

# Rozmawiając o dźwięku – badanie łączące opisy barwy dźwięku w języku polskim z parametryczną analizą sygnału

Miron Markowski<sup>1\*</sup>, Urszula Kaleta<sup>1</sup>, Jan Jasiński<sup>2</sup>, Bartosz Głowacki<sup>2</sup>

<sup>1</sup> Uniwersytet Jagielloński w Krakowie

<sup>2</sup> AGH Akademia Górniczo-Hutnicza w Krakowie

\* [miron.markowski@student.uj.edu.pl](mailto:miron.markowski@student.uj.edu.pl)

## Streszczenie

Celem niniejszej pracy było zbadanie powiązań pomiędzy parametrami akustycznymi określającymi barwę dźwięków a określeniami semantycznymi wykorzystywanymi do jej opisu w języku polskim. Zostały przeprowadzone dwa eksperymenty. W pierwszym słuchacze proszeni byli o opisanie słuchanych dźwięków jak największą ilością określeń. Na podstawie zebranych odpowiedzi utworzony został zbiór określeń najpowszechniej wykorzystywanych do opisu barwy dźwięku w języku polskim. Drugi eksperyment polegał na prezentowaniu słuchaczom dźwięków i proszeniu ich o umiejscowienie ich na skalach pomiędzy antonimicznymi określeniami semantycznymi. Wyniki ankiety porównane zostały ze zbiorem parametrów akustycznych uzyskanych z analizy wykorzystanych dźwięków i wyznaczono korelację pomiędzy tymi wartościami. Z zastosowaniem analizy czynnikowej utworzony został również trójskładnikowy model przestrzeni semantycznej opisu barwy dźwięku: jasność, wydźwięk, faktura/temperatura.

## 1. Wstęp

Barwa jest właściwością dźwięku o wielowymiarowym charakterze. Najczęściej jest ona definiowana jako zbiór wszystkich cech dźwięku, które umożliwiają słuchaczowi rozróżnienie dwóch dźwięków o tej samej głośności, wysokości i długości trwania [1], jednakże definicja ta jest często krytykowana [2]. Donnadiu [3] sugeruje dwa inne, niekoniecznie wykluczające się, podejścia do definicji barwy dźwięku. Pierwsze z nich skupia się na abstrakcyjnych własnościach sygnału i definiuje barwę jako trójwymiarową przestrzeń, w której skład wchodzi własności czasowe, widmowe i widmowo-czasowe sygnału. W podejściu środowiskowym natomiast uważa się, iż układ słuchowy za pomocą barwy dźwięku koduje właściwości jego źródła. Barwa opisuje wtedy fizyczny związek przyczynowy pomiędzy źródłem dźwięku a jego percepcją. Niezależnie od przyjętej definicji pytaniem kluczowym w tym kontekście jest to, w jaki sposób odbiorca odczuwa dane dźwięki i opisuje je. Czy słuchacze robią to

w sposób podobny i spójny w reakcji na konkretny dźwięk o określonych właściwościach? Badania nad percepcją barwy dźwięku prowadzone są na całym świecie od wielu lat, jednak można zaobserwować wiele zróżnicowanych hipotez. Kluczowym problemem jest rozłam pomiędzy fizycznymi aspektami odsłuchiwanego dźwięku a odczuciami percepcyjnymi, które są wzbudzane w słuchaczu w związku z dźwiękiem oraz semantycznymi określeniami, które zostaną wykorzystane przez niego do opisu doznań [4]. Jedynie dzięki zrozumieniu powiązań i relacji pomiędzy tymi elementami możliwe jest zrozumienie barwy dźwięku.

Metodyka badawcza wykorzystywana do prowadzenia badań różni się znacząco w zależności od tego, jaki aspekt wyżej wymienionych powiązań jest rozważany. Zrozumienie neurologicznych zjawisk odpowiedzialnych za przetwarzanie muzyki stało się w przeciągu ostatniej dekady częstym zagadnieniem poruszonym w dziedzinie neurobiologii poznawczej. Rozwój technik obrazowania w neurologii umożliwia badanie aktywności poszczególnych struktur mózgu pod wpływem muzyki oraz jej barwy, jednak jest to kwestia niezwykle trudna w związku z wielowymiarowym charakterem tych zjawisk [5, 6]. Badania te pozwalają nam w większym stopniu zrozumieć powiązanie pomiędzy cechami odsłuchiwanego dźwięku a sposobem, w jaki jest on odbierany.

Ponadto, ważnym obszarem w badaniach nad barwą dźwięku jest lingwistyka poznawcza, która eksploruje m.in. empiryczny aspekt opisu barwy dźwięku. Ważne jest tu wzięcie pod uwagę wszystkich elementów tego zjawiska – fizycznego, percepcyjnego oraz werbalnego. W pracach badawczych tworzone są schematy tego, w jaki sposób zachodzi percepcja dźwięku, począwszy od sygnału, przez jego odbiór, przetwarzanie zmysłowe, konceptualizację i interpretację, a skończywszy na werbalizacji wrażenia [7].

Wielu badaczy wzięło sobie jednak za cel próbę zmapowania semantycznej przestrzeni barwy dźwięku, zróżnicowanej pod względem języka oraz metodologii [8–10]. Wyniki ich prac umożliwiają analizę wpływu cech dźwięku na sposób, w jaki będzie on opisywany przez słuchaczy. Powszechnie stosowaną metodą, skuteczną w tworzeniu i eksploracji przestrzeni semantycznej barwy dźwięku, jest *semantic differential* [8, 11]. Polega ona na podaniu listy par przeciwstawnych sobie przymiotników określających końce skali, na których oznaczane zostaje odczuwane przez badanego brzmienie. Badając wiele dźwięków i wiele par słów, uzyskuje się obraz odczuć barwy dźwięku. Następnie można dokonać redukcji liczby osi tej przestrzeni dzięki znalezieniu korelacji pomiędzy różnymi parami. Alternatywnym podejściem jest wykorzystanie metodologii badawczej VAME (*Verbal Attribute Magnitude Estimation*), która zamienia pary określeń antonimicznych na pojedyncze określenia. Uczestnicy są zatem proszeni o określenie swoich odczuć na skali pomiędzy obecnością a brakiem danej cechy (np. brak jasności – jasność). Podejście to jest przez niektórych preferowane w związku z tym, że nie ma potrzeby wyznaczania antonimicznych par określeń i problemu związanego z brakiem możliwości określenia, czy pary mają dokładnie przeciwstawne znaczenie w kontekście opisu barwy dźwięku [12]. Wiąże się jednak ze znaczącym wydłużeniem badania przy analizie takiej samej liczby określeń semantycznych.

Eksperymenty przy użyciu tych metod przeprowadzane były m.in. w języku niemieckim. Uzyskano wówczas osie *full–empty*, *dull–sharp*, *colorful–colorless* oraz *compact–diffuse* [8]. W języku angielskim zadeklarowano z kolei istnienie trzech wymiarów *bright–dull*, *warm–cold* oraz *rich–pure* [13] lub *texture*, *luminence* i *mass* [14], natomiast w języku czeskim otrzymano osie *bright/clear–gloomy/dark*, *hard/sharp–delicate/soft*, *wide–narrow* oraz *hot/heartly* [10].

W niniejszym badaniu wykorzystano metodę *semantic differential* w celu rozszerzenia obecnego stanu wiedzy na temat percepcji barwy o dane dotyczące języka polskiego oraz sprawdzenia, czy wyniki i wnioski wyciągane z wcześniejszych badań są również prawidłowe. Postawione zostały hipotezy o istnieniu skończonej liczby osi, czyli zbioru określeń semantycznych, które tworzą przestrzeń barw umożliwiającą modelowanie opisu percepcji dźwięku wśród słuchaczy, oraz o występowaniu związku pomiędzy zmianami w wyliczanych parametrach sygnałowych a tym, jak ludzie odczuwają i opisują barwę słyszanego dźwięku. Dodatkowo obecny stan wiedzy opierany jest na badaniach, które skupiały się na analizowaniu percepcji dźwięku bardzo homogenicznej próby badawczej osób z wykształceniem i doświadczeniem muzycznym. Założeniem badania jest uzyskanie odpowiedzi od grupy heterogenicznej pod względem płci oraz doświadczenia muzycznego, odzwierciedlającej bardziej reprezentatywny przekrój społeczeństwa.

## 2. Przygotowania badań

### 2.1. Badania wstępne – stworzenie zbioru określeń semantycznych

Przed przeprowadzeniem badań eksperymentalnych konieczne było stworzenie zbioru określeń semantycznych wykorzystywanych do opisu barwy dźwięków w języku polskim. Wykonany został eksperyment wstępny, w którym osoby badane poproszone zostały o opisanie usłyszanych dźwięków jak największą liczbą określeń, które według nich właściwie opisują ich barwę. Badanie wykonane zostało w formie zdalnej przez internet, co oznacza, że każdy uczestnik odsłuchiwał dźwięki w innych warunkach odsłuchowych i przy wykorzystaniu innych słuchawek. Uznane zostało to za dopuszczalne w związku z subiektywną naturą eksperymentu i brakiem potrzeby porównywania wyników różnych słuchaczy. Ponadto odtwarzane dźwięki stanowiły jedynie zachętę, zaś przedmiotem tego eksperymentu były same słowa, których używali badani.

W głównym badaniu wzięło udział 60 osób, których językiem ojczystym jest język polski. Ich dane demograficzne przedstawione są w tabelach 1–3.

**Tabela 1**

Dane demograficzne uczestników – doświadczenie

Doświadczenie	Liczba uczestników
żadne	24
amatorskie	31
profesjonalne/formalne	5

**Tabela 2**

Dane demograficzne uczestników – wiek

Wiek	Liczba uczestników
18–24	31
25–34	16
35–44	6
45–54	5
55–64	2

**Tabela 3**

Dane demograficzne uczestników – płeć

Płeć	Liczba uczestników
kobieta	27
mężczyzna	32
pozostałe	1

Wyniki ankiety poddane zostały analizie i klasyfikacji. Dokonana została selekcja, w wyniku której odrzucone zostały określenia nieodnoszące się do barwy dźwięku. Poniżej przedstawione są kategorie przymiotników, które zostały odrzucone:

- Określenia przyjemności dźwięku (np. nieprzyjemny, przyjemny, miły, drażniący, irytujący) – przymiotniki te wykorzystywane są do przedstawienia subiektywnej oceny przyjemności dźwięku przez słuchaczy i zatem nie odnoszą się bezpośrednio do barwy słuchanego dźwięku.
- Określenia opisujące cechy dźwięku niebędące jego barwą (np. wysoki, krótki, ciągły, wibrujący, monotony, wygaszający) – przymiotniki te opisują głośność, wysokość i czas trwania dźwięków wraz z uwzględnieniem, jak te parametry się zmieniają w czasie. Według większości definicji barwa dźwięku jest określana jako niezależna od tych cech.
- Określenia wzbudzanych emocji (np. niepokojący, straszny, tajemniczy, radosny, smutny) – słowa te odnoszą się do emocji, które są wzbudzane przez dźwięk u słuchacza. Są one niewątpliwie powiązane z barwą dźwięku, jednak nie są to słowa bezpośrednio ją opisujące.
- Określenia opisujące źródło dźwięku (np. strunowy, elektroniczny, gitarowy, dęty, perkusyjny) – przymiotniki te odnoszą się do sposobu produkcji dźwięku. Opisują instrument, który wytworzył dźwięk, jednak nie opisują samego dźwięku i jego barwy.

Odrzucone zostały również słowa, które w odpowiedziach ankietowanych pojawiły się rzadziej niż trzykrotnie. Pozostałe odpowiedzi przedstawione są w tabeli 4.

Porównując uzyskane wyniki z korpusami słów wykorzystywanych do opisu barwy dźwięku uzyskanych w przypadku języka angielskiego, możemy zaobserwować wiele podobieństw [15]. Odnosnie do wszystkich najczęściej występujących słów polskich możemy znaleźć odpowiednik w języku angielskim, który również został określony jako znaczący. Wyniki tych tłumaczeń przedstawione są w tabeli 5. Oczywiście nie można na tej podstawie stwierdzić, czy koncepty te są jednakowo rozumiane w różnych językach pomimo ich zgodności przy tłumaczeniu. Te same słowa mogą mieć różne konotacje i znaczenia zależnie od rodzimej kultury. Wskazuje to jednak na fakt, że przy opisie odczuwanej barwy dźwięku różne kultury sięgają do tych samych porównań i konceptów przy próbie klasyfikacji. Podobny poziom zgodności został również wykazany w przypadku porównania języka greckiego z językiem angielskim [9].

Tabela 4

Występowanie przymiotników w ankiecie badań wstępnych

Przymiotnik	N	Przymiotnik	N	Przymiotnik	N
ostry	29	delikatny	28	ciepły	28
spokojny	25	jasny	21	czysty	19
dźwięczny	18	wybrzmiewający	18	metaliczny	18
łagodny	17	miękki	16	stłumiony	14
syntetyczny	13	lekki	13	głęboki	13
kosmiczny	12	klasyczny	10	przenikliwy	10
szorstki	10	płytki	10	drewniany	9
prosty	9	brzęczący	9	głuchy	8
naturalny	8	powietrzny	7	mechaniczny	7
przytłumiony	7	świdrujący	6	świszczący	6
ciemny	6	płaski	6	neutralny	6
szumiący	6	ostrzegawczy	5	szybki	5
intensywny	5	twardy	5	zimny	5
mały	5	przeszywający	5	melodyjny	5
matowy	5	sztuczny	5	zgrzytliwy	5
konkretny	5	dudniący	5	energiczny	5
akustyczny	5	przestrzenny	4	rezonujący	4
harmonijny	4	gładki	4	obły	4
tępy	4	industrialny	4	nosowy	4
pobudzający	4	szklany	4	poszarpany	4
wyraźny	4	intrygujący	3	silny	3
wieloskładnikowy	3	wietrzny	3	chłodny	3
chropowaty	3	drżący	3	brudny	3
obojętny	3	pusty	3	orientalny	3
uporczywy	3	słodki	3	dmuchany	3
piłujący	3	odbijający	3	sprężysty	3
nieczysty	3	złożony	3	dziwny	3
sentymentalny	3	selektywny	3	piszczący	3
wyrazisty	3	klaksonowy	3	metalowy	3
zwyczajny	3	psychodeliczny	3	ogłuszający	3

**Tabela 5**

Tłumaczenia najczęściej występujących słów polskich na język angielski [9, 15]

Słowo w języku polskim	Tłumaczenie na język angielski
ostry	sharp
spokojny	mellow
czysty	clean
wybrzmiewający	ringing
łagodny	mild
stłumiony	muted
lekki	light
delikatny	gentle
jasny	bright
dźwięczny	resonant
metaliczny	metallic
miękki	soft
syntetyczny	synthetic
głęboki	deep

Najczęściej występujące słowa opisujące dźwięk zostały przeanalizowane i wybrane tak, aby utworzyć osie antonimiczne. Uwzględnione zostało również występowanie określeń zbliżonych znaczeniem [16] we wcześniejszych badaniach [2]. Wybrane pary słów pokazane zostały w tabeli 6. Występowanie słów z wybranych par antonimicznych w badaniu wstępnym oznaczone zostało na kolorowo w tabeli 4.

**Tabela 6**

Utworzone osie semantyczne

Wybrane antonimiczne pary przymiotników	
ciemny	jasny
ostry	łagodny
miękki	twardy
ciepły	zimny
syntetyczny	naturalny
dźwięczny	głuchy
zabrudzony	czysty
wybrzmiewający	stłumiony

## 2.2. Wybór dźwięków

Do wykorzystania w badaniach wybrane zostały zarówno instrumenty wchodzące w skład tradycyjnej europejskiej orkiestry symfonicznej, tradycyjne instrumenty innych kultur światowych, jak również współczesne instrumenty rozrywkowe. Celem było uniknięcie zbioru dźwięków, których barwa mogłaby być kojarzona z pojedynczym kontekstem kulturowym, co mogłoby wpływać na sposób, w jaki byłyby one postrzegane i opisywane. Wybór został również podjęty z uwzględnieniem instrumentów występujących we wcześniejszych badaniach [9], a także różnorodności pomiędzy wartościami

parametrów akustycznych dla nich wyznaczonych. Ton c1 (261,63 Hz) [17] został uznany za podstawową wysokość dźwięków, a głośność wszystkich dźwięków została znormalizowana do poziomu  $-24,5$  LUFS [18].

Ostateczny zbiór dźwięków prezentowany uczestnikom badania zawierał 18 instrumentów (fletnia pana, syntezator FM, gitara klasyczna, gitara elektryczna, gitara elektryczna z nałożonym przesterowaniem typu fuzz, klarnet, klawesyn, kontrabas, koto, marimba, syntezator moog, organy, pianino, pojedynczy sinus z nałożonym szumem białym, sitar, skrzypce, trąbka, wibrafon.

### 3. Procedura eksperymentalna

W głównym badaniu wzięło udział 75 osób, których językiem ojczystym jest język polski. Uczestnicy nie zgłosili żadnych ubytków w słuchu. Ich dane demograficzne przedstawione są w tabelach 7–10.

**Tabela 7**

Dane demograficzne uczestników – wiek

Wiek	N
18–24	43
25–34	32

**Tabela 8**

Dane demograficzne uczestników – płeć

Płeć	N
kobieta	39
mężczyzna	34
pozostałe	2

**Tabela 9**

Dane demograficzne uczestników – doświadczenie

Doświadczenie	N
żadne	42
amatorskie	25
profesjonalne / formalne	8

**Tabela 10**

Dane demograficzne uczestników – liczba godzin słuchanej muzyki w tygodniu

Ekspozycja	N
0	21
10	22
20	13
30	7
40	9
50	2
60	1

W związku z ograniczeniami pandemicznymi badania odsłuchowe musiały zostać wykonane w warunkach zdalnych. Zestaw odsłuchowy, składający się z słuchawek zamkniętych Beyerdynamic DT 770 Pro, interfejsu M-audio Fasttrack pro oraz komputera, przekazywany był uczestnikom. Uczestnicy proszeni byli o złożenie stanowiska w cichym miejscu oraz otworzenie programu, który został specjalnie stworzony do przeprowadzania badań. Po wypełnieniu kwestionariusza demograficznego wykonywana była kalibracja głośności odtwarzania dźwięków, w trakcie której słuchaczom prezentowany był zakres barwowy dźwięków, a następnie rozpoczynała się właściwa część eksperymentalna. Uczestnicy mogli dowolną liczbę razy odtworzyć każdy dźwięk i proszeni byli o ustawienie na skalach pomiędzy dwoma określeniami semantycznymi suwaka tak, aby odzwierciedlał on sposób, w jaki opisałiby barwę słuchanego dźwięku. Interfejs użytkownika oprogramowania badawczego przedstawiony został na rysunku 2. W celu uniknięcia zaburzenia wyników badania w związku ze zmęczeniem i adaptacją słuchu [19] kolejność dźwięków, kolejność skal semantycznych oraz ich kierunek były generowane losowo dla każdego uczestnika. Czas trwania badania różnił się znacząco w przypadku poszczególnych uczestników, jednak nie przekraczał 30 minut, co oznacza, że nie istniała konieczność zrobienia przerwy w trakcie jego wykonywania.



Rys. 1. Interfejs graficzny programu badawczego

## 4. Analiza wyników badań

### 4.1. Analiza rzetelności

W trakcie prowadzenia badań wielu uczestników zgłaszało, że nie rozumie rozróżnienia na skali naturalny–syntetyczny. Te obserwacje spowodowały, że w dalszej analizie skala ta nie będzie brana pod uwagę. Na potrzeby prezentacji wyników skale będą nazwane jednym z antonimicznych określeń. Wyniki dotyczące poszczególnych skal semantycznych zostały poddane analizie rzetelności dzięki wyznaczeniu współczynnika Alfa Cronbacha. Wyniki przedstawione są w tabeli 11.

Dla każdej skali  $\alpha > 0,8$ , co określa spójność uzyskanych odpowiedzi jako dopuszczalną. Aby uzyskać lepszą spójność i dokładność analizy oraz usunąć skrajne odpowiedzi badanych, odrzucone zostały wartości nienależące do przedziału średnia arytmetyczna  $\pm$  odchylenie standardowe. Po analizie histogramów określono, że wszystkie skale mają rozkład jednomodalny oraz wyliczone dla nich parametry – skośność i kurtoza – przyjmują wartości bezwzględnie mniejsze niż 4 (wartość kurtozy dla rozkładu normalnego przyjmujemy jako 0).

Tabela 11

Wyniki analizy rzetelności wyników

Cecha	Współczynnik Alfa Cronbacha
Jasny	0,89
Czysty	0,98
Zimny	0,91
Głuchy	0,96
Łagodny	0,98
Twardy	0,97
Stłumiony	0,96

Oznacza to, że pomimo negatywnego wyniku testu normalności Lillieforsa i testu Shapiro–Wilka (w obu testach wartość  $p < 0,01$ ) można zastosować dla tych danych analizę czynnikową.

## 4.2. Analiza czynnikowa

Korelację pomiędzy poszczególnymi skalami semantycznymi (wyznaczone dla poziomu istotności  $p < 0,05$ ) zaprezentowane są w tabeli 12.

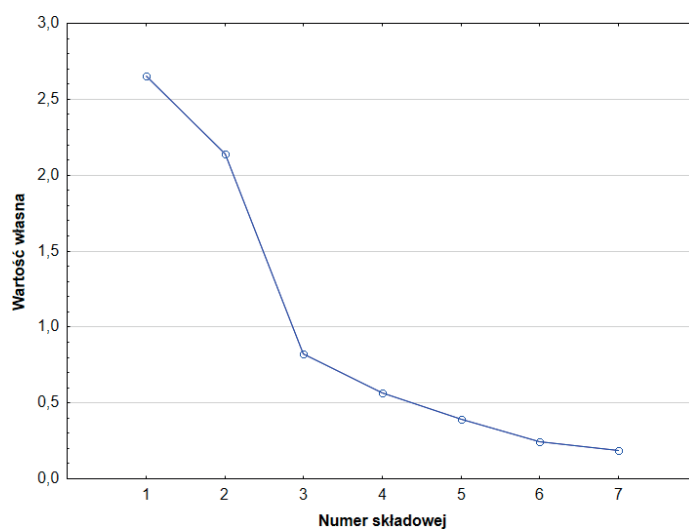
**Tabela 12**  
Korelacje pomiędzy skalami semantycznymi

Zmienna	Jasny	Czysty	Zimny	Głuchy	Łagodny	Twardy	Stłumiony
Jasny	1	0,33	-0,13	-0,27	0,04	-0,08	-0,21
Czysty	-	1	-0,37	-0,37	0,44	-0,39	-0,21
Zimny	-	-	1	0,09	-0,53	0,56	-0,05
Głuchy	-	-	-	1	0,15	-0,16	0,74
Łagodny	-	-	-	-	1	-0,81	0,30
Twardy	-	-	-	-	-	1	-0,28
Stłumiony	-	-	-	-	-	-	1

Można zaobserwować:

- bardzo wysoką korelację ( $0,7 < |r| < 0,9$ ) pomiędzy skalami:
  - głuchy-dźwięczny ↔ stłumiony-wybrzmiewający,
  - ostry-łagodny ↔ miękki-twardy;
- wysoką korelację ( $0,5 < |r| < 0,7$ ) pomiędzy skalą zimny-ciepły a skalami ostry-łagodny / miękki-twardy;
- znaczącą korelację ( $0,3 < |r| < 0,5$ ) pomiędzy skalą czysty-zabrudzony a skalami jasny-ciemny / zimny-ciepły / głuchy-dźwięczny / ostry-łagodny / miękki-twardy;
- słabą korelację lub jej brak pomiędzy pozostałymi skalami.

Celem analizy czynnikowej z metodą składowych głównych jest wyznaczenie zredukowanej liczby nieskorelowanych czynników, które wyjaśniają zaobserwowaną wariację. Analizując wykres osypiska przedstawiony na rysunku 2, ustalamy liczbę czynników, którą powinien mieć model. Ograniczenie modelu do dwóch czynników umożliwia wyjaśnienie 68% wariacji, ograniczenie do trzech czynników umożliwia wyjaśnienie 81% wariacji, natomiast ograniczenie do czterech czynników do 88% wariacji. W związku z takimi wynikami podjęta została decyzja o przyjęciu trzech czynników jako najbardziej odpowiedniego wyboru zapewniającego przejrzystość modelu, przy jednoczesnym wyjaśnieniu znacznej części wariacji. W tabeli 13 widoczny jest wpływ poszczególnych osi semantycznych na wyznaczone czynniki z przyjętą minimalną wartością własną  $\lambda = 0,82$  i po zastosowaniu rotacji Varimax.



Rys. 2. Wykres osypiska



**Tabela 13**  
Wpływ poszczególnych osi semantycznych na wyznaczone czynniki

Zmienna	Czynnik 1	Czynnik 2	Czynnik 3
Jasny	0,04	0,13	0,97
Czysty	0,59	0,42	0,37
Zimny	-0,77	-0,13	-0,02
Łagodny	0,89	-0,18	0,02
Twardy	-0,88	0,19	-0,05
Stłumiony	0,19	-0,89	-0,08
Głuchy	0,006	-0,92	-0,14
Wyjaśniona wariancja	0,37	0,28	0,16

Na czynnik pierwszy w największym stopniu składają się następujące skale: zimny–ciepły, ostry–łagodny, twardy–miękki, czysty–zabrudzony, na drugi zaś skale głuchy–dźwięczny oraz stłumiony–wybrzmiewający, natomiast na czynnik trzeci głównie skala jasny–ciemny. Po analizie powyższych wyników utworzone zostały oznaczenia czynników – kolejno faktura/temperatura, wydźwięk, jasność. Nie jest możliwe idealne oddanie charakteru czynników za pomocą pojedynczych słów, zatem nazwy te należy traktować jako aproksymacje ułatwiające dalszą dyskusję i rozważania.

### 4.3. Analiza korelacji parametrów fizycznych i skal semantycznych

Przy próbie znalezienia powiązań pomiędzy dźwiękiem a jego percepcją i opisem kluczowym krokiem jest parametryzacja i ekstrakcja cech sygnału audio. Jest to próba izolacji poszczególnych aspektów dźwięku i sprowadzenia niezwykle złożonego i wielowymiarowego charakteru dźwięku do niewielkiej liczby parametrów, które możliwe są do wykorzystania w analizie i poszukiwaniu powiązań z odpowiedziami osób badanych. Zbiór parametrów, który jest wykorzystywany i opisywany w literaturze, jest niezwykle szeroki i pozwala na opis zarówno właściwości częstotliwościowych, jak i czasowych sygnału. Zbiór użytych w badaniu parametrów wybrany został na podstawie tych, które we wcześniejszych badaniach określone zostały jako mające największy wpływ na percepcję barwy [2, 4, 5] oraz analizujące różne aspekty barwy dźwięku. Zbiór parametrów składa się z następujących elementów: *Spectrum Centroid* [20], *Spectrum Spread* [20], *High Energy-Low Energy Ratio* [20], *Tristimulus* [20], *Inharmonicity* [20], *Noisiness* [20], *Odd-Even Ratio* [20], *Zero Crossing Rate* [2], *Entropy* [2]. Cała analiza wykonana została w języku Python z wykorzystaniem bibliotek do przetwarzania audio Librosa [21] i Iracema [22]. W tabeli 14 zaprezentowane są wyniki wartości współczynnika korelacji z poszczególnymi czynnikami przestrzeni semantycznej ( $p < 0,05$ ).

**Tabela 14**  
Korelacje parametrów fizycznych z czynnikami przestrzeni semantycznej

Parametr	Faktura/temperatura	Wydźwięk	Jasność
<i>Spectrum Centroid</i>	-0,31	-0,14	0,14
<i>Spectrum Spread</i>	-0,12	-0,31	0,16
<i>High Energy-Low Energy Ratio</i>	-0,28	0,16	0,12
<i>Tristimulus 1</i>	0,38	-0,28	0,02
<i>Tristimulus 2</i>	-0,24	0,29	-0,10
<i>Tristimulus 3</i>	-0,36	0,19	0,09
<i>Inharmonicity</i>	-0,23	0,02	0,09
<i>Noisiness</i>	-0,15	0,07	0,27
<i>Odd-Even Ratio</i>	0,04	-0,44	-0,18
<i>Zero Crossing Rate</i>	-0,35	0,12	0,14
<i>Entropy</i>	-0,37	0,10	0,08

Powyższe wyniki pokazują, że z czynnikiem faktura/temperatura przeciętnie skorelowane są parametry *Spectrum Centroid*, *Tristimulus 1* i *3* oraz *Zero-Crossing Rate*, z czynnikiem wydźwięku przeciętnie skorelowane są parametry *Spectrum Spread*, *Odd-Even Ratio*, a z czynnikiem jasności przeciętnie skorelowany jest parametr *Noisiness*.

#### 4.4. Analiza regresji

Wykonana została analiza regresji krokowej wstecznej (poziom istotności  $p < 0,05$ ). Na początku w modelu były wszystkie zaprezentowane cechy (z wyłączeniem tych, które wykazywały dużą współliniowość z pozostałymi), a później stopniowo eliminowano zmienne tak, aby utrzymać jak największą wartość współczynnika determinacji, przy zachowywaniu istotności tych cech. Do analizy wykorzystane zostały wszystkie parametry fizyczne, jednak konieczne było odrzucenie parametrów *Tristimulus 1* i *Zero-Crossing Rate* w związku z ich bardzo dużą korelacją z pozostałymi parametrami, tj. odpowiednio z *Tristimulus 2* ( $-0,80$ ), *Tristimulus 3* ( $-0,80$ ) oraz z *Tristimulus 3* ( $0,88$ ), *Noisiness* ( $0,83$ ), *Entropy* ( $0,84$ ). Wartości współczynników Beta zamieszczone zostały w tabeli 15, a wartości współczynników determinacji w tabeli 16. Wpływ parametrów *Tristimulus 2* oraz *Spectrum Spread* okazał się statystycznie nieistotny.

**Tabela 15**  
Wartości współczynników Beta

Parametr fizyczny	Faktura/temperatura	Wydźwięk	Jasność
<i>Spectrum Centroid</i>	-0,33	-0,51	-
<i>HE-LE Ratio</i>	-0,43	-	-
<i>Tristimulus 3</i>	-	0,54	-
<i>Inharmonicity</i>	-	-0,27	-
<i>Noisiness</i>	0,77	0,21	0,58
<i>Odd-Even Ratio</i>	-	-0,36	-0,22
<i>Entropy</i>	-0,41	-	-0,37

**Tabela 16**  
Współczynniki determinacji

Czynnik	$R^2$ <sup>a</sup>	$F$ <sup>b</sup>	$p$ <sup>c</sup>
Faktura/temperatura	0,31	27,9 <sup>d</sup>	<0,001
Wydźwięk	0,33	30,5 <sup>e</sup>	<0,001
Jasność	0,15	11,2 <sup>f</sup>	<0,001

<sup>a</sup> Wartości współczynników determinacji

<sup>b</sup> Test Fishera-Snedecora

<sup>c</sup> Poziom istotności

<sup>d</sup>  $F(4,440)$

<sup>e</sup>  $F(5,440)$

<sup>f</sup>  $F(3,440)$

Na podstawie tych wartości można wnioskować, że czynnik faktura/temperatura można wyjaśnić w 31% za pomocą parametrów fizycznych. W największym stopniu dodatnio wpływa na niego parametr *Noisiness*, ale możemy także zaobserwować, iż w mniejszym stopniu ujemnie wpływają na niego parametry *Spectrum Centroid*, *High Energy-Low Energy Ratio* i *Entropy*. Czynnik wydźwięku możliwy jest do wyjaśnienia w 33% przez dodatni wpływ parametrów *Tristimulus 3* i *Noisiness* oraz ujemny wpływ *Spectrum Centroid*, *Odd-Even Ratio* i *Inharmonicity*. Czynnik jasności można wyjaśnić w 15% przez dodatni wpływ *Noisiness* i ujemny wpływ *Entropy* i *Odd-Even Ratio*.

## 5. Dyskusja i wnioski

Przedmiotem pracy było wykonanie badania mającego na celu odkrycie struktury opisującej semantyczny sposób określania barwy dźwięków. W wyniku badań wstępnych utworzony został korpus określeń, które w języku polskim wykorzystywane są do opisu barwy dźwięku. W celu jego weryfikacji konieczne byłoby powtórzenie ankiety z większym zbiorem dźwięków oraz większą próbą badawczą, niemniej może on służyć jako pomoc przy wyborze nomenklatury stosowanej w tekstach i wystąpieniach specjalistycznych lub przy planowaniu dalszych badań dotyczących kwestii semantycznego opisu barwy w języku polskim.

Uzyskane w wyniku badań *semantic differential* odpowiedzi mogą wskazywać na to, że istnieje zbiór określeń semantycznych w języku polskim wykorzystywany do opisu barwy dźwięków, który jest jednoznaczny i powszechnie stosowany. Analiza rzetelności uzyskanych odpowiedzi wskazuje na wysoki poziom precyzyjności i spójności.

W wyniku analizy czynnikowej wyznaczone zostały trzy składowe semantycznej przestrzeni opisu barwy, którym przypisano następujące nazwy: jasność, faktura/temperatura, wydźwięk. Wyjaśniają one 82% zaobserwowanych wariacji. Wiele prac, w których rozważano tematykę opisu barwy w różnych językach, określało trzy albo cztery osie jako wystarczające do modelowania semantycznej przestrzeni reprezentującej percepcję barwy. Niniejsze badanie potwierdziło te odkrycia. Dodatkowo uzyskane w języku polskim osie zbliżone są do tych, które określone zostały w przypadku języka angielskiego i greckiego w postaci *Luminance*, *Texture* i *Mass* [4, 9], chociaż wyraźnie widać, że pewne koncepty nie są możliwe do przetłumaczenia z jednego języka na inny. Wyniki te umacniają jednak koncepcję interlingwistycznej zgodności semantycznych sposobów opisu barwy dźwięku [4].

Na podstawie wyników eksperymentu można wnioskować, że parametry fizyczne dźwięków w umiarkowanym stopniu umożliwiają wyjaśnienie uzyskanych czynników semantycznych. Nie jest to spójne z wynikami wcześniejszych badań w tej dziedzinie. Jednym z aspektów, który może mieć na to wpływ, jest brak heterogeniczności grupy badanej. W większości badań grupę osób badanych stanowili profesjonalni muzycy, natomiast w niniejszym eksperymencie grupa była mocno zróżnicowana. Kwestia ta powinna zostać poddana ewaluacji w kolejnych badaniach, w których uwzględniona zostanie większa liczba parametrów fizycznych, a w analizie wyników zostanie wzięty pod uwagę poziom edukacji muzycznej ankietowanych.

## Bibliografia

- [1] ANSI, *Psychoacoustical Terminology*, American National Standards Institute, New York, 1973
- [2] V. Alluri, P. Toiviainen, *Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre*, *Music Perception* 2010, vol. 27, s. 223–242
- [3] S. Donnadieu, *Mental Representation of the Timbre of Complex Sounds*, w: J.W. Beauchamp (ed.), *Analysis, Synthesis, and Perception of Musical Sounds*, Modern Acoustics and Signal Processing, Springer, New York, 2007, [https://doi.org/10.1007/978-0-387-32576-7\\_8](https://doi.org/10.1007/978-0-387-32576-7_8)
- [4] A. Zacharakis, K. Pantiadis, J.D. Reiss, *An Interlanguage Unification of Musical Timbre*, *Music Perception* 2015, vol. 32, s. 394–412, <https://doi.org/10.1525/mp.2015.32.4.394>
- [5] V. Alluri, P. Toiviainen, I. Jääskeläinen, E. Glerean, M. Sams, E. Brattico, *Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm*, *NeuroImage* 2011, vol. 59(4), s. 3677–3689, <https://doi.org/10.1016/j.neuroimage.2011.11.019>
- [6] A. Caclin, E. Brattico, M. Tervaniemi, R. Näätänen, D. Morlet, M.-H. Giard, S. McAdams, *Separate Neural Processing of Timbre Dimensions in Auditory Sensory Memory*, *Journal of Cognitive Neuroscience* 2006, vol. 18, s. 1959–1972, <https://doi.org/10.1162/jocn.2006.18.12.1959>
- [7] Z. Wallmark, R. Kendall, *Describing sound: The cognitive linguistics of timbre*, w: E.I. Dolan, A. Rehdig (eds.), *The Oxford Handbook of Timbre*, Oxford University Press, 2018, s. 579–608, <https://doi.org/10.1093/oxfordhb/9780190637224.013.14>
- [8] G. von Bismarck, *Timbre of steady sounds: A factorial investigation of its verbal attributes*, *Acta Acustica united with Acustica* 1974, vol. 30, s. 146–159
- [9] A. Zacharakis, K. Pantiadis, J.D. Reiss, *An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates*, *Music Perception* 2012, vol. 31, s. 339–358, <https://doi.org/10.1525/mp.2014.31.4.339>

- [10] O. Moravec, J. Štěpánek, *Verbal description of musical sound timbre in Czech language*, w: *Proceedings of the Stockholm Music Acoustics Conference (SMAC03)*, 2003, s. 643–645
- [11] C.E. Osgood, G.J. Suci, P.H. Tannenbaum, *The measurement of meaning*, University of Illinois Press, Urbana, 1957
- [12] G. Darke, *Assessment of Timbre Using Verbal Attributes*, w: *Proceedings of the Conference on Interdisciplinary Musicology*, 2005
- [13] R. Pratt, P. Doak, *A subjective rating scale for timbre*, *Journal of Sound and Vibration* 1976, vol. 45, s. 317–328, [https://doi.org/10.1016/0022-460X\(76\)90391-6](https://doi.org/10.1016/0022-460X(76)90391-6)
- [14] A. Zacharakis, K. Pasiadis, *A confirmatory approach of the Luminance-Texture-Mass model for musical timbre semantics*, w: *AM '15: Proceedings of the Audio Mostly 2015 on Interaction with Sound. October 2015*, 2015, s. 1–5, <https://doi.org/10.1145/2814895.2814898>
- [15] Z. Wallmark, *A corpus analysis of timbre semantics in orchestration treatises*, *Psychology of Music* 2019, vol. 47, s. 585–605, <https://doi.org/10.1177/0305735618768102>
- [16] K. Szlifirski, *Pro-audio: angielsko-polski słownik terminologii nagrań dźwiękowych = English-Polish dictionary of sound engineering terms*, Audiologos, Warszawa, 2008
- [17] M. Pluta, *Zasady muzyki i notacja muzyczna*, Wydawnictwa AGH, Kraków, 2012
- [18] International Telecommunications Union, *ITU-R-BS.1770: Algorithms to Measure Audio Programme Loudness and True-peak Audio Level BS Series*, 2006, <https://www.itu.int/rec/R-REC-BS.1770-0-200607-S/en>, dostęp 10.10.2021
- [19] P. Kleczkowski, *Percepcja dźwięku*, Wydawnictwa AGH, Kraków, 2013
- [20] G. Peeters, B.L. Giordano, P. Susini, N. Misdariis, S. McAdams, *The Timbre Toolbox: Extracting audio descriptors from musical signals*, *The Journal of the Acoustical Society of America* 2011, vol. 130, s. 2902–2916, <https://doi.org/10.1121/1.3642604>
- [21] B. McFee, A. Metsai, M. McVicar, S. Balke, C. Thomé, C. Raffel, F. Zalkow et al., *librosa/librosa: 0.8.1rc2*, 2021, <https://doi.org/10.5281/zenodo.4792298>
- [22] T. Magalhaes, F. Barros, M. Loureiro, *Iracema: a Python library for audio content analysis*, *Revista de Informática Teórica e Aplicada* 2020, vol. 27(4), s. 127–138, <https://doi.org/10.22456/2175-2745.107202>