XVIII Międzynarodowe Sympozjum
Inżynierii i Reżyserii Dźwięku

18th International Symposium
on Sound Engineering and Tonmeistering

ISSET 2021

Kraków, 27–29 czerwca 2021

Kraków, 27–29 June 2021

Materiały konferencyjne

Conference proceedings

© Wydawnictwo Katedry Mechaniki i Wibroakustyki
Wydziału Inżynierii Mechanicznej i Robotyki
w Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie

ISBN 978-83-908774-7-1

Redakcja
Aleksandra Król-Nowak

Skład i redakcja techniczna
Teresa Makuch
Michał Kmiecik

http://aes.org.pl/isset2021/
http.//aes.org.pl

Audio Engineering Society Polish Section

Złożone przy pomocy LaTeX 2ε oraz INKSCAPE™

WSPIERAJĄ NAS:

Kraków

ESTRADA STUDIO

konsbud audio
AUDIO AND MEDIA SOLUTIONS

LEAPWING

CYCLING '74

Spis treści

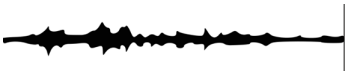
Referaty zaproszone 7
Sesja 1 – SI w inżynierii i reżyserii dźwięku 9
Sesja 2 – Konkurs referatów studenckich – część 1 22
Sesja 2 – Konkurs referatów studenckich – część 2 52
Sesja 3 – Psychoakustyka 76
Sesja 4 – Dźwięk przestrzenny 106
Sesja 5 – Rozwiązania akustyczne i elektroakustyczne . . 112

Komitet Naukowy

Przewodniczący
prof. dr hab. inż. Piotr Kleczkowski

Komitet

prof. dr hab. inż. Jan Adamczyk
dr hab. Mateusz Bień
prof. dr hab. Marek Chołoniewski
prof. dr hab. inż. Adam Dąbrowski
prof. dr hab. Rafał Delekta
prof. dr hab. inż. Andrzej Dobrucki
prof. dr hab. inż. Bożena Kostek
John Krivit
dr inż. Ewa Łukasik
prof. dr hab. Andrzej Miśkiewicz
prof. Agnieszka Roginska
prof. dr hab. inż. Krzysztof Szlifirski
dr hab. Alicja Wieczorkowska
dr hab. inż. Sławomir Zieliński
prof. dr hab. inż. Jan Żera



Komitet Organizacyjny

Przewodniczący
dr hab. Marek Pluta

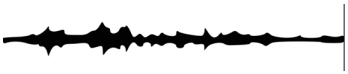
Wiceprzewodnicząca
dr inż. Dorota Czopek

Sekretarz ds. naukowych
dr inż. Aleksandra Król-Nowak

Sekretarz ds. organizacyjnych
mgr inż. Bartłomiej Chojnacki

Komitet

mgr inż. Karol Czesak
mgr inż. Jan Jasiński
mgr inż. Michał Kmiecik
mgr inż. Anna Kosiek
mgr inż. Katarzyna Kotarba
mgr inż. Teresa Makuch
mgr inż. Wojciech Wronka
mgr inż. Kamil Zimny



Referaty zaproszone

niedziela, 27 czerwca 2021
16:30
John Krivit
How the audio industry has changed during the COVID-19 crisis

poniedziałek, 28 czerwca 2021
10:00
Roger Roshnik
Challenges on designing professional monitors

poniedziałek, 28 czerwca 2021
15:00
Agnieszka Roginska
Immersive sound expands our reality

Referaty

Sesja 1

SI w inżynierii i reżyserii dźwięku

niedziela, 27.06.2021
17:30 – 18:10

ROBUST, HYBRID ALGORITHMS IN AI-AIDED AUTOMATIC MUSIC PRODUCTION

P. Skrzyński[1], J. Kwiecień[1], M. Pluta[1], A. Dąbrowski[2], and B. Szadkowski[2]

[1]AGH University of Science and Technology, Kraków, Poland
[2]Independent Digital Sp. z o.o., Warszawa, Poland

skrzynia@agh.edu.pl

Abstract

Music market is currently vastly dominated by streaming services. Growing popularity of streamed music consumption enables each music genre to find its niche. Significant position in this respect belongs to so-called relaxation music. This genre, along with several others, is defined well enough in terms of musical form and structure, to carry out an attempt at making the process of their production fully automated. This paper presents selection of algorithms implemented in a system designed to produce relaxation, dance, and electronic music. A hybrid solution combines robustness and controllability of classic approaches with ability to generate and evaluate complex structures inherent for AI-based methods.

Acknowledgments

This paper is a part of dissemination of the results of R&D project titled "The research and development work on the sound processing and creation of compositions and music recordings based on the artificial intelligence, in the integrated system of the digital content distribution". Project was co-finansed by the ROP Regional Operational Programme for Mazowieckie Voivodeship 2014-2020, Poland (RPO WM 2014-2020) and realised by Independent Digital Sp. z o. o.

1. Introduction

The era of streaming services and advances in audio engineering brought better, easier to use, broadly available tools for musicians and music producers, matched by convenient means to quickly publish new music worldwide. It enables more people than ever to be involved on both sides: firstly – in, or around production of music, and secondly – in reception, or as it is nowadays more often referred to, in consumption of music. It is a very favourable time for musical experiments – vast domination of streaming services, and popularity of streamed music consumption enables each music genre to find its niche. Some of more interesting experiments, that have been attempted times and again before, but only recently received appropriate means to succeed, are the experiments in automatic creation of music: not only algorithmic composition, but full process of music production.

This paper presents a selection of algorithms implemented in a system designed to produce music in three different genres: relaxation, dance, and electronic. While an automatic composition of music is a relatively old idea [1, 2, 3], the system proposed is a fully-automated tool not only for composing, but also for producing music – from a concept to a recording. It has been designed to be reliable, and produce usable results. Therefore a hybrid solution has been applied. It combines robustness and controllability of classic approaches with ability to generate and evaluate complex structures inherent for AI [4, 5].

2. Solution

2.1. Overview

Music production is often described as a process consisting of composition, sound design, arrangement, mixing, and mastering. What is not always mentioned, but shall be regarded a key stage, is a critical evaluation of the effect. In an attempt to recreate this process in automatic approach, a model has been designed, referred to as "generator-critic" (Fig. 1).

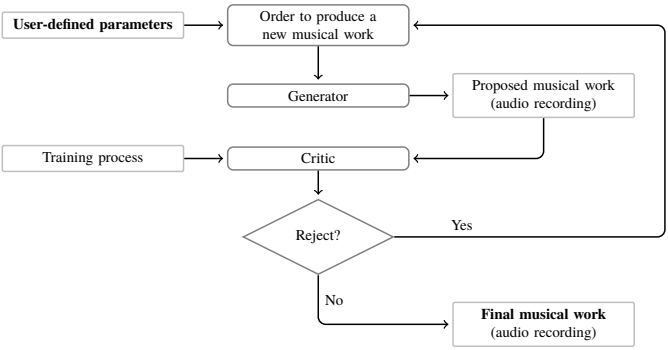


Figure 1 The generator-critic model

The generator is responsible for the creative tasks. Given a set of initial directions regarding general characteristics of music to produce, it composes a symbolic musical score, arranges it, and transforms it to an audio form, thus producing a ready-to-use recording. The output might not always be up to the requirements of a user. It is a task of the critic to evaluate the final recording, and in case of negative result, to order the generator to start over, with the same user-settings, but with a different internal decisions. The process is repeated as long as the evaluation is negative. The output is produced to the user only after receiving positive evaluation from the critic.

2.2. Critic and Generator

The critic is based on an artificial neural network [6], and the data it uses are the parameters extracted from final audio signal produced by the generator. The set includes low level audio descriptors (such as dynamic and spectral complexity, or pitch salience), rhythm descriptors (such as beats count and loudness, or onset rate), and tonal descriptors (chords changes rate, or key strength) [7].

While a neural network proved to be a viable solution for the critic, no such straightforward solution was available for the generator. In recent years one could observe a fast development and growing interest in deep learning (DL) techniques applied to art-creation process [8]. While such approach has been considered and tried for the system, the results were often erratic, with unusable output. Moreover, deep networks tend to mimic music used for their training. The problem originates from insufficient amount of music in a score form for particular genres for deep networks to generalize upon. Even with some promising results obtained during development, in its current state DL is not stable and reliable enough. Therefore a different approach has been applied.

The generator performs a chain of tasks, using classical, and AI-based techniques such as fuzzy logic, genetic algorithms, and rule-based systems. A set of user-provided input parameters is interpreted using fuzzy logic. The set includes genre, duration, tempo, mood, oddity parameter, and fine pitch-tuning information (Table 1). They are used to design a form of a musical work to be produced, including harmonic layer and structure of a composition. Rule based systems produce a set of lead motifs, assembled into phrases

Table 1 User-provided input to generator for a single musical work

Input	Interpretation
Genre	Single-choice from the following list: relaxation, dance, electronic
Duration	Target duration [s]
Tempo	Target tempo [BPM]
Mood	Target mood setting in range from *sad* to *happy*
Oddity	Amount of unconventional elements in range from none (setting *normal*) to many (*odd*)
Tuning	Pitch tuning reference point – fundamental frequency of A4 [Hz]

with genetic algorithm – both techniques apply selected rules of counterpoint. Other parts, such as bass, drums, pads, ambient sounds, etc. adapt various combinations of these techniques. A score is converted to MIDI and passed to a sequencer controlling a sampler that produces audio parts to be mixed, and finally – mastered.

3. Algorithms

A key feature of a process applied in a production environment is its reliability and robustness. While experimental approaches based entirely on DL techniques may produce interesting results, in current state they pose serious problems limiting their usability. A primary limitation is a requirement for very large database of digital scores categorised into styles to learn and generalise upon, which – apart from a few exceptions such as Bach or Mozart – is not, and will not be available. Even provided with appropriate scores, DL-based solutions rarely produce usable output. Finally, control over such processes is limited, so that a user might not affect features of produced music in a meaningful way. On the other hand, traditional approaches, based entirely on predefined rules or statistics, tend to produce predictable results.

Table 2 Use of AI methods

Task	FL	GA	RBS	ANN
Form design	+			
Harmonic progression generation	+			
Lead motifs generation	+		+	
Lead phrases design		+	+	
Drum patterns design		+	+	
Bass lines generation	+			
Accompanying tracks design	+		+	
Applying effects and mixing			+	
Classification (critic)				+

As a way to solve above mentioned problems, a process of automatic music production has been divided into stages (Fig. 2) handled by various algorithms integrated into a hybrid system. A few key stages that involve creative tasks are handled by AI algorithms such as genetic algorithms (GA), fuzzy logic (FL), rule-based expert systems (RBS), and artificial neural networks (ANN), as pointed in Table 2.

3.1. Lead-melody generator

The system is designed to generate various genres of music. Therefore, the melody generator needs to be flexible and configurable. The approach applied is to combine RBS and FL to handle and mix meaningful music features with a GA to combine small elements into varied and evolving, but consistent musical phrases.

Lead melody is generated within two main stages:

- generation of motifs (basic building blocks),
- modification and arrangement of motifs into longer sequences.

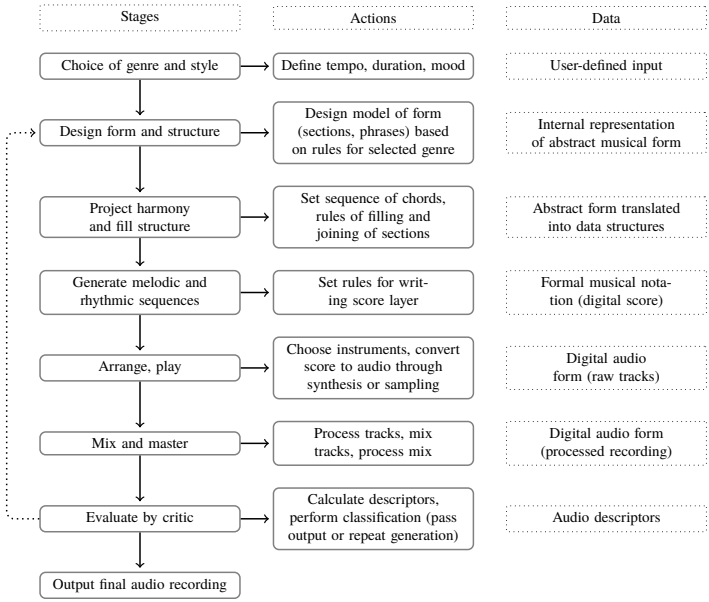


Figure 2 Stages of automated music production process

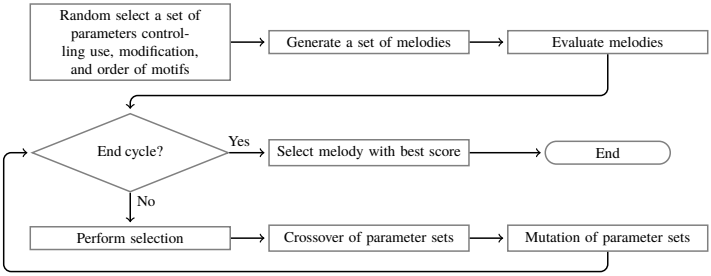


Figure 3 Lead instrument phrase generator algorithm

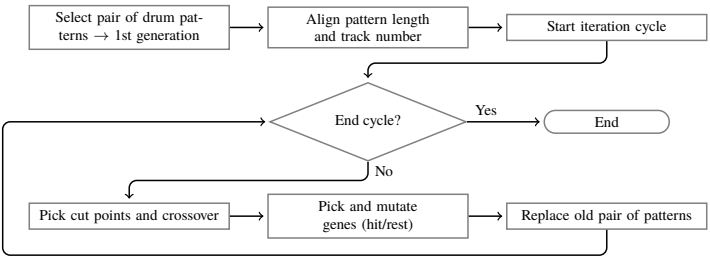


Figure 4 Drum pattern generator algorithm

Motifs are the most basic melo-rhythmic sequences, typically few-note long. A set of motifs can be arranged into a sequence as long as required. Generation of motifs is guided by rules of counterpoint that prevent some, while rewarding other pitch intervals, depending on vertical and horizontal note context. Such rules lead to fluent, natural sequences. A small set of motifs is arranged into a longer sequence using a GA-based algorithm presented in Fig. 3. The GA encoding is a set of parameters controlling transformations and order of motifs, while fitness is estimated using accordance of output melody itself with rules of counterpoint. As a result, phrases are fluent due to counterpoint restrictions, and mostly homogeneous due to common motif-base – both features are musically desirable in lead melodies.

3.2. Rhythmic sequence generator

Generator of rhythmic sequences utilises a large set of initial rhythmic patterns. Pattern is a 2-dimensional binary array with 1 representing instrument hit, and 0 – rest. Rows represent instruments, columns – subsequent time-steps. GA-based algorithm picks two predefined patterns to produce a new one, as shown in Fig. 4. In this case the GA encoding is a direct representation of pattern.

3.3. Classifier

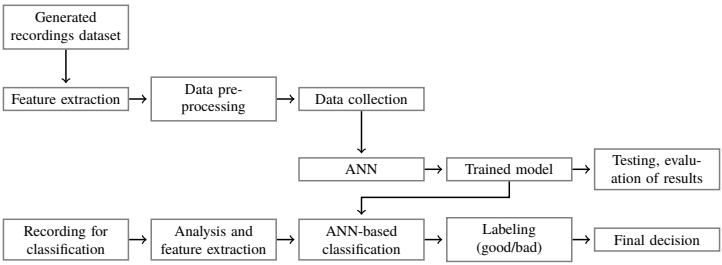


Figure 5 Critic algorithm

Assuming that only a portion of recordings produced by the generator is acceptable, the critic has to classify every new recording into one of two groups: good (acceptable) or bad. The algorithm applied to carry out this task is ANN with supervised learning (Fig. 5).

Initially, a large collection of recordings of each genre is produced by the generator and passed to human experts for binary evaluation. For every recording a set low-level, rhythmic, and tonal descriptors is calculated. Descriptors with expert evaluation assigned are used as a training set for the multilayer

feedforward neural network, with backpropagation algorithm as a training technique. Once trained, ANN is able to generalise and perform classification for new recordings belonging to trained genres.

4. Conclusions

At current stage of development the system produces varied and complete examples of three distinct genres. All the necessary stages of music production process have been implemented. Due to algorithms applied, the procedure is reliable, configurable, and fast – entire production process is by order of magnitude shorter than duration of generated music.

However, there is always a room for improvement. Further works will involve tuning some of the generator stages, and obtaining broader training set for the critic – the latter involves significant effort of human experts to mark acceptable examples. Quality and diversity of the output may be improved by adding more varied, and higher-quality sound samples to the sampler. Finally, a broader range of means may be applied on the side of signal effects, in addition to the current set of filters, amplifiers, low frequency oscillators and envelope generators.

References

[1] L. A. Hiller Jr. and L. M. Isaacson: "Musical Composition with a High-Speed Digital Computer", in: *J. Audio Eng. Soc* 6.3, 1958, pp. 154–160

[2] Roberto De Prisco, Gianluca Zaccagnino, and Rocco Zaccagnino: "A Genetic Algorithm for Dodecaphonic Compositions", in: Apr. 2011, pp. 244–253, ISBN: 978-3-642-20519-4, DOI: *10.1007/978-3-642-20520-0_25*

[3] Maciej Komosinski and Piotr Szachewicz: "Automatic species counterpoint composition by means of the dominance relation", in: *Journal of Mathematics and Music* 9.1, 2015, pp. 75–94, DOI: *10.1080/17459737.2014.935816*, eprint: *https://doi.org/10.1080/17459737.2014.935816*

[4] J.D. Fernandez and F. Vico: "AI Methods in Algorithmic Composition: A Comprehensive Survey", in: *Journal of Artificial Intelligence Research* 48, Nov. 2013, pp. 513–582, ISSN: 1076-9757, DOI: *10.1613/jair.3908*

[5] Filippo Carnovalini and Antonio Rodà: "Computational Creativity and Music Generation Systems: An Introduction to the State of the Art", in: *Frontiers in Artificial Intelligence* 3, 2020, p. 14, ISSN: 2624-8212, DOI: *10.3389/frai.2020.00014*

[6] Andries P Engelbrecht: *Computational intelligence: an introduction*, John Wiley & Sons, 2007

[7] Thomas Lidy, Andreas Rauber, Antonio Pertusa, and José Manuel Inesta Quereda: "Improving Genre Classification by Combination of Audio and Symbolic Descriptors Using a Transcription Systems.", in: *ISMIR*, Citeseer, 2007, pp. 61–66

[8] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet: *Deep Learning Techniques for Music Generation – A Survey*, 2019, arXiv: *1709.01620* [cs.SD]

INTERNETOWY SYSTEM ROZPOZNAWANIA ZAPISU NUTOWEGO WYKORZYSTUJĄCY GŁĘBOKIE SIECI NEURONOWE

M. Kałamoniak[1], Z. Piniarska[1], M. Subocz[1] i E. Łukasik[1]

Instytut Informatyki, Politechnika Poznańska
mateusz.kalamoniak@student.put.poznan.pl

Streszczenie

W artykule przedstawiono prototyp internetowego systemu odczytu zeskanowanych zapisów nutowych (*Optical Music Recognition, OMR*) wykorzystującego głębokie sieci neuronowe. Za pośrednictwem interfejsu w przeglądarce internetowej następuje przesłanie obrazu z zapisem nutowym na serwer w celu jego analizy i odbiór linków do pobrania przetworzonych wyników w formacie MusicXML i MIDI. Aplikacja serwerowa jest odpowiedzialna za wykrywanie i klasyfikowanie symboli muzycznych w obrazie oraz odwzorowanie ich kompozycji w formacie odczytywalnym maszynowo. Eksperymentalnie wykorzystana sieć neuronowa YOLO V5 dała obiecujące wyniki rozpoznawania symboli zapisu nutowego z przesłanego obrazu.

1. Wstęp

Mimo długiej historii przemysłu fonograficznego, wciąż jednak większość spuścizny muzycznej jest zapisana w postaci nut, a nie w postaci nagrań. Aktualnie zbiory muzykaliów w bibliotekach cyfrowych mogą być w pełni docenione jedynie przez użytkowników znających nuty. Dla innych użytkowników byłaby interesująca możliwość posłuchania muzyki reprezentowanej za pomocą zapisu nutowego. Niezwykle przydatne byłoby również automatyczne odtworzenie zapisu nutowego, które zastąpiłoby wykonywaną do dziś ręczną transkrypcję materiałów nutowych. *Optical Music Recognition*, w skrócie OMR, to dziedzina naukowa badająca problem komputerowego dekodowania i interpretacji notacji muzycznej zapisanej w postaci obrazu (skan, fotografia, rękopis), w sposób analogiczny do interpretacji człowieka [1]. Dla prawidłowej interpretacji zapisu nutowego potrzebne jest spojrzenie z perspektywy wielu dziedzin nauki. Rozpoznanie zapisu nutowego musi prowadzić nie tylko do odtworzenia zapisu nutowego w formie zgodnej ze wzorcem, ale właśnie do odegrania rozpoznanej muzyki zapisanej nutowo. Nie jest to łatwe ze względu na fakt, że na interpretację danego symbolu mogą wpływać znaki oddalone od danej nuty, a jest to tylko jedna z trudności interpretacyjnych automatycznego zapisu nutowego.

Wydajny algorytm OMR o dużej niezawodności pozwoliłby na zrewolucjonizowanie sposobu archiwizacji ogromnych zasobów utworów muzycznych, jakie są zebrane w bibliotekach. Dałby również możliwość odsłuchania zapisów nutowych w czasie ich przeglądania. Stąd od kilku dziesiątków lat podejmowane są próby realizacji systemów optycznego rozpoznawania zapisów nutowych OMR. Początkowo systemy OMR były oparte na regułach i obejmowały potok operacji, na które składały się: i) binaryzacja dostarczonego obrazu, detekcja pięciolinii, iii) lokalizacja symboli muzycznych, iv) klasyfikacja symboli muzycznych prezentacja wyników w postaci graficznej lub/i muzycznej (możliwość odegrania) [2]. W ostatnich latach pojawiły się propozycje zastosowania głębokich sieci neuronowych dla realizacji zadania optycznego rozpoznawania zapisu nutowego (OMR) oraz szereg anotowanych baz dokumentów muzycznych służących trenowaniu sieci neuronowych. Są to rozwiązanie otwarte, a ich twórcy zachęcają społeczność informatyków muzycznych do rozwijania systemów OMR.

Niniejszy artykuł opisuje projekt wykorzystujący i rozwijający wspomniane badania z wykorzystaniem sieci neuronowych *YOLO V5 oraz MungLinker*. Sieć *YOLO* zalicza się do grupy tak zwanych *One Stage Detector*, czyli sieci neuronowych, które wykrywają wszystkie obiekty w jednej iteracji, co skraca czas potrzebny do wytrenowania sieci oraz czas rozpoznania obiektów nutowych.

2. Systemy OMR

Badania w dziedzinie OMR mają długą historię, która rozpoczęła się w latach sześćdziesiątych ubiegłego wieku [3,4]. Prace rozwojowe prowadzili Bainbridge [1] i Fujinaga [5] oraz inni [6,7] pozostawiając wciąż otwarte kwestie do rozwiązania.

Zaawansowane algorytmy są wykorzystywane nie tylko do drukowanych partytur, ale także do pisma odręcznego notacji muzycznej [6,8,9]. Stosowano różne metody przetwarzania obrazów, a do klasyfikacji - różne metody statystyczne z wykorzystaniem maszyn wektorów nośnych (SVM), sieci neuronowych (NN), k-najbliższych sąsiadów (kNN) i ukrytych modeli Markowa (HMM). Metody te porównali Rebelo i in. w jednej z częściej cytowanych prac w dziedzinie OMR [10].

Pomimo tego, że wiele problemów badawczych nie zostało jeszcze w pełni rozwiązanych, istnieją aplikacje komputerowe, które wykonują to zadanie dla użytkowników. Dobrze znane są m.in. programy komercyjne, takie jak SmartScore, PhotoScore Ultimate, SharpEye oraz oprogramowanie otwarte – Audiveris [11]. Każdy z tych programów przetwarza obraz nuty i dokonuje jego transkrypcji do formatu MusicXML.

W ostatnich latach w wielu problemach uczenia maszynowego, stosuje się z powodzeniem głębokie sieci neuronowe (ang. *Deep Neural Networks, DNN)*. Sieci te mają wiele warstw z funkcjami aktywacyjnymi, które są wykorzystywane przed propagacją informacji do następnej warstwy. Im więcej warstw ma model, tym lepiej może wykryć ukryte nieliniowe relacje między danymi. Dane mogą również reprezentować obiekty notacji muzycznej, stąd popularne staje się zastosowanie głębokich sieci neuronowych w dziedzinie OMR. Głębokie sieci neuronowe potrzebują bardzo dużych oznakowanych zbiorów danych uczących. W przypadku OMR jest to poważny problem do rozwiązania, gdyż takie zbiory dopiero powstają, a wymagają oznaczenia wielu drobnych obiektów na jednej stronie i oznaczenia właściwych powiązań między nimi. Niedawne prace Pachy i in. [12] pokazują zastosowanie sieci Faster R-CNN [13] do rozpoznawania odręcznej notacji muzycznej z wykorzystaniem bazy MUSCIMA++, ze średnią precyzję sięgającą 80%.

3. Zbiory danych uczących dla OMR

Symbole muzyczne opisujące elementy dzieła muzycznego ułożone są na pięciolinii oraz w jej bliskim sąsiedztwie. Każdą pięciolinię definiuje klucz, który określa jakie dźwięki znajdą się na polach i liniach. Nuty niosą ze sobą informację o długości trwania dźwięku oraz jego wysokości. Symbolami wpływającymi na podwyższenie lub obniżenie dźwięku o pół tonu są bemol oraz krzyżyk. Stosuje się również podwójne bemole lub podwójne krzyżyki oraz kasowniki służące do anulowania działania wcześniej wymienionych znaków przygodnych. Wokół nut mogą znajdować się również inne symbole związane z pozostałymi elementami dzieła muzycznego. Specyfika baz danych reprezentujących obrazy zawierające zapis nutowy jest taka, że obraz zawiera bardzo dużą liczbę bardzo małych obiektów, jak laska i główka nuty, znaki przygodne itp., których położenie musi być określone z dokładnością do jednego piksela oraz muszą być zdefiniowane relacje między obiektami (np. nuta i kropka). Do najliczniejszych zbiorów oznakowanych w powyższy sposób należą DeepScores V2 [14] oraz MUSCIMA++ [15].

4.1. DeepScores V2

Zbiór danych DeepScores V2 [14] zawiera 300 tysięcy stron wygenerowanych nut , renderowanych w rozdzielczości 400 dpi, co czyni go jednym z największych tego typu zbiorów danych. Każdy symbol w notacji jest opisany poprzez klasę, która identyfikuje symbol oraz jego położenie na zdjęciu. Zawiera dziesiątki milionów obiektów, podzielonych na 123 klasy wraz z informacją o rozmieszczeniu kolejnych symboli za pomocą współrzędnych, unikatowych identyfikatorów i zdefiniowanych klas. Ten zbiór danych jest również kompatybilny ze zbiorem MUSCIMA++ [15].

4.2. MUSCIMA++

Zbiór danych MUSCIMA++ [15] składa się ze 140 stron ręcznie zapisanych i anotowanych nut. Każdy symbol jest oznaczony prostokątną obwiednią (ang. *bounding box)* i jej współrzędnymi na

płaszczyźnie oraz jest przypisany do odpowiedniej kategorii. Dodatkowo MUSCIMA++ zawiera informację o relacjach między symbolami.

Zbiór zawiera 91255 symboli, składających się zarówno z elementów składowych notacji (główki nut, laski), jak i obiektów notacji wyższego poziomu, takich jak sygnatury kluczy lub sygnatury czasowe.

5. Głębokie sieci neuronowe

Głębokie sieci neuronowe są narzędziem informatycznym szeroko wykorzystywanym do przetwarzania obrazów. Można wydzielić kilka typów zagadnień, na które dzieli się zastosowania sieci neuronowych, w tym przede wszystkim:

- Klasyfikacja obrazu,
- Lokalizacja obiektów,
- Segmentacja obrazu,
- Segmentacja obiektów,
- Detekcja obiektów.

W opisywanym projekcie, po analizie efektywności różnych architektur sieci neuronowych, do detekcji elementów zapisu nutowego eksperymentalnie wykorzystano sieć *YOLO V5* [16] (inaczej niż w projekcie A. Pachy i in. [12], na którym się wzorowano, a którzy stosowali sieć Fast R-CNN) ze względu na szybkość działania oraz szczegóły implementacyjne. Sieć zalicza się do grupy tak zwanych *One Stage Detector*, czyli sieci neuronowych, które wykrywają wszystkie obiekty w jednej iteracji, co przyśpiesza fazę trenowania sieci oraz skraca czas predykcji. Do oceny, czy istnieje zależność pomiędzy sąsiednimi obiektami na obrazie wykorzystano konwolucyjną sieć *MungLinker* dedykowaną do zastosowań OMR. Na wyjściu wystarcza jeden neuron zwracający wartość sigmoidalnej funkcji unipolarnej informujący, czy wykryto poszukiwaną zależność. Pozyskane informacje pozwalają zbudować graf notacji muzycznej, który informuje o zależnościach pomiędzy elementami wykrytymi wcześniej przez sieć *YOLO V2* [16]. Graf notacji opisuje relacje pomiędzy wszystkimi elementami wykrytymi na obrazie. Wierzchołki reprezentują symbole notacji muzycznej, a krawędzie relacje miedzy nimi. Graf tworzony jest zgodnie z wczytanymi regułami, które opisują formalne relacje pomiędzy poszczególnymi elementami notacji muzycznej. Na przykład główka nuty może być powiązana z laską i flagą, krzyżykiem, bemolem lub kasownikiem, ale nie może być powiązana z pauzą.

5.1. YOLO V5

YOLO V5 (skrót od angielskiego *You Only Look Once* – „patrzysz tylko raz") [16] jest jedną z wiodących architektur sieci neuronowych wykorzystywanych do detekcji obiektów. Zalicza się do grupy tak zwanych *One Stage Detectors*, czyli sieci neuronowych, które wykrywają wszystkie obiekty w jednej iteracji. Wpływa to korzystnie na czas potrzebny do wytrenowania sieci oraz czas rozpoznawania.

Wykorzystana w tej pracy wersja piąta [16] została opracowana przez firmę Ultralytics. Sieć została zaimplementowana w języku Python z wykorzystaniem biblioteki PyTorch. Sieć jest w pełni konwolucyjna. Zdjęcia są jedynie skalowane do zadanej długości dłuższej z krawędzi. Taki zabieg jest konieczny do skutecznego trenowania sieci z wykorzystaniem pól kotwiczących.

W architekturze YOLO można wyróżnić trzy kluczowe elementy.

1. *Backbone* - jest to część sieci odpowiadająca za agregację i formowanie wyekstrahowanych z obrazu cech. Formuje cechy na trzech różnych poziomach ‚abstrakcji", od najbardziej szczegółowych, do znacząco generalizujących.

2. *Neck* - łączy cechy obrazu w zestawy tych samych cech o rożnej wielkości i skalowaniu. Pozwala to sieci rozpoznawać obiekty danej klasy bez względu na ich rozmiar na zdjęciu. Jest to szczególnie przydatne w skutecznym rozpoznawaniu obiektów z tej samej klasy, które występują w różnym rozmiarze, jak na przykład belki w notacji muzycznej.

3. *Head* - ten fragment sieci na podstawie cech utworzonych przez *Neck* wykrywa pola otaczające symbol oraz przypisuje im klasy. Sieć dzieli zdjęcia na siatkę kwadratowych pól. Następnie jeżeli w którymś z nich wykrywa obiekt, przypisuje mu trzy informacje :

a) pole kotwiczące,

b) wartości [cx, cy] opisujące przesunięcie pola w siatce pól względem początku układu współrzędnych oraz [tx, ty, tw, th], które opisują współrzędne przesunięcie danej wartości względem domyślnej dla przypisanego pola kotwiczącego,
c) rozkład prawdopodobieństwa dla wszystkich możliwych klas obliczony za pomocą regresji logistycznej.

5.2. MungLinker

Sieć neuronowa MungLinker [17] wykorzystana w zaimplementowanym modelu ma architekturę konwolucyjną. Na wejściu zostaje wprowadzona próbka wycięta z obrazu, na podstawie której sieć ma ocenić, czy pomiędzy dwoma obiektami pokazanymi na obrazie jest zależność, czy jej nie ma. W tym celu na wyjściu wystarcza jeden neuron zwracający wartość sigmoidalnej funkcji unipolarnej informujący, czy wykryto zależność. Pozyskane informacje pozwalają zbudować graf notacji muzycznej, który wskazuje zależności pomiędzy elementami wykrytymi wcześniej przez sieć YOLO V5.

6. System OMR z wykorzystaniem sieci YOLO V5 oraz MungLinker

Realizowany system OMR opiera się na wcześniej wytrenowanym module sieci neuronowych *YOLO V5* oraz *MungLinker*. Załadowany w aplikacji obraz notacji muzycznej zostaje najpierw przetworzony pod kątem wyszukania pięciolinii, następnie wykonywana jest predykcja na sieci YOLO, a potem zadanie łączenia symboli wykonuje sieć *MungLinker*. Z finalnie stworzonej struktury połączeń między symbolami budowany jest plik MusicXML oraz MIDI. Procedura jest następująca:

1. Zdjęcie zapisu nutowego trafia do sieci *YOLO V5*, która zwraca wykryte symbole,
2. Zdjęcie przekazywane jest do modułu wykrywania pięciolinii, który zwraca wykryte linie pięciolinii,
3. Zdjęcie oraz wykryte symbole i pięciolinie przekazywane są do sieci *MungLinker*, która tworzy graf notacji.
4. Moduł odpowiadający za przetworzenie grafu notacji tworzy plik MusicXML i MIDI.

Do wykrycia na obrazie pięciolinii został użyty algorytm opierający się na przetwarzaniu obrazu z pomocą biblioteki OpenCV.

Wyniki testów zaproponowanej architektury na zbiorze *DeepScores V2* okazały się bardzo dobre, porównywalne z osiągniętymi w badaniach przeprowadzonych przy tworzeniu tego zbioru danych [14]. Autorzy [14] przetestowali dwie architektury - *Deep Watershed Detecto*r oraz *Faster R-CNN*. Pierwsza z nich uzyskała wynik mAP@0.5 = 0.422, czyli o jeden punkt procentowy gorszy od osiągniętego przez wykorzystaną tej pracy architekturę *YOLO V5*. Sieć w architekturze *Faster R-CNN* uzyskała wynik mAP@0.5 = 0.676, który jest lepszy od osiągniętego przez zastosowaną architekturę. Autorzy zbioru *DeepScores* V2 wskazują, że metryka mAP nie jest najlepsza do oceny działania sieci w badanym problemie. Ocena modelu dla progu IoU = 0.5 nie jest dobrym pomysłem szczególnie dla małych obiektów, dla których IoU nie jest odporne na drobne przesunięcie wykrytej anotacji względem oryginalnej.

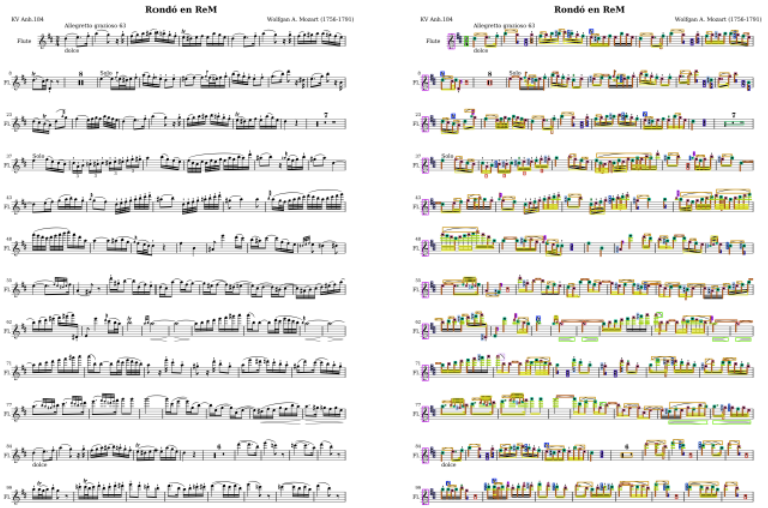
Do oceny klasyfikatora *MungLinker* wybrano metryki *Precision, Recall* oraz *F1-Score*. Podczas trenowania modelu wykorzystano moduł *Early Stopping*, który wstrzymuje trenowanie modelu po 10 epokach bez poprawy wartości funkcji straty na zbiorze walidacyjnym. Uzyskano w ten sposób najlepsze wyniki po 10 epokach, przy czym do przerwania doszło po 20 epokach. Wyniki dla zbioru walidacyjnego i testowego okazały się być podobne i dla zbioru testowego kształtowały się następująco:

- *Precision* - 0.955
- *Recall* - 0.658
- *F1-Score* - 0.778

Bardzo podobne wartości metryk dla zbioru walidacyjnego i testowego świadczą o niewystąpieniu zjawiska *overfitting*'u, czyli przesadnego dopasowania danych do zbioru treningowego i walidacyjnego. Wysoka wartość *Precision* wskazuje, że praktycznie wszystkie przypisania do klas zostały wykonane prawidłowo. Tylko niewielka część danych została błędnie sklasyfikowana jako pozytywna. Niższa, aczkolwiek i tak wysoka wartość *Recall* świadczy o dobrym poziomie rozpoznawania rzeczywiście istniejących połączeń. Około jednej trzeciej istniejących połączeń nie zostaje utworzonych, co jest mniej istotne, niż gdyby połączenia zostały wykonane nieprawidłowo. Wartość metryki *F1-Score* wskazuje na

wysoką skuteczność klasyfikatora, który w zdecydowanej większości przypadków poprawnie klasyfikuje dane wejściowe.

Na Rysunku 1. Przedstawiono przykład testowy: oryginalne zdjęcie oraz zaznaczone wykryte i skategoryzowane i oznaczone za pomocą kolorów elementy notacji muzycznej ze zbioru *DeepScores V2* o dużej gęstości symboli.



a) b)

Rysunek 1 Przykład testowy ze zbioru *DeepScores V2* o dużej gęstości symboli: a) oryginał, b) zaznaczone elementy wykryte i skategoryzowane (oznaczone kolorami).

7. Podsumowanie

W artykule przedstawiono wyniki badań optycznego rozpoznawania notacji muzycznej za pomocą głębokich sieci neuronowych. Praca wymagała analizy wielu architektur sieci neuronowych, jako potencjalnie najlepszych do postawionego zadania. Eksperymentalnie wybrano do detekcji obiektów zapisu muzycznego sieć *YOLO V5* w połączeniu z siecią *MungLinker*. Uzyskano bardzo obiecujące wyniki, które dają podstawę do kontynuowania prac z wykorzystaniem sieci YOLO. Wykonano prototypową aplikację internetową, w której wszystkie opisane działania były wykonane na serwerze. Z poziomu przeglądarki internetowej przesyłao na serwer zdjęcia i odbierano linki do wygenerowanych plików MusicXML i MIDI. Perspektywy rozwoju oprogramowania obejmują: rozszerzenie zbioru danych o skany autentycznych zachowanych dokumentów nutowych, opracowanie bardziej efektywnej metody łączenia elementów zapisu nutowego, niż to oferuje *MungLinker*, trenowanie sieci neuronowej na wyspecjalizowanym sprzęcie bez ograniczeń wydajnościowych orze przygotowanie bardziej rozbudowanego interfejsu użytkownika.

8. Bibliografia

[1] J. Calvo-Zaragoza, D. Rizo. „End-to-end neural optical music recognition of monophonic scores", *Applied Sciences*, 8(4), 2018.

[2] D. Bainbridge, *Extensible Optical Music Recognition*, Ph.D. thesis. Department of Computer Science, University of Canterbury, Christchurch (1997)

[3] D. Pruslin, *Automatic Recognition of Sheet Music*. Sc.D. dissertation, Massachusetts Institute of Technology, Cambridge (1966)

[4] D.S. Prerau, *Computer Pattern Recognition of Standard Engraved Music Notation*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge (1970)

[5] I. Fujinaga, *Adaptive Optical Music Recognition*. Ph.D. thesis, Department of Theory, Faculty of Music, McGill University, Montreal (1997)

[6] K. Ng, "Optical music analysis for printed music score and handwritten music manuscript", [w:] S.E. George (ed.) *Visual Perception of Music Notation: On-Line and Off Line Recognition*, IGI Global (2004), str.108–127

[7] W. Homenda, W. Lesinski, "Optical music recognition: case of pattern recognition with undesirable and garbage symbols", [w:] Choras, et al. (eds.) *Image Processing and Communications Challenges*, Exit, Warszawa (2009), str. 120–127.

[8] K.C., Ng, D. Cooper, E. Stefani, R.D. Boyle, N. Bailey, "Embracing the composer: optical recognition of handwritten manuscripts", *Proc. of the ICMC* (1999), str. 500–503

[9] J. Hajič jr., M. Dorfer, G. Widmer, P. Pecina, "Towards Full-Pipeline Handwritten OMR with Musical Symbol Detection by U-Nets", *ISMIR* (2018)

[10] A. Rebelo, I. Fujinaga, i in. "Optical music recognition: state-of-the-art and open issues", *Int. J. Multimedia Information Retrieval*, 1(3), (2012), str. 173–190

[11] H. Bitteur. Audiveris, https://github.com/Audiveris/audiveris.

[12] A. Pacha, Jorge Calvo-Zaragoza, Jan Hajic, "Learning notation graph construction for full-pipeline optical music recognition", *ISMIR* (2019)

[13] S. Ren, K. He, R. B. Girshick, J. Sun. "Faster R-CNN: towards real-time object detection with region proposal networks", CoRR, abs/1506.01497, 2015.

[14] L. Tuggener, Y. Satyawan, A. Pacha, J. Schmidhuber, T. Stadelmann, "The deepscoresv2 dataset and benchmark for music object detection", 25th *International Conference on Pattern Recognition (ICPR)*, (2020), str. 9188-9195

[15] J. Hajic Jr., P. Pecina, "In search of a dataset for handwritten optical music recognition: Introducing MUSCIMA++" CoRR, abs/1703.04824, 2017

[16] Ultralytics. Yolov5. https://github.com/ultralytics/yolov5

[17] A. Pacha. Munglinker. [on-line] https://github.com/OMR-Research/MungLinker, 2020.

Sesja 2

Konkurs referatów studenckich – część 1

poniedziałek, 28.06.2021
11:00 – 12:40

KLASYFIKACJA EMOCJI W MUZYCE FILMOWEJ Z WYKORZYSTANIEM TESTÓW SUBIEKTYWNYCH

T. Ciborowski[1], S. Reginis[1], A. Kurowski[1, 2], D. Weber[1, 2] i B. Kostek[1, 2]

[1]Katedra Systemów Multimedialnych, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska
[2]Laboratorium Akustyki Fonicznej, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska

s165501@student.pg.edu.pl

Streszczenie

Celem referatu było przedstawienie testów odsłuchowych, w których zadaniem osób ankietowanych było przypisanie danego fragmentu muzycznego do odpowiedniej klasy emocji. Kolejne kroki eksperymentu obejmowały wybór muzyki filmowej do testów (baza Epidemic Sound), przygotowanie założeń ankiety oraz modelu emocji wykorzystywanych w testach odsłuchowych, jak również konstrukcję ankiety. Ankieta została zrealizowana za pomocą formularzy Google. W ramach analizy wyników testów odsłuchowych w pierwszej kolejności sprawdzone zostały korelacje pomiędzy parami klas emocji przypisanych do katalogów utworów w bazie Epidemic Sound. Następnie zmapowano uzyskane wyniki ankiety na wykorzystany model, który zawiera emocje odniesienia (autorski model). W referacie zawarte też zostały histogramy zawierające przykłady list rankingowych przypisań kolorów do emocji. Przedstawiony również został zarys dalszych eksperymentów.

1. Wstęp

Muzyka filmowa stanowi ważny fragment przekazu artystycznego towarzyszącemu obrazowi. Jej celem jest podbudowanie lub wywołanie emocji związanych z daną sceną, ale przede wszystkim staje się wizytówką filmu, często określając jego gatunek. Klasyfikacja emocji w muzyce filmowej – w uproszczeniu – obejmuje zagadnienia związane z analizą muzyki, kolorystyki obrazu, narracji filmu oraz przekazu emocji, które mają towarzyszyć projekcji filmu. Celem pracy jest przedstawienie eksperymentów związanych z przeprowadzeniem testów subiektywnych określających emocje w muzyce filmowej. W pierwszej kolejności przygotowano założenia ankiety internetowej, która została następnie wykorzystana w testach subiektywnych. Ankietę skonstruowano z wykorzystaniem formularzy Google (z wykorzystaniem narzędzia "allocate monster" do dystrybucji). Opracowana ankieta jest dostępna pod adresem: https://allocate.monster/RVCTOXRP.

1.1. Modele emocji

Modele emocji wykorzystywane w algorytmach klasyfikacji można podzielić na dwie grupy: kategorialne (ang. *categorical*) i wymiarowe (ang. *dimensional*). Podejście kategorialne zakłada, że w modelu znajdują się dyskretne deskryptory emocji, które mogą być podzielone na klasy. Modele wymiarowe definiują przestrzeń emocjonalną za pomocą osi, a każda emocja traktowana jest jako punkt w tej przestrzeni [1], [2].

Do modeli kategorialnych dla przykładu można zaliczyć listę 66 deskryptorów podzielonych na 8 grup - są one wynikiem jednych z pierwszych badań nad ekspresyjnością w muzyce przeprowadzonych przez K. Hevner [3]. Innym przykładem może być model wykorzystywany w międzynarodowym konkursie MIREX, który zawiera 29 określeń podzielonych na 5 klastrów. Został on zaproponowany przez X. Hu oraz J. S. Downie [4]. Modele kategorialne badane były również w kontekście muzyki (i etykiet w j. polskim) w pracach M. Plewy oraz B. Kostek [5], [6].

Model emocji Russela [7] jest przykładem modelu wymiarowego. Zakłada on, że emocje rozmieszczone są na płaszczyźnie, która zdefiniowana jest za pomocą dwóch wymiarów - walencji oraz pobudzenia. Model emocji Thayera [8] prezentuje się w podobny sposób, ale przestrzeń dwuwymiarowa aktywacji opisana jest w nim za pomocą dwóch rodzajów pobudzeń - energetycznego oraz napięciowego.

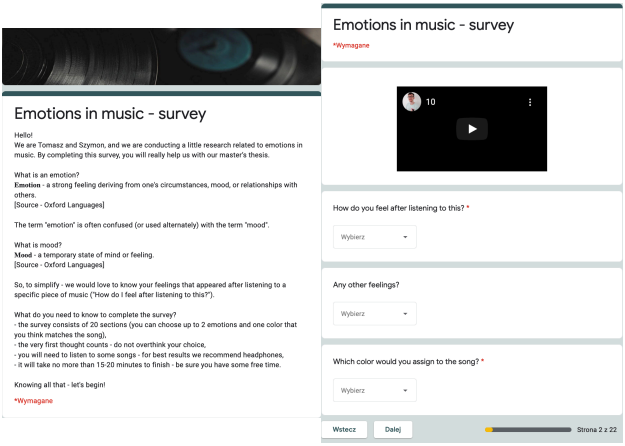
1.2. Rozpoznawanie emocji w muzyce

W literaturze można zauważyć zróżnicowane podejścia do zadania rozpoznawania emocji w muzyce. Istnieje też obszar badań (*Music Emotion Recognition*, MER) związany z klasyfikacją emocji w muzyce i nastroju, który muzyka może wywoływać [6], [9]. Wybór metody klasyfikacji jest zależny w dużej mierze od przyjętych założeń - emocje można bowiem rozpoznawać w kilkusekundowym fragmencie utworu muzycznego lub analizować go jako całość. Ważnym aspektem wymagającym uwzględnienia jest również sposób reprezentacji sygnału fonicznego, na podstawie którego opierać się będzie praca algorytmu. Utwory mogą być przedstawione w formie surowej (próbki dźwięku postaci czasowej [*wavenet*]), parametrycznej (np. wektor parametrów mel-cepstralnych czy parametrów wykorzystujących standard MPEG 7) [10], [11], [12], [13], [14], [15] lub też obecnie - w przypadku głębokiego uczenia – reprezentacji graficznej (spektrogram, chromagram) lub parametrów uzyskanych przez strukturę głęboką [16], [17], [18], [19], [20]. Kolejnym, ale bardzo istotnym elementem rozpoznawania emocji w muzyce, jest wybrany model emocji [9], [16], [21], [22]. Percepcja muzyki jest bardzo subiektywnym i indywidualnym procesem, co sprawia, że problem klasyfikacji emocji nie należy do trywialnych, zwłaszcza, że należy brać pod uwagę emocje, które zostały intencjonalnie zawarte przez kompozytora muzyki, jak również te, które towarzyszą odbiorcy muzyki i obrazu.

2. Eksperymenty

2.1. Testy subiektywne określające emocje w muzyce filmowej

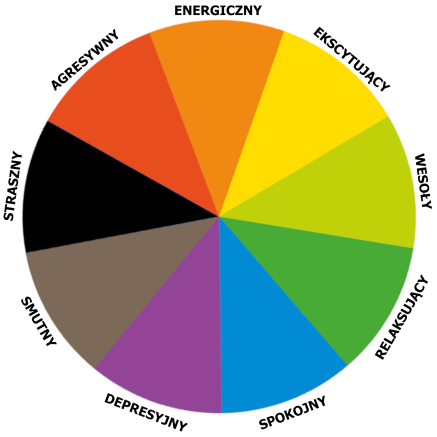
W testach subiektywnych określających emocje zawarte w danym fragmencie muzyki wzięło udział 176 osób. Zadaniem słuchaczy było przypisanie danego fragmentu muzycznego do odpowiedniej klasy emocji. Założono, że w ankiecie internetowej (rys. 1) znajdą się takie klasy, jak: *dark, epic, floating, glamorous, mysterious, hopeful, laid back, euphoric, smooth, sneaking, dreamy, senitmental, sad, romantic, happy, sexy, angry, relaxed, suspense,* zgodnie z katalogami zawartymi w bazie Epidemic Sound (epidemicsound.com). Ponieważ wykorzystywane klasy są w niektórych przypadkach bliskoznaczne, dlatego w pierwszej kolejności sprawdzono korelacje pomiędzy parami klas emocji. Niektóre przypadki na podstawie analizy korelacyjnej oraz testu chi-kwadrat (wyniki statystycznie istotne) wskazały na bliskie powiązania pojęć. Dotyczy to m.in. etykiet: *dark-mysterious, dreamy-laid back, floating-relaxed*.



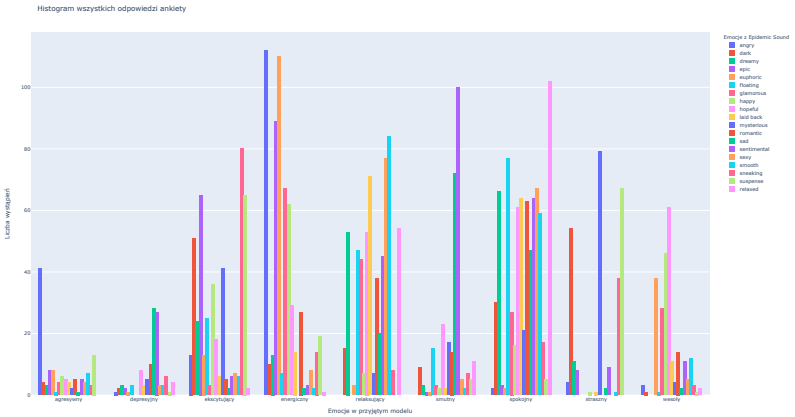
Rysunek 1 Strona wejściowa skonstruowanej ankiety internetowej oraz pytania dotyczące odsłuchiwanego fragmentu muzyki filmowej

2.2. Analiza wyników

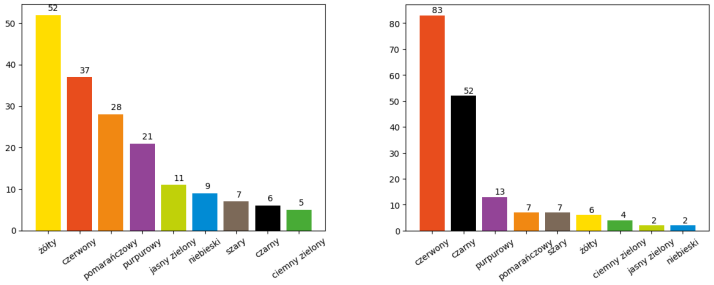
Wyniki uzyskane w ankiecie internetowej, w której zadaniem słuchaczy było przypisanie klasy emocji do danego utworu, przedstawiono na rys. 2. W dalszym kroku analiz zmapowano uzyskane wyniki na wykorzystany model, który zawiera emocje odniesienia (rys. 3): *energiczny, ekscytujący, wesoły, relaksujący, spokojny, depresyjny, smutny, straszny, agresywny*. Histogramy przedstawione na rys. 4 zawierają przykłady list rankingowych przypisań emocji do kolorów.



Rysunek 2 Model emocji wykorzystywany w pracy



Rysunek 3 Wyniki uzyskane w ankiecie internetowej, w której zadaniem słuchaczy było przypisanie klasy emocji do danego utworu (zmapowanie klas z bazy Epidemic Sound na własny model)



(a) Odpowiedzi dla *euphoric* **(b)** Odpowiedzi dla *angry*

Rysunek 4 Przykłady list rankingowych przypisania emocji do kolorów w postaci wykresów słupkowych

Na podstawie analiz histogramów można wyszczególnić pojęcia, które słuchacze wiążą ze sobą. Dla przykładu emocja *sad* budzi - w proporcji przypisań - następujące skojarzenia:

- *smutny* + *spokojny*: 39
- *depresyjny* + *smutny*: 36
- *relaksujący* + *spokojny*: 17

Z kolei, dla emocji *romantic* te typowania słuchaczy są następujące:

- *relaksujący* + *spokojny*: 34
- *smutny* + *spokojny*: 13
- *relaksujący* + *wesoły*: 13

3. Wnioski i dalsze prace

Na podstawie analiz można zauważyć, że pomimo iż występują rozbieżności w adnotacjach słuchaczy, większość typowań jest zgodna z oczekiwaniem. Głównym problemem - często wskazywanym w literaturze - jest trudność jednoznacznej definicji poszczególnych pojęć opisujących emocje. Niektóre z nich wydają się badanym osobom bliskoznaczne, zaś inne oddalone pojęciowo od siebie w sposób znaczący. Kolejnym etapem eksperymentów będzie klasyfikacja emocji z wykorzystaniem uczenia głębokiego. W tym celu zostanie wykorzystany algorytm sieci splotowej (ang. CNN – *Convolutional Neural Network*). Ze względu na wymaganie związane z głębokim uczeniem, algorytm będzie wykorzystywać reprezentacje 2D sygnałów fonicznych. W danych wejściowych zostaną uwzględnione emocje z zaproponowanego modelu (rys. 2) oraz typowania słuchaczy.

Bibliografia

[1] T. Eerola i J.K. Vuoskoski: „A comparison of the discrete and dimensional models of emotion in music”, w: *Psychology of Music* 39.1, 2011, s. 18–49, DOI: *10.1177/0305735610362821*

[2] J. Grekow: *From Content-Based Music Emotion Recognition to Emotion Maps of Musical Pieces*, Warszawa, 2017

[3] K. Hevner: „Experimental Studies of the Elements of Expression in Music”, w: *The American Journal of Psychology* 48.2, 1936, s. 246–268

[4] X. Hu i J. S. Downie: „Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata.”, w: *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007*, 2007, s. 67–72

[5] M. Plewa i B. Kostek: „Music Mood Visualization Using Self-Organizing Maps”, w: *Archives of Acoustics* 40.4, 2015, s. 513–525, DOI: *10.1515/aoa-2015-0051*

[6] B. Kostek i M. Plewa: „Rough Sets Applied to Mood of Music Recognition”, w: *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, red. M. Ganzha, L. Maciaszek i M. Paprzycki, t. 8, ACSIS, 2016, s. 71–78, DOI: *10.15439/2016F548*

[7] J. Russell: „A Circumplex Model of Affect”, w: *Journal of Personality and Social Psychology* 39, 1980, s. 1161–1178, DOI: *10.1037/h0077714*

[8] R. E. Thayer: *The Biopsychology of Mood and Arousal*, 1989

[9] M. Barthet, G. Fazekas i M. Sandler: „Music Emotion Recognition: From Content- to Context-Based Models”, w: *From Sounds to Music and Emotions. CMMR 2012. Lecture Notes in Computer Science*, red. M. Aramaki, M. Barthet, R. Kronland-Martinet i Ystad S, t. 7900, Berlin, Heidelberg: Springer, 2013, DOI: *10.1007/978-3-642-41248-6_13*

[10] B. Kostek i M. Plewa: „Parametrization and Correlation Analysis Applied to Music Mood Classification”, w: *International Journal of Computational Intelligence Studies* 2.1, 2013

[11] B. Kostek i M. Plewa: „Testing a Variety of Features for Music Mood Recognition”, w: *166th Meeting Acoustical Soc. of America* 2.1, 2013, s. 3994

[12] O. Lartillot i P. Toiviainen: „Mir in matlab (II): A toolbox for musical feature extraction from audio”, w: *Proceedings of the 8th International Society for Music Information Retrieval Conference, ISMIR 2007*, 2007, s. 127–130

[13] R. Panda, B. Rocha i R. P. Paiva: „Music emotion recognition with standard and melodic audio features”, w: *Appl. Artif. Intell.* 29, 2015, s. 313–334, DOI: *10.1080/08839514.2015.1016389*

[14] Y. Song, S. Dixon i M. Pearce: „Evaluation of musical features for emotion classification”, w: *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012*, 2012, s. 523–528

[15] F. Zhang, H. Meng i M. Li: „Emotion extraction and recognition from music”, w: *2016 12th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov. ICNC-FSKD 2016.* 2016, s. 1728–1733

[16] S. Hizlisoy, S. Yildirim i Z. Tufekci: „Music emotion recognition using convolutional long short term memory deep neural networks”, w: *Engineering Science and Technology, an International Journal* 24.3, 2021, s. 760–767, ISSN: 2215-0986, DOI: *10.1016/j.jestch.2020.10.009*

[17] Y.C. Lin, Y.H. Yang i H.H. Chen: „Exploiting Online Music Tags for Music Emotion Classification”, w: *ACM Trans. Multimedia Comput. Commun. Appl.* 7S.1, 2011, ISSN: 1551-6857, DOI: *10.1145/2037676.2037683*

[18] T. Liu, L. Han, L. Ma i D. Guo: „Audio-based deep music emotion recognition”, w: *AIP Conference Proceedings* 1967.1, 2018, DOI: *10.1063/1.5039095*

[19] R. Sarkar, S. Choudhury, S. Dutta, A. Roy i S.K. Saha: „Audio-based deep music emotion recognition”, w: *AIP Conference Proceedings* 1967.1, 2018, DOI: *10.1063/1.5039095*

[20] H. Liu, Y. Fang i Q. Huang: „Music Emotion Recognition Using a Variant of Recurrent Neural Network”, w: *Proceedings of the 2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018)*, Atlantis Press, 2019, s. 15–18, ISBN: 978-94-6252-661-7, DOI: *10.2991/mmssa-18.2019.4*

[21] M. Plewa i B. Kostek: „Multidimensional Scaling Analysis Applied to Music Mood Recognition”, w: *134 Audio Engineering Society Convention*, 2013, preprint 8876

[22] M. Plewa i B. Kostek: „Creating Mood Dictionary Associated with Music”, w: *132 Audio Engineering Society Convention*, 2012, preprint 8607

“MASZYNA KONCERTOWA” W INTERAKTYWNYCH TRANSMISJACH WYDARZEŃ MUZYCZNYCH

J. Skorupa[1], M. Głowiak[1], R. Zapała[2]

[1]Instytut Chemii Bioorganicznej PAN – Poznańskie Centrum Superkomputerowo-Sieciowe, ul Noskowskiego 12/14, 61-704 Poznań

[2]Studio Telematyczne, Akademia Muzyczna im. Ignacego Jana Paderewskiego w Poznaniu ul. Święty Marcin 87 61-808 Poznań

jskorupa@man.poznan.pl

Streszczenie

Projekt „Maszyna Koncertowa” [1] jest otwartą platformą umożliwiającą tworzenie autorskich, telematycznych wydarzeń muzycznych za pośrednictwem sieci. Narzędzie to pozwala na włączenie do klasycznej formy transmisji audio-wideo czynnej interakcji pomiędzy odbiorcą wydarzenia a jego twórcą. Odbywa się to poprzez udostępnienie uniwersalnego indywidualnie wykreowanego interface’u zbliżonego w swojej formie do prostego sieciowego kontrolera MIDI. W ramach artykułu twórcy omówią zasady działania portalu, sposoby komunikacji między z zewnętrznym oprogramowaniem, szczegóły interakcji oraz praktyczne metody zastosowania „Maszyny Koncertowej” na przykładzie koncertu inaugurującego.

1. Wstęp

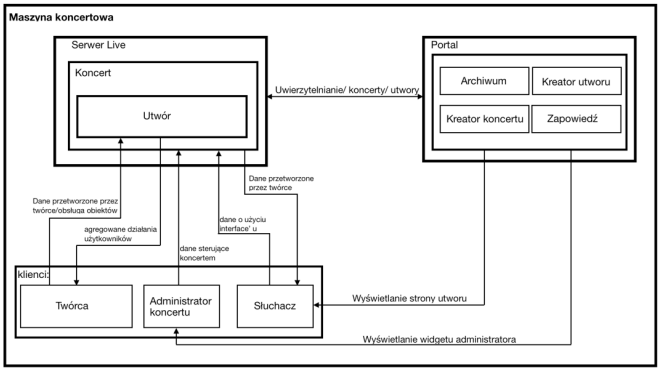
W obliczu globalnego kryzysu związanego z pandemią wirusa SARS-CoV2 obszar kultury i sztuki został szczególnie dotknięty. Wynika to przede wszystkim z faktu, iż odbieranie kultury nie jest zorientowane tylko na namacalnych produktach (tj. zakup płyt CD, książek, filmów DVD itd.). Percypowanie kultury jest doświadczeniem wieloobszarowym, często wymagającym od zainteresowanych fizycznej obecności w salach koncertowych, teatrach czy galeriach.[2]. Aspekt ten jest szczególnie istotny w przypadku funkcjonowania kultury instytucjonalnej czy wysokiej. Ogólnoświatowy *lockdown* na długi czas uniemożliwił sprawne funkcjonowanie tego sektora, zaś negatywny wpływ sytuacji pandemicznej miał szczególny wpływ na muzykę. Artyści, wykonawcy, kompozytorzy oraz instytucje kultury musieli mierzyć się z zupełnie odmienną rzeczywistością, niż ta, w której dotychczas funkcjonowali. Znane do tej pory formy prezentowania wydarzeń muzycznych okazały się bezużyteczne. Twórcy oraz instytucje chcąc dalej realizować swoje założenia zostali zmuszeni do poszukiwania alternatywnych sposobów prezentowania swojej działalności. Szczególnie popularnym zamiennikiem stały się więc cyfrowe transmisje na żywo, wirtualne galerie czy inne sposoby udostępniania zbiorów online.[2] Formy te w pierwszych miesiącach pandemii cieszyły się dużą popularnością gromadząc niejednokrotnie znacznie większą liczbę widzów niż w przed pandemicznej rzeczywistości. Niestety zamienniki te szybko przestały być wystarczające, a wręcz ich nadmiar w sieci zaczął zniechęcać potencjalnych odbiorców.

Muzyka jest dziedziną ściśle związaną z kolektywizmem czy kolaboracją. Tyczy się to zarówno relacji pomiędzy muzykami. jak również wykonawców z odbiorcami, pomiędzy którymi wcześniej zachodziła istotna interakcja wpływająca na specyficzną atmosferę w czasie koncertów czy spektakli. Transmisje online wydarzeń muzycznych pozbawiły artystów tej istotnej możliwości interakcji z publicznością. Odpowiedzią na to stała się “Maszyna koncertowa”, która powstała w wyniku refleksji nad przemianą form wydarzeń muzycznych. Aplikacja i portal stanowią interaktywną platformą pozwalającą na stworzenie sieciowego muzycznego performance’u, w którym wirtualna publiczność dołącza do wydarzenia w sposób czynny na zasadach określonych przez samego twórcę. Idea portalu odnosi się do wcześniejszych prób twórczych implementacji technologii informatycznych czy telekomunikacyjnych na gruncie artystycznym w takich projektach jak: FMOL, IMNs [3], League of Automatic Music Composers [4]. A. Barabosa w swoim artykule: *Displaced Soundscapes: A Survey of Network Systems for Music and Sonic Art Creation* przedstawia klasyfikacje oraz charakteryzuje kilka głównych sposobów wykorzystania technologii teleinformatycznych w tzw. Sztuce dźwięku (ang. *Sonic Art*) obserwowanych od początku lat 70. XX w., z których idea „Maszyny Koncertowej” jest najbardziej zbliżona do konceptu “*Musical Composition Support Systems*” [4] czyli systemu będącego rozszerzeniem tradycyjnej formy

komponowania muzyki o kolaboracje asynchroniczną wielu odbiorców (użytkowników) rozporoszonych geograficznie. Projekt „Maszyna Koncertowa" rozszerza przytoczoną definicję o pełną otwartość środowiska, w którym każdy twórca może skorzystać z sieciowych narzędzi i zbudować swój autorski *performance*.

2. Architektura portalu

Schemat architektury platformy przedstawiono na **Rys. 1. „**Maszynę Koncertową" można podzielić na dwa główne elementy. Serwer Live komunikujący się ze wszystkimi uczestnikami i agregujący dane oraz Portal umożliwiający twórcom aranżowanie i publikowanie Utworów (wraz z dedykowanymi interfejsami dla słuchaczy) i Koncertów, a widzom dostęp do nich. Obydwa elementy są ze sobą dwustronnie skomunikowane, a Serwer dodatkowo uwierzytelnia użytkowników. Portal jest zatem platformą obsługującą serwer za pomocą graficznego *interface' u*.



Rysunek 1 Schemat architektury platformy Maszyna Koncertowa

2.1. Portal

Portal jest elementem „Maszyny Koncertowej" odpowiadającym przede wszystkim za zarządzanie bazą danych użytkowników oraz obsługę wszystkich webowych usług „Maszyny Koncertowej". Zawiera takie elementy jak: kreator koncertu, kreator utworu i panele klientów. Z pozycji portalu możliwe jest używanie większości dostępnych funkcji „Maszyny Koncertowej" w łatwy i intuicyjny sposób.

2.2. Klienci

Portal Maszyna Koncertowa obsługuje trzy klasy użytkowników. Są to:

1. Słuchacze, stanowiący wirtualną publiczność koncertu, czyli klienci korzystający z portalu z pozycji użytkowników nieposiadających subskrypcji. Osoby te łączą się z Serwerem Live za pośrednictwem Portalu. Korzystają oni z webowych *interface'ów* zdefiniowanych i przygotowanych przez twórcę.
2. Twórcy, którzy definiują możliwości interakcji i przygotowują konkretny *interface* Utworu za pośrednictwem biblioteki osadzonej w środowisku *Max/MSP* [5].
3. Administrator koncertu, który tworzy Koncert ze swoich lub udostępnionych jemu Utworów i ma możliwość sterowania przebiegiem koncertu to znaczy może sterować strumieniem audio-video, rozpoczynać i kończyć Koncert oraz zmieniać bieżący Utwór.

2.3. Serwer live

Serwer live „Maszyny Koncertowej" jest najważniejszym elementem odpowiadającym za obsługę danych oraz przekazywanie ich między wszystkimi rodzajami klientów. Narzędzie odpowiada także za agregację i uśrednianie danych pochodzących od interakcji słuchaczy. Wszyscy klienci, zarówno Słuchacze, Twórcy jak i Administratorzy łączą się z serwerem przy wykorzystaniu innych narzędzi. Słuchacze oraz Administrator Koncertu wykorzystują do tego webowe aplikacje obsługiwane z pozycji Portalu natomiast Twórca wykorzystuje do tego środowisko *Max/MSP* oraz zbudowaną do tego celu bibliotekę. Każda z klas ma przydzieloną inną rangę. Najniżej w hierarchii jest Słuchacz, który może skorzystać z *interface'u* tylko poprzez używanie różnych typów interaktywnych obiektów w celu uczestnictwa w *performance'ie.* Administrator Koncertu będący wyżej w hierarchii od Słuchacza posiada dodatkowy *widget* pozwalający mu na operowanie wydarzeniami składającymi się z kilku utworów, tworzenie odnośnika do strumienia sieciowego oraz kontrolowanie przepływu danych między *Max/MSP* a Twórcą. Twórca jako jedyny ma dostęp do agregowanych oraz uśrednionych danych z aktualnie obsługiwanego *interface'u*. Ma on możliwość wpływania na ich wartości oraz wykorzystanie ich w dowolny kreatywny sposób jak np. wizualizacja uczestnictwa innych słuchaczy czy wskazanie sposobu korzystania z poszczególnych obiektów. Twórca jako jedyny ma możliwość dwustronnej komunikacji z serwerem. Oprócz przywilejów wynikających z dostępu do różnych funkcji "Maszyny Koncertowej" poszczególne klasy są dywersyfikowane pod kątem częstotliwości próbkowania przesyłania danych między klientami a serwerem. Najszybsze odświeżanie danych jest zaimplementowane w połączeniu między twórcą a serwerem natomiast najwolniejsze między słuchaczami a serwerem. Serwer zbudowany jest w taki sposób, aby w jednym momencie mogło odbywać się symultanicznie kilka niezależnych wydarzeń dostępnych pod różnymi adresami URL.

3. Opis portalu

Od strony użytkowej portal "Maszyna Koncertowa" składa się z kilku głównych narzędzi pozwalających na budowanie autorskich internetowych wydarzeń artystycznych takich jak Utwór oraz Koncert.

3.1. Utwór

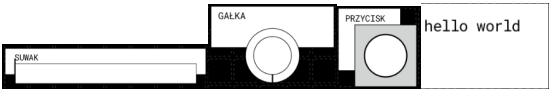
Utwór jest sieciowym *performancem* tworzonym przez zalogowanego użytkownika. Składa się z interaktywnego *interface'u* oraz opcjonalnego okna strumieniowania. W ramach Utworu Twórca określa zasady oraz wygląd swojego *performance'u* poprzez wykorzystanie interaktywnych obiektów. Dodatkowo, Twórca ma możliwość swobodnej edycji swojego Utworu za pośrednictwem edytora Utworu.

Utwór może zostać udostępniony publiczności za pomocą odpowiedniego linku URL. Publiczność uczestniczy w nim na zasadach określonych przez autora wynikających z tego, w jaki sposób została zaprogramowana aplikacja Max/MSP obsługująca *interface*. Utwór jest jednocześnie obsługiwany przez dwa rodzaje klientów: pierwszym z nich są Słuchacze, którzy mają możliwość interakcji z poszczególnymi elementami *interface'u,* drugim natomiast jest twórca, który operuje Utworem z pozycji zewnętrznej aplikacji stworzonej w *Max/MSP*. Ma on także dostęp do uśrednionych danych z poszczególnych obiektów używanych przez publiczność, przesyłania prostych komunikatów tekstowych oraz obsługiwania strumieni wideo.

3.2. Edytor Utworu

Tworzenie oraz edytowanie Utworu, odbywa się za pomocą graficznego edytora dostępnego na stronie Portalu. Przy jego pomocy użytkownik w intuicyjny sposób ma możliwość umieszczenia poszczególnych obiektów na siatce edytora, ich skalowanie oraz nadawanie nazw za pośrednictwem inspektora obiektów. Użytkownik ma do dyspozycji 6 typów obiektów takich jak: suwak, gałka, przycisk, okno wideo, obiekt tekstowy oraz okno strumieniowania. W trybie edycji Utworu Twórca może nadać tytuł oraz opis Utworu, który zostanie następnie wyświetlony Słuchaczom. Każdy Utwór ma przypisane dwa unikatowe linki. Jeden służący do udostępnienia gotowego *interface'u* dla Słuchaczy, drugi zaś pozwalający na połączenie się z pozycji oprogramowania *Max/MSP* z docelowym *interface'em*.

3.3. Charakterystyka obiektów



Rysunek 2 Obiekty interakcji dostępne w „Maszynie Koncertowej od lewej: suwak, gałka, przycisk, obiekt tekstowy

W ramach edytora Utworu Twórca ma do dyspozycji kilka typów obiektów pozwalających na interakcję pomiędzy publicznością a twórcą. Obiekty te można podzielić ze względu na sposób komunikacji. Suwak oraz gałka przedstawione na **Rys. 2** pozwalają na komunikację dwustronną, gdzie uśredniona wartość aktywności wszystkich użytkowników przekierowana jest do użytkownika *Max/MSP*. W drugą stronę użytkownik *Max/MSP* może przesłać dowolne zmienne w zakresie od 0 do 1, które są reprezentowane w formie obwiedni, mogącej służyć jako informacja o aktywności innych klientów bądź też być wskazówką określającą sposób korzystania z obiektu. Przycisk pozwala wyłącznie na komunikacje jednostronną, gdzie informacja o wszystkich kliknięciach od wszystkich klientów trafia do *Max/MSP*. Okno wideo oraz obiekt tekstowy zapewniają komunikację jednostronną od użytkownika *MaxMSP* do klientów *interface'u*. W przypadku obiektu tekstowego możliwe jest dynamiczne przesyłanie informacji tekstowej wprowadzanej przez użytkownika *Max/MSP*. Okno wideo pozwala na odtwarzanie dowolnych treści audio-wizualnych z portalu Youtube. Twórca ma możliwość dynamicznej zmiany linków w trakcie *performance'u,* jak również kontrolowanie głośności oraz momentu rozpoczęcia bądź zatrzymania projekcji. Okno strumieniowania pozwala na dołączenie do *interface'u* obszaru, w którym wyświetlana jest transmisja pochodząca z portali Youtube lub Twitch. Odmiennie do pozostałych obiektów jego obsługa odbywa się z pozycji Administratora koncertu a nie z pozycji klienta Max/MSP. Wszystkie obiekty są umieszczane w polu edytora utworu na zasadzie *drag and drop*.

3.4. Koncert

Koncert jest formą dynamicznej *playlisty* składającej się z pojedynczych Utworów. Administrator koncertu tworzy go z własnych Utworów oraz tych udostępnionych jemu przez innych użytkowników. Założeniem koncertu było stworzenie wirtualnej sali posiadającą jedną wspólną scenę dla wszystkich *performance'ów*. Po stronie Administratora koncertu możliwa jest:

1. dynamiczna zmiana poszczególnych Utworów tak aby publiczność w ramach jednego linku mogła brać udział w kilku Utworach, z drugiej strony pozwala to również na wykorzystanie kilku rodzajów interface'ów w ramach jednego *performance' u;*
2. kontrolowanie statusu przepływu danych między Twórcą a publicznością (włączanie/ wyłączanie komunikacji);
3. embedowanie bądź zmiana na żywo linku do streamingu. W przypadku koncertu we wszystkich poszczególnych utworach w obiektach okna strumieniowania, zostanie wyświetlone jedno i to samo źródło strumieniowania. Dzięki temu podczas zmiany utworu przez Administratora koncertu strona nie zostaje przeładowana, a pozycja okna streamingu jest zgodna z założeniem poszczególnych Twórców.

4. Zewnętrzna biblioteka *Max/MSP*

W celu umożliwienia jak najprostszej a jednocześnie otwartej formy komunikacji między Portalem a oprogramowaniem zewnętrznym została zaimplementowana oddzielna biblioteka dla wizualnego środowiska programistycznego *Max/MSP*. Decyzja o wykorzystaniu tego oprogramowania jako pomostu pomiędzy Twórcą a Portalem została uwarunkowana kilkoma faktami:

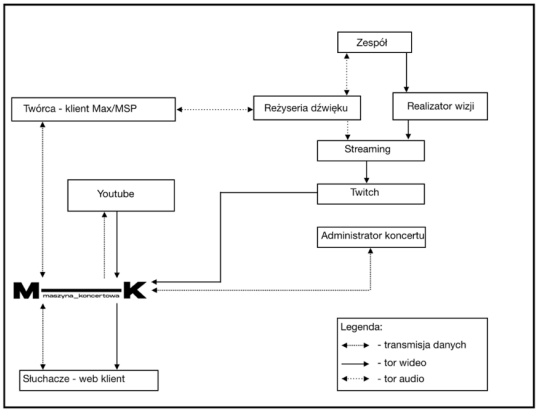
1. wersja środowiska *Max/MSP* 8.0 pozwala na implementację skryptów node.js, co znacząco ułatwia komunikację tworzonych aplikacji ze stronami WWW za pośrednictwem API;
2. max/MSP jest najpopularniejszym narzędziem z pogranicza kreatywnego programowania dedykowanego zastosowaniom multimedialnym;
3. otwartość środowiska pozwala na bardzo swobodne i autorskie wykorzystanie możliwości portalu "Maszyna Koncertowa".

Biblioteka docelowo pozwala na podłączenie się z dowolnym *interfacem* stworzonym za pośrednictwem Portalu. Biblioteka składa się z obiektu: *music-machine-client* oraz dedykowanego *help-patcha* będącego instrukcją obsługi dla twórcy. Obiekt *music-machine-client* za pośrednictwem zaimplementowanego skryptu JavaScript pozwala na połączenie z serwerem agregującym dane w portalu Maszyna Koncertowa. Komunikacja ta odbywa się za pośrednictwem protokołu *Web Socket* pozwalając na dynamiczną wymianę danych między serwerem a klientem. Cała komunikacja odbywa się w czasie rzeczywistym. Oprócz transmisji danych, *patch Max/MSP* odpowiada za translację konkretnych pakietów danych na słowniki [6] oraz utrzymuje połączenie z serwerem live.

5. Koncert inauguracyjny

Pierwszym wydarzeniem, podczas którego narzędzie zostało wykorzystane w pełnej funkcjonalności, był koncert inaugurujący publiczne otwarcie Portalu. W ramach zdarzenia piątka kompozytorów napisała 5 utworów dedykowanych oraz wykorzystujących założenia portalu "Maszyna Koncertowa". Kompozycje zostały wykonane podczas transmisji na żywo przez 8-osobowy *ensemble* (2 x skrzypce, altówka, wiolonczela, kontrabas, gitara elektryczna, saksofon, perkusja). Docelowa transmisja była realizowana za pośrednictwem portalu Twitch przy latencji ok. 4 sekund. Docelowy *set-up* koncertu został przedstawiony na **Rys. 3**.

Każdy z twórców wykorzystał możliwości portalu w odmiennej formie, co można pogrupować do trzech ogólnych kategorii sposobu interakcji słuchaczy w skomponowanych utworach:



Rysunek 3 Schemat przedstawiający *set-up* koncertu inaugurującego

5.1. Interaktywna spatializacja

Kompozytorzy Michał Janocha oraz Aleksandra Słyż zaprosili słuchaczy do budowania przestrzeni dźwiękowej transmitowanego koncertu. W przypadku M. Janochy poszczególni użytkownicy za pomocą kilku suwaków mieli możliwość dynamicznego budowania przestrzeni stereo poszczególnych ścieżek audio poprzez dowolne rozmieszczanie poszczególnych instrumentów. A. Słyż natomiast zaimplementowała w swojej kompozycji technologię dźwięku Ambisonicznego. Za pośrednictwem enkoderów B-formatu oraz dekoderów binaurialnych słuchacze mogli percypować transmisję w wirtualnej

przestrzeni 3D. Poszczególne ścieżki w transmisji permanentnie obracały się w wirtualnej przestrzeni 3D, a słuchacze za pośrednictwem *interface'u* Utworu kontrolowali poszczególne parametry encoderów Ambisonicznych takie jak: *speed* (prędkość obrotu), *spread* (szerokość źródła), *eleveation* (elewacja źródła).

5.2. Teleobecność

Kompozytor Sebastian Dembski w swoim wirtualnym *performace'ie* nawiązał do tradycji związanej ze sztuką telematyczną [7]. W pierwszej części swojego utworu zaprosił on słuchaczy do interaktywnej gry nawiązującej formą do znanego *talk show* p.t. "Millioners". W drugiej części artysta sam zaaranżował sytuację alternatywnego strumienia wideo gdzie połączył się ze słuchaczami ze swojej pracowni. Otworzył on bramkę telefoniczną, poprzez którą odbiorcy mogli rozmawiać z artystą w wirtualnym studio. Pozostali użytkownicy komunikowali się z twórcą za pomocą zbiorowego głosowania klikając w dwa przyciski zatytułowane: TAK lub NIE.

5.3. Interaktywna partia wykonawcy online

Kompozytorzy Jan Skorupa oraz Katarzyna Taborowska w swoich pracach wykorzystali czynny udział publiczności tworząc tym samym przestrzeń do kolektywnego sieciowego muzykowania za pośrednictwem specjalnie przygotowanych instrumentów. K. Taborowska w swojej kompozycji stworzyła sieciowy instrument, który pozwalał na odgrywanie konkretnych wysokości dźwięku wkomponowanych w warstwę instrumentalną, natomiast J. Skorupa zbudował fizyczną instalacje składającą się z przerobionego gramofonu. Użytkownicy w tym utworze za pośrednictwem portalu mogli spowalniać bądź przyspieszać prędkość obrotu płyty gramofonowej wpływając jednocześnie na rozstrojenie wszystkich instrumentów.

6. Podsumowanie

Za sprawą implementacji platformy „Maszyny Koncertowej" oraz otwartego dostępu do jej funkcjonalności, twórcy mają możliwość w bardzo prosty sposób, adaptacji aplikacji sieciowych angażujących interakcję zewnętrznych użytkowników w swoich autorskich *performance' ach* sieciowych. Idea „Maszyny Koncertowej" nie jest wyłącznie odpowiedzią na sytuację pandemiczną oraz wiążącym się z nią aktualnym stanem dostępności do kultury za pośrednictwem sieci. Platforma jest również formą eksperymentu z różnymi formami technologii teleinformatycznych na gruncie kompozycji muzycznej, w której odbiorcy dzieła muzycznego stają się równoprawnymi wykonawcami bądź współautorami

Wydarzenie inaugurujące obejrzało blisko 200 widzów z czego ponad 80 uczestniczyło w koncercie w sposób czynny. Twórcy oprogramowania oraz artyści związani z projektem mają zamiar w dalszym ciągu rozwijać swoje narzędzie i zainteresować nim kolejnych użytkowników. Planowane są kolejne koncerty, a doświadczenie zebrane pozwoli na rozwój możliwości prezentowania muzyki poprzez Internet i platformy streamingowe.

Bibliografia

[1] https://maszynakoncertowa.art/pl/ dostęp 07.06.21

[2] A. S. Radermeckert: *Art and culture in the COVID-19 era: for a consumer-oriented approach, 2020*

[3] *G. Weinberg: Interconected Musical Networks, w: Interconected Musical Networks – Bringing Expression and Thoughtfulness to Collaborative Group Playing*, Masschusetts, 2003

[4] A. Barabosa: *Displaced Soundscapes: A Survey of Network Systems for Musica and Sonic Art. Creation* w: Leonardo Music Journal (13), 2003, s. 53-59

[5] https://cycling74.com/get-started dostęp 07.06.2021

[6] https://docs.cycling74.com/max7/refpages/dict dostęp, 09.06.2021

[7] R. Ascott: Is *There Love in Telematic Embrace*

PORÓWNANIE DOKŁADNOŚCI POMIARÓW WYKONYWANYCH PRZEZ APLIKACJE NA SMARTFONY Z MIERNIKIEM POZIOMU DŹWIĘKU KLASY 1

E. Sieradzka[1], B. Kruk[1]

[1]Politechnika Wrocławska, Wyspiańskiego 27, 50-370 Wrocław

eliza.sieradzka.eka@gmail.com

Streszczenie

Hałas jest poważnym zagrożeniem, które towarzyszy nam na co dzień. Według WHO [1] (ang. World Health Organization), około 466 milionów ludzi posiada ubytek słuchu, a w ciągu następnych 30 lat ta liczba może ulec podwojeniu. Aplikacje mobilne mierzą poziom głośności w decybelach za pomocą wbudowanego w urządzenie mikrofonu. Dokładność pomiaru zależna jest od parametrów mikrofonu umieszczonego w smartfonie oraz od sposobu rejestracji i przetwarzania otrzymanego sygnału przez urządzenie. Celem pracy jest przeprowadzenie pomiarów za pomocą darmowych aplikacji na smartfonu oraz miernika poziomu dźwięku klasy 1, porównanie uzyskanych rezultatów oraz uzyskanie odpowiedzi czy dostępne aplikacje mobilne są wiarygodne i mogą w warunkach domowych zastępować profesjonalny sprzęt.

1. Wstęp

Działanie hałasu nie ogranicza się tylko do narządu słuchu, obejmuje także inne skutki zdrowotne. Hałas wpływa na układ krążenia, układ pokarmowy, układ nerwowy, psychikę, powoduje wcześniejsze starzenie się organizmu. Badania kliniczne potwierdzają występowanie istotnego związku między natężeniem hałasu, a ryzykiem zawału serca i nadciśnienia tętniczego.

Na rynku jest dostępne wiele aplikacji służących do pomiaru natężenia dźwięku. Wszystkie z badanych posiadały, według użytkowników, ocenę co najmniej 4 gwiazdki na 5. Jak udowodniono w trakcie pomiarów, ocena ta ma niewiele wspólnego z faktycznym działaniem aplikacji i jej dokładnością. Wszystkie smartfony biorące udział w badaniu były pozbawione dodatkowych obudów jednak nie można wykluczyć wewnętrznych, niewidocznych gołym okiem, uszkodzeń, zabrudzeń czy naleciałości na mikrofonie lub otworze w obudowie, które znacznie mogły obniżać uzyskiwane wartości.

1.1. Sposób przetwarzania sygnału przez różne modele smartfonów.

Warto rozróżnić dwa najbardziej popularne systemy operacyjne- Android i iOS. Najlepsze urządzenia z oprogramowaniem Android bazują na procesorach z serii np. Qualcomm Snapdragon, Kirin czy Exynos . Topowe produkty od Apple zawierają procesory z serii Bionic czy Fusion.

W urządzeniach z oprogramowaniem Android posiadamy różne rodzaje procesorów. W Qualcomm Snapdragon zastosowano kodek Aqstic. Pozwala ona między innymi na stosowanie funkcji audio zoom (skupianie kierunkowości mikrofonu na źródle dźwięku i niwelowanie dźwięków dookoła) czy nagrywanie dźwięków z efektem przestrzenności 5.1, dzięki zastosowaniu trzech mikrofonów. Kodek Aqstic [2] wyróżnia się wysokim zakresem dynamiki (123 dB) i niskim współczynnikiem zawartości harmonicznych z szumem (THD + N).

Urządzenia z iOS w zaawansowanych opcjach deweloperskich pozwalają na konfigurację urządzeń, np. na ustawienie częstotliwości próbkowania, czasu trwania bufora I/O czy liczby kanałów [3].

Trzeba mieć na uwadze przetworniki AC/DC, jaki jest wgrany system operacyjny smartfonu, używane oprogramowanie oraz parametry wbudowanego mikrofonu. W większości urządzeń z systemem Android posiadamy 2 wbudowane mikrofony, a w urządzeniach z systemem iOS – trzy.

Aby uzyskać wiarygodne wyniki pomiarów należy skalibrować urządzenia, dotyczy to również smartfonów. Najprostszym rozwiązaniem jest więc porównanie poziomu dźwięku z profesjonalnym miernikiem poziomu dźwięku określonej klasy. Niektóre aplikacje zawierają wartości nominalne czułości dla wybranych modeli smartfonów, dzięki którym użytkownik posiada skalę odniesienia bez konieczności użycia dodatkowego profesjonalnego sprzętu pomiarowego. W poniższej pracy smartfony nie były kalibrowane by jak najskuteczniej odzwierciedlić jak działają w funkcji miernika dźwięku dla osoby, która nie posiada możliwości wykonania profesjonalnej kalibracji.

Do badań zostały wykorzystane następujące komponenty:

- miernik poziomu dźwięku klasy 1 SVAN 955 firmy SVANTEK ,

- pojemnościowy przedwzmacniacz mikrofonowy 1⁄2” SV 12L,

- mikrofon pojemnościowy prepolarised 1⁄2” condenser ACO PACIFIC’S z czułością nominalną 22mV/Pa,

- 5 modeli telefonów z oprogramowaniem Android- Oppo A1K, OnePlus 5T, Samsung S6, Xiaomi Redmi Note 8 Pro (2 szt.), Samsung J7 (2 szt.))

- 2 modele smartfonów z oprogramowaniem iOS (iPhone 7 (2 szt.) oraz iPhone XR).

2. Pomiary

2.1. Przygotowanie do pomiarów

W celu zapewnienia powtarzalności pomiarów wszystkie sygnały testowe zostały nagrane i odtwarzane przez system elektroakustyczny. Źródłem dźwięku dla części pomiarów był zestaw głośnikowy z subwooferem marki TEAC CD-X9 o charakterystyce full range flat response, a dla kolejnych głośnik JBL Flip 4. Pomiary przeprowadzone zostały w warunkach adekwatnych dla przeciętnego użytkownika smartfonów, tzn. w domu jednak po uprzednim zaznajomieniu się z PN-EN 61672-2:2014-03/A1:2017-10 Mierniki poziomu dźwięku - Część 1 [4] oraz PN-EN 61672-2:2014-03/A1:2017-10 Mierniki poziomu dźwięku - Część 2 [5]. Źródło dźwięku było oddalone o 1 m od smartfonów.

W celu sprawdzenia jak badane aplikacje przetwarzają sygnał autorzy pracy przetestowali szerokie spektrum sygnałów pomiarowych. Analizowanymi sygnałami były:

- szum biały,

- szum różowy,

- sygnał impulsowy,

- fragment utworu muzycznego Guns N’Roses „November Rain”,

- nagrany fragment rozmowy mężczyzny z kobietą.

Przed przystąpieniem do badań każdorazowo w pierwszej kolejności trzykrotnie dokonano kalibracji miernika za pomocą dwuzakresowego kalibratora akustycznego klasy 1 SV30 [6] o wartości 93,5 dB.

Przebadane zostały następujące aplikacje:

Aplikacja nr 1 - Miernik dźwięku autorstwa EXA Tools [7]

Aplikacja nr 2 - Miernik dźwięku autorstwa Splend Apps [8]

Aplikacja nr 3 - Sound Meter 2020 autorstwa HMXS Apps [9]

Aplikacja nr 4 - Miernik dźwięku autorstwa KTW Apps [10]

Aplikacja nr 5 - Sound Meter – Decibel Meter! autorstwa 黎丽陈 [11]

Aplikacja nr 6- Decibel X autorstwa SkyPaw Co. Ltd [12]

W tabeli 1 zostały przedstawione wyniki dla aplikacji nr 1-3. Aplikacja prawdopodobnie dokonują pomiarów z krzywą korekcyjną Z oraz stałej czasowej slow ponieważ uzyskane wartości dla telefonu i miernika są najbardziej zbliżone. Na podstawie uzyskanych rezultatów pierwszej aplikacji można zaobserwować, że działa najlepiej dla pomiarów sygnałów szumowych, a najgorzej dla sygnałów impulsowych lub rozmowy.

Tabela 1 Zestawienie wyników uzyskanych w aplikacji nr 1-3 dla smartfonów z wynikami z miernika klasy 1.

model telefonu	OnePlus			Oppo			Samsung			iPhone	Miernik
aplikacja	1	2	3	1	2	3	1	2	3	1	
Wartość minimalna [dB(Z)]											
Szum biały	56	79	64	67	72	57	45	79	67	39	76,9
Szum różowy	51	75	44	65	42	13	43	55	67	70	76,6
Sygnał impulsowy	40	50	30	18	32	3	33	42	39	35	59,2
Muzyka	67	54	57	61	32	52	67	42	65	44	72,4
Rozmowa	53	49	42	17	34	17	50	42	63	51	67,9
Wartość maksymalna [dB(Z)]											
Szum biały	79	83	70	72	79	64	78	85	60	78	78,3
Szum różowy	79	81	71	70	76	64	79	89	59	74	80,7
Sygnał impulsowy	65	78	66	52	67	68	62	87	19	70	66,3
Muzyka	81	84	70	73	79	67	82	81	58	80	79,4
Rozmowa	81	83	71	76	84	70	85	87	56	83	86,4
Wartość średnia [dB(Z)]											
Szum biały	75	82	67	69	75	70	74	82	66	69	77,6
Szum różowy	73	83	67	68	72	70	71	81	67	78	77,2
Sygnał impulsowy	49	59	39	24	41	63	42	78	35	48	62,8
Muzyka	75	78	65	66	72	72	74	53	64	72	76,3
Rozmowa	73	78	63	64	73	75	72	77	62	70	79,6

W drugiej części badań sprawdzano powtarzalność pomiarów dla smartfonów tego samego modelu oraz porównywano uzyskane wyniki do wartości otrzymanych z miernika klasy 1.

Tabela 2 Zestawienie wyników powtarzalności tego samego modelu smartfonów.

	Xiaomi 1	Xiaomi 2	Samsung 1	Samsung 2	iPhone 7 1	iPhone 7 2	Miernik
	Wartość minimalna [dB(Z)]						
Szum biały	71	70	78	58	69	80	76,9
Szum różowy	70	68	79	79	72	47	76,6
Sygnał impulsowy	32	33	36	34	44	71	59,2
Muzyka	33	45	35	55	68	71	72,4
Rozmowa	42	43	45	45	50	69	67,9
	Wartość maksymalna [dB(Z)]						
Szum biały	73	73	81	81	76	81	78,3
Szum różowy	72	71	81	81	75	76	80,7
Sygnał impulsowy	59	58	87	67	75	74	66,3
Muzyka	73	73	81	83	76	74	79,4
Rozmowa	81	78	85	85	82	75	86,4
	Wartość średnia [dB(Z)]						
Szum biały	72	72	80	79	72	72	77,6
Szum różowy	71	70	80	80	73	71	77,2
Sygnał impulsowy	40	39	45	43	55	73	62,8
Muzyka	65	67	74	78	71	73	76,3
Rozmowa	69	68	75	74	72	72	79,6

3. Zestawienie wyników

3.1. Klasyfikacja według aplikacji

Analizując uzyskane wartości dla aplikacji 1 jasno widać, że aplikacja najlepiej radzi sobie z wartościami maksymalnymi. Najdokładniejsze wyniki uzyskano z telefonu OnePlus, najmniej dokładne dla telefonu Oppo. Aplikacja najgorzej bada sygnały impulsowe.

Tabela 3 Zestawienie wyników dla aplikacji nr 1 – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	Wartość maksymalna [dB(Z)]				Wartość średnia [dB(Z)]			
	OnePlus	Oppo	Samsung	iPhone	OnePlus	Oppo	Samsung	iPhone
Szum biały	0,7	6,3	0,3	0,3	2,6	8,6	3,6	8,6
Szum różowy	1,7	10,7	1,7	6,7	4,2	9,2	6,2	0,8
Sygnał impulsowy	1,3	14,3	4,3	3,7	13,8	38,8	20,8	14,8
Muzyka	1,6	6,4	2,6	0,6	1,3	10,3	2,3	4,3
Rozmowa	5,4	10,4	1,4	3,4	6,6	15,6	7,6	9,6

Aplikacja nr 2 najdokładniej wskazuje poziom dźwięku dla sygnału rozmowy, a najmniej dokładnie dla sygnału impulsowego. Analizując uzyskane wyniki można uznać ją jednak za aplikację generalnie mało wiarygodną, niezależnie od użytego urządzenia.

Tabela 4 Zestawienie wyników dla aplikacji nr 2 – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	Wartość maksymalna [dB(Z)]			Wartość średnia [dB(Z)]		
	OnePlus	Oppo	Samsung	OnePlus	Oppo	Samsung
Szum biały	4,7	0,7	6,7	4,4	2,6	4,4
Szum różowy	0,3	4,7	8,3	5,8	5,2	3,8
Sygnał impulsowy	11,7	0,7	20,7	3,8	21,8	15,2
Muzyka	4,6	0,4	1,6	1,7	4,3	23,3
Rozmowa	3,4	2,4	0,6	1,6	6,6	2,6

Aplikacja nr 3 wskazuje wartości dość dokładne (maksymalnie 3 dB różnicy) dla smartfonów OnePlus i Samsung. Gorzej radzi sobie z wartościami średnimi. Smartfon Oppo bardzo niedokładnie mierzy żądane parametry.

Tabela 5 Zestawienie wyników dla aplikacji nr 3 – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	Wartość maksymalna [dB(A)]			Wartość średnia [dB(A)]		
	OnePlus	Oppo	Samsung	OnePlus	Oppo	Samsung
Szum biały	0,5	6,5	0,5	2,5	9,5	3,5
Szum różowy	1,3	8,3	2,3	4,8	12,8	4,8
Sygnał impulsowy	1,7	3,7	1,3	13,3	33,3	17,3
Muzyka	3,1	6,1	1,1	3,4	10,4	4,4
Rozmowa	4,7	5,7	0,7	4,4	11,4	5,4

Aplikacja nr 4 najgorzej wskazała pomiary dla sygnału impulsowego, z rozmową też nie radzi sobie najlepiej. Jako jedyna, ze stuprocentową dokładnością, wskazała wynik dla szumu białego smartfonem OnePlus. Ponownie widać, że najgorsze wartości uzyskiwała dla smartfonu Oppo.

Tabela 6 Zestawienie wyników dla aplikacji nr 4 – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	Wartość maksymalna [dB(C)]			Wartość średnia [dB(C)]		
	OnePlus	Oppo	Samsung	OnePlus	Oppo	Samsung
Szum biały	0	5,1	6,4	1,1	7,7	0,6
Szum różowy	1,6	6	5,1	0,7	7,8	0,2
Sygnał impulsowy	12,2	12,3	2,4	6,8	20,1	12,5
Muzyka	0,7	0,2	3,7	1	9,3	1,6
Rozmowa	7,7	6,2	4,3	5,8	12,2	6,9

Aplikacja nr 5 dość skutecznie mierzy wartości maksymalne, niedokładnie jednak wartości średnie. Najlepiej radzi sobie z szumami, najgorzej z sygnałem impulsowym.

Tabela 7 Zestawienie wyników dla aplikacji nr 5 – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	Wartość maksymalna [dB(C)]	Wartość średnia [dB(C)]
Szum biały	0,2	2,2
Szum różowy	0,9	0,7
Sygnał impulsowy	11,8	3,9
Muzyka	0,2	8,4
Rozmowa	3,8	9,3

3.2. Klasyfikacja według smartfonu

Analizując pozyskane wartości dla smartfonu OnePlus zauważono, że aplikacja nr 1 najlepiej radziła sobie z wartościami maksymalnymi, a aplikacja nr 4 z wartościami średnimi.

Każdy program najmniej dokładnie dokonywał pomiaru sygnału impulsowego, różnica między wskazaniami wynosiła jednak maksymalnie 13,8dB.

Tabela 8 Zestawienie wyników dla smartfonu OnePlus – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	OnePlus							
	Aplikacja 1		Aplikacja 2		Aplikacja 3		Aplikacja 4	
	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(A)]	Wartość średnia [dB(A)]	Wartość maksymalna [dB(C)]	Wartość średnia [dB(C)]
Szum biały	0,7	2,6	4,7	4,4	0,5	2,5	0	1,1
Szum różowy	1,7	4,2	0,3	5,8	1,3	4,8	1,6	0,7
Sygnał impulsowy	1,3	13,8	11,7	3,8	1,7	13,3	12,2	6,8
Muzyka	1,6	1,3	4,6	1,7	3,1	3,4	0,7	1
Rozmowa	5,4	6,6	3,4	1,6	4,7	4,4	7,7	5,8

Smartfon Oppo najczęściej wskazywał największe błędy, łatwo zauważyć że mało dokładnie mierzył wartości średnie. Najgorzej radzi sobie z sygnałem impulsowym błędy są naprawdę znaczące. Dobre rezultaty uzyskał jedynie dla wartości maksymalnych drugiej aplikacji.

Tabela 9 Zestawienie wyników dla smartfonu Oppo – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	Oppo							
	Aplikacja 1		Aplikacja 2		Aplikacja 3		Aplikacja 4	
	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(A)]	Wartość średnia [dB(A)]	Wartość maksymalna [dB(C)]	Wartość średnia [dB(C)]
Szum biały	6,3	8,6	0,7	2,6	6,5	9,5	5,1	7,7
Szum różowy	10,7	9,2	4,7	5,2	8,3	12,8	6	7,8
Sygnał impulsowy	14,3	38,8	0,7	21,8	3,7	33,3	12,3	20,1
Muzyka	6,4	10,3	0,4	4,3	6,1	10,4	0,2	9,3
Rozmowa	10,4	15,6	2,4	6,6	5,7	11,4	6,2	12,2

Smartfon firmy Samsung najmniej dokładnie mierzy wartości poziomu sygnału impulsowego, słabo radzi sobie również z szumem białym i szumem różowym. Dobrą dokładność osiągnął dla wartości maksymalnych przy użyciu aplikacji nr 3.

Tabela 10 Zestawienie wyników dla smartfonu Samsung – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

Samsung							
Aplikacja 1		Aplikacja 2		Aplikacja 3		Aplikacja 4	
Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(A)]	Wartość średnia [dB(A)]	Wartość maksymalna [dB(C)]	Wartość średnia [dB(C)]
0,3	3,6	6,7	4,4	0,5	3,5	6,4	0,6
1,7	6,2	8,3	3,8	2,3	4,8	5,1	0,2
4,3	20,8	20,7	15,2	1,3	17,3	2,4	12,5
2,6	2,3	1,6	23,3	1,1	4,4	3,7	1,6
1,4	7,6	0,6	2,6	0,7	5,4	4,3	6,9

Smartfon marki Apple dokonuje mało dokładnych pomiarów poziomu dźwięku. Najgorzej radzi sobie z sygnałem impulsowym, zadziwiająco niedokładnie mierzy również wartości maksymalne. Pomimo niewielkich różnic dla wartości poziomu szumu białego, z dużym błędem dokonano pomiaru wartości średniej. Żadnego z badanych sygnałów nie można uznać za wiarygodnie zmierzony.

Tabela 11 Zestawienie wyników dla smartfonu iPhone – błąd bezwzględny pomiarów smartfonem w odniesieniu do miernika klasy 1.

	iPhone			
	Aplikacja 1		Aplikacja 5	
	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]	Wartość maksymalna [dB(Z)]	Wartość średnia [dB(Z)]
Szum biały	0,3	8,6	0,2	2,2
Szum różowy	6,7	0,8	0,9	0,7
Sygnał impulsowy	3,7	14,8	11,8	3,9
Muzyka	0,6	4,3	0,2	8,4
Rozmowa	3,4	9,6	3,8	9,3

4. Podsumowanie i wnioski

Podczas przeprowadzania badań i analizy wyników należy mieć na uwadze błąd użytkownika. Część aplikacji działała z opóźnieniem, możliwe więc było włączenie pomiaru zbyt wcześnie lub zbyt późno. Należy zwrócić uwagę również na mechaniczne bariery przeprowadzania pomiarów. Wszystkie smartfony biorące udział w badaniu były pozbawione dodatkowych obudów jednak nie można wykluczyć wewnętrznych, niewidocznych gołym okiem, uszkodzeń, zabrudzeń czy naleciałości na mikrofonie lub otworze w obudowie, które znacznie mogły obniżać uzyskiwane wartości.

Podsumowując aplikację nr 1, zauważono że dość trafnie pokazuje ona wartości maksymalne dla telefonów z oprogramowaniem Android, niestety zaniża wartości minimalne. Odczyt wartości średniej pomiaru jest dość dokładny. W przypadku telefonów z oprogramowaniem iOS aplikacja lepiej radzi sobie z wartościami średnimi, wartości minimalne są znacznie zaniżane. Różnice w poziomach odczytanych z aplikacji a miernika są jednak na tyle wysokie, że nie pozwalają uznać by aplikacja mogła zostać określona wiarygodną w odczytywaniu poziomów natężenia dźwięku.

Aplikacja nr 2 z większą dokładnością mierzy poziom minimalny i maksymalny, gorzej radzi sobie z pomiarem wartości średniej pomiaru. Niestety aplikacja nie jest wiarygodna szczególnie, że w części pomiarów zaniżała wynik, w innych go zawyżała.

Aplikacja nr 3 zauważalnie zaniża wyniki, dyskwalifikując część wartości z dalszej analizy. Dla osoby, która ma niewielkie pojęcie o akustyce może to być mylące. Niestety spore zakłamanie w odczycie wartości minimalnych dyskwalifikuje tą aplikację jako wiarygodną. Wartości maksymalne zostały zmierzone dość dokładnie.

Aplikacja nr 4 dla smartfonu OnePlus jest zdecydowanie najlepsza spośród reszty propozycji. Niestety nawet ona pokazuje wartości odbiegające o ponad 12dB od wyników uzyskanych za pomocą miernika

Aplikacja nr 5 zawyża uzyskane dane w odniesieniu do wyników z miernika klasy 1, a różnice przekraczają nawet 10dB co pozwala stwierdzić iż aplikacja jest niedokładna.

Wszystkie aplikacje najmniej dokładnie dokonują pomiarów sygnału impulsowego, co wynika ze zbyt niskiego próbkowania co doprowadzania do znacznych błędów przy dalszym przetwarzaniu próbek.

W drugiej części pomiarów dokonano porównania dwóch tych samych urządzeń tego samego modelu. W analizie wyników zauważono, że urządzenia z oprogramowaniem Android dokonują zbliżonych pomiarów (maksymalnie ±4dB różnicy między urządzeniami). Niestety urządzenia z oprogramowaniem iOS posiadały między sobą większe różnice.

Podsumowując przebieg badań, wybór aplikacji jest bardzo ważny jednak każda aplikacja działa zgoła inaczej dla każdego modelu telefonu. Pomijając błędy grube, większość aplikacji dokonywała pomiarów z dokładnością około ±10dB, niestety część telefonów, np. Samsung S6 zawyżała wyniki, a smartfon Oppo A1K raz zawyżał, innym razem zaniżał wartości. Część aplikacji pozwalała użytkownikowi na ręczną kalibrację czy wybór krzywej korekcyjnej z jaką dokonuje ona pomiarów. W celu kalibracji należało wtedy uruchomić pomiar testowy w cichym pomieszczeniu jednak aby ujednolicić wyniki pomiarów, krok ten pominięto.

W aspekcie ochrony słuchu lepiej oczywiście gdy wyniki aplikacji są zawyżone aniżeli zaniżone jednak gdy chcielibyśmy bardziej wiarygodne wyniki, nie zaleca się używanie aplikacji mobilnych, a profesjonalnego sprzętu. Niektórzy producenci programów w opisie informują użytkownika iż są to jedynie narzędzia poglądowe i nie powinno się zanadto sugerować ich wskazaniami, co udowodniono w powyższej pracy.

Sytuacją idealną byłoby gdyby aplikacje samodzielnie dobierały odpowiednią krzywą korekcyjną i stałą czasową na podstawie rodzaju badanego sygnału. Niemożliwe jest jednak przyrównanie

otrzymanych wartości do pomiarów z jedną określoną krzywą korekcyjną ponieważ twórcy aplikacji nie podają publicznie do wiadomości z jaką korekcją ich aplikacje dokonują badań.

Bibliografia

[1] World Health Organization, „https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.,” 1 marzec 2020. [Online].

[2] Qualcomm, „https://www.qualcomm.com/products/features/aqstic.,” listopad 2020. [Online].

[3] Apple, listopad 2020. [Online].

[4] Norma, „PN-EN 61672-2:2014-03/A1:2017-10 Mierniki poziomu dźwięku --Część 1.

[5] Norma, PN-EN 61672-2:2014-03/A1:2017-10 Mierniki poziomu dźwięku - Część 2.

[6] Svantek, „http://svantek.com/lang- pl/product/17/sv_30a_kalibrator_akustyczny_klasy_1.html,” listopad 2020. [Online].

[7] E. Mobile, „https://examobile.pl/,” wrzesień 2020. [Online].

[8] S. Apps, „http://splendapps.com/,” lipiec 2020. [Online].

[9] H. Apps, „https://www.facebook.com/hmxsapps/,” wrzesień 2020. [Online].

[10] K. Apps, „https://ktwapps-ad94a.firebaseapp.com/,” wrzesień 2020. [Online].

[11] 黎丽陈„,https://appgrooves.com/publisher/431471807/%E9%BB%8E%E4%B8%BD-%E9%99%88.,” październik 2020. [Online].

[12] S. C. Ltd, „https://www.skypaw.com/,” wrzesień 2020. [Online].

ROZMAWIAJĄC O DŹWIĘKU – BADANIE ŁĄCZĄCE OPISY BARWY DŹWIĘKU W JĘZYKU POLSKIM Z PARAMETRYCZNĄ ANALIZĄ SYGNAŁU

M. Markowski[1], U. Kaleta[1], J. Jasiński[2] i B. Głowacki[2]

[1] Uniwersytet Jagielloński w Krakowie
[2]Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie

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 wykorzystywany do opisu barwy dźwięku w języku polskim. Drugi eksperyment polegał na prezentowaniu słuchaczom dźwięków i proszeniu ich o umiejscowieniu 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] [3]. 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, a ich osią są eksperymenty nad wyznaczaniem przestrzeni semantycznej barwy dźwięku, zróżnicowanych pod względem języka oraz metodologii [4] [5] [6].

Wielu badaczy wzięło sobie za cel próbę zmapowania przestrzeni barwy dźwięku. Powszechnie stosowaną metodą, skuteczną w tworzeniu i eksploracji przestrzeni semantycznej barwy dźwięku, jest *semantic differential* [7] [4], której wynikiem jest określenie zredukowanej ilości osi tej przestrzeni. Eksperymenty przy użyciu tej metody przeprowadzane były m.in. w języku niemieckim, gdzie uzyskano osie *full-empty, dull-sharp, colorful-colorless* oraz *compact-diffuse* [4], w języku angielskim gdzie zadeklarowano istnienie trzech wymiarów *bright-dull, warm-cold oraz rich-pure* [8] lub *texture, luminence i mass* [5] i w języku czeskim gdzie otrzymano osie *bright/clear-gloomy/dark, hard/sharp-delicate/soft, wide-narrow* oraz *hot/hearty* [6].

Celem niniejszego badania było wykorzystanie metody semantic differential w celu rozszerzenia obecnego stanu wiedzy na temat percepcji barwy o język polski oraz sprawdzenie czy wyniki i wnioski wyciągane we wcześniejszych badaniach są również adekwatne. Postawione zostały hipotezy o istnieniu skończonej liczby osi, czyli zbioru określeń semantycznych, które tworzą przestrzeń barwową, 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 poprzez internet co oznacza, że każdy uczestnik odsłuchiwał dźwięków 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 pomiędzy różnymi słuchaczami. Ponadto odtwarzane dźwięki stanowiły jedynie zachętę, zaś przedmiotem tego eksperymentu były same słowa, których używali badani.

Badanie zostało przeprowadzone na 60 osobach, spośród których: 47 w wieku 18-34 lat, a 13 powyżej 34; 27 kobiet, 32 mężczyzn i 1 osoba określająca swoją płeć jako pozostałe; 24 osoby bez doświadczenia muzycznego, 31 amatorów i 5 profesjonalistów.

Wyniki ankiety poddane zostały filtracji, w wyniku której odrzucone zostały słowa nie odnoszące się do barwy dźwięków. Były to określenia subiektywnej oceny przyjemności dźwięków, opisy instrumentów grających, terminy opisujące głośność lub długość trwania oraz określenia emocji wzbudzanych u słuchaczy. Najczęściej występujące słowa opisujące dźwięk, które pozostały zostały przeanalizowane i wybrane tak, aby utworzyć osie antonimiczne. Uwzględnione zostało również występowania określeń zbliżonych znaczeniem [9] w poprzednich badaniach [3]. Wybrane zostały podane poniżej pary słów, w nawiasach znajduje się liczba wystąpień danego słowa w trakcie badania.

ciemny (6) – jasny (21), ostry (28) – łagodny (17), miękki (16) – twardy (5), ciepły (28) – zimny (5), syntetyczny (13) – naturalny (8), dźwięczny (18) – głuchy (8), zabrudzony (3) – czysty (19), wybrzmiewający (18) – stłumiony (14).

2.2. Wybór dźwięków

Dźwięki wykorzystane do badania zostały ostrożnie dobrane na podstawie zbiorów instrumentów występujących we wcześniejszych badaniach [5] jak i również różnorodności pomiędzy wartościami parametrów akustycznych dla nich wyznaczonych. Jako podstawową wysokość dźwięków wybrany został ton c^1 (261,63 Hz) [10] oraz wszystkie dźwięki zostały znormalizowane do poziomu -24,5 LUFS. 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 polski. Uczestnicy nie zgłosili żadnych ubytków w słuchu. Spośród nich: 43 w wieku 18-24, a 31 w wieku 25-34 lat; 39 kobiet, 33 mężczyzn i 2 osoby określające swoją płeć jako pozostałe; 41 bez doświadczenia muzycznego, 25 amatorów i 8 profesjonalistów.

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 do uczestników. Uczestnicy proszeni byli o złożenie stanowiska w cichym miejscu oraz otworzenie programu, który został specjalnie stworzony do przeprowadzania badań. Po wypełnieniu formularza 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 ustawieniu na skalach pomiędzy dwoma określeniami semantycznymi suwaka tak, aby odzwierciedlał on sposób w jaki opisaliby barwę słuchanego dźwięku. W celu uniknięcia zaburzenia wyników badania w związku ze zmęczeniem i adaptacją słuchu [11], kolejność dźwięków, kolejność skal semantycznych oraz ich kierunek był generowany losowo dla każdego uczestnika. Badanie zajmowało przeciętnie 15 minut.

4. Analiza wyników badań

4.1. Analiza rzetelności

W trakcie prowadzenia badań wiele uczestników zgłaszało brak zrozumienia skali *naturalny – syntetyczny*. Te uwagi, wraz z wynikami analizy wstępnej spowodowały, że w dalszej analizie skala ta nie będzie brana do pod uwagę.

Na potrzeby prezentacji wyników skale będą nazywane przez jeden z antonimicznych członów. Wyniki dla poszczególnych skal semantycznych zostały poddane analizie rzetelności poprzez wyznaczenie współczynnika Alfy Cronbacha. Wyniki przedstawione są w tabeli 1.

Tabela 1 Wyniki analizy rzetelności wyników

Cecha	$CORR_1$ [a]	α [b]
Jasny	0,13	0,89
Czysty	0,46	0,98
Zimny	0,14	0,91
Głuchy	0,33	0,96
Łagodny	0,50	0,98
Twardy	0,39	0,97
Stłumiony	0,33	0,96

[a] Średnia korelacja pomiędzy pozycjami skali semantycznej
[b] Alfa Cronbacha

Dla każdej skali $\alpha > 0{,}8$ co określa spójność uzyskanych odpowiedzi jako dopuszczalną. W celu usunięciu wartości odstających 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 skosność i kurtoza przyjmują wartości bezwzględne mniejsze niż 4. Oznacza to, że pomimo negatywnego wyniku testu normalności Lillieforda i testu Shapiro-Wilka 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 2.

Można zaobserwować:

- Bardzo wysoką korelację ($0{,}7 < |r| > 0{,}9$) pomiędzy skalami
  - ***głuchy – dźwięczny*** $\leftrightarrow$ ***stłumiony – wybrzmiewający***
  - ***ostry – łagodny*** $\leftrightarrow$ ***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.

Tabela 2 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

Celem analizy czynnikowej z metodą składowych głównych jest wyznaczenie zredukowanej ilości nieskorelowanych czynników, które wyjaśniają zaobserwowaną wariancję. Ograniczenie modelu do dwóch czynników umożliwia wyjaśnienie 68% wariancji, do trzech czynników umożliwia wyjaśnienie 81%, natomiast do czterech czynników do 88%. W związku z takimi wynikami podjęta została decyzja o przyjęciu trzech czynników jako optymalnej konfiguracji pomiędzy przejrzystością modelu, a ilością wyjaśnionej wariancji. W tabeli 3 widoczny jest wkład poszczególnych osi semantycznych na czynniki po zastosowaniu rotacji Varimax z przyjętą minimalną wartością własną $\lambda = 0{,}82$.

Tabela 3 Wkład osi semantycznych do wyznaczonych czynników

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 skalę *zimny – ciepły, ostry – łagodny, twardy – miękki, czysty – zabrudzony*, na drugi skalę *głuchy – dźwięczny, stłumiony – wybrzmiewający* natomiast na czynnik trzeci skala *jasny – ciemny*. Po analizie powyższych wyników utworzone zostały oznaczenia czynników jako kolejno **Faktura/Temperatura, Wydźwięk, Jasność**. Nie jest możliwe idealne oddanie charakteru czynników poprzez pojedyncze słowa, zatem nazwy te należy traktować jako aproksymacje ułatwiające dalszą dyskusję i rozważania.

4.3. Analiza korelacji parametrów fizycznych i skal semantycznych

Wykorzystane w badaniu dźwięki poddane zostały analizie parametrycznej. Zbiór parametrów wybrany został poprzez wykorzystanie tych, które we wcześniejszych badaniach określone zostały jako mające największy wpływ na percepcję barwy [3] [12]. Zbiór parametrów składa się z: *Spectrum Centroid [13], Spectrum Spread [13], High Energy-Low Energy Ratio [13], Tristimulus [13], Inharmonicity [13], Noisiness [13], Odd-Even Ratio [13], Zero Crossing Rate [3], Entropy [3]*. W tabeli 4 zaprezentowane są wyniki wartości współczynnika korelacji z poszczególnymi czynnikami przestrzeni semantycznej.

Powyższe wyniki pokazują, że z czynnikiem faktury/temperatury 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.

Tabela 4 Korelacje parametrów fizycznych z czynnikami przestrzeni semantycznej

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

4.4. Analiza regresji

Wykonana została analiza regresji krokowej wstecznej (poziom istotności $p < 0{,}05$), której wyniki zaprezentowane są w tabeli 5. Do analizy wykorzystane zostały wszystkie parametry fizyczne, jednak konieczne było odrzucenie parametrów Tristimulus 1 i Zero-Crossing Rate w związku z ich bardzo dużą współliniowością z pozostałymi parametrami.

Tabela 5 Analiza regresji

	Współczynniki Beta		
Parametr fizyczny	Fakt./Temp.	Wydźwięk	Jasność
Spectrum Centroid	-0,33	-0,51	-
HE–LE Ratio	-0,43	-	-
Tristimulus 3	-	0,54	-
Inharmonicity	-	-0,27	-
Noisiness	0,77	0,21	0,58
Odd-Even Ratio	-	-0,36	-0,22
Entropy	-0,41	-	-0,37

Czynnik	R^{2a}	F^b	p^c
Fakt./Temp.	0,31	27,9[d]	$< 0{,}001$
Wydźwięk	0,33	30,5[e]	$< 0{,}001$
Jasność	0,15	11,2[f]	$< 0{,}001$

[a] Współczynnik determinacji
[b] Test Fishera-Snedecora
[c] Poziom istotności
[d] $F(4{,}440)$
[e] $F(5{,}440)$
[f] $F(3{,}440)$

Na podstawie powyższych wartości można wnioskować, że czynnik faktury/temperatury można wyjaśnić w 31% poprzez parametry fizyczne. 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% poprzez dodatni wpływ parametrów Tristimulus 3 i Noisiness oraz ujemny wpływ Spectrum Centroid, Odd-Even Ratio i Inharmonicity. Czynnik jasności można wyjaśnić w 15% poprzez dodatni wpływ Noisiness i ujemny wpływ Entropy i Odd-Even Ratio.

5. Dyskusja i Wnioski

Celem pracy było wykonanie wstępnego badania mającego na celu odkrycie struktury opisującej semantyczny sposób opisywania barwy dźwięków. Uzyskane w badaniu odpowiedzi mogą wskazywać na to, że istnieją określenia semantyczne w języku polskim, wykorzystywanych do opisu barwy dźwięków, który jest spójnie rozumiany i powszechnie stosowany. Analiza rzetelności uzyskanych odpowiedzi wskazuję na ich wysoki poziom precyzyjności i spójności, a jej porównanie do innych prac w dziedzinie pozwala wnioskować, że wpływ zdalnego stanowiska odsłuchowego nie powinien być znaczący na uzyskane wyniki.

W wyniku analizy czynnikowej wyznaczone zostały trzy składowe semantycznej przestrzeni opisu barwy, którym przypisano nazwy: jasność, faktura/temperatura, wydźwięk. Wyjaśniają one 82% zaobserwowanej wariancji. Wiele prac, które rozważało tematykę opisu barwy w różnych językach określało trzy albo cztery osie jako wystarczające do modelowania semantycznej przestrzeni reprezentującej percepcję barwy i niniejsze badanie potwierdziło te odkrycia. Dodatkowo uzyskane w języku polskim osie zbliżone są do tych, które określone zostały dla języka angielskiego i greckiego w postaci Luminance, Texture i Mass [5] [12] chociaż wyraźnie widać, że pewne koncepty nie są możliwe do przetłumaczenia pomiędzy językami. Wyniki te umacniają jednak koncepcję interlingwistycznej zgodności semantycznych sposobów opisu barwy dźwięku [12].

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 dziedzinie. Jednym z aspektów, który może mieć na to wpływ to brak heterogeniczność 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 ilość parametrów fizycznych oraz analiza wyników ze względu na poziom edukacji muzycznej ankietowanych.

Bibliografia

[1] ANSI: *Psychoacoustical Terminology*, New York: American National Standards Institute, 1973

[2] S. Donnadieu: *Mental Representation of the Timbre of Complex Sounds*, Springer, New York, NY, 2007, rozd. In: Beauchamp J.W. (eds) Analysis, Synthesis, and Perception of Musical Sounds. Modern Acoustics and Signal Processing. DOI: *https://doi.org/10.1007/978-0-387-32576-7_8*

[3] Vinoo Alluri i Petri Toiviainen: „Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre", w: *Music Perception: An Interdisciplinary Journal* 27 (3), 2010, s. 223–242

[4] G. von Bismarck: „Timbre of steady sounds: A factorial investigation of its verbal attributes", w: *Acta Acustica united with Acustica* 30.(3), 1974, s. 146–159

[5] Asterios Zacharakis, Konstantinos Pastiadis i Joshua D. Reiss: „An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates", w: *Music Perception* 31.4, 2012, s. 339–358, DOI: *10.1525/mp.2014.31.4.339*

[6] O. Moravec i J. Štěpánek: „Verbal description of musical sound timbre in Czech language", w: *Proceedings of the Stockholm Music Acoustics Conference (SMAC03)*, 2003, s. 643–645

[7] C.E. Osgood, G.J. Suci i P.H. Tannenbaum: *The measurement of meaning*, University of Illinois Press, Urbana, 1957

[8] R. Pratt i P. Doak: „A subjective rating scale for timbre", w: *Journal of Sound and Vibration* 45, 1976, s. 317–328

[9] Krzysztof Szlifirski: *Pro-Audio - Słownik terminologii nagrań dźwiękowych*, Audiologos, 2008

[10] Marek Pluta: *Zasady muzyki i notacji muzycznej*, Wydawnictwa AGH, 2012

[11] Piotr Kleczkowski: *Percepcja dźwięku*, Wydawnictwa AGH, 2013

[12] Asterios Zacharakis, Konstantinos Pastiadis i Joshua D. Reiss: „An Interlanguage Unification of Musical Timbre", w: *Music Perception* 32.4, 2015, s. 394–412, DOI: *10.1525/mp.2015.32.4.394*

[13] Geoffroy Peeters, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis i Stephen McAdams: „The Timbre Toolbox: Extracting audio descriptors from musical signals", w: *The Journal of the Acoustical Society of America* 130.5, 2011, s. 2902–2916, DOI: *10.1121/1.3642604*

BETTER

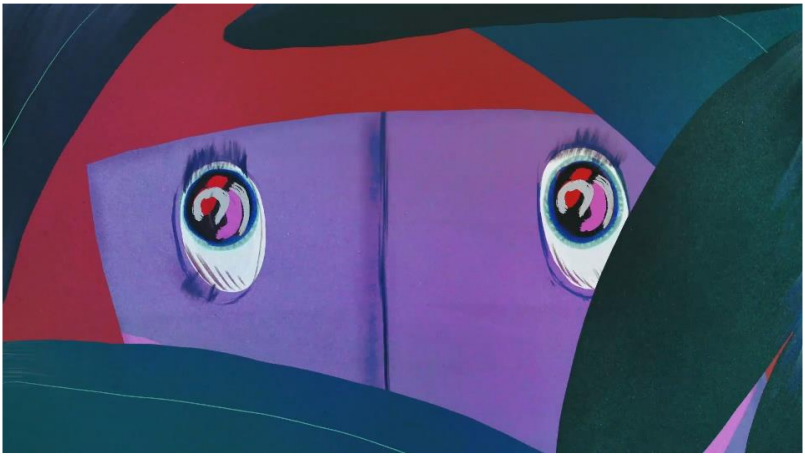
J. Zajączkowski[1]

[1] Uniwersytet Muzyczny Fryderyka Chopina

zajac19pl@gmail.com

Praca zgłoszona na konkurs nagrań.

1. Materiał źródłowy



Rysunek 1 Klatka z filmu „Better”

„Better” to film animowany w reżyserii Emily Downe, który został stworzony na potrzeby projektu „Life Rewired”. Projekt ten skupia się na znaczeniu człowieczeństwa w czasach intensywnego rozwoju technologii.

2. Pomysł i historia

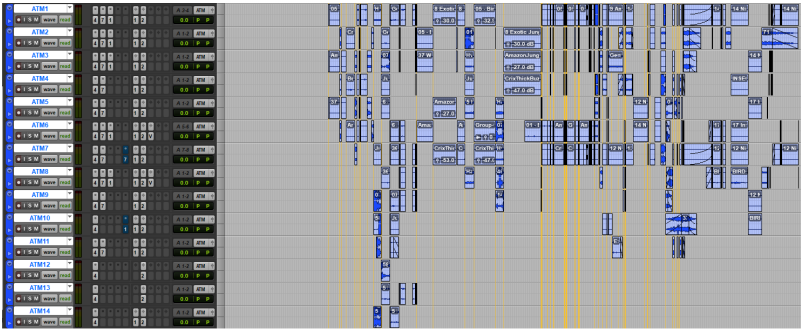
Pierwszy raz natknąłem się na tą animację przeglądając Vimeo Staff Pick. Film od razu przykuł moją uwagę. Zakochałem się w rysunkowym stylu warstwy wizualnej, która w mojej opinii stwarza ogromne możliwość do interpretacji w ścieżce dźwiękowej, dodatkowo, fabularnie animacja również pozwala na dużą swobodę w kreacji artystycznej. Moja interpretacja skupia się na dezintegracji fikcyjnej dżungli symbolizującej bałagan współczesnego świata oraz na problemach psychicznych z którymi mierzy się bohaterka przedstawiona w roli narratora.

Moim pomysłem, którym kierowałem się przy tworzeniu warstwy dźwiękowej było stworzenie dyskretnego kontrastu między brudnym i technologicznie nasyconym światem zewnętrznym a spokojnym i cichym światem wewnętrznym.

3. Proces kreacji dźwięku

Swoją pracę zacząłem od warstw atmosfer, dzięki którym dżungla ożyje. Mając na uwadze kontrast, który chcę osiągnąć, starałem się aby brzmienie dżungli było spokojne i kojące. Z drugiej strony, przebłyski świata zewnętrznego musiały wytworzyć impakt na słuchaczu – nie było to proste, gdyż pojawiają się one najczęściej na kilka klatek obrazu. W większości przypadków wykorzystywałem nagrania różnych, zaludnionych miejsc w Polsce, nakładałem je na siebie a potem mixowałem z różnymi industrialnymi

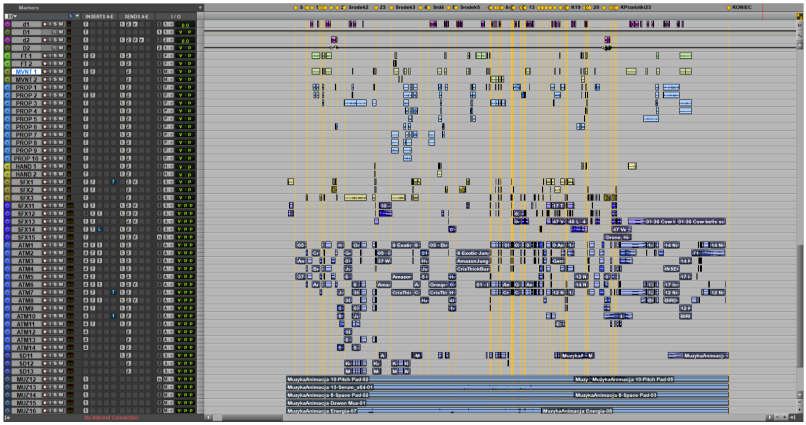
brzmieniami dla lepszego efektu dźwiękowego. Najbardziej ciekawa część prac natomiast rozpoczęła się kiedy dwa światy animacji zaczęły się ze sobą łączyć, co doprowadziło do wykorzystania przeze mnie dużej ilości fade'ów i crossfade'ów, różnych pogłosów i ich automatyzacji, zmiany wysokości dźwięku i wielu innych technik sound-design'owych.



Rysunek 2 Ścieżki atmosfer

Końcowa ścieżka dźwiękowa składa się z:

- nagrań plenerowych wykonanych przeze mnie na rejestratorze Zoom H2N
- ścieżki narratorki nagranej na mikrofon Neumann TLM 103, oraz innych dialogów tła
- dźwięków i atmosfer pochodzących z bibliotek, często w dużym stopniu zmodyfikowanych przeze mnie
- efektów synchronicznych nagranych na uczelni na mikrofon Neumann KMR 81 i
- ścieżki muzycznej i sound design'owej skomponowanej i wyprodukowana przeze mnie

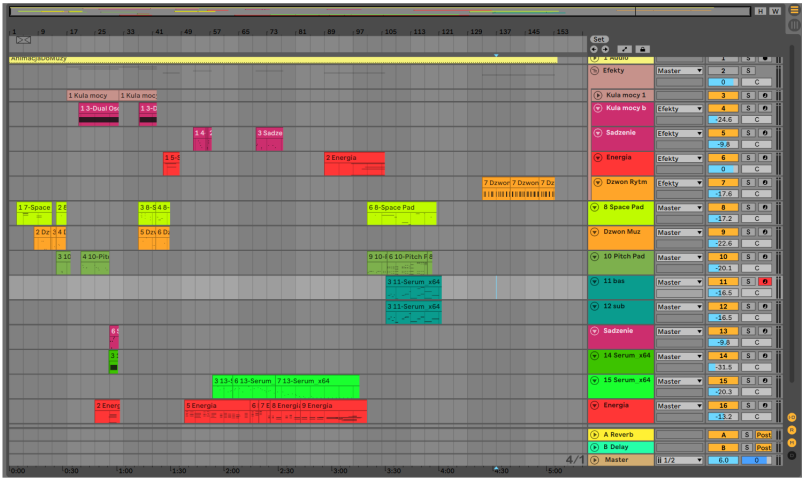


Rysunek 3 Prawie cała sesja

4. Muzyka i Sound-design

Pomimo traktowania całej ścieżki dźwiękowej jako artystycznej kompozycji, ścieżka muzyczna którą stworzyłem dla animacji była dla mnie niezwykle istotna. Do procesu jej tworzenia wykorzystałem DAW Ableton Live w którym utworzyłem osobną sesję stworzoną do pracy na wielu różnych instrumentach MIDI z wykorzystaniem dwóch kontrolerów: M-Audio Code 61 oraz Novation Launchpad Mini. Mój ulubiony wirtualny syntezator, którego używałem najczęściej w tej pracy to XFER SERUM. Każda ścieżka

instrumentalna została osobno przeze mnie opracowana za pomocą różnych efektów i przeniesiona do sesji w Pro Tools, w którym następnie dokonałem mixu tych śladów z resztą ścieżki dźwiękowej. W Ableton Live powstało również wiele dźwięków z dziedziny sound-design'u które znalazły swoje miejsce w finalnej wersji animacji.



Rysunek 4 Sesja w Ableton Live

5. Mix

Podczas mixu ważne dla mnie było utrzymanie klimatu intymnej opowieści, ale również zachowanie dynamicznego brzmienia świata zewnętrznego. Kluczowe było dla mnie aby wyraźnie doprowadzić widza do punktu kulminacyjnego, a potem, gdy bohaterka godzi się ze światem zewnętrznym, odpowiednio uspokoić ścieżkę dźwiękową. Do monitorowania poziomu głośności wykorzystałem wtyczkę MeldaProduction MLoudnessAnalyzer, a większość mixu została wykonana na uczelni z pomocą Avid S6 Control Surface na platformie Pro Tools HD 12 i odsłuchu Genelec 8050.

Różne techniki sound-design'owe, w głównej mierze wykorzystujące equalizacje i filtry zostały zastosowane w ujęciach w których dżungla jest zamykana przez magiczną bańkę świata zewnętrznego, co pozwoliło na uzyskanie ciemnego i brudnego brzmienia. Jedną z moich ulubionych scen jest moment w którym bańka się domyka – zautomatyzowanie EQ oraz wtyczek pogłosowych dało możliwość uzyskania efektu zwiększającej się pustki. Dodatkowo, w tym fragmencie, tak jak w wielu innych pojawią się duża ilość modulacji wysokości dźwięku.

Kolejna scena którą uwielbiam, to punkt kulminacyjny. Aby osiągnąć w niej maksymalną dramaturgię zastosowałem w niej przester, saturacje oraz emulację taśmy, a efekty te zostały oczywiście również zautomatyzowane.

Głos narratorki i dodatkowe dialogi zostały odpowiedni opracowane przeze mnie w iZotope RX7 Audio Suite.

Podczas końcowego zgrania wykorzystałem wtyczki:

- L2, L3 Ultra, RCompressor, RDeEsser, LinMB
- EQ3-7Band, EQ3-1Band, Q10, Q4,
- Altiverb 7, RVerb
- LoFi, Pitch II, Kramer Tape, MannyM Distortion, MDMX Overdrive, Time Shift, Time Compression Expansion, Speakerphone
- iZotope RX7

Każdy mikrofon w szkolnym studio został nagrany bezpośrednio poprzez Avid HD Omni interface.

Sesja, jak również wszystkie pliki audio wykorzystywane były w formacie 48 KHz, 24 bit. Głośność ogólna została zmierzona przez wtyczkę MLoudnessAnalyzer z wynikiem na poziomie -18 LUFS i True Peak -1.09dB.

Plik wideo z którym pracowałem utrzymany był w formacie 25FPS 1920x1080 DNxHD.

Sesja 2

Konkurs referatów studenckich – część 2

poniedziałek, 28.06.2021
13:20 – 14:40

ODWZOROWANIE BRZMIENIA UTWORU MUZYCZNEGO WYKONANEGO ZA POMOCĄ APARATURY ANALOGOWEJ, PRZY UŻYCIU CYFROWYCH NARZĘDZI DŹWIĘKOWYCH

B. Czubak[1]

[1] Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

bartekczubak@o2.pl

Praca zgłoszona na konkurs nagrań.

1. Cel pracy

Celem pracy było odwzorowanie nagrania muzycznego, wykonanego w latach sześćdziesiątych za pomocą aparatury analogowej, przy użyciu dzisiejszego sprzętu studyjnego ze szczególnym uwzględnieniem cyfrowych narzędzi, które emulują brzmienie legendarnych, używanych w studiu Abbey Road, analogowych urządzeń audio. W ramach badań wykonano wielośladowe nagranie utworu Roll Over Beethoven zespołu The Beatles, pochodzącego z początku lat sześćdziesiątych. W badaniach starannie uwzględniono dawne techniki rejestracji instrumentów oraz sposób miksu materiału, który wykonano w programie do wielościeżkowej edycji dźwięku.

2. Wnioski

Wykonanie nagrań muzycznych oraz ich analiza posłużyły do sprawdzenia czy istnieje możliwość, aby we współczesnych warunkach, osiągnąć charakterystyczne i niepowtarzalne brzmienie, słyszane na archiwalnych nagraniach z przeszłości. Ponadto podjęte działania umożliwią określenie czy i jakie korzyści niosła za sobą, spowodowana rozwojem technologii, modyfikacja technik rejestracji instrumentów muzycznych.

3. Warunki techniczne

Sesja nagraniowa została przeprowadzona 5 grudnia 2020 roku, w warszawskim studio muzycznym „W Dobrym Tonie Studio". Zdecydowano się na wybór powyższego miejsca ze względu na duży (576 m3), żywy akustycznie live room, który miał imitować warunki akustyczne, jakie panowały w Studio Two – Abbey Road. Wszystkie partie zostały nagrane „na setkę" w jednym pomieszczeniu. Zarejestrowanego materiału, nie poddano żadnym zabiegom edycji.

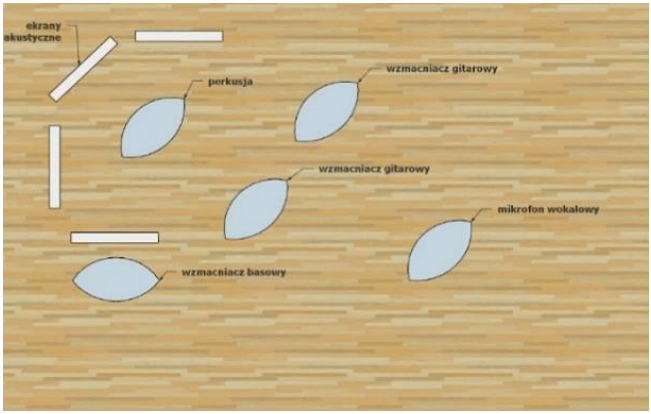
Do nagrania użyto wielokanałowej konsoli cyfrowej Soundcraft Ui24, z możliwością pracy jako wielośladowy interfejs audio. Mikser został połączony z komputerem – stacją roboczą DELL Precision M4800, kablem USB A-B. Rejestracja materiału została wykonana za pomocą programu DAW – Steinberg Cubase 5.

4. Rejestracja

Proces rejestracji utworu przebiegał w następujący sposób:

1) Rejestracja materiału „na setkę": perkusja, bas, gitary, wokal.
2) Rejestracja oklasków metodą superimpositon.
3) Ponowna rejestracja wokalu, metodą superimposition.
4) Double-tracking solowej partii gitarowej metodą superimpostion.

Podczas nagrywania zadbano o charakterystyczne dla wczesnych nagrań zespołu The Beatles, rozmieszczenie muzyków w studiu. Usytuowanie poszczególnych instrumentów w studio muzycznym pokazano na rysunku 1.



Rysunek 1 Rozmieszczenie instrumentów podczas nagrań utworu

W celu nagrania perkusji zdecydowano się wykorzystać dwa główne mikrofony – Shure Super 55, jako mikrofon dynamiczny o charakterystyce kardioidalnej do bębna basowego oraz mikrofon wstęgowy o charakterystyce ósemkowej Golden Age Project R1, jako overhead. Ten minimalistyczny setup, znajdujący się na rysunku 2 miał na celu odwzorowanie typowego dla 1963 roku ustawienia mikrofonów perkusyjnych, kolejno: STC 4033-A – bęben basowy oraz wstęgowego STC 4038 – overhead. W poniższej tabeli 1, znajduje się lista wejściowa instrumentów z dedykowanymi mikrofonami, podczas nagrań utworu.

Tabela 1 Lista instrumentów podczas nagrań pierwszego utworu

Kanał	Instrument	Źródło	Mikrofon	Uwagi
1	Bęben basowy	-	Shure Super 55	
2	Bęben basowy	-	AKG D112	Użyty jako backup
3	Overhead	-	Golden Age Project R1	
4	Gitara basowa	Fender	Shure Sm57	
5	Gitara basowa	Fender	Audix D6	Użyty jako backup
6	Gitara rytmiczna	VOX AC30	Rode NT1A	
7	Gitara solowa	VOX AC15	Studio Projects C4	
8	Gitara solowa	DAW – AMP	Sygnał liniowy	Użyty jako backup
9	Wokal	-	Rode NT2A	

Wzmacniacz basowy został nagrany za pomocą dwóch mikrofonów – Shure Sm57 oraz Audix D6, pokazanych na rysunku 3. Ze względu na bardziej zbliżoną charakterystykę mikrofonu Shure Sm57 do charakterystyki STC 4033-A, używanego masowo w latach sześćdziesiątych przez The Beatles, zdecydowano się na użycie popularnego Sm57.

Do nagrań gitar zdecydowano się na użycie dwóch mikrofonów pojemnościowych, wielkomembranowego Rode NT1A oraz małomembranowego Studio Projects C4, ukazanego na rysunku 4. Ta konfiguracja miała z kolei odwzorować często stosowane w 1963 roku ustawienie Neumann KM54 dla jednego wzmacniacza oraz Neumann U48 dla drugiego.

Do nagrań wokalu, aby odwzorować mikrofon Neumann U48, wykorzystano mikrofon o zmiennej charakterystyce kierunkowej Rode NT2A. Ze względu, że w nagrywanym utworze jest jeden, dominujący wokal, mikrofon ustawiono w charakterystyce kardioidalnej. Podobnie jak w 1963, nie użyto pop-filtra.

Rysunek 2 Omikrofonowanie perkusji podczas rejestracji utworu

Kolejnym krokiem była rejestracja oklasków (rysunek 3). W badanym przypadku zastosowano ponownie mikrofon Rode NT2A, tym razem ustawiony w charakterystyce ósemkowej. Kluczowym elementem tego etapu nagrań było wykorzystanie szerokopasmowej kolumny jako odsłuchu. Jej lokalizacja została wybrana dokładnie tak, jak w roku 1963 – w niewielkiej odległości od muzyków, po boku, na osi najmniejszej skuteczności mikrofonu. Opisywana metoda nagrywania nosi nazwę superimposition. Ten sposób pozwolił nagrać oklaski z możliwie małym poziomem tła – podkładu.

Rysunek 3 Nagrywanie oklasków, Rode NT2A w charakterystyce ósemkowej jako Neumann U48

Ostatnim elementem rejestracji danego utworu była ponowna rejestracja głównej partii wokalu oraz gitary solowej, w momencie partii solowej. Proces zastosowania metody superimposition, ukazano na rysunku 4.

Rysunek 4 Wykorzystanie metody superimposition do ponownego nagrania wokalu

Miks zarejestrowanego materiału, ze względu na próbę dokładnego odtworzenia brzmienia i toru audio, był bardzo wymagającym przedsięwzięciem. Ten etap pracy odbył się w dwóch lokalizacjach – pierwszej, obejmującej zasadniczy miks materiału w domowym studiu, wyposażonym w profesjonalną adaptację akustyczną, monitory

odsłuchowe EVE Audio SC205, słuchawki referencyjne Beyerdynamic DT770 Pro; drugi, obejmujący finalny odsłuch i mastering, w krakowskim studiu Kotłownia – Studio Muzyczne AGH. Aby uzyskać satysfakcjonujące efekty brzmieniowe, korzystano ze standardowych suwaków głośności, potencjometrów panoramy oraz z następujących, cyfrowych wtyczek VST:

• Waves REDD37-51, jako odwzorowanie toru wejściowego konsoli EMI REDD.37-51;

• Softube RS127 Box i Rack, jako odwzorowanie korektora EMI RS127;

• Waves RS56, jako odwzorowanie korektora EMI RS56;

• Softube RS124, jako odwzorowanie kompresora EMI RS124;

• Waves J37, jako odwzorowanie rejestratora taśmowego Studer J37;

• Waves Abbey Road Chambers, jako odwzorowanie komór pogłosowych, między innymi Chamber Two;

• Waves Abbey Road Vinyl, jako odwzorowanie finalnego procesu utrwalania nagrania na nośniku analogowym;

5. Miks

Podczas tworzenia miksu wzorowano się na autentycznym zgraniu utworu, pochodzącego z oryginalnej płyty winylowej „With The Beatles" z 1963 roku.

W ramach miksu przyjęto następującą strategię:

- jeśli barwa instrumentu wymagała ingerencji, pierwszym efektem w łańcuchu był korektor RS127 Box lub RS127 Rack,

- na każdej rejestrowanej w pierwszym podejściu ścieżce jako insert wpinano następnie wtyczkę REDD 37-51, będącej symulacją „wejścia" sygnału do preampu konsoli. W tym momencie możliwa była kolejna, dokładniejsza korekcja barwy,
- utworzono dwie grupy „main fader1" i „main fader2", które odwzorowywały rozmieszczenie instrumentów na kanałach konsoli REDD. Partie rytmiczne skierowano na lewą, partie wokalne i solowe – na prawą. W grupy wpięto kolejno RS124 oraz J37. To moment, który imitował w rzeczywistości zgranie pierwszego podejścia na rejestator dwuścieżkowy, po wcześniejszej kompresji dwóch kanałów,
- następnie utworzono dwie kolejne grupy „track1" i „track2", na które oprócz grup „main fader1" i „main fader2", wysłano zdublowane partie. Na grupach zapięto kolejno REDD37-51, RS124, J37. W ten sposób ścieżki pierwotne, tak jak w rzeczywistości, były więc ponownie przepuszczane przez konsolę, a ścieżki dublowane – „doklejane" do wcześniejszego podkładu,
- późniejszym krokiem był właściwy mix materiału, który polegał na skierowaniu grup „track1" i „track2", do finalnych grup „mainL" i „mainR", na których zainsertowano następujące wtyczki REDD37-51, RS124 i J37. W tym momencie zastosowano również wysyłkę do Abbey Road Chambers, z których powrót kierowany był na poszczególne grupy „mainL" i „mainR",
- finalny proces obejmował ewentualną korekcję nagrania przy użyciu Waves RS56 oraz utrwalenie go na płycie winylowej, przy pomocy Waves Abbey Road Vinyl.

6. Uwagi

Należy zaznaczyć, że elementy nagrania podlegające ocenie w konkursie, między innymi takie jak zakres dynamiczny, elementy techniczne (SNR, obecność zakłóceń), przestrzeń i panoramowanie itd. zostały osiągnięte w sposób naturalny, tj. poprzez dokładne odwzorowanie toru audio i wykorzystanie emulacji rzeczywistych urządzeń hardware'owych ze studia Abbey Road z lat 60.

PROBLEMY GŁOSU WOKALNEGO W PROCESIE TRANSKRYPCJI

A. Konieczna

Akademia Górniczo – Hutnicza w Krakowie

angelikakonieczna97@gmail.com

Streszczenie

Głównym zadaniem transkrypcji głosu wokalnego jest uzyskanie z sygnału informacji o czasie trwania poszczególnych dźwięków oraz ich częstotliwości podstawowej. Następnie informacje te zostają przekształcone na odpowiadającą im wysokość muzyczną i wartość rytmiczną. Problemy głosu wokalnego, występujące w procesie transkrypcji, można podzielić na te związane z dziedziną czasu (wyznaczenie momentu rozpoczęcia i zakończenia dźwięku, między którymi można przypisać mu określoną wysokość muzyczną, wahania tempa utworu) i wysokości (detekcja F0, utrzymanie wysokości dźwięku, błędy oktawowe, problemy z intonacją). Niektóre z trudności następują też w związku z obecnością tekstu melodii lub w przypadku wybranych artykulacji. Praca skupia się na przedstawieniu oraz analizie problemów, co stanowi podstawę zarówno do tworzenia nowych rozwiązań w metodach segmentacji, jak i do poprawy jakości wyników samej transkrypcji.

1. Wstęp

Transkrypcja głosu wokalnego jest szczególnym przypadkiem transkrypcji muzycznej. Posiada ona wiele cech zarówno indywidualnych dla wokalisty (predyspozycje wokalne, barwa głosu), jak i zależnych od niego w trakcie śpiewu (dynamika, sposób artykulacji, stosowanie ozdobników czy interpretacja melodii). Dążenie do jak najlepszych wrażeń słuchowych, za pomocą wymienionych środków, jest w środowisku muzycznym pożądane, jednak w przypadku transkrypcji może okazać się ono sporym utrudnieniem ze względu na swoją nieprzewidywalność. Warto również nadmienić, że osiągane przez głos wokalny wysokości muzyczne dźwięków lub ich wartości rytmiczne mogą często różnić się od oczekiwanych lub tych zawartych w zapisie nutowym. Analiza przyczyny tego i innych problemów głosu wokalnego w procesie transkrypcji jest przedmiotem poniższej pracy. Dodatkowo uwzględniono również wpływ omawianych sytuacji na poszczególne etapy transkrypcji, takie jak segmentacja nagrania, analiza wysokościowa lub zapis wyników.

2. Wybrane etapy transkrypcji

Transkrypcja muzyczna jest złożeniem procesów analizujących sygnał audio, których celem jest uzyskanie informacji muzycznych zawartych w nagraniu. W przypadku transkrypcji głosu wokalnego za kluczowe uznaje się wydobycie informacji o wysokościach muzycznych, wartościach rytmicznych oraz kolejności występowania dźwięków zaśpiewanej melodii. Wartość rytmiczna jest wyznaczana na podstawie czasu trwania poszczególnych dźwięków oraz wartości tempa utworu. Z kolei na wysokość muzyczną wskazuje wartość częstotliwości podstawowej (F0) otrzymywanej w procesie analizy widma częstotliwościowego. Ostatnim etapem transkrypcji jest analiza muzyczna oraz jej zapis np. w postaci graficznego zapisu nutowego.

Najczęściej spotykaną formą transkrypcji wśród muzyków jest zapis rozpoznanych struktur melodii na podstawie jej nagrania lub zapamiętanej formy. Podczas pierwszych odtworzeń ocenie poddawany jest charakter melodii, określa się ilość taktów, najniższy i najwyższy dźwięk oraz rozpoznane grupy rytmiczne. Melodia zostaje odtwarzana lub powtarzana głosem do momentu, w którym uzyskany zapis nutowy uznany zostanie za kompletny lub wystarczający dla muzyka. Zaletą tej metody jest dowolność powtarzania procesów analizy, wybranych fragmentów melodii lub uproszczanie jej formy, jednak wyniki transkrypcji są mocno uzależnione od umiejętności osoby, która jej dokonuje.

Początkowym, a zarazem kluczowym etapem transkrypcji automatycznej jest segmentacja nagrania ze względu na wartości rytmiczne pojawiające się w melodii. Określenie momentów rozpoczynania

i kończenia się poszczególnych dźwięków stanowi podstawę dla analizy wysokościowej, której zadaniem jest przypisanie wysokości muzycznej wskazanym fragmentom sygnału. Skuteczność metod użytych do wspomnianych procesów przekłada się na otrzymane informacje o nutach, a tym samym na stopień odwzorowania ostatecznych wyników transkrypcji i satysfakcję użytkownika. Istnieje wiele rodzajów segmentacji, jednak większość z nich można zakwalifikować do tych pracujących w dziedzinie czasu lub częstotliwości. Trudno jest również wskazać taką metodę segmentacji, która byłaby odporna na większość problemów głosu wokalnego. Ich wpływ na wybrane aspekty segmentacji zostanie przytoczony przy okazji ich omawiania.

3. Problemy głosu wokalnego w procesie transkrypcji

W środowisku muzycznym stosowanie różnorodnych artykulacji, obecność tekstu lub ozdobników melodii pomaga wokalistom w przekazywaniu emocji oraz własnej interpretacji utworu, a także wzbogaca wrażenia odsłuchowe słuchacza. Jednak w przypadku głosu wokalnego wspomniane procesy mogą okazać się trudne do skategoryzowania i doprowadzać do błędnie rozpoznanej wartości rytmicznej lub wysokości dźwięku. Warto również rozróżnić warunki nagrywania głosu wokalnego poddawanego późniejszej transkrypcji. Nagrania dostarczone z występów na żywo mogą okazać się trudniejsze do analizy ze względu na przywiązywanie większej uwagi do wykonania utworu, niż do wiernego odwzorowania zapisu nutowego. Dodatkowo mogą one zawierać dźwięki niepożądane lub pochodzące od innych źródeł ze sceny (np. instrumentów lub innych wokalistów). Podczas rejestracji głosu wokalnego w warunkach kontrolowanych możliwe jest znaczne uproszczanie melodii, wielokrotne powtórzenia wymagających tego fragmentów lub zamiana tekstu na sylabę, co może znacznie ułatwić analizę w dalszych procesach transkrypcji.

Różnorodność wykonania oraz interpretacji motywu są zagadnieniami, które również należy rozpatrzeć przy ocenie uzyskanych wyników transkrypcji. Są one silnie związane z doświadczeniem i umiejętnościami muzyka, jednak nie zawsze należy traktować je w kategorii błędu. Badania przeprowadzone przez Honinga [1] wskazują, że użycie innej struktury rytmicznej lub ich rozpoznawanie nie zawsze powinny być z nim utożsamiane. Zależą one od pamięci muzycznej oraz dopasowania struktur melodii do tych, które są najbliższe wykonawcy. Niektóre parametry, takie jak określenie tonacji, mogą również zostać odebrane w sposób indywidualny przez muzyków. Skutkiem tego jest np. otrzymanie transkrypcji w różnych tonacjach, jednak przy zachowaniu właściwych dla danej melodii odległości między dźwiękami, przeniesienie jej do oryginalnej tonacji nie powinno stanowić dodatkowej trudności.

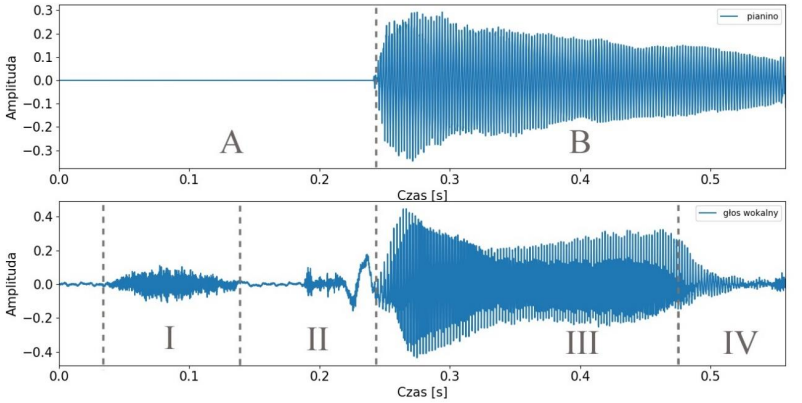
W celu wstępnego pogrupowania problemów wykorzystano podział, który zakłada w pierwszej kolejności analizę amplitudy, a następnie częstotliwości zawartych w sygnale. Stanowią one podstawę do dalszych procesów, takich jak grupowanie w kontekście zapisu nutowego [2]. Trudnym zadaniem jest jednoznaczna klasyfikacja wybranych problemów głosu wokalnego do jednej z dziedzin – czasu lub częstotliwości – dlatego są one omawiane przy okazji tej, na którą silniej wpływają.

3.1. Obwiednia amplitudowa i czas trwania dźwięku

Przebieg czasowy sygnału jest najprostszą formą jego prezentacji. Określenie dynamiki zmian pozwala na wstępne określenie rodzaju sygnału (bez konieczności dodatkowej analizy widmowej) oraz sposobu artykulacji. Dodatkowo wyróżniane są poszczególne fazy obwiedni amplitudowej, takie jak narastania, opadania, podtrzymania i wybrzmiewania dźwięku. Przykładowo obwiednia amplitudowa instrumentów perkusyjnych charakteryzuje się dużą dynamiką zmian oraz stosunkowo krótkimi czasami narastania i wybrzmiewania dźwięku. Z kolei większość obwiedni instrumentów klawiszowych stanowi faza podtrzymania oraz wybrzmiewania. Trudnym zadaniem jest określenie charakterystycznych cech dla obwiedni amplitudowej głosu wokalnego. W zależności od sposobu artykulacji oraz obecności słów melodii może ona przyjmować najróżniejsze formy. Niekiedy zdarza się też, że imitacja głosem wokalnym instrumentów (np. trąbki) osiąga bardzo zbliżony kształt do obwiedni zaobserwowanej podczas gry na tym instrumencie.

Poniżej (Rys. 1) przedstawiono przebieg czasowy dźwięku syntetycznego pianina, a także dźwięku wokalnego z użyciem słowa „skąd". Oba te dźwięki posiadają tą samą wysokość muzyczną równą *gis*[1] oraz

reprezentują identyczną (w kontekście całej melodii) wartość rytmiczną. Na podstawie obserwacji obwiedni amplitudowej pianina można jednoznaczne wskazać moment rozpoczęcia dźwięku wraz z linią oznaczającą fragment „B". Z kolei wskazanie go dla głosu wokalnego nie jest oczywiste.



Rysunek 1 Przebieg czasowy dźwięku syntezowanego pianina oraz dźwięku wokalnego z użyciem słowa „skąd".

Wyznaczenie miejsca rozpoczęcia i zakończenia dźwięku wokalnego jest zagadnieniem towarzyszącym początkom przetwarzania mowy, a jego wskazanie jest często arbitralne. Wybór błędnych fragmentów może wiązać się z niewłaściwym określeniem częstotliwości podstawowej dźwięku w przypadku metod pracujących na zmianach amplitudy sygnału (np. analiza przejść przez zero lub analiza funkcji autokorelacji). [3] Na podstawie powyższego wykresu fragment „III" jest najbardziej odpowiedni dla analizy częstotliwościowej. Sąsiadujące z nim „II" i „IV" są momentami przejściowymi, które wymagają przygotowania aparatu mowy do wykonania kolejnego dźwięku. W kontekście ostatniego fragmentu widoczne jest znaczne skrócenie czasu trwania dźwięku, co w zbyt dużym stopniu może skutkować przypisaniem błędnej wartości rytmicznej nuty niż oczekiwana.

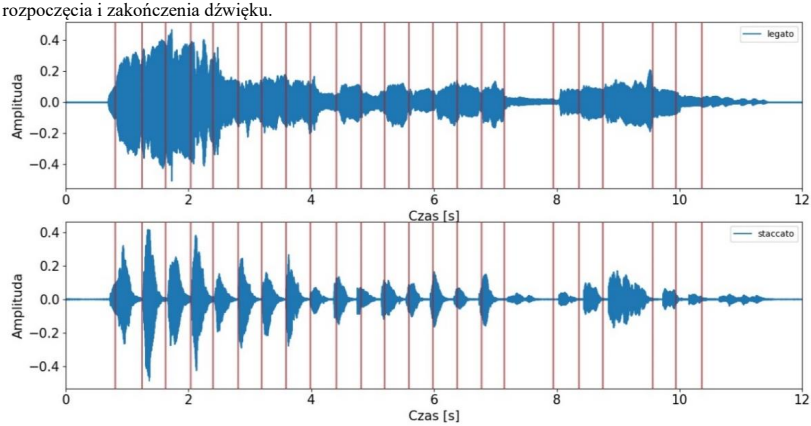
Obecność tekstu melodii stanowi kolejną warstwę analizy mającą wpływ na definiowanie czasu trwania dźwięku. Pierwszy fragment przebiegu czasowego dźwięku wokalnego dotyczy samogłoski szczelinowej „s". Decyzja o dołączeniu jej do analizy może skutkować zarówno przypisaniem błędnej wartości rytmicznej, jak i wysokości całego dźwięku. Detekcja wartości częstotliwości podstawowej zależy mocno od umiejscowienia wydobycia głoski w aparacie głosowym, a jej wartości niekiedy sięgają rzędu kilku tysięcy Hz, co jest wartością nieosiągalną podczas śpiewu [4].

3.2. Rodzaje artykulacji

Warto poddać analizie również przebieg czasowy całego sygnału. Poniżej (Rys. 2) przedstawiono wykonanie tej samej melodii z wykorzystaniem techniki artykulacji *legato* oraz *staccato.* Liniami czerwonymi oznaczono oczekiwaną segmentację na podstawie nagrania wygenerowanego zapisu nutowego – miejsca te powinny zostać uznane za początki nowych nut.

Pierwsze wykonanie melodii charakteryzuje się płynnymi przejściami między kolejnymi dźwiękami, a najbardziej wyraźny wzrost amplitudy wskazuje się na początku frazy. Dodatkowo melodia została zaśpiewana na jednym oddechu, co skutkuje powrotem amplitudy do wartości początkowych dopiero po jej zakończeniu. Z kolei dla wykonania *staccato* każdy dźwięk charakteryzuje się stosunkowo krótkim czasem narastania oraz wybrzmiewania, a także jest oddzielony przerwą. Można powiedzieć, że kształt

obwiedni jest niekiedy zbliżony do instrumentów perkusyjnych, co ułatwia wskazanie momentu rozpoczęcia i zakończenia dźwięku.



Rysunek 2 Przebieg czasowy melodii z artykulacją *legato* oraz *staccato* wraz z zaznaczonymi liniami oczekiwanej segmentacji.

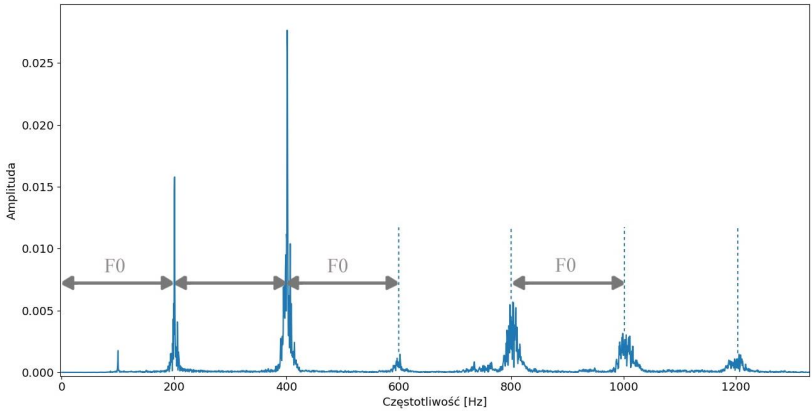
Wybór artykulacji jest znaczący w kontekście wyboru metody segmentacji. Część z nich pracujących w dziedzinie czasu polega na obserwacji obwiedni amplitudowej i jej energii. W tym celu tworzona jest nowa funkcja, której zadaniem jest śledzenie dynamiki zmian sygnału i przypisywanie miejsc występowania jej maksimów jako moment rozpoczęcia kolejnych dźwięków. Dla przedstawionych nagrań zdecydowanie drugie wykonanie jest dedykowane tej metodzie segmentacji. Cechy charakterystyczne dla artykulacji *staccato* wskazują na rozpoznanie większości występujących dźwięków melodii. Z kolei dzięki płynnym przejściom w pierwszym wykonaniu początki nut zostałyby wskazane tylko w nielicznych miejscach. Należy pamiętać, że dodatkowym utrudnieniem jest często nieznajomość melodii wymagającej transkrypcji, co utrudnia odgórne określenie miejsc występowania nut na podstawie przebiegu czasowego. Wpływ wybranych artykulacji na wahania wysokości – segmentację ze względu na wysokości – zostanie przedstawiony w kolejnym rozdziale.

3.3. Detekcja F0 i błędy oktawowe

Kolejna grupa problemów głosu wokalnego w procesie transkrypcji dotyczy dziedziny częstotliwości. Warto zaznaczyć, że aby przypisać wysokość muzyczną do danego dźwięku niezbędna jest informacja o jego wartości częstotliwości podstawowej (F0). Jej wyznaczenie jest możliwe zarówno na podstawie analizy przebiegu czasowego, jak i składowych harmonicznych występujących w widmie częstotliwościowym. Wartość F0 jest równa odległości między kolejnymi składowymi harmonicznymi, o ile takowe występują. Często, błędnie, przyjmuje się, że wartość częstotliwości podstawowej jest równa składowej o największej amplitudzie. Poniższy wykres (Rys. 3) stanowi kontrprzykład tego podejścia prezentując widmo dźwięku o F0 = 200 Hz i drugiej składowej harmonicznej posiadającej największą amplitudę.

Oprócz błędnej detekcji wartości F0, równie często spotykane są tzw. błędy oktawowe. Oznaczają one przypisanie iloczynu wartości częstotliwości podstawowej i kolejnych liczb parzystych, a w efekcie klasyfikację dźwięku do niewłaściwej oktawy. Przyczyną wspomnianych trudności może być nieprawidłowy sposób wydobycia dźwięku, zmiana rejestru wokalnego czy obecność głosek szczelinowych przypisywanych wysokościom niemożliwym do osiągnięcia przez wokalistę.

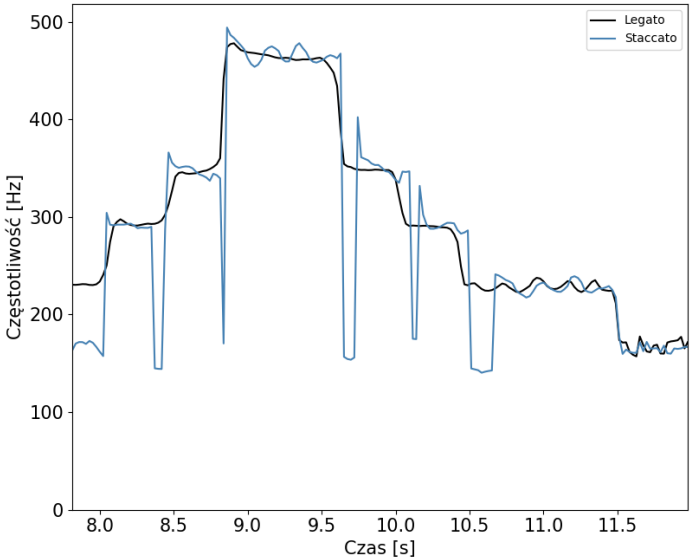
Dodatkowo w zakresie detekcji częstotliwości podstawowej trudnym do wyjaśnienia jest przypadek, gdy niektóre składowe harmoniczne nie występują w widmie, a mimo to słyszalna jest taka sama wysokość, jak w przypadku ich występowania. Jednak w przypadku głosu wokalnego zdarza się on rzadko i równie rzadko jest on uwzględniany.



Rysunek 3 Widmo dźwięku wokalnego ze składową częstotliwości podstawowej równą 200 Hz.

3.4. Wahania wysokości dźwięku

Najczęstszym problemem spotykanym wśród wokalistów jest utrzymanie wysokości muzycznej śpiewanego dźwięku. Trudno wskazać jedną przyczynę jego występowania jednak wiążą się one np. z zaśpiewaniem dużej zmiany wysokości, niewystarczającym podparciem oddechowym, zmianą rejestru lub niewystarczającą do zaśpiewania melodii skalą.



Rysunek 4 Zmiany częstotliwości dźwięków głosu wokalnego z artykulacją *legato* oraz *staccato*.

Powyżej (Rys.4) przedstawiono fragment melodii, której cały przebieg czasowy zaprezentowano w rozdziale 3.2 (Rys. 2). Kształt prowadzenia linii oznaczonej kolorem czarnym (*legato*) przejawia cechy charakterystyczne dla tego rodzaju artykulacji. Jest to stabilna wysokość dźwięku i zaniżanie lub zawyżanie granicznych fragmentów dźwięku w zależności od kierunku, w którym zmierza melodia. Dla artykulacji *staccato* wyróżnia się zawyżenia wysokości wraz z początkiem każdego dźwięku oraz jej stabilizację widoczną podczas całego czasu trwania. Jest to między innymi spowodowane większą dynamiką zmian i ilością procesów, do których zmuszony jest aparat głosowy wraz z pojawianiem się kolejnych dźwięków.

Niekiedy zdarza się, że linia prowadzenia wysokości może stopniowo ulec zaniżaniu lub zawyżaniu wraz z czasem trwania melodii. Jest to sytuacja mogąca wystąpić przy wykonywaniu melodii przez niedoświadczonych muzyków, dodatkowo posiadających problemy z intonacją. W szczególności zdarza się ona przy braku punktu odniesienia (np. śpiewu *a capella* – bez towarzyszenia instrumentów) lub w przypadku występowania dźwięków melodii nieosiągalnych dla wykonawcy i sprowadzania przebiegu melodii w kierunku posiadanej przez nich skali wokalnej.

4. Podsumowanie

Należy pamiętać, że praca z sygnałem wytwarzanym przez aparat głosowy człowieka jest niemal zawsze obarczona niedoskonałościami procesu jego wytwarzania a parametry śpiewanych dźwięków różnić się będą od wartości idealnych. Różnorodność wykonania tej samej melodii, dobór artykulacji, ozdobników oraz interpretacji wokalisty skutkować może otrzymaniem różniących się od siebie wyników transkrypcji. Wymienione aspekty w środowisku muzycznym nie są zazwyczaj określane jako problemy, jednak w kontekście transkrypcji mogą stanowić utrudnienie dla wybranych procesów, takich jak segmentacja lub analiza wysokościowa.

Znalezienie równowagi między sygnałem bezpiecznym dla transkrypcji, a posiadającym cechy sprzyjające odsłuchowi muzycznemu stanowi spore wyzwanie. Świadomość występowania oraz analiza problemów głosu wokalnego sugeruje nowe drogi doskonalenia metod analizy czasowej, wysokościowej, celem poprawy wyników samej transkrypcji.

Bibliografia

[1] Honing H, *Structure and Interpretation of Rhythm in Music*, Cognitive Science Center Amsterdam, Universiteit van Amsterdam, The Netherlands, 2013

[2] Gerhard D, *Computer Music Analysis*, Simon Fraser University, 2002

[3] Rabiner L.R, Cheng M.J, *A Comparative Performance Study of Several Pitch Detection Algorithms*. 1976

[4] Chen C. J., *Elements of human voice*, Columbia University, USA, *2016*

OBIEKTYWNA OCENA JAKOŚCI ORAZ PRZESTRZENNOŚCI SZTUCZNEGO POGŁOSU

A. Korytowski

adamkoryt@interia.pl

Streszczenie

Celem pracy było stworzenie obiektywnych parametrów do oceny jakości sztucznego pogłosu. Zdefiniowanie parametrów było odpowiedzią na powtarzające się w literaturze podkreślenia cech sygnału pogłosowego wpływających na jego jakość oraz odczucie wrażenia przestrzenności. Do ich stworzenia wykorzystano znane parametry statystyczne stosowane do analizy widmowej i czasowej sygnału oraz do oceny podobieństwa sygnałów. Wartości liczbowe stworzonych parametrów mogą być wyznacznikami jakości i przestrzenności różnych wersji pogłosu w celu wyodrębnienia najlepszej. Parametry te należy traktować jako próbę wskazania drogi, która docelowo ma prowadzić do pokrywania się wartości parametrów oceniających pogłos z subiektywnymi wrażeniami słuchaczy.

1. Wstęp

Zjawisko pogłosu akustycznego towarzyszy człowiekowi w niemal każdym momencie życia. Pełni kluczową rolę w odczuwaniu przestrzeni akustycznej oraz lokalizacji źródła dźwięku w pomieszczeniu. Bardzo często pogłos towarzyszący dźwiękowi bezpośredniemu wiąże się z jego atrakcyjnością dla człowieka – zarówno naturalny pogłos pomieszczenia (np. w sali koncertowej), jak i przy odsłuchu we własnych warunkach odsłuchowych (nagranie z naturalnym, zarejestrowanym pogłosem lub pogłosem sztucznym dodanym na etapie miksu utworu).

Badania związane z oceną jakości lub przestrzenności sztucznego pogłosu prowadzone pod koniec XX wieku często wiązały się z koniecznością wykorzystania do tego celu akustycznego pola swobodnego lub dyfuzyjnego oraz stosunkowo skomplikowanego toru audio [1]. Współczesne możliwości znacznie to ułatwiają. Ułatwienia te to przede wszystkim możliwość odwzorowania większości zjawisk w domenie cyfrowej, jak również większe możliwości obliczeniowe. Pozwalają one na realizację większej liczby testów w tym samym czasie, a więc również na przetestowanie większej liczby rozwiązań; przybliżają nas także do potencjalnej możliwości oceny jakości i przestrzenności pogłosu parametrami obiektywnymi.

2. Cel i zakres pracy

Celem pracy było stworzenie parametrów do obiektywnej oceny jakości oraz przestrzenności sztucznego pogłosu. Zdefiniowanie parametrów było odpowiedzią na powtarzające się w literaturze podkreślenia cech sygnału pogłosowego wpływających na jego jakość [2, 3, 4] oraz odczucie wrażenia przestrzenności [1, 5]. Do ich stworzenia wykorzystano znane parametry statystyczne stosowane do analizy widmowej i czasowej sygnału oraz do oceny podobieństwa sygnałów.

3. Znane parametry statystyczne zastosowane w definicji parametrów

- Płaskość widmowa – parametr ten pozwala na rozróżnienie sygnału tonalnego od szumowego [6]. Wartości parametru przedstawiane są w decybelach, gdzie im widmo sygnału jest bardziej płaskie, tym wartość parametru wyższa. Wartość szczytowa (0 dB) oznacza, iż sygnał ma widmo stałe [7]. Definicja płaskości widmowej to stosunek średniej geometrycznej i średniej arytmetycznej mocy widma sygnału [6, 7]:

$$Flatness = \frac{exp(\frac{1}{N}\sum_{n=0}^{N-1}\ln x(n))}{\frac{1}{N}\sum_{n=0}^{N-1}x(n)} \quad (1)$$

gdzie:

N – liczba próbek sygnału,

x(n) – moc widma sygnału,

- koherencja sygnałów – wielkość ta to miara relacji pomiędzy dwoma sygnałami i informuje o zgodności ich faz w badanej częstotliwości [8]. Zdefiniowana jest jako [9]:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} \quad (2)$$

gdzie:

$C_{xy}(f)$ – koherencja sygnałów x i y,

$G_{xy}(f)$ – wzajemna gęstość widmowa między sygnałami x i y,

$G_{xx}(f)$, $G_{yy}(f)$ – wzajemne gęstości widmowe sygnałów x i y samych z sobą,

- widmowa gęstość mocy – wielkość ta określa, jak moc sygnału jest rozprowadzona w dziedzinie częstotliwości. Zdefiniowana jest jako [10]:

$$ø(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-i\omega k} \quad (3)$$

gdzie:

$ø(\omega)$ – widmowa gęstość mocy,

$r(k)$ – funkcja autokowariancji sygnału.

4. Definicja parametrów do oceny jakości i przestrzenności pogłosu

Pierwszy, zdefiniowany przez autora parametr w założeniu służy do oceny jakości pogłosu. Definicja parametru jest odpowiedzią na powtarzające się w literaturze [3, 4, 9] podkreślenia pożądanej wysokiej widmowej gęstości mocy oraz płaskości widma na jakość pogłosu. Parametr może być używany do oceny wpływu doboru wartości czasów opóźnienia linii opóźniających, późnego pogłosu oraz sposobu filtracji na jakość pogłosu. Zdefiniowany przez autora parametr przedstawia się jako zależność:

$$B_x := \frac{\frac{H_{xy}+H_{xz}}{2}E(\emptyset_x)}{\sigma(flatness_x)*\sigma(\emptyset_x)} \tag{4}$$

gdzie:

H_{xy} – współczynnik wyrażony liczbą z zakresu [0;1] będący stosunkiem liczby punktów, w których płaskość widmowa sygnału *x(n)* jest większa, niż płaskość widmowa sygnału *y(n)* do liczby wszystkich punktów wektora płaskości widmowej. Porównywane punkty to wartości wektora płaskości widmowej sygnału – zawiera on wartości tego parametru dla jego kolejnych okien czasowych,

x, y, z – kolejne wersje pogłosu,

σ – średnie odchylenie standardowe,

E – średnia arytmetyczna,

$\emptyset_x$ – widmowa gęstość mocy sygnału *x(n)*,

$flatness_x$ – płaskość widmowa sygnału *x(n)*.

Drugi, zdefiniowany przez autora parametr w założeniu służy do oceny przestrzenności pogłosu. Stopień wrażenia przestrzenności związany jest ze stopniem niekoherencji sygnałów z obu kanałów [5]. Dlatego też, w przypadku chęci spowodowania u słuchacza wrażenia przestrzenności, pożądane jest, aby koherencja sygnałów z obu kanałów była jak najmniejsza. Równie pożądana jest jednak wysoka widmowa gęstość mocy. Przyjęto założenie, iż im bardziej wartości koherencji są skupione wokół średniej tym lepiej. Podobnie, jak odchylenie od wartości średniej widmowej gęstości mocy, która pożądana jest na stałym poziomie. Założenia opisano zdefiniowanym na potrzeby pracy parametrem Pr. Zdefiniowanie parametru jest odpowiedzią na powtarzające się w literaturze podkreślanie istotności cech sygnału wpływające na jakość pogłosu [2, 3, 4] oraz jego przestrzenność [1, 5]. Z pomocą stworzonego parametru możliwe jest porównanie sygnałów pogłosowych z obecnością różnych zjawisk przestrzennych. Cechy te to: wysoka gęstość widma oraz stopień

niekoherencji sygnałów z obu kanałów. Zdefiniowany przez autora parametr Pr przedstawia się jako:

$$Pr := \frac{E(\emptyset_x)}{E(C_{x,LR}) * \sigma(\emptyset_x) * \sigma(C_{x,LR})} \quad (5)$$

gdzie:

σ – średnie odchylenie standardowe,

$\emptyset_x$ – widmowa gęstość mocy sygnału,

$C_{x,LR}$ – koherencja lewego i prawego kanału sygnału,

E – średnia arytmetyczna.

5. Sposób porównywania różnych wersji sygnałów pogłosowych

- Porównując wersje pogłosu w ramach któregoś z jego elementów wszystkie inne parametry pogłosu powinny być identyczne dla wszystkich wersji sygnału – jest to konieczne, aby na wartość liczbową parametru miała wpływ tylko jedna zmienna,
- w celu uzyskania większej dokładności należy porównywać fragmenty sygnałów zawierające tylko sygnał z pogłosem (bez dźwięku bezpośredniego),
- zalecana jest normalizacja wartości liczbowych zdefiniowanych parametrów do wartości maksymalnej w ramach porównywanych sygnałów (na przykład przyjmując wartość maksymalną jako 1),
- w założeniu wyższa wartość parametru oznacza, iż sygnał charakteryzuje się wyższą jakością (wzór 4) lub przestrzennością (wzór 5),
- zalecane narzędzia pomocne w ich wyliczaniu to funkcje biblioteki *scipy* języka Python [11, 12, 13],
- w celu większej świadomości, jakie parametry mają wpływ na ostateczną wartość liczbową, można zapoznać się z graficzną reprezentacją przebiegów czasowych lub częstotliwościowych parametrów opisanych wzorami 1-3,
- zalecane jest wyznaczenie wartości parametrów w ramach różnych rodzajów sygnałów (np. mowy ludzkiej, instrumentów muzycznych, sygnałów pomiarowych), gdyż dla różnych sygnałów wskazania parametrów mogą się różnić.

Parametry mogą mieć różne zastosowania: ich wartości liczbowe można potraktować jako wyznaczniki jakości oraz przestrzenności różnych wersji pogłosu i na tej podstawie wybrać najlepsze. Parametry mogą być również pomocne w przypadku chęci oceny jakości pogłosu bez konieczności badań subiektywnych. Podejście takie

wymaga jednak weryfikacji czy wartości liczbowe parametrów do obiektywnej oceny pogłosu pokrywają się ze wskazaniami subiektywnymi.

Autor zastosował parametry w praktyce, implementując różne wersje pogłosu w ramach jego różnych elementów (wpływu doboru wartości czasów opóźnienia linii opóźniających, sposobu tworzenia późnego ogona pogłosowego i sposobu filtracji na jakość pogłosu oraz obecności różnych zjawisk przestrzennych na jego przestrzenność) i wyznaczył wartości liczbowe stworzonych parametrów dla każdej wersji pogłosu. Następnie, na podstawie wartości liczbowych, wybrał najlepsze wersje oraz połączył wybrane warianty w ostateczny pogłos. Jest to przykład zastosowania tych parametrów.

6. Podsumowanie

Celem pracy było stworzenie parametrów do obiektywnej oceny jakości i przestrzenności pogłosu. Zdefiniowanie parametrów było odpowiedzią na powtarzające się w literaturze podkreślenia cech sygnału pogłosowego wpływających na jego jakość oraz odczucie wrażenia przestrzenności. Przykład ich zastosowania to porównanie różnych wersji pogłosu w celu wyłonienia tej, która charakteryzuje się najlepszą jakością lub przestrzennością.

Historia realizacji tego typu projektów pokazuje, iż czynnik oceny subiektywnej jest nieodłącznym jej elementem. Ma to związek z niewielkimi możliwościami obliczeniowymi w czasach publikacji kluczowych dla tematyki prac, jak również z powtarzalnością ocen przy testach odsłuchowych. Przez powtarzalność opinii osób badanych na temat sygnałów, metodę badawczą uznawano za wiarygodną. Stworzone na potrzeby pracy parametry i sposób ich zastosowania należy traktować jako próbę wskazania drogi, która docelowo ma prowadzić do pokrywania się wartości parametrów oceniających pogłos z subiektywnymi wrażeniami słuchaczy. Nasuwa się więc pomysł zweryfikowania, czy jakość pogłosu wyrażona otrzymanymi wartościami pokryje się z jakością wskazaną przez słuchaczy, lub zaproponowania nowych parametrów i ich weryfikacji również porównując otrzymane wartości ze wskazaniami subiektywnymi. W przypadku znalezienia parametrów zbieżnych co do oceny ze wskazaniami subiektywnymi zostałaby wyeliminowana konieczność przeprowadzania subiektywnych badań, które są bardziej czasochłonne, a ponadto niezbędne jest wykorzystanie do nich dużej grupy osób. Będzie bowiem istniała możliwość oceny jakości pogłosu z dużą szansą, iż wskazania parametrów pokrywają się z wrażeniami słuchowymi.

Bibliografia

[1] Barron M.: The subjective effects of first reflections in concert halls – the need for lateral reflections, Kwiecień, 1969

[2] Schroeder M. R., Logan B. F. – Colorless Sounding Artificial Reverberation, Journal of the Audio Engineering Society, Bell Telephone Laboratories, Incorporated, Murray Hill, vol. 9, lipiec 1961

[3] Schroeder M. R.: Natural Sounding Artificial Reverberation, Journal of the Audio Engineering Society, Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey, Vol. 10, No. 3, lipiec 1962

[4] Jot J. – M., Chaigne A.: Digital Delay Networks for Designing Artificial Reverberators, 90th Convetion of Audio Engineering Society, Paryż, luty 1991

[5] Smyth T.: Time and Space, Department of Music, University of California, San Diego (UCSD), Kwiecień 2019. Dostępny: http://musicweb.ucsd.edu/~trsmyth/space175/space175.pdf (odwiedzona 25.08.20)

[6] Tsihrintzis G. A., Sotiropoulos D. N., Jain L. C.: Machine Learning Paradigms, Springer, Szwajcaria, 2019

[7] Tobias J. V., Broen P., Carey S. W., Casper J., Chapman R., Costello J. M. et al.: Proceedings of the Conference on the Assessment of Vocal Pathology, Bethesda, Maryland, Kwiecień 1979

[8] Kramer M. A.: An Introduction to Field Analysis Techniques: The Power Spectrum and Coherence, Department of Mathematics and Statistics, Boston University, Boston, 2013

[9] Lessard C. S., Ph.D.: Signal processing of random physiological signals, Texas A&M University, College Station, Morgan & Claypool, USA, 2006

[10] Priemer R.: Introductory Signal Processing, Advanced Series in Electrical and Computer Engineering – Vol. 6, Chicago, USA, 1991

[11] Dokumentacja funkcji coherence z biblioteki scipy w języku Python. Dostępny: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.coherence.html (odwiedzona 02.05.20)

[12] Dokumentacja funkcji spectral_flatness z biblioteki scipy w języku Python. Dostępny: https://librosa.org/librosa/generated/librosa.feature.spectral_flatness.html (odwiedzona 02.05.20)

[13] Dokumentacja funkcji welch z biblioteki scipy w języku Python. Dostępny: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.welch.html#scipy.signal.welch (odwiedzona 27.08.20)

SOFTWARE EMULATION OF MIDI ARRANGERS

Kamil Zimny[1]

[1]AGH University of Science and Technology

kzimny@agh.edu.pl

Abstract

MIDI arrangers are the type of keyboard instruments, used for song arrangement and playing live. It's done by using the so-called automatic accompaniment, controlled by the user in real time. With auto-accompaniment play mode, some song-creation operations are moved to the MIDI arrangers' algorithms. Therefore the user can focus on manually played part, while not giving up the following of other instruments. So far, the idea of MIDI arrangers has been used only for hardware devices and its software emulation. Therefore, the use of MIDI arranger features in any MIDI-device systems became the subject of this paper. For this purpose, an application which emulates the operation of MIDI arranger's algorithms has been designed.

1. Introduction

1.1. MIDI Arrangers

Among the many keyboard instruments, there is a separate group of them called MIDI arrangers. They are a multi-timbral synthesizers which, in addition to manual play, allow to use the so-called automatic accompaniment (auto-accompaniment). This is nothing more than following the user's play by the other instrumental parts, which enables the creation and performance of extensive musical arrangements. What's important, the course of the entire song is controlled by the user in real-time.

When playing with auto-accompaniment, the device's keyboard is split into two zones (Figure 1):

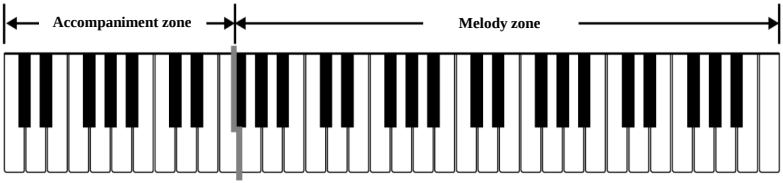


Figure 1 Keyboard zones in MIDI Arrangers in the auto-accompaniment play mode

When pressing a keys in *Accompaniment zone*, the MIDI arranger recognizes the chord and transforms the accompaniment's MIDI sequence to the new chord. The *Melody zone* is used for manual play, like in other keyboard instruments (piano, synthesizer).

MIDI arrangers fit perfectly into the One-Man-Band playing concept. Thanks to the advanced control of the song course in real-time, and better and better sounds, these instruments successfully replace the conventional music bands. Interestingly, the ideology of work and use MIDI arrangers has so far been used only in hardware devices, and virtual emulations of them.

1.2. MIDI

MIDI (Musical Instrument Digital Interface) is an digital standard covering the data transmission protocol, connectors and file format [1]. This protocol is used for communication between electronic musical instruments and other compatible devices. The appearance of the MIDI was an response to the needs of musicians, who want to use their instruments as an integrated system. MIDI gave them this feature and much more like remote control of devices, recording the MIDI data, etc.

In the MIDI standard, data is sent in the form of so-called MIDI messages [2]. Some of them are sent on one of the 16 MIDI channels (Channel Messages), i.e. independent MIDI data sub-streams; the others work as global messages. Among the Channel messages, there is one group, which is mainly responsible for the musical and harmonic content of a MIDI sequence (MIDI song). There are called Note-On and Note-Off messages. Note-On message consist of three bytes of data (**Figure *2***)[1]:

Figure 2 Structure of Note-On MIDI message

where: ***n*** – number of MIDI Channel, ***kk*** – number of MIDI Note, ***vv*** – velocity value.

In MIDI standard each note is assigned a numerical value, within the range 0_D-127_D. The velocity parameter relates to the speed at which a key is pressed, and is usually assigned to control the volume of the sound. In this case the higher velocity value, the louder the sound.

The Note-Off messages are sent when note (key) is released. In most applications, Note-Off messages are replaced by the Note-On, with number of released note and velocity value equal to zero.

2. Project assumptions

The aim of this paper was to design a system which allows to use the features of MIDI arrangers in any MIDI-device system. For this reason, research and tests were carried out on a hardware MIDI arranger KETRON SD9 PRO LIVE STATION [3][4]. Based on it, a general MIDI arranger scheme has been designed (**Figure *3***):

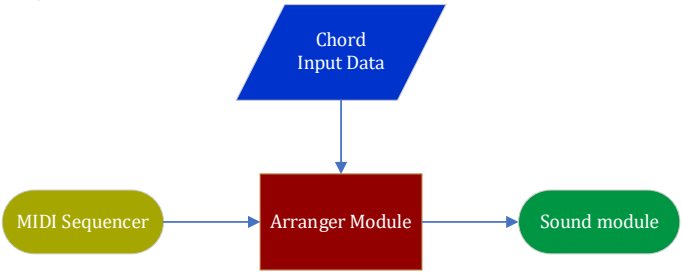


Figure 3 General scheme of MIDI arranger's data flow

Each MIDI arranger is equipped with a large number of auto-accompaniments, in a variety of musical styles[2]. After selecting the desired one, an appropriate MIDI sequence is loaded into the internal MIDI sequencer. It consist of several parts corresponding to different parts of the song (verse, chorus, etc), which are chosen by the user during playing using dedicated buttons on device's panel. What's more, the entire harmonic content of the MIDI sequence is based on one so-called base chord – most often C major.

Then the arranger module – which receives MIDI data from the sequencer – transposes the harmony of the auto-accompaniment, according to the so-called current chord and selected operating modes. The current chord is defined by the user during playing via pressing the appropriate keys in the Accompaniment keyboard (*Chord Input Data*). Finally the processed data is sent to the internal sound module, which executes the MIDI commands, e.g. play sounds.

[1] Two hexadecimal digits (two characters) represent the 8-bit byte content, one character is half of the content (4 bits).
[2] For this reason, the auto-accompaniments are often called *styles*.

There are many MIDI devices that can act as MIDI sequencers, sound modules, or generate the Chord Input Data. For creating the software emulation of MIDI arrangers, the only missing and key element is the emulation of Arranger module. That has been realised as Author's software application named **MIDI Arranger Emulator (MAE)**. This app receives MIDI data from any MIDI sequencer, process it according to recognized chord and selected modes and sent to the outputs. An example system which emulates the working of MIDI arrangers has been shown on Figure 4:

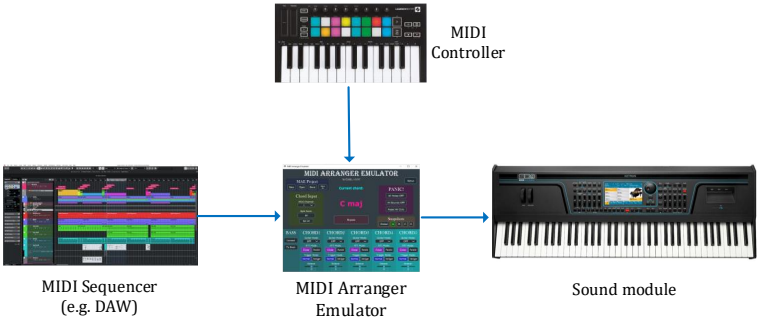


Figure 4 An example MIDI arranger emulation system with MAE app[3]

3. Implementation

The MIDI Arranger Emulator app emulates the operation of the Arranger module. Thus at the beginning, a detailed scheme of it was designed (Figure 5):

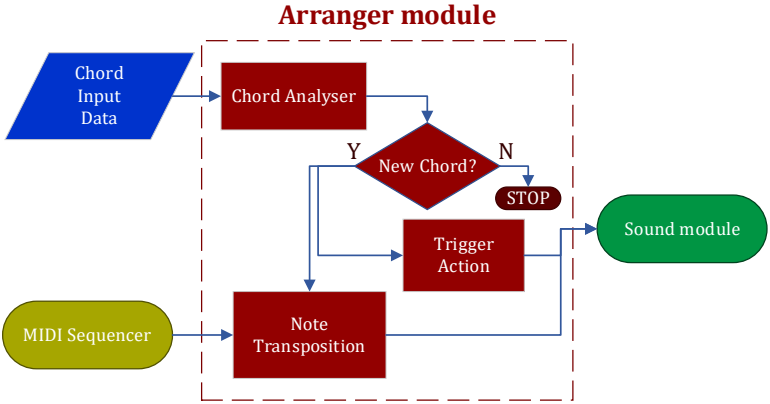


Figure 5 Detailed Arranger module scheme

Details of the implementations of Arranger module's blocks are discussed in the following sections.

[3] Sources: https://www.ketron.it/en/prodotto/sd9-2#&gid=1&pid=1, access 11.06.2021; https://novationmusic.com/sites/novation/files/LKM-overhead-1018-530.png, access 11.06.2021

3.1. Chord Analyser

As mentioned before, each MIDI note has its own number. Thus chord recognition algorithm is based on calculation of the distances (intervals) between keys pressed by the user (Figure 6):

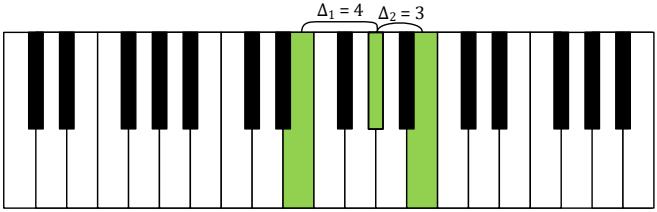


Figure 6 Counting the intervals between pressed keys

After calculating the differences, the algorithm searches the defined database for the pattern. If found, the chord is recognized. The latest version of MAE supports 13 chord types with up to 10-key combinations, and also the simplified chord patterns (1- or 2-keys). The lowest chord note can be recognized separately as the chord root note, if needed (in the *Lowest* mode).

The decision block *New Chord?* checks whether the newly recognized chord differs from the previous one. If so, appropriate actions are performed, otherwise the algorithm terminates.

3.2. Note Transposition

The Note Transposition block is responsible for the note transposition, so moving them to new pitches, according to the current chord. In the implementation, this block works in two modes:

- **Close (fixed) mode** – the transposition is performed so that the intervals between processed and original notes are as small as possible. This is done taking into account the rules of conventional harmony. An example transposition from C major to F major chord in close mode has been shown in Figure 7:

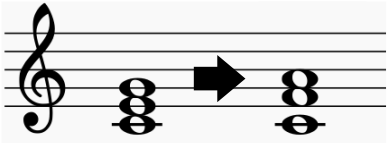


Figure 7 Close mode transposition

- **Parallel mode** – the notes are transposed without changing the intervals between them. To avoid too high transposition values, the so-called inversion chord root point is defined; if root of current chord exceeds it, transposed notes are shifted down by an octave. An example transposition from C major to F major chord in parallel mode has been shown in Figure 8:

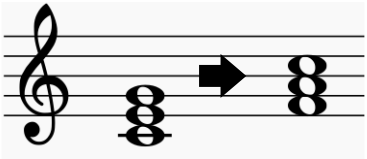


Figure 8 Parallel mode transposition

Due to the finite number of chromatic sound pitches, supported chords and transposition modes, in practical implementation of Note Transposition block the matrix of all possible transposition values is used. This matrix is called Note Transposition Table (NTT), and the role of the algorithm is to find the appropriate values in the NTT and transpose the notes according to them.

3.3. Trigger Action

In general, the recognized chord may change anytime, also when some notes are sounding (they are called active notes). In order for the auto-accompaniment to correctly follow the chords, an action for active notes has to be taken. That is the task for Trigger Action block, which acts in two modes:

- **Normal mode** – after changing the current chord, all active notes are turned off (Note-Off);
- **Retrigger mode –** after changing the current chord, all active notes are turned off (Note-Off), transposed to a new chord, and then turned on (Note-On)

4. MAE distribution

The key element of MIDI arranger's emulation – MIDI Arranger Emulator app – has been designed with JUCE framework in C++ language. It's an standalone app, which could be run on PC with Windows operating system. MAE connects to other MIDI devices using physical or virtual MIDI cables, for receiving and sending the MIDI data. All project assumptions have been successfully implemented – the application works stable and has low resource consumption. In addition to the processing blocks, MAE offers several useful features like saving/recalling an MIDI Input/Output configurations, banks of memory for different configurations, ability to save the MAE configuration to file (so-called MAE project), remote control and much more. The application is being prepared for commercial distribution. In the current version it has an Windows installer, icons, user's manual and security mechanisms which allows running MAE only on the PC's with matched licence file (Figure 9).

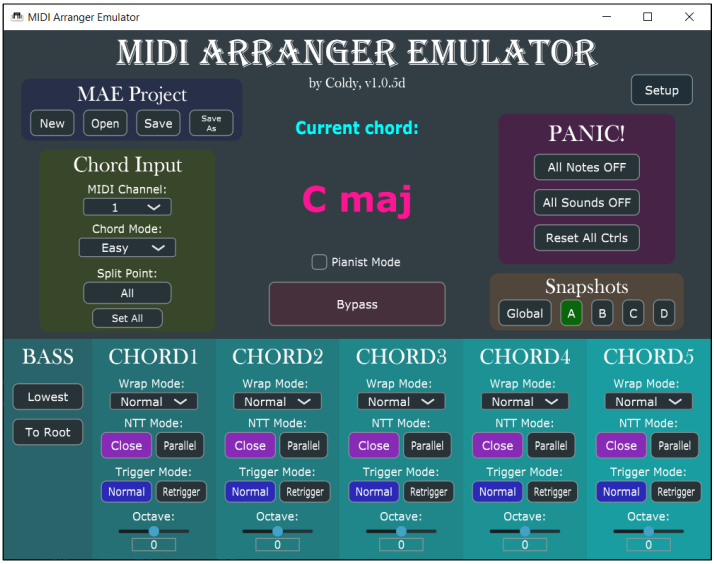


Figure 9 MIDI Arranger Emulator (MAE) app window

5. Conclusion

This paper presented an software emulation of MIDI arrangers system, created with MIDI Arranger Emulator application. Thanks to its modularity, this amazing musical workflow can be used outside of hardware MIDI arrangers, in any MIDI device system, giving new ways of music creation and performance. Moreover, this system is a big help in auto-accompaniment creation process, when it comes to emulate how the created MIDI sequence will act as MIDI arranger's accompaniment. With MAE-based system, this work could be significantly speeded up.

In the future, further expansion of MAE is planned. New processing modes will be added for playing with intervals like third or fifth. Also, the integration with the Steinberg® Cubase software will be added. Thus it will be possible to create the emulation of hardware MIDI arrangers, which use the Cubase DAW as sequencer and any sound modules (hardware and/or software like VST instruments). The operation of this system will be controlled from MAE app. As for security, the ability to run the application using a hardware USB key will be added. Also the number of supported chords will gradually increase, and many other improvements will be added for increasing the stability and convenience of using MAE.

Thanks to the operability only on MIDI signals, the MAE-based emulation of MIDI arrangers finds the above-mentioned and many other applications. The only limit is an human imagination in terms of expressing themselves during creating and making the music performances.

References

[1] The MIDI Manufactures Association, *The Complete MIDI 1.0 Detailed Specification*, Los Angeles 2014

[2] Paul D. Lehrman [PhD], *What is MIDI?*, Sections adapter from „MIDI For Professional" by Paul D. Lehrman & Tim Tully, 2017

[3] Kamil Zimny, Inżynieria wsteczna algorytmów instrumentu KETRON SD9 PRO LIVE STATION, AGH, Kraków 2020

[4] KETRON s.r.l., KETRON SD9 PRO MANUAL, HINTS & TIPS, Ancona 2016

Sesja 3

Psychoakustyka

wtorek, 29.06.2021
10:30 – 12:10

QUALITY ASSESSMENT OF SPEECH SIGNALS BY TEENAGERS AND YOUNG ADULTS

P. Falkowski-Gilski[1], S. Brachmański[2], A. Dobrucki[2], and M. Kin[2]

[1]Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics,

Narutowicza 11/12, 80-233 Gdańsk

[2]Wroclaw University of Science and Technology, Faculty of Electronics,

Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

przemyslaw.falkowski@eti.pg.edu.pl

Abstract

The subjective assessment of speech signals takes into account previous experiences and habits of an individual. Since the perception process deteriorates with age, differences should be noticeable among people from dissimilar age groups. In this work, we investigated the difference of speech quality assessment between high school students and university students. The study involved 60 participants, with 30 people in both the adolescents group aged 16-18 years, as well as the young adults aged 20-25 years, without any hearing disorders. The speech samples included 3 sets: American English, British English, and Polish. Results of research indicated that evaluation of young adults is characterized by higher scores and lower variations of obtained results.

1. Introduction

The task of any speech transmission system is to provide high quality and intelligibility of speech signals. Due to variable parameters of transmission links and/or broadcasting system, e.g. bandwidth, delay, packet loss, etc., the network parameters may cause degradation. We then speak of two concepts of quality: objective QoS (Quality of Service), expressed as a set of technical parameters, and subjective QoE (Quality of Experience), linked with feelings of the user.

The subjective evaluation of listeners is a complex process that takes into account previous experience and habits of individuals. Since the perception process deteriorates with age, e.g. due to harmful working conditions, listening to too loud music, etc., differences should occur for people from different age groups.

The popularity of various online social networking sites, as well as those designed for music distribution and language teaching, enforces newer and newer signal processing solutions. One of such treatments is the process of lossy coding, widely used due to effective reduction of the volume of transmitted information. An important part of this coding scheme is the selection of the correct bitrate [1]. It is known from literature that speech signals can be transmitted at a lower bit rate, compared to music, if only one can provide different criteria for the perception of music signals [2]. However, excessive reduction of bits in case of speech transmission is not desirable, not to mention educational purposes, where the effectiveness of learning a foreign language is based both on intelligibility and aesthetic criteria [3], [4].

Despite many years of research on speech signal perception, still the concept of how the human brain interprets complex acoustic signals, which may be perceived as linguistic elements, is not fully understood and clarified. It remains unknown what is the main element in the process of speech signal perception and speech quality evaluation [5]. Furthermore, knowledge and familiarity of a particular language provides a basis for the actual use of this language, both speaking and listening [3], which may affect the differences in assessment of speech quality itself, due to redundancy of the speech signal.

In view of the above, it seemed interesting to examine the way of ratings in terms of perception and quality assessment of speech signals spoken in various languages, including both the American and British English dialects, as well as Polish, coded at different bit rates. The aim of this work was to investigate the process of perception and evaluation of speech quality among two groups of young people with no hearing impairment.

2. Experiment

2.1. Stimuli and procedure

The test signals were all speech samples sourced from ITU-T P.501 [6], among which there are numerous so-called sentence lists, available in different languages. Each list consisted of sentences spoken by two male and two female lectors, and one sentence list was read by one lector. After the initial analysis, 3 sets were selected: American English (AE), British English (EN) and Polish (PL).

The original files were available in the WAV 16-bit PCM format. Each file was processed using the Ogg Vorbis codec, which is utilized in Spotify. According to a preliminary questionnaire (see Fig. 1), this streaming service proved to be the most popular one in case of young people.

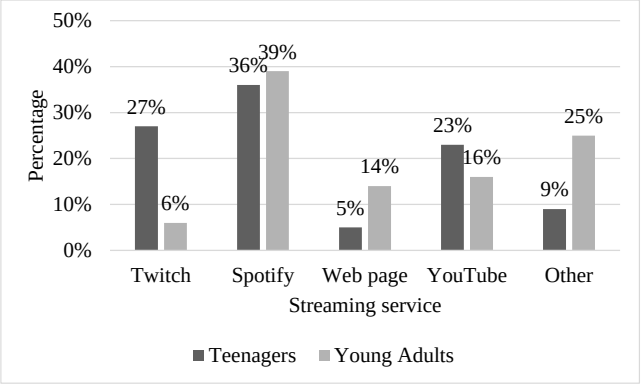


Figure 1 Popular streaming services among teenagers and young adults.

The audio files were processed at 3 bitrates: 8, 16, and 24 kbit/s. The sampling rate was set to 44.1 kHz, as in case of many popular online services and audio transmission systems.

The listening tests were carried out using closed headphones. The Absolute Category Rating (ACR) method was applied, and participants were asked to rank the overall speech quality of each sentence in a 5-degree scale as the MOS (Mean Opinion Score) notes, from 1 (bad quality) to 5 (excellent quality), according to the ITU-R BS.1284 recommendation [7]. Each sentence was presented 3 times in random order. A single session, preceded by a training phase, during which it was possible to adjust the volume according to one's needs, took about 10 minutes. Each person performed the assessment independently, and was not informed about the actual bitrate of the currently presented speech signal.

2.2. Participants

The study involved 60 individuals divided among 2 groups, with 30 people in each of them, i.e. teenagers aged 16-18 years, and young adults aged 20-25 years. Each one started and continues his or her education in a Polish educational institution with the Polish language of instruction.

The group of teenagers included high school students. Each person was fluent in Polish (as their native language) and English (at an advanced level), which was their favorite one in which they communicate with foreigners. The group of young adults consisted of university students for whom Polish was also their native language. As in the younger group, English was their favorite foreign language as well, which they learned at an advanced level.

The subjects of experiment have no experience in listening tests, so it can be said that the research was provided with amateur participants thus reflecting average preferences in a targeted audiences.

3. Results and discussion

The results of experiment, considering the assessment of speech quality and dispersion of the results, for both groups: teenagers (16-18 years old) and young adults (20-25 years old), utilizing samples of speech signals spoken in American English (AE), British English (EN) and Polish (PL), processed at 3 bitrates (8, 16 and 24 kbit/s), are presented in Figs. 2-4. The results were statistically analyzed with the use of Bartlett test and the variances in both groups were found to be homogeneous for particular signals used (the values of the χ^2 statistic were less than the critical one (χ^2_α = 24.996) at a significance level of α = 0.05, allowing the answers to be averaged across all listeners in a given group.

According to obtained results, it may be clearly indicated that for the lowest bit rate of 8 kbit/s, teenagers were more critical and sensitive to signals distortions compared to young adults. For the medium bit rate of 16 kbit/s, the difference in MOS ratings between the two groups was smaller, regardless of the language and/or lector. Nevertheless, the younger group proved to be more demanding once again. The only exception appeared to be for the Polish (PL) language sample from the male lector. In case of the highest bit rate of 24 kbit/s, teenagers once again provided lower scores. Only British English (EN) samples were rated at a similar level, both for female and male lectors. The rest of the samples, including American English (AE) and Polish (PL), were rated higher by university students.

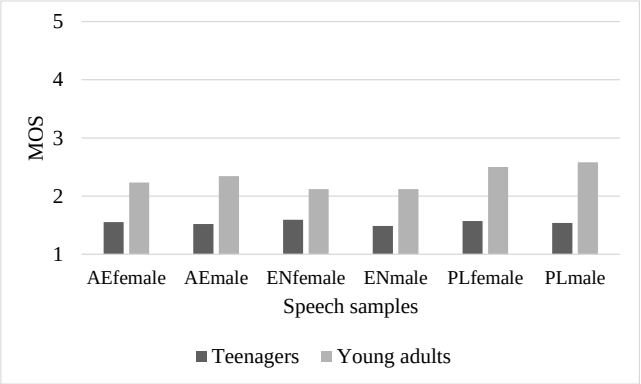


Figure 2 Subjective quality assessment of speech samples coded at 8 kbit/s.

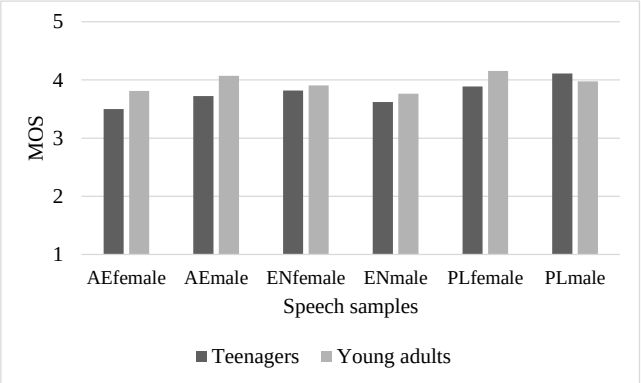


Figure 3 Subjective quality assessment of speech samples coded at 16 kbit/s.

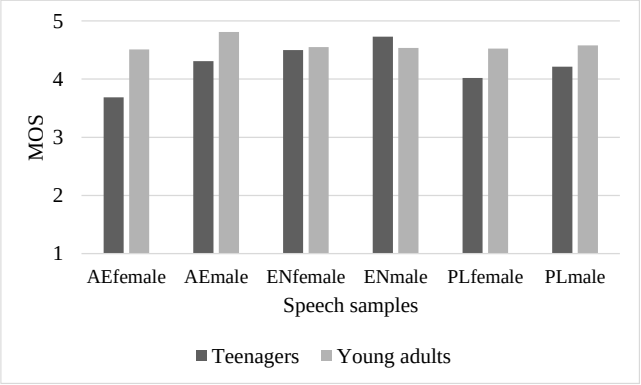


Figure 4 Subjective quality assessment of speech samples coded at 24 kbit/s.

In order to examine the degree to which evaluations in the two groups converge, it was decided to compare the variance values of the scores obtained in the two listening groups for given speakers and specific languages. A similar measure had been used to research on the pitch strength of residual pitch [8] and the impression of spatiality of recordings made in different rooms [9]. This would allow us to estimate the stability and unambiguity of judgments for specific stimuli as a function of listeners' age.

Figs. 5-7 shows the variance values of scores obtained by both groups of listeners in case of a given bit rate. It can be observed that the variance is smallest for the lowest bit rate of 8 kbit/s, and thus characterized by the lowest MOS ratings. It may be said that the values of variances testify the high confidence of ratings and the homogeneous criterion, regardless of individuals.

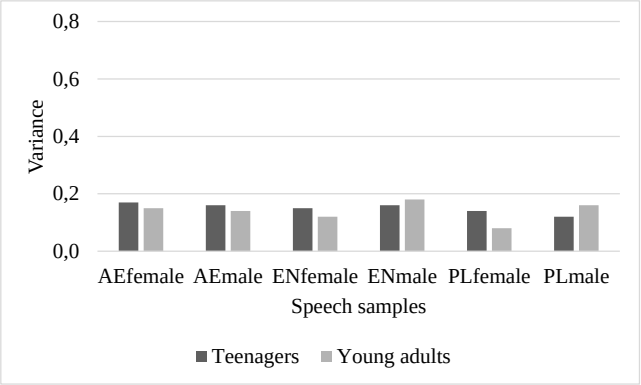


Figure 5 Variance of MOS scores for speech samples coded at 8 kbit/s.

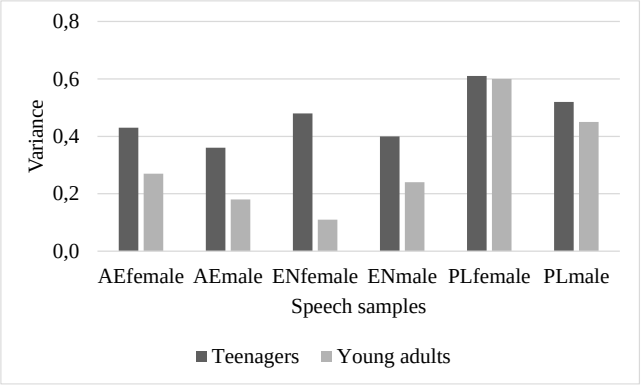


Figure 6 Variance of MOS scores for speech samples coded at 16 kbit/s.

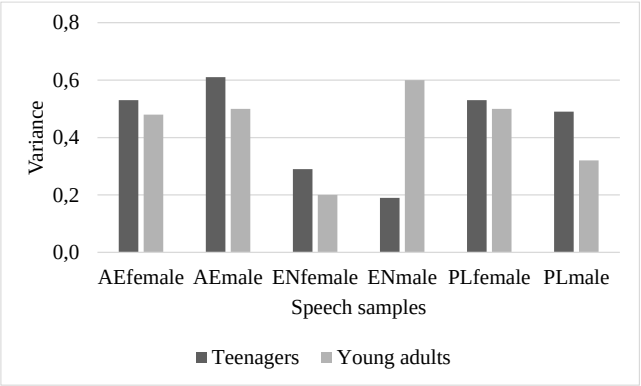


Figure 7 Variance of MOS scores for speech samples coded at 24 kbit/s.

For higher bitrate values, an increase in variance was observed, with higher MOS rates, which may be due to additional influence of the aesthetic experience on the quality evaluation. The aesthetic layer "supports", so to speak, the information content in the evaluated sample [10], resulting in an increase of the overall quality rating, without worsening the intelligibility of the speech signal, while the quality criterion is not less clear [11]. This is especially visible in case of the medium bit rates of 16 kbit/s, and for teenage listeners during the evaluation of English samples in both American and British dialects. Polish language samples were rated for this bit rate in both groups with similar dispersion. As the bit rate increases to 24 kbit/s, the variance values in both groups become equal, indicating greater homogeneity in the rating criterion.

4. Conclusions

Obtained results showed some differences between both groups, including e.g. sensitivity to noise and distortions, related to lossy coding, habits regarding the acceptable level of audio signal quality, as well as language preferences. According to the obtained results, it can be concluded that younger users proved to be more demanding listeners. This was especially seen in case of the lowest bit rate, equal to 8 kbit/s, and the highest one, equal to 24 kbit/s. On the other hand, it may be said that young adults are able to appreciate

audio content of higher quality, as they assessed all speech samples at a level above 4.0 (good quality) in case of the highest investigated bit rate of 24 kbit/s.

It should be also mentioned that the homogeneity of subjective assessments, in addition to experience, is also determined by the emotional attitude of the listeners to the presented stimuli. When one grows up with what he or she embraces with consciousness, along with goals, one sets the values that he or she has learned to know and has experienced [12]. On the other hand, speech signals are the most useful auditory stimuli and these sounds have their roots in everyday life what may suggest that this kind of stimuli does not influence the overall impression of sound image from semantic point of view [13]. However, the overall quality of speech signals may depend on degradation of original sounds while evaluated without the involvement of the semantic layer, despite a clear understanding of the message.

To sum up, there are noticeable differences in the assessment of speech signal quality among two examined age groups, that is teenagers (16-18 years old) and young adults (20-25 years old). These differences are independent of the language sample, being either one of the English dialects, or their native Polish language, when the MOS ratings exceed values of 3.5. For the lowest observed quality, there are no significant differences between the MOS results as well as the variances of scores obtained for teenagers and young adults.

References

[1] P. Falkowski-Gilski: “On the consumption of multimedia content using mobile devices: a year to year user case study”, in: *Archives of Acoustics* 45(2), 2020, pp. 321–328.

[2] M. Kin and S. Brachmański: “Quality assessment of musical and speech signals broadcasted via single frequency network DAB”, in: *International Journal of Electronics and Telecommunications* 66 (1), 2020, pp. 139–144.

[3] J. Greene: *Psycholinguistics. Chomsky and Psychology*, New York, 1972.

[4] S. Brachmański: *Selected issues of speech signal transmission quality assessment*, Wrocław, 2015 [in Polish].

[5] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy: “Perception of the speech code”, in: *Psychological Review* 74 (6), 1967, pp. 431-461.

[6] ITU Recommendation P.501: *Test signals for telecommunication systems*, Geneva, 2017.

[7] ITU Recommendation BS.1284: *General methods for the subjective assessment of sound quality*, Geneva, 2003.

[8] H. Fastl, G. Stoll: “Scaling of Pitch Strength”, in: *Hearing Research* 1 (4), 1979, pp. 293-301.

[9] A. Jasińska, M. Kin: “The Influence of Room Acoustic Parameters on The Perception of Room Characteristics”, in: *Akustika* 37 (12), 2020, pp. 29-32.

[10] A. Moles: *Theorie de l’information et perception esthetique*, Paris, 1958.

[11] H. Fastl: “Psycho-acoustics and sound quality”, in: *Communication acoustics*, ed. J. Blauert, Berlin, 2005.

[12] S. Szuman: “On the important role of art in adult education”, in: *Selection of aesthetical papers*, ed. M. Kielar-Turska, Krakow, 2008 [in Polish].

[13] G. B. Kempster, B. R. Gerratt, K. Verdolini-Abbott, J. Barkmeier-Kraemer, R. E. Hillman: “Consensus Auditory-Perceptual Evaluation of Voice”, in: *American Journal of Speech-Language Pathology* 18, 2009, pp. 124-132.

RESEARCH ON THE PERCEPTION OF SOUND COLOR BY PEOPLE WITH PRESBYACUSIS

A. Dobrucki[1], M. Kin[1]

[1]Department of Acoustics, Multimedia and Signal Processing, Faculty of Electronics, Wrocław University of Science and Technology

andrzej.dobrucki@pwr.edu.pl

Summary

The paper presents the results of a study on the perception of timbre by people with presbycusis and compares them to the results obtained for otologically healthy subjects. It was found that for certain sounds the timbre attribute was evaluated differently, but there were cases where the evaluation of timbre did not depend on hearing loss and was the same as for healthy subjects. In the case of timbre brightness ratings for some cues, subjects with presbycusis rated these cues as brighter compared to those with normal hearing. In contrast, other signals were rated as darker by these subjects compared to the ratings of the control group. It was also found that the sensitivity to spectral change in the higher frequency range in people with hearing loss is greater than in normal-hearing people, as long as these changes are significant, while small spectral changes are perceived worse by this group.

1. Introduction

According to the World Health Organization, more than 450 million people worldwide suffer from hearing problems. Hearing loss can be caused by genetics, complications during childbirth, disease complications, consumption of ototoxic drugs, exposure to noise and due to aging. The human ear, like other human organs, ages over the years. High-frequency sensorineural hearing loss that worsens with age is called sensorineurale deafness (Latin *presbyacusis*) [1][2].

Presbyacusis is most often caused by involutionary changes in the sensorineural part of the hearing organ, i.e. in the inner ear, in the spiral ganglion and cochlear nerve, and in the auditory centers and pathways of the central nervous system. Results obtained by tonal audiometry of individuals with presbyacusis show sensorineural loss, usually symmetrical, and no or little cochlear reserve. With such kind of hearing loss the worsening of speech discrimination is noted [3]. Thus, it can be speculated how the perception of sound differs for people with presbyacusis compared to normal hearing. Modern methods of hearing aids partially compensate for such conditions, nevertheless, the measures used are primarily aimed at improving the communication of the hearing impaired person with the environment, while for technological reasons there is no emphasis on improving the aesthetic experience. It is obvious that with large high-frequency losses, a person cannot hear a sound with a frequency in the region of their hearing loss. But what happens when complex sound signals are heard by people with presbyacusis? Does this kind of deafness affect the perception of tone change? These are the questions the authors tried to answer in this paper.

The aim of this study was to perform the test of timbre perception changes for individuals with presbycusis and a control group of normal-hearing subjects. The tests were conducted using musical signals as well as sounds commonly encountered in everyday life (e.g., doorbell and car horn). The results were compared to each other and statistical analysis was performed.

2. Experiments

2.1. Stimuli and their presentation

The study was performed in 2 ways: as a perception of the timbre changes made in the sound spectrum, and as an evaluation of the sharpness and brightness of the timbre of various signals (including everyday life signals and signals of pink noise, white noise, and male and female speech). To examine the perception of timbre changes, a sequence of pairs of music signals from different genres was used, in which timbre correction was performed for 3 bands with center frequencies of 125 Hz, 1 kHz, and 10 kHz. The test procedure was based on the pair presentation method. The goodness of the filters used was 0.7, and the corrections were made in the range of ±1.5 dB, ±3.0 dB, and ±6.0 dB. The listeners' task was to answer the

question whether they could notice a difference in the sound of the first and second samples in the pair, without the rating. The second part of the test consisted of evaluating the brightness and sharpness impressions of the presented signals. Samples of sounds ranging in duration from 0.5 s (warning and noise signals) to about 1.5 (speech samples) were presented to the listeners. The parameter rating scale was in accordance with the recommendations of ITU-RBS.1284-1 [4]. A five-scale rating system was used by the subject: from -2 (dark tone, indistinct sound), through 0 (neutral, neutral sound) to +2 (light tone, sharp sound). The listeners noted their responses on prepared sheets.

Tests of timbre perception were conducted in room with low background sound levels. It was the same room in which hearing thresholds were tested. The test signals were played back using AKG K240 headphones. Listening was done individually and only one listener and the testing person were present in the room during each measurement. The duration of a single test block was 15 min. for the detection of spectral change, and 5 min. for the assessment of the brightness and sharpness of the tested sounds. Each person repeated the test 3 times, and the interval between each test block was 15 min. Before the start of the listening sessions, each participant underwent a hearing test using thresholded tonal audiometry. Then, each time, the key concepts that were tested were explained and the experimental setup was presented. The teaching sequence was presented to the listeners.

2.2. Listeners

The experiment was performed for 10 individuals with presbyacusis, aged 60-70 years. Six women and four men participated in the tests. All subjects in this group had hearing losses between 20 dB and 35 dB at 4 kHz and 8 kHz. These subjects had not previously participated in similar listening tests. A control group of 10 subjects with normal hearing (less than or equal to 5 dB) was also tested. The subjects in this group had extensive experience in auditory testing, including sound quality assessment. In addition, all of the control subjects were familiar with music on a daily basis - at work, at university, and in their everyday lives.

3. Results

3.1. Perception of spectra changes

The results obtained were statistically analyzed and the variances in both groups were found to be homogeneous for all music signals used (the values of the χ^2 statistic were less than the critical one (χ^2_α = 5.1456) at a significance level of α = 0.05, allowing the answers to be averaged across all listeners in a given group [5].

Fig. 1 shows the results of a test on the detection of changes in the sound spectrum of musical signals. It is noticeable that in the low and medium frequencies, the detection of spectral changes of musical signals by people with presbyacusis does not differ significantly from the results obtained by healthy subjects. However, some differences were observed for spectral changes in the high-frequency range: subjects with presbyacusis perceived small changes in the spectrum (±1.5 dB) worse than the subjects from control group. On the other hand, large ± 6 dB changes in this frequency range are perceived better in comparison to subjects with normal hearing. This phenomenon may be due to the fact that the auditory area in the high-frequency range becomes narrower, which, in the context of even isophonic curves or Weber-Fechner's law, results in a greater sensitivity to changes in sound sharpness and brightness [6]. However, it should be kept in mind that these subjects do not listen to all musical genres used in the experiment in their daily lives, which may be reflected in the behavioral-verbal data obtained [7], which, despite the homogeneity of response variance across genres, may cause subjective ratings to be exaggerated.

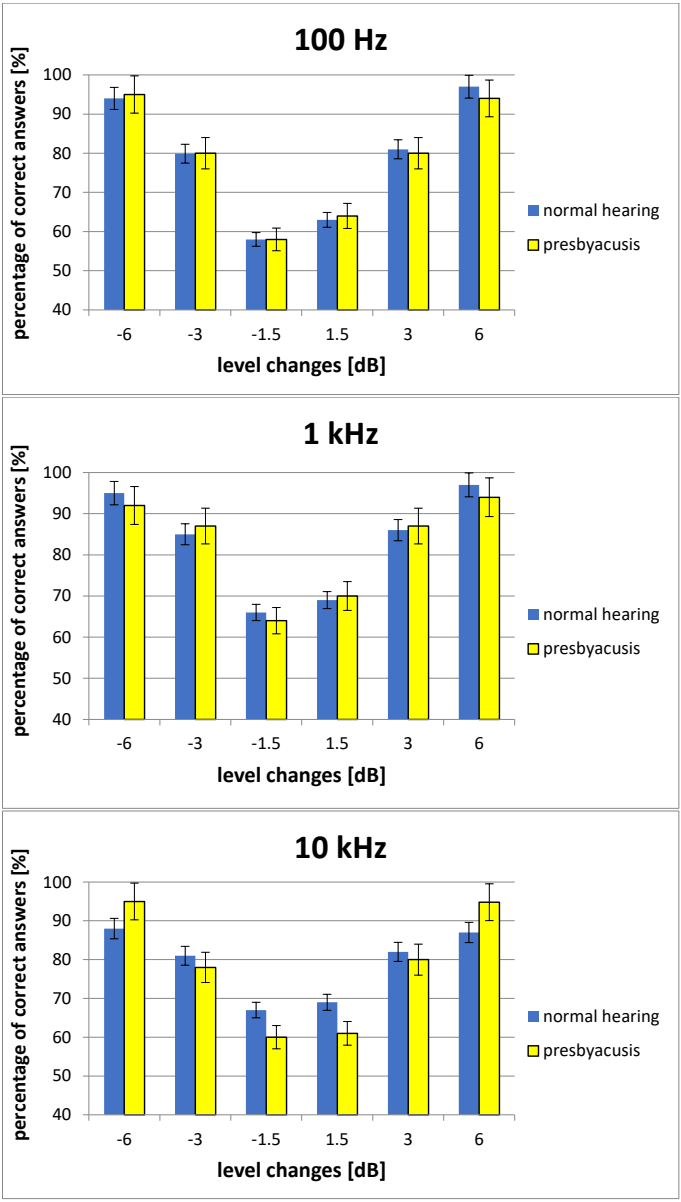


Figure 1 Percentages of correct tone change detection responses for selected frequencies.

3.2. Evaluation of sound brightness and sharpness

Figure 2 shows the results of the brightness and sharpness evaluation of the tested sounds. On the basis of the conducted test for homogeneity of variances (Bartlett's test) and the value of the χ^2 statistic, only in the case of the car horn signal and the pink noise the calculated value was greater than the critical one ($\chi^2_\alpha = 3.8419$) at the significance level $\alpha = 0.05$, which means that in the case of the remaining signals, the variances of the responses of both groups are equal and the different degree of dispersion of the evaluations for both groups was not confirmed [5]. The obtained responses in both the test and control groups can therefore be treated as homogeneous for all signals except for the above mentioned ones. In order to examine whether the responses of the two groups are statistically significantly different, *t*-Student test for two mean values was performed. It turned out that, except for the bell sound and pink noise, the statistical significance of the differences in the ratings of the brightness of the tested sounds was confirmed for all other signals, while statistically significant differences in the ratings of the sharpness of the timbre between both groups were reported for male speech and female speech (the calculated values of the statistic are greater than the critical value of $t_\alpha = 1.093$ at the significance level of $\alpha = 0.05$).

Figure 2 Results of the evaluation of the sharpness and brightness impressions of the test sounds.

In the case of tone brightness ratings for most of the signals tested, listeners with presbyacusis rated them as darker compared to the ratings of the control group, the exceptions being male speech and the horn sound. On the other hand, when it came to tone sharpness, listeners with presbyacusis rated the sounds of the horn, bell, and both types of noise as sharper, while both speech signals were rated as less sharp. It is also notable that, with the exception of the speech signals, the other sounds were rated as sharp by both groups of listeners. This fact may be due to the way in which the stimuli were presented, as well as to their content: the participants reported that this group of signals reminded them of warning signals. Speech signals, on the other hand, presented in the form of single sentences, contained information that may have been partially difficult to perceive, especially for people with hearing loss, which may have influenced the evaluation of this attribute and value assessments according to the criterion assumed by the listeners [8].

It should also be noted that the response variance values in the control group are higher, for both brightness and sharpness ratings, compared to the presbyacusis group, which may indicate either that individuals with hearing loss are less sensitive to temporal changes in sensation or that the criterion is more heterogeneous in healthy subjects, depending on signal content [9].

4. Discussion

The results of this study allow us to conclude that presbyacusis influenced the value and choice of timbre evaluation criteria for most of the presented signals. One might expect that presbyacusis, as a high-frequency hearing loss, would result in decreased timbre brightness and sharpness impressions, due to decreased high-frequency audibility, compared to timbre perception by normal-hearing subjects. For brightness ratings for two types of cues (speech and horn), subjects with presbyacusis rated these sounds as brighter compared to normal-hearing subjects. However, the other signals were rated as darker by these subjects compared to the ratings of the control group. It is interesting to note that the spectra of horn sounds and male speech are different, so an explanation for this phenomenon based on ratings based on spectrum only is not valid. A probable explanation for this effect may be related to the mode of aesthetic evaluation of natural signals containing information (speech) and other sounds that may be more abstract in a given context [10].

The ratings for timbre sharpness and the ability to recognize timbre changes are more similar to expected values, i.e., the ratings obtained in the aging hearing loss group are higher than those of the control group for signals whose nature is similar to warning signals. In contrast, speech signals (both male and female) were rated as less sharp. This means that people with presbyacusis perceive these signals as softer compared to people with normal hearing, which may be related to the perception and processing of the semantic content of speech signals: if there is a decrease in intelligibility, these signals are unwittingly rated lower, often regardless of the criterion assumed [8][11]. Thus, it can be concluded that individuals with presbyacusis perceive warning-like signals as sharper.

The results of the signal spectrum change detection experiment indicate that the degree of detectability of signal spectrum changes from hearing impairment varies: for the group with presbyacusis, in the higher frequency range (10 kHz octave band) a decrease in detectability of small spectrum changes (± 1.5 dB) was noted, while for significant changes (±6 dB) these changes are more noticeable, which may be related to the narrowing of the auditory area resulting from presbyacusis [6]. Comparing the results of the experiment in review with the results of a study on the discrimination of sound color change under listeners' fatigue caused by a temporary increase in the hearing threshold (TTS) [12], it can be concluded that the results of detection of spectral changes for frequencies with a value of ±6 dB obtained for a group of listeners with hearing loss are similar to those obtained for healthy listeners after 90 min of exposure to loud music (with an equivalent sound level of 92.6 dB). Thus, it can be said that in the high-frequency range, the impressions of listeners with hearing loss correspond to those of persons with normal hearing but fatigued by long-term exposure to noisy music. It should also be noted that a slight changes in timbre in the middle and treble range are similarly noticeable to hearing fatigue following 60 minutes of listening, resulting in an increase in the hearing threshold of 5 - 7 dB, depending on frequency, which is consistent with literature reports [13][14].

5. Conclussions

The results of the experiments allow us to conclude that the listeners with age-related deafness have similar perceptual properties in the high-frequency range in terms of detecting changes in the sound spectrum as persons with normal hearing, in whom a temporary threshold shift by about 5-7 dB occurred. Although the permanent threshold shift for this frequency range in these subjects ranged from 20 to 35 dB, it should be noted that these listeners did not exhibit the mental fatigue to which healthy listeners were affected [12]. Thus, it can be predicted that fatigue in this group of listeners would cause additional degradation in the detection of spectral changes of various signals.

It has been reported that elderly required from 10 to 20 dB higher sound level than young people to obtain a good speech intelligibility scores [15]. This fact may support the explanation of presented results of sharpness impression which depends on the semantic content of sound: if the test signals contain such information, listeners with presbyacusis rated these stimuli as less sharp, while the other signals were treated similarly than by the others. Thus, it may be concluded that good intelligibility of speech for elderly people can be achieved by the higher values of speech transmission index [16] what is connected to the sharpness impression of speech signal.

Bibliography

[1] E. Hojan: *Hearing Care,* UAM Press, Poznań, 2015 [in Polish].

[2] M. Bergman: „Hearing and aging: implications of recent research findings", in: *Audiology*, 10 (3), 1971, pp164-171.

[3] J.R. Dubno, D.D. Dirks, D.E. Morgan: "Effect of Age and Mild Hearing Loss on Speech Recognition in Noise", in: *Journal of the Acoustical Society of America*, 76 (1), 1984, pp. 87-96.

[4] ITU-RBS.1284-1, Methods for the subjective assessment of sound quality – general requirements.

[5] J. Greń: *Mathematical Statistics. Models and tasks,* PWN, Warszawa, 1974 [in Polish].

[6] H. Fastl, E. Zwicker: *Psychoacoustics. Facts and Models,* Springer, Berlin, 2007.

[7] T. Pawłowski: „Subjectivity", in: *Selection of aesthetical papers,* G. Sztabiński – ed., Univeritas, Kraków, 2010 [in Polish].

[8] S, Brachmański: *Selected issues of speech signal transmission quality assessment,* Wroclaw University of Science and Technology Press, Wrocław, 2015 [in Polish].

[9] H. Fastl: Psycho-Acoustics and Sound Quality, in: *Communication Acoustics,* J. Blauert – ed., Springer, Berlin, 2005.

[10] S. Morawski: „On the objectivity of aesthetic judgments", in: *Selection of aesthetical papers*, P. Przybysz and A. Zeidler-Janiszewska - eds, Univeritas, Kraków, 2007 [in Polish].

[11] A. Perry: *Fundamentals of Voice-Quality Engineering in Wireless Networks,* Cambridge University Press, Cambridge, 2007.

[12] M. Kin, A. Dobrucki: "Perception of Changes in Spectrum and Envelope of Musical Signals vs Auditory Fatigue", in: *Archives of Acoustics*, 41 (2), 2016, pp. 323 – 330.

[13] S. M. Marcora, W. Staiano, V. Manning: "Mental fatigue impairs physical performance in humans", in: *Journal of Applied Physiology*, 106 (3), 2009, pp. 857‑864.

[14] B.C.J. Moore: "Effects of sound-induced hearing loss and hearing aids on the perception of music", in: *Journal of Audio Engineering Society*, 64 (3), 2016, pp. 112‑123.

[15] A. K. Nabelek, P. K. Robinson: „Monoaural and binaural speech perception in reverberation for listeners of various ages", in: *Journal of the Acoustical Society of America*, 71 (5), 1982, pp. 1242-1248.

[16] J. Peng, J. Zeng, Y. Zhao: "Relationship Between Chinese Speech Intelligibility of Elderly and Speech Transmission Index", in: *Archives of Acoustics*, 46 (2), 2021, pp. 229-235.

PROGI DYSKRYMINACJI WYSOKOŚCI TONÓW NISKOCZĘSTOTLIWOŚCIOWYCH

Piotr Rogowski[1], Andrzej Miśkiewicz[1]

[1] Uniwersytet Muzyczny Fryderyka Chopina, Wydział Reżyserii Dźwięku,

Katedra Akustyki Muzycznej i Multimediów, ul. Okólnik 2, 00–368 Warszawa

piotr.rogowski@chopin.edu.pl

Streszczenie

Praca przedstawia wyniki pomiarów progu dyskryminacji wysokości tonów przeprowadzonych przy różnych częstotliwościach w przedziale 20–100 Hz oraz przy częstotliwościach 250 i 1000 Hz. Wyniki wskazują, że przy zmniejszaniu częstotliwości tonu próg zwiększa się od 0,6–0,8 Hz przy 250 Hz do ponad 3 Hz przy częstotliwości 20 Hz. Próg, wyrażony jako interwał w centach, znacząco wzrasta poniżej 250 Hz i przekracza półton przy częstotliwości 31,5 Hz. Między słuchaczami wystąpił znaczny rozrzut wyników. U niektórych badanych próg przekroczył 400 centów przy częstotliwości 20 Hz. Znaczący wzrost progu dyskryminacji wysokości przy najmniejszych częstotliwościach zakresu słyszenia przemawia na rzecz hipotezy wiążącej mechanizm kodowania częstotliwości dźwięku w ślimaku z położeniem krańcowego punktu wierzchołkowego zbocza obwiedni wychylenia błony podstawnej.

1. Wprowadzenie

Opublikowane badania wskazują, że próg dyskryminacji wysokości, zdefiniowany jako najmniejsza słyszalna różnica częstotliwości między dwoma tonami, wyrażona na skali liniowej w hercach, wynosi 0,7–1 Hz w przedziale częstotliwości 125–500 Hz i stopniowo zwiększa się do około 70–80 Hz przy częstotliwości 8000 Hz. Próg dyskryminacji wysokości, wyrażony jako interwał w centach, maleje od około 10 ct przy częstotliwości 125 Hz do około 3 ct w przedziale 1000–2000 Hz i wzrasta powyżej 2000 Hz do około 15 ct przy 8000 Hz [1, 2, 3]. Według nielicznych danych odnoszących się do tonów w przedziale poniżej 125 Hz próg dyskryminacji wysokości wynosi w tym przedziale około 1,0–1,3 Hz [4]. Obserwacja, że w liniowej skali częstotliwości próg dyskryminacji zmienia się w przedziale poniżej 500 Hz jedynie w niewielkim stopniu oznacza, że wartość progu, przedstawiona na skali logarytmicznej w centach, znacząco wzrasta przy zmniejszaniu częstotliwości tonu poniżej 500 Hz.

W niniejszej pracy zmierzono wartości progu dyskryminacji wysokości tonów o częstotliwościach pokrywających, w niewielkich odstępach, zakres zbliżony do dolnej granicy słyszenia. Badania składały się z dwóch eksperymentów przeprowadzonych różnymi metodami pomiaru progu, z użyciem podobnego zestawu tonów testowych.

Badania miały na celu: (1) oszacowanie rozdzielczości słuchu w dziedzinie wysokości dźwięku, w odniesieniu do tonów niskoczęstotliwościowych, (2) stwierdzenie, czy przy częstotliwościach zbliżonych do dolnej granicy zakresu słyszenia mierzone wartości progu są zgodne ze zgromadzoną w literaturze wiedzą na temat procesów dyskryminacji częstotliwości dźwięku w układzie słuchowym, (3) odniesienie wyników pomiaru progu dyskryminacji wysokości do zagadnień percepcji wysokości w najniższych oktawach muzycznej skali wysokości.

2. Metoda badań

2.1. Tony odtwarzane w eksperymentach

W eksperymencie I częstotliwości tonów pomiarowych wynosiły 20, 25, 31,5, 40, 63, 100, 250 i 1000 Hz. W eksperymencie II zestaw ten powiększono o częstotliwość 22,36 Hz. W obu eksperymentach, w przedziale częstotliwości 31,5–1000 Hz, poziom głośności tonów wynosił 40 fonów. W przedziale poniżej 31,5 Hz poziom głośności nieco zmniejszono by uniknąć odtwarzania tonów przy poziomach ciśnienia akustycznego powyżej 90 dB, przy których znacząco zwiększały się zniekształcenia dźwięku w słuchawkach. Poziomy ciśnienia akustycznego odpowiadające przy poszczególnych częstotliwościach poziomowi głośności 40 fonów określono na podstawie normy ISO 226:2003 [5].

Czas trwania tonów wynosił 1,5 s przy częstotliwości do 100 Hz włącznie i 750 ms powyżej 100 Hz. Obwiednie tonów, w stadium narastania i stadium zanikania, ukształtowano według funkcji Hanna. W przedziale 20–100 Hz, czas narastania i czas zanikania, mierzone w zakresie od 0 do 100% wartości sygnału w stanie ustalonym, wynosiły odpowiednio 50 i 150 ms. Przy częstotliwości 250 i 1000 Hz czasy narastania i czasy zanikania tonów były jednakowe i wynosiły 20 ms.

W obu eksperymentach tony prezentowane były w parach. W każdej parze jeden ton (standard) był bodźcem odniesienia o stałej w danej serii częstotliwości, a drugi bodźcem zmiennym o odstrojonej częstotliwości. Czas przerwy pomiędzy tonami w parze wynosił 500 ms, przy czym kolejność tonów ustalano w każdej parze losowo.

2.2. Aparatura

Generowanie i odtwarzanie tonów pomiarowych oraz rejestracja odpowiedzi słuchaczy sterowane były za pomocą oprogramowania komputerowego. Sygnały cyfrowe przetwarzano na postać analogową przy użyciu 24-bitowego przetwornika mAudio FireWire Audiophile wyposażonego we wzmacniacz słuchawkowy. Tony odtwarzano dwuusznie poprzez słuchawki nauszne zamknięte, Sennheiser HD-25-13-II.

By zapewnić pełną szczelność przylegania poduszek słuchawek do uszu położenie słuchawek było starannie dopasowywane do głów słuchaczy, zwłaszcza osób o szczupłej budowie czaszki. Słuchaczy pouczono by w przypadku, gdyby barwa dźwięku była ich zdaniem nienaturalna, przerywali sesję pomiarową i poprawili ułożenie słuchawek. W trakcie badań odnotowano kilka takich sytuacji przy odtwarzaniu tonu o częstotliwości 20 Hz.

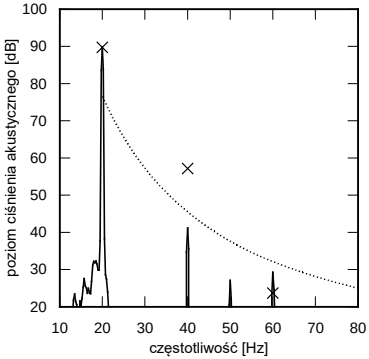
Charakterystyki częstotliwościowe słuchawek zostały zmierzone przy użyciu sztucznego ucha B&K 4153 wyposażonego w mikrofon ½" typu 4134. Wyniki pomiarów posłużyły do korekcji poziomów odtwarzanych tonów. Pomiary wykazały również, że w przypadku najniższych tonów (20,0–31,5 Hz) występowały znaczne zniekształcenia harmoniczne, przejawiające się głównie obecnością drugiej harmonicznej. Słuchacze mogliby zatem dokonywać ocen nie na podstawie różnicy wysokości składowych podstawowych, ale słyszalnych drugich harmonicznych. Problemu tego nie można było wyeliminować poprzez wybór innych słuchawek, ponieważ wśród dostępnych i przebadanych pod tym kątem przez autorów kilkunastu modeli słuchawki HD-25 charakteryzowały się największą stabilnością odtwarzania dźwięku w zakresie bardzo małych częstotliwości. Z tego względu, w zakresie częstotliwości 20,0–31,5 Hz zamiast tonów użyto dwutonów utworzonych poprzez dodanie do składowej podstawowej drugiej harmonicznej o poziomie i fazie tak dobranych, by skompensować zniekształcenia wprowadzane przez słuchawki. Tak przygotowane bodźce słuchacze zgodnie ocenili je jako bardziej „czyste" tonalnie od bodźców nieskorygowanych. Pomiary zniekształceń zmodyfikowanych bodźców wykonane przy użyciu sztucznego ucha potwierdziły, że poziomy harmonicznych nie przekraczają progu słyszenia (rys. 1).

2.3. Słuchacze

W badaniach wzięło udział sześcioro słuchaczy – dwie kobiety i czterech mężczyzn w wieku 23–25 lat. Wszyscy słuchacze byli studentami Wydziału Reżyserii Dźwięku UMFC i posiadali duże doświadczenie jako uczestnicy testów słuchowych. Jeden ze słuchaczy (LZ) dysponował słuchem absolutnym. Słuchacze zostali przebadani audiometrycznie przy użyciu audiometru Interacoustics AC40. U większości z nich, z wyjątkiem słuchacza (JN), nie odnotowano obniżenia progu słyszenia w pasmie 125–1500 Hz przekraczającego 4 dB względem audiometrycznego zera, przy czym progi słyszenia były na ogół lepsze niż przeciętne. W przypadku słuchacza JN próg słyszenia był obniżony do 6 dB HL w przedziale częstotliwości powyżej 500 Hz.

2.4. Procedura pomiaru progu dyskryminacji wysokości

Pomiary przeprowadzono metodą stałych bodźców w eksperymencie I i adaptacyjną metodą schodkową (ang. *adaptive, staircase up-down method*) w eksperymencie II [6]. W obu eksperymentach zadanie słuchacza polegało na wskazaniu, poprzez naciśnięcie odpowiedniego klawisza na klawiaturze komputera, wyższego tonu w odtwarzanej parze. Słuchacze zostali poinstruowani by w przypadkach, gdy nie zauważali różnicy wysokości między tonami, losowo wybierali odpowiedź.



Rysunek 1 Widmo bodźca pomiarowego o częstotliwości 20 Hz (linia ciągła) odniesione do progu słyszenia wg ISO 226 [4] (linia kropkowa). Krzyżykami oznaczono poziomy trzech pierwszych harmonicznych tonu bez kompensacji zniekształceń wprowadzanych przez słuchawki.

Czas na udzielenie odpowiedzi nie był ograniczony. Prezentacja kolejnego zadania rozpoczynała się automatycznie po upływie 1,5 s od udzielenia odpowiedzi, jednakże nie wcześniej niż 1,5 s od wybrzmienia ostatniego tonu w poprzedniej parze.

W eksperymencie I, przeprowadzonym metodą stałych bodźców, wprowadzono siedem poziomów odstrojenia, o które zwiększano częstotliwość bodźca standardowego. Odstrojenia dobrano indywidualnie przy każdej częstotliwości pomiarowej w eksperymencie wstępnym, który pełnił jednocześnie rolę treningu przed właściwymi pomiarami. Przy poszczególnych częstotliwościach bodźca standardowego odstrojenia wynosiły: częstotliwość **20 Hz**: 50, 100, … 350 ct, **25 Hz**: 30, 60, … 210 ct, **31,5 Hz**: 20, 45, 70, … 170 ct, **40 Hz**: 10, 28, 46, … 118 ct, **63 Hz**: 5, 15, 25, … 65 ct, **100 Hz**: 2, 7, 12, … 32 ct, **250 Hz**: 2, 3,5, 5… 11 ct, **1 kHz**: 1, 2, … 7 ct. Zestaw bodźców pomiarowych zawierał 14 par dźwięków – siedem poziomów odstrojenia bodźca zmiennego w dwóch wariantach kolejności prezentacji bodźców (standard jako pierwszy albo jako drugi w parze). Seria pomiarowa obejmowała prezentację dwóch zestawów bodźców – łącznie 28 zadań. Kolejność prezentacji par w serii pomiarowej była losowa. Eksperyment I składał się z 26 serii pomiarowych, przy czym wyniki pierwszej serii były odrzucane. Łącznie zgromadzono 100 odpowiedzi na każdy punkt pomiarowy.

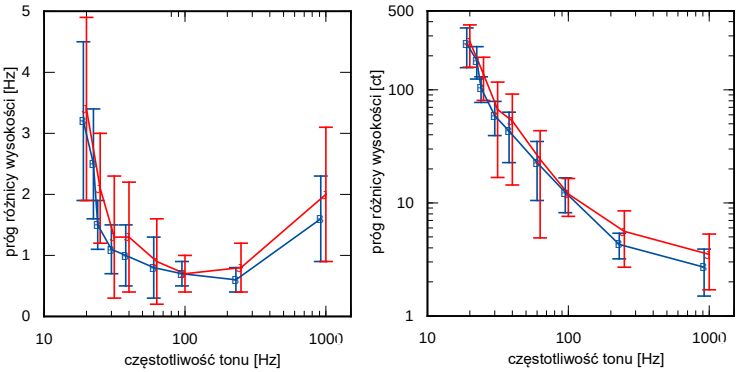
W eksperymencie II, adaptacyjną serię pomiarową rozpoczynano od wyraźnie słyszalnego odstrojenia częstotliwości. Odstrojenie zmniejszano o współczynnik skoku poziomu bodźca po trzech kolejnych poprawnych odpowiedziach słuchacza i zwiększano o ten sam współczynnik po każdej błędnej odpowiedzi. Zasada ta prowadzi do oszacowania progu na krzywej psychometrycznej przy poziomie 79,4% poprawnych odpowiedzi [6]. Współczynnik skoku poziomu bodźca czyli współczynnik, o jaki zmieniała się różnica częstotliwości między tonami w parze wynosił 1,58 (log 1,58 = 0,2) na początku serii. Po wystąpieniu w serii dwóch nawrotów poziomu bodźca, czyli dwukrotnego odwrócenia kierunku odstrojenia, współczynnik ten zmniejszano do 1,26 (log 1,26 = 0,1). Seria pomiarowa składała się z 60 zadań. Jako wynik pomiaru w serii przyjęto średnią geometryczną wyrażonego w centach interwału wysokości pomiędzy standardem i bodźcem zmiennym w tzw. punktach zwrotnych, czyli prezentacjach pary tonów, w których odwracał się kierunek zmiany odstrojenia częstotliwości z malejącego na rosnący lub odwrotnie. W obliczeniach wyniku nie brano pod uwagę pierwszych czterech punktów zwrotnych w serii. Każdy słuchacz wykonał 15 serii zadań przy każdej częstotliwości bodźca standardowego.

3. Wyniki

Na rys. 2 przedstawiono średnie arytmetyczne oraz odchylenia standardowe wyników pomiaru progu różnicy wysokości w eksperymencie I oraz w eksperymencie II, w grupie sześciorga słuchaczy. Dane uzyskane w eksperymencie I przedstawiają wartości progu odpowiadające punktowi 75% poprawnych odpowiedzi na krzywej psychometrycznej a dane z eksperymentu II są oszacowaniem progu w punkcie

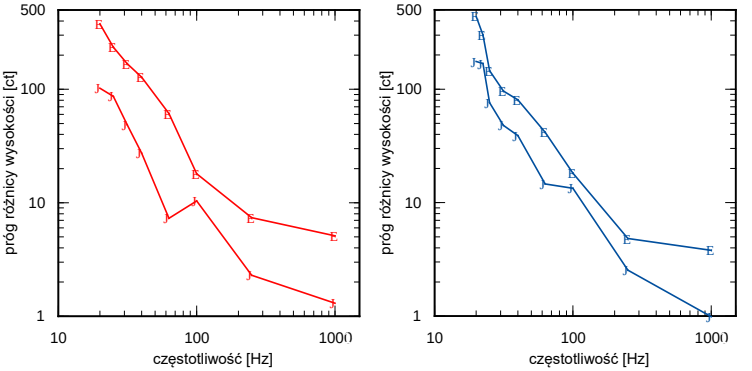
79,4% poprawnych odpowiedzi. Na wykresie po lewej stronie wartości progu przedstawiono w skali liniowej, w hercach, a na wykresie po prawej stronie – w skali logarytmicznej, jako interwał w centach.

Na wykresie po lewej stronie (rys. 2) widoczne jest, że w obu eksperymentach, próg dyskryminacji wysokości wynosił 0,6–1,5 Hz, w przedziale częstotliwości 31,5–500 Hz. W przedziale poniżej 31,5 Hz wartość progu znacząco wzrosła i przekroczyła 3 Hz przy częstotliwości 31,5 Hz. Z danych tych wynika, że próg wyrażony jako najmniejszy słyszalny interwał w centach (wykres po prawej stronie), bardzo znacząco zwiększa się przy zmniejszaniu częstotliwości tonu, od 12–15 ct przy częstotliwości 100 Hz do blisko 300 ct (trzech półtonów temperowanych) przy częstotliwości 20 Hz. Należy przy tym dodać, że zaobserwowano znaczną dyspersję wartości progu między słuchaczami. Na przykład, w eksperymencie I, najmniejsza osobnicza wartość progu dyskryminacji wysokości tonu o częstotliwości 20 Hz wyniosła 102 ct, a u pozostałych słuchaczy przekraczała 200 ct, sięgając 392 ct. W eksperymencie II, w tym samym punkcie pomiarowym, najmniejsza wartość progu wyniosła 175 ct a największa 439 ct.



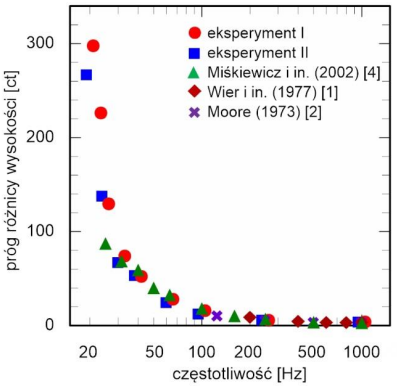
Rysunek 2 Wyniki pomiaru progu różnicy wysokości. Średnie i odchylenia standardowe w grupie słuchaczy w eksperymencie I (koła) i eksperymencie II (kwadraty). Na wykresie po lewej stronie oś rzędnych zwymiarowano w hercach a na wykresie po prawej stronie w centach.

Na rys. 3 przedstawiono osobnicze wyniki pomiaru progu dyskryminacji wysokości uzyskane w eksperymentach I i II dla słuchaczy, u których wystąpiły najniższe (LZ) i najwyższe (JN) wartości progu w największej liczbie punktów pomiarowych.



Rysunek 3 Osobnicze wyniki pomiaru progu różnicy wysokości słuchaczy LZ (koła zamknięte) i JN (koła otwarte). Na wykresie po lewej stronie przedstawiono wyniki uzyskane w eksperymencie I, a na wykresie po prawej stronie wyniki eksperymentu II.

Wykres na rys. 4 przedstawia średnie wartości progów wyznaczone w obu eksperymentach na tle danych z literatury [1, 2, 4]. Dane pochodzące z pracy [4] dotyczą zakresu częstotliwości od 25 Hz natomiast dane z pozostałych prac [1, 2] nie obejmują częstotliwości mniejszych niż 125 Hz.



Rysunek 4 Wyniki pomiaru progu różnicy wysokości (średnie grup słuchaczy) na tle danych z literatury. W celu zwiększenia czytelności wykresu symbole odnoszące się do danych z eksperymentów I i II rozsunięto względem siebie wzdłuż osi odciętych.

4. Omówienie wyników i wnioski

Z danych zestawionych na rys. 4 wynika, że wartości progu dyskryminacji wysokości wyznaczone dla grup słuchaczy w obecnej pracy są zgodne z wartościami progu zebranymi z literatury. Obecne wyniki są potwierdzeniem obserwacji uzyskanych w pracy Miśkiewicza i in. [4] świadczących o tym, że próg dyskryminacji wysokości tonów, wyrażony w centach, bardzo znacznie zwiększa się w przedziale poniżej 250 Hz. W odniesieniu do wyników pracy [4] nasuwało się przypuszczenie, że jedną z przyczyn gorszej dyskryminacji wysokości mogło być rozmycie widma tonów niskoczęstotliwościowych o czasie trwania 500 ms. W obecnej pracy wykluczono ten czynnik poprzez zastosowanie w pasmie poniżej 100 Hz dłuższych tonów, o czasie trwania 1500 ms.

Z badań przeprowadzonych z użyciem tonów o częstotliwości 125–8000 Hz wynika, że próg, wyrażony w hercach, monotonicznie zwiększa się z częstotliwością tonu [1, 2, 3]. Sęk i Moore [3] zauważyli, że iloraz progu dyskryminacji wysokości i częstotliwości tonu pomiarowego osiągał przy częstotliwości 500 Hz minimum wynoszące 0,002. W przedziale powyżej 500 Hz iloraz ten monotonicznie zwiększał się do blisko 0,03 przy częstotliwości 8000 Hz. Przy częstotliwości 250 Hz, jedynej użytej w eksperymencie w zakresie poniżej 500 Hz [3], iloraz progu dyskryminacji i częstotliwości tonu pomiarowego wynosił 0,04. Podobny, ogólny przebieg miała zależność ilorazu progu i wskaźnika ERB – odpowiednika częstotliwości na skali będącej odwzorowaniem szerokości pasm pasmowoprzepustowych filtrów słuchowych [3]. Wyniki obecnego eksperymentu poszerzają opisane powyżej obserwacje o zakres poniżej 250 Hz. Przy zmniejszaniu częstotliwości poniżej 250 Hz, próg różnicy wysokości, wyrażony jako interwał częstotliwości na skali logarytmicznej w centach, bardzo znacznie zwiększa się (rys. 4).

W interpretacji wyników obecnej pracy podstawowe znaczenie ma odniesienie zmierzonych wartości progu do teorii opisujących mechanizmy dyskryminacji częstotliwości dźwięku w układzie słuchowym. Według współczesnej wiedzy podstawą dyskryminacji częstotliwości są złożone procesy obejmujące zjawiska opisywane w teoriach słyszenia, tradycyjnie dzielonych na teorie miejsca i teorie czasowe. Jedna z hipotez, sformułowana przez Zwisłockiego [7] zakłada, że podstawą mechanizmu kodowania częstotliwości tonu pobudzającego jest położenie na błonie podstawnej krańcowego punktu wierzchołkowego zbocza obwiedni wychylenia spowodowanego pobudzeniem. Wyniki niniejszej pracy przemawiają na rzecz tej hipotezy ponieważ przy częstotliwości poniżej około 200 Hz błona podstawna wychyla się na całej długości i krańcowy punkt zbocza obwiedni nie zmienia swego położenia przy

zmianie częstotliwości tonu pobudzającego. Mechanizm postulowany przez Zwisłockiego [7] nie może zatem działać i częstotliwość tonu kodowana jest jedynie poprzez synchronizację potencjałów czynnościowych z fazą tonu pobudzającego, w sposób opisany w teorii czasowej [8].

Wyniki badań przeprowadzonych w niniejszej pracy dostarczyły danych, które odnoszą się również do zagadnień percepcji dźwięku w muzyce. Fakt, że próg różnicy wysokości tonów, wyrażony jako interwał na logarytmicznej skali częstotliwości, znacznie zwiększa się w przedziale poniżej 100 Hz jest zgodny z wynikami badań w których oszacowano, że najniższa słyszalna wysokość muzyczna odpowiada częstotliwości około 30 Hz [9]. Obserwacje uzyskane w niniejszej pracy mają odniesienie również do percepcji interwałów muzycznych. W eksperymencie przeprowadzonym z użyciem dwutonów tworzących interwały muzyczne stwierdzono, że muzycy popełniają znacznie więcej błędów w identyfikacji interwałów w oktawie wielkiej (zakres częstotliwości 65,4–130,8 Hz) niż w oktawie razkreślnej (261,6–523,3 Hz) i oktawie trzykreślnej (1047–2093 Hz).

Obserwacją, której wyjaśnienie wymaga dalszych badań jest bardzo duże zróżnicowanie osobnicze progu dyskryminacji wysokości tonów niskoczęstotliwościowych uzyskanych w grupie słuchaczy z wykształceniem muzycznym, w tej samej grupie wiekowej. Jedną z potencjalnie możliwych przyczyn dyspersji wyników są szczegółowe różnice mające związek z procesami poznawczymi percepcji dźwięku, wynikające z posiadanej przez słuchaczy praktyki w wykonywaniu muzyki. Słuchacz LZ, u którego stwierdzono najmniejsze wartości progu dyskryminacji wysokości, był w grupie badanych jedynym muzykiem grającym na instrumencie o swobodnym stroju, którego skala, rozpoczynająca się od dźwięku E_1 (41,2 Hz), obejmuje znaczną część zakresu częstotliwości tonów użytych w eksperymencie.

Bibliografia

[1] C. C. Wier, W. Jesteadt, D. M. Green: „Frequency discrimination as a function of frequency and sensation level", w: *Journal of the Acoustical Society of America* 61 (1), 1977, s. 178–184.

[2] B. C. J. Moore: "Frequency-difference limens for short duration tones" w: *Journal of the Acoustical Society of America* 54 (3), 1973, s. 610–619.

[3] A. Sęk, B. C. J. Moore: „Frequency discrimination as a function of frequency, measured in several ways", w: *Journal of the Acoustical Society of America* 97 (4), 1995, s. 2479–2486.

[4] A. Miśkiewicz, A. Rakowski, P. Rogoziński, M. Kocańda: „Progi różnicowe częstości i progi różnicowe wysokości dźwięków muzycznych", w: *Kształtowanie i percepcja sekwencji dźwięków muzycznych*, A. Rakowski (red.), Wydawnictwo Akademii Muzycznej im. Fryderyka Chopina, Warszawa, 2002, s. 125–154.

[5] ISO 226: 2003(E): *Acoustics—Normal equal-loudness-level contours*, 2003.

[6] H. Levitt: „Transformed up-down methods in psychoacoustics", w: *Journal of the Acoustical Society of America* 49, 1971, s. 467–477.

[7] J. J. Zwislocki: „What is the cochlear code for pitch?", w: *Acta Otolaryngologica* (Stockholm) 111 (2), 1991, s. 256–262.

[8] A. R. Palmer, I. J. Russel: „Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of the inner hair-cells", w: *Hearing Research* 24 (1), 1986, s. 1–15.

[9] D. Pressnitzer, R. Patterson, K. Krumbholtz: „The lower limit of melodic pitch", w: *Journal of the Acoustical Society of America* 109 (5 Pt.1), 2001, s. 2074–2084.

[10] T. Rogala, A. Miśkiewicz, P. Rogowski: „Identification of harmonic musical intervals: The effect of pitch register and tone duration", w: *Archives of Acoustics* 42 (4), 2017, s. 591–600.

W JAKI SPOSÓB BODŹCE AKUSTYCZNE MOGĄ WPŁYWAĆ NA ROZWÓJ I ZACHOWANIE DZIECI?

A. Kosiek[1]

[1]Akademia Górniczo – Hutnicza im. Stanisława Staszica w Krakowie

kosiek@agh.edu.pl

Streszczenie

Badania polegały na poznaniu związków pomiędzy bodźcami dźwiękowymi i ich wpływem na rozwój i zachowanie małych dzieci. Przeprowadzone badania oraz obserwacje pokazały, że dla dzieci szczególnie rozpraszające są te bodźce, które wyjątkowo im się podobają. Dodatkowo jeśli dla dzieci jakaś muzyka lub hałas nie są ani szczególnie interesujące ani szczególnie nieprzyjemne, to mają one dużo mniejszy wpływ na poprawność i szybkość wykonywania zadań w stosunku do odgłosów pochodzących od innych dzieci. Kolejną ciekawą obserwacją jest to, że dzieci podczas wykonywania zadań prawie w ogóle nie zwracają uwagi na hałas drogowy lub domowy. Można wysunąć pewien ogólny wniosek mówiący o tym, że w przedszkolach, w których dzieci zachowują się spokojniej i bawią się bez robienia niepotrzebnego hałasu wyniki są lepsze, dzieci odpowiadają szybciej na pytania oraz potrafią skupić się na nauce oraz wykonywanym zadaniu znacznie dłużej.

1. Wstęp

Obecność dźwięków, zarówno tych, na których obecność nie mamy wpływu, np. hałas przejeżdżających za oknem samochodów czy szczekanie naszego psa, jak i tych, które sami „wygenerowaliśmy" w postaci włączonej płyty czy odkurzacza nie są dla nas bez znaczenia. Każdy bodziec dźwiękowy może wpływać na poprawę lub osłabienie koncentracji, na jakość snu czy poziom naszego stresu [1]. Temat ten jest szczególnie istotny w przypadku dzieci, dla których sen jest wyjątkowo ważny dla prawidłowego rozwoju, a ich układ nerwowy dopiero się kształtuje. Co więcej, małe dzieci niemal przez cały czas uczą się czegoś nowego. Obecnie w literaturze możemy znaleźć już pewne informacje na temat wpływu dźwięków na rozwój dzieci [2] oraz ich zachowanie, ale wciąż istnieje wiele niezbadanych obszarów. Z tego powodu podjęłam próbę znalezienia pewnych zależności pomiędzy dźwiękami, które otaczają dzieci (zarówno takimi, w obecności których przebywają w nocy, jak i takich, które towarzyszą im w tle podczas wykonywania pewnych zadań) a zachowaniem i rozwojem dzieci oraz przyswajaniem przez nie wiedzy.

Badania zostały podzielone na trzy główne obszary. Pierwszym z nich był szczegółowy wywiad z rodzicami i nauczycielami na temat warunków akustycznych w jakich dzieci przebywają na co dzień oraz zachowania tych dzieci i tempa przyswajania przez nie nowych umiejętności. Celem wywiadu było znalezienie bodźców dźwiękowych, które mogą mieć istotny wpływ na zachowanie i rozwój intelektualny dzieci. Drugim etapem było wykonanie przez dzieci trzech typów zadań w obecności ośmiu różnych rodzajów dźwięków oraz w ciszy. Ostatnim etapem był pomiar poziomu natężenia dźwięku w przedszkolach, do których uczęszczają badane dzieci oraz poznanie dominujących tam częstotliwości. Etap drugi i trzeci zostały powtórzone dwukrotnie w celu weryfikacji uzyskanych wyników.

W badaniach zdecydowałam się na przeanalizowanie wpływu dźwięków na wykonywanie zadań pamięciowych, z zakresu arytmetyki oraz słuchania ze zrozumieniem. Obecnie w psychologii badania nad Irrelevant Sound Effect, czyli wpływem nieistotnych dźwięków w tle na wykonywane przez człowieka zadania [3] skupiają się głównie na zagadnieniach związanych z zapamiętywaniem. Niewiele jest natomiast badań, w których sprawdzano wpływ nieistotnych dźwięków na wykonywanie innych typów zadań np. z zakresu arytmetyki lub rozumienia tekstu. Zbadanie wpływu bodźców akustycznych również na te czynności jest istotna m.in. ze względu na fakt, że są to typowe zadania wykonywane w różnych miejscach pracy, a także w szkołach i przedszkolach. Zazwyczaj badania z zakresu psychoakustyki przeprowadzane są w specjalnych pomieszczeniach. Często są one dźwiękoszczelne, a uczestnicy badań dźwięki mają dostarczane przez słuchawki. Z jednej strony takie rozwiązanie pozwala na bardzo dobre

wyeliminowanie zmiennych zakłócających w postaci m.in. innych dźwięków, rozpraszaczy wzrokowych itp. Jednak stwarza to warunki niemal całkowicie nienaturalne – nikt z nas przecież na co dzień nie jest całkowicie odizolowany od reszty świata i nie ma do czynienia jedynie z jednym bodźcem. Bardzo często bodźce zakłócające są krótkie albo przerywane innymi bodźcami i niemal stale się mieszają [4]. W związku z tym pojawia się ryzyko, czy jeśli stworzymy zupełnie nierealne na co dzień warunki w trakcie badań, to czy ich wyniki będą miały jakiekolwiek zastosowanie w praktyce. Oczywiście konieczne jest wyeliminowanie możliwie wielu zmiennych zakłócających, jednak w moich badaniach zdecydowałam się na przeprowadzenie badań w warunkach naturalnych dla dzieci, czyli w przedszkolach. Dodatkowo warto mieć na uwadze, że w przypadku badań w specjalnie przygotowanym laboratorium pojawia się także zmienna zakłócająca w postaci zainteresowania uczestnika badań nowym miejscem. Może to być wyjątkowo dużą przeszkodą w przypadku badań z udziałem dzieci, dla których nowe miejsce może wręcz uniemożliwić skupienie się na wykonywanym zadaniu. Biorąc pod uwagę dziecięcą ciekawość świata i fakt, że nawet sama moja obecność w ich sali wzbudzała ogromne zainteresowanie i poruszenie przypuszczam, że przeprowadzenie badań w całkowicie nowych warunkach wiązałoby się z dużymi problemami ze skupieniem się na wykonywanych zadaniach, a w przypadku niektórych dzieci byłoby wręcz niemożliwe.

2. Projektowanie badań

Projektując badania z zakresu psychofizyki należy pamiętać, że nie mogą trwać one zbyt długo, ponieważ jeśli za bardzo zmęczymy lub znudzimy uczestnika, to w pewnym momencie przestanie on rzetelnie udzielać odpowiedzi. Będzie chciał jedynie jak najszybciej skończyć, co spowoduje, że będzie odpowiadał niedokładnie, losowo albo bez wystarczającego zastanowienia przez co uzyskane wyniki nie będą wiarygodne i będzie trzeba je powtórzyć, co często jest niemożliwe przy udziale małych dzieci, ponieważ nie będą chciały wykonać ponownie tego samego ćwiczenia. Ze względu na fakt, że dzieci szczególnie szybko się męczą, a przede wszystkim szybko się nudzą zadania, które miały wykonać nie tylko musiały pozwolić na poznanie odpowiedzi na postawione pytania, lecz także musiały być tak dobrane, żeby dzieci chciały wziąć udział w badaniu, dotrwały do jego końca oraz udzieliły konkretnych odpowiedzi. W związku z tym jedno ćwiczenie, które miały wykonać dzieci trwało nie więcej niż 30/40 minut łącznie z częścią organizacyjną. Z tego względu w testach ograniczyłam się jedynie do ośmiu rodzajów dźwięków obecnych w tle podczas wykonywanych zadań. Zdecydowałam się na dwa najczęściej występujące na co dzień rodzaje hałasu (domowy i drogowy), ludzką mowę oraz 5 rodzajów muzyki (muzykę klasyczną, jazzową, współczesną muzykę z tekstem, współczesną muzykę instrumentalną oraz rock). Dodatkowo jeden pomiar referencyjny przeprowadziłam bez żadnych dodatkowych dźwięków w tle. Niestety większa liczba dźwięków za bardzo wydłużyłaby badanie. Poza tym rodzaj ćwiczenia musi być interesujący dla dziecka. Z tego powodu chcąc sprawdzić jak różne dźwięki wpływają na poprawność liczenia przez dzieci zamiast prosić je o kilkukrotne policzenie kilku elementów, prosiłam je o kolorowanie odpowiedniej ilości elementów, co dzieciom wydało się bardziej atrakcyjne.

W trakcie badań pojawiły się pewne problemy związane głównie z faktem, że uczestnikami były czteroletnie dzieci, które łatwo się rozpraszają, nie zawsze chcą współpracować, a czasem niektóre z nich celowo przeszkadzają innym w wykonywaniu zadań przez co cała grupa nie jest w stanie się skupić. Z tego powodu warto zadbać o dobrą relację z dziećmi biorącymi udział w badaniu, gdyż wtedy są bardziej chętne do współpracy i nie przeszkadzają celowo podczas pracy. Kilka razy zdarzyło się również, że badanie należało przerwać i kontynuować w inny dzień lub po przerwie na poruszanie się, lub zabawę. Z tego samego powodu dobrym rozwiązaniem było realizowanie kolejnych części badań z minimum dwutygodniowymi przerwami. Niestety powoduje to, że badania znacznie rozciągają się w czasie, ale za to dzieci są bardziej chętne do współpracy. Z otrzymanych wyników można zaobserwować, że w tych przedszkolach, w których jest najgłośniej ze względu na dokazywanie przez dzieci badania zajęły najwięcej czasu, a czasami nie udało się doprowadzić badania do końca.

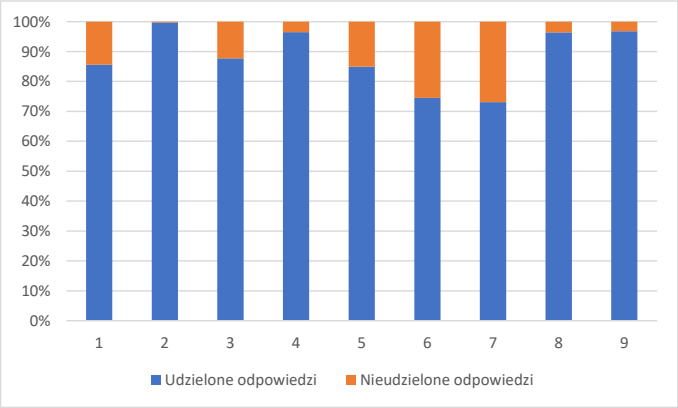
Planując badania z zakresu psychoakustyki, w których udział będą brać dzieci (zwłaszcza małe) warto również wziąć pod uwagę, że część z nich będzie oszukiwać albo zwyczajnie odmówi wzięcia udziału

w badaniu. Dlatego też wskazane jest stworzenie grupy badawczej o większej liczebności niż potrzeba nam wyników. W przypadku moich badań około 5% wyników należy z tego powodu odrzucić. Chociaż badania z udziałem dzieci mogą wydawać się bardziej skomplikowane i wymagające, to są one niezwykle ważne i potrzebne, ponieważ na chwilę obecną większość badań z zakresu psychoakustyki skupia się na osobach w przedziale wiekowym od 18 do 55 roku życia, zdarzają się również badania z udziałem nastolatków, natomiast badań z udziałem małych dzieci jest niezwykle mało. Są one tak istotne, ponieważ dzieci często odbierają bodźce dźwiękowe nieco inaczej niż dorośli i nastolatkowie oraz inaczej na nie reagują.

3. Obserwacje

Na tym etapie badań udało mi się zaobserwować kilka istotnych tendencji, które wpływają na wynik moich badań. Przede wszystkim dzieci mają spore problemy ze skupieniem się na wykonywanej czynności, gdy muzyka, dźwięk, które słyszą bardzo im się podobają. W takiej sytuacji skupiają się one głównie na atrakcyjnych dla nich bodźcach – tańczą w rytm muzyki, podskakują, śpiewają, nucą, chcą wiedzieć jaki jest tytuł lub wykonawca danego utworu albo jaki to dźwięk, skąd pochodzi. Zadanie, które miały wykonać niestety schodzi na dalszy plan. Dzieci albo całkowicie rezygnują z robienia go albo wykonują je źle, niedokładnie albo wykonują je z dużym opóźnieniem lub zajmuje im znacznie więcej czasu niż normalnie.

Kolejną rzeczą, którą można zauważyć jest fakt, że jeśli dla dzieci dana muzyka lub hałas nie są ani szczególnie interesujące ani szczególnie nieprzyjemne, to mają one dużo mniejszy wpływ na poprawność i szybkość wykonywanych zadań w stosunku do odgłosów pochodzących od innych dzieci (ich rozmowy, wzajemne zaczepianie się, krzyki, czasem płacz). Być może wynika to z tego, że dźwięki w tle w postaci muzyki lub hałasu pochodzącego z ulicy są o wiele mniej zmienne niż dźwięki od grupy przedszkolaków. Odgłosy grupy są również często nagłe, niespodziewane i o bardzo wysokich poziomach natężenia. Oczywiście nie na każde dziecko takie dźwięki będą tak samo silnie oddziaływać, niektóre z nich będą na takie dźwięki dosyć odporne lub będą do nich już przyzwyczajone, ale takich dzieci jest zdecydowanie mniej. Część dzieci jest tak podatna na rozpraszanie ze strony kolegów i koleżanek, że nawet może całkowicie zaprzestać wykonywania danego zadania. Zostało to zaprezentowane na rysunku 1.



Rysunek 1 Procentowa zależność udzielonych i nieudzielonych odpowiedzi w przedszkolach o kolejnych numerach

W przedszkolach nr 5, 6 oraz 7 dzieci rozpraszały się wzajemnie wyjątkowo często – mówiły coś głośno, dokazywały, czasami zaczepiały się wzajemnie - i to właśnie w tych przedszkolach części najczęściej rezygnowały z wykonania zadań lub nie mogły się na nim skupić i w konsekwencji nie udzielały

żadnej odpowiedzi. Natomiast w przedszkolach nr 2 i 9, w których jest bardzo spokojnie oraz w przedszkolach nr 4 i 8, gdzie sporadycznie ktoś zakłócał spokój praktycznie nie zdarza się, żeby pytania pozostały bez odpowiedzi. Niestety to wszystko są bodźce niezwykle trudne do wyeliminowania. O ile jesteśmy w stanie wyłączyć muzykę, która brzmi w tle i przeszkadza dziecku w nauce lub odpoczynku albo możemy zamknąć okno, gdy hałas dobiegający z ulicy jest zbyt duży, to praktycznie niemożliwe jest, aby dzieci w przedszkolu wzajemnie sobie nie przeszkadzały w trakcie nauki. Zawsze znajdzie się ktoś, kto będzie o coś głośno pytał, ktoś inny skończy wcześniej i pójdzie się bawić (niekoniecznie będzie to cicha zabawa), jeszcze ktoś inny zacznie dokazywać lub będzie coś krzyczał do kogoś. Przy licznej grupie przedszkolaków bardzo ciężko jest to opanować. W efekcie stworzy się mieszanina różnych bodźców – nie tylko dźwiękowych – które będą stanowiły silny efekt rozpraszający dla dzieci. Zdarza się również, że nawet jeśli dziecko cały czas próbuje się skupić na wykonywaniu zadania i ignoruje otoczenie, to może po prostu nie usłyszeć tego, co jest niezbędne do wykonania zadania. Oczywiście w zależności od tego, jak grzeczne i zdyscyplinowane dzieci należą do danej grupy efekt ten będzie albo bardzo duży albo mało zauważalny, ale zawsze trzeba wziąć go pod uwagę. Dlatego też przeprowadzanie badań z dziećmi indywidualnie mogłoby mieć niewielki sens, ze względu na fakt, że stworzylibyśmy warunki dalekie od rzeczywistych.

Dodatkowo warto podkreślić, że liczba nieudzielonych odpowiedzi na zadane przeze mnie w trakcie badań pytania jest największa, gdy rozpraszającym bodźcem dźwiękowym jest mowa ludzka. Obserwacja ta jest zgodna z wynikami uzyskanymi podczas badań nad Irrelevant Sound Effect [3].

W trakcie pomiaru poziomu natężenia dźwięku wyznaczone zostały średnie wartości Leq (ekwiwalentnego poziomu dźwięku) w każdym z przedszkoli. We wszystkich przedszkolach poziom ten mieści się w zakresie 74 – 79 dB. Szczegółowe wyniki przedstawione zostały w tabeli 1.

Tabela 1 Średnie wartości poziomu Leq w przedszkolach

Przedszkole	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Leq [dB]	74,2	74,3	76,7	79,0	76,1	76,9	76,1	76,7	77,2

Jednak co zaskakujące chociaż poziomy te są zbliżone do siebie, to zarówno dzieci, jak i dorośli przebywający w przedszkolach są wrażliwi nie tylko na sam poziom natężenia dźwięku, lecz także przede wszystkim na to skąd ten dźwięk pochodzi. Odgłosy rozmów, zabawy czy muzyka włączona w tle nie działa na dzieci szczególnie rozpraszająco, ale gdy pojawiają się niespodziewane krzyki, hałas rozrzucanych zabawek lub alarm samochodu za oknem dzieci od razu przerywają swoje zajęcia i skupiają się na nieoczekiwanym dźwięku. Ponadto zdarza się, że niektóre dzieci na takie nagłe dźwięki reagują bardzo nerwowo i stają się niespokojne. Jest to ściśle powiązane z wynikami przedstawionymi na wykresie 1. W przedszkolach, w których dzieci często krzyczą bez powodu lub zaczepiają innych podczas nauki, więcej pytań w trakcie badania pozostało bez odpowiedzi. Powodem było to, że dzieciom w obecności nagle pojawiających się rozpraszających dźwięków znacznie trudniej było skupić się na teście.

Kolejną ciekawą obserwacją jest to, że dzieci prawie w ogóle nie zwracają uwagi na hałas drogowy lub domowy. Być może wynika to z faktu, że otoczenie z roku na rok staje się coraz głośniejsze, po drogach jeździ coraz więcej samochodów, w domach bardzo często mamy uruchomione coraz liczniejsze sprzęty, słuchamy radia, oglądamy telewizję. Coraz mniej czasu udaje nam się spędzić w ciszy lub jedynie przy dźwiękach o niskich poziomach natężenia. To powoduje, że dzieci do hałasu są po prostu przyzwyczajone, często nie znają nawet życia w ciszy. Niestety życie w ciszy jest obecnie niemal niemożliwe. Dla nich bardzo często jest to naturalne i nie rozprasza ich, nie przeszkadza w nauce. Oczywiście są również dzieci, które potrzebują wyciszenia, hałas je męczy i najlepiej żyje im się i uczy w ciszy, ale jest ich stosunkowo niewiele. Natomiast prawdopodobnie ma to wpływ na jakość ich wypoczynku czy układ nerwowy.

4. Wnioski

Porównując wyniki uzyskane przez dzieci w testach, można wysunąć pewien ogólny wniosek mówiący o tym, że w przedszkolach, w których dzieci zachowują się grzeczniej, spokojniej i bawią się bez robienia niepotrzebnego hałasu (np. nie krzyczą bez powodu) wyniki są zdecydowanie lepsze, dzieci odpowiadają szybciej i chętniej na pytania, a dodatkowo potrafią się skupić na nauce oraz wykonywanym zadaniu znacznie dłużej (około dwa razy dłużej). Oczywiście w każdym przedszkolu jest gwar, jednak w niektórych z nich są to jedynie odgłosy zabawy lub nakładające się na siebie rozmowy wielu dzieci jednocześnie, ale nie są to wyłącznie zbędne krzyki. O ile ten typowy przedszkolny „hałas" nie stanowi dla dzieci problemu, nie jest szczególnie rozpraszający i jest odbierany przez nie jako coś normalnego w przedszkolu (tak samo jak np. odgłos samochodów na ulicy lub śpiew ptaków w parku) i nie zwracają na niego uwagi, to częste krzyki, piski lub inne dźwięki wydawane przez niektóre dzieci w niespodziewanych momentach są dość mocno uciążliwe, powodują problemy ze skupieniem się na wykonywanych zadaniach oraz bywają na tyle męczące, że dzieci same się na nie skarżą.

Otrzymane wyniki pokazują jak istotne dla dzieci są warunki akustyczne w jakich przebywają, zwłaszcza jeśli chodzi o bodźce dźwiękowe, które pojawiają się nagle lub są w jakiś sposób nietypowe. Dlatego ważne jest, aby zapewnić im warunki „stabilne" pod względem akustycznym. Oczywiście nie da się wyeliminować wszystkich nieoczekiwanych dźwięków takich jak dzwonek do drzwi, klakson przejeżdżającego samochodu czy krzyk wystraszonego kolegi, jednak dobrym rozwiązaniem mogłoby się okazać np. zapewnienie odpowiedniej adaptacji akustycznej w przedszkolach, dzięki czemu te nagłe dźwięki byłyby chociaż trochę wytłumione.

Podziękowania

Badania były częściowo finansowane z subwencji badawczej Wydziału Inżynierii Mechanicznej i Robotyki AGH, nr: 16.16.130.942 oraz z projektu Zintegrowany Program Rozwoju Akademii Górniczo-Hutniczej w Krakowie II, nr: POWR.03.05.00-00-Z309/18-00

Bibliografia

1. C. M. T. Tiesler, „Exposure to road traffic noise and children's behavioural problems and sleep disturbance: Results from the GINIplus and LISAplus studies", w: *Environmental Research* 123, 2013, pp. 1–8.
2. J. Saffran, „Background noise may hinder toddlers' ability to learn words", w: *The ASHA Leader* 21 (10), 2016.
3. W. Ellermeier, K. Zimmer, „The psychoacoustics of the irrelevant sound effect", w: *Acoust. Sci & Tech.* 35 (1), 2014, pp. 10-16.
4. S. Banbury, D. C. Berry, „Disruption of office-related tasks by speech and office noise", w: *British Journal of Psychology* 89 (3), 2011, pp. 499-517.

WPŁYW MODYFIKACJI FAZY SYGNAŁU NA PERCEPCJĘ POGŁOSOWOŚCI W NAGRANIACH AUDIO

T. Makuch[1] i P. Kleczkowski[1]

[1]Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie, Katedra Mechaniki i Wibroakustyki

tmakuch@agh.edu.pl

Streszczenie

Zależności fazowe pomiędzy składowymi częstotliwościowymi fali akustycznej są w kontekście percepcji dźwięku zazwyczaj kojarzone przede wszystkim z barwą dźwięku oraz lokalizacją źródła fali o niskiej częstotliwości. Istnieją jednak badania sugerujące, że zmiany fazy w czasie mogą wpływać na postrzeganie pogłosowości. W referacie przedstawiony zostanie pierwszy etap badań nad wywoływaniem wrażenia pogłosowości przez losowe zmiany fazy składowych sygnału. W nagraniach bezechowych zmodyfikowano fazę, w ramach przyjętych ograniczeń. Następnie przeprowadzono nieformalne i formalne testy słuchowe. Rozważono kilka aspektów modyfikacji, takich jak szerokość pasma, zakres zmian fazy oraz częstość ich występowania; niektóre z nich mają istotny wpływ na pogłosowość.

1. Wprowadzenie

Powszechnie stosowana w akustyce analiza widmowa dźwięku dostarcza informacji o amplitudach poszczególnych składowych oraz zależnościach fazowych pomiędzy nimi. Jednak zdecydowanie więcej uwagi poświęcane jest amplitudzie i energii sygnału (zazwyczaj mówimy o sygnale napięciowym, zarejestrowanym przez przetwornik akustoelektryczny), niż fazie. Zależności fazowe są uwzględniane m.in. przy projektowaniu filtrów oraz układów elektroakustycznych – w tych przypadkach dąży się do osiągnięcia liniowej charakterystyki fazowej, aby możliwie zredukować zniekształcenia fazowe oraz wprowadzenie zakolorowania dźwięku.

Zależności fazowe są pomijane często w akustyce pomieszczeń – do jej oceny stosowane są głównie metody energetyczne, zarówno podczas pomiarów, jak i w symulacjach (metody geometryczne, analiza modalna). Należy jednak pamiętać, że jeden z kluczowych parametrów akustycznych powierzchni, fizyczny współczynnik pochłaniania dźwięku (α_s), wywodzi się ze współczynnika odbicia fali R, który z kolei ma bezpośredni związek z impedancją akustyczną powierzchni [1]. Ta zaś jest wielkością zespoloną, z członem amplitudowym i fazowym. W chwili odbicia fali dźwiękowej od powierzchni następuje zmiana fazy sygnału; wartość wprowadzanego przesunięcia zależna jest od parametrów materiału, częstotliwości fali oraz kąta padania [2]. W polu pogłosowym przyjmuje się, że każdy kierunek padania fali jest jednakowo prawdopodobny; co więcej, często powierzchnie pochłaniające i odbijające przybliża się odpowiednio jako idealnie podatne i idealnie sztywne. Efektem tych przybliżeń jest założenie o pomijalności zmiany fazy przy odbiciu; prowadzone są jednak badania związane np. z uwzględnieniem fazy np. w metodzie promieniowej [3]. W połączeniu z rosnącymi mocami obliczeniowymi być może pozwoli to na korzystanie z dokładniejszych symulacji, uwzględniających także zjawiska fazowe.

Jednym z argumentów pomijania zmian fazy jest ich znikoma zauważalność przez człowieka – zgodnie z obecnym stanem wiedzy dla ludzkiego ucha bardziej zauważalne są zmiany sygnału związane z amplitudą składowych, mniej z ich przesunięciami w czasie. Najlepiej zbadanym z mechanizmów percepcyjnych związanych z zależnościami fazowymi jest międzyuszna różnica czasów/faz (ITD, ang. *interaural time difference*), którą możemy rozpatrywać tylko w kontekście pary sygnałów. Na podstawie różnicy w czasie dotarcia sygnału do lewego i prawego ucha układ słuchowy określa kierunek (w płaszczyźnie horyzontalnej), z którego następuje imisja dźwięku. Mechanizm ITD jest najskuteczniejszy dla fal sinusoidalnych o częstotliwościach poniżej 1500 Hz; działa też w przypadku dźwięków złożonych, zawierających wyższe

częstotliwości – wówczas jego skuteczność zależy od struktury czasowej (obwiedni czasowej) dźwięku [4]. Poza tym aspektem percepcyjnie zjawiska fazowe przekładają się najbardziej na odbiór barwy dźwięku; w zależności od struktury dźwięku, przesunięcia między jego składowymi mogą być postrzegane jako szorstkość lub wywoływać wrażenie obecności dodatkowego sygnału tonalnego w widmie harmonicznym [5].

W kontekście odbioru przestrzeni dźwiękowej zależności fazowe nie są zbyt dobrze zbadane. Na podstawie subiektywnych eksperymentów D. Griesingera można wysnuć hipotezę, że relacje fazowe pomiędzy składowymi dźwięku o strukturze harmonicznej mają wpływ na to, czy jest on postrzegany jako pogłosowy, odległy od słuchacza [6, 7]. Wiąże się to z oceną pola akustycznego na podstawie stosunku dźwięku bezpośredniego i pogłosowego (DRR, ang. *direct to reverberant ratio*); zgodnie z tą teorią, w dźwięku pogłosowym pierwotne relacje między harmonicznymi są zaburzone na skutek odbić, co wywołuje wrażenie „rozmytego"dźwięku, utratę klarowności.

Przedmiotem pracy opisanej w niniejszym artykule jest zbadanie, czy randomizacja fazy sygnału faktycznie wywołuje wrażenie większego pogłosu. W tym celu w nagraniach dźwiękowych dokonano modyfikacji fazy bez ingerencji w amplitudę sygnału, a następnie przeprowadzono formalne testy odsłuchowe z udziałem 31 osób. W kolejnej sekcji opisano, w jaki sposób przygotowane zostały sygnały testowe; następnie przedstawiono przebieg i wyniki badań psychoakustycznych.

2. Metodyka

2.1. Przygotowanie próbek

W badaniach psychoakustycznych często wykorzystywane są dźwięki syntetyczne – pozwala to na precyzyjną kontrolę wszystkich elementów sygnału, jednak bywa nienaturalne dla słuchaczy.

Z tego powodu zdecydowano się na wykorzystanie nagrań „żywych" sygnałów – instrumentów muzycznych oraz mowy. Przy wyborze metody ich przetwarzania na potrzeby badań nadrzędną kwestią była możliwość prezycyjnej zmiany fazowej przy minimalnej (najlepiej: żadnej) ingerencji w amplitudę. Dlatego zdecydowano się na skorzystanie z krótkoczasowej transformaty Fouriera: w każdym okienku modyfikowano jedynie widmo fazowe sygnału, pozostawiając widmo aplitudowe bez zmian. Następnie, korzystając z odwrotnej transformaty Fouriera i metody *overlap-add*, rekonstruowano sygnał w dziedzinie czasu. Nakładkowanie okien wynosiło 75%, a długość w zależności od wariantu (opisane dalej) 1024 lub 2048 próbek (tj. odpowiednio 21,3 i 42,7 ms – częstotliwość próbkowania dla wszystkich sygnałów wynosiła $\mathrm{kilo48\,Hz}$

Widmo sygnału w m-tym okienku można zatem opisać wzorem:

$$\hat{S}_m(f) = |\hat{S}_m(f)|\left[\cos\left(\phi_{m0}(f) + \Delta\phi_m(f)\right) + \mathrm{i}\cdot\sin\left(\phi_{m0}(f) + \Delta\phi_m(f)\right)\right] \quad , \tag{1}$$

gdzie ϕ_{m0} to oryginalna faza sygnału, a $\Delta\phi_m(f)$ – wprowadzane przesunięcie fazowe. Jego wartość losowano z przedziału $< -\Delta\phi_{max}; \Delta\phi_{max} >$ dla co drugiego pasma częstotliwości – następnie wartości przesunięć w pozostałych pasmach były interpolowane wielomianem 3. stopnia z uwzględnieniem wartości maksymalnych (zmodyfikowana metoda interpolacji Akimy) [8]. Taki zabieg miał na celu uniknięcie skokowych przesunięć między sąsiadującymi pasmami częstotliwości.

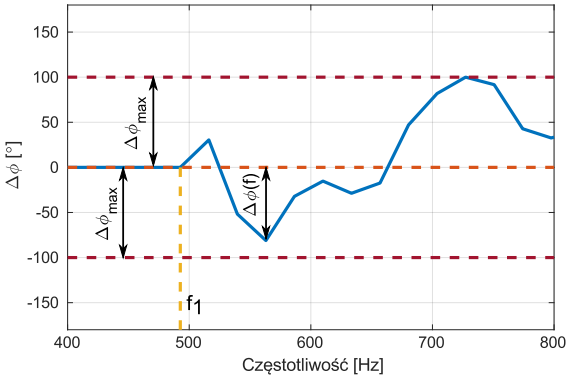
Drugim ograniczeniem zakresu losowania była różnica między przesunięciami w kolejnych oknach – jego wprowadzenie pozwoliło na uniknięcie bardzo gwałtownych zmian, które mogłyby być słyszane jako zniekształcenia sygnału. Te dwa warunki na losowaną wartość $\Delta\phi_m(f)$ można zapisać jako:

$$\begin{cases} |\Delta\phi_m(f)| \leq \Delta\phi_{max} \\ |\Delta\phi_m(f) - \Delta\phi_{m-1}(f)| \leq \Delta\phi_{win} \end{cases} , \tag{2}$$

gdzie $\Delta\phi_{max}$ to maksymalna dopuszczalna zmiana fazy, a $\Delta\phi_{win}$ – maksymalna dopuszczalna różnica między kolejnymi oknami. Na rysunku 1 przedstawiono przykładowy wynik losowania w jednym oknie.

2.2. Badania psychoakustyczne

W badaniu brało udział 31 osób w wieku od 16 do 58 lat, z czego 87% poniżej 28 roku życia. 58% badanych ma doświadczenie/wykształcenie muzyczne lub związane z inżynierią dźwięku.



Rysunek 1 Przykład uzyskanych zmian fazy dla jednego okna sygnału; zaznaczono amplitudę zmian $\Delta\phi_{max}$ (przedział losowania), dolną granicę zakresu częstotliwości podlegającego modyfikacjom f_1 oraz wartość wylosowaną dla wybranego pasma częstotliwości $\Delta\phi(f)$

Aby sprawdzić, czy wprowadzane modyfikacje wywołują wrażenie większego pogłosu, przeprowadzono test oparty na metodzie porównania dwóch bodźców, z zastosowaniem skali interwałowej (ang. *paired scaling*). W każdej próbie, używając przygotowanej na potrzeby badania aplikacji komputerowej w środowisku LabVIEW, słuchacze mieli za zadanie wysłuchać dwóch próbek dźwiękowych, a następnie odpowiedzieć na pytanie: **Który dźwięk wywołuje wrażenie bardziej pogłosowego (większego) pomieszczenia?** Sformułowanie „większego" skierowane było do osób mniej obeznanych z terminologią akustyczną, gdyż zazwyczaj pogłos jest kojarzony z pomieszczeniem o dużej kubaturze; w przypadku wątpliwości zawsze precyzowano, że ocenianą cechą jest pogłosowość. Skalę odpowiedzi przedstawiono w tabeli 1: górny wiersz to odpowiedzi dostępne dla słuchaczy, a dolny – ich interpretacja liczbowa na potrzeby analizy wyników.

Tabela 1 Reprezentacja liczbowa skali ocen użytej w testach odsłuchowych

zdecydowanie A	raczej A	brak różnicy	raczej B	zdecydowanie B
-2	-1	0	1	2

Wykorzystane do badań nagrania instrumentów pochodziły ze zbiorów autorów, zostały wykonane w studiu Katedry Mechaniki i Wibroakustyki AGH z bliskim mikrofonowaniem [9]. Są to kilkudźwiękowe sekwencje zagrane na flecie (Fl), oboju (Ob) i wiolonczeli (Vc). Sygnały mowy pochodzą z dostępnej na licencji *Creative Commons* bazy *TSP Speech Database* Uniwersytetu McGill [10]; każde nagranie to jedno zdanie w języku angielskim. Do badań wykorzystano jedno zdanie wypowiedziane przez kobietę (K) i jedno przez mężczyznę (M).

Pytania były podzielone na cztery kategorie: w pierwszej porównywano wpływ długości okna STFT (10 prób), w drugiej maksymalne odchylenie od pierwotnej wartości fazy (15 prób). W kategoriach trzeciej i czwartej badano odpowiednio maksymalną różnice pomiędzy kolejnymi oknami (4 próby) i zakres częstotliwości, w którym faza była modyfikowana (5 prób). Poza kategorią pierwszą, wszędzie stosowano okno o długości 2048 próbek. W teście znalazły się też 3 próby kontrolne, w których obie próbki były tym samym sygnałem, wplecione między pozostałe pytania. Kolejność prób była jednakowa dla wszystkich słuchaczy.

3. Wyniki

Na potrzeby analizy statystycznej przeskalowano odpowiedzi tak, aby zawsze próbka A była sygnałem, w którym wprowadzono mniejsze modyfikacje fazy (lub nie wprowadzono żadnych). Do analizy skorzystano z testu znakowanych rang Wilcoxona, który jest nieparametrycznym testem mediany i nie wymaga założenia o ciągłym rozkładzie badanej zmiennej [11]. Na poziomie ufności 95% testowano następujące hipotezy statystyczne:

$\boldsymbol{H_0: Me = 0}$ udzielone odpowiedzi nie wskazują na występowanie różnic percepcyjnych między próbkami.

$\boldsymbol{H_1: Me \neq 0}$ odpowiedzi wskazują na możliwość wystąpienia istotnych statystycznie różnic percepcyjnych między próbkami.

Wykonano dwie analizy: na pełnej próbie słuchaczy, oraz na zawężonym zbiorze. Próby kontrolne potraktowano jako wyznacznik dla wiarygodności wyników dla danego słuchacza – w zbiorze zawężonym odrzucono wyniki wszystkich badanych, którzy przynajmniej w 2 z 3 prób kontrolnych wskazali różnicę między dźwiękami. Eliminacja tych wyników nie wpłynęła jednak na wyniki testu istotności – tylko dla jednej próby wynik był inny dla zbioru zawężonego (w zestawie badającym długość okna).

3.1. Długość okna

Wstępne, nieformalne testy wykazały, że przy czasie trwania okna $21{,}3\,\mathrm{ms}$ (1024 próbki) nawet duże modyfikacje fazy są trudno zauważalne. Dlatego zbadano słyszalność zmian dla tych samych parametrów zmian fazy, a różnych długości okna. Wyniki analizy statystycznej pokazują, że istotną różnicę między próbkami zauważono w 2 próbach z 10 (dla zawężonego zbioru wyników – 3 próby).

3.2. Maksymalna zmiana fazy

W tabeli 2 przedstawiono wyniki analizy statystycznej dla kategorii drugiej, związanej z maksymalną wprowadzaną zmianą fazy. Jako pierwszy parametr próbki podana jest amplituda zmian fazy w stopniach, jako drugi (w nawiasie) – maksymalna zmiana między kolejnymi oknami. Na szaro zaznaczono próby, w których percepcyjna różnica między próbkami była statystycznie istotna według przyjętych kryteriów – były to nagrania mowy męskiej oraz wiolonczeli, w których zakres losowania był największy. Dla tych samych zakresów w przypadku nagrań oboju słuchacze nie wskazywali istotnych różnic pogłosowości.

Tabela 2 Wyniki analizy statystycznej dla zestawienia amplitud zmiany faz; na szaro zaznaczono próby o istotnym statystycznie wyniku testu

Próba			Zbiór pełny, $N = 31$				Zbiór zawężony, $N = 14$			
Nagranie	**Próbka A**	**Próbka B**	m_1	Me	W	p	m_1	Me	W	p
M	0	60 (10)	0,45	1,0	199	**0,0507**	0,50	0,5	32	0,0625
M	0	90 (10)	0,68	1,0	192	0,0041	0,71	1,0	36	0,0078
M	0	120 (10)	1,35	2,0	436	0,0000	1,57	2,0	105	0,0001
M	60 (10)	120 (10)	1,35	2,0	397	0,0000	1,64	2,0	105	0,0001
Ob	0	90 (10)	-0,13	0,0	87	0,4951	0,07	0,0	9	1,0000
Ob	0	60 (10)	0,29	-0,0	113	0,0791	0,07	0,0	4	1,0000
Ob	0	120 (10)	0,35	-0,0	126	0,0795	0,00	0,0	5	1,0000
Ob	60 (10)	120 (10)	0,29	0,0	164	0,2238	0,21	0,0	20	0,4531
Vc	0	60 (10)	0,00	-0,0	114	0,9999	0,21	0,0	29	0,6133
Vc	0	90 (10)	0,68	1,0	236	0,0091	0,93	1,0	45	0,0039
Vc	0	120 (10)	0,81	1,0	281	0,0006	0,86	1,0	55	0,0020
Vc	60 (10)	120 (10)	0,90	1,0	327	0,0004	0,71	1,0	48	0,0430

3.3. Maksymalna zmiana między oknami

Wstępne, nieformalne testy słuchowe sugerowały, że różnica zmian faz między kolejnymi oknami $\Delta\phi_{win}$ ma wpływ na postrzeganie pogłosowości; jednak w badanym zakresie testy nie potwierdziły tej zależności. Badano dwie próbki głosu kobiecego i dwie fletu; w każdej maksymalna modyfikacja wynosiła $90°$, dla próbki A $\Delta\phi_{win}$ wynosiła zawsze $10°$, a dla próbki B – $15°$ i $30°$. Jedynie dla pary nagrań fletu 90 (10) i 90 (30) postrzegane różnice były statystycznie istotne.

3.4. Zakres częstotliwości

Tę część badań przeprowadzono jedynie z użyciem nagrań wiolonczeli, których widmo obejmowało najszerszy zakres częstotliwości (najniższa częstotliwość: 98 Hz). W przypadku porównywania zakresów $0 - f_g$ oraz $f_g - f_s/2$, gdzie f_s to częstotliwość próbkowania, a f_g częstotliwość podziału wynosząca 400 lub 1500 Hz, za bardziej pogłosową uznawana była próbka, w której faza była zmodyfikowana w zakresie wyższych częstoliwości ($f_g - f_s/2$). Potwierdza to hipotezę D. Griesingera o największym znaczeniu tego zjawiska powyżej 1000 Hz [7].

Przy porównaniu zmian w pełnym paśmie oraz w zakresie $1500\,\mathrm{Hz} - f_s/2$ za bardziej pogłosową uznawano próbkę zmodyfikowaną w szerszym zakresie częstotliwości. Porównanie modyfikacji w zakresie $0-400$ Hz i $0-1500$ Hz oraz $400\,\mathrm{Hz} - f_s/2$ i $1500\,\mathrm{Hz} - f_s/2$ nie wykazało istotnych statystycznie różnic percepcyjnych.

4. Dyskusja i podsumowanie

Fakt, że modyfikacje fazy nie były zauważalne dla oboju jest niezgodny z oczekiwaniami, gdyż dźwięk oboju charakteryzuje się harmoniczną strukturą widma, gdzie harmoniczne są wyraźnie wyróżnialne, można się zatem było spodziewać, że zmiany fazy będą wyraźnie słyszalne. Na kształt widma wiolonczeli wpływają dodatkowo rezonanse korpusu i nieharmoniczność strun; jest ono więc bardziej „rozmyte". Z kolei widmo sygnału mowy opisywane jest przez układ formantów. Badania pokazują, że struktura widmowa ma wpływ na percepcję zależności fazowych w sygnale [5]; dobór nagrań do kolejnych etapów badań powinien być poprzedzony sprawdzeniem tej zależności.

Po zakończeniu badania słuchacze byli poproszeni o komentarz i ewentualne uwagi co do jego przebiegu – kilka osób wskazało na istnienie różnic między próbkami, które jednak nie dawały wrażenia pogłosu, a także na nienaturalne brzmienie prezentowanych nagrań. Są to kwestie, które należy wziąć pod uwagę przy projektowaniu kolejnych eksperymentów; być może należy rozważyć inne metody wprowadzania modyfikacji fazy (np. filtr wszechprzepustowy o określonej charakterystyce fazowej).

W przeprowadzonych badaniach wykazano, że modyfikacja fazy sygnału może wywoływać wrażenie pogłosu. Przeprowadzono formalne testy słuchowe, z wykorzystaniem nagrań, w których faza modyfikowana była w kolejnych oknach STFT. Sprawdzono wpływ różnych aspektów tej modyfikacji na wynik: wartości zmiany, maksymalnej zmiany miedzy kolejnymi oknami czasowymi, długości okna czasowego, w ograniczonym zakresie także zakres częstotliwości. Przedstawione wyniki skłaniają do przeprowadzenia dalszych badań nad tym zagadnieniem, po wcześniejszej rewizji metodyki.

Bibliografia

[1] Anna Snakowska: *Teoria pola akustycznego zastosowana do badania układów o symetrii cylindrycznej*, Kraków: Wydawnictwa AGH, 2018, ISBN: 978-83-66016-08-8

[2] Cheol-Ho Jeong, Doheon Lee, Sébastien Santurette i Jeong-Guon Ih: "Influence of Impedance Phase Angle on Sound Pressures and Reverberation Times in a Rectangular Room", w: *The Journal of the Acoustical Society of America* 135.2, lut. 2014, s. 712–723

[3] Cheol-Ho Jeong, Jeong-Guon Ih i Jens Rindel: „A study on the characteristics of phased beam tracing method for the acoustic simulation of an enclosure at mid frequencies", w: *The 9th Western Pacific Acoustics Conference*, Seoul, Korea, czer. 2006

[4] Edward Ozimek: *Dźwięk i jego percepcja. Aspekty fizyczne i psychoakustyczne*, Warszawa: Wydawnictwo Naukowe PWN, 2018, ISBN: 978-83-01-20116-6

[5] Mikko-Ville Laitinen, Sascha Disch i Ville Pulkki: "Sensitivity of Human Hearing to Changes in Phase Spectrum", w: *J. Audio Eng. Soc* 61.11, list. 2013, s. 860–877

[6] David Griesinger: "Pitch Coherence as a Measure of Apparent Distance in Performance Spaces and Muddiness in Sound Recordings", w: *Audio Engineering Society Convention 121*, paź. 2006

[7] David Griesinger: "Pitch, Timbre, Source Separation, and the Myths of Loudspeaker Imaging", w: *Audio Engineering Society Convention 132*, kw. 2012

[8] MathWorks: *Matlab Documentation:* `interp1`, 2019, URL: *https://uk.mathworks.com/help/matlab/ref/interp1.html*, term. wiz. 10. 05. 2019

[9] Teresa Makuch, Zuzanna Kusal i Krzysztof Mikulec: „Śpiewające instrumenty", w: *Studium badawcze młodych akustyków 2017*, red. Jarosław Rubacha, 4. Ogólnopolska Studencka Konferencja Akustyków OSKA 2017, Kraków, czer. 2018, ISBN: 978-83-908774-5-7

[10] Peter Kabal: *TSP Speech Database*, list. 2018, URL: *http://www-mmsp.ece.mcgill.ca/Documents/Data/*

[11] Vijay K. Rohatgi i A. K. Md. Ehsanes Saleh: *An Introduction to Probability and Statistics*, New Jersey: John Wiley & Sons Inc., 2015, ISBN: 978-1-118-79964-2

Sesja 4

Dźwięk przestrzenny

wtorek, 29.06.2021
13:40 – 14:20

SPATIAL AUDIO IN MUSIC PRODUCTION – TOOLS, METHODOLOGY AND PERCEPTION

K. Sochaczewska[1], M. Piotrowska[2], Sz. Piotrowski[3], P. Małecki[1]

[1]AGH University of Science and Technology

[2]Music Academy in Kraków

[3] Learn How to Sound

ksocha@agh.edu.pl

Abstract

Spatial audio systems along with the available tools create a superior environment for creative musical production. Rapid expansion of immersive audio technology opened the new ways of expression, especially regarding localisation,perception of distance and movement of the sound sources. Paper includes the outline of selected spatial formats and tools. Practical implementation can be basis for the research on human auditory perception within multichannel and binaural systems. Remarks on the case study production process in ambisonic system including technical aspects, the creative factor and aesthetics are followed by listening tests. Outcomes and discussion are aimed for future development of the workflow in music production in 3D audio systems.

1. Introduction

The rapid expansion of immersive audio technology has facilitated new and innovative methods of music production. Through 3D mixing practices, various possibilities can be explored, as well as limitations encountered. The research presented here refers to a previous article by the authors, where some aspects of 3D music production were discussed [1].Since there is no established workflow nor standards for spatial music productions, case studies are used as the basis for research on human auditory perception within multichannel and binaural systems. In parallel, an improved understanding of the psychoacoustics of perception in spatial systems is a step towards a more coherent translation of artists' ideas and their successful execution. This paper explores some of the possibilities, challenges and limitations associated with spatial music production.

2. Spatial audio systems and tools

The range of spatial sound technologies is already wide and expanding. Depending on the application, Dolby Atmos and Auro3D is exploited in film and tv industry, ambisonic technology can be found behind the gaming engines such as Unity, not to mention hybrid methods encrypted within commercial systems (AirPods alongside with Netflix - Spatial Audio, YouTube and Facebook360). Some of the technologies are not exclusive and can be mixed, both within each other and as an extension to standard stereo.

Multichannel systems can also be of great value in educational purposes. Until very recently (early 2000-ies [2]) only afew attempts have been made in order to determine psychoacoustic phenomena in all three planes. Just Noticeable Difference of parameters such as Minimum Audible Angle, Max Velocity and others are crucial for perception of the source direction and movement. Resolving these values is relevant in order to deliver the content in the most coherent way for diversified audience [3].

Within the tools used in spatial audio production one can encounter following:

- open source, free and usually ambisonic-based, developed at the universities
- commercial, paid for
- licensed formats

- can include only transformations (encoding, decoding, beaming, downmixing etc.)
- as well as creative ones (implemented simulations such as virtual acoustics, air absorption, doppler).

Figure 1 Spatial ambisonic reproduction system at AGH UST Laboratory of Auralization

We are using ambisonic technology for this specific purpose for a number of reasons. Not only it offers a great flexibility in terms of delivery format, which includes binaural downmix, but also in terms of combining tools, methods and techniques. This is aimed to provide the best of both: stereo and spatial qualities.

In the case of the binaural playback, the acoustics of the room are not being taken into consideration, which play a role in, among others, localisation within ambisonic speaker playback systems (dryness of the space, positions of the speakers with regard to wall adjacency and early reflections follow). Considering moving sound sources, the consistency of loudness, timbre, width is crucial to preserve in panning. It is worth noting that for the commercial stereophonic playback the simple principle called pan law (3 dB, 4.5 dB or 6 dB) is in use. For spatial audio the panning methods can be divided into two main categories: amplitude- and psychoacoustics based.

An encoded ambisonic signal needs to be decoded before being played back. The fact that those two processes – encoding and decoding - are separate, makes one of the biggest advantages of ambisonic technology – independence of any fixed speaker arrangement (although it is possible to merge those two steps together resulting in channel-based panning type).

3. Music production in spatial systems

Most of all, mixing in spatial systems provides amazing capabilities in terms of the space that can be utilized. At the same time, sources that have been packed into a 60° stereo set-up now can cover up to the whole sphere (depending on the system). A 3D systems provide new opportunities but also introduce limitations, which result in some compromises. Immersive mixing and traditional stereo mixing share a multitude of basic rules and requirements, i.e. frequency balance, selectivity, but meet different challenges. In case of the spatial content, overloading the piece with movement (either the amount as well as the velocity) may result in unconvincing effects. Therefore, psychoacoustic tests allowing to determine such parameters as MAA (Minimum Audible Angle) or MAMA (Minimum Audible Movement Angle) can be implemented into the conscious process of spatial production, in order to provide the most efficient

result. Creativity in sound sources placement is definitely a great opportunity to emphasise musical content but, at the same time, provides a chance to take the attention away from the music.

The other demanding part is restrained options in ambisonic bus processing. Although ambisonic dynamic and reverb plugins provide great opportunities, their palette of sound is limited when compared to plugins available in the stereo format.

4. Case study

In the presented research, the authors introduce the case study of original electronic music production and describe an ambisonic remix of a stereo piece. The main emphasis was placed on the utilisation of all dimensions and extended space. Remarks on the mixing process, composition for ambisonic system, a creative factor and aesthetics are followed by listening tests performed by a panel of experts. The subjective evaluation included stereo and ambisonic playback, as well as binaural rendering. In particular, binaural rendering is currently a promising topic, with more information regarding this matter, such as localisation, front-back-up-down confusion and sound quality [1].

Working on the binaural version of the composition consisted of the three stages. The first was getting familiar with the original stereo mix and focusing on the strong points of the stereo mix that would not be straightforward to translate to the spatial version. Those included for instance low end and strong attack of the percussive elements. We took into consideration the psychoacoustic phenomena such as perceiving the high frequencies to appear above the horizontal plane. Another way of increasing the sense of space and depth is to use filtering (dampening) and simulate reflections(reverb). They were used to enrich the auditory perception, especially for the planes where it becomes difficult to distinguish the movement and location of the sound sources otherwise.

Beside the precise, high resolution elements, two types of soundscape are included in the stems: natural environmentrecordings captured through the stereo AB pair (Schoeps MK2S) and electronic pads that spread across the tridimensional space. They give the foundation for the track and also importantly create a spatial anchor for the listener. It is said to be more efficient if there is a movement relative to some stationary object. It grabs attention and enhances the perceptual sensation and being aware of the sound sources' localisation. In the mixing process we used 3rd ambisonic order to provide high resolution related to refined directional localisation precison. It also allows to preserve decorrelation. It is estimated by the covariance between uncorrelated directions [4].

The ambisonic tools used for latter binaural part included both free plugins such as IEM Suite and O3 Blue Ripple Sound Package that mostly focus on spatial translation and transformations as well as commercial tools with implemented creative functions. For the binaural render, the IEM Binaural Decoder that implements HRTF from Neumann KU 100 dummy head [5] was utilized. In contrast to the majority of the binaural decoders, IEM uses direct convolution with HRTF, omitting the virtual speakers in decoding chain, resulting in maintaining the HRTF response in the best possible manner [6]. The last part of the process was combining the stereo tracks with the binaural renders.

The authors performed preliminary listening experiment, which methodology was based on previous research including a 3D audio evaluation performed by the authors [1,7,8,9,10]. As described in Section 3, remixing electronic music produced in stereo into a spatial system is associated with many opportunities, but in some cases also compromises, therefore subjective evaluation involving a group of 10 experts was performed. For this purpose, both versions were loudness normalized to -18 LUFS with app. 10 LUFS Loudness Range. The test consisted of preference choice between the spatial binaural render versus the stereo mix (Beyerdynamic DT770 closed headphones and Sennheiser HD 650 oped headphones). The group of listeners consisted of highly trained listeners- audio engineers, sound engineering teachers, musicians and acoustical engineering students. Most of them reported previous experience with surroundand/or immersive set-ups. They rated their preference in four categories: Spaciousness, Selectivity, Timbre, Dynamics and General impression. The participants were also asked to provide their impressionsand comments. In some cases, they reported that it was difficult to pick one version as they

were so different from each other and that they appreciated both mixes. Some listeners also emphasized the spectacular impression of the sound source movement above the listener in the binaural version.

As the outcome (Fig. 2), the spaciousness was preferred in the binaural version, but the results are the opposite in the case of selectivity. The assessment of timbre and dynamics is slightly higher in stereo, while the general preference is in favor of the binaural variant. It is worth recalling that was no HRTF individualization or equalization.

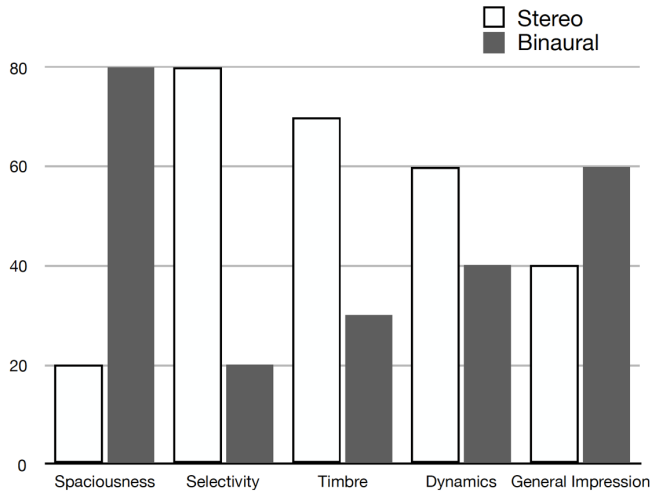


Figure 2 Listeners' preference results of comparison of the stereo mix to the binaural render of the spatial remix

5. Final conclusions

Outcomes of the presented research include:

- Outline of some selected formats and tools
- Discussion focused on the technical workflow, aesthetic possibilities, potential applications, improvements of localisation of the sound sources
- Elaboration on a case study, including production processes and listening evaluation

All these points are aimed at the future development of the workflow in music production in spatial systems.

The approach proposed in this paper is a combination of fixed stereo and spatial techniques, which preserves the best of both approaches. Moreover, this approach allows for conventional bus processing techniques to be applied to the stereo tracks only, as advancements in spacial audio technology in this area continue to be made and standards and practices are established.

Preliminary listening tests indicate a general preference for the binaural versions, as well as a significant preference for the spaciousness of the binaural versions. Due to the prevalence of headphone playback, the obtained results show thepotential of spatial productions with targeted binaural playback. It

is important to note that despite some statistical significance, these results cannot be treated as general outcomes due to the subjective nature of the experiment. The work presented in this paper is based on a case study, and the results represent the preference of the selected group of experts. Therefore the results cannot be interpreted as general rules, but are important qualitative remarks that can be used to supplement existing understanding of the area.

The results highlight the dominance of the binaural mixes in the terms of spaciousness, but also show that they are not highly rated in the remaining 3 categories (timbre, selectivity and dynamics). The findings thus demonstrate the need for future work to explore more fully the possibilities of merging the advantages of a classic stereo approach with those offered by spatial techniques, thus enabling the creation of more engaging musical content.

References

[1] P. Małecki, M. Piotrowska, K. Sochaczewska, Sz. Piotrowski: "Electronic music production in ambisonics – case study", in: Journal of the Audio Engineering Society (2020)

[2] P. Power, Ch. Dunn, B. Davies, J. Hirst: "Localisation of Elevated Sources in Higher-Order Ambisonics", in: BBC Research & Development White Paper, WHP 261 (2013)

[3] J. Francombe, T. Brookes, and R. Mason, "Evaluation of Spatial Audio Reproduction Methods(Part 1): Elicitation of Perceptual Differences", in: J. AudioEng. Soc., vol. 65, pp. 198–211 (2017 Mar.),https://doi.org/10.17743/jaes.2016.0070

[4] F. Zotter, M. Frank: "Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality", Springer Nature (2019)

[5] IEM Plug-in Suite (2018 (accessed June 13, 2021)), URL https://plugins.iem.at/docs/plugindescriptions/

[6] Schoerkhuber, Christian; Zaunschirm, Markus; Hoeldrich, Robert. "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares", Fortschritte der Akustik, DAGA (2018)

[7] P. Kleczkowski, A. Krol, P. Malecki, "Multichannelsound reproduction quality improves with angular separationof direct and reflected sounds", in: Journal of the Audio Engineering Society, vol. 63, no. 6, pp. 427–442 (2015),URL https://doi.org/10.17743/jaes.2015.0054

[8] K. Sochaczewska, P. Malecki, D. Czopek, J. Wiciak, "Soundscape Recording: Review of Approaches", presented at the Audio Engineering Society Convention 142 (2017 May), URL http://www.aes.org/e-lib/browse.cfm?elib=18687

[9] K. Sochaczewska, "The study of recording of Klezmer music using ambisonics and binaural microphone techniques" (2017)

[10] K. Sochaczewska, Consumer grade surroundsound system - calibration, mix and mastering, Master's thesis, AGH University of Technology, Faculty of Mechanical Engineering and Robotics, Kraków, Poland (2017)

Sesja 5

Rozwiązania akustyczne i elektroakustyczne

wtorek, 29.06.2021
14:20 – 15:40

EKSTRAKCJA PARAMETRÓW STANU GŁOŚNIKA MAGNETOELEKTRYCZNEGO WPŁYWAJĄCYCH NA ZNIEKSZTAŁCENIA

M. Kmiecik[1]

[1]AGH Akademia Górniczo-Hutnicza

miszkoo@gmail.com

Streszczenie

Znaczący wpływ na jakość dźwięku mają zniekształcenia wprowadzane przez głośniki. Redukcja zniekształceń jest możliwa dzięki zastosowaniu cyfrowego przetwarzania sygnałów (DSP), ale wymaga dobrego numerycznego modelu ruchu membrany, w którym uwzględniono zjawisko nieliniowości. W pracy opisano główne przyczyny nieliniowości głośników oraz możliwe metody szacowania parametrów potrzebnych do wykonania odpowiedniego modelu. Opisana propozycja nie wykorzystuje dodatkowych czujników, takich jak laserowy czujnik przemieszczenia czy mikrofon – proponowane źródło informacji to siła elektromotoryczna cewki głośnika. Rozwiązanie to wymaga zastosowania odpowiedniej metody analizy sygnału oraz obwodu elektrycznego, który zapewnia pomiary impedancji. Dodatkowo, system nie będzie działał w oparciu o sygnał testowy, taki jak sinus przemiatany lub MLS, ale w sposób ciągły, w oparciu o sygnał audio.

1. Wstęp

Głośniki dynamiczne są najczęściej stosowanym rodzajem przetworników elektroakustycznych służących zamianie sygnału elektrycznego na falę akustyczną. Ze względu na rozwój materiałów oraz wiedzy na temat zjawisk zachodzących w poszczególnych elementach głośnika wprowadzono szereg modyfikacji, które zmieniają parametry tych urządzeń; jednak na przestrzeni lat gówna zasada działania nie została zmieniona. Co więcej, budowa miniaturowych przetworników stosowanych w urządzeniach mobilnych bazuje na tej samej zasadzie działania, co konwencjonalne głośniki [1]. Na tej podstawie można sądzić, że konstrukcja ta mimo posiadanych niedoskonałości zapewnia najlepsze parametry przetwarzania ze wszystkich znanych przetworników elektroakustycznych służących do przetwarzania sygnału elektrycznego w dźwięk. Wzrost szybkości i możliwości nowoczesnych układów o architekturze przystosowanej do cyfrowego przetwarzania sygnałów (DSP) pozwala na implementacje rozwiązań poprawiających jakość i bezpieczeństwo przetwarzania elektroakustycznego.

Zestawy głośnikowe są poddawane czynnikom o szerokim zakresie zmienności. Głównym czynnikiem jest szeroki zakres mocy przetwarzanego sygnału. Od głośników magnetoelektrycznych wymaga się pracy w liniowym zakresie – bez zniekształceń – od dziesiątych części wata do wartości rzędu kilowatów w przypadku przetworników średnio i niskotonowych stosowanych w urządzeniach nagłośnieniowych o przeznaczeniu profesjonalnym, od których wymagany jest szeroki zakres generowanych poziomów akustycznych. Parametry głośnika podczas pracy zmieniają się w zależności od mocy dostarczanego sygnału. Większa moc sygnału dostarczanego do głośnika przekłada się na wzrost amplitudy ruchu membrany, z którym nasilają się oddziaływania zjawisk będących źródłem nieliniowości przetwarzania. Amplituda ruchu jest również związana z częstotliwością sygnału. Dla niskich częstotliwości amplituda jest znacznie większa niż dla wysokich, wartość maksymalną osiąga dla częstotliwości rezonansowej, Dokładność odwzorowania charakterystyki ruchu membrany zależy od uwzględnianych źródeł nieliniowości. Zniekształcenia nieliniowe związane z przetwornikiem mają swoje źródła w skończonych wymiarach szczeliny magnetycznej, właściwościach mechanicznych elementów sprężystych zawieszenia membrany np. zjawisku określanym

jako efekt pełzania (ang. *Creep Effect*) [2, 3, 4], a także – w niewielkim stopniu – w zmianie temperatury uzwojenia cewki.

W literaturze wskazuje się na kilka istotnych elementów pod względem nieliniowości np.: zmienność współczynnika siły Bl (iloczyn gęstości strumienia magnetycznego w szczelinie układu magnetycznego oraz długości przewodnika cewki na który działa strumień magnetyczny), podczas pracy z dużymi wychyleniami [5]. Stała Bl określa siłę mechaniczną jaką może wywrzeć na korpus cewki przepływający przez jej uzwojenie prąd i wyrażana jest w $\mathrm{N/A}$. Wspomniane zjawisko można zaobserwować w przykładowych modelach numerycznych do oprogramowania *COMSOL* [6]. Innym istotnym elementem wpływającym na ruch są: budowa (kształt) i parametry sprężystości materiałów, z których wykonano zawieszenie membrany, na co wskazuje Klippel w [7], a potwierdzenie można znaleźć w przykładowych modelach numerycznych do ww. oprogramowania [8]. Parametry te mają bezpośredni wpływ na nieliniowość stałej sprężystości wchodzącej do modelu zaproponowanego przez [9, 10, 11]. W [12] R. Ravaud wskazuje zmienność indukcyjności cewki L_e jako istotny czynnik nieliniowości, w szczególności, w zakresie niskich częstotliwości. Zmienność L_e jest czynnikiem w największym stopniu przyczyniającym się do powstawania drugiej harmonicznej. Natomiast na powstawanie trzeciej harmonicznej najmiększy wpływ ma rezystancja mechaniczna R_{ms}.

Próbę wskazania parametru modelu *Thiele-Small'a* w największym stopniu wpływającego na nieliniowość głośnika podjęto w wcześniej wspomnianej pracy R. Ravaud [12]. Autorzy wskazali rankingi udziału parametrów dla dwóch częstotliwości 50 i $150\,\mathrm{Hz}$.

Powyższe rozważania, prowadzą do wniosku, że nie jest możliwe jednoznaczne wskazanie parametrów, mających najistotniejszy wpływ na zniekształcenia w szerokim zakresie napięć i częstotliwości sygnału doprowadzonego do głośnika, wpływ ten jest uzależniony od wartości napięcia oraz częstotliwości sygnału.

2. Metodyka

W literaturze można znaleźć wiele badań na temat zniekształceń wywołanych nieliniowością głośników oraz prób ich wyeliminowania. Rozwiązania można podzielić na dwie grupy: systemy pętli zamkniętej oraz systemy pętli otwartej [13].

Systemy pętli zamkniętej cechuje konieczność pomiaru parametrów opisujących stan głośnika za pomocą dodatkowych czujników. W celach badawczych do pomiaru ruchu membrany głośnika stosuje się laserowe czujniki przemieszczeń. Jednak takie rozwiązanie nie znajduje uzasadnienia do stosowania w systemach dźwiękowych ze względów ekonomicznych i czysto praktycznych.

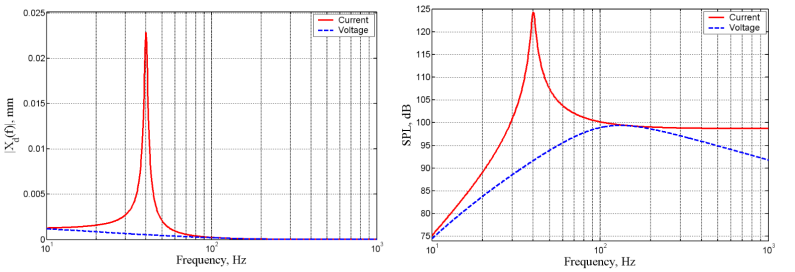
Innym rozwiązaniem są systemy pętli otwartej; bazują na modelu głośnika, którego parametry zostały wprowadzone do systemu i są stałe [13]. Algorytm wprowadzający zmiany w sygnale dostarczanym do głośnika nie ma informacji o aktualnym stanie pracy, stąd nie jest odporny na zmiany zachodzące w układzie. Takie rozwiązanie nie wymaga dodatkowych czujników, wszystkie parametry są pozyskiwane na podstawie pomiarów laboratoryjnych.

2.1. Sterowanie źródłem prądowym

Badania na temat zastąpienia konstrukcji wzmacniacza w układzie źródła napięciowego, rozwiązaniem odpowiadającym źródłu prądowemu zostały zaprezentowane w [14, 15]. Autorom udało się zredukować poziom zniekształceń wyłącznie przez zmianę konstrukcji wzmacniacza na pracującą w układzie źródła prądowego. Rozwiązanie to ma jednak istotną wadę – dla częstotliwości rezonansowej układu amplituda drgań rośnie. W przypadku sterowania źródłem napięciowym wzrost mocy w stanie rezonansu redukowany jest przez wzrost impedancji (amplituda prądu maleje); dzięki temu nie jest konieczne stosowanie korekcji amplitudy sygnału.

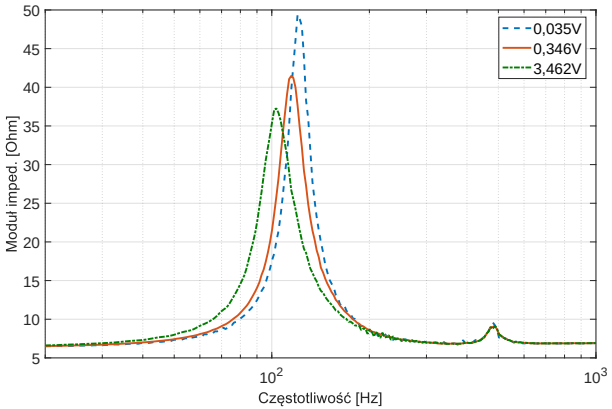
Powyższe rozważania prowadzą do wniosku, że w przypadku zasilania głośnika za pomocą wzmacniacza mocy audio w układzie źródła prądowego, w stanie rezonansu moc dostarczana do głośnika rośnie, co przekłada się na gwałtowny wzrost amplitudy przemieszczeń, rys. 1a, a za tym idzie wzrost poziomu ciśnienia akustycznego generowanego przez głośnik rys. 1b.

Budowa wzmacniacza mocy w układzie źródła prądu przyczynia się do redukcji zniekształceń, jednak jak wykazano w [17] wymaga do poprawnej pracy korekcji amplitudy prądu dla częstotliwości rezonansowej. Przedstawione rozwiązanie wymaga dopasowania filtru do konkretnej konfiguracji głośnika w obudowie



Rysunek 1 Głośnik zamontowany w nieskończonej odgrodzie zasilany za pomocą źródła napięcia lub prądu: a) amplituda przemieszczenia cewki, b) poziom ciśnienia akustycznego [16]

– ze względu na wpływ obudowy na częstotliwość rezonansową. Inaczej jest w przypadku typowo stosowanego źródła napięcia, które nie wymaga tak precyzyjnej korekcji. Na podstawie dotychczasowych wstępnych badań zauważono, że częstotliwość rezonansowa zmienia się z amplitudą sygnału dostarczonego do przetwornika, rys. 2. Stąd wniosek, że w przypadku pracy głośnika z różnymi amplitudami sygnału oraz konstrukcji wzmacniacza opartej na układzie źródła prądowego wymagana jest precyzyjna korekcja zależna od przetwarzanego sygnału. Problem sprowadza się więc do ciągłego monitorowania stanu ruchu głośnika co jest jednym z problemów podjętych w ramach dalszych prac.



Rysunek 2 Charakterystyki impedancji wyznaczona na podstawie sygnału sinusa przestrajanego dla trzech amplitud napięcia: 35mV; 346mV; 3,462V

2.2. Model matematyczny ruchu membrany

Głośnik magnetoelektryczny zalicza się do grupy przetworników liniowych odwracalnych; w obwodzie cewki powstaje siła elektromotoryczna (SEM), której źródłem jest ruch cewki w polu magnetycznym

wywołany siłą bezwładności w układzie drgającym. Wypadkowa siły elektromotorycznej dostarczanej z źródła napięcia oraz siły elektromotorycznej powstałej w cewce jest obserwowana w zmianach charakterystyki impedancji przetwornika. Zawarte w charakterystyce impedancji informacje mogą posłużyć do określenia właściwości ruchu membrany.

Do uwzględnienia zjawisk nieliniowych wykorzystuje się modele matematyczne uwzględniające zmienność parametrów w funkcji np. prądu cewki. Autorzy [12, 18] prezentują metodę wyznaczenia parametrów modelu Thiele-Smalla:

$Bl(i)$ – współczynnik siły (iloczyn gęstości strumienia magnetycznego w szczelinie układu magnetycznego oraz długości przewodnika cewki na który działa strumień magnetyczny),
$R_{ms}(i)$ – rezystancja mechaniczna strat,
$C_{ms}(i)$ – podatność mechaniczna,
$M_{ms}(i)$ – masa układu drgającego,
$L_e(i)$ – indukcyjność cewki;
z uwzględnieniem zmienności tych parametrów w funkcji prądu obwodu cewki. W celu uwzględnienia zmienności ww. parametry zostały przedstawione w postaci wielomianów, których współczynniki wyznaczono za pomocą algorytmu sympleksowego (ang. Simplex Algorithm) – iteracyjna metoda optymalizacji. Funkcją celu algorytmu sympleksowego był błąd między teoretyczną, a rzeczywistą zmierzoną charakterystyką impedancji [18]:

$$Z_e^{teor.}(i) = R_e + \frac{jR_\mu L_e(i)\omega}{jL_e(i)\omega + R_\mu} + \frac{Bl(i)^2}{R_{ms}(i) + jM_{ms}(i)\omega + \frac{k(i)}{j\omega}} \quad , \tag{1}$$

$$\Delta Z_e(i) = \sum ||Z_e^{zm.}(i) - Z_e^{teor.}(i)||^2 \quad , \tag{2}$$

gdzie:
i – prąd w obwodzie cewki,
ω – pulsacja ($2\pi f$),
R_μ – rezystancja związana z występowaniem prądów wirowych,
$k = \frac{1}{C_{ms}}$.

Przedstawione podejście może zostać zastosowanie do ciągłej kontroli parametrów głośnika. Na podstawie wzoru [12]:

$$x = \frac{Bl(i)}{\left((M_{ms}(i)(j\omega))^2 + R_{ms}(i)(j\omega) + \frac{1}{C_{ms}(i)}\right)} i \quad , \tag{3}$$

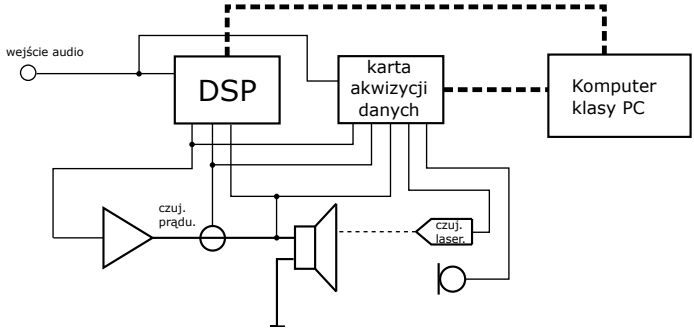
możliwe jest wyznaczanie sygnału odpowiadającego przemieszczeniu x głośnika z uwzględnieniem nieliniowości ww. parametrów.

2.3. Stanowisko badawcze

Przedstawione rozwiązanie wymaga dokładnego pomiaru charakterystyki impedancji głośnika. Typowe laboratoryjne układy pomiaru impedancji wykorzystują do pomiaru prądu rezystancję szeregową włączoną w obwód cewki; pomiar prądu odbywa się poprzez pomiar napięcia na rezystancji zgodnie z prawem Ohma. Zastosowanie czujnika prądu działającego w oparciu o zjawisko Halla pozwoli zwiększyć stosunek czułość/rezystancja szeregowa około 100-krotnie. Układ stanowiska badawczego został przedstawiony na rys. 3.

Układ cyfrowego przetwarzania sygnału wyznacza parametry modelu uwzględniające nieliniowości. Na podstawie sygnału ruchu cewki wyznaczonego za pomocą modelu zostaną ustalone parametry filtrów kompensujących nieliniowości.

Przedstawiony na rys. 3 komputer klasy pc służy wyłącznie do kontroli DSP oraz rejestracji sygnałów w układzie. Na podstawie zarejestrowanych sygnałów z laserowego czujnika przemieszczeń oraz mikrofonu możliwa będzie ocena działania układu poprzez porównanie sygnału ruchu membrany pozyskanego na podstawie modelu z informacją zarejestrowaną metodą optyczną za pomocą triangulacyjnego czujnika przemieszczeń.



Rysunek 3 Schemat stanowiska badawczego

2.4. Sygnały testowe

Jednym z celów jest stworzenie rozwiązania, które będzie w sposób ciągły kontrolować zmieniające się parametry pracy głośników; by to osiągnąć parametry modelu muszą być pozyskiwane na podstawie sygnałów muzycznych lub sygnału mowy. Wyznaczenie charakterystyki impedancji może zostać zrealizowane przez wyznaczenie ilorazu szybkiej transformaty Fouriera napięcia i prądu. Jednak ze względu na zdeterminowanie widma częstotliwościowego sygnałów mogą one nie zapewniać wymaganej dokładności.

3. Podsumowanie

Przedstawione rozwiązania skupiają się na uzyskaniu zakładanego celu jakim jest percepcyjne zauważalna poprawa jakości przetwarzania sygnału elektrycznego w akustyczny przy jednoczesnym zachowaniu poziomu dźwięku. Wymagają jednak weryfikacji na drodze doświadczalnej, która pozwoli określić faktyczny wpływ zastosowanych rozwiązań na jakość dźwięku oraz wskaże najistotniejsze źródła błędów. Jakość działania jest głównie uzależniona od dokładności odwzorowania sygnału ruchu membrany za pomocą otrzymanego modelu matematycznego. Zaproponowane podejście pozwala na eliminację dodatkowych czujników ruchu, w tym celu zastosowany zostanie algorytmu optymalizacji, który pozwala wyznaczyć niezbędne parametry modelu matematycznego opisującego ruch na podstawie charakterystyki impedancji.

Badania były częściowo finansowane z subwencji badawczej Wydziału Inżynierii Mechanicznej i Robotyki AGH, nr: 16.16.130.942/Kmiecik, oraz z projektu Zintegrowany Program Rozwoju Akademii Górniczo-Hutniczej w Krakowie II, nr: POWR.03.05.00-00-Z309/18-00

Bibliografia

[1] V. Lemarquand, G. Lemarquand, E. Lefeuvre, I. Shahosseini, R. Ravaud, J. Moulin i in.: „Electrodynamic MEMS: Application to mobile phone loudspeakers”, w: *IEEE Transactions on Magnetics* 48.11, 2012, s. 3684–3687

[2] M. H. Knudsen i J. Grue Jensen: „Low-frequency loudspeaker models that include suspension creep”, w: *AES: Journal of the Audio Engineering Society* 41.1-2, 1993, s. 3–18

[3] Finn Agerkvist, Knud Thorborg i Carsten Tinggard: „A Study of the Creep Effect in Loudspeaker Suspension”, w: *Audio Engineering Society Convention 125*, 2008

[4] Ulf Seidel i Wolfgang Klippel: „Fast and Accurate Measurement of the Linear Transducer Parameters”, w: *Audio Engineering Society Convention 110*, 2001

[5] Earl Geddes i Alan Phillips: „Efficient Loudspeaker Linear and Nonlinear Parameter Estimation", w: *Audio Engineering Society Convention 91*, 1991

[6] Comsol-Multiphysics: *Topology Optimization of a Magnetic Circuit*

[7] Wolfgang Klippel: „Dynamical Measurement of Non-Linear Parameters of Electrodynamical Loudspeakers and their Interpretation", w: *Audio Engineering Society Convention 88*, 1990

[8] Comsol-Multiphysics: *Loudspeaker Spider Optimization*

[9] Leo L. Beranek i Tim Mellow: *Acoustics: Sound Fields and Transducers*, Oxford: Elsevier, 2012

[10] H F Olson: *Acoustical Engineering*, Van Nostrand, 1957

[11] Richard H. Small: „Direct-Radiator Loudspeaker System Analysis", w: *J. Audio Eng. Soc* 3.4, 1972, s. 57–84

[12] Romain Ravaud, Guy Lemarquand, Valérie Lemarquand i Tangi Roussel: „Ranking of the nonlinearities of electrodynamic loudspeakers", w: *Archives of Acoustics* 35.1, 2010, s. 49–66

[13] Wolfgang Klippel: „The Mirror filter - A New Basis for Reducing Nonlinear Distortion and Equalizing Response in Woofer Systems", w: *AES: Journal of the Audio Engineering Society* 40.9, 1992, s. 675–691

[14] P. G.L. Mills i M. O.J. Hawksford: „Distortion reduction in moving-coil loudspeaker systems using current-drive technology", w: *AES: Journal of the Audio Engineering Society* 37.3, 1989, s. 129–148

[15] Gaël Pillonnet, Eric Sturtzer, Timothé Rossignol, Pascal Tournier i Guy Lemarquand: „Distortion Improvement in the Current Coil of a Loudspeaker", w: *Audio Engineering Society Convention 134*, 2013

[16] Current Sources, Rosalfonso Bortoni, Sidnei Noceti Filho i Rui Seara: „Comparative Analysis of Moving-Coil Loudspeakers Driven by Voltage and Current Sources", w: *Audio Engineering Society Convention 115*, 2003, s. 1–14

[17] Rosalfonso Bortoni, Sidnei Noceti Filho i Rui Seara: „Comparative Analysis of Moving-Coil Loudspeakers Driven by Voltage and Current Sources", w: *Audio Engineering Society Convention 115*, 2003, s. 1–14

[18] „Time-varying non linear modeling of electrodynamic loudspeakers", w: *Applied Acoustics* 70.3, 2009, s. 450–458

METODYKA WYZNACZANIA CHARAKTERYSTYK KIERUNKOWOŚCI GŁOŚNIKÓW MODÓW ROZPROSZONYCH

K. Czesak[1], P. Kleczkowski[1]

[1]Akademia Górniczo – Hutnicza im. Stanisława Staszica w Krakowie

czesak@agh.edu.pl

Streszczenie

Głośniki modów rozproszonych charakteryzują się o w wiele szerszym kątem promieniowania, niż przetworniki tłokowe. Ich działanie przypomina niekoherentne źródła dźwięku i wytwarzają one pole akustyczne o bardziej rozproszonej charakterystyce. Umieszczenie tego typu głośnika w rzeczywistym pomieszczeniu prowadzi do uzyskania pola akustycznego o innych właściwościach, niż uzyskiwane w wyniku działania konwencjonalnego przetwornika.

Charakterystyki amplitudowo – częstotliwościowe uzyskiwane na drodze pomiarów głośników modów rozproszonych, w sąsiednich punktach pomiarowych na sferze, mogą znacząco różnić się między sobą. Co więcej – w przeciwieństwie do konwencjonalnych przetworników tłokowych – głośniki modów rozproszonych nie wytwarzają maksymalnych wartości ciśnienia akustycznego na osi przetwornika. Celem niniejszej pracy było ujawnienie szczegółowych charakterystyk kierunkowego promieniowania głośników modów rozproszonych w trzech wymiarach.

Niniejsza praca poświęcona została planowaniu oraz przygotowywaniu procedury pomiaru charakterystyk amplitudowo - częstotliwościowych promieniowania głośników modów rozproszonych w przestrzeni, a także problemom napotkanym w czasie dokonywania tego typu pomiarów oraz analizy zebranych danych.

1. Wstęp

Cechą charakterystyczną głośników modów rozproszonych (DML, ang.: *Distributed Mode Loudspeakers)*, która w największym stopniu odróżnia je od konwencjonalnych przetworników tłokowych jest o wiele mniejsze nasilenie zjawiska zawężania szerokości promieniowanej wiązki wraz ze wzrostem częstotliwości pobudzenia [1]. Także poziom ciśnienia akustycznego, mierzony na osi prostopadłej do powierzchni płaskiego panela głośnikowego nie osiąga najwyższych wartości spośród wszystkich punktów pomiarowych rozłożonych na półsferze. Wartym odnotowania jest też fakt, że charakterystyki amplitudowo-częstotliwościowe uzyskiwane w dwóch sąsiednich punktach pomiarowych (dla tego samego kąta inklinacji, jednak przy zmiennym kacie azymutalnym) bardzo często znacząco różnią się od siebie, co znacznie utrudnia wykreślenie charakterystyk kierunkowości dla głośnika modów rozproszonych, przy potraktowaniu go jak konwencjonalnego przetwornika [2].

Odnotowane właściwości DML nie pozostają bez wpływu na ich zachowanie w rzeczywistym pomieszczeniu odsłuchowym, gdzie występują odbicia dźwięku od ścian, podłogi, sufitu i elementów wyposażenia pomieszczenia. Przetworniki tego typu charakteryzuje niższa sprawność niż konwencjonalnych głośników tłokowych, jednak dzięki bardziej rozproszonej charakterystyce promieniowanego przez nie dźwięku, ich działanie prowadzi do innych rozkładów postrzeganej głośności w pomieszczeniu, niż przy zastosowaniu konwencjonalnych przetworników elektroakustycznych [3]. Powyższe rozważania prowadzą do konkluzji, że zjawiska odbicia dźwięku w pomieszczeniu stają się mniej destrukcyjne dla słuchacza w przypadku działania w nim głośników modów rozproszonych. Emisja bardziej rozproszonej wiązki teoretycznie pozwala więc na reprodukcję dźwięku w jakości lepszej niż przy pomocy głośników tłokowych. Uzyskanie jednak zadowalającego poziomu ciśnienia akustycznego w miejscu odsłuchu, w przypadku zastosowania DML wymaga znacząco mniejszej amplitudy pobudzenia niż przy korzystaniu z konwencjonalnych przetworników.

2. Rozmieszczenie punktów pomiarowych na półsferze

Każdy z badanych głośników modów rozproszonych poddawany był pomiarowi charakterystyki amplitudowo-częstotliwościowej w 325 punktach, rozmieszczonych na półsferze, z rozdzielczością kątową równą 10º (1 punkt na osi prostopadłej do płaszczyzny przetwornika oraz po 36 punktów dla dziewięciu kątów inklinacji). Pomiar na półsferze był wystarczający, ponieważ badane były głośniki DML umieszczone w obudowach zamkniętych. Realizacja pomiarów charakterystyk amplitudowo-częstotliwościowych na półsferze miała miejsce w dużej komorze bezechowej Laboratorium Akustyki Technicznej w Katedrze Mechaniki i Wibroakustyki na Wydziale Inżynierii Mechanicznej i Robotyki Akademii Górniczo – Hutniczej im. S. Staszica w Krakowie. Kąt inklinacji zadawany był poprzez podnoszenie i opuszczanie mikrofonu z użyciem zainstalowanego w komorze zautomatyzowanego ramienia, zaś kat azymutalny - poprzez obrót stolika, do którego przytwierdzony był DML. Oba wymienione urządzenia sprzężone były z komputerem klasy PC zaprogramowanym w środowisku LabVIEW, co umożliwiało sprawne i dokładne (do dziesiątych części stopnia) zadawanie kolejnych punktów pomiarowych. System umożliwiał też precyzyjną kontrolę promienia półsfery, który w omawianym przypadku wynosił 2,5 metra Była to maksymalna odległość mikrofonu pomiarowego od środka stolika obrotowego, technicznie możliwa w dostępnym systemie pomiarowym. Decyzja o zastosowaniu możliwie największej odległości mikrofonu od DML podyktowana była założeniem o ograniczeniu warunków pola bliskiego. Przykładowa konfiguracja do pomiaru w jednym z punktów o niewielkim kącie inklinacji przedstawia rysunek 1.

Rysunek 1 Rozmieszczenie badanego przetwornika oraz mikrofonu pomiarowego w komorze bezechowej.

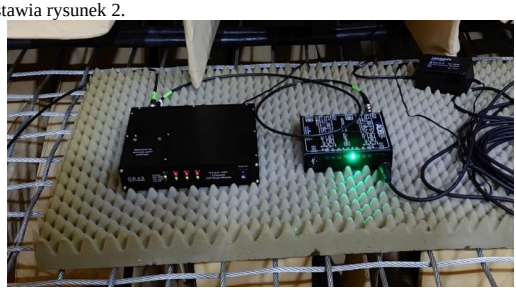
Z zachowania stałej rozdzielczości kątowej prowadzonych pomiarów wynika jednak potrzeba korekty przy obliczaniu częstotliwościowej charakterystyki mocy promieniowanej przez głośnik. Otóż – wraz z kątem inklinacji, zmienia się odległość sąsiednich punktów pomiarowych od siebie, co przekłada się na różne – co do wielkości - części powierzchni półsfery pomiarowej, reprezentowane przez poszczególne punkty pomiarowe [4]. Z uwagi na niejednoznaczność podejścia do tego tematu w literaturze, obecnie autorzy pracują nad własnym modelem obliczenia wag. Przykładowe wartości literaturowe wag zostały zestawione w tabeli 1.

Tabela 1 Wagi dla punków pomiarowych o poszczególnych inklinacjach [5].

l. p	kąt inklinacji [º]	waga
0	0	0,002417944
1	10	0,004730189
2	20	0,008955027
3	30	0,012387354
4	40	0,014989611
5	50	0,016868154
6	60	0,018165962
7	70	0,019006744
8	80	0,019477787
9	90	0,019629373

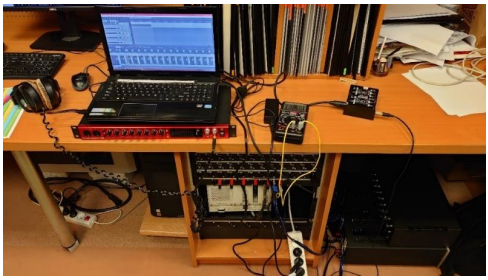
3. Tor pomiarowy i sygnał wymuszający

Pomiary charakterystyk amplitudo-częstotliwościowych głośników modów rozproszonych, w funkcji kierunkowości prowadzone były za pomocą mikrofonu pola swobodnego G.R.A.S 46AE. Sygnał z mikrofonu (przytwierdzonego do ramienia pomiarowego) trafiał za pośrednictwem przewodu BNC do kondycjonera G.R.A.S 12AK, który również zapewniał zasilanie mikrofonu. Sygnał z kondycjonera podawany był do aktywnego urządzenia typu direct injection box – ART dADB, zapewniającego symetryzację sygnału (pożądaną w przesyle analogowego sygnału audio na odległości większe niż 2m) oraz jego separację galwaniczną od przetwornika analogowo-cyfrowego. Urządzenia zainstalowane w komorze bezechowej przedstawia rysunek 2.



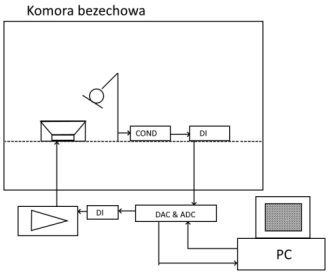
Rysunek 2 Elementy toru pomiarowego zainstalowane w komorze bezechowej.

Poza komorą, w pomieszczeniu kontrolnym użyte zostało urządzenie Focusrite Clarett 8Pre USB, łączące w sobie przedwzmacniacze mikrofonowe, przetworniki analogowo-cyfrowe (A/C), cyfrowo-analogowe (C/A) oraz wielokanałowy zewnętrzny interfejs dźwiękowy. Urządzenie to przyjmuje na wejście sygnały symetryczne oraz ma możliwość zasilania przyłączonych urządzeń wejściowych napięciem +48V. Wyjścia sygnałowe tegoż urządzenia również są symetryczne. Do jednego z wyjść, za pośrednictwem pasywnego urządzenia typu direct injection box RH-Sound DB-02 (użytego w celu desymetryzacji sygnału oraz separacji galwanicznej wzmacniacza mocy od przetwornika C/A), przyłączony został wzmacniacz mocy Anthem PVA-7, zasilający badane przetworniki elektroakustyczne sygnałem wymuszającym. Równolegle z płaskim panelem głośnikowym, do wyjścia wzmacniacza przyłączony był woltomierz. Sygnał wracający z komory bezechowej kontrolowany był przy użyciu słuchawek Beyerdynamic DT-990. Urządzenia zainstalowane w pomieszczeniu kontrolnym przedstawia rysunek 3.



Rysunek 3 Elementy toru pomiarowego zainstalowane w pomieszczeniu kontrolnym.

Urządzenie Focusrite Clarett 8Pre USB zostało sprzęgnięte z komputerem klasy PC, na którym zainstalowane było oprogramowanie Cockos Reaper. W programie tym dokonywana była rejestracja mierzonych przebiegów do plików wav o częstotliwości próbkowania 96kHz i rozdzielczości bitowej 24 bity. Zapis wyzwalany był synchronicznie z odtwarzaniem przebiegu wymuszającego. Uproszczony schemat toru pomiarowego przedstawia rysunek 4.

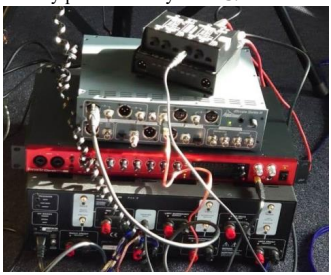


Rysunek 4 Schemat blokowy toru pomiarowego.

Celem minimalizacji błędów pomiarowych jako sygnał wymuszający zastosowano sinus przestrajany liniowo w zakresie częstotliwości 70Hz – 20kHz. Sygnał ten zapisany był w pliku wav o częstotliwości próbkowania 96kHz i rozdzielczości bitowej 24 bity. Długość wektora próbek wynosiła 2^{22} = 4 194 304 próbki. Przy takim doborze parametrów aktywny czas trwania pojedynczego sygnału wymuszającego wynosił 36 s. Dolne ograniczenie zakresu częstotliwości wymuszenia wynikło z dolnej częstotliwości granicznej badanych głośników, która leży wyraźnie powyżej częstotliwości 70Hz. Badanie skuteczności poniżej dolnej częstotliwości granicznej było niewskazane, ponieważ producent zaleca ostrożność przy pobudzaniu tych głośników w niskich częstotliwościach. Przy impedancji każdego z przetworników równej 4Ω, mającej charakter głownie rezystancyjny, wykazano szczególną dbałość, by woltomierz podłączony do wyjścia wzmacniacza mocy – równolegle z płaskim panelem głośnikowym wskazywał wartość ~2,83V. Zauważone wahania tej wartości stały się powodem weryfikacji elementów toru pomiarowego.

4. Weryfikacja elementów toru pomiarowego

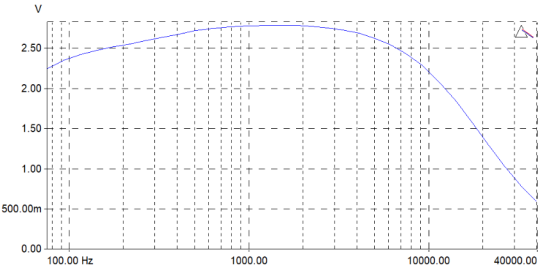
Zauważone wahania napięcia na wyjściu wzmacniacza mocy spowodowały konieczność weryfikacji charakterystyki amplitudowo częstotliwościowe elementów toru pomiarowego. O ile płaskość tychże charakterystyk w przypadku urządzeń z założenia przeznaczonych do pomiarów akustycznych (mikrofon i kondycjoner) nie budziła wątpliwości, o tyle pojawiły się one w przypadku samego wzmacniacza mocy, przetwornika (C/A) oraz pasywnego urządzenia typu DI-box separującego galwanicznie przetwornik C/A i wzmacniacz mocy. Pomiary powadzone były przy użyciu systemu Prism dScope III i objęły następujące układy: sam wzmacniacz mocy, wzmacniacz mocy + DI-box, wzmacniacz mocy + przetwornik C/A, wzmacniacz mocy + przetwornik C/A + Di-box. Każdy z pomiarów został powtórzony dla obciążenia czysto rezystancyjnego o wartości 4Ω, oraz dla obciążenia jednym z badanych płaskich paneli głośnikowych. Zestawiony tor pomiarowy przedstawia rysunek 5.



Rysunek 5 Tor pomiarowy do badania charakterystyk amplitudowo – częstotliwościowych elementów toru pomiarowego wykorzystywanych w pomiarze DML.

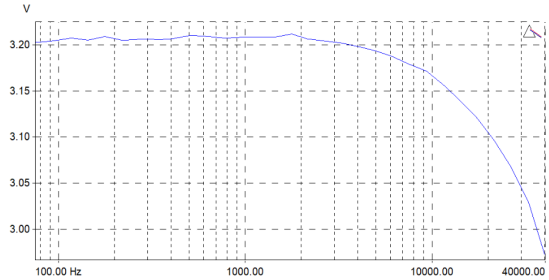
Pomiary wykazały istotny spadek poziomu napięcia przy częstotliwości 20 kHz na wyjściu wszystkich torów. Istotny spadek przy częstotliwości 70Hz względem częstotliwości 1 kHz zaobserwowano tylko w przypadku urządzenia DI-box.. W przypadku wszystkich badanych torów różnica poziomów między obciążeniem rezystancyjnym a obciążeniem głośnikiem DML nie przekroczyła 0,02 dB co potwierdza rezystancyjny charakter impedancji głośników DML.

Główny wkład w odchylenie poziomów na krańcach pasma wnosiło urządzenie DI-box, a w konsekwencji największe odchylenie wnosił tor: przetwornik C/A + DI-box + wzmacniacz mocy. Charakterystykę amplitudowo-częstotliwościową tego układu przedstawiono na rysunku 6. Wszystkie wyniki pomiarów charakterystyk amplitudowo-częstotliwościowych głośników DML zostały skorygowane za pomocą odpowiednio odwróconej krzywej z tego wykresu.



Rysunek 6 Charakterystyka amplitudowo – częstotliwościowa układu przetwornik C/A + DI-box + wzmacniacz mocy, obciążonego głośnikiem modów rozproszonych. Liniowa skala napięcia.

Na rysunku 7 zamieszczona została charakterystyka amplitudowo-częstotliwościowa powyższego układu, z pominięcie urządzenia typu Di-box:



Rysunek 7 Charakterystyka amplitudowo – częstotliwościowa układu: (C/A) + wzmacniacz mocy, obciążonego głośnikiem modów rozproszonych. Liniowa skala napięcia.

Wartości napięć zmierzonych w funkcji częstotliwości na wyjściu wzmacniacza mocy dla konfiguracji: przetwornik C/A + DI-box + wzmacniacz mocy (C/A + DI) oraz przetwornik C/A + wzmacniacz mocy (C/A) przedstawia tabela 2.

Tabela 2 Napięcie w funkcji częstotliwości mierzone na wyjściu wzmacniacza mocy, obciążonego DML

L.p.	0	1	2	3	4	5	6	7
Częstotliwość [Hz]	75	92	114	141	173	214	263	325
Napięcie (C/A) [V]	3,2029	3,2051	3,2079	3,2052	3,2091	3,2054	3,2064	3,2059
Napięcie (C/A + DI) [V]	2,2572	2,3584	2,4314	2,4794	2,5262	2,5619	2,6031	2,6432

L.p.	8	9	10	11	12	13	14	15
Częstotliwość [Hz]	400	493	608	750	924	1140	1405	1732,1
Napięcie (C/A) [V]	3,2069	3,2104	3,2094	3,2079	3,2087	3,2087	3,2091	3,2123
Napięcie (C/A + DI) [V]	2,6851	2,725	2,7514	2,7725	2,7845	2,7927	2,7946	2,7903
L.p.	16	17	18	19	20	21	22	23
Częstotliwość [Hz]	2135,3	2632,5	3245,3	4000,9	4932,4	6080,8	7496,5	9241,9
Napięcie (C/A) [V]	3,2067	3,2044	3,2019	3,1982	3,1937	3,1876	3,1792	3,1714
Napięcie (C/A + DI) [V]	2,7795	2,7626	2,7348	2,6937	2,6339	2,5497	2,4331	2,2808
L.p.	24	25	26	27	28	29	30	
Częstotliwość [Hz]	11394	14046	17317	21348	26318	32446	40000	
Napięcie (C/A) [V]	3,1565	3,1389	3,1203	3,0957	3,0667	3,0288	2,9684	
Napięcie (C/A + DI) [V]	2,0821	1,8448	1,5795	1,3008	1,0309	0,7861	0,5762	

5. Wnioski

Przeprowadzenie dokładnych pomiarów przetwornika elektroakustycznego niekonwencjonalnego rodzaju, wymaga szczegółowego planowania, ponieważ standardowo stosowane techniki mogą być nieprzydatne. Poznanie właściwości głośnika DML wymaga precyzyjnego pomiaru jego charakterystyk częstotliwościowych na całej półsferze. Za wskazane uznano na przykład zastosowanie separacji galwanicznej elementów toru pomiarowego aby uniknąć pojawienia się w rejestrowanym synale składowej 50Hz oraz jej harmonicznych. To z kolei spowodowało konieczność wprowadzenia odpowiedniej korekty charakterystyki częstotliwościowej. Jednak korekta ta byłaby niezbędna nawet gdyby nie zastosowano separacji, ponieważ inne elementy toru również wprowadzają znaczące zniekształcenia liniowe. W przypadku wzmacniacza mocy okazały się one większe niż deklarowane przez producenta..

6. Podziękowania

Autorzy pragną złożyć podziękowania pracownikom Laboratorium Akustyki Technicznej w Katedrze Mechaniki i Wibroakustyki na Wydziale Inżynierii Mechanicznej i Robotyki Akademii Górniczo – Hutniczej imienia Stanisława Staszica w Krakowie, w szczególności panom dr inż. Arturowi Flachowi, dr inż. Jarosławowi Rubasze oraz mgr inż. Bartłomiejowi Chojnackiemu za udostępnienie komory bezechowej oraz aparatury, jak również przeszkolenie z użytkowania systemu pozycjonowania mikrofonu pomiarowego. Badania były częściowo finansowane z subwencji badawczej Wydziału Inżynierii Mechanicznej i Robotyki AGH, nr: 16.16.130.942/Czesak, oraz z projektu Zintegrowany Program Rozwoju Akademii Górniczo-Hutniczej w Krakowie II, nr: POWR.03.05.00-00-Z309/18-00

Bibliografia

[1] N. Harris: „Spatial Bandwidth of Diffuse Radiation in Distributed-Mode Loudspeakers" , w: *111th AES Convention, 2001 September 21-24 New York, NY, USA*, AES, USA, 2001

[2] V. P. Gontcharov, N. P. R. Hill, V. J. Taylor: „Measurement Aspects of Distributed Mode Loudspeakers", w: *106th AES Convention, 1999 May 8-11 Munich, Germany*, AES, USA, 1999

[3] S. Flanagan, N. Harris: "Loudness – a study of the subjective Loudness Difference between DML and Conventional Loudspeakers", w: *106th AES Convention, 1999 May 8-11 Munich, Germany*, AES, USA, 1999

[4] M. C. Heilemann D. A. Anderson, S. Roesner, M. F. Bocko: "Quantifying Listener Preference of Flat-Panel Loudspeakers" , w: *145th AES Convention, 2018 October 17-20 New York, NY, USA*, AES, USA, 2018

[5] G. L. Wilson: "Directivity Factor: Q and R_θ? Standard Terminology and Measurement Methods", w: *Journal of Audio Engineerig Society*, 21 (10), 1973, s. 828 - 833

WPŁYW ZMIANY OBJĘTOŚCI PUDŁA REZONANSOWEGO NA BRZMIENIE CHORDOFONU SZARPANEGO

J. Jasiński[1]

[1]Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie

jjasinsk@agh.edu.pl

Streszczenie

Tematem pracy jest analiza możliwości wpływania na barwę dźwięku chordofonu szarpanego poprzez modyfikację objętości jego komory rezonansowej. Celem pracy jest zaobserwowanie czy zakres możliwych zmian charakterystyk brzmieniowych jest wystarczająco istotny, aby zasadny był dalszy rozwój tej technologii. Zaprojektowany i zbudowany został prototyp instrumentu z komorą rezonansową o zmiennej objętości. Został on stworzony na bazie gitary klasycznej, w której ruchoma płyta tylna umożliwia zmianę głębokości jej komory rezonansowej. Przeprowadzone zostały badania sprawdzające zakres możliwych modyfikacji brzmieniowych mechanizmu poprzez nagranie pojedynczych wymuszeń struny dla różnych objętości komory. Analiza wyników pomiarów, z wykorzystaniem wielu parametrów barwowych, wykazała, że zmiana głębokości pudła instrumentu w znaczący i słyszalny sposób może wpłynąć na brzmienie instrumentu. Przeprowadzone badania wskazują, że zaprojektowane rozwiązanie pozwala osiągnąć zmianę brzmienia instrumentu strunowego w stopniu, który uzasadnia dalszy rozwój zaproponowanej technologii.

1. Wstęp

Proces rozwoju instrumentów akustycznych trwa zazwyczaj wiele wieków. Pokolenia lutników doskonaliło wzorce konstrukcyjne tworząc standardy, które obecnie można wykorzystywać przy budowie instrumentów klasycznych. W związku z rozwojem technologii zjawisko to zostało jednak w znaczącym stopniu przyśpieszone. Szczególnym przykładem tego procesu może być gitara, której postać klasyczna zaczęła być modyfikowana wraz z wzrostem zapotrzebowania na głośniejsze instrumenty. Zawdzięczamy temu powstanie gitary akustycznej w jej wielu postaciach, a następnie gitary elektrycznej. W trakcie tego rozwoju doszło zatem do zastąpienia pojedynczych form gitar wieloma różnymi modelami, które teraz znajdują swoje zastosowanie w różnych kontekstach muzycznych, zależnie od barwy i dynamiki, których potrzebuje muzyk. Spowodowało to jednak, że twórcy wykorzystują wiele osobnych instrumentów. Pojawia się jednak pytanie, czy nie byłoby możliwe zastąpienie wielu różnorodnych barwowo instrumentów akustycznych pojedynczym instrumentem, którego charakterystykę można modyfikować zależnie od potrzeb.

Celem pracy było stworzenie instrumentu akustycznego, który zawiera mechanizm umożliwiający modyfikację jego brzmienia poprzez modyfikację parametrów fizycznych instrumentu oraz sprawdzenie zakresu zmian, jaki zostanie przez takie rozwiązanie umożliwiony. Regulacja brzmienia instrumentu nie ma z założenia następować w trakcie gry, lecz w ramach przygotowań do nagrania lub występu. Jako podstawa do modyfikacji została wybrana gitara klasyczna w związku z jej powszechnym zastosowaniem w muzyce rozrywkowej oraz prostotą konstrukcyjną, która ułatwiła proces budowy prototypu. Warto zaznaczyć, iż wyniki wykonywanych w ramach tej pracy badań nie są przeznaczone do wykorzystywania w trakcie procesu projektowania i budowy samych gitar. Konstruowany w ramach tej pracy instrument będzie w zdecydowanie zbyt dużym stopniu zmodyfikowany, by móc na jego podstawie wyciągać wnioski dotyczące gitary, na której bazuje.

2. Opis rozwiązania

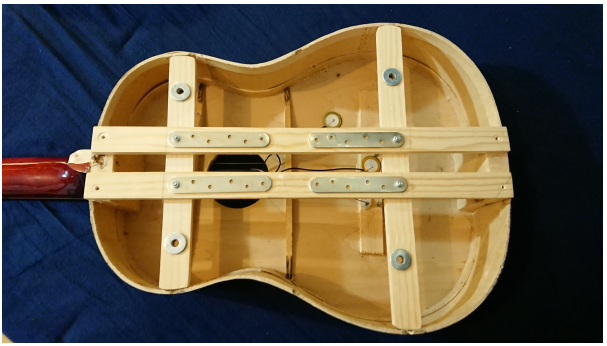
2.1. Projektowanie mechanizmu modyfikującego brzmienie chordofonu

Spośród wszystkich parametrów fizycznych, które mogą być zmieniane wewnątrz systemu, jakim jest instrument, w niniejszej pracy analizie został poddany wpływ zmiany objętości jego komory rezonansowej na brzmienie. Rozważane jednak były również inne rozwiązania, takie jak zmiana wielkości otworu rezonansowego, przykładanie dodatkowego napięcia mechanicznego do płyty wierzchniej, regulacja sztywności płyty wierzchniej oraz modyfikacja rozkładu masy w kluczowych miejscach instrumentu. Wszystkie te rozwiązania zostały jednak ocenione jako mające mniejsze prawdopodobieństwo na wywarcie satysfakcjonującego wpływu na barwę instrumentu [1] i będą rozważane w kolejnych pracach badawczych. Objętość instrumentu może być zmieniana na wiele sposobów, jednak wybrana została regulacja ruchomą płytą tylną instrumentu i jej odległością od płyty wierzchniej. Zasada działania mechanizmu przedstawiona została na schemacie na rysunku 1.

Rysunek 1 Schemat działania mechanizmu modyfikującego objętość komory rezonansowej instrumentu

2.2. Skonstruowany prototyp

Prototyp rozwiązania został skonstruowany drogą modyfikacji gitary klasycznej firmy String Instruments. Po usunięciu płyty tylnej, na jej miejsce wprowadzona została konstrukcja umożliwiająca zmianę objętości. Pierwszym jej elementem jest drewniany szkielet, widoczny na rysunku 2, którego zadaniem jest zapewnienie mechanicznej sztywności instrumentu po usunięciu płyty tylnej oraz stworzenie punktu zaczepu dla mechanizmu zmiany objętości instrumentu.



Rysunek 2 Zdjęcie drewnianego szkieletu wprowadzonego w miejsce usuniętej płyty tylnej

Drugim elementem tego mechanizmu jest płyta wykonana ze sklejki, wycięta w kształcie wnętrza pudła rezonansowego, przedstawiona na rysunku 3. Płyta ta otoczona jest uszczelką, która po wprowadzeniu płyty do wnętrza pudła rezonansowego zapewnia szczelność komory rezonansowej od strony tylnej.

Rysunek 3 Zdjęcie wewnętrznej płyty instrumentu zamontowanej do szkieletu konstrukcyjnego

Płyta zamontowana jest do szkieletu przy pomocy czterech śrub, których regulacja umożliwia zmianę głębokości pudła rezonansowego.

3. Opis eksperymentu

3.1. Stanowisko pomiarowe

W celu zwiększenia wiarygodności otrzymanych wyników, podjęto szereg działań mających na celu wyeliminowanie potencjalnych czynników zewnętrznych, które mogłyby zaburzyć wykonywany eksperyment. Kluczowym elementem stanowiska pomiarowego jest stworzony specjalnie spust, który wykorzystuje sprężynę piórową z zamocowanym piórkiem gitarowym tak, aby w powtarzalny sposób szarpać strunę podczas kolejnych pomiarów. Skonstruowany został również statyw zapewniający montaż spustu w stałej pozycji względem struny chordofonu oraz utrzymujący instrument w stabilnym ułożeniu względem mikrofonów pomiarowych, jednocześnie umożliwiając regulację objętości komory rezonansowej bez potrzeby jego zdejmowania. Rysunek 4 przedstawia zdjęcie stanowiska pomiarowego.

Rysunek 4 Zdjęcie konfiguracji stanowiska badawczego do pomiarów zależności brzmienia chordofonu szarpanego od głębokości pudła rezonansowego

Celem eksperymentu jest obserwacja zmian w barwie instrumentu w muzycznym kontekście, zatem do pomiarów zostały wybrane mikrofony estradowe: Rode NT2 ustawiony w odległości 20 cm od instrumentu naprzeciw łączenia pudła rezonansowego z gryfem[2] oraz Shure SM57 umieszczony w odległości 15 cm nad mostkiem instrumentu i skierowany na środek przestrzeni pomiędzy mostkiem, a otworem rezonansowym[2]. Dodatkowo wewnątrz instrumentu zamontowany został przetwornik piezoelektryczny. Sygnały zarejestrowane zostały przy wykorzystaniu karty dźwiękowej Focusrite Scarlett 18i8.

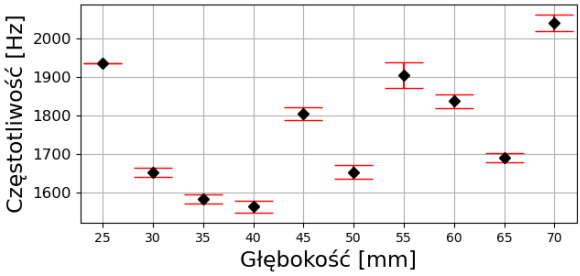
3.2. Wykonanie pomiarów

W ramach eksperymentu przeprowadzone zostały serie pomiarowe dla dziesięciu różnych głębokości pudła rezonansowego instrumentu dla dźwięku E (82,41 Hz) i dla trzech ustawień głębokości dla dźwięków G (98,0 Hz), A# (116,54 Hz) i c# (138,59 Hz). W ramach każdej serii rejestrowano dziesięć wymuszeń instrumentu przy pomocy spustu wraz z pełnym wybrzmieniem dźwięku.

4. Analiza wyników

Uzyskane próbki poddane zostały analizie parametrycznej w ramach której wyznaczony został następujący zbiór parametrów: *Spectrum Centroid [3], Spectrum Spread [3], High Energy-Low Energy Ratio [3], Tristimulus [3], Inharmonicity [3], Noisiness [3], Odd-Even Ratio [3], Zero Crossing Rate [4], Entropy [4]*. Parametry te zostały wybrane na podstawie prac analizujących korelację pomiędzy percepcją barwy dźwięku, semantycznymi określeniami wykorzystywanymi do jej opisu oraz zmianami w analizie parametrycznej [4][5][6]. Na potrzeby niniejszej pracy przeanalizowane zostaną nagrania uzyskane mikrofonem Shure SM57 dla wysokości dźwięku E. Analiza pozostałych sygnałów potwierdza wyniki i wnioski jakie zostaną przedstawione poniżej.

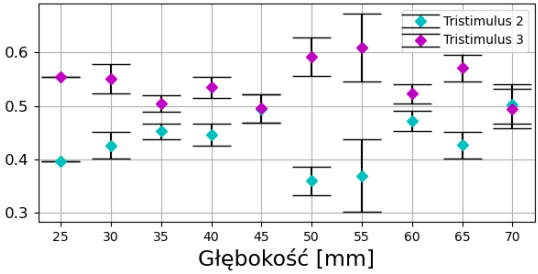
Jednym z najistotniejszych parametrów wykorzystywanych do analizy barwy instrumentów jest środek ciężkości widma, który w wielu badaniach pokazany został jako silnie korelujący z percepcją jasności barwy dźwięku [7]. Rozważając wartości Spectrum Centroid dla różnych objętości komory rezonansowej instrumentu, przedstawione na wykresie 5, możemy zaobserwować znaczące różnice pomiędzy nimi.



Rysunek 5 Wykres wartości środka ciężkości widma w zależności od głębokości komory rezonansowej instrumentu wraz z zaznaczonymi odchyleniami standardowymi w ramach serii pomiarowych

Zmiany dźwięku, powodujące zróżnicowanie wartości tego parametru w zakresie 1600–2000 Hz będą niezwykle znaczące z punktu widzenia jego barwy i sposobu w jaki instrument będzie mógł być wykorzystywany w kontekście muzycznym. Podobny zakres zmian możemy zaobserwować w kolejnych parametrach barwowych: zero crossing rate (0,04-0.07s) i noisiness (0.025-0.175), które opisują szumowość sygnału i stosunek energii szumowej do energii harmonicznych w dźwięku [3]. Oba te parametry również zostały określone jako skorelowane z odczuwaną jasnością i głębią barwy dźwięku [4] [5].

Warte rozważenia są również parametry opisujące struktury harmoniczne widoczne w widmie amplitudowym sygnału. Takim parametrem jest Tristimulus, definiowany jako stosunek amplitud poszczególnych harmonicznych sygnału do sumy amplitud wszystkich harmonicznych. Na rysunku 6 przedstawiono wartości Tristimulus 2 (stosunek amplitud harmonicznych 2, 3 i 4 do sumy amplitud wszystkich harmonicznych sygnału) i Tristimulus 3 (stosunek amplitud harmonicznych powyżej piątej do sumy amplitud wszystkich harmonicznych) [3].



Rysunek 6 Wykres wartości parametrów tristimulus 2 i tristimulus 3 w zależności od głębokości komory rezonansowej instrumentu wraz z zaznaczonymi odchyleniami standardowymi w ramach serii pomiarowych

Zakres różnic zaobserwowanych przy różnych objętościach badanego chordofonu wynosi kolejno 0.36-0.5 i 0.5-0.6. Parametry te wykazują silną korelację z odczuwaną ostrością barwy dźwięku i jego fakturą [5] co wskazuję na to, że opisane różnice mogą mieć wpływ na to jak instrument jest odbierany przez słuchaczy. Analiza widm amplitudowych impulsów instrumentu z podziałem na fazy wybrzmienia pokazała również, że stosunki pomiędzy amplitudami harmonicznych ewoluują wraz z trwaniem dźwięku. Sposób w jaki dochodzi do ewolucji brzmienia również jest zależny od objętości komory rezonansowej.

Na podstawie analizy pozostałych parametrów wyciągnąć można dodatkowe wnioski:

- Dokonywane modyfikacje nie mają znaczącego wpływu na nieharmoniczność dźwięku,
- Zmiana objętości komory rezonansowej w znaczący sposób wpływa na czas wybrzmiewania instrumentu - czasy pomiędzy osiągnięciem maksimum obwiedni sygnału a jej opadnięciem poniżej wartości granicznej różniły się ponad dwukrotnie dla różnych konfiguracji mechanizmu,
- Skonstruowany mechanizm może zauważalnie wpływać na głośność instrumentu - energia impulsów (RMS) różniła się o 140% dla różnych głębokości komory,
- Analiza strojenia instrumentu pomiędzy poszczególnymi konfiguracjami wykazała, że maksymalne odchylenie pomiędzy strojeniami wynosiło poniżej sześciu centów. Wartość ta jest znacząco poniżej progu rozróżnialności przeciętnego słuchacza [8] i wskazuje na to, że nie powinien to być czynnik wpływający na wyniki analizy,
- Niskie wartości odchylenia standardowego uzyskane dla serii pomiarowych pokazują satysfakcjonujący poziom powtarzalności skonstruowanego mechanizmu wzbudzającego strunę.

5. Wnioski

Celem niniejszej pracy było zbadanie możliwości modyfikacji brzmienia instrumentu akustycznego przez zmianę jego właściwości fizycznych. Zaprojektowany został instrument strunowy, w którym objętość komory rezonansowej jest regulowana poprzez zmianę pozycji ruchomej płyty tylnej. Stworzony został prototyp opisanego instrumentu i wykonane zostały nagrania dla różnych jego konfiguracji.

Analiza uzyskanych nagrań pokazuje, że różnice w brzmieniu instrumentu uzyskane poprzez zmianę objętości jego komory rezonansowej są wyraźnie widoczne. Znaczące różnice w wartościach środka ciężkości widma, noisiness i tristimulus wskazują na to, że mechanizm może modyfikować barwę instrumentu, a w szczególności jego jasność i fakturę [4] [5] [6]. Różnice w czasie ataku i wybrzmienia pokazują wpływ objętości komory na cechy czasowe instrumenty, natomiast zmiany w parametrach energetycznych przedstawiają możliwość modyfikacji jego głośności.

Zasadne jest stwierdzenie, że poprzez wykorzystanie zaprojektowanego mechanizmu możliwa jest modyfikacja wszelkich aspektów brzmienia instrumentu dla szerokiego zakresu dźwięków. Uzyskane wyniki pokazują jednak niezwykle złożony charakter opisywanych zmian i nie umożliwiają przewidywania jak modyfikowana będzie tonalność instrumentu w związku z konkretnymi konfiguracjami mechanizmu. Można wnioskować zatem, że mechanizmy modyfikujące brzmienie chordofonów poprzez wpływ na ich parametry fizyczne mają potencjał znalezienia zastosowania muzycznego, jednak proces ich projektowania i zrozumienia wymaga dodatkowych badań.

Bibliografia

[1] S. Schwerman, N. Fletcher i T. Rossing: „The Physics of Musical Instruments", w: *Computer Music Journal* 18, 1991

[2] B. Owsinski: *The Recording Engineers Handbook*, Thomson Course Technology PTR, 2005

[3] Geoffroy Peeters, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis i Stephen McAdams: „The Timbre Toolbox: Extracting audio descriptors from musical signals", w: *The Journal of the Acoustical Society of America* 130.5, 2011, s. 2902–2916, DOI: *10.1121/1.3642604*

[4] Vinoo Alluri i Petri Toiviainen: „Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre", w: *Music Perception: An Interdisciplinary Journal* 27 (3), 2010, s. 223–242

[5] Asterios Zacharakis, Konstantinos Pastiadis i Joshua D. Reiss: „An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates", w: *Music Perception* 31.4, 2012, s. 339–358, DOI: *10.1525/mp.2014.31.4.339*

[6] Asterios Zacharakis, Konstantinos Pastiadis i Joshua D. Reiss: „An Interlanguage Unification of Musical Timbre", w: *Music Perception* 32.4, 2015, s. 394–412, DOI: *10.1525/mp.2015.32.4.394*

[7] Emery Schubert, Joe Wolfe i Alex Tarnopolsky: „Spectral centroid and timbre in complex, multiple instrumental textures", w: 2004

[8] Zarate J. M., Ritson C. R. i Poeppel D.: „Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary?", w: *Acoustical Society of America* 132(2), 2012

IMPLEMENTACJA I ZASTOSOWANIA TRÓJOSIOSWEGO OSCYLOSKOPU W WIZUALIZACJI DŹWIĘKU

W. Wronka[1]

[1]Katedra Mechaniki i Wibroakustyki, Akademia Górniczo-Hutnicza

wronka@agh.edu.pl

Streszczenie

Praca opisuje implementację systemu wizualizacji dźwięku czasu rzeczywistego, który przedstawia cechy sygnału w użyteczny i intuicyjny sposób. Opiera się na pomyśle przeniesienia obrazu tradycyjnego oscyloskopu w trzeci wymiar przestrzenny, w celu poszerzenia kontekstu czasowego odczytu lub prezentacji dodatkowych cech sygnału. Istotnym założeniem było zachowanie spójnej i intuicyjnej formy tradycyjnego oscyloskopu. System wizualizacji został stworzony w graficznym języku programowania Max, w szczególności w przeznaczonym do tworzenia trójwymiarowej grafiki i animacji środowisku Jitter, wykorzstującym OpenGL API. System został przetestowany w różnych scenariuszach, takich jak oparte na krzywych Lissajous wizualizacjach trójdźwięków lub trójwymiarowy goniometr do zastosowań w miksie sygnałów ambisonicznych.

1. Wstęp

Choć pierwszy oscyloskop powstał ponad 100 lat temu, podstawowa koncepcja jego działania pozostaje niezmieniona. Szerokie zastosowanie w laboratoriach elektronicznych i studiach nagraniowych dowodzi przydatności tego urządzenia.

Oscyloskop służy do obserwacji i badania przebiegów zależności pomiędzy dwoma wielkościami fizycznymi, interpretowanych jako funkcja napięcia elektrycznego w czasie. Sygnały mogą pochodzić z dowolnych przetworników, termometru, czujnika drgań, mikrofonu. Wyświetlanie odbywa się w dwóch trybach. Pierwszy z nich to synchroniczne przedstawianie sygnału w zależności od czasu. W tym trybie można skalować zarówno oś Y, jak i oś X, a zatem podstawę czasu, przedstawiając przykładowo kształt okresu sygnału lub jego obwiednię.

Drugi ze sposobów użycia oscyloskopu polega na wyświetleniu zamiast sygnału podstawy czasowej drugiego sygnału wejściowego (tryb X-Y). W takim wypadku różnice pomiędzy sygnałami objawiają się obecnością punktów wykresu poza prostą $y = x$. Konfiguracja ta jest w akustyce najczęściej używana do obserwacji przestrzenności sygnału stereofonicznego (goniometr), jak również do odczytu stosunku częstotliwości i przesunięcia fazowego za pomocą krzywych Lissajous.

Synchroniczny charakter obrazu wyświetlanego przez oscyloskop powoduje łatwość połączenia wizji z fonią (wizualizowanym dźwiękiem), co umożliwia połączenie intuicyjności odbioru z bogactwem zawartych informacji. Zalety te w połączeniu z możliwością zastosowania technologii cyfrowej spowodowały powstanie nowych urządzeń bazujących na oscyloskopie, dodających nowe funkcjonalności. Przykładem może być oscyloskop przedstawiający prawdopodobieństwo wartości sygnału jako jasność punktu wykresu [1].

2. Założenia

Powyższe spostrzeżenia przywiodły autora do zamiaru rozszerzenia obrazu oscyloskopu o trzeci wymiar przestrzenny. Przeznaczenie tego dodatkowego kanału informacji byłoby zależne od potrzeb użytkownika w konkretnym zastosowaniu wizualizacji. W związku z tym opracowano dwie podstawowe koncepcje:

- dwa sygnały akustyczne wyświetlane w osiach *X* i *Y*, z sygnałem w osi *Z* w roli podstawy czasowej, prezentujące obraz goniometru umożliwiający obserwację zmian w dłuższej perspektywie czasowej (rys. 1 i 3;
- trzy sygnały akustyczne wyświetlane w osiach *X*, *Y* i *Z*; do zastosowania w roli trójwymiarowego goniometru (rys. 6) lub wizualizacji trójdźwięków (rys. 4).

Istotnym założeniem było zachowanie podstawowych zalet oscyloskopu, a zatem wyświetlania w czasie rzeczywistym, spójności generowanych kształtów i intuicji odczytu. Dlatego też już na etapie koncepcyjnym wzięto pod uwagę złożoność aspektów percepcyjnych wizualizacji, wpływających na czytelność wykresów i w konsekwencji łatwy odczyt przedstawianych parametrów.

3. Implementacja

Projekt zrealizowano w graficznym języku programowania Max, z wykorzystaniem przeznaczonego do grafiki i animacji środowiska Jitter [2]. Na jego wybór wpłynął szereg pożądanych cech i funkcjonalności: takich jak szybkie i intuicyjne prototypowanie, działanie w czasie rzeczywistym i mnogość przydatnych w przedsięwziętym zadaniu funkcji. Dużą zaletą była również możliwość generowania trójwymiarowej animacji z wykorzystaniem biblioteki OpenGL [3], dokonując obliczeń na karcie graficznej.

W toku rozwijania systemu i testowania pierwszych prototypów dostrzeżono potrzebę wyraźniejszego zaakcentowania głębi obrazu trójwymiarowego. Wyświetlanie kolejnych próbek sygnału jako punkty (osobno lub połączone w linie) osadzone w przestrzeni było mało czytelne, a różnice odległości w osi prostopadłej do ekranu niewyraźne.

W związku z tym podjęto próbę nadania punktom wykresu fizycznych właściwości. Współrzędne kolejnych próbek sygnału posłużyły za źródło pozycji trójwymiarowych sfer, o ustalonym wymiarze i rzucanym cieniu. Efekt daje wrażenie ciągłości, jednak przy nagłych zmianach wypadkowej amplitudy (a zatem - zawartości wyższych składowych częstotliwościowych) można zaobserwować pojedyncze sfery, co może przypominać korale nawleczone na sznurek (rys. 1).

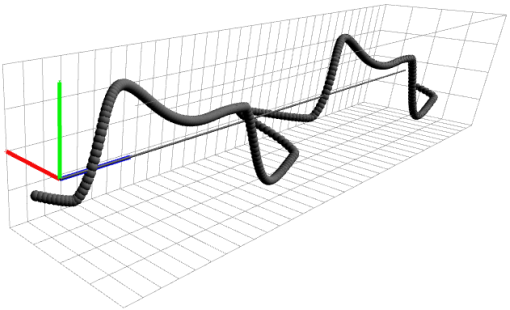
Zastosowanie powyższej techniki dało zauważalny efekt, problemem pozostała jednak nieoczywista perspektywa wyświetlanego obrazu. Interpretacja kształtów zawieszonych w wirtualnej przestrzeni przedstawia sobą pewną trudność z uwagi na brak punktów odniesienia, osadzających je w ustalonej perspektywie. Prostym rozwiązaniem tego problemu jest imlementacja prostopadłościennej siatki z podziałką, okalającej układ współrzędnych (rys. 1).

Trzecia z zastosowanych technik ułatwiających percepcję trójwymiarowego obrazu opiera się na zdolności rozróżniania barw. Korzystając z kolorów oznaczających osie wykresu X, Y i Z, odpowiednio czerwony, zielony i niebieski, przeniesiono Współrzędne punktów w przestrzeń barw RGB. W efekcie rzut punktu na daną oś oznacza zawartość danej barwy podstawowej w jego kolorze. Tęczowe barwy powstałej w konsekwencji wizualizacji mają niewiele wspólnego z naturalnymi mechanizmami percepcji głębi obrazu, ich interpretacja jest jednak stosunkowo intuicyjna i zdecydowanie pomocna. Technika okazała się najbardziej przydatna w wizualizacji trójdźwięków (rys. 4).

4. Zastosowania

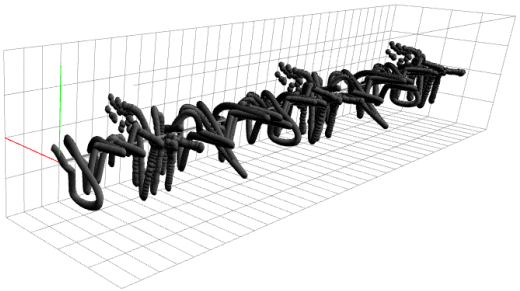
W celu dokonania subiektywnej oceny przydatności opracowanego rozwiązania dokonano szeregu testów w określonych scenariuszach zastosowań, związanych z obserwacją zmienności przestrzennej sygnału stereofonicznego. Przeprowadzono również próbę czytelności i przydatności systemu do wizualizacji sygnału ambisonicznego, jak i graficznej reprezentacji trójdźwięków i narzędzia pomocnego w treningu słuchu muzycznego.

Pierwsze z zastosowań dotyczy obserwacji periodycznego sygnału stereofonicznego w krótkim kontekście czasowym, porównywalnym z jego okresem. Ta metoda może być pomocna w tworzeniu przestrzennych syntezatorów analogowych lub algorytmów syntezy dźwięku. Rys.1 przedstawia sygnał periodyczny, którego lewy i prawy kanał stanowią przebieg piłokształtny i prostokątny, co odpowiada osiom *X* i *Y* wykresu, oznaczonych kolorami czerwonym i zielonym. Przebiegi zostały przefiltrowane dolnoprzepustowe, co wygładziło ich ostre wierzchołki.



Rysunek 1 Wizualizacja okresu periodycznego sygnału stereofonicznego. Osie X i Y odpowiadające kanałom lewemu i prawemu oznaczono kolorami czerwonym i zielonym,sygnał w niebieskiej osi Z to sygnał podstawy czasowej.

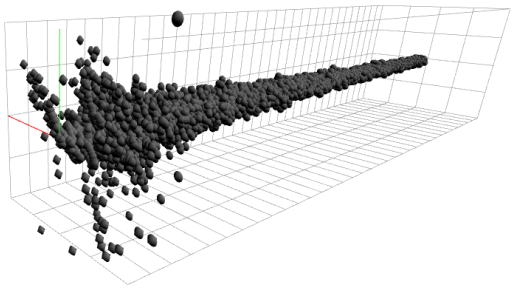
Analiza wizualizacji, w szczególności zaś odległości punktów od płaszczyny $y = x$, dostarcza informacji o różnicach pomiędzy składowymi sygnału, a w konsekwencji o obrazie przestrzennym dźwięku. W odróżnieniu od obrazu tradycyjnego oscyloskopu w trybie X-Y, wizualizacja wzbogacona jest o konteks czasowy. Warto nadmienić, że w zależności od punktu obserwacji można obserwować osobne przebiegi czasowe obu sygnałów (ustawiając widok równolegle z osiami X lub Y) lub obraz trybu X-Y oscyloskopu. Po odpowiednim przeskalowaniu sygnału podstawy czasowej wizualizacja umożliwia obserwację przestrzenności dłuższego dźwięku, na przykład fragmentu utworu (rys. 2.



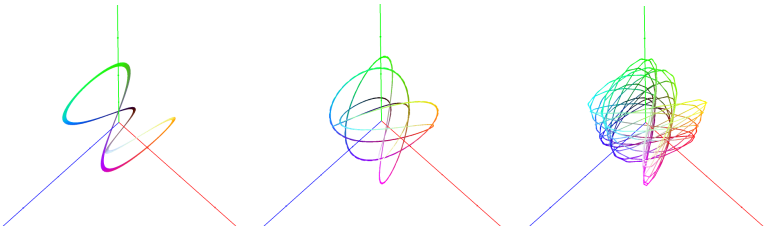
Rysunek 2 Wizualizacja fragmentu utworu muzycznego o długości 250 ms.

Wizualizacja może stanowić użyteczny element interfejsu graficznego wirtualnych instrumentów muzycznych i efektów w postaci wtyczek, w szczególności tych związanych z kreowaniem lub przetwarzaniem obrazu przestrzennego sygnału. Przykład może stanowić trójosiowy obraz stereofonicznej odpowiedzi impulsowej (rys. 3) generowanej przez pogłos algorytmiczny, zmieniającej się przy każdorazowym przestawieniu parametrów.

Jednym z ciekawszych i bardziej abstrakcyjnych przykładów użycia trójosiowego oscyloskopu jest wizualizacja trójdźwięków. W tym celu należy podać na wejście układu trzy sygnały sinusoidalne, a w efekcie wyświetlony zostanie kształt charakterystyczny dla stosunku ich częstotliwości. Przykład wizualizacji trzech wybranych trójdźwięków przedstawiono na rys. 4.



Rysunek 3 Przykład zastosowania wizualizacji w interfejsie graficznym pogłosu algorytmicznego, fragment stereofonicznej odpowiedzi impulsowej



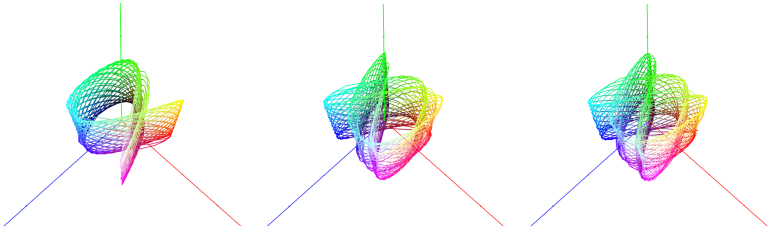
(a) 1:2:3 **(b)** 1:1,25:1,5 **(c)** 1:1,1875:1,5

Rysunek 4 Wizualizacje trójdźwięków w stroju naturalnym: trzy pierwsze częstotliwości szeregu harmonicznego (a), akord durowy (b), akord molowy (c). Pod rysunkami zawarto stosunki częstotliwości sygnałów składowych

Narastający stopień ich skomplikowania odpowiada zawartym w trójdźwiękach interwałów o charakterze dysonansowym. Co istotne, uzyskany kształt jest stabilny z uwagi na zastosowanie interwałów pochodzących ze stroju naturalnego. Wyrażone są one wówczas stosunkowo małymi liczbami naturalnymi, a w konsekwencji charakter interferencji sygnałów składowych jest niezmienny w czasie. Zastosowanie stroju równomiernie temperowanego powoduje drastyczne zmniejszenie stabilności kształtów (5), odpowiadające wrażeniu dudnienia częstotliwościowego. Dalsze odstrajanie częstotliwości powoduje wyraźne, okresowe spłaszczenie kształtu w osi prostopadłej do osi odstrajanego sygnału.

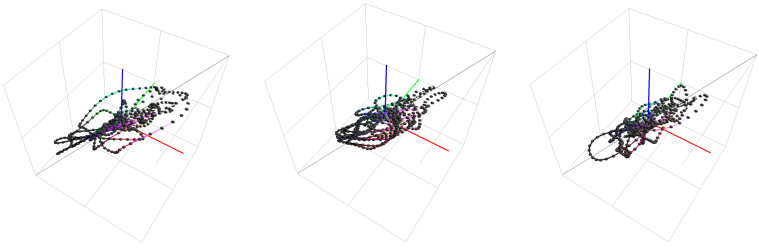
Jakkolwiek praktyczne zastosowania graficznej interpretacji trójdźwięków mogą zdawać się wątpliwe, tak z pewnością może służyć za wartościową pomoc naukową w treningu słuchu muzycznego i podstaw harmonii, angażującą intuicyjne analogie konsonansu ze stabilnością. Może również stanowić narzędzie wizualnego porównania różnych systemów tonalnych, uwzględniające nie tylko charakter występujących w nich interwałów, lecz również trójdźwięków.

Ostatnie z przedstawionych zastosowań przewiduje użycie trójwymiarowego goniometru w podstawowej obserwacji sygnału ambisonicznego [4]. W tym celu wykorzystano sygnał ambisoniczny w *B-formacie*, składający się z komponentów X, Y i Z o wartościach proporcjonalnych prędkości akustycznej w odpowiednich osiach i składowej W o wartości proporcjonalnej do ciśnienia, którą dodano do każdego z trzech sygnałów. Efekt można zaobserwować na rys. 6. W ten sposób wykres zawiera informacji o amplitudzie sygnału dochodzącego z każdego kierunku, lecz nie uwzględnia jego zwrotu. Pomimo tego, narzędzie może być pomocne w generalnej ocenie obrazu przestrzennego czy dynamiki sygnału.



(a) (b) (c)

Rysunek 5 Trzy klatki (a, b, c) pochodzące z wizualizacji akordu durowego w stroju równomiernie temperowanym (stosunek częstotliwości 1:1,259921:1,498307)



(a) (b) (c)

Rysunek 6 Trzy klatki (a, b, c) pochodzące z wizualizacji sygnału ambisonicznego

5. Podsumowanie

Przytoczone sposoby użycia trójosiowego oscyloskopu nie wyczerpują tematu, a ich dokładna realizacja może z pewnością zostać udoskonalona. Mimo to, ich różnorodność zdaje się wskazywać na wszechstronność zaproponowanego rozwiązania.

Bibliografia

[1] Kuojun Yang, Shulin Tian, Hao Zeng, Lei Qiu i Lianping Guo: „A seamless acquisition digital storage oscilloscope with three-dimensional waveform display", w: *Review of Scientific Instruments* 85.4, 2014, s. 045102

[2] *Jitter Tutorial*, *https://docs.cycling74.com/max7/tutorials/jitindex*, Dostęp: 2021-06-13

[3] Donald Hearn, M Pauline Baker i M Pauline Baker: *Computer graphics with OpenGL*, t. 3, Pearson prentice hall Upper Saddle River, NJ: 2004

[4] Michael A Gerzon: „Periphony: With-height sound reproduction", w: *Journal of the audio engineering society* 21.1, 1973, s. 2–10