikacji Doktoranta. Jest to aż 5 publikacji w renomowanych międzynarodowych konferencjach z wykazu MNiSW oraz 1 rozdział w monografii

wydawnictwa Springer (Dodatek A). Tym samym Doktorant zawiązką wywiązał się z działalności publikacyjnej.

### **Jakie są słabe strony rozprawy i jej główne wady?**

Rozprawa doktorska nie zawiera żadnych istotnych uchybień. Jedyne uwagi natury dyskusyjnej dotyczą łatwości z jaką Doktorant prześlizguje się nad niektórymi stwierdzeniami, podczas gdy dla czytelnika rozprawy, słabiej zaznajomionego z charakterystyką treści muzycznych, nie są one zawsze oczywiste. Przykładowe kwestie, które mogłyby zostać rozszerzone, lecz nie ujmują wysokiej oceny rozprawy, dotyczą:

- dokładnych rozmiarów wektora cech przestrzeni wejściowej i wyjściowej oraz sposobu uczenia modelu autoenkodera w problemie transferu stylu brzmieniowego;
- ilościowego porównania rozmiarów (złożoność pamięciowa lub liczba parametrów definiująca wielkość modelu) dyskutowanych modeli sztucznych sieci neuronowych, skoro Autor traktuje swoje rozwiązanie jako nisko zasobo-chłonne (ang. *lightweight solution*);
- ilościowych podstaw analizy jakościowej spójności rytmu, harmonii, melodii czy struktury utworów dla generatywnego komponowania treści muzycznych, które zostały stwierdzone na podstawie autorskiej analizy, ale nie zostały jawnie poparte wynikami ilościowymi;
- porównania autorskich wyników z metodami referencyjnymi, które w każdym z rozważanych obszarów tworzenia i analizy treści muzycznych zostały przytoczone, ale nie zostały zestawione w porównaniach ilościowych/jakościowych;

Szczegółowe spostrzeżenia dotyczące aspektów dyskusyjnych zostały dokładniej opisane w części recenzji poświęconej opisowi poszczególnych rozwiązań.

### **Konkluzja**

Uważam, że cele rozprawy doktorskiej zostały w pełni zrealizowane. Autor przedstawił w rozprawie nowe i oryginalne metody tworzenia oraz analizowania treści muzycznych. Efektywność i użyteczność zaproponowanych rozwiązań wykazał w prawidłowo zaprojektowanych i przeprowadzonych eksperymentach badawczych, z użyciem odpowiednio dobranych zbiorów danych oraz zgodnie z przyjętą w literaturze metodologią badawczą. Uzyskane przez Doktoranta wyniki uważam za oryginalne, wartościowe i ciekawe poznawczo. Tym samym rozprawa prezentuje ogólną wiedzę teoretyczną w wybranym obszarze Dyscypliny Informatyka Techniczna i Telekomunikacja oraz potwierdza umiejętność samodzielnego prowadzenia przez Doktoranta pracy naukowej. Przedmiotem rozprawy doktorskiej są oryginalne i wartościowe rozwiązania interdyscyplinarnego problemu naukowego.

Ustosunkowując się do kwestii kategorii, do której zaliczyłbym dysertację: 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; bez wątpliwości stwierdzam, że recenzowana rozprawa doktorska Pana Mateusza Modrzejewskiego spełnia wymagania stawiane rozprawom doktorskim, przez obowiązującą ustawę, z wyraźnym nadmiarem (d), głównie za sprawą zakresu przeprowadzonych badań oraz ich interdyscyplinarnego charakteru. Tym samym wnoszę o dopuszczenie rozprawy do publicznej obrony.

Adam Wojewchowski