



XVIII Międzynarodowe Sympozjum Nowości w Technice Audio i Wideo NTAV2020

Wrocław, 15 października 2020



Materiały konferencyjne

ORGANIZATORZY ▪ ORGANIZERS

Polska Sekcja Audio Engineering Society
Polskie Towarzystwo Akustyczne oddział we Wrocławiu
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów
Politechnika Wrocławska

KOMITET NAUKOWY ▪ SCIENTIFIC COMMITTEE

Przewodniczący: Krzysztof Opieliński
Honorowy przewodniczący: Andrzej Dobrucki

Andrzej Czyżewski
Adam Dąbrowski
Tadeusz Kamisiński
Piotr Kleczkowski
Bożena Kostek
Miroslaw Meissner
Janusz Piechowicz
Anna Preis
Aleksander Sęk
Ewa Skrodzka
Paweł Strumiłło
Jerzy Wiciak
Jan Żera

KOMITET ORGANIZACYJNY ▪ ORGANIZING COMMITTEE

Przewodniczący:	Przemysław Plaskota
Sekretarz:	Michał Łuczyński
Skarbnik:	Paweł Dziechciński
Członkowie:	Bartłomiej Kruk
	Maciej Walczyński
	Maurycy J. Kin
	Piotr Kozłowski
	Jędrzej Borowski

Redakcja:	Przemysław Plaskota
Edycja:	Michał Łuczyński

Wydawnictwo InfoArt
ISBN 978-83-921663-9-9

WSPIERAJĄ NAS



SPIS TREŚCI

Jan SKORUPA, Maciej GŁOWIAK

- *Produkcja dźwięku immersyjnego. Praktyczne metody i zastosowania dźwięku ambisonicznego wyższego rzędu do tworzenia produkcji audiowizualnych VR/360°* 5

Marcin WALCZAK, Ewa ŁUKASIK

- *Generowanie, edycja i transmisja wieloźródłowych strumieni audio z wykorzystaniem WEB AUDIO API, WEBRTC i WEB MIDI API* 9

Bartłomiej MRÓZ, Bożena KOSTEK

- *Techniki wielokanałowe w zastosowaniu do koncertów i nagrań muzycznych na odległość*..... 14

Michał ŁUCZYŃSKI

- *Stanowisko laboratoryjne kolumny z niezależnie sterowanymi głośnikami*..... 23

Cezary PIETRZAK, Piotr ODYA

- *Ambisoniczna mapa wybranych miejsc w Trójmieście* 30

Tomasz NOWAK, Bartłomiej KRUK

- *Praktyczne zastosowanie przetwornika o stałej szerokości wiązki (CBT) do projektowania monitorów studyjnych*..... 40

Maciej WALCZYŃSKI, Piotr RUSZKOWSKI

- *Lokalizacja źródeł dźwięku z wykorzystaniem sztucznych sieci neuronowych* 46

Paweł DZIECHCIŃSKI

- *Wykorzystanie wskaźnika odbicia dźwięku do oceny właściwości pochłaniających elementów budowlanych*..... 50

Maciej BLASZKE, Damian KOSZEWSKI, Bożena KOSTEK

- *Porównanie skuteczności klasyfikacji gatunków muzycznych w przy pomocy sieci neuronowej w zależności od typu danych wejściowych* 58

Kamila PIETRUSIŃSKA, Adam KUROWSKI, Bożena KOSTEK

- *Automatyczne generowanie kolejności list utworów muzycznych* 60

Adam ROSIŃSKI

- *Wpływ edukacji muzycznej na grupowanie dźwięków sekwencji ABA-ABA w rytmie galopowym*..... 70

Karol CZESAK, Piotr KLECZKOWSKI

- *Wpływ parametrów kompresji dynamicznej na subiektywną głośność zapisu instrumentu muzycznego* 74

Przemysław PLASKOTA, Małgorzata GAWLIŃSKA

- *Pomiar głośności nagrań dźwiękowych przechowywanych na różnych nośnikach*... 81

Maurycy KIN, Stefan BRACHMAŃSKI

- *Ocena jakości dźwięku sygnałów mowy i muzyki nadawanych za pomocą radiofonii cyfrowej DAB+* 84

Przemysław FALKOWSKI-GILSKI, Stefan BRACHMAŃSKI

- *Subiektywna ocena jakości sygnałów mowy i muzyki emitowanych w lokalnych multiplexach radiofonii DAB+ w Gdańsku i Wrocławiu* 91

Marcin ŁUCZAK, Karolina JARUSZEWSKA, Andrzej DOBRUCKI

- *Acoustics Course for Industry (ACI)*..... 96

PRODUKCJA DŹWIĘKU IMMERSYJNEGO. PRAKTYCZNE METODY I ZASTOSOWANIE DŹWIĘKU AMBISONICZNEGO WYŻSZEGO RZĘDU DO TWORZENIA PRODUKCJI AUDIOWIZUALNYCH VR/360°.

Jan SKORUPA¹

Maciej GŁOWIAK²

¹Instytut Chemii Bioorganicznej PAN – Poznańskie Centrum Superkomputerowo-Sieciowe,
ul. Noskowskiego 12/14, 61-704 Poznań
jskorupa@man.poznan.pl

² Instytut Chemii Bioorganicznej PAN – Poznańskie Centrum Superkomputerowo-Sieciowe,
ul. Noskowskiego 12/14, 61-704 Poznań
mac@man.poznan.pl

Systemy wirtualnej i poszerzonej rzeczywistości stają się coraz bardziej popularne. Rosnące wymagania konsumentów odnośnie treści i jakości treści mają ogromny wpływ, zarówno na rozwój w obszarze immersyjnego wideo i przemysłu gier, jak również w dziedzinie systemów dźwięku przestrzennego. Tradycyjne systemy stereo czy *surround* (np. 5.1, 7.1 itd.) nie są jednak wystarczające do zaspokojenia potrzeb współczesnego odbiorcy. Aby sprostać wymaganiom, konieczne jest rozszerzenie tradycyjnie pojętej produkcji w obszarze dźwięku przestrzennego. W niniejszym opracowaniu podjęto dyskusję na temat produkcji dźwięku przestrzennego w oparciu o Ambisonię wyższego rzędu (ang. High Order Ambisonics), w tym nagrywanie i postprodukcję dźwięku towarzyszącego materiałom wideo o wysokiej rozdzielczości VR/360°. W niniejszym artykule omówiono szczegółowo budowę i konfigurację instalacji Ambisonicznej w oparciu o doświadczenia z 24-kanalowym systemem odsłuchowym zbudowanym w PCSS, a także kwestie techniczne miksowania i nagrań Ambisonicznych wyższego rzędu. Cały proces produkcji został opisany na podstawie trzech różnych przykładów muzycznych, z których każdy został zrealizowany w różnych warunkach akustycznych. Wszystkie przykłady zostały przygotowane przez Dział Nowych Mediów PCSS w ramach projektu badawczego Immersify finansowanego z programu Horizon 2020 w latach 2017-2020.

1. WSTĘP

Historia dźwięku przestrzennego jest dziedziną od samego początku ściśle związaną z przemysłem rozrywkowym. Już w latach 40. ubiegłego stulecia mogliśmy zaobserwować pierwsze próby rozszerzenia treści wideo za pomocą dźwięku wielokanałowego. Mowa tu przede wszystkim o pierwszym systemie projekcji dźwięku wielokanałowego o nazwie „Fantasound”, zastosowanego przez wytwórnię Disney. Schemat ten, w którym dźwięk jako medium uzupełniał doznania wizyjne, był w kolejnych latach rozbudowywany o nowe możliwości techniczne. Zrozumiałe jest więc, iż kierunek rozwoju tej dziedziny inżynierii dźwięku jest warunkowany poprzez nowe rozwiązania wizyjne. Obecnie rynek rozrywki, napędzany wymaganiami użytkowników szuka nowych rozwiązań i obserwuje się odchodzenie od tradycyjnego obrazu dwuwymiarowego na rzecz systemów immersyjnych. Wkraczamy w erę wirtualnej oraz rozszerzonej rzeczywistości, która zapewnia widzom nowe doznania audiowizualne, co wymaga jednak szczególnego i innowacyjnego podejścia do kwestii dźwięku otaczającego.

Powszechnie stosowane systemy *surround* takie jak 5.1, 7.1 czy 7.4.2 nieznacznie różnią się tylko od tych opracowywanych jeszcze w drugiej połowie XX w. Oparte są one przede wszystkim o głośniki

umieszczone wokół słuchacza, z których każdy posiada specjalny identyfikator oraz stałe miejsce w przestrzeni (np. Center, Front-Left, Front-Right itd). Systemy te mają swoje poważne ograniczenia, szczególnie w połączeniu z technologią VR. W takich systemach *surround*, każdy kanał audio jest bowiem na stałe przypisany tylko do jednego głośnika. W celu uzyskania właściwego efektu przestrzennego ścieżki dźwiękowe muszą być przygotowane zgodnie z predefiniowanym rozmieszczeniem głośników. Nie ma możliwości odtworzenia plików audio 7.4.2 w konfiguracji 5.1 bez utraty istotnych informacji i pominięcia czy ponownego zmiksowania niektórych kanałów, co utrudnia adaptację raz przygotowanego dźwięku do różnych rodzajów instalacji dźwiękowych. Innym aspektem mówiącym o niedoskonałości takiego podejścia jest stosunkowo niska rozdzielczość przestrzenna oraz mały obszar tzw. *sweet-spot* [1]. Istnieje jednak inny sposób myślenia o dźwięku przestrzennym, opracowany w latach 70. przez brytyjskiego akustyka i matematyka Michaela Gerzona nazwany przez niego Ambisonią. [2].

2. PODSTAWY TEORII

Ambisonia jest przełomową koncepcją reprezentacji pola dźwiękowego, w której w przeciwieństwie do typowych systemów wielokanałowych, poszczególne kanały audio nie przenoszą informacji o sygnale odpowiadającym jednemu tylko głośnikowi, a w zamian reprezentują one właściwości pola akustycznego. Docelowym formatem pozwalającym na zapis tych informacji jest B-format [3]. Charakteryzuje się on różną rozdzielczością przestrzenną regulowaną przez tzw. rząd Ambisoniczny (*Ambisonic order*). Dzięki takiemu podejściu nagranie zrealizowane w B-formacie może być dekodowane dynamicznie do dowolnego układu głośników z zachowaniem informacji o kierunku dźwięku. Rozwiązanie to pozwala inżynierowi dźwięku skupić się wyłącznie na rzeczywistym kierunku dźwięku nieograniczonego poprzez lokacje głośników w docelowym systemie odsłuchowym.

3. NAGRANIA AMBISONICZNE

Pomimo, że minęło prawie 50 lat od powstania teorii Ambisonii, nagrania wykorzystujące jej podstawy zostały powszechniej zastosowane wraz z adaptacją dekodatorów binauralnych dopiero w świecie obrazu immersyjnego (wirtualna rzeczywistość). Aktualnie możliwe jest zastosowanie wielu sprzętowych jak i programowych rozwiązań w celu uzyskania nagrań Ambisonicznych, począwszy od tych klasycznych, opracowanych jeszcze przez Michaela Gerzona. Było nimi m.in. zastosowanie mikrofonu typu *soundfield* rejestrującego zmiany pola akustycznego za pomocą czterech kardioidalnych kapsuł, w których wyjściowym formatem nagrania był A-format [2] wymagający dalszego dekodowania. Kolejną metodą stosowaną przez Gerzona było wykorzystanie rozszerzonego do nagrań przestrzennych koincydentalnego ustawienia mikrofonów bipolarnych typu *blumlein*.

Obecnie możliwe jest zastosowanie nowszej generacji mikrofonów Ambisonicznych również opartych o pionierskie próby Gerzona, rozwijających metody pozyskiwania dźwięku Ambisonicznego w dziedzinie rozdzielczości przestrzennej. Możliwe jest również zastosowanie software'owych enkoderów i przeniesienie ścieżek zarówno monofonicznych jak i stereofonicznych do dziedziny B-formatu oraz stworzenie wirtualnej przestrzeni akustycznej. W referacie, na podstawie praktycznych doświadczeń i obserwacji związanych z różnymi technikami nagrań Ambisonicznych omówimy różne sposoby pozyskiwania i przetwarzania dźwięku. W odniesieniu do przytoczonych przykładów nagrań zostanie rozpatrzona możliwość kombinacji poszczególnych metod w celu osiągnięcia jak najlepszych efektów, zarówno pod kątem jakości nagrań, jak i najdoskonalszego odtworzenia pola akustycznego. Dyskusji zostaną poddane finalne efekty dotyczące reprodukcji pola akustycznego pod względem rzeczywistych sytuacji nagraniowych oraz realizacji wysokiej rozdzielczości w dziedzinie kierunkowości. Przykładami

nagrań będą tutaj: rejestracja septetu jazzowego „Anomalia” w studio PCSS, czy kwintetu smyczkowego na dachu budynku wraz z pozyskaniem obrazu VR/360° w wysokiej rozdzielczości 20K. Dodatkowo zaprezentowana zostanie produkcja udźwiękowania Ambisonicznego do filmu opartego o wizualizację skanów laserowych [4].

4. POSTPRODUKCJA DŹWIĘKU AMBISONICZNEGO

W zakresie miksowania dźwięku Ambisonicznego, poza tym, że choć w ogólności wykorzystuje się podobne narzędzia jak w przypadku postprodukcji nagrań stereo, wymagane jest specyficzne podejście oraz adaptacja podstawowych założeń teorii Ambisonicznej odnośnie wielkości rzędu, normalizacji poszczególnych składowych, różnych metod kodowania i dekodowania B-formatu, jak również ilości głośników systemu, w którym nagranie zostanie odtworzone. Każdy z tych wymienionych aspektów ma znaczący wpływ na finalny efekt w dziedzinie rozdzielczości przestrzennej nagrań. W referacie zostanie rozwinięty wpływ poszczególnych założeń na kwestie brzemienia oraz przestrzenności miksowanego materiału, jak również dopasowanie poszczególnych rozwiązań do konkretnych scenariuszy nagranych. Omówiona zostanie również postprodukcja materiału dźwiękowego przy uwzględnieniu docelowych warunków odsłuchowych, a w szczególności nakreślone zostaną różnice pomiędzy wielogłośnikowym a binauralnym odsłuchem dźwięku.

5. DEKODOWANIE ORAZ ODSŁUCH

Ostatnim, choć równie istotnym etapem mającym decydujący wpływ na finalną jakość i przestrzenność nagrania jest sposób dekodowania B-formatu do docelowych warunków odsłuchowych. Zgodnie z teorią Ambisonii, jednym z podstawowych założeń jest to, iż każde nagranie Ambisoniczne może zostać odtworzone na dowolnej instalacji wielogłośnikowej, jak również w domenie binauralnej za pomocą tradycyjnych słuchawek stereo. Pomijając kluczowe aspekty jakościowe takie jak: rodzaj oraz liczba głośników w docelowej instalacji odsłuchowej czy warunki akustyczne pomieszczenia odsłuchowego, w referacie rozpatrzone zostaną różne sposoby dekodowania B-formatu oraz ich wpływ na subiektywne wrażenia słuchowe, a także przestrzenność oraz kierunkowość.

Zaprezentowana zostanie również sferyczna instalacja Ambisoniczna zbudowana w PCSS, która składa się z 24 głośników typu Genelec 8010A umieszczonych w trzech okręgach po 8 głośników każdy. Pozwala ona na odtwarzanie dźwięku przestrzennego w pełnym azymucie oraz elewacji. Przedstawione zostaną obiektywne wyniki badań oceniające lokalizacje wirtualnego źródła w przestrzeni horyzontalnej docelowej instalacji oraz wpływ wyników tych badań na proces produkcji materiałów dźwiękowych.

PODZIĘKOWANIA

Prace nad dźwiękiem Ambisonicznym zostały przeprowadzone przez Dział Nowych Mediów PCSS w ramach projektu badawczego Immersify finansowanego przez Unię Europejską z programu Horyzont 2020 w latach 2017-2020 (numer umowy grantowej 762079) [5].

LITERATURA

- [1] FRANK M., *How to make ambisonics sound good*, 2014, https://www.researchgate.net/publication/315797065_How_to_make_Ambisonics_sound_good, dostęp [28.09.20]
- [2] GERZON M., *What is wrong with quadraphonics*, 1974, <https://www.michaelgerzonphotos.org.uk/ambisonics.html> dostęp [28.09.20]

- [3] CORCUERA A., Ambisonics W: A real time encoding tool for Higher Order Ambisonics, A. Corcuera, Barcelona, University Pompeu Fabra, 2014, 16-20
- [4] SKORUPA J., GŁOWIAK M., *Content production guidelines: Ambisonic recordings and postproduction*, Immersify 2020
- [5] Projekt Immersify, www.immersify.eu

IMMERSIVE AUDIO PRODUCTION. PRACTICAL METHODS AND APPLICATIONS OF THE HIGH ORDER AMBISONICS IN MUSICAL PRODUCTION FOR VR/360° VIDEO CONTENT.

Virtual and augmented reality systems are becoming increasingly popular. Growing consumer demands for content itself and the content quality have a huge impact, both on the development of immersive video and gaming industry, as well as in the field of surround sound systems. However, traditional stereo or surround systems (e.g. 5.1, 7.1, etc.) are not sufficient to meet the needs of the modern consumer. In order to meet the requirements, it is necessary to extend the traditionally understood production in the area of surround sound. This paper discusses spatial sound production based on High Order Ambisonics, including recording and post-production of sound accompanying the high resolution VR/360° video. This article discusses in detail the designing and configuration of an Ambisonic installation based on the 24-channel loudspeaker system constructed in PSNC, as well as the technical issues of mixing and recording of high-order Ambisonics. The whole production process was described on the basis of three different music examples, each of which was realized in different acoustic conditions. All the examples were prepared by the PSNC New Media Department as part of the Immersify research project financed from the Horizon 2020 programme in 2017-2020.

GENEROWANIE, EDYCJA I TRANSMISJA WIELOŹRÓDŁOWYCH STRUMIENI AUDIO Z WYKORZYSTANIEM WEB AUDIO API, WEBRTC I WEB MIDI API

Marcin WALCZAK¹

Ewa ŁUKASIK²

Instytut Informatyki, Politechnika Poznańska, ul. Piotrowo 2, 60-965 Poznań

¹kontakt@marcin-walczak.pl

²ewa.lukasik@cs.put.poznan.pl

Streszczenie. W czasach gdy konieczne jest stosowanie dystansu społecznego istnieje szczególnie zapotrzebowanie na niezawodne, gwarantujące bardzo małe opóźnienia systemy komunikacji audio/wideo online do przeprowadzania zdalnych prób, nagrywania i nadawania muzyki online. W artykule dokonano przeglądu systemów wykorzystywanych przez muzyków do zdalnej współpracy. Omówiono cechy technologii Web Audio API umożliwiającej syntezę, reprodukcję i przetwarzanie dźwięku w przeglądarce internetowej. W połączeniu z technologią WebRTC oraz Web MIDI API istnieje możliwość tworzenia zaawansowanych aplikacji webowych audio. W artykule opisano prototypowe wdrożenie i testy przykładowego systemu wykorzystującego Web Audio API, Web RTC i Web MIDI do tworzenia i przesyłania strumieni audio z wielu źródeł.

1. WPROWADZENIE

Komunikacja i wymiana treści multimedialnych za pośrednictwem Internetu stała się powszechna i ogólnie akceptowana, choć ostatnie miesiące koniecznej izolacji związanej z rozprzestrzenianiem się choroby COVID-19 sprawiło, że wykorzystanie środków komunikacji internetowej stało się wręcz nieodzowne we wszystkich dziedzinach życia. Bezpłatne komunikatory internetowe, np. Skype, Facebook Messenger, WhatsApp, Signal, Google Hangout stosowane są do komunikacji prywatnej, a dla celów zawodowych do listy tej dopisują się Zoom, Webex czy otwarte systemy oparte na BigBlueButton [2]. Istnieją również darmowe komunikatory głosowe dedykowane komunikującym się ze sobą uczestnikom internetowych gier komputerowych.

Internet nie został pierwotnie opracowany w celu przesyłania ruchu w czasie rzeczywistym. Utrata pakietów i opóźnienia wpływają negatywnie na jakość przesyłanego sygnału audio. Usługi wideokonferencji rozwiązują problemy związane z utratą pakietów, stosując długie ramki audio, duże bufora sieciowe i zasadę retransmisji pakietów. W konsekwencji powoduje to dodatkowe opóźnienia rzędu kilkuset milisekund. Takie opóźnienia nie stanowią problemu w kontekście komunikacji głosowej, jednak dla wymiany sygnałów muzycznych w czasie rzeczywistym jest to znaczne ograniczenie. Tymczasem muzycy jeszcze w czasach przed pandemią byli zainteresowani wspólnym muzykowaniem online i wspólnymi próbami. Dlatego od początku XXI-go wieku trwały prace nad realizacją sieci audio zoptymalizowanej pod kątem opóźnienia. Celem było umożliwienie zespołowego wykonywania muzyki przez artystów rozproszonych geograficznie za pośrednictwem sieci internetowej, tzw. Networked Music Performance (NMP)[9] z zapewnieniem wysokiej jakości interakcji. Problemy, jakie należało rozwiązać w celu realizacji NMP zostały przedstawione w rozprawie doktorskiej Carot'a [4]. W efekcie wieloletnich prac powstało kilka systemów wykorzystywanych przez muzyków do prowadzenia zdalnych prób i przygotowania nagrań za pomocą sieci. Są to zarówno systemy otwarte, jak np. Jamulus [5], SoundJack [11], LoLa [6] jak i komercyjne, np. Artsmesh [1].

Mimo wspomnianych wyżej zaawansowanych rozwiązań, przetwarzanie audio nie było przez długi czas wspierane w przeglądarkach internetowych. Wprowadzenie elementu `<audio>` w HTML5 umożliwiło strumieniowe odtwarzanie dźwięku, ale nie pozwalało na budowę bardziej skomplikowanych aplikacji audio, ani na obsługę audio w zaawansowanych grach internetowych i aplikacjach interaktywnych. Nie zapewniało miksowania, przetwarzania i filtrowania na równi z nowoczesnymi rozwiązaniami dostępnymi w komputerach stacjonarnych. Takie możliwości zapewnia Web Audio API [12], którego pierwsza wersja pojawiła się w 2011 roku i nad którym prace trwają do dnia dzisiejszego. Podstawowym paradygmatem dla tej technologii jest graf routingu audio, w którym wiele obiektów *AudioNode* jest połączonych ze sobą, aby zapewnić renderowanie dźwięku. Web Audio API jest wspomagany przez szereg bibliotek JavaScript ułatwiających realizację bardziej skomplikowanych zadań, np. Tone.js.

2. OPIS TECHNOLOGII

Podstawowe funkcje, jakie spełnia WebAudio API są następujące [12]: modularny routing dla prostych lub złożonych architektur i efektów dźwiękowych, miksowanie i efekty dźwiękowe, 32-bitowe przetwarzanie zmiennoprzecinkowe zapewniające wysoki zakres dynamiki dźwięku, odtwarzanie dźwięku z niskim opóźnieniem do zastosowań muzycznych wymagających bardzo wysokiego stopnia precyzji rytmicznej, takich jak automaty perkusyjne i sekwencery, elastyczna obsługa kanałów w strumieniu audio, umożliwiająca dzielenie i łączenie strumieni, przetwarzanie źródeł dźwięku z obiektu multimedialnego `<audio>` lub `<video>`, przetwarzanie strumienia audio generowanego na żywo, przetwarzanie dźwięku odebranego od zdalnego użytkownika i wysyłanie przetworzonego strumienia, a także obsługę dźwięku przestrzennego, efektywna analiza muzyki w dziedzinie czasu i częstotliwości w czasie rzeczywistym, wydajna filtracja dolnoprzepustowa, górnoprzepustowa i za pomocą innych popularnych filtrów, kształtowanie fali i inne efekty nieliniowe.

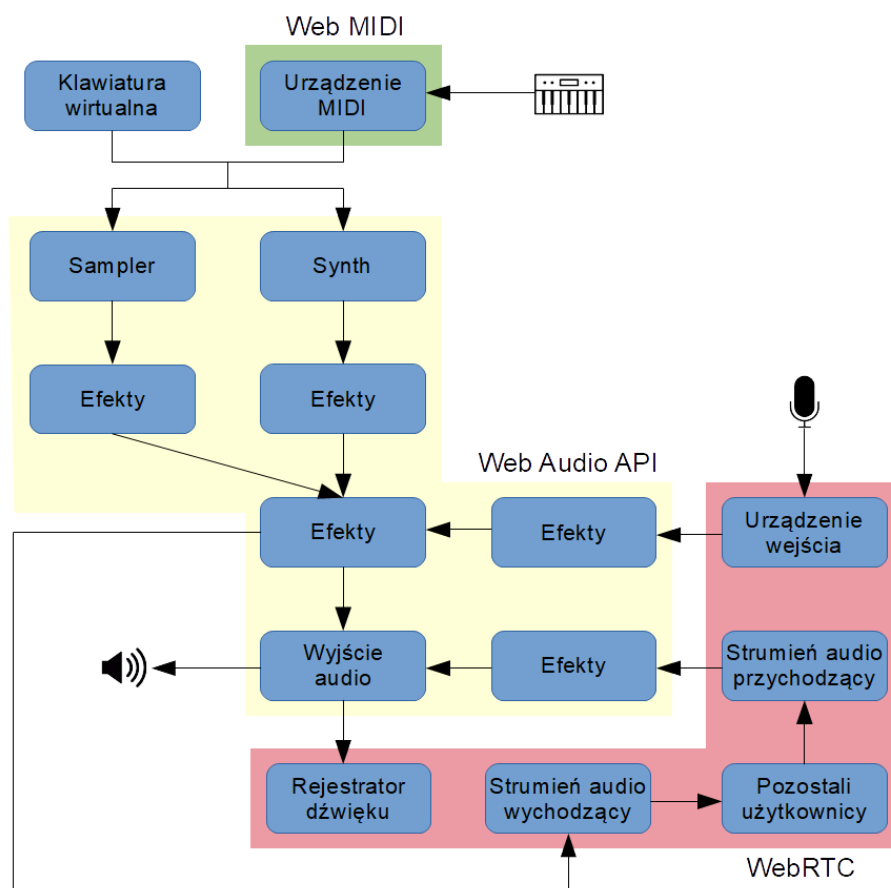
Web Audio API umożliwia również integrację z WebRTC [7]. WebRTC (ang. *Web Real-Time Communication*) jest technologią umożliwiającą aplikacjom internetowym przechwytywanie i przesyłanie strumienia audio i wideo oraz innych danych bezpośrednio między urządzeniami metodą *peer-to-peer*.

WebRTC obejmuje trzy zasadnicze API [10]:

- `MediaStream` API zapewniający dostęp do strumieni danych audiowizualnych z wielu źródeł, jak np. mikrofony, i kamery,
- `RTCPeerConnection` do tworzenia i zarządzania połączeniami *peer-to-peer*,
- `RTCDataChannel` obsługujący komunikację *peer-to-peer* danych.

Architektura WebRTC jest oparta na trapezoidzie SIP (ang. *Session Initiation Protocol*) [7]. W modelu trapezoidalnym WebRTC dwie komunikujące się przeglądarki uruchamiają aplikacje sieciowe, pobrane z osobnych serwerów. Za pośrednictwem serwerów sygnalizacyjnych, wstępnie wymieniają się danymi w celu wynegocjowania połączenia *peer-to-peer*. Dane sygnalizacyjne mogą być przesyłane dowolną metodą, jak np. protokołem HTTP lub WebSocket. Po nawiązaniu połączenia, dalsza komunikacja pomiędzy przeglądarkami następuje według modelu *peer-to-peer*, bez udziału serwerów.

Protokół MIDI (ang. *Musical Interface for Digital Instruments*) jest powszechnie stosowanym w przemyśle muzycznym protokołem umożliwiającym komunikację pomiędzy kompatybilnymi urządzeniami muzycznymi. Celem Web MIDI API [13] jest komunikacja pomiędzy przeglądarką, a urządzeniami MIDI, zarówno wejścia jak i wyjścia.



Rys.1. Przykładowy schemat przepływu strumieni dźwiękowych wewnątrz aplikacji wraz z oznaczeniem elementów wykorzystujących technologie Web Audio API, WebRTC i Web MIDI. Opracowanie własne

3. REALIZACJA SYSTEMU

3.1 SZCZEGÓŁY IMPLEMENTACYJNE

Celem opisanego w artykule projektu była budowa prototypowego systemu umożliwiającego generowanie i modyfikację przez użytkownika ścieżek dźwiękowych z wykorzystaniem przeglądarki internetowej zarówno strumieni wejściowych, jak i wyjściowych. Strumienie dźwiękowe podlegają konfiguracji oraz modyfikacji w czasie rzeczywistym poprzez filtry i inne elementy zmieniające strumień dźwiękowy. Możliwa jest również rejestracja sygnałów dźwiękowych pochodzących z urządzeń wejścia (jak np. mikrofon) i generowanych przez aplikację strumieni dźwiękowych. Zarejestrowane w ten sposób ścieżki dźwiękowe można odtworzyć i wykorzystać w aplikacji. Dodatkowo dostępna jest funkcja eksportu zarejestrowanej ścieżki dźwiękowej w formie wybranego przez użytkownika formatu audio. Sterowanie warstwą dźwiękową odbywa się za pomocą wirtualnej klawiatury dostępnej w oknie przeglądarki, fizycznej klawiatury urządzenia oraz urządzenia MIDI. Poza warstwą dźwiękową aplikacja umożliwia komunikację sieciową pomiędzy użytkownikami. Użytkownik ma możliwość tworzenia i zarządzania wirtualnymi pokojami pozwalającymi innym użytkownikom na dołączenie do nich. Osoby korzystające z tego samego pokoju mogą komunikować się między sobą poprzez połączenia *peer-to-peer* [8]. Schemat przepływu strumieni dźwiękowych wewnątrz aplikacji wraz z oznaczeniem elemen-

tów wykorzystujących technologie Web Audio API [12], WebRTC [10] i Web MIDI API [13] przedstawiono na rysunku 1.

Podstawowym zadaniem serwera jest pośrednictwo przy komunikacji między klientami aplikacji. Nawiązanie połączenia *peer-to-peer* za pomocą technologii WebRTC wymaga kanału komunikacyjnego do negocjacji właściwego połączenia. Rolę tę przejmuje przygotowany w tym celu serwer.

Zaimplementowano i skonfigurowano serwer STUN w celu pozyskania przez aplikacje klienckie swoich globalnych adresów IP. Umożliwiło to poprawne ustanawianie bezpośrednich połączeń między użytkownikami w przypadku, gdy co najmniej jeden z nich znajduje się wewnątrz sieci NAT.

Zgodnie z przedstawionymi wymaganiami aplikacja powinna pozwalać na uzyskanie połączenia *peer-to-peer* pomiędzy użytkownikami przy różnych konfiguracjach sieciowych. Ustanawianie połączeń pomiędzy użytkownikami może zostać utrudnione m.in. z uwagi na zastosowanie przez sieć lokalną klienta NAT lub ograniczenia w ustanawianiu połączeń nałożone przez lokalną zaporę sieciową.

W przypadku, gdy nie jest możliwe nawiązanie połączenia *peer-to-peer* z uwagi na zabezpieczenia sieci, zastosowano serwer TURN. Został on wdrożony jako zastępcza forma komunikacji w sytuacji blokowania przez sieć połączenia *peer-to-peer* z innym klientem aplikacji. Działa on wtedy jako pośrednik, przesyłając dane pomiędzy użytkownikami. Z uwagi na przechodzenie strumieni danych przez serwer TURN zamiast bezpośrednio poprzez połączenie *peer-to-peer*, opóźnienia przy wykorzystaniu tego rodzaju połączenia mogą ulec zwiększeniu.

3.1 TESTY

Przeprowadzono testy aplikacji metodą białej skrzynki. Skupiono się w nich na sprawdzeniu działania poszczególnych elementów aplikacji na wybranych przeglądarkach wspierających Web Audio API i WebRTC.

Testy przeprowadzono na podanych poniżej przeglądarkach:

- Google Chrome w wersji 79.0.3945.88,
- Mozilla Firefox w wersji 71.0,
- Opera w wersji 65.0.3467.69,
- Safari w wersji 13.0,
- Microsoft Edge w wersji 44.18362.449.0.

Przetestowano różne aspekty komunikacji między użytkownikami, generowanie dźwięku z syntezyzatora, obsługę urządzeń wejścia: mikrofonów i urządzeń MIDI, przetwarzanie dźwięku: dodawanie i edycję efektów w strumieniach dźwięku oraz rejestrację i eksport ścieżki dźwiękowej. Wszystkie rezultaty były zgodne z założeniami.

Natomiast przeprowadzone testy aplikacji w sieci wykazały, że w przypadku części sieci nie było możliwe nawiązanie połączenia *peer-to-peer* wyłącznie za pomocą serwera sygnalizacyjnego. Przypadki niemożności ustanowienia połączenia między użytkownikami były spowodowane brakiem dostępu aplikacji klienckich do ich globalnych adresów IP lub blokowanie przez zaporę sieciową połączeń *peer-to-peer*. Powodowało to oczywiście generowanie dodatkowych opóźnień.

PODSUMOWANIE

W artykule została opisana aplikacja internetowa demonstrująca możliwości wykorzystania technologii Web Audio API, WebRTC API oraz Web MIDI. Aplikacja pozwala na ustanawianie połączeń *peer-to-peer* pomiędzy wieloma użytkownikami za pomocą technologii WebRTC. Wdrożono możliwość generowania, odtwarzania, modyfikacji, przechwytywania oraz przesyłania ścieżek dźwiękowych drogą sieciową. Mimo że aplikacja została opracowana jako demonstracja możliwości opisanych technologii, posiada potencjał rozwojowy. W oparciu o już zaimplementowane technologie możliwe jest

wdrożenie nowych funkcji z zakresu generowania i przetwarzania dźwięku, komunikacji między użytkownikami oraz rozwinięcie już istniejących. Aplikacja może zostać rozbudowana o m.in. sekwencjonowanie ścieżek dźwiękowych, graficzną reprezentację cech strumieni dźwięku w czasie rzeczywistym, czy transmisję strumieni dźwięku między wieloma użytkownikami za pomocą serwera mediów. Wtedy należy przeprowadzić badania jakości transmisji, w tym opóźnień systemowych.

Doświadczenie zdobyte przez jednego z autorów przy realizacji przedstawionego projektu stało się podstawą do podjęcia prac w ramach komercyjnego międzynarodowego projektu Branch [3] udostępniającego wirtualną przestrzeń biurową. System pozwala na komunikację audiowizualną między użytkownikami z wykorzystaniem technologii WebRTC [10]. Jedną z cech wyróżniających projekt jest sposób, w jaki użytkownicy komunikują się między sobą. Aplikacja pozwala na dynamiczne nawiązywanie dwustronnych połączeń w topologii siatki między użytkownikami znajdującymi się w swoim "zasięgu wzroku". W przypadku utraty użytkownika z pola widzenia, połączenie między nimi zostaje zamknięte. Takie rozwiązanie umożliwia emulację rzeczywistej interakcji między członkami zespołu pozwalając na dynamiczne tworzenie grupowych spotkań w sposób odpowiadający temu w rzeczywistym życiu. Dodatkową zaletą jest ograniczenie liczby aktywnych połączeń WebRTC [10], co pozwala redukować generowane przez nie obciążenie łącza.

Omówione technologie nie są jeszcze w pełni wspierane przez wszystkie nowoczesne przeglądarki internetowe. Ich standardy i rozwiązania wciąż podlegają rozwojowi i mogą ulec zmianie w przyszłości.

LITERATURA

- [1] ARTSMESH <https://www.artsmesh.com/>,
- [2] BIGBLUEBUTTON <https://bigbluebutton.org>
- [3] BRANCH <https://branch.gg/>
- [4] CARÔT A., Musical Telepresence - A Comprehensive Analysis Towards New Cognitive and Technical Approaches, PhD Thesis, Lübeck, 2009.
- [5] JAMULUS <http://llcon.sourceforge.net/>,
- [6] LOLA <https://lola.conts.it/>
- [7] LORETO S.; ROMANO S.P.: Real-Time Communication with WebRTC: Peer-to-Peer in the Browser. O'Reilly Media, Inc. Sebastopol, 2014
- [8] NAYYEF Z., AMER S., HUSSAIN Z., Peer to Peer Multimedia Real-Time Communication System based on WebRTC Technology, Int. J. for the History of Engineering & Technology, 2019, s. 125 – 130
- [9] ROTTONDI C, CHAFE Ch., ALLOCCHIO C., SARTI A.: An Overview on Networked Music Performance Technologies. IEEE Access 4: 8823-8843 2016,
- [10] WEB RTC, Real Time Communication for the Web, webrtc.org
- [11] SOUNDJACK <https://www.soundjack.eu/>,
- [12] Web Audio API, W3C Candidate Recommendation, 11 June 2020, <https://www.w3.org/TR/webaudio/>
- [13] Web MIDI API, W3C Working Draft, March 2015, <https://www.w3.org/TR/2015/WD-webmidi-20150317>

GENERATING, EDITING AND MULTI-SOURCE TRANSMISSION OF AUDIO STREAMS USING WEB AUDIO API, WEBRTC AND WEB MIDI API

In the time of social distancing, there is a special need for reliable low-latency audio/video online communication systems for online music rehearsing, recording and broadcasting. In the paper, first, a review of existing systems used by remote users will be provided. Then the potential of the Web Audio API will be discussed. The Web Audio API introduces extensive functionality related to the synthesis, reproduction and processing of sound. In combination with the WebRTC technology enabling real-time audio streaming and capturing, and with the Web MIDI API technology ensuring communication with MIDI devices, there is a possibility of creating advanced Web audio applications. We will describe prototype implementation and tests of an exemplary system using the above mentioned technologies to form and transmit audio streams from multiple sources.

TECHNIKI WIELOKANAŁOWE W ZASTOSOWANIU DO KONCERTÓW I NAGRAŃ MUZYCZNYCH NA ODLEGŁOŚĆ

Bartłomiej MRÓZ^{1,2}

Bożena KOSTEK²

¹Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Katedra Systemów
Multimedialnych, Narutowicza 11/12, 80-233 Gdańsk
bartlomiej.mroz@pg.edu.pl

²Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Laboratorium Aku-
styki Fonicznej, Narutowicza 11/12, 80-233 Gdańsk
bokostek@audioakustyka.org

W czasie pandemii koronawirusa COVID-19 nowego znaczenia nabrały możliwości transmisji dźwięku z obrazem, zwłaszcza do pracy zdalnej. Zwiększenie zapotrzebowania na zdalne, wieloźródłowe połączenia zwróciło uwagę na wielokanałowość i uprzestrzennienie dźwięku celem łatwiejszej lokalizacji źródeł dźwięku. Zdalna praca jest szczególnym wyzwaniem dla muzyków; zarówno w kontekście wspólnych ćwiczeń i prób, jak i koncertów. Tworzenie zdalnych nagrań muzycznych jest niepowtarzalną okazją do produkcji wielokanałowych, przestrzennych, wykorzystujących techniki ambisoniczne i binauralne. Techniki te pozwalają na stworzenie nowych, immersyjnych doznań dla słuchaczy. W pracy przedstawiono zrealizowane nagrania ambisoniczne Akademickiego Chóru Politechniki Gdańskiej, zawarto opis związany z warsztatem realizatora dźwięku i obrazu oraz przedyskutowano problemy związane z synchronizacją dźwięku.

1. WPROWADZENIE

W ciągu ostatnich lat rozwój techniki pozwala na coraz szersze zastosowanie urządzeń do pracy zdalnej w czasie rzeczywistym. Tym niemniej, zdalna praca twórcza, a w szczególności wykonawstwo muzyczne, nadal nie jest prostym do rozwiązania wyzwaniem. Jednokierunkowe połączenia, dzięki którym wydarzenia artystyczne są transmitowane w czasie rzeczywistym, stały się dostępne dzięki sieciom zdolnym obsługiwać dużą przepustowość, routing pakietów o małych opóźnieniach i gwarantowaną jakość usług (QoS – *Quality of Service*) [12,13,9]. Można powiedzieć, że jest to połączenie jednokierunkowe, nie pozwalające artystom na zdalne połączenie między sobą, a jedynie między miejscem wykonania a publicznością. Dużo trudniejsze jest połączenie, w którym wykonawcy w dwóch lub więcej odległych lokalizacjach próbują razem wykonać ustaloną kompozycję lub improwizację w czasie rzeczywistym, a co Weinberg [45] określa jako podejście "pomostowe" ("*bridge approach*"). W takich przypadkach wiadomo, że nieuniknione opóźnienie spowodowane fizycznym czasem tranzytu pakietów sieciowych ma wpływ na wydajność [4,5,11,24,35,36]. W związku z tym próbowano uwzględnić te opóźnienia poprzez projekt lub kompozycję [6,8,26]. Warto też nadmienić, iż jako pierwsze tego typu realizacje zastosowano w dedykowanych sieciach akademickich, a eksperymentalny koncert z artystami z różnych miast w Polsce miał miejsce już w 2001 roku [34]. Jednakże, w transmisjach strumieniowych warstwa przestrzenna źródeł dźwięku dopiero od niedawna nabiera znaczenia. Techniki binauralne i ambisoniczne jako narzędzie do umiejscowienia zdalnych źródeł dźwięku najszybciej zadomowiły się w zastosowaniach telekonferencyjnych [3,17,16,46,33,7]. Warto wspomnieć o platformie AltSpaceVR [2], w której można przeprowadzać spotkania telekonferencyjne, pokazy, prezentacje czy zajęcia, jak

również spotkania towarzyskie; dźwięk wszystkich obiektów w platformie jest przestrzenny i zbinauralizowany.

Jednakże, w przypadku tzw. sieciowych/zdalnych wykonań muzycznych (NMP – *Networked Music Performance*) [41], już wcześniej zostały podjęte udane próby transmitowania dźwięku wielokanałowego [47,49,40]. Zostały także przeprowadzone transmisje z wykorzystaniem ambisonii [25]. Ambisonię można traktować jako format produkcyjny, który jest wykorzystywany tworzenia dźwięku 3D. Opiera się na reprezentacji pola dźwiękowego poprzez rozłożenie go na podstawowe funkcje ortonormalne, zwane harmonicznymi sferycznymi. Taka reprezentacja pozwala na elastyczny proces produkcji, który jest niezależny od docelowego systemu odtwarzania (zestaw głośników czy słuchawki). Koncert „PURE Ambisonics Concert & the Night of Ambisonics” zorganizowany w ramach międzynarodowej konferencji (3rd International Conference on Spatial Audio) w Grazu we wrześniu 2015 roku był z pewnością jednym z prekursorów tego typu wydarzeń muzycznych [38,42,22]. Podczas wieczoru koncertowego wykorzystano format ambisoniczny do dystrybucji koncertu do różnych miejsc i transmisji w czasie rzeczywistym (w tym transmisji radiowej: ogólnokrajowej naziemnej i satelitarnych programów radiowych). Dźwięk z sali koncertowej był przygotowany do nagrania i transmisji z wykorzystaniem 23-kanałowego systemu głośników, miksu 5.1, a także miksu binuaralnego do odsłuchu na słuchawkach. Koncert z konferencji ICSA 2015 zachęcił autorów tej idei do przeprowadzenia koncertu 3D na żywo z Al Di Meola w lipcu 2016 roku, obejmującego przestrzenne efekty w czasie rzeczywistym i transmisję do innego wnętrza [22].

Wszystkie te prezentacje miały jednak charakter eksperymentalny. Aby stały się standardem nagrań czy transmisji, muszą zostać zaimplementowane w szeroko dostępnych platformach. Dla przykładu, w platformie YouTube możliwa jest jedynie transmisja na żywo z samym obrazem w technice 360° – dźwięk ambisoniczny 1. rzędu można zastosować tylko przy przesłaniu gotowego nagrania [48]. Kolejna platforma, Facebook, wspiera transmisję na żywo z obrazem 360° oraz ambisonią 1. rzędu [48], natomiast przesłane gotowe nagranie może zawierać dźwięk ambisoniczny 2. rzędu [19]. Powstała też zupełnie nowa platforma, zorientowana na prezentację treści z obrazem dookólnym oraz dźwiękiem zmiksowanym w ambisonii wyższych rzędów – HOAST [15]. Warto jednak podkreślić, że wszystkie te technologie nie dotyczą *stricte* połączenia muzyków znajdujących się w oddzielnych miejscach, a raczej wirtualnego odwzorowania sceny z muzykami dla publiczności w oddległych salach koncertowych. Przeszkodą dla pełnego NMP nadal są opóźnienia rzędu 25 ms w sieci internetowej, wynikające z jej architektury. Dopiero technologia 5G daje nadzieję na pokonanie tych trudności [10].

Z tego względu, zdalne wykonania muzyczne zaistniały w formie przesyłania nagrań poszczególnych partii utworu, w celu złożenia ich w jedną całość w post-produkcji. Najbardziej znane są dokonania kompozytora Erica Whitacre’a, który od 2009 roku organizuje tzw. wirtualny chór [18]; powstało 6 edycji tych sesji, a w każdej z nich uczestniczyły tysiące osób z całego świata. Zwłaszcza ostatnia, szósta edycja przyciągnęła uczestników, gdyż odbyła się wiosną 2020 roku, kiedy COVID-19 stał się pandemią. W tym czasie muzycy szczególnie intensywnie zaczęli szukać alternatywnych, zdalnych metod wykonawczych; zapotrzebowanie na takie rozwiązania może potwierdzić fakt, iż w 6. edycji wirtualnego chóru Erica Whitacre’a wzięło udział ponad 40 000 chórzystów z 145 krajów. Niemniej jednak, żadne zdalnie wykonane nagranie nie zostało uprzestrzennione z wykorzystaniem ambisonii. Dopiero w czasie pandemii koronawirusa COVID-19 powstały pierwsze tego typu produkcje [23,27,1,44,32]. Jedną z nich jest utwór Giovanniego Pierluigi da Palestriny *Sicut Cervus* wykonany przez Akademicki Chór Politechniki Gdańskiej w pionierskim nagraniu pt. *Wirtualna Katedra* [43][43X]. Problematyka związana z produkcją tego nagrania zostanie przedstawiona w następnych rozdziałach.

2. REALIZACJA ZDALNEGO NAGRANIA MUZYCZNEGO

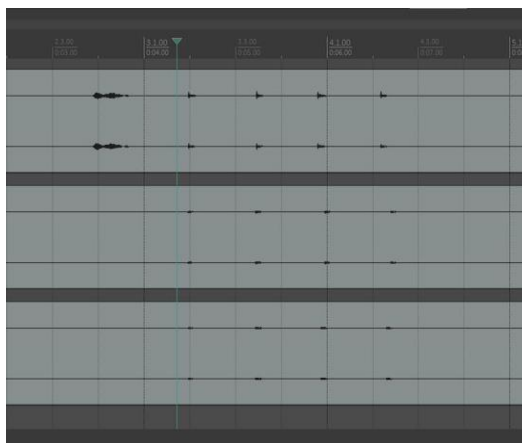
2.1. PRZYGOTOWANIE DO NAGRANIA

Pierwszym krokiem w tego typu przedsięwzięciu jest omówienie materiału muzycznego z zespołem; przedstawienie interpretacji muzycznej materiału przez dyrygenta zespołu, wskazanie miejsc w nutach (oraz ew. ich zaznaczenie) wymagających szczególnej uwagi, itp. Dodatkowo ważnym elementem jest omówienie techniki nagrania, estetyki kadru, spójnego ubioru, umiejscowienia urządzenia rejestrującego i wszelkich szczegółów z tym związanych. Po stronie realizatora kluczowe jest, aby zapewnić repozytorium do nadsyłania plików z nagraniami przez uczestników projektu.

Następnym krokiem jest stworzenie „nagrania wzorcowego”, swoistej macierzy, na której będą wzorować się wykonawcy utworu. Bardzo dobrą podpowiedzią może być nagranie wizualne, w którym dyrygent dyryguje utworem oraz wykonuje dany utwór na pianinie z transkrypcji fortepianowej (warstwę dźwiękową może wykonać też akompaniator). Następnie do skonstruowanego w ten sposób materiału – wizja+fonia = dyrygent+akompaniament – własne nagrania audiowizualne tworzą muzycy reprezentujące swoje partie. Dla przykładu: dla chóru będą to sopran, alt, tenor i bas; dla kwintetu smyczkowego będą to: I skrzypce, II skrzypce, altówka, wiolonczela, kontrabas. Ważne jest, aby te nagrania były nieco bardziej ekspresyjne niż typowo; w tym celu odtwórcy takiego „wzorca” będą mieli dodatkową warstwę informacji, z której będą mogli odtworzyć wspólną interpretację utworu. Rys. 1 prezentuje kadr z przykładowego „wzorcowego” nagrania. Dodatkową, niezwykle pomocną mogą być ustalenia z wykonawcami, aby przed wykonaniem swojej partii w umówiony sposób wystukali 1-2 takty rytmu (np. 1-2 takty przed rozpoczęciem utworu). Pozwoli to kolejnym odtwórcom „wzorca” wczuć się w rytm oraz przygotować do nagrania. Ponadto, w post-produkcji będzie to dobry punkt odniesienia do synchronizacji ścieżek; co więcej, zostawi to odpowiednie miejsce przed rozpoczęciem wykonania do przycięcia w post-produkcji – dzięki temu uniknie się sytuacji, gdy dane nagranie jest za krótkie, bo np. ktoś jeszcze nie był gotowy do wykonania swojej partii. Rys. 2 przedstawia obrazowo ideę tego zabiegu.



Rys. 1. Przykładowy kadr nagrania wzorcowego partii basu



Rys. 2. Wstępna synchronizacja ścieżek do 4 kłaśnieć przed rozpoczęciem wykonania

2.2. NAGRANIE PRZEZ MUZYKÓW

W ten sposób przygotowane nagrania wzorcowe powinny stanowić bazę, do której dogrywać swoje partie będą pozostali członkowie zespołu. Bardzo wygodne jest umieszczenie wzorców w platformie YouTube jako niepubliczne nagrania; dzięki temu w trakcie wykonywania swojej partii, muzyk będzie mógł podłączyć niewielkie słuchawki oraz trzymać telefon w zasięgu wzroku (np. w pobliżu nut) i na bieżąco kontrolować ruchy rąk dyrygenta oraz osobę wzorcową dla swojej partii muzycznej. Rys. 3 przedstawia sytuację nagraniową w opisanych warunkach. Nagrania nadesłane przez muzyków powinny być zweryfikowane przez dyrygenta i/lub producenta projektu pod względem artystycznym, ale też technicznym, w szczególności z punktu widzenia estetyki kadru oraz wierności odtworzenia wzorca, zwłaszcza od strony rytmicznej. Im wierniejsze wykonanie, tym mniej korekt trzeba będzie zastosować na późniejszym etapie procesu wydania nagrania.

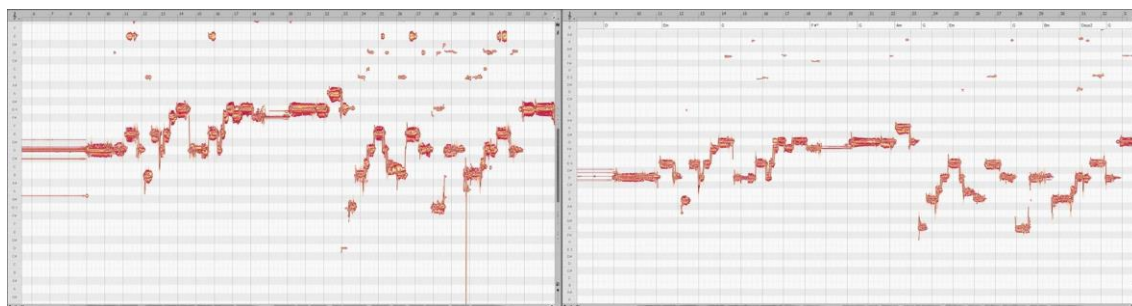


Rys. 3. Przykładowe ustawienie nagrania w warunkach domowych; na ekranie widoczne nuty oraz „wzorzec”

2.3. POST-PRODUKCJA WARSTWY DŹWIĘKOWEJ

Dla osiągnięcia najlepszego efektu konieczne są ręczne poprawki każdej z nadesłanych ścieżek dźwiękowych. Niestety, większość muzyków nie posiada w domu warunków do nagrania. W przypadku zespołu amatorskiego, jakim jest chór akademicki, większość uczestników wykonało swoje nagrania na telefon komórkowy. W ten sposób wykonane nagranie najczęściej wymaga usunięcia szumów. Można

go dokonać za pomocą bramek szumów oraz filtrów FIR. Kolejną sprawą są korekty samej warstwy melodyczno-rytmicznej. O ile nierówności intonacyjne, w dużych grupach wykonawczych, mogą zostać przykryte (lub nawet ściszone) przez większość czystszych wykonawców, o tyle nierówności rytmiczne są dużo bardziej zauważalne i psują dobre wrażenie odbioru produkcji. W tym celu można wykorzystać wiele dostępnych wtyczek VST, m.in. wbudowaną w DAW (*Digital Audio Workstation*) Reaper [39], wtyczkę ReaTune. Bardzo dobrym narzędziem jest też wtyczka Melodyne Studio [37].



Rys. 4. Przed i po korekcie nagrania w programie Melodyne Studio (partia altowa)

Duże grupy wykonawcze najczęściej nagrywa się w dużych salach koncertowych, nie zaś indywidualnie, każdego z osobna. Z tego względu, w ten sposób przeprowadzone nagranie pozwala na potraktowanie ścieżek obiektowo i uprzestrzennienie ich w ambisonii. W tym celu można wykorzystać wtyczki VST z IEM Plug-in Suite [31]. Aby jak najwierniej oddać wirtualną akustykę, należy najpierw użyć wtyczki DirectivityShaper [28]. W tej wtyczce można zaprojektować kierunkowość źródła dźwięku, aby jak najwierniej oddawał charakterystykę rzeczywistego źródła dźwięku. Następnie można użyć wtyczki RoomEncoder [30]. W tej wtyczce można zaprojektować pozycję źródła dźwięku w przestrzeni; można również opracować charakterystykę pomieszczenia, w tym liczbę odbić. Jednakże, zbyt duża liczba źródeł pozornych (wynikających z dużej liczby odbić) może zaciemnić obraz dźwiękowy; nie należy zatem ustawić zbyt dużej liczby odbić. Brakujący pogłos można dodać za pomocą wtyczki FDNReverb [29]. Należy tylko odpowiednio ustawić czas *fade-in*, tak aby był zgodny z czasem ostatniego odbicia z wtyczki RoomEncoder. To zapewni naturalne brzmienie tych dwóch pogłosów. Rys. 5 przedstawia przykładowe ustawienie tych trzech wtyczek VST.

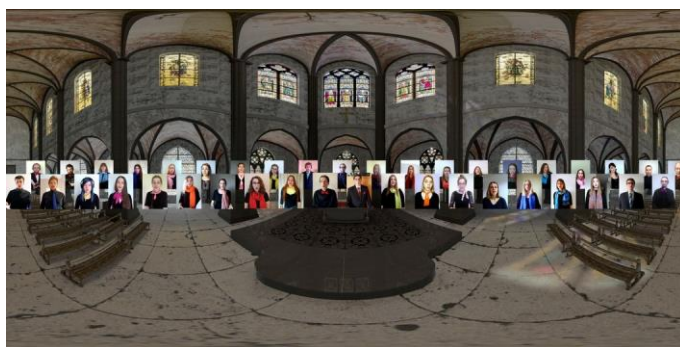


Rys. 5. Wtyczki VST z IEM Plug-in Suite: DirectivityShaper, RoomEncoder, FDNReverb

2.4. POST-PRODUKCJA WARSTWY WIZUALNEJ

Przy produkcji klasycznego chóru wirtualnego dużo uwagi wymaga zaplanowanie wszystkich ujęć, a także przejść pomiędzy nimi. W omawianym nagraniu zdecydowano się na statyczny obraz 360°, na którym zawieszone będą nagrania chórzystów. Od początku do końca w nagraniu widnieją te same osoby, gdyż przejścia między nimi byłyby nieestetyczne; ponadto, w nagraniu 360° twórca nie ma kontroli nad tym, gdzie w danym momencie spogląda widz. Dlatego obraz w całej sferze został niezmienny; zdecydowano się natomiast na stworzenie dopracowanego modelu, który posłużył do wygenerowania tła. Postanowiono użyć gotowy model 3D katedry, a następnie za pomocą programu Blender wykonać rendering wnętrza katedry z zadanego punktu – konkretnego ustawienia kamery. Kamera była ustawiona w tryb sferyczny, a dokładniej w odwzorowanie walcowe równoodległościowe. W ten sposób otrzymany render mógł posłużyć jako tło do ścieżek obrazu w programie do montażu nieliniowego – DaVinci Resolve [14]. Rys. 6 przedstawia układ chórzystów rozmieszczonych na tle modelu.

W ten sposób stworzony render można następnie złączyć z dźwiękiem w *b*-formacie, czyli dla ambisonii pierwszego rzędu jest to 4-kanałowa ścieżka audio. W tym celu posłużyć może program FB360Encoder [21], udostępniony przez platformę Facebook, gdyż wspiera on także kodowanie audio-video na platformę YouTube. Wynikowy plik, po załadowaniu do platformy, zostanie automatycznie rozpoznany jako film 360° z dźwiękiem ambisonicznym (dzięki meta-danym pliku), a po przetworzeniu przez platformę będzie on dostępny do odtwarzania.



Rys. 6. Rozmieszczenie chórzystów na tle wyrenderowanego ujęcia z wnętrza modelu katedry

3. PODSUMOWANIE

W niniejszym tekście przedstawiono i krótko omówiono dotychczasowe dokonania w dziedzinie zdalnych połączeń audio-wizualnych. Przedstawiono trudności związane z wykonawstwem muzycznym w warunkach zdalnych. Omówiono także przypadki przestrzennych transmisji oraz zdalnych produkcji. W sposób szczegółowy zaprezentowano proces produkcyjny zdalnego nagrania chóralnego, które wykonał Akademicki Chór Politechniki Gdańskiej. Nagranie to prezentuje wysoki poziom artystyczny, zapewniając słuchaczom nowe doznania wynikające z immersji w wirtualnej rzeczywistości. Podkreślają one walory chórów wirtualnych, a zwłaszcza ich innowacyjny charakter, których możliwości twórcze są wartością kultury XXI wieku.

Kolejnym etapem tego projektu artystycznego będą badania subiektywne ocena różnych aspektów nagrania ambisonicznego w kontekście akceptacji przez odbiorców.

LITERATURA

- [1] A socially-distanced, 360 performance of Puccini's Turandot (Royal Opera House Chorus and Orchestra). Online: <https://youtu.be/VwOpNf8eHeY> (dostęp dnia 04.10.2020)
- [2] AltSpaceVR – Try spatialized audio in AltSpaceVR. Online: <https://altvr.com/try-spatialized-audio-in-altspacevr/> (dostęp dnia 04.10.2020)
- [3] Aoki S., Cohen M., Koizumi N.: Design and control of shared conferencing environments for audio telecommunication using individually measured HRTFs. *Presence* 3, No 1 (1994) 60–72.
- [4] Bartlette C., Headlam D., Bocko M., Velikic G.: Effect of network latency on interactive musical performance. *Music Perception* 24, No 1 (2006) 49–62.
- [5] Bouillot N., Cooperstock J. R.: Challenges and performance of High-Fidelity audio streaming for interactive performances. *Proceedings of the 9th International Conference on New Interfaces for Musical Expression*, (2009), 135–140.
- [6] Bouillot N.: nJam user experiments: enabling remote musical interaction from milliseconds to seconds. *Proceedings of the 7th International Conference on New Interfaces for Musical Expression* (2007) 142–147.
- [7] Buxton W.: Telepresence: integrating shared task and person spaces. *Proceedings of Graphics Interface '92* (1992) 123–129.
- [8] Caceres J.-P., Hamilton R., Iyer D., Chafe C., Wang G.: To the edge with China: Explorations in network performance. *ARTECH 2008: Proceedings of the 4th International Conference on Digital Arts* (2008).
- [9] Caceres J.-P., Chafe C.: JackTrip/SoundWIRE meets server farm. *Computer Music Journal* 34, No 3 (2010) 29–34.
- [10] Carôt A., Sardis F., Dohler M., Saunders S., Uniyal N., Cornock R.: Creation of a Hyper-Realistic Remote Music Session with Professional Musicians and Public Audiences Using 5G Commodity Hardware. *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, London, United Kingdom (2020) 1–6. doi: 10.1109/ICMEW46912.2020.9105995
- [11] Chafe C., Caceres J. P., Gurevich M.: Effect of temporal separation on synchronization in rhythmic performance. *Perception* 39, No 7 (2010) 982–992.
- [12] Chafe C., Wilson S., Leistikow R., Chisholm D., Scavone G.: A simplified approach to high quality music and sound over IP. *COST-G6 Conference on Digital Audio Effects* (2000) 159–164.
- [13] Chafe C.: Tapping into the internet as an acoustical/musical medium. *Contemporary Music Review* 28, No 4 (2009) 413–420.
- [14] DaVinci Resolve. Online: <https://www.blackmagicdesign.com/products/davinciresolve/> (dostęp dnia 04.10.2020).
- [15] Deppisch T., Meyer-Kahlen N., Hofer B., Latka T., Zernicki T.: HOAST: A Higher-Order Ambisonics Streaming Platform. *Proceedings of the 148th Audio Engineering Society Convention* (2020).
- [16] Durlach N. I., Shinn-Cunningham B. G., Held R. M.: Supernormal auditory localization. *Presence* 2, No 2 (1993) 89–103.
- [17] Durlach N.: Auditory localization in teleoperator and virtual environment systems: ideas, issues, and problems. *Perception* 20, No 4 (1991) 543–554.
- [18] Eric Whitacre's Virtual Choir: <https://ericwhitacre.com/the-virtual-choir/about> (dostęp dnia 04.10.2020)
- [19] Facebook 360 spatial workstation – Creating Videos with Spatial Audio for Facebook 360. Online: <https://facebookincubator.github.io/facebook-360-spatial-workstation/KB/CreatingVideosSpatialAudioFacebook360.html> (dostęp dnia 04.10.2020)
- [20] Facebook 360 spatial workstation – Using an Ambisonic Microphone With Your Live 360 Video on Facebook. Online: <https://facebookincubator.github.io/facebook-360-spatial-workstation/KB/UsingAnAmbisonicMicrophone.html> (dostęp dnia 04.10.2020)
- [21] Facebook 360 Spatial Workstation. Online: <https://facebook360.fb.com/spatial-workstation/> (dostęp dnia 04.10.2020).
- [22] Frank M., Sontacchi A.: Case Study on Ambisonics for Multi-Venue and Multi-Target Concerts and Broadcasts. *J. Audio Eng.Soc.* 65, No 9 (2017) 749–756. doi: 10.17743/jaes.2017.0026
- [23] Georgia Symphony Chorus – Georgia On My Mind – 360° Virtual Choir with adaptive audio in 8K. Online: <https://youtu.be/BrXZ63nOUhU> (dostęp dnia 04.10.2020)
- [24] Gu X., Dick M., Kurtisi Z., Noyer U., Wolf L.: Network-centric music performance: practice and experiments. *IEEE Communications Magazine* 43, No 6 (2005) 86–93.
- [25] Gurevich M., Donohoe D., Bertet S.: Ambisonic spatialization for networked music performance. *17th International Conference on Auditory Display* (2011).

- [26] Gurevich M.: JamSpace: a networked real-time collaborative music environment. CHI'06 Extended Abstracts on Human Factors in Computing Systems (2006) 821-826.
- [27] I(solace)ion (Juliana Kay & Exaudi) | 360° – Exaudi. Online: <https://youtu.be/HkiIUeuugk8> (dostęp dnia 04.10.2020)
- [28] IEM Plug-in Suite – DirectivityShaper. Online: <https://plugins.iem.at/docs/directivityshaper/> (dostęp dnia 04.10.2020).
- [29] IEM Plug-in Suite – FDNReverb. Online: <https://plugins.iem.at/docs/pluginDescriptions/#fdnreverb> (dostęp dnia 04.10.2020).
- [30] IEM Plug-in Suite – RoomEncoder. Online: <https://plugins.iem.at/docs/pluginDescriptions/#roomencoder> (dostęp dnia 04.10.2020).
- [31] IEM Plug-in Suite. Online: <https://plugins.iem.at/> (dostęp dnia 04.10.2020).
- [32] J. S. Bach – Koncert na dwa skrzypce BWV 1043 [360°]. Online: <https://youtu.be/mQXNneuRG3s> (dostęp dnia 04.10.2020)
- [33] Jouppe N. P., Pan M. J.: Mutually-immersive audio telepresence. Proceedings of the 113th Audio Engineering Society Convention (2002).
- [34] Jubileusz 10-lecia Internetu w Polsce – Koncert internetowy. Online: <http://www.internet10.pl/koncert.html> (dostęp dnia 04.10.2020)
- [35] Kapur A., Wang G., Davidson P., Cook P. R.: Interactive network performance: A dream worth dreaming?. Organised Sound 10, No 3 (2005) 209-219.
- [36] Lazzaro J., Wawrzynek J.: A case for network musical performance. Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video (2001) 157-166.
- [37] Melodyne Studio. Online: <https://www.celemony.com/en/melodyne/what-is-melodyne> (dostęp dnia 04.10.2020).
- [38] PURE Ambisonics Concert & the Night of Ambisonics. Online: <https://ambisonics.iem.at/icsa2015/pure-ambisonics-concert> (dostęp dnia 04.10.2020).
- [39] Reaper. Online: <https://www.reaper.fm/> (dostęp dnia 04.10.2020).
- [40] Rosen T.: Is it live or is it Internet2?: Miro quartet shows how technology may change the future of live performances (2004). Online: <http://www.utexas.edu/features/archive/2004/internet.html>
- [41] Rottondi C., Chafe C., Allocchio C., Sarti A.: An Overview on Networked Music Performance Technologies. IEEE Access 4 (2016) 8823-8843. doi: 10.1109/ACCESS.2016.2628440.
- [42] Rudrich D., Zotter F., Frank M.: Efficient Spatial Ambisonic Effects for Live Audio. Proceedings of 29th Tonmeistertagung (2016).
- [43] Sicut Cervus – Wirtualna Katedra #ZostańWDomu #ŚpiewajWDomu [4k 360°]. Online: <https://youtu.be/4dwSRNxUrlU> (dostęp dnia 04.10.2020)
- [43X] Giovanni Pierluigi da Palestrina – Sicut Cervus, Akademicki Chór Politechniki Gdańskiej, dyryguje Mariusz Mróz, zdalne nagranie wirtualne (maj 2020)
- [44] Socially Distant Orchestra plays "Jupiter" in 360°. Online: <https://youtu.be/eiouj6HkjfA> (dostęp dnia 04.10.2020)
- [45] Weinberg G.: Interconnected musical networks: Toward a theoretical framework. Computer Music Journal 29, No 2, 23-39, 2005.
- [46] Wenzel E. M., Wightman F. L., Kistler D. J.: Localization with non-individualized virtual acoustic display cues. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (1991) 351-359.
- [47] Xu A., Woszczyk W., Settler Z., Pennycook B., Rowe R., Galanter P., Bary J., Martin G., Corey J., Cooperstock J. R.: Real-time streaming of multichannel audio data over internet. J. Audio Eng. Soc. 48, No 7-8, 627-641, 2000.
- [48] YouTube help – Use spatial audio in 360-degree and VR videos. Online: <https://support.google.com/youtube/answer/6395969> (dostęp dnia 04.10.2020)
- [49] Zimmermann R., Chew E., Ay S. A., Pawar M.: Distributed musical performances: Architecture and stream management. ACM Transactions on Multimedia Computing, Communications, and Applications 4, No 2 (2008) 1-23.

MULTICHANNEL TECHNIQUES IN THE APPLICATION OF REMOTE CONCERTS AND MUSIC RECORDINGS AT A DISTANCE

During COVID-19 pandemic, the possibility of transmitting sound with video, especially for remote working, has gained new significance. The increased demand for remote, multi-source connections has brought attention to multi-channel sound and spatialization, for example in order to facilitate the location of sound sources – but not only. Remote working is a challenge especially for musicians; both in terms of joint exercises and rehearsals, or even more in terms of concerts and performances. Creating remote music recordings is a unique opportunity for multi-channel, spatial productions using ambisonic and binaural techniques. These techniques allow creating new, immersive experiences for the audience. This publication discusses in detail the use of binaural and ambisonic techniques in recorded music based on remote performance of a choir using Internet network. In particular, the latest projects of the Academic Choir of Gdańsk University of Technology are discussed in the context of spatialization techniques utilized as well as synchronization issues.

STANOWISKO LABORATORYJNE KOLUMNY Z NIEZALEŻNIE STEROWANYMI GŁOŚNIKAMI

Michał LUCZYŃSKI¹

¹Politechnika Wrocławska, Wyspiańskiego 27, 50-370 Wrocław
michal.luczynski@pwr.wroc.pl

W ramach pracy przedstawiono stanowisko dydaktyczne składającego się z kolumny głośnikowej wyposażonej w 8 przetworników, zestawu mikrofonów pomiarowych i interface'u audio wraz z oprogramowaniem DAW. Przeznaczeniem stanowiska jest demonstracja zasady działania sterowanych cyfrowo kolumn głośnikowych z niezależnie sterowanymi przetwornikami. Dzięki zastosowaniu dowolnego oprogramowania DAW zaproponowane stanowisko (set-up) jest zarówno łatwy do zestawienia jak i prosty w obsłudze. Dzięki ustawianiu odpowiednich poziomów i opóźnień na sygnałach na poszczególnych ścieżkach i obserwować wpływ na kierunkowość źródła. Obserwacja kierunkowości może odbywać organoleptycznie (słuchowo) oraz za pomocą obserwacji zmian poziomów na mikrofonach pomiarowych. Materiały dydaktyczne wsparte są symulacjami komputerowymi rozkładu poziomu ciśnienia akustycznego dla różnych konfiguracji sterowania wiązką. Zaprezentowana metody może jednocześnie być interesująca dydaktycznie w celu demonstracji sterowania wiązką w przypadku przetworników ultradźwiękowych.

1. WPROWADZENIE

1.1. KOLUMNY GŁOŚNIKOWE

Kolumna głośnikowa jest specyficznym rozwiązaniem wśród urządzeń głośnikowych o bardzo nietypowych właściwościach. Jednocześnie to pojęcie często jest mylnie przypisywane zestawom głośnikowym. Należy mieć na względzie, że jest to trudne, aby zmienić świadomość użytkowników nieprofesjonalnych. Jednak trzeba zadbać, aby studenci kierunków związanych z akustyką rozróżniali te pojęcia oraz potrafili zdefiniować możliwości jakie stoją za zastosowaniem tego typu rozwiązań. W literaturze, która stanowi podstawowe źródło wiedzy w dziedzinie [1,2] pojęcie kolumny głośnikowej jest bardzo skutecznie wytłumaczone a przedstawiony opis matematyczny pozwala na obliczenie ich parametrów akustycznych. Wiedza teoretyczna najlepiej jest zapamiętywana i przyswajana kiedy jest związana z praktyką. W ramach tej pracy przedstawiono propozycję prostego w zbudowaniu i obsłudze stanowiska laboratoryjnego, który pozwoli na praktyczne zapoznanie się z układami kolumn głośnikowych. Przedstawione rozwiązanie pozwala zarówno załączanie różnej liczby przetworników, niezależne sterowania poszczególnymi głośnikami oraz łatwe rozbudowanie konstrukcji o kolejne moduły. Kolejne moduły można rozbudowywać tworząc nie tylko dłuższe kolumny głośnikowe, ale również dwu-wymiarowe matryce głośnikowe.

Kolumnę głośnikową tworzy kilka lub kilkanaście jednakowych głośników, które są umieszczone w jednej linii. Przeważnie są umieszczone w jednej obudowie. Dzięki temu kolumna głośnikowa tworzy źródło liniowe. W przypadku, gdy wszystkie przetworniki są zasilane takim samym sygnałem, kolumna głośnikowa wykazuje się zawężeniem charakterystyki kierunkowości w płaszczyźnie osi kolumny. W płaszczyźnie prostopadłej do osi kolumny charakterystyka kierunkowości jest zgodna z charakterystyką pojedynczego przetwornika. Jednocześnie ta charakterystyka zależy od częstotliwości a dokładnie od stosunku odległości między osiami przetworników a długością fali oraz liczby przetworników tworzą-

cych kolumnę. Dodatkowo na charakterystykę kierunkowości kolumny można wpłynąć poprzez indywidualne sterowanie każdym z przetworników. Sygnały sterujące różnią się wzmocnieniem i biegunością (przesunięcia fazowe). Głównym zadaniem jest sztuczne uzyskanie efektu pochylenia kolumny głośnikowej.

Oczywistą wadą urządzeń głośnikowych opartych o jeden rodzaj przetwornika jest ograniczenie pasma częstotliwości wynikające z trudności w efektywnym pobudzeniu całego pasma akustycznego przez pojedynczy głośnik. Jednym ze współczesnych rozwiązań jest zastosowanie dodatkowego głośnika niskotonowego [3] który może stanowić jedno modułowe urządzenie. Dzięki temu możliwe jest uzyskanie większego użytecznego zakresu częstotliwości. Należy jednak pamiętać, że cechy takie rozwiązuje wykazuje cechy kolumny tylko w tym zakresie częstotliwości w jakim pracują głośniki tworzące szereg jednakowych przetworników. Mimo wszystko urządzenia składające się z wielu jednakowych przetworników w których można niezależnie sterować pracą każdego z nich są coraz bardziej popularne. Mowa tutaj oczywiście o listwach dźwiękowych (ang. Soundbar). W tym przypadku stosowane są dużo bardziej skomplikowane algorytmy przetwarzania sygnałów niż w typowych kolumnach głośnikowych, gdzie sterowanie odbywa się głównie w celu zmiany kierunku promieniowania dźwięku.

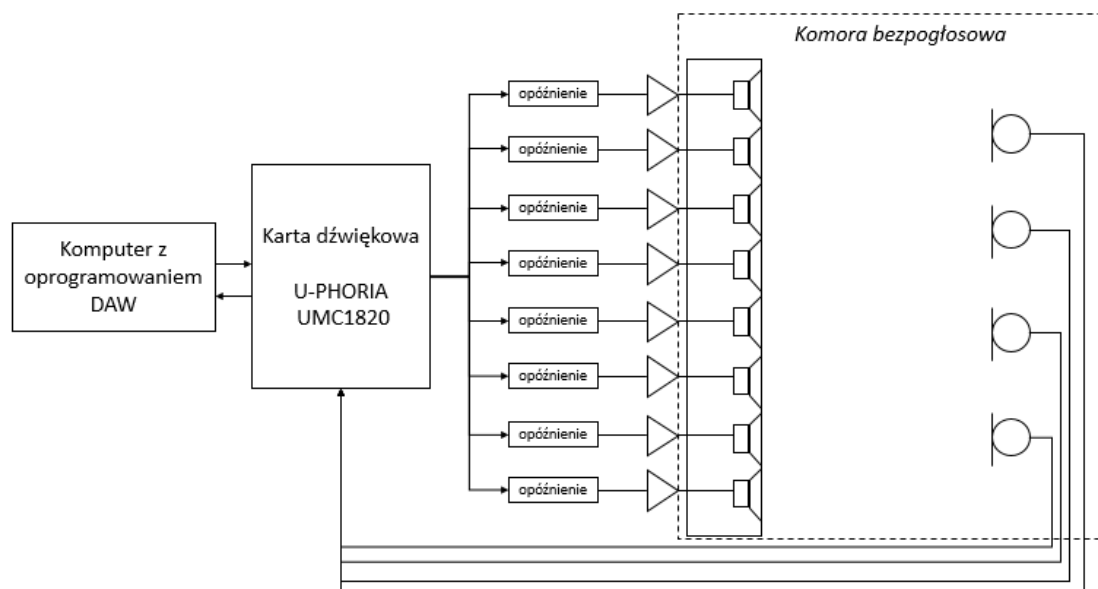
Dodatkowo kształtowanie sygnału za pomocą odpowiedniego opóźniania i wzmacniania sygnałów sterujących wykorzystywane jest w innych dziedzinach akustyki. W technice ultradźwiękowej stosuje się takie przetwarzanie w celu uzyskania odpowiedniego ogniskowania wiązki ultradźwiękowej [4]. Testowanie takich rozwiązań w zakresie częstotliwości słyszalnych może pomóc zrozumieć to zagadnienie studentom tej dziedziny.

2. PROPONOWANA KONSTRUKCJA

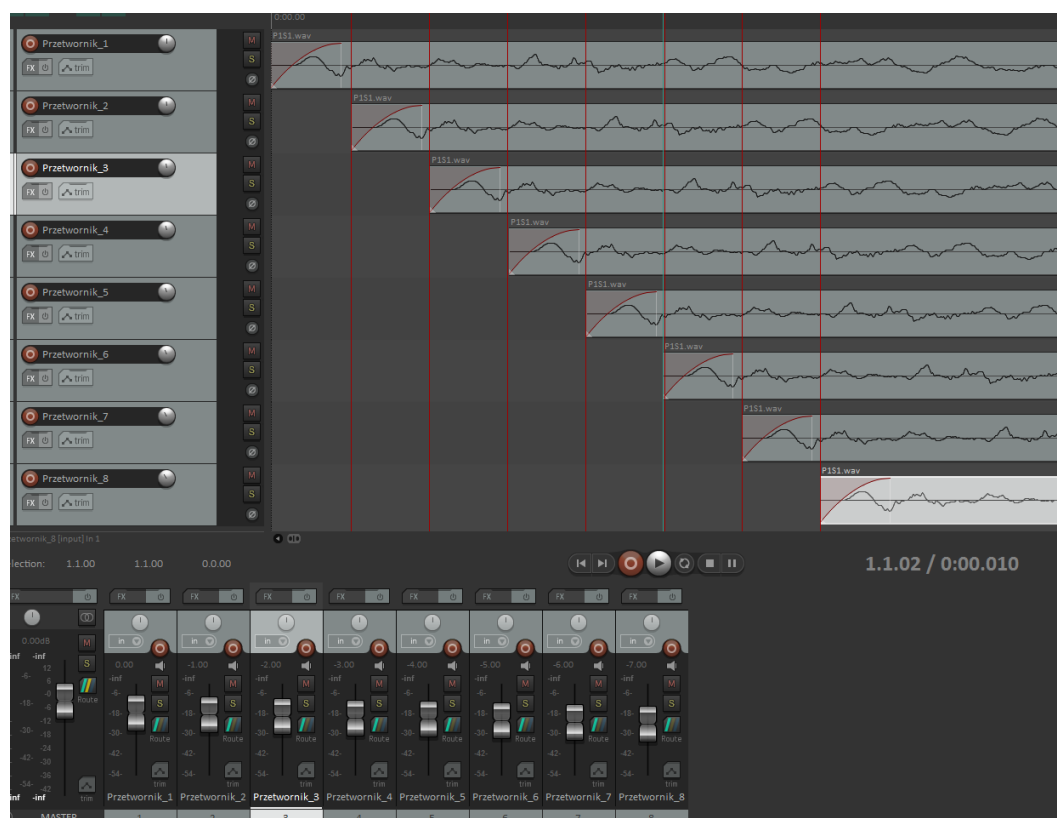
Podstawowe założenia projektowe tworzenia przedstawionego stanowiska laboratoryjnego to możliwie prosta i tania konstrukcja i obsługa prosta dla studentów kształcących się na kierunkach związanych z akustyką. W ramach pracy przedstawiono propozycję budowy stanowiska wraz z przykładowymi symulacjami i propozycją pomiarów i eksperymentów jakie można przeprowadzić z jego użyciem.

System składa się z kolumny głośnikowej wyposażonej w 8 przetworników 2,5 calowych i 8 wzmacniaczy 3-watowych, 8-kanalowego interfejsu audio, komputera z oprogramowaniem DAW oraz okablowania. Konfiguracja systemu polega na przygotowaniu 8 ścieżek w programie DAW i przypisaniu ich do odpowiednich wyjść. Następnie wyjścia sygnałowe należy podłączyć do wejść wzmacniaczy, które zasilają przetworniki. Schemat blokowy proponowanego systemu przedstawiono na rysunku 1.

Interfejs wykorzystany do systemu to U-PHORIA UMC1820 [5]. Jest to interfejs z możliwością pracy do 24 bitów i częstotliwością próbkowania do 96 kHz. Posiada on 8 wejść typu combo (wejścia mikrofonowe/liniowe), 10 wyjść liniowych i 2 wyjścia słuchawkowe. Ma również możliwość rozszerzenia liczby wejść i wyjść za pomocą protokołu ADAT (8 wejść i 8 wyjść z częstotliwością próbkowania 44,1 lub 48 kHz). Interfejs jest podłączany do komputera z wykorzystaniem przewodu USB. Używanym sterownikiem jest ASIO [6], a wybrany program do obróbki dźwięku to Reaper [7]. Każda ze ścieżek odpowiada sygnałowi, który zasilą każdy z przetworników. Takie rozwiązanie pozwala na proste ustawienie wzmocnień za pomocą suwaków odpowiadającym ścieżkom. Opóźnienie może być ustawiane poprzez ręczne przesunięcie fragmentu materiału na ścieżce. Dodatkowo takie rozwiązanie pozwala w prosty sposób zastosować dodatkowe efekty na każdej ścieżce, np. filtrację częstotliwościową. Porównanie różnych ustawień również może być proste dzięki ustawieniu kolejnej konfiguracji sygnałów zasilających jedna po drugiej. Na rysunku 2 przedstawiono przykładową konfigurację w ramach, które każdy kolejny przetwornik jest zasilany sygnałem o poziomie 1 dB mniejszym a opóźnienie na każdym z przetworniku jest o 2 ms większe.



Rysunek 1 Schemat blokowy systemu stanowiska laboratoryjnego

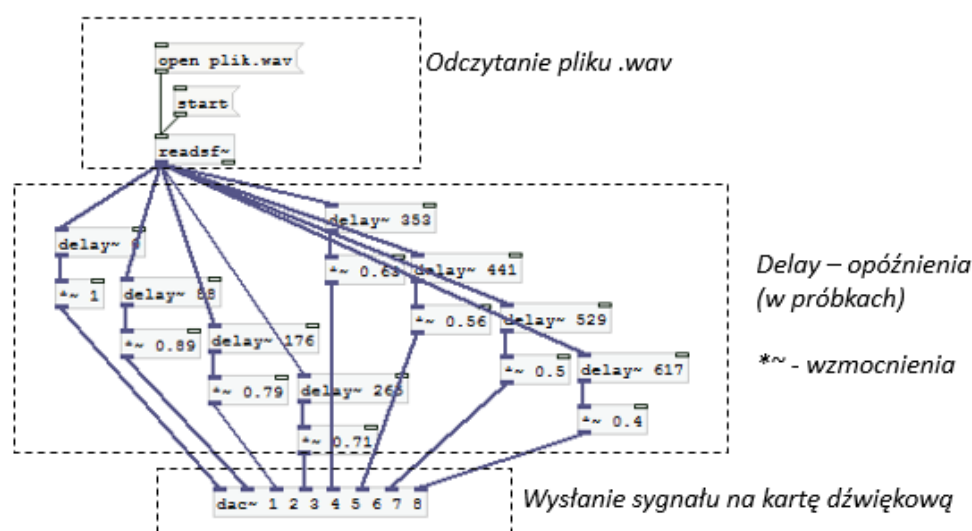


Rysunek 2 Przykładowa konfiguracja opóźnień i wzmacnień z poziomu programu DAW

Zaproponowana konstrukcja kolumny głośnikowej składa się z 8 przetworników 2,5 calowych o impedancji znamionowej równej 8 Ohm. Każdy z głośników posiada osobny wzmacniacz o mocy 3 W.

Zastosowane wzmacniacze to gotowe układy scalone dostępne w niskiej cenie i zasilane napięciem stałym równym 5 V. Takie napięcie zasilające o dobrych parametrach (wystarczające natężenie prądu, niskie napięcia tętnień) jest łatwe do uzyskania, np. dzięki użyciu starej ładowarki do telefonu. Wszystkie przetworniki znajdują się w jednej obudowie. Odległość między osiami przetworników jest stała i wynosi 10 cm. Odległość między osiami skrajnych przetworników a krawędzią obudowy wynosi 5 cm (połowa odległość między osiami przetworników). Dzięki temu można rozbudować układ poprzez dołożenie dodatkowego urządzenia składającego się z kolejnym 8 przetworników pozostawiając stałą odległość między osiami przetworników. Układ planowo można rozbudowywać poprzez łączenie urządzeń dłuższymi bokami. W ten sposób można utworzyć dwu-wymiarową macierz głośnikową. W przypadku zestawienia ze sobą 8 kolumny głośnikowych tworząc macierz o wymiarach 8x8 potrzebna jest karta dźwiękowa, która jest w stanie obsługiwać 64 niezależne kanały. Niestety takie urządzenia nie są ani popularne, ani dostępne w niskich cenach. Z tego powodu przewiduje się uproszczone rozwiązanie w ramach, którego przetworniki mogą być łączone w grupy i zasilane wspólnie (po 2, po 4 lub po 8 przetworników).

Podstawowym planowym źródłem sygnału w przypadku zastosowania dydaktycznego planuje się DAW (np. Reaper). Jednak możliwości takiego systemu można skutecznie rozszerzyć poprzez zastosowanie środowisk programistycznych przeznaczonych do przetwarzania sygnałów audio (MAX MSP[8] lub PureData[9]). Przykładowy kod programu w PureData realizujący opóźnienie zwiększane co 0,2 ms dla każdego przetwornika i ze zmniejszaniem poziomu co 1 dB przedstawiono na rysunku 3. Możliwości eksperymentowania z zastosowaniem m.in. losowych lub zmienianych opóźnień i wzmocnień oraz efektów takich jak filtry może przynieść dużo obserwacji i efektów zarówno w zakresie informacji o przetwarzaniu sygnałów jak i efektów czysto artystycznych.

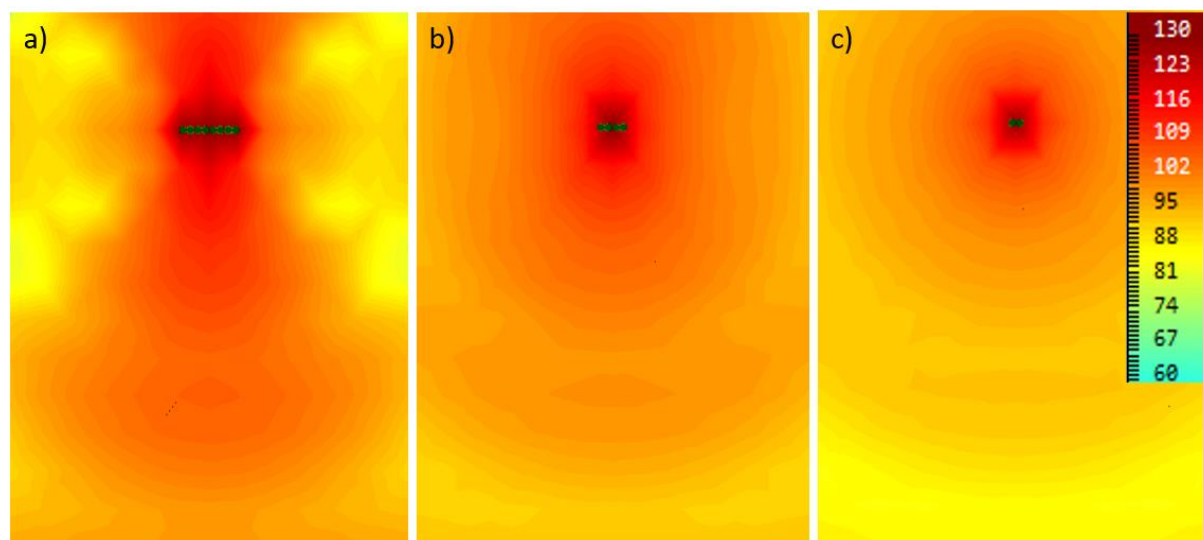


Rysunek 3 Kod programu PureData z zastosowanymi różnymi opóźnieniami i wzmocnieniami dla różnych kanałów wraz z zaznaczonymi blokami przetwarzania sygnałów

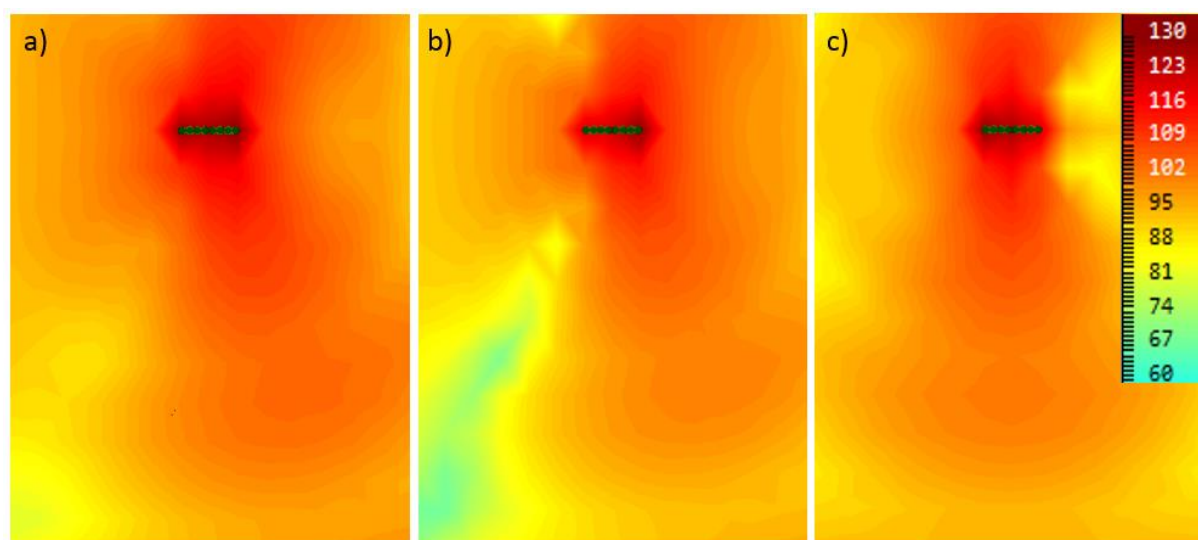
3. SYMULACJE

W tym rozdziale przedstawiono przykładowe symulacje rozkładu poziomu ciśnienia akustycznego generowanego przez badaną kolumnę głośnikową. Symulacje przeprowadzono w polu swobodnym. Wybrano środowisko symulacyjne Acoustic Boundary Element Calculator (ABEC) [10] korzystające z Metody Elementów Brzegowych. ABEC jest środowiskiem obliczeniowym pozwalającym na bardzo dokładne modelowanie urządzeń głośnikowych, ale również pozwala na prostsze obliczenia rozkładu

poziomu ciśnienia. Pole obliczeniowe co do wymiarów jest zgodne z powierzchnią podłogi komory bezekowej Politechniki Wrocławskiej. W ramach przedstawionych symulacji przedstawiono uproszczone modele zbudowane z szeregu źródeł punktowych. Uproszczenie ma na celu wskazanie prostych możliwości uzyskania symulacji, które można porównywać z wynikami pomiarów poziomu ciśnienia akustycznego i wrażeniami subiektywnymi. Rysunek 4 przedstawia symulacje rozkładu poziomu dla trzech różnych długości kolumny. Następnie rysunek 5 przedstawia symulacje dla dwóch różnych konfiguracji opóźnień i wzmocnień sygnałów sterujących przetworniki.



Rys. 4 symulacje dla różnych długości kolumny



Rys. 5 symulacje dla różnych konfiguracji opóźnień i wzmocnień

4. WYKORZYSTANIE W DYDAKTYCE

Przedstawiona konstrukcja ma duże możliwości zastosowania w dydaktyce. Głównie dzięki niezależnemu sterowaniu każdego z głośników można przygotować szereg eksperymentów. W tym rozdziale przedstawiono skrótowo kilka propozycji ćwiczeń laboratoryjnych.

4.1. WZAJEMNA IMPEDANCJA PROMIENIOWANIA

Jednym z problemów urządzeń głośnikowych składających się z więcej niż jednego przetwornika jest wzajemne wpływanie ich na siebie. To zjawisko nazywa się wzajemną impedancją promieniowania, która zależy m.in. od odległości między przetwornikami i częstotliwością. Korzystając z przedstawionego systemu można porównywać wartości obliczone teoretycznie z wartościami zmierzonymi. Można również badać wpływ pracy urządzenia dwu-drożnego dla sygnałów szerokopasmowych podając na przetworniki sygnały odpowiednie filtrowane.

4.2. LINIA ŹRÓDEŁ PUNKTOWYCH

Zagadnienie związane z teorią źródeł liniowych i linii źródeł punktowych realizowane na przedstawionym stanowisku można rozpatrywać dla kilku różnych wariantów. Wartości teoretyczne spadku poziomu ciśnienia akustycznego w funkcji odległości od źródła dźwięku można realizować dla różnej liczby przetworników budujących linię źródeł punktowych. Można wykorzystać wszystkie przetworniki lub skracać długość źródła liniowego. Można również zmniejszać liczbę źródeł punktowych przeprowadzając pomiary dla większej odległości między poszczególnymi źródłami.

4.3. KOLUMNA GŁOŚNIKOWA O RÓŻNEJ DŁUGOŚCI

Jednym z istotnych aspektów przedstawionego stanowiska jest możliwość badania jak liczba przetworników budujących kolumnę głośnikowych wpływa na jej charakterystykę kierunkową. Możliwe jest wykonywanie pomiarów dla kolumn o różnej długości. Dzięki prostemu sterowaniu źródłami za pomocą programów DAW można również przygotować ścieżkę audio w ramach której dołączane są kolejne źródła. Następnie można w łatwy sposób słuchowo ocenić jak dana kolumna zmienia swoją kierunkowość. Takie porównanie łatwo przeprowadzić dla różnego typu sygnałów. Można również przeprowadzić porównanie kolumn głośnikowych zbudowanych z tej samej liczby przetworników, lecz rozmieszczonych w różnych odległościach. Przeprowadzenie takiego porównania sprowadza się do utworzenia sygnałów sterujących na odpowiednie przetworniki.

4.4. ZMIANA KIERUNKOWOŚCI KOLUMNY GŁOŚNIKOWEJ

Najważniejszym aspektem przedstawionego stanowiska jest możliwość prezentowania zmiany kierunkowości kolumny głośnikowej. Jak już zostało wcześniej wspomniane, za pomocą prostego ustawienia opóźnień i amplitud można spowodować, że kolumna skieruje dźwięk w innym kierunku. Modyfikacje opóźnień i wzmocnień odbywają się bez ponownego wgrywania ustawień. Efekty modyfikacji mogą być odsłuchiwane na bieżąco. Korzystając z przedstawionego układu przetworników można prezentować zasadę działania ogniskowania fali ultradźwiękowej, która realizowana jest dla wielu przetworników. Analogia opierająca się na przeniesieniu rozpatrywanego zjawiska do pasma akustycznego pomoże w zrozumieniu zagadnienia. Odpowiednie sterowanie przedstawionego urządzenia głośnikowego pozwoli również zaprezentować efekty jakie są uzyskiwane za pomocą soundbar'ów. To wszystko jest możliwe do uzyskania w prosty sposób dzięki bezpośredniemu dostępowi do sygnałów podawanych na dane ścieżki. Sytuacji można w łatwy sposób rozbudować o trudniejsze eksperymenty korzystając ze środowisk programistycznych takich jak PureData.

5. PODSUMOWANIE

W ramach pracy przedstawiono koncepcję wraz z symulacjami stanowiska laboratoryjnego składającego się z kolumny głośnikowej z niezależnie sterowanymi przetwornikami. Zgodnie z założeniem konstrukcji jest wykorzystana ogólnie dostępna na rynku karta dźwiękowa zapewniająca transmisję przynajmniej 8 kanałów. Całość może być sterowana za pomocą dowolnego oprogramowania DAW lub środowisk programistycznych typu MaxMSP lub PureData. Główną zaletą konstrukcji jest jej niska cena, łatwość w budowie i konfiguracji oraz szeroki zakres zastosowania w celach dydaktycznych. Dzięki niej studenci mogą się w praktyce zapoznać z pojęciem linii źródeł punktowych, kolumny głośnikowej i niezależnego sterowania poszczególnymi przetwornikami. Stanowiska stwarza również możliwość badania możliwości efektywnych jakie można uzyskać na urządzeniach typu soundbar. Kolumna głośnikowa jest przygotowana w taki sposób, aby w łatwy sposób można było ją rozbudować zarówno w celu uzyskania kolumny o większej długości jak i matrycy głośnikowej. Taka matryca głośnikowa pozwala na odtworzenie czoła fali.

LITERATURA

- [1] Żyszkowski Z., *Podstawy Elektroakustyki*, Wydanie II, Wydawnictwo Naukowo-Techniczne Warszawa 1965
- [2] Dobrucki A. *Przetworniki Elektroakustyczne*, Wydawnictwo Naukowo-Techniczne Warszawa 2007
- [3] Alexander Pietschmann, *Active Loudspeaker Column System*, Patent nr US20170171649A1 <https://patents.google.com/patent/US20170171649A1/en>
- [4] Shuki Vitek, *Systems and methods for testing and calibrating a focused ultrasound transducer array*, Patent nr US6543272B1, <https://patents.google.com/patent/US6543272B1/en>
- [5] U-PHORIA UMC1820 Quick Start Guide, https://www.bhphotovideo.com/lit_files/155647.pdf
- [6] ASIO4ALL - Universal ASIO Driver For WDM <http://www.asio4all.org/>
- [7] Reaper Digital Audio Workstation <https://www.reaper.fm/>
- [8] Cycling '74: Tools for sound, graphics, and interactivity <https://cycling74.com/products/max/>
- [9] Pure Data – Pd Community Site <https://puredata.info/>
- [10] Joerg Panzer. *Coupled lumped and boundary element simulation for electro-acoustics*. Acoustics 2012, Apr 2012, Nantes, France. fhal-00811256f

LINE SOURCE WITH INDEPENDENTLY CONTROLLED LOUDSPEAKERS FOR DIDACTIC PURPOSE

In this paper, a laboratory set consisting of a loudspeaker with 8 transducers, a set of measuring microphones, and an audio interface with DAW software was presented. The purpose of the set is to demonstrate the operation of a digitally controlled line sound source with independently controlled transducers. Due to the use of any DAW software, the proposed system is both easy to set up and easy to use. By setting the appropriate levels and delays of the signals on the individual paths the effect on the source directivity could be observed. Observation of directivity can be carried out subjectively (listening test) and objectively (using measuring microphones). Didactic materials are supported by computer simulations of the sound pressure level distribution for various beam control configurations. The presented methods can also be interesting didactically for the demonstration of beam control in the case of ultrasonic transducers.

AMBISONICZNA MAPA WYBRANYCH MIEJSC W TRÓJMIĘŚCIE

Cezary PIETRZAK¹

Piotr ODYA²

Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki,

Katedra Systemów Multimedialnych, ul. Gabriela Narutowicza 11-12, 80-233 Gdańsk

¹cezarypiotrpietrzak@gmail.com

²pioodya@pg.edu.pl

Projekt miał na celu stworzenie ambisonicznej mapy Trójmiasta w formie aplikacji internetowej. Materiały wideo w technologii 360 z dźwiękiem w postaci sygnału ambisonicznego zostały zarejestrowane w lokalizacjach Trójmiasta, które uznano za charakterystyczne dla tej aglomeracji. Celem badawczym projektu było porównanie dostępnych algorytmów miksowania sygnałów ambisonicznych poprzez przeprowadzenie testów odsłuchowych. Przeprowadzono test porównań parami, aby uzyskać informacje o preferencjach odnośnie występowania w nagraniach ambisonicznych dodatkowego podkładu stereo oraz jego poziomu głośności. W drugim z testów zbadano jaki wpływ na odbiór nagrania ambisonicznego ma sposób oraz stopień jego przetwarzania. Wnioski otrzymane poprzez analizę wyników obu testów posłużyły jako wskazówki przy postprodukcji nagrań. Otrzymane materiały zostały zamieszczone na interaktywnej mapie w aplikacji internetowej.

1. WPROWADZENIE

Pojęcie systemu ambisonicznego po raz pierwszy pojawiło się w latach siedemdziesiątych ubiegłego wieku. W roku 1974, w magazynie *Studio Sound* ukazał się artykuł zatytułowany *What's wrong with quadrophonics?*, w którym Michael Gerzon wskazał wady popularnego wówczas systemu kwadrofonicznego. Wśród nich znalazły się między innymi problemy z odtwarzaniem lokalizacji poszczególnych źródeł dźwięku, stosunkowo mała powierzchnia dobrego odsłuchu czy podatność na zmiany tego obszaru, który był silnie uzależniony od pozycji słuchacza. W tym samym artykule Gerzon przedstawił również alternatywną koncepcję dla kwadrofonii, która dziś znana jest jako system ambisoniczny. Pomimo pracy włożonej przez Gerzona i jego współpracowników w rozwój ambisonii oraz jej zalet względem systemów kwadrofonicznych, nie przyjęła się ona wówczas w przemyśle nagraniowym [6] [11].

Jednak postępujący rozwój technologiczny oraz powszechna dostępność materiałów audiowizualnych w Internecie sprawiły, że nagrania ambisoniczne znalazły swoje miejsce w wydawałoby się dostatecznie złożonym i różnorodnym globalnym zbiorze multimedialnych. Szczególnie w ostatniej dekadzie nastąpił wzrost zainteresowania tym formatem dźwięku przestrzennego o czym świadczą jego implementacje w filmach 360 oraz rzeczywistości wirtualnej w serwisie społecznościowym *Facebook* oraz na platformie *YouTube*. W obu przypadkach określone zostały również wytyczne dotyczące publikowania materiałów zawierających dźwięk w postaci ambisonicznej odnośnie wspieranych formatów oraz wymaganych dodatkowych metadanych [3] [16].

W niniejszym artykule opisany został projekt, w którym za cel postawiono stworzenie ambisonicznej mapy Trójmiasta w formie aplikacji internetowej. Część badawcza pracy polegała na porównaniu dostępnych algorytmów miksowania sygnałów ambisonicznych. W tym celu przeprowadzono dwa testy odsłuchowe, których wyniki zostały następnie przeanalizowane i posłużyły jako wskazówki w dalszym procesie obróbki nagrań. Otrzymane materiały zostały zamieszczone w aplikacji internetowej z interaktywną mapą, dostępną online.

2. REALIZACJA NAGRAŃ

Wybór urządzeń do przeprowadzenia nagrań podyktowany był mobilnością sprzętu, przy zachowaniu dostatecznie wysokiej jakości materiałów audio i wideo. Podczas realizacji skorzystano z mikrofonu ambisonicznego pierwszego rzędu *ZOOM H3-VR* z wbudowanym rejestratorem oraz z kamery *Insta360 ONE X*, które przedstawiono na rysunku 1. Oba urządzenia w czasie nagrań mocowano w jednej osi na dwustronnym uchwycie, kamera była skierowana do góry, a mikrofon do podłoża. Natomiast sam uchwyt montowany był na statywie.



Rys. 1. Mikrofon ambisoniczny pierwszego rzędu - ZOOM H3-VR oraz kamera 360 – Insta360 ONE X [9] [21]

W przypadku kamery zdecydowano się na ustawienie rozdzielczości 5.7K przy 30 klatkach na sekundę. Dodatkowo włączony został również tryb HDR, aby zwiększyć zakres dynamiki obrazu w nagraniu i uniknąć potencjalnych prześwietleń np. podczas rejestracji na pełnym słońcu. Natomiast dla sygnałów audio jako format wyjściowy ustawiono B-format przy częstotliwości próbkowania 96kHz i rozdzielczości 24 bity. Automatyczna konwersja rejestrowanego nagrania do odpowiedniej postaci, uwzględniającej orientację mikrofonu, była możliwa dzięki wbudowanemu w mikrofon żyroskopowi. Podczas nagrań korzystano również z możliwości monitorowania sygnału w czasie rzeczywistym poprzez wyjście słuchawkowe urządzenia, co pozwoliło na kalibrowanie sygnału [9] [21].

Materiały zarejestrowano w dwunastu wybranych miejscach, które uznano za charakterystyczne dla krajobrazu Trójmiasta:

- Gdynia: Port Gdynia (przy Muzeum Emigracji), Skwer Kościuszki, Bulwar Nadmorski im. Feliksa Nowowiejskiego, Gdynia Orłowo,
- Sopot: Park Północny, Molo, ul. Bohaterów Monte Cassino,
- Gdańsk: Park Oliwski, Nowy Port - Kapitanat Portu Gdańsk, Westerplatte, Długi Targ, Stogi - Plaża.

Aby w pełni oddać klimat danego miejsca przy jak najmniejszych zakłóceniach spowodowanych potencjalną obecnością ludzi i wzmożonym ruchem ulicznym, nagrania realizowano w godzinach porannych. W efekcie otrzymano szesnaście zestawów nagrań dźwiękowych oraz wideo zarejestrowanych we wszystkich wybranych wcześniej lokalizacjach.

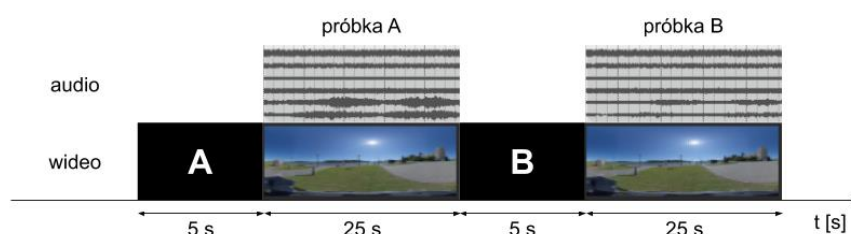
3. TEST PORÓWNAŃ PARAMI

Serwisem, w którym umieszczone miały zostać docelowe nagrania była platforma YouTube. Poza standardową opcją przesyłania plików wideo ze ścieżką audio mono lub stereo, serwis ten umożliwia również publikowanie materiałów 360 z dźwiękiem przestrzennym. Obsługiwane są dwa następujące formaty: ambisonia pierwszego rzędu oraz ambisonia pierwszego rzędu z dźwiękiem stereo. Przy czym drugi z wymienionych formatów charakteryzuje się stałą pozycją ścieżki stereo, bez względu na ruch głowy słuchacza w danej chwili. Z uwagi na tę możliwość zdecydowano się na przeprowadzenie testu

odsluchowego, mającego na celu sprawdzenie preferencji użytkowników odnośnie występowania podkładu stereo oraz poziomu jego głośności w materiale 360 z dźwiękiem przestrzennym. Hipotezą postawioną w teście było założenie, że potencjalny odbiorca nagrania ambisonicznego, zarejestrowanego w przestrzeni publicznej, preferuje wersję pozbawioną dodatkowego podkładu muzycznego [16].

Jako formułę testu wybrano metodę porównań parami, jednak zamiast porównywania jakości próbek przed uczestnikami postawiono pytanie, która z prezentowanych próbek jest lepszej jakości. Dodatkowo zastosowano również powtórzną rundę z zamienioną kolejnością próbek w celu zweryfikowania ocen testujących. Było to podyktowane założeniem, że uczestnikami testu będą osoby niebędące ekspertami. Również z tego względu licznosc grupy testowej określono na dwadzieścia osób [2].

Zdecydowano się na przygotowanie sześciu próbek audiowizualnych, z których utworzono osiem par, w tym cztery z zamienioną kolejnością. Każda para prezentowana była w układzie przedstawionym na rysunku 2. Z zarejestrowanych wcześniej nagrań utworzono po cztery próbki testowe dla dwóch wybranych miejsc o różnych poziomach głośności podkładu muzycznego, które zostały oznaczone literami ABCD oraz numerem miejsca (1 lub 2). Informacje o poszczególnych próbkach oraz dobranych parach przedstawiono w tabelach 1 oraz 2.



Rys. 2. Struktura prezentacji próbek w przeprowadzonym teście preferencji dwójkowych

Tab. 1. Próbkı użyte w teście porównań parami

Próbka	Podkład stereo	Miejsce nagrania
A1	brak podkładu	Gdynia – Bulwar Nadmorski im. Feliksa Nowowiejskiego
B1	z podkładem	
C1	z podkładem -6 dB	
A2	brak podkładu	Gdańsk – Kapitanat Portu Gdańsk
B2	z podkładem -3 dB	
C2	z podkładem -6 dB	

Tab. 2. Pary próbek testowych dla testu porównań parami

Nr pary	1	2	3	4	5	6	7	8
Próbka A	A1	A2	B1	B2	C1	A2	B1	C2
Próbka B	B1	C2	C1	A2	B1	B2	A1	A2

Ze względu na brak dedykowanej platformy do przeprowadzania oceny subiektywnej materiałów 360, test został utworzony przy pomocy aplikacji *Google Forms*, pozwalającej na stworzenie formularza, który może zawierać materiały wideo zamieszczone wcześniej w serwisie *YouTube*. Test podzielono na dziesięć sekcji. Pierwsza z nich, traktowana jako sekcja wstępna, zawierała opis przebiegu testu oraz dodatkowe informacje dotyczące konieczności odsluchu za pośrednictwem słuchawek, obracania obrazem podczas prezentacji próbek oraz ograniczeń *Google Forms* związanych z brakiem możliwości odtwarzania próbek testowych w trybie pełnego ekranu. Kolejne osiem sekcji przeznaczono do prezentacji par próbek testowych, z czego w każdej z nich znajdował się odtwarzacz *YouTube* oraz pytanie jednokrotnego wyboru dotyczące preferowanej przez słuchacza wersji. Na końcu formularza zamieszczono również ankietę z pytaniami dotyczącymi doświadczeń uczestników testu w zakresie realizacji i

odbioru nagrań 360. Łączny czas trwania testu dla jednej osoby wynosił 8 minut. Między obiema częściami (rundami) nie stosowano przerwy.

W teście udział wzięło łącznie dwadzieścia jeden osób. Zgodnie z informacjami udzielonymi przez respondentów w ankiecie umieszczonej na końcu testu, jedynie sześciu z nich miało wcześniej doświadczenia w realizacji materiałów audiowizualnych, w tym tylko jedna osoba przy rejestracji wideo w technologii 360. Można więc uznać, że grupa testowa składała się z osób niebędących ekspertami. W tabeli 3 przedstawiono zsumowane odpowiedzi wszystkich uczestników w każdej prezentowanej parze, przy czym każda z próbek mogła otrzymać maksymalnie 21 głosów w kontekście jednej pary.

Tab. 3. Wyniki przeprowadzonego testu porównań parami

I część testu								II część testu							
numer pary															
1		2		3		4		5		6		7		8	
A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
A1	B1	A2	C2	B1	C1	B2	A2	C1	B1	A2	B2	B1	A1	C2	A2
13/2	8/2	14/2	7/2	7/2	14/2	7/2	14/2	13/2	8/2	16/2	5/2	7/2	14/2	9/2	12/2
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Na podstawie oceny stabilności odpowiedzi każdego z uczestników z dalszej analizy wykluczono wyniki osób, które udzieliły odpowiedzi zgodnych w mniej niż trzech przypadkach. Następnie dla tak ograniczonego zbioru obserwacji wyznaczono liczbę odpowiedzi oddanych na każdą z próbek testowych, zarówno osobno dla obu części testu, jak i dla całego testu.

Z zaprezentowanych w teście par można wyodrębnić ich dwa rodzaje: pary, w których należało dokonać wyboru pomiędzy próbką bez podkładu muzycznego a próbką, w której podkład występuje oraz pary zawierające obie próbki z podkładem muzycznym o różnym poziomie głośności. W pierwszym z wymienionych rodzajów par uczestnicy wybierali wersję bez podkładu średnio w 14 na 21 przypadkach. W takim samym stosunku wybór ekspertów padał na próbki zawierające cichszą wersję podkładu dla par, w których nie występowała próbka bez podkładu.

Korzystając z otrzymanych danych wyznaczono statystykę chi kwadrat, porównującą wyniki obu części testu. Na jej podstawie można określić, czy istnieje istotna różnica między odpowiedziami uczestników w obu częściach testu. Wartość chi kwadrat obliczono według wzoru [10]:

$$X^2 = n \cdot \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i \cdot} \cdot n_{\cdot j}}{n}\right)^2}{n_{i \cdot} \cdot n_{\cdot j}} = 0,735 \quad (1)$$

gdzie:

r – liczba części testu,

s – liczba badanych obiektów,

n_{ij} – liczba wyborów j-tego obiektu w i-tej części testu,

$n_{i \cdot}$ – suma wyborów w i-tej części testu,

$n_{\cdot j}$ – suma wyborów j-tego obiektu we wszystkich częściach testu,

n – suma wszystkich wyborów podczas testu.

Uzyskana wartość statystyki chi kwadrat udowadnia, że przyjęto odpowiedni minimalny poziom zgodnych odpowiedzi udzielonych przez osoby predystynowane do bycia ekspertem równy 75%, a tym samym potwierdza właściwy dobór ekspertów do dalszej analizy wyników.

Statystyczną zgodność wyników określono poprzez porównanie obliczonej wartości chi kwadrat z wartością krytyczną, odczytywaną z tablic rozkładu tej statystyki dla wyznaczonej liczby stopni swobody oraz obranego poziomu istotności. Dla przyjętego poziomu istotności równego 0,05 oraz wyznaczonej liczby stopni swobody odczytano wartość krytyczną statystyki chi kwadrat równą 11,0705. Obliczona wartość statystyki chi kwadrat nie przekracza wartości krytycznej, zatem można potwierdzić hipotezę o zgodności wyników. Następnie określono istotności różnic w wynikach dla poszczególnych par w obu turach, które zostały wyliczone według wzoru [10]:

$$z_{ij} = \frac{|p_i - p_j|}{\sqrt{\frac{(p_i + p_j) \cdot (2 - p_i - p_j)}{2 \cdot N}}} \quad (2)$$

gdzie:

z_{ij} – określone prawdopodobieństwo,

p_i – względna liczba głosów na i -te obiekty,

p_j – względna liczba głosów na j -te obiekty,

N – maksymalna liczba głosów, którą może otrzymać jeden obiekt,

n – liczba porównywanych obiektów,

m – liczebność grupy ekspertów.

Otrzymane wartości porównano z wartością graniczną z tablic rozkładu normalnego, która dla przyjętego poziomu istotności równego $\alpha = 0,05$ wynosi $z(\alpha) = 1,96$. Jeżeli wartość z_{ij} jest mniejsza od $z(\alpha)$, to nie istnieją podstawy do odrzucenia hipotezy dotyczącej braku istotnej różnicy między próbkami porównywanymi w danej parze (znak „-”). Natomiast w sytuacji odwrotnej należy uznać, że różnica ta jest statystycznie istotna (znak „+”). W tabeli 4 przedstawiono wartości prawdopodobieństwa z_{ij} oraz wynik porównania z wartością $z(\alpha)$ [10].

Tab. 4. Istotności różnic pomiędzy próbkami w danych parach

	pary próbek			
	1, 7	2, 8	3, 5	4, 6
z_{ij}	0,128	0,107	0,149	0,149
$z(\alpha) > z_{ij}$	—	—	—	—

Z poczynionych obliczeń jednoznacznie wynika, że nie ma podstaw do odrzucenia hipotezy o braku istotnej różnicy między porównywanymi obiektami w danej parze. Uzyskane wyniki można więc uznać za istotne i stabilne. Przeprowadzona analiza potwierdziła założoną wcześniej hipotezę, zatem można uznać, że docelowy odbiorca nagrania ambisonicznego, zrealizowanego w przestrzeni publicznej, będzie preferował wersję „naturalną”, bez podkładu muzycznego.

4. TEST PARAMETRYCZNY

Ze względu na wielość dostępnych wtyczek przeznaczonych do postprodukcji sygnałów ambisonicznych, zdecydowano się na przeprowadzenie testu parametrycznego, który miał na celu uzyskanie wytycznych dotyczących wartości parametrów poszczególnych wtyczek oraz ich potencjalnej przydatności podczas obróbki materiału w procesie tworzenia docelowej bazy nagrań, co miało zostać sprawdzone poprzez przetestowanie wpływu różnych ustawień na odbiór materiału przez słuchaczy.

Podczas wyboru testowanych wtyczek kierowano się charakterem zarejestrowanych nagrań. Z uwagi na to, że materiały były realizowane w różnych warunkach i pojawiły się w nich niepożądane składowe o różnych zakresach częstotliwości, zdecydowano się na przetestowanie działania korektora barw w zakresie niskich i wysokich częstotliwości, a także kompresora wielopasmowego w zakresie niskich częstotliwości, celem sprawdzenia wpływu na odbiór materiału przy zminimalizowaniu występowania częstotliwości odpowiadających powiewom wiatru oraz powstających niepożądanych składowych o wysokich częstotliwościach. Przetestowano również działanie obu powyższych wtyczek przy wzmacnianiu charakterystycznych dźwięków otoczenia, znajdujących się w zakresie średnich częstotliwości. Postanowiono także w ramach eksperymentu sprawdzić jak na odbiór nagrania wpłynie wyrównanie poziomów dla kanałów w sygnale ambisonicznym. W tym celu przetestowano dwie wtyczki wyrównujące wzmocnienia kanałów ortogonalnych względem kanału dookólnego. Materiały realizowano rów-

nież w miejscach, w których znajdował się bodziec dźwiękowy, charakterystyczny dla danego otoczenia. Dlatego też przetestowano wpływ działania wtyczek wzmacniających dany kierunek lub przestrzeń w nagraniu.

Analogicznie jak w przypadku testu porównań parami, z uwagi na docelowe przeznaczenie materiałów, założono, że w teście będą uczestniczyć osoby nie będące ekspertami. Minimalną liczebność grupy testowej określono na 35 osób. Formułę testu zdefiniowano kierując się wymaganiami stosowanymi w podejściu typu MUSHRA [14]. Na ich podstawie przyjęto długość próbek testowych równą 10 sekund. Zadaniem uczestników testu w przypadku każdej z próbek była jej subiektywna ocena w skali od 1 do 10, gdzie wartość 10 traktowana była jako najwyższa możliwa ocena.

Do stworzenia formularza testowego ponownie posłużyła platforma *Google Forms*. Dla każdego z testowanych parametrów wyodrębniono osobną sekcję prezentującą przypisane do niej próbki. Ponieważ serwis ten nie służy docelowo do przeprowadzania odsłuchowych testów parametrycznych, nie było możliwe spełnienie warunków sposobu prezentacji próbek w formie dowolnego przełączania się między nimi oraz oceny próbek w skali ciągłej, co wymagane jest w przypadku testów typu MUSHRA. Z tego powodu zdecydowano, aby dla każdej próbki przeznaczyć osobną instancję odtwarzacza *You-Tube* oraz pytanie o jej ocenę. Podobnie jak w formularzu przygotowanego dla testu porównań parami, na jego początku zamieszczono sekcję wstępną zawierającą informacje dotyczące przebiegu testu [14].

Na potrzeby testu stworzono łącznie 41 próbek testowych. Mając na uwadze wyniki testu porównań parami, zrezygnowano z dodatkowego podkładu muzycznego. Rolę kotwicy w każdym z prezentowanych zestawów próbek stanowił sygnał niepoddany działaniu wtyczki w badanym aspekcie.

W teście udział wzięło łącznie trzydzieści pięć osób. Spełnione zostało zatem założenie dotyczące liczebności grupy testowej. W pierwszej kolejności wyłoniono grupę ekspertów, czego dokonano na podstawie wariancji obliczonej dla każdego uczestnika w kontekście całego testu. Z dalszej analizy wykluczono wyniki sześciu osób, dla których wariancja wyniosła poniżej założonego poziomu równego 0,75. Wyodrębniono w ten sposób grupę ekspertów liczącą 29 osób.

Następnie dla otrzymanej grupy ekspertów wygenerowano ponownie histogramy rozkładu ocen poszczególnych próbek testowych oraz obliczono dla nich statystyki opisowe. W każdym z badanych przypadków wartości średniej arytmetycznej oraz mediany dla poszczególnych próbek były zbliżone do siebie. Największe różnice między tymi statystykami zauważono dla próbek bez przetwarzania w sekcjach nr 4, 5, 6 oraz 9, które oscylowały w granicach od 0,7 do 0,9, co świadczy o symetryczności rozkładu ocen względem jego środka [8].

Natomiast analizując poziomy wariancji, będącej miarą rozproszenia wyników względem wartości średniej, najwyższe wartości zaobserwowano dla sekcji, w których testowane było działanie wtyczek uwydatniających wybrane kierunki lub przestrzenie w nagraniu. Ponadto porównano wartości wariancji ocen ekspertów z wartościami uzyskanymi dla wszystkich uczestników testu. Po wyodrębnieniu grupy ekspertów dla ocen próbek testowych znajdujących się w pierwszych siedmiu sekcji testowych wartość wariancji wzrosła. Jednak w tych sekcjach, w których badano wtyczki uwydatniające wybrane kierunki w nagraniu, miara rozproszenia nieznacznie spadła w ośmiu na dwanaście przypadków badanych w tych sekcjach. Takie zależności pomiędzy wartościami statystyki wariancji otrzymanymi dla obu grup potwierdzają prawidłowy dobór ekspertów [8].

Z uwagi na występowanie miar porządkowych oraz fakt, że otrzymane rozkłady odpowiedzi nie były rozkładami normalnymi, zdecydowano się na przeprowadzenie testu Kruskala-Wallisa dla wyników ekspertów. Hipotezą zerową w tym teście jest założenie, że średnie rangi lub mediany dwóch serii danych są takie same. Przeciwny przypadek stanowi zaś hipotezę alternatywną. Jako hipotezę do przeprowadzanego testu przyjęto domyślną hipotezę zerową. W pierwszej kolejności obliczono wartość statystyki T dla całego zbioru odpowiedzi według następującego wzoru [15]:

$$T = \frac{(N-1)(\sum_{i=1}^k \frac{s_i^2}{n_i} - C)}{s_r - C} = 49,872 \quad (3)$$

gdzie:

N – liczba elementów zbioru,
 k – liczba grup, na które został podzielony zbiór,
 s_i – suma rang elementów grupy i ,
 n_i – liczebność grupy i ,
 S_r – suma kwadratów rang wszystkich elementów zbioru,
 C – wartość korekty dla średniej.

Hipotezę zerową testu sprawdzono poprzez porównanie obliczonej statystyki T z wartością krytyczną chi kwadrat, odczytaną z tablic rozkładu chi kwadrat dla wyznaczonej liczby stopni swobody równej 40 oraz obranego poziomu istotności – 0,05. Dla tych wartości odczytano poziom krytyczny statystyki chi kwadrat równy 55,7585, zatem nie można odrzucić hipotezy zerowej, ponieważ wyznaczona wartość statystyki T nie przekroczyła wartości krytycznej [15].

W dalszej analizie sekcje testu zostały podzielone na grupy według przeznaczenia przetwarzania danych wtyczek. W celu sprawdzenia hipotezy zerowej dla każdej z grup wyznaczono wartość statystyki T oraz liczbę stopni swobody, dla których odczytano wartość krytyczną statystyki chi kwadrat. Według otrzymanych obliczeń nie istnieją podstawy do odrzucenia hipotezy zerowej, ponieważ dla żadnej z wydzielonych grup wartość statystyki T nie przekroczyła wartości krytycznej chi kwadrat odczytanej dla założonego poziomu istotności równego 0,05 oraz określonej liczby stopni swobody.

Kolejne kroki analizy przeprowadzono przy użyciu skryptu w języku R oraz pakietu *fBasics* zawierającego zbiór funkcji do analizy danych, w tym również funkcję dedykowaną dla testu Kruskala-Walisa. Serie odpowiedzi wyłonionych ekspertów, stanowiące dane wejściowe, zapisano w pliku CSV. Podając kolejne wektory odpowiedzi do funkcji w kombinacji „każdy z każdym”, otrzymano macierz statystyki chi kwadrat oraz macierz p-wartości dla całego testu [20].

Następnie z otrzymanych macierzy wyodrębniono zbiory wartości statystyk dla każdej z grup. Użyte wartości statystyki chi kwadrat dla wszystkich badanych próbek w kolejnych grupach są mniejsze od wyznaczonej dla każdej z nich wartości krytycznej, co świadczy o istotności różnic między udzielanymi odpowiedziami. Analizując obliczone p-wartości można uznać, że nie ma podstaw do odrzucenia hipotezy zerowej, ponieważ wszystkie są większe od założonego poziomu istotności [8].

Dodatkowo porównano sumy ocen oddanych na próbki testowe w celu uzyskania informacji o preferencjach wyznaczonej grupy ekspertów dotyczących charakteru oraz stopnia przetworzenia dźwięku ambisonicznego w prezentowanych przykładach. Przygotowane zestawienie nie wskazało jednoznacznie faworytów wśród nagrań ocenianych przez ankietowanych. Z tego powodu sumy ocen porównano ponownie, jednak tym razem pod uwagę wzięto jedynie oceny powyżej średniej arytmetycznej dla danego obiektu testowego ocenianego przez grupę wyznaczonych ekspertów. Na podstawie wyników dla każdej sekcji wyznaczono najwyższe oceniane próbki oraz odpowiadające im wartości testowanych parametrów, które zamieszczono w tabeli 5.

Tab. 5. Preferowane próbki oraz przypisane do nich wartości testowanych parametrów

nr sekcji	nr próbki	testowany parametr	wartość parametru
1	1 lub 3	częstotliwość odcięcia filtra górnoprzepustowego	bez lub 120Hz
2	3	poziom tłumienia górnopasmowego korektora półkowego lub filtra pasmowo-zaporowego ustawionego w górnym paśmie	korektor półkowy: -12dB
3	1 lub 4	poziom wzmocnienia filtra pasmowo-zaporowego ustawionego w środkowym paśmie	bez lub 4dB
4	1	stopień kompresji	bez
5	1 lub 4	stopień kompresji	bez lub 8:1
6	1 lub 4	stopień kompresji	bez lub 8:1
7	3	parametr Envelopment	0,27
8	3	stopień kompresji	4:1
9	2	wzmocnienie	2dB
10	1	stopień przekształcenia	bez

Wyniki analizy przeprowadzonego testu są w wielu aspektach wyraźnie niejednoznaczne. Przyczyn takiego stanu rzeczy należałoby szukać w kilku czynnikach. Pierwszym z nich jest niewątpliwie nieodpowiedni dobór uczestników, o czym świadczą m.in. wartości wariancji uzyskane przez uczestników w poszczególnych sekcjach. W przypadku przeprowadzania testu, w którym oceniane próbki przetworzone zostały w stopniu średnim lub znacznym, rekomenduje się dobranie grupy testowej składającej się z osób, które można określić mianem ekspertów. Jednak z uwagi na ograniczenia występujące podczas realizacji niniejszej pracy dyplomowej możliwe było zorganizowanie testu wyłącznie w formie zdalnej, co niosło za sobą dodatkowe komplikacje w postaci braku kontroli nad przebiegiem sesji odsłuchowych oraz dysproporcji w odbiorze próbek testowych ze względu na zróżnicowany sprzęt, z którego korzystali uczestnicy. W związku z tym wnioski otrzymane podczas przeprowadzonej analizy wyników należy traktować raczej jako punkt wyjścia do dalszych rozważań na temat przetwarzania nagrań ambisonicznych niż jako konkretne wskazówki dotyczące tego zagadnienia [14].

5. STWORZENIE BAZY NAGRAŃ ORAZ APLIKACJI

Dla każdego z miejsc, w którym zarejestrowano materiały otrzymano zestaw dwóch nagrań: wideo oraz audio, odpowiednio w formatach INSV i WAV. Należy jednak zaznaczyć, że pliki z rozszerzeniem INSV nie są domyślnie wspierane przez programy służące do postprodukcji materiałów wideo. Rozwiązaniem tego problemu okazała się być wtyczka dostępna w pakiecie z darmowym oprogramowaniem *Insta360 Studio* udostępnianym przez producenta kamery, która jest przeznaczona do użycia wyłącznie we współpracy z oprogramowaniem *Adobe Premiere Pro* w wersjach 2018, 2019 lub 2020. Dzięki niej możliwa jest obsługa oraz edycja plików w formacie INSV. W związku z ograniczoną możliwością wyboru programu do edycji materiałów wideo, zdecydowano się na użycie *Adobe Premiere Pro* w wersji 2019 [9] [12].

Po zapoznaniu się z zarejestrowanymi materiałami wideo uznano, że poza standardową obróbką w postaci przycięcia ścieżek do docelowej długości, konieczna będzie ich dodatkowa edycja ze względu na przechylenia obrazu względem podłoża oraz na widoczny w kadrze statyw oraz mikrofon. W celu wypoziomowania obrazu użyto wtyczki *VR Sphere Rotation* dostępnej w ramach oprogramowania *Adobe Premiere Pro*. Przykład porównania przed i po wprowadzonej korekcie zaprezentowano na rysunku 3. Natomiast znajdujące się w kadrze statyw oraz mikrofon zasłonięto, nakładając na obraz maskę z logiem Katedry Systemów Multimedialnych. Materiały wideo wyeksportowano bez ścieżki dźwiękowej w formacie MPEG-4 dla parametrów jednakowych z parametrami zarejestrowanego nagrania [12].



Rys. 3. Efekt wypoziomowania obrazu za pomocą wtyczki *VR Sphere Rotation*

Do postprodukcji nagrań audio wybrano cyfrową stację roboczą *Reaper* w wersji 6.02, która pozwala na miksowanie oraz eksportowanie sygnałów wielokanałowych. Dla każdego z projektów, w którym przeprowadzono obróbkę dźwięku, ustawione zostały parametry tożsame z parametrami zarejestrowanych sygnałów [13].

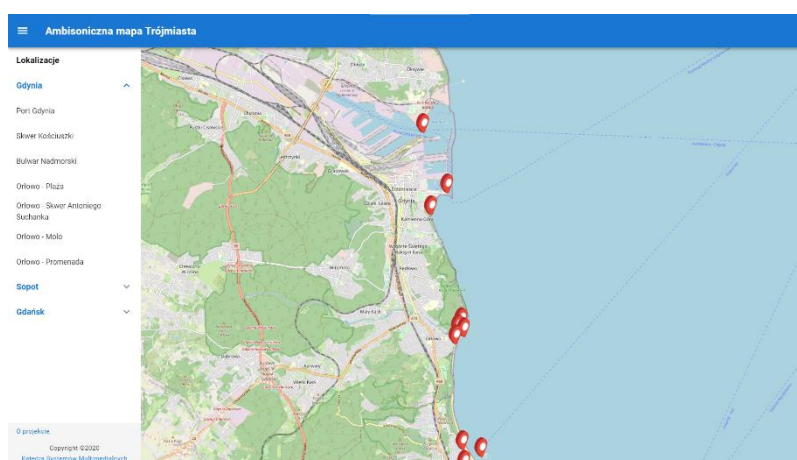
Uwzględniając wyniki testu porównań parami zrezygnowano z dodatkowego podkładu stereo, więc

routing kanałów został ustawiony tak, aby zarówno dla kanału głównego jak i pozostałych sygnał wyjściowy był w postaci czterokanałowej. Z uwagi na fakt, że nagrania zostały zarejestrowane w ambisonicznym B-formacie, nie było konieczne użycie enkodera. Monitorowanie sygnału poprzez słuchawki umożliwiło ustawienie na kanale głównym wtyczki dekodującej sygnał do postaci binauralnej [1] [5].

Ze względu na niejednoznaczne wyniki testu parametrycznego, zdecydowano się jedynie na korekcję częstotliwości w dolnym oraz górnym paśmie, sugerując się wariantami, które były najlepiej oceniane przez ekspertów. Pliki dźwiękowe zostały wyeksportowane w formacie WAV dla ustawień projektu.

Przygotowane w ten sposób pliki audio oraz wideo połączono w kontenerze MPEG-4 za pomocą kodeka FFMPEG. Następnie przy użyciu narzędzia *Spatial Media Metadata Injector* dodano również wymagane metadane, zgodnie z wymaganiami platformy YouTube [4] [7] [16].

Ostatnim etapem projektu było stworzenie aplikacji internetowej z interaktywną mapą, zawierającą przygotowane wcześniej nagrania. W tym celu wykorzystano pakiet *Vue.js* i kompatybilne z nim biblioteki komponentów: *Vuetify* oraz *VueLayers*. Finalny wygląd aplikacji, która została umieszczona na serwerze Katedry Systemów Multimedialnych, przedstawiono na rysunku 4 [17] [18] [19].



Rys. 4. Aplikacja internetowa z interaktywną mapą

6. WNIOSKI

W ramach niniejszego projektu zrealizowano serię materiałów audiowizualnych dla wybranych lokalizacji Trójmiasta. Nagrania zostały poddane procesowi postprodukcji i zamieszczone na interaktywnej mapie w aplikacji internetowej. Istnieje przy tym możliwość łatwej rozbudowy mapy o nowe nagrania, wykonywane w kolejnych miejscach w Trójmieście.

Przeprowadzony test porównań parami wykazał, że docelowi odbiorcy materiałów zdecydowanie preferują nagrania pozbawione dodatkowego podkładu muzycznego. Natomiast w przypadku drugiego testu wnioski pochodzące z analizy jego wyników są niejednoznaczne. Prawdopodobnym tego powodem były ograniczenia związane ze zdalną formą testu, co wpłynęło bezpośrednio na brak kontroli nad przebiegiem sesji odsłuchowych oraz na dysproporcje w odbiorze próbek testowych przez poszczególnych uczestników.

LITERATURA

- [1] 5 Tips for Mixing Ambisonics | 360° | VR | Spatial Audio | Part 5/7 | Berklee Online, <https://www.youtube.com/watch?v=1dvK4ojh2b0>, (data dostępu: 05.10.2020).
- [2] ANSTENDIG M. B., AB Testing, *A misapplication of visual criteria in audio*, 2006. <http://www.anstendig.org/ABTesting.html>, (data dostępu: 05.10.2020).

- [3] Facebook 360 Video, <https://facebook360.fb.com/>, (data dostępu: 05.10.2020).
- [4] FFmpeg Codecs Documentation, <https://www.ffmpeg.org/ffmpeg-codecs.html>, (data dostępu: 05.10.2020).
- [5] FRANK M., ZOTTER F.: *Ambisonics – A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality*, Springer Topics in Signal Processing 2019.
- [6] GERZON M.: *What's wrong with quadrophonics*, Studio Sound, May 1974, [https://www.audiosignal.co.uk/Resources/What is wrong with quadraphonics A4.pdf](https://www.audiosignal.co.uk/Resources/What_is_wrong_with_quadraphonics_A4.pdf), (data dostępu: 05.10.2020).
- [7] google/spatial-media: Specifications and tools for 360° video and spatial audio., <https://github.com/google/spatial-media>, (data dostępu: 05.10.2020).
- [8] HEUMANN C., SCHOMAKER M., SHALABH, *Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R*, Springer International Publishing, Switzerland, 2016.
- [9] Insta360 ONE X - Own the moment., <https://www.insta360.com/product/insta360-onex>, (data dostępu 05.10.2020).
- [10] KOSTEK B., Wykład z przedmiotu Technologia nagrań – Testy subiektywne, Politechnika Gdańska – Wydział Elektroniki, Telekomunikacji i Informatyki – Katedra Systemów Multimedialnych, https://multi-med.org/student/tm/testy_subiektywne.pdf, (data dostępu: 05.10.2020).
- [11] Michael Gerzon Audio Pioneer – Ambisonics, <https://www.michaelgerzonphotos.org.uk/ambisonics.html>, (data dostępu: 05.10.2020).
- [12] Profesjonalny edytor wideo | Adobe Premiere Pro, <https://www.adobe.com/pl/products/premiere.html>, (data dostępu: 05.10.2020).
- [13] REAPER | Audio Production Without Limits, <https://www.reaper.fm/>, (data dostępu: 05.10.2020).
- [14] Recommendation ITU-R BS.1534-1 – Method for the subjective assessment of intermediate quality level of coding systems, https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-1-200301-S!!PDF-E.pdf, (data dostępu: 05.10.2020).
- [15] SPRENT P., SMEETON N.C., *Applied Nonparametric Statistical Methods, Third Edition*, Champan & Hall/CRC, 2001.
- [16] Używanie dźwięku przestrzennego w filmach 360° i rzeczywistości wirtualnej (VR), <https://support.google.com/youtube/answer/6395969>, (data dostępu: 05.10.2020).
- [17] Vue.js, <https://vuejs.org/>, (data dostępu: 05.10.2020).
- [18] Vuetify, <https://vuetifyjs.com/en/>, (data dostępu: 05.10.2020).
- [19] VueLayers :: Homepage, <https://vuelayers.github.io/>, (data dostępu: 05.10.2020).
- [20] WUERTZ D., SETZ T., CHALABI Y., MAECHLER M., Package fBasics, Rmetrics – Markets and Basic Statistics, 2020.
- [21] ZOOM H3-VR Handy Recorder | Zoom, <https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder>, (data dostępu: 05.10.2020).

AN AMBISONIC MAP OF SPECIFIC PLACES IN THE TRI-CITY

The project aimed to create an ambisonic map of the Tri-City in the form of a web application. The video materials were recorded in 360 technology with sound in the form of an ambisonic signal in locations of the Tri-City, which were considered to be characteristic for this agglomeration. The research objective of the study was to compare the available algorithms for mixing ambisonic signals by conducting listening tests. The pairwise comparison test was performed to obtain information about the preferences regarding the presence of additional head-locked stereo background music and its level. In the second test the impact on the reception of the ambisonic recording was examined in terms of the method and degree of its processing. The conclusions obtained by analyzing the results of both tests served as guidelines for the post-production of the recordings. The prepared materials were placed on an interactive map in the web application.

PRACTICAL IMPLEMENTATION OF CONSTANT BEAMWIDTH TRANSDUCER (CBT) FOR STUDIO MONITOR DESIGN

Tomasz NOWAK¹

Bartłomiej KRUK¹

¹ Wrocław University of Science and Technology, Faculty of Electronics, Department of Acoustics and Multimedia, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

T.Nowak@pwr.edu.pl

Bartlomiej.Kruk@pwr.edu.pl

The first research on the Constant Beamwidth Transducer (CBT) dates back to the 1980s and involved military research on underwater transducers. Later Don Keele applied this theory to speakers, presenting the results in AES journals.

This paper presents the advantages of using CBT to build monitors. Like a linear source, it has the same advantage in reducing spl losses over a distance of 3dB per double the distance compared to 6dB for a point source. The second important advantage is the achievement of a larger listening area with a much more consistent SPL compared to a point source.

1. INTRODUCTION AND BACKGROUND

CBT ORIGINS

Research on CBT describes a curved surface transducer in the form of a spherical cap with frequency-independent Legendre shading. Shading is critical to achieve a broadband constant directionality that provides extremely consistent beam width and directional behaviour with virtually no side lobes.

This theory, extended by the use of circular arc-shaped speaker arrays, has been studied by Keele since 2000 [2], [3], [4].

CONCEPT OF OPERATION

CBT theory dictates that each transducer in the board should be driven at different levels that are compatible with Legendre's continuous shading function. The Legendre shading function gradually reduces the levels of each transducer. In this example, the two lowest speakers in the array were given full power, the next two levels reduced by 3 dB and the highest driver reduced by 6dB. Stepped gain changes were realized passively using voltage dividers and by changing the impedance of the transducer (a combination of serial-parallel connections).

The big advantage of shading is that it can be done without digital signal processing that is, it can be done passively.

ADVANTAGES OVER POINT SOURCE

The sound pressure level attenuation formula for point source is as follows:

$$\text{SPL}_2 = \text{SPL}_1 - 20 * \log (R_2 / R_1) \quad (1)$$

Where:

SPL_1 is the Sound Pressure Level at point 1,

SPL_2 is the Sound Pressure Level at point 2,

R_1 is the distance from the sound source to point 1, and

R_2 is the distance from the sound source to point 2.

Attenuation formula for line source:

$$\text{SPL}_2 = \text{SPL}_1 - 10 * \log (R_2 / R_1) \quad (2)$$

The above formulas indicate a 3 dB of attenuation per doubling of distance from the line source (within the critical distance) compared to 6 dB in the case of a point source. For a curved array and CBT design, even at a greater than critical distance, the apparent acoustical center is further away from the device's front baffle resulting in higher SPL than a point source placed in the location of the front of a CBT.

2. SMALL FACTOR CBT DESIGN GUIDELINES AND CONSTRAINTS

TRANSDUCER SELECTION

The loudspeaker device is divided into a low frequency and a CBT section. The former, with a compact 'bookshelf' format in mind, consists of a high displacement, 125 mm transducer in an 8 liter chamber with a helmholtz resonator tuned to 45 Hz. The latter is composed of 5 pieces of 72 mm full range transducers with measured frequency response of 200 Hz to 18.5 kHz (-6 dB) and nominal dispersion of 90 degrees (-6 dB, conical).

The diameter of full range drivers determines minimum acoustical center-to-center spacing on the front baffle. Size of the diaphragm also influences lower frequency limit and high frequency beam width narrowing (beaming). The choice of full range transducer was a compromise between mentioned parameters.

POSITION OF THE APPARENT ACOUSTICAL SOURCE

The loudspeaker cabinet is designed to house the full range transducers in line with each other on a curved front baffle. The tangent of this curvature is perpendicular to the ground plane, when the wave reflection from this plane is taken into account, the acoustical axis is parallel to the surface the device is placed on. The radius of a front baffle is 1.1352 m which is also an apparent acoustical source behind the loudspeaker device (fig. 1).

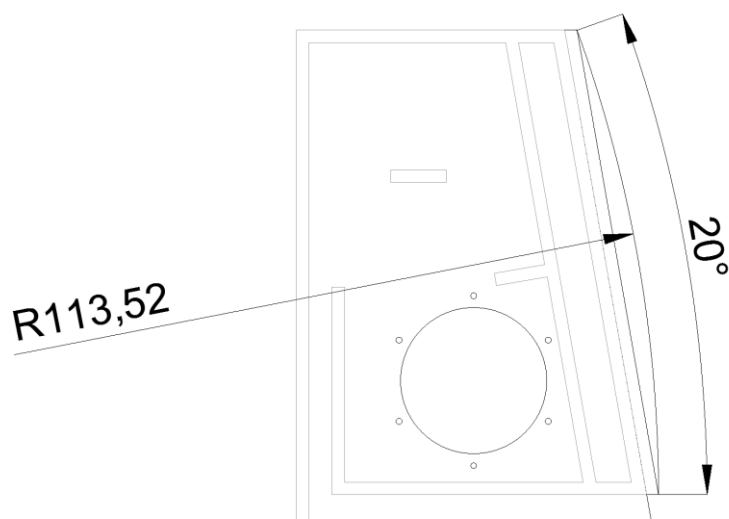


Figure 1 Cabinet schematic

ELECTRICAL FILTER CONCEPT

Legendre shading was implemented for improvement of vertical directivity consistency over frequency (fig. 2).

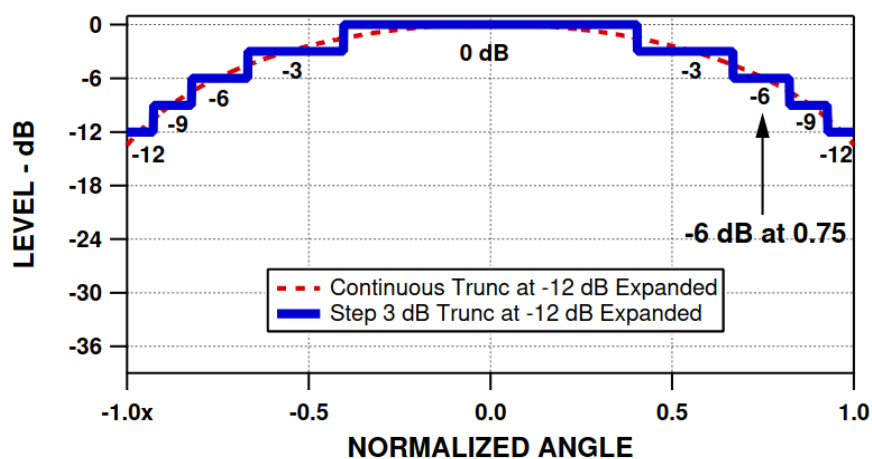


Figure 2 Legendre shading - stepped approximation (angle in radians) [2]

The 5-loudspeaker array was divided into 3 sections: 2 bottom transducers connected in series (0 dB), 2 middle transducers also in series with a voltage divider (-3 dB), one top transducer connected with voltage divider (-6 dB).

The drivers had a nominal impedance of 4 ohms, the array should have a resulting impedance of at least 4 ohms as well, lower value could be a problematic load for common power amplifiers. The input impedance of voltage dividers was adjusted to achieve resulting impedance of 4 ohms (fig. 3).

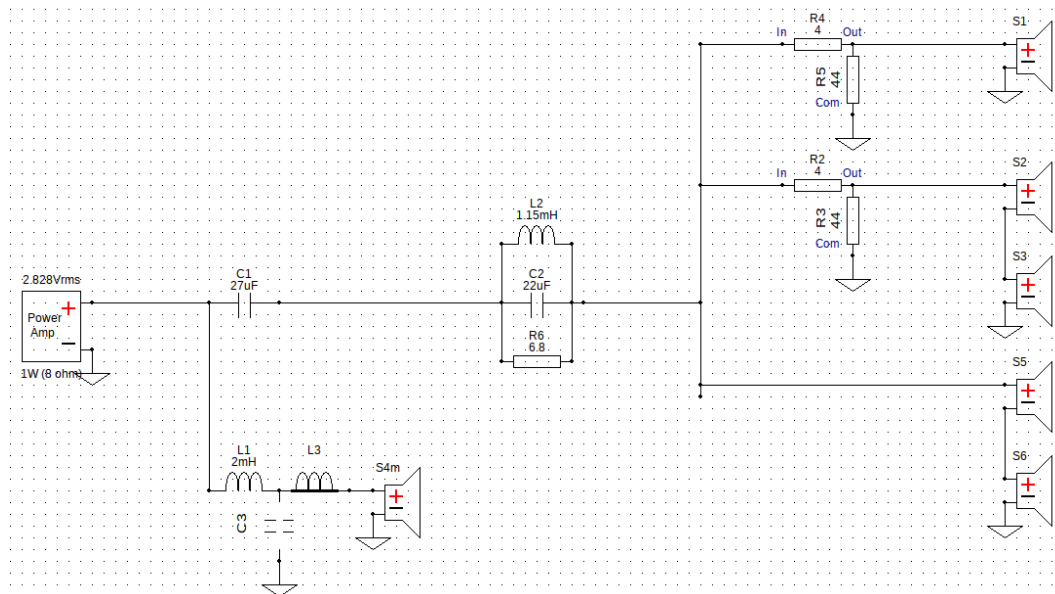


Figure 3 Electrical schematic of shading implementation

3. PROTOTYPE MEASUREMENTS

The frequency response measurements were taken on a ground plane (fig. 4) at distances from 10 to 320 cm. The data is presented as an average SPL (in 500 Hz - 20 kHz band) over distance relative to the level measured at 10 cm from the device. (fig. 5) The measured attenuation was compared with theoretical slopes: a line source and a point source.

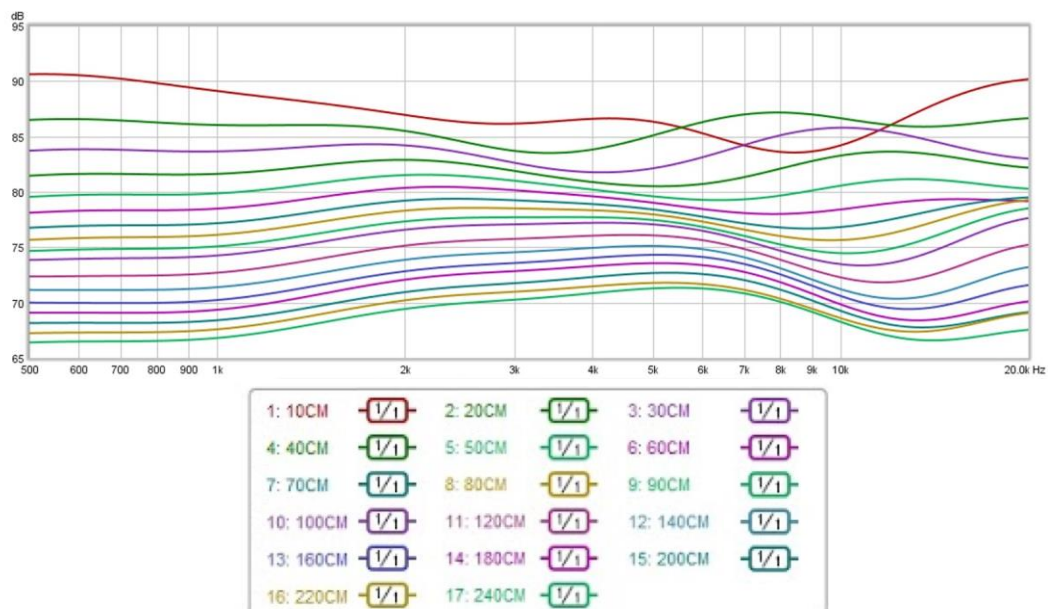


Figure 4 Subsequent frequency response measurements over distance from the device

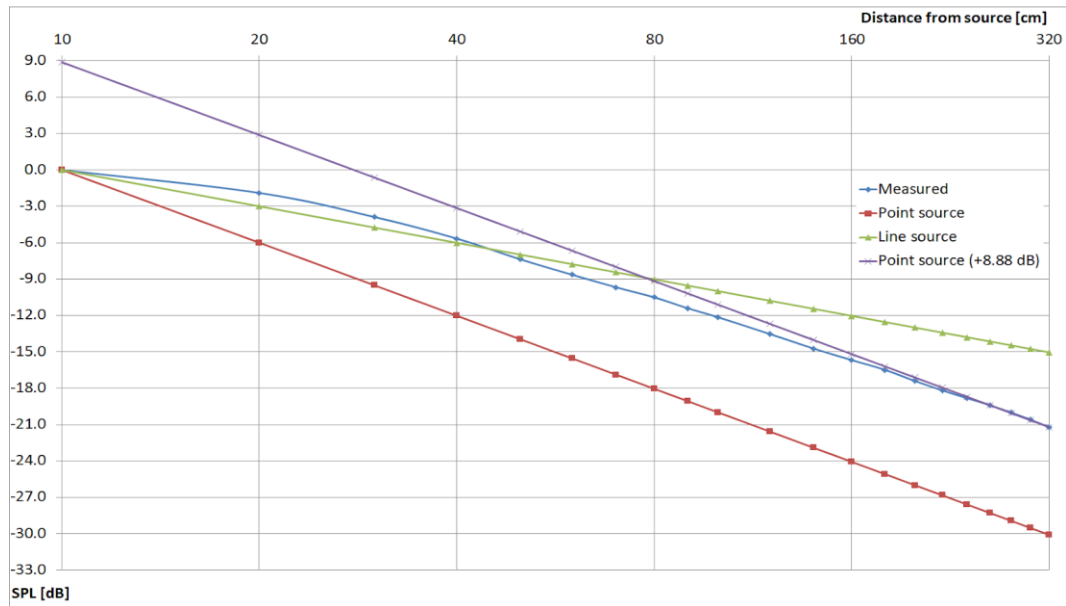


Figure 5 Measured SPL over distance compared to theoretical behaviour of a line and point sources

The tilt of the measured slope is smaller than line source's slope in near field (distance up to 40 cm) then slowly transitioning to being parallel to the point source's slope tilt at 320 cm. Attenuation over distance is between the line source's and a point source's theoretical values in the medium field.

For clearer presentation the additional point source attenuation curve was raised by 8.88 dB to match the SPL of the measured device at 320 cm distance. It indicates CBT's advantage over a point source loudspeaker, which has higher difference in sound pressure levels over distance and significantly lower efficiency in far field conditions.

The advantage of a CBT can be presented as a SPL gain over point source in function of distance (fig. 6) and compared to equally calculated gain of a line source.

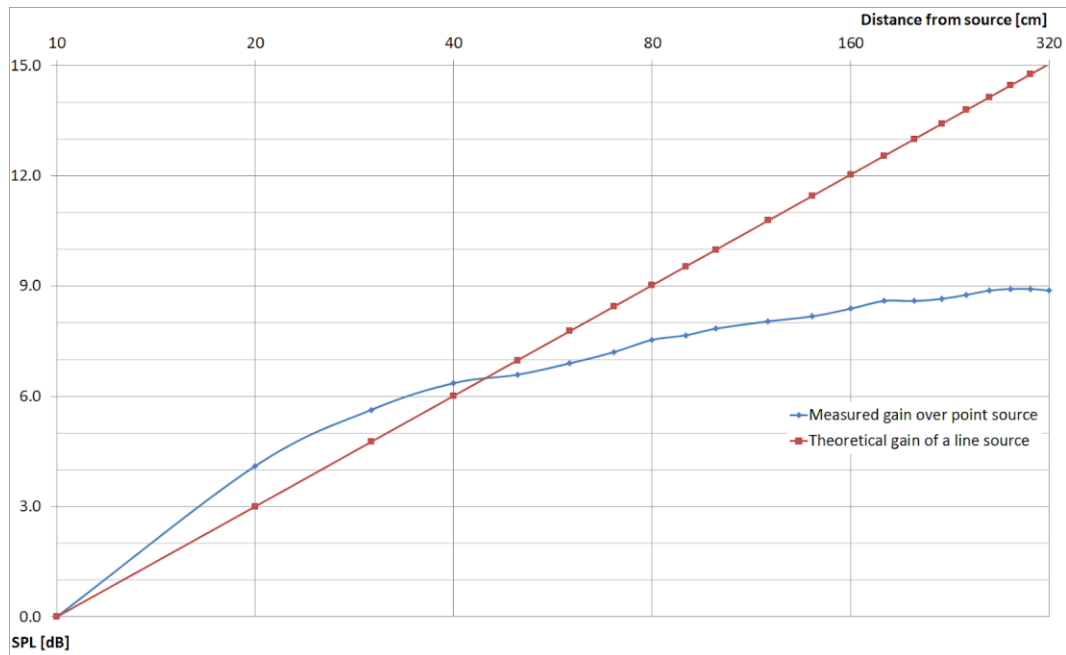


Figure 6 Measured SPL gain over a point source over distance with theoretical line source gain reference

4. CONCLUSIONS

Considering the results presented in this paper: the small format curved loudspeaker array provides measurable advantage over a point source loudspeaker in sound pressure level uniformity over distance. This could potentially lead to larger 'sweet spot' in the listening area for the CBT devices. Decreased sound pressure level between the near field and far field listening positions could provide flexibility in the sound engineer's positioning in the studio.

Further measurements of directivity behaviour and device optimization is scheduled.

BIBLIOGRAPHY

- [1] P. H. Rogers, and A. L. Van Buren, "New Approach to a Constant Beamwidth Transducer," *J. Acous. Soc. Am.*, vol. 64, no. 1, pp. 38-43 (1978 July)
- [2] Keele, Jr., D. B., "The application of broadband constant beamwidth transducer (CBT) theory to loudspeaker arrays," in *Audio Engineering Society Convention 109*, Audio Engineering Society, 2000.
- [3] Keele, Jr., D. B., "Practical implementation of constant beamwidth transducer (CBT) loud- speaker circular- arc line arrays," in *Audio Engineering Society Convention 115*, Audio Engineer- ing Society, 2003
- [4] Keele, Jr., D. B. and Button, D. J., "Ground- plane constant beamwidth transducer (CBT) loud- speaker circu- lar-arc line arrays," in *Audio Engineering Society Convention 119*, Audio Engineering Society, 2005

LOKALIZACJA ŹRÓDEŁ DŹWIĘKU Z WYKORZYSTANIEM SZTUCZNYCH SIECI NEURONOWYCH

Maciej WALCZYŃSKI¹

Piotr RUSZKOWSKI

¹Politechnika Wrocławska, Wyspiańskiego 27, 50-370 Wrocław
maciej.walczynski@pwr.edu.pl

W pracy przedstawiono wyniki badań dotyczących możliwości detekcji lokalizacji źródeł dźwięku oraz jej skuteczności z użyciem nagrań wykonanych z użyciem sztucznej głowy z wykorzystaniem sztucznych sieci neuronowych. Wykorzystanie w pracy algorytmów sztucznej inteligencji oraz modelu sztucznej głowy wynikało z próby odwzorowania percepcji sygnałów dźwiękowych przez ludzi z wykorzystaniem współczesnej technologii informatycznej. W problemie odwzorowania lokalizacji źródła dźwięku wykorzystano metody cyfrowego przetwarzania sygnałów oraz algorytmy uczenia maszynowego.

1. WSTĘP

Algorytmy uczenia maszynowego, a także sztuczne sieci neuronowe są z niemałym powodzeniem wykorzystywane w wielu zagadnieniach, jak na przykład w problemie automatycznej detekcji obrazów [1,2]. W takim przypadku z wykorzystaniem kamery czy aparatu algorytm dokonuje klasyfikacji obiektów znajdujących się na obrazie. Ponadto algorytmy uczenia maszynowego wykorzystywane są przy realizacji samouczących modeli rozpoznawania mowy czy tekstu.

Celem niniejszej pracy było zbadanie możliwości detekcji lokalizacji źródeł dźwięku oraz jej skuteczności z użyciem nagrań wykonanych z użyciem sztucznej głowy z wykorzystaniem sztucznych sieci neuronowych.

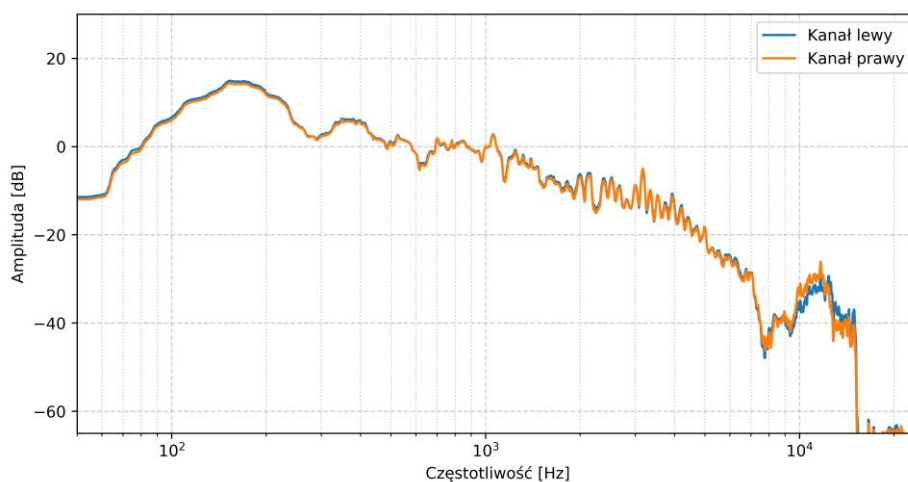
Umiejętność lokalizacji źródła dźwięku przez zwierzęta i ludzi jest bardzo istotną umiejętnością. Pozwala ona bowiem na wskazanie kierunku interesującego nas obiektu lub natychmiastowe wskazanie kierunku zagrożenia. Informacja o lokalizacji źródła fali akustycznej pozwala nam na błyskawicznym skupieniu wzroku na zdetekowanym obiekcie. Lokalizacja dźwięku, to jest ocena z jakiego kierunku z jakiego fala akustyczna dochodzi, oraz informacja o odległości źródła fali akustycznej zachodzi w oparciu o tak zwane czynniki lokalizacyjne [3]. Proces lokalizacji dokonywany jest na podstawie jedno-usznego przetwarzania sygnału oraz z wykorzystaniem porównania sygnałów dochodzących do obydwu uszu. Czynniki lokalizacyjne zależne są od fizycznych właściwości dźwięku. Możliwa jest analiza czynników w zależności od tego, czy sygnałem jest pobudzenie sinusoidalne o stałej częstotliwości oraz amplitudzie (ton prosty), czy sygnał złożony o zmiennych częstotliwościach składowych oraz modulowanej amplitudzie [4]. Użyteczność konkretnych czynników lokalizacyjnych zależy również od kierunku dochodzenia dźwięku. W płaszczyźnie horyzontalnej detekcja jest bardzo precyzyjna (sięgająca dokładności 1°). Rozdzielczość lokalizacji w płaszczyźnie środkowej czy przedniej jest mniejsza (wynosi około 4°). Detekcja w płaszczyźnie horyzontalnej jest dokonywana przede wszystkim z wykorzystaniem różnicy czasu dojścia (ITD), oraz na podstawie międzyusznej różnicy poziomów (ILD) [5]. W przypadku detekcji na płaszczyźnie przedniej oraz środkowej istotną informację stanowi widmo sygnału kształtowane przez ludzkie małżowiny oraz odbicia od torsu.

Eksperymenty polegające na automatycznej lokalizacji dźwięku zostały przeprowadzone dla zbioru

piętnastu różnych pozycji źródła względem głowy. W celu uniknięcia zwielokrotniania podobnych pomiarów, źródło dźwięku umiejscowiono po lewej strony sztucznej głowy. W procesie cyfrowego przetwarzania sygnałów akustycznych oraz w implementacji sztucznej sieci neuronowej wykorzystany został język *Python* wraz z modułami *Numpy* (odpowiedzialnym za obliczenia numeryczne), *Scipy* (przetwarzanie sygnałów) oraz *Keras* i *Tensorflow* (realizacja implementacji sztucznej sieci neuronowej). Wybór języka *Python* został podyktowany ze względu na jego duże możliwości obliczeniowe, jak i dużą liczbę wysokiej jakości implementacji narzędzi matematycznych, które zostały pod tym kątem odpowiednio zoptymalizowane.

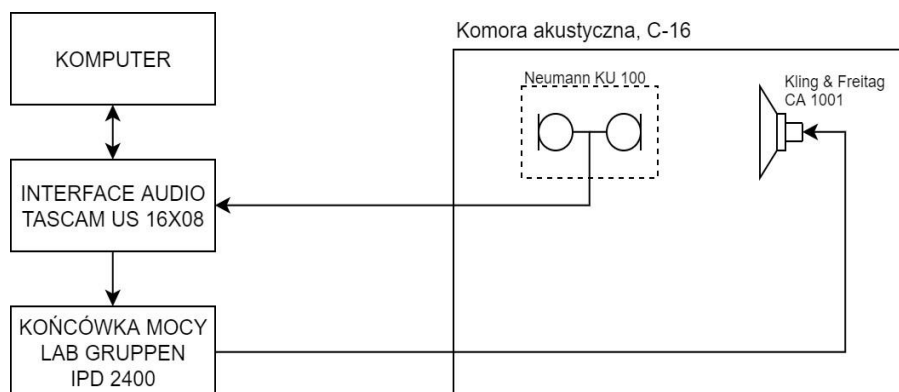
2. PRZYGOTOWANIE SYGNAŁÓW TESTOWYCH

Niezwykle ważny elementem badań stanowiło przygotowanie sekwencji sygnałów, która była odtwarzana przez źródło. Aby możliwa była analiza transmitancji całego systemu elektroakustycznego, na samym początku sekwencji umieszczony został sygnał sinusoidalny o logarytmicznie przestrajanej częstotliwości. Do trenowania sieci, a także jej testowania jej skuteczności, wykorzystano sygnały, mowy oraz sygnały muzyczne.



Rysunek 1 Widmo przykładowego sygnału muzycznego używanego do trenowania sieci.

Jako źródło fali akustycznej zastosowano urządzenie głośnikowe. Rysunek 2 przedstawia schemat połączeń toru elektroakustycznego wykorzystanego w celu przygotowania sygnałów testowych.



Rysunek 2 Schemat połączeń podczas nagrywania próbek testowych

3. REALIZACJA SZTUCZNEJ SIECI NEURONOWEJ

Model sztucznej sieci neuronowej używany do lokalizacji źródła dźwięku zaimplementowano opierając o bibliotekę *Keras*. Biblioteka ta, wykorzystuje z kolei pakiet *Tensorflow*. Liczba neuronów w warstwie wejściowej stworzonej sztucznej sieci neuronowej jest determinowana poprzez zadaną liczbę cech. Z kolei liczba neuronów w warstwie wyjściowej określona jest poprzez liczbę klas, tj. Liczbę pozycji źródła względem sztucznej głowy.

Stworzona na potrzeby tej pracy sieć posiadała 6 neuronów w warstwie wejściowej oraz 15 neuronów w warstwie wyjściowej. Role funkcji aktywacji neuronów warstwy wyjściowej pełniła funkcja *softmax*. Liczba warstw ukrytych, a także ich rozmiar zostały dobrane w sposób empiryczny. W tym celu przeprowadzono szereg eksperymentów obliczeniowych mających na celu poprawę skuteczności modelu. Model w swej ostatecznej formie posiada dwie warstwy ukryte, z których każda zawiera 64 neurony. Rolę funkcji aktywacji pełnił tangens hiperboliczny.

Pomiędzy każdą z warstw ukrytych, znajdowała się warstwa typu drop-out. Dla pierwszej warstwy drop-out wartość współczynnika wynosiła 0,1, natomiast wartość drugiego współczynnika wynosiła 0,3.

4. MODEL SYSTEMU AUTOMATYCZNEJ LOKALIZACJI

Jednym z elementów prezentowanej pracy było opracowanie algorytmu pozwalającego na estymację międzyususznej różnicy poziomów, jak i międzyususznej różnicy czasu dojścia. Wartość, którą wzorowano na ILD obliczono jako różnicę wartości skutecznej sygnału zarejestrowanego przez ucho lewe oraz wartości skutecznej sygnału zarejestrowanego przez ucho prawe. Implementacji algorytmu dokonano z wykorzystaniem języka *Python*. Listing omawianej funkcji przedstawiony jest na Rysunku 3. W omawianym algorytmie dokonywana jest normalizacja wartości skutecznej obu sygnałów, przy czym czynione jest to względem sygnału odbieranego przez ucho lewe. Proces normalizacji przeprowadzono ze względu na późniejsze przetwarzanie wartości przez blok aktywacji w sztucznym neuronie.

```

1  def get_ild(path_tuple):
2      LEFT_FILE_PATH, RIGHT_FILE_PATH = path_tuple
3
4      data_left, l_fp = sf.read(LEFT_FILE_PATH)
5      rms_l = np.sqrt(np.mean(data_left**2))
6
7      data_right, r_fp = sf.read(RIGHT_FILE_PATH)
8      rms_r = np.sqrt(np.mean(data_right**2))
9
10     norm_ild = rms_l/rms_l - rms_r/rms_l
11     return norm_ild
12

```

Rysunek 3. Implementacja funkcji ILD dla dwóch sygnałów wejściowych.

5. PODSUMOWANIE

Stworzenie modelu oraz symulacja systemu lokalizacji źródła dźwięku w przestrzeni z wykorzystaniem sztucznych sieci neuronowych i metod cyfrowego przetwarzania sygnałów pozwoliło na detekcję kierunku z bardzo dużą - osiagającą niemal 90% - skutecznością.

Ukształtowanie głowy oraz małżowin usznych, w znaczny sposób wpływa na transmitancję toru, dzięki czemu możliwe jest rozróżnienie kierunku z jakiego dochodzi fala akustyczna. Detekcja lokalizacji źródła dźwięku przez algorytm z wykorzystaniem sztucznej głowy, jest możliwa do realizacji dla wybranych założeń.

Pomimo wpływu transmitancji całego zestawionego toru elektroakustycznego poprzedzającego sztuczną głowę, udało się dokonać analizy istotnych cech sygnału oraz przeprowadzić ewaluację sztucznej sieci neuronowej.

Istotnym problemem wynikającym z wykorzystania sztucznych sieci neuronowych, jest konieczność wykonania bazy sygnałów testowych; wiąże się to z dużym nakładem czasu na przygotowanie ciągu uczącego oraz jego rejestrację w warunkach rzeczywistych, z wykorzystaniem sztucznej głowy.

LITERATURA

- [1] Nowell, D. & Nowell, P.W.. (2019). A machine learning approach to the prediction of fretting fatigue life. *Tribology International*. 141. 105913. 10.1016/j.triboint.2019.105913.
- [2] Karkanis, Stavros & Magoulas, George & Theofanous, N. (2000). Image recognition and neuronal networks: Intelligent systems for the improvement of imaging information. *Minimally invasive therapy & allied technologies : MITAT : official journal of the Society for Minimally Invasive Therapy*. 9. 225-30. 10.1080/13645700009169652.
- [3] Bronkhorst, Adelbert. (1995). Localization of real and virtual sound sources. *Journal of the Acoustical Society of America*. 98. 10.1121/1.413219.
- [4] Majdak, Piotr & Goupell, Matthew & Laback, Bernhard. (2010). 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training. *Attention, perception & psychophysics*. 72. 454-69. 10.3758/APP.72.2.454.
- [5] Dobrucki A., Pruchnicki P., and Plaskota P. System for high speed measurement of head-related transfer function. 2012.

LOCALIZATION OF SOUND SOURCES USING ARTIFICIAL NEURAL NETWORKS

The paper presents the results of research concerning the possibility of sound source localization detection and its effectiveness with the use of recordings made with artificial head with the use of artificial neural networks. The use of artificial intelligence algorithms and artificial head model in the work resulted from an attempt to reproduce the perception of sound signals by people using modern information technology. In the problem of the mapping of sound source location, digital signal processing methods and machine learning algorithms were used.

WYKORZYSTANIE WSKAŹNIKA ODBICIA DŹWIĘKU DO OCENY WŁAŚCIWOŚCI POCHŁANIAJĄCYCH ELEMENTÓW BUDOWLANYCH

Paweł DZIECHCIŃSKI

Politechnika Wrocławska, wyb. Stanisława Wyspiańskiego 27, 50-370 Wrocław
pawel.dziehcinski@pwr.edu.pl

W pracy omówiono możliwości wykorzystania wskaźnika odbicia dźwięku do oceny właściwości pochłaniających elementów budowlanych. Metodyka pomiarów wskaźnika odbicia dźwięku znana jako "QUIESST" jest między innymi przedmiotem normy PN-EN 1793-5, która dotyczy ekranów akustycznych. Poza wyjaśnieniem podstaw algorytmu pomiarowego, w pracy omówiono również wymagania co do systemu pomiarowego zarówno w kontekście wymagań normy jak i własnych doświadczeń autora. Przedstawiono wyniki wykonanych pomiarów i ich porównanie z wynikami uzyskanymi innymi metodami. Pomiary wykonano za pomocą systemu zbudowanego z wykorzystaniem powszechnie dostępnych urządzeń takich jak wielokanałowy interfejs USB przeznaczony do zastosowań w realizacji dźwięku czy samochodowego wzmacniacza mocy. Dostępność elementów składowych systemu sprawia, że taka metoda pomiarowa wydaje się być atrakcyjna dla wszystkich, którzy są zainteresowani oceną właściwości pochłaniających elementów budowlanych, a uzyskane wyniki, potwierdzają, że pomimo szeregu ograniczeń jest ona do tego celu przydatna.

1. WPROWADZENIE

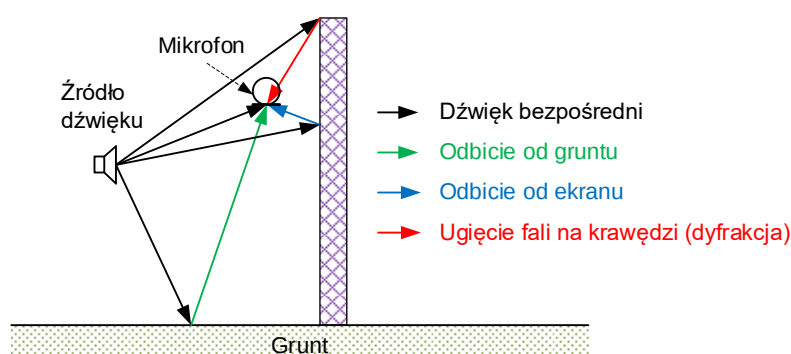
Jednym z elementów inwentaryzacji akustycznej obiektów budowlanych jest ocena ich elementów składowych w zakresie właściwości pochłaniania dźwięku. W warunkach laboratoryjnych właściwości te ocenia się za pomocą współczynnika pochłaniania dźwięku wyznaczanego w warunkach pola pogłosowego (PN-EN ISO 354:2005 [11], PN-EN 1793-1:2017-05 [8]), albo współczynnika pochłaniania dźwięku przy prostopadłym padaniu fali dźwiękowej (PN-EN ISO 10534-2:2003 [12]). W pracy omówiono metodykę pomiaru wskaźnika odbicia dźwięku RI (*ang.* reflection index) rozwijaną głównie na potrzeby badań właściwości ekranów akustycznych w miejscu ich zainstalowania. Najnowszą normą, której przedmiotem jest pomiar RI , jest PN-EN 1793-5:2016-05+AC:2018-08 [9]. Metodyka pomiarów wskaźnika odbicia dźwięku RI jest wynikiem badań prowadzonych od początku lat 90-tych 20-tego wieku. Pierwsza wersja zaproponowana przez Garai [1] oparta była o pomiary za pomocą jednego mikrofonu i dla prostopadłego do badanej płaszczyzny kierunku padania fali. Mommertz [6] zaproponował modyfikację tej metody polegającą na zastosowaniu techniki odejmowania odpowiedzi impulsowych w celu rozdzielenie impulsu fali padającej i odbitej w sygnale, i wykonanie pomiarów dla kilku kątów padania fali. Kontynuację badań w tej tematyce zrealizowano w ramach projektu Adrienne, w wyniku którego powstała specyfikacja techniczna CEN/TS 1793-5:2003. Metodyka będąca przedmiotem tego standardu znana jest jako metoda Adrienne i polega na wykonywaniu pomiarów odpowiedzi impulsowych w 9 punktach, z rozdzielczością 10° za pomocą głośnika, z którym na sztywno połączony jest mikrofon. Szereg prac mających na celu doskonalenie metody [2,3,4,5] zaowocowało wprowadzeniem normy EN 1793-5:2016 (i jej polskiego odpowiednika [9]). W normie tej wprowadzono zmodyfikowaną metodykę pomiarową znaną jako QUIESST (Quietening the Environment for a Sustainable Surface Transport).

Pomimo, że norma omawia pomiary ekranów akustycznych, jednak wydaje się, że metoda ta może być również wykorzystana do badania innych elementów budowlanych. Między innymi ze względu na zróżnicowane warunki pomiarowe nie należy oczekiwać, że wartości współczynników pochłaniania

bezpośrednio uzyskane tymi metodami będą zgodne, jednak dla niektórych przypadków mogą być one zbliżone. W pracy przedstawiono wyniki pomiarów dwóch elementów budowlanych o zróżnicowanych właściwościach odbijających. Badano płaskie materiały o powierzchni jednorodnej w zakresie właściwości pochłaniających dźwięk. Z doniesień literaturowych [13,14] i doświadczeń autora wynika, że w innych przypadkach metoda ta może się nie sprawdzać, pomimo, że w PN-EN 1793-5 nie podano takich ograniczeń. Uzyskane wyniki porównano z wartościami uzyskanymi na podstawie pomiarów czasu pogłosu oraz z danymi literaturowymi. Ponadto omówiono wymagania co do systemu pomiarowego zarówno w kontekście wymagań normy jak i własnych doświadczeń autora. Na potrzeby badań wykonano system pomiarowy wykorzystujący powszechnie dostępne urządzenia. Wykorzystanie takich urządzeń może stanowić o powszechności a tym samym atrakcyjności metody. Ograniczeniem metody jest zwłaszcza zakres badanych częstotliwości. Zakres ten wiąże się z rozmiarami badanej próbki oraz nierównomiernościami jej powierzchni. Inne ograniczenia, na które zwrócono uwagę to wymagania co do systemu pomiarowego oraz badanych obiektów. Niektóre z wymagań zdaniem autora nie zostały w należyty sposób podkreślone w normie.

2. PODSTAWY ALGORYTMU POMIARU METODĄ RÓŻNICOWĄ

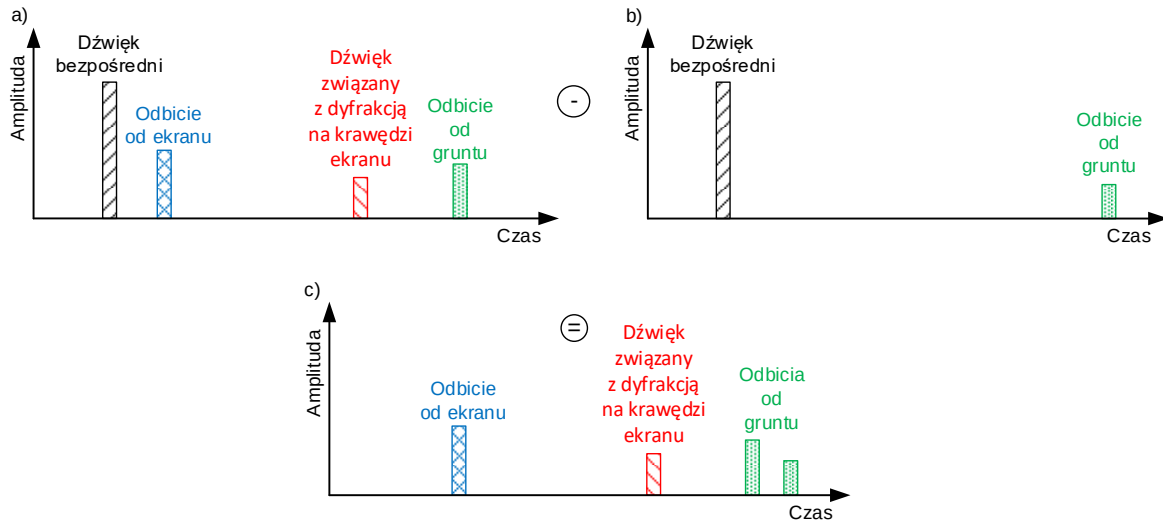
Pomiary wskaźnika odbicia omówione w normie PN-EN 1793-5:2016 oparte są o ideę tzw. metody różnicowej. Metoda ta oparta jest o analizę podstawowych zjawisk akustycznych zachodzących w obecności przeszkody po stronie źródła (rys. 1).



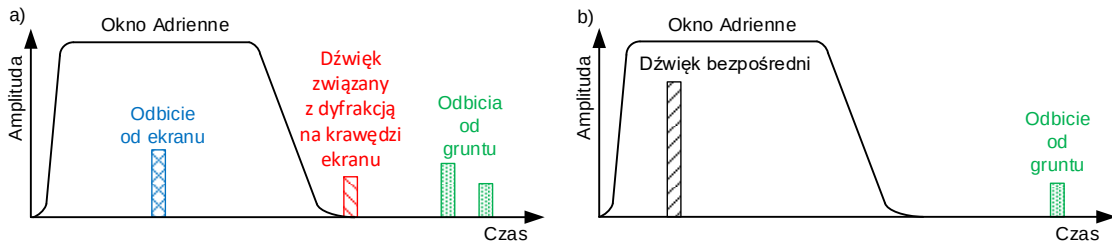
Rys. 1. Podstawowe zjawiska obrazujące pole akustyczne w obecności przeszkody po stronie źródła

Do punktu odbiorczego znajdującego się w obszarze między źródłem dźwięku a płaskim ekranem, dźwięk może docierać: bezpośrednio ze źródła, w wyniku odbić od ekranu i od gruntu, a także może to być dźwięk związany z dyfrakcją na krawędzi ekranu. W efekcie w punkcie obserwacji uzyskujemy odpowiedź impulsową, której schematyczny rozkład przedstawiono na rys. 2a. Ogólna idea pomiaru polega na tym, że w wyniku dwóch niezależnych pomiarów należy uzyskać dwie odpowiedzi impulsowe – dla dźwięku bezpośredniego i odbitego od badanego ekranu oraz dla dźwięku bezpośredniego. Obie odpowiedzi impulsowe należy wyznaczyć dla tej samej odległości głośnika od mikrofonu i nie zmieniając wzmocnienia systemu. W praktyce pomiar odpowiedzi impulsowej dźwięku bezpośredniego również wykonuje się nad powierzchnią gruntu (ewentualnie zwiększając wysokość układu), ale w znacznej odległości od ekranu. W związku z tym uzyskuje się odpowiedź impulsową, której schematyczny rozkład przedstawiono na rys. 2b. Odpowiedź impulsową związaną z odbiciem od badanego ekranu wyznacza się jako różnicę między pierwszą a drugą z powyższych odpowiedzi. W wyniku uzyskuje się odpowiedź impulsową, której schemat przedstawiono na rys. 2c. Zarówno odpowiedź impulsowa z dźwiękiem bezpośrednim (rys. 2b) jak i odbitym od ekranu (rys. 2c) zawierają niepożądane składowe związane z odbiciem od gruntu i z dyfrakcją na krawędzi ekranu. Rozwiązanie tego problemu

wymaga takiego doboru geometrii układu pomiarowego, aby niepożądane efekty odbić i dyfrakcji pojawiały się w uzyskiwanej odpowiedzi impulsowej odpowiednio późno w stosunku do dźwięku bezpośredniego i odbicia od ekranu. Dzięki temu, niepożądane składowe można wyeliminować poprzez specjalnie zaprojektowane okno czasowe zwane oknem Adrienne. Operację tę dla uzyskanej odpowiedzi różnicowej przedstawiono na rys. 3a. To samo okno czasowe wykorzystuje się w celu eliminacji odbicia od gruntu dla odpowiedzi impulsowej mającej na celu wyznaczenie dźwięku bezpośredniego rys. 3b. Wskaźnik odbicia wyznacza się poprzez wyznaczenie stosunku energii dźwięku odbitego do dźwięku bezpośredniego w pasmach 1/3 oktawy, uwzględniając odpowiednie współczynniki korekcyjne.



Rys. 2. Schematyczny rozkład amplitud odpowiedzi impulsowej: a) w obecności przeszkody po stronie źródła, b) dźwięku bezpośredniego nad gruntem, c) uzyskanej w wyniku odejmowania odpowiedzi z rys. a) i b)



Rys. 3. Eliminacja niepożądanych składowych odpowiedzi impulsowych dla dźwięku: a) odbitego, b) bezpośredniego

W metodzie QUIESST układ pomiarowy składa się z głośnika i 9-cio mikrofonowej matrycy odpowiednio umieszczonej w stosunku do badanego ekranu. Do przeprowadzenia pomiarów w zakresie częstotliwości określonym przez normę PN-EN 1793-5 minimalna wysokość oraz szerokość badanego elementu powinny wynosić co najmniej 4 m. Z podaną geometrią wiąże się podstawowa długość okna czasowego Adrienne, wynosząca 7,9 ms. W przypadku, kiedy najmniejszy wymiar badanej próbki jest mniejszy od 4 m konieczne jest zastosowanie okna czasowego o mniejszej długości. Przyjmuje się, że badany obszar jest kołem o promieniu r , który można wyznaczyć ze wzoru (1):

$$r = \frac{1}{d_S + d_M + cT_w} \sqrt{\left(d_S + d_M + \frac{cT_w}{2}\right) \left(d_S + \frac{cT_w}{2}\right) (2d_M + cT_w) cT_w} \quad (1)$$

gdzie: c – prędkość dźwięku, T_w – długość okna czasowego, d_S – odległość między głośnikiem i badaną próbką, d_M – odległość między mikrofonami i badaną próbką.

3. SYSTEM POMIAROWY

Do badań wykorzystano system pomiarowy wykonany we własnym zakresie. System ma być wykorzystywany zarówno w warunkach laboratoryjnych - z dostępem do sieci elektrycznej, jak i w warunkach terenowych, gdzie będzie wymagał zasilania akumulatorowego. Dziewięciokanałowe odpowiedzi impulsowe stanowiące podstawę pomiarów wyznaczano za pomocą programu EASERA, a przetwarzano je za pomocą programu DIRAC oraz własnych programów i arkuszy kalkulacyjnych.

Zgodnie z wymaganiami normy PN-EN 1793-5:2016, cały system pomiarowy powinien spełniać wymagania dla mierników pierwszej klasy dokładności [10], za wyjątkiem mikrofonów, które powinny spełniać wymagania dla urządzeń klasy drugiej i powinny mieć średnicę nie większą niż $\frac{1}{2}$ ". Matryca mikrofonowa składa się z dziewięciu mikrofonów równomiernie rozmieszczonych w rastrze 3×3 z rozstawem 40 cm. Do wykonania matrycy mikrofonowej zastosowano mikrofony o średnicy $\frac{1}{4}$ " i spełniające wymagania stawiane mikrofonom klasy pierwszej. W przypadku pomiarów wskaźnika odbicia dźwięku bardzo istotnym elementem systemu jest rama na której zamocowane są mikrofony. Elementy tej konstrukcji mogą być źródłem niepożądanych odbić, dyfrakcji oraz zasłaniania dźwięk bezpośredni lub odbity dochodzący do mikrofonu i tym samym mogą mieć wpływ na wyniki pomiarów. W ramach badań sprawdzono kilka wersji ramy. W opracowaniu przedstawiono wyniki uzyskane z dwoma wersjami ramy – rama #2 wykonana z elementów o przekroju zbliżonym do kwadratu o boku 20 mm oraz rama #5 wykonana z rur okrągłych o średnicy 12 mm.

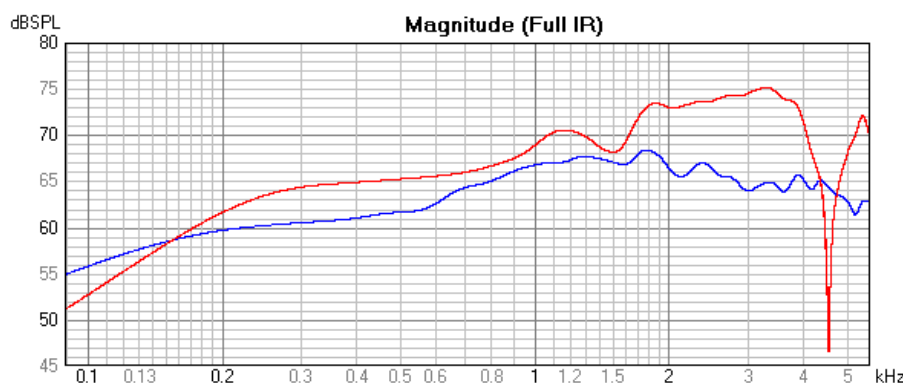
Układ do pomiarów odpowiedzi impulsowych przedstawiony w PN-EN 1793-5:2016 i część związanych z nim wymagań wydają się być nieco archaiczne. Wprawdzie w zapisach w najnowszym wydaniu normy pojawia się informacja o możliwości wyznaczania odpowiedzi impulsowej zarówno za pomocą sygnału MLS jak i e-sweep, ale zamieszczony w normie układ dotyczy sytuacji, w której sygnał MLS generowany jest analogowo. Specyficzne są też zapisy dotyczące wymagań stawianych filtrom antyaliasingowym. Norma wymaga, aby częstotliwości graniczne filtrów antyaliasingowych f_{co} były relatywnie małe w stosunku do częstotliwości próbkowania f_s . Częstotliwości graniczne filtrów f_{co} mają być mniejsze lub równe $1/3f_s$ dla filtrów Czebyszewa i $1/4f_s$ dla filtrów Butterwortha. Współczesne techniki [7] i systemy do pomiarów odpowiedzi impulsowych wykorzystują sygnały pomiarowe generowane cyfrowo. Zarówno przetworniki A/C jak i C/A współczesnych systemów pomiarowych wykorzystywanych w technice fonicznej są oparte o architekturę $\Sigma\Delta$, a więc pracują z dużym nadpróbkowaniem i w związku z tym filtracja antyaliasingowa jest realizowana dwustopniowo – filtr analogowy niskiego rzędu plus filtr cyfrowy o dużym nachyleniu zbocza stanowiący element układu scalonego przetwornika. Takie systemy przetwarzania A/C i C/A w praktyce uniemożliwiają użytkownikowi ustawienie częstotliwości granicznych zgodnie z przedmiotowymi normami. W wykonanym systemie wykorzystano popularny foniczny interfejs USB z 12 wejściami mikrofonowymi, który dla pomiarów z dziewięcioma kanałami umożliwia przetwarzanie sygnałów z częstotliwością próbkowania do 96 kHz. Ograniczeniem wybranego interfejsu była zbyt mała wydajność prądowa zasilania „Phatnom”, w związku z tym dokonano jego odpowiedniej modyfikacji.

W systemie wykorzystano wzmacniacz samochodowy, który można zasilać z akumulatorów co zgodnie z założeniami ułatwi pomiary w warunkach terenowych.

Zmierzono parametry elektryczne wykorzystywanego interfejsu i wzmacniacza mocy, potwierdzające, że urządzenia te w odpowiednim zakresie spełniają wymagania dla mierników klasy pierwszej według normy PN-EN 61672-1:2014 [10].

Urządzenie głośnikowe wykorzystywane do badań musi być urządzeniem jednogłośnikowym w obudowie zamkniętej, o charakterystyce częstotliwościowej bez ostrych nieregularności w zakresie pasm $1/3$ oktawy z zakresu od 100 Hz do 5 kHz, a jego odpowiedź impulsowa powinna być nie dłuższa niż 3 ms. Do badań początkowo wykorzystywano zmodyfikowany tzw. projektor głośnikowy z systemów rozgłoszeniowych (obudowa w kształcie walca o średnicy 14 cm) - w opracowaniu oznaczany jako głośnik #1. Okazało się jednak, że jego nierównomierności charakterystyki w zakresie dużych częstotliwości miały znaczący wpływ na wyniki pomiarów. W związku z tym, zaprojektowano i wykonano

urządzenie głośnikowe w kształcie prostopadłościanu - w opracowaniu oznaczany jako głośnik #2. W stosunku do projektora urządzenie to zapewnia większe poziomy ciśnienia akustycznego oraz posiada bardziej wyrównaną charakterystykę. Duże poziomy ciśnienia akustycznego są wymagane ze względu na konieczność zapewnienia odpowiednio dużego stosunku sygnału do szumu w warunkach terenowych. Z kolei ich uzyskanie jest trudne za pomocą urządzenia jednodrożnego w obudowie zamkniętej mającego wyrównaną charakterystykę w zakresie częstotliwości od 89 Hz do 5,6 kHz. Normy nie definiują kryterium, zgodnie z którym należy określić koniec odpowiedzi impulsowej głośnika. Wymóg wyrównanej charakterystyki częstotliwościowej jest jak najbardziej zasadny. Znaczna nierównomierność charakterystyki częstotliwościowej pierwszego z głośników między 4 i 5 kHz (rys. 4) wpływa na wyniki pomiarów co zostanie omówione w dalszej części pracy.



Rys. 4. Charakterystyki częstotliwościowe: głośnika #1 (czerwona) i głośnika #2 (niebieska)

4. BADANE OBIEKTY I WYNIKI BADAŃ

Do badań wskaźnika odbicia dźwięku wytypowano dwa obiekty o bardzo zróżnicowanych właściwościach odbijających. Badanym obiektem mocno odbijającym była betonowa podłoga komory akustycznej o wymiarach $13,2 \times 9,3 \times 6,5$ m, a pochłaniającym była ściana tej komory składająca się z następujących warstw (kolejność od strony padania fali akustycznej):

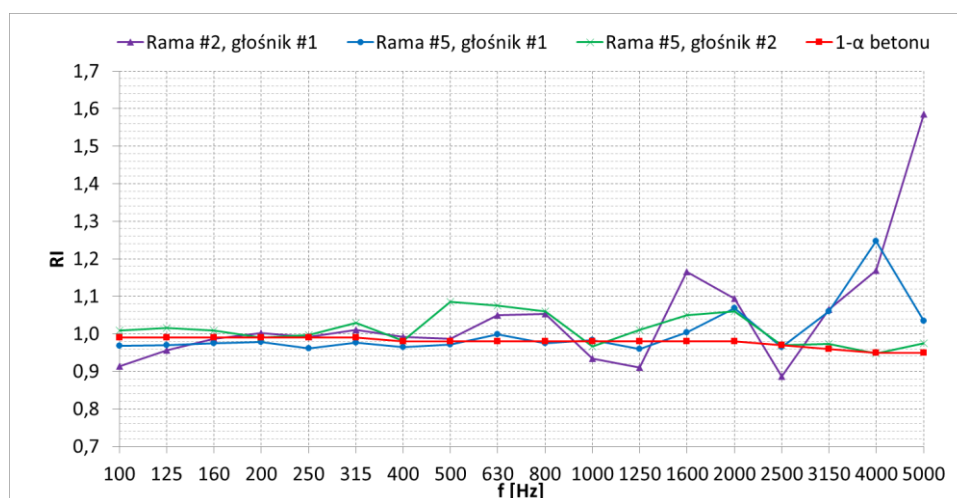
- 1) Ecophon Wall Panel SuperG C o grubości 40 mm,
- 2) wełna mineralna z jednostronnym welonem o gęstości 80 kg/m^3 i grubości 100 mm,
- 3) podkonstrukcja z pustką powietrzną o grubości 200 mm,
- 4) betonowa ściana nośna.

Uzyskane w wyniku pomiarów zgodnych z normą PN-EN 1793-5 wartości RI porównano z wartościami odbicia dźwięku wyznaczonymi na podstawie pomiarów czasu pogłosu pomieszczenia. Wartości odbicia dźwięku uzyskane na podstawie pomiarów czasu pogłosu wyznaczono jako $1 - \alpha$, gdzie α oznacza pogłosowy współczynnik pochłaniania dźwięku. Parametr ten w dalszej części pracy oznaczany będzie jako $1 - \alpha$. Przyjęto założenie, że ściany i sufit badanego pomieszczenia wykonane są z takiego samego materiału. Współczynniki pochłaniania betonowej podłogi zaczerpnięto z literatury. W rzeczywistości współczynniki pochłaniania dźwięku sufitu mogą się różnić od współczynników pochłaniania ścian ze względu na inną strukturę:

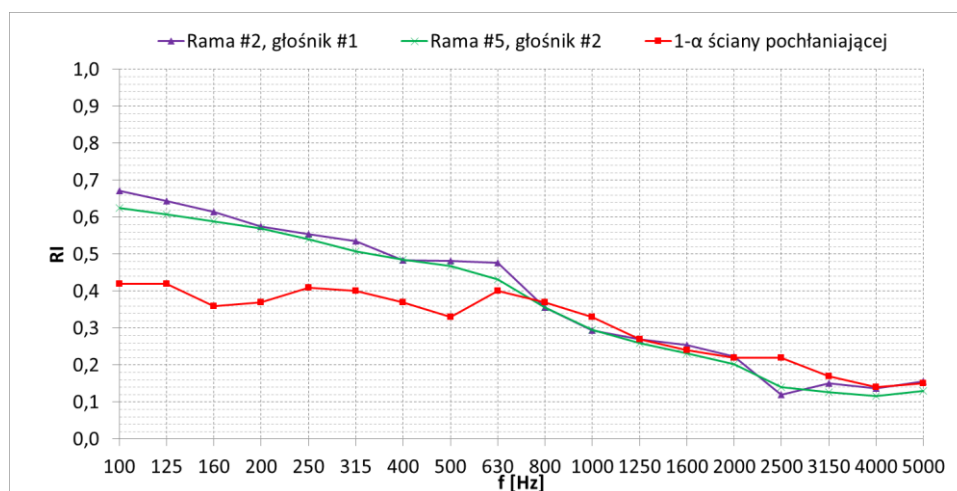
- 1) Ecophon Master DS o grubości 40 mm,
- 2) pustka powietrzna z profilami usztywniającymi o grubości 45 mm,
- 3) wełna mineralna z jednostronnym welonem o gęstości 80 kg/m^3 i grubości 100 mm,
- 4) podkonstrukcja z pustką powietrzną 400 mm (brak w obszarach pod belkami konstrukcyjnymi),
- 5) strop betonowy.

Wyniki pomiarów betonowej podłogi uzyskane z wykorzystaniem dwóch wersji głośników i ramy

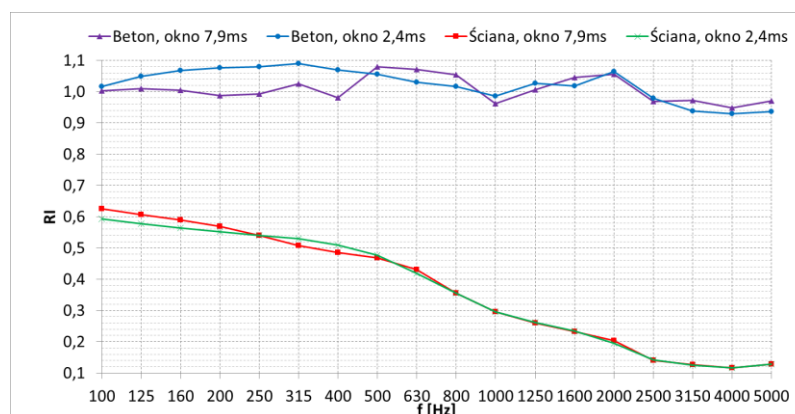
mocującej mikrofony przedstawiono na rys. 5, natomiast ściany o dużym współczynniku pochłaniania dźwięku na rys 6. W przypadku powierzchni mocno odbijającej wyraźnie uwypukliły się niedoskonałości ramy #2 wykonanej z elementów o przekroju w kształcie zbliżonym do kwadratu. Porównując wyniki dla obu ram uzyskane z głośnikiem #1 widać znaczne różnice uzyskanych wartości RI , które dla częstotliwości 1,6 kHz wynoszą 0,16, a dla 5 kHz 0,54. Ponadto dla wielu częstotliwości uzyskano wartości RI znacznie różniące się od 1,00 co nie odzwierciedla rzeczywistych właściwości odbijających podłogi. Szczególne odstępstwa od spodziewanych wartości RI odnotowano dla częstotliwości 4 i 5 kHz. Zgodnie z przewidywaniami były one wynikiem znaczącej nierównomierności charakterystyki częstotliwościowej głośnika #1 (rys. 4). Zastosowanie głośnika #2 z ramą #5 praktycznie wyeliminowało ten problem i zgodność wartości RI uzyskanych w zakresie 2,5-5 kHz z wartościami $1-\alpha$ betonu jest bardzo duża. Głośnik #2 był z kolei źródłem relatywnie dużych różnic między RI a $1-\alpha$ betonu w zakresie 0,5-1,6 kHz. Szczególnym przypadkiem jest częstotliwość 2 kHz, dla której niezależnie od badanego przypadku uzyskano podobne wartości RI większe od 1,00.

Rys. 5. Wyniki badań RI dla podłogi betonowej

Wpływ ramy i głośnika na uzyskiwane wartości RI był znacznie mniejszy dla ściany pochłaniającej. Dla tej struktury uzyskano jednak relatywnie duże różnice między wartościami RI a $1-\alpha$ dla częstotliwości mniejszych od 630 Hz.

Rys. 6. Wyniki badań RI dla ściany o dużym współczynniku pochłaniania dźwięku

W przypadku badanego pomieszczenia zarówno podłoga jak i ściany posiadały wymiary większe od 4×4 m i możliwe było wykonanie pomiarów dla geometrii pomiarowej, dla której odległość między mikrofonami a elementami wyposażenia pomieszczenia (poza badanym) była większa od 2 m. Umożliwiało to wykonanie analiz z wykorzystaniem okna czasowego Adrienne o standardowej długości 7,9 ms. Zgodnie z normą PN-EN 1793-5, właściwym zakresem pomiarowym dla tej długości okna czasowego jest zakres pasm 1/3 oktawy od 200 Hz do 5 kHz. W przypadku kiedy badany element ma mniejsze wymiary niż 4×4 m, lub geometria pomieszczenia ogranicza możliwość odsunięcia mikrofonów na 2 m od innych elementów pomieszczenia konieczne jest zastosowanie krótszego okna czasowego zgodnie ze wzorem (1). W pracy poza standardową długością okna czasowego wykorzystano okno Adrienne o długości 2,4 ms co odpowiada $r = 0,82$ m, a więc umożliwia ono badanie próbek, na których da się wyznaczyć okrąg o średnicy 1,64 m. Przyjmując w przybliżeniu, że dolna częstotliwość graniczna analiz jest odwrotnością długości okna, uzyskuje się w rezultacie zakres pomiarowy powyżej 420 Hz. Porównanie wyników pomiarów dla obu badanych obiektów i obu długości okna czasowego wykorzystywanego do analiz przedstawiono na rys. 7. Okno czasowe wygładza charakterystykę częstotliwościową i w przypadku badanych próbek, których charakterystyka częstotliwościowa RI była relatywnie równomierna wpływ okna czasowego na wyniki poniżej pasma 1/3 oktawy 500 Hz był relatywnie mały w stosunku do niepewności metody pomiarowej podawanej w normie PN-EN 1793-5, zwłaszcza w przypadku próbki mocno pochłaniającej.

Rys. 7. Wpływ długości okna czasowego na RI

5. PODSUMOWANIE

W pracy przedstawiono wyniki pomiarów wskaźnika odbicia dźwięku RI dla dwóch elementów budowlanych o zróżnicowanych właściwościach pochłaniania dźwięku. Badano płaskie elementy o jednorodnych powierzchniowo właściwościach pochłaniających. Okazuje się, że metoda stawia szczególne wymagania co do niektórych elementów systemu pomiarowego, zwłaszcza jeśli badane są elementy o dużych wartościach RI . Tymi elementami są głośnik i rama, na której mocowane są mikrofony pomiarowe. Norma wprowadzie zaznacza, że charakterystyka częstotliwościowa głośnika ma być wyrównana, ale nie określono tych wymagań precyzyjnie. Jak pokazano w pracy, nawet zastosowanie głośnika o nierównomiernościach ± 4 dB w zakresie od 0,5 do 2 kHz może wprowadzać w tym zakresie znaczne błędy pomiarowe. Oznacza to, że należałoby dokładniej określić wpływ właściwości głośnika na uzyskiwane wyniki i doprecyzować w tym zakresie wymagania. Wprowadzie norma wyklucza stosowanie aktywnych bądź pasywnych elementów, które mogą wpływać na pasmo przenoszenia systemu, ale dobrym rozwiązaniem wydaje się cyfrowa korekcja zniekształceń liniowych wykorzystywanego głośnika. Sprawdzenie skuteczności takiego rozwiązania planowane jest przez autora w przyszłości.

W normie nie zaznaczono, że istotnym elementem systemu jest rama, na której mocowane są mikrofony pomiarowe, a jak pokazano w pracy jej konstrukcja może mieć istotne znaczenie.

Wprawdzie badano obiekt o bardzo specyficznych właściwościach akustycznych (mocno odbijająca podłoga i mocno pochłaniające pozostałe ściany), ale wydaje się, że metodyka pomiarowa przedstawiona w normie PN-EN 1793-5 może być wykorzystywana nie tylko do oceny właściwości pochłaniających ekranów akustycznych, ale również do oceny w warunkach terenowych płaskich elementów budowlanych o jednorodnych powierzchniowo właściwościach pochłaniających. W przypadku elementu o silnych właściwościach pochłaniających uzyskano istotne rozbieżności zmierzonych wartości RI z $1-\alpha$ w zakresie częstotliwości poniżej 630 Hz. Wynika to między innymi z różnic w określaniu pogłosowych i fizycznych właściwości pochłaniających, z niedokładności przyjętych jako odniesienie wartości $1-\alpha$ (różne właściwości ścian i sufitu) oraz ze znacznej głębokości badanej struktury.

Istotnym ograniczeniem metody są wymagania co do wymiarów badanego elementu i jego otoczenia. W przypadku mniejszych elementów, lub kiedy w pobliżu mikrofonów znajdują się elementy mogące być źródłem znaczących odbić, konieczne jest stosowanie okna czasowego o długości ograniczającej zakres analizowanych częstotliwości. Okazało się jednak, że przypadku elementu mocno pochłaniającego wpływ długości okna czasowego był relatywnie mały w stosunku do niepewności pomiaru omawianej metody. Wydaje się, że kiedy charakterystyka częstotliwościowa RI nie zawiera wyraźnych nierównomierności to do orientacyjnej oceny właściwości pochłaniających badanego elementu można wykorzystywać również wyniki uzyskane w zakresie częstotliwości poniżej częstotliwości granicznej wynikającej z długości okna czasowego.

LITERATURA

- [1] GARAI M., Measurement of the sound-absorption coefficient in situ: The reflection method using periodic pseudo-random sequences of maximum length, *Applied Acoustics* 39:119-139 (1993).
- [2] GARAI M., GUIDORZI P., In situ measurements of the intrinsic characteristics of the acoustic barriers installed along a new high speed railway line. *Noise Control Eng. J.* 56 (2008) 342-355.
- [3] GARAI M., GUIDORZI P., Sound reflection measurements on noise barriers in critical conditions, *Building and Environment*, Volume 94, Part 2, December 2015, Pages 752-763.
- [4] GUIDORZI P., KLEPÁČEK J., GARAI M., On the repeatability of reflection index measurements on noise barriers. *Proceeding of Euronoise 2012*, Prague, 2012, 1314–1319.
- [5] GUIDORZI P., GARAI M., Advancements in sound reflection and airborne sound insulation measurement on noise barriers. *Open Journal of Acoustics* 3 (2013) 25–38.
- [6] MOMMERTZ E. Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique” -. *Appl. Acoust.* 1995, 46 pp. 251–263.
- [7] MÜLLER S., MASSARANI P., Transfer-Function Measurement with Sweeps, *J. Audio Eng. Soc.*, Vol.49, No.6, s. 443 471, 2001 June.
- [8] PN-EN 1793-1 -- Drogowe urządzenia przeciwhałasowe -- Metoda oznaczania właściwości akustycznych -- Część 1: Podstawowe właściwości pochłaniania dźwięku w warunkach rozproszonego pola akustycznego
- [9] PN-EN 1793-5:2016-05+AC:2018-08 -- Drogowe urządzenia przeciwhałasowe -- Metoda oznaczania właściwości akustycznych -- Część 5: Właściwości wewnętrzne -- Wartości odbicia dźwięku w warunkach bezpośredniego pola akustycznego w miejscu zamontowania
- [10] PN-EN 61672-1:2014-03 -- Elektroakustyka -- Mierniki poziomu dźwięku -- Część 1: Wymagania
- [11] PN-EN ISO 354:2005 -- Akustyka -- Pomiar pochłaniania dźwięku w komorze pogłosowej
- [12] PN-EN ISO 10534-2:2003 -- Akustyka -- Określanie współczynnika pochłaniania dźwięku i impedancji akustycznej w rurach impedancyjnych -- Część 2: Metoda funkcji przejścia.
- [13] SIPAR P., JALKANEN T., SIPONEN D., Sound reflection from different noise barriers, *Research reports from the Finnish Transport Agency*, 29eng/2017, Helsinki 2017
- [14] TRONCHIN L., VENTURI A., FARINA A., VARANI C., In situ measurements of Reflection Index and Sound Insulation Index of noise barriers, *Proceedings of 20th International Congress on Acoustics, ICA* 2010, Sydney, Australia

PORÓWNANIE SKUTECZNOŚCI KLASYFIKACJI GATUNKÓW MUZYCZNYCH W PRZY POMOCY SIECI NEURONOWEJ W ZALEŻNOŚCI OD TYPU DANYCH WEJŚCIOWYCH

Maciej BLASZKE^{1,2}

Damian KOSZEWSKI^{1,2}

Bożena KOSTEK²

¹Katedra Systemów Multimedialnych, Wydział Elektroniki, Telekomunikacji i Informatyki,
Politechnika Gdańska, Narutowicza 11/12, 80-233 Gdańsk

²Laboratorium Akustyki Fonicznej, Wydział Elektroniki, Telekomunikacji i Informatyki,
Politechnika Gdańska, Narutowicza 11/12, 80-233 Gdańsk

mblaszke@multimed.org; damkosze@multimed.org; bokostek@audioakustyka.org

Rozpoznawanie gatunku muzycznego jest jednym z podstawowych elementów inteligentnych systemów tworzenia automatycznych list muzyki. Platformy strumieniowe oferujące taką usługę wymagają rozwiązań, które pozwolą jak najdokładniej określić przynależność utworu do gatunku muzycznego. Aktualny stan wiedzy wskazuje, że najskuteczniejszym klasyfikatorem są sztuczne sieci neuronowe (w tym w wersji uczenia głębokiego), dla których wejście stanowi spektrogram (postać 2D wektora wejściowego), współczynniki MFCC czy wektor parametrów.

Ogólnodostępne narzędzia (np. MIRTtoolbox, czy Python/libROSA) pozwalają na wyznaczenie redundantnej listy parametrów. Wynikowy wektor jest zawsze nadmiarowy ze względu na obecność wysoce skorelowanych parametrów. Przykładem dokładnie wyspecyfikowanego wektora parametrów może być standard MPEG-7 opisany w normie ISO/IEC 15938. Jego zastosowanie pozwala uzyskać nawet 100% skuteczność w kontekście klasyfikacji instrumentu muzycznego. Zawiera on w sobie deskryptory opisujące zarówno przebieg czasowy sygnału, jak i jego strukturę. Można je podzielić na 6 głównych grup:

1. *Basic* - bazujące na wartości próbek sygnału fonicznego (*AudioWaveform Descriptor*, *AudioPower Descriptor*)
2. *Basic Spectral* - Prosta czasowo-częstotliwościowa analiza sygnału (*AudioSpectrumEnvelope*, *AudioSpectrumCentroid*, *AudioSpectrumSpread*, *AudioSpectrumFlatness*)
3. *Spectral Basis* - jednowymiarowa widmowa projekcja sygnał przygotowana w szczególności w celu łatwiejszej klasyfikacji sygnału (*AudioSpectrumBasis*, *AudioSpectrumProjection*)
4. *Signal Parameters* - Informacja o okresowości sygnału (*AudioFundamentalFrequency*, *AudioHarmonicity*)
5. *Timbral Temporal* - charakterystyka czasowo oraz barwa muzyczna (*LogAttackTime*, *TemporalCentroid*)
6. *Timbral Spectral* - deskryptory opisujące zależności liniowo-częstotliwościowe w sygnale (*SpectralCentroid*, *HarmonicSpectralCentroid*, *HarmonicSpectralDeviation*, *HarmonicSpectralSpread*, *HarmonicSpectralVariation*).

Główną motywacją do prowadzenia badań jest potrzeba każdorazowego tworzenia narzędzi do wyznaczania deskryptorów standardu MPEG-7. Istniejące oprogramowanie (różne środowiska programistyczne) zwykle nie zapewnia pełnej możliwości ich wykorzystania. Ponadto, często stosuje się dodatkowe, tzw. dedykowane do rozwiązywanego problemu deskryptory. Kolejnym problemem, który się pojawia jest nadmiarowość wektora cech. Aby uzyskać wysoką skuteczność klasyfikacji niezależnie od stosowanego algorytmu, należy dodatkowo stosować metody pozwalające na redukcję proponowanych

cech (np. analizę składowych głównych, PCA – *Principal Component Analysis* czy analizę korelacyjną). Autorzy poprzednich prac opracowali sztuczną sieć neuronową, która z 5% błędem je wyznacza. Zaproponowaną sieć można wykorzystać zarówno jako niezależną metodę wyznaczania parametrów, jak również jako część bardziej złożonej architektury, np. jako jej wejście. Przykładem takiej sieci może być między innymi klasyfikator gatunku muzycznego.

W pracy została wykorzystana baza danych SYNAT zawierająca przykłady utworów muzycznych wraz z wyznaczonymi dla nich zgodnie z definicją deskryptorami MPEG-7 oraz dokładną informację o gatunku do którego należy dane nagranie. Ze względu na ograniczenia sieci neuronowej do aproksymacji niskopoziomowych deskryptorów do dalszych eksperymentów wybrano tylko nagrania o długości przynajmniej 8192 próbek. Następnie dokonano losowego podziału na trzy podzbiory: treningowy, walidacyjny i ewaluacyjny.

Do porównania wykorzystano 5 typów danych reprezentujących sygnał:

1. Przebieg czasowy;
2. Spektrogram;
3. Współczynniki MFCC;
4. Deskryptory MPEG-7 z bazy SYNAT;
5. Deskryptory MPEG-7 wyznaczone przy pomocy sieci neuronowej.

Dla całego wybranego zbioru wyznaczono wszystkie wymienione wyżej typy danych wykorzystując 8192 próbek ze środka nagrania. Dalsze eksperymenty wykonano z wykorzystaniem głębokiej sieci neuronowej dokonującej klasyfikacji do którego gatunku należy dane nagranie. W każdym przypadku budowa sieci była jednakowa w obrębie warstw ukrytych, a różnice były w obrębie warstw wejściowych, co było wymagane ze względu na różną długość i typy wektora danych.

ABSTRACT

Genre recognition is one of the core components of intelligent automatic music list creation systems. Streaming platforms offering such a service require solutions that will allow for the most accurate determination to what music genre a given song belongs. The current state of the art shows that the most effective classifier is an artificial neural network (including its deep learning version). In such cases, the input is often a 2d representation, i.e., a spectrogram, MFCC coefficients, or a feature vector. In the previous work, the authors developed an artificial neural network which, with a 5% error, determines a set of the MPEG-7 descriptors. They can be used, among other things, as input data to the music genre classifier. This work presents a comparison of the effectiveness of such a classifier depending on the type of input data such as raw waveform or spectrogram representation, MFCC, feature vector containing MPEG-7 standard low-level descriptors, or descriptors calculated with the use of a deep neural network.

AUTOMATYCZNE GENEROWANIE KOLEJNOŚCI LIST UTWORÓW MUZYCZNYCH

Kamila PIETRUSIŃSKA¹

Adam KUROWSKI^{1,2}

Bożena KOSTEK²

¹ Katedra Systemów Multimedialnych, Wydział Elektroniki, Telekomunikacji i Informatyki,
Politechnika Gdańska, ul. Gabriela Narutowicza, 80-233 Gdańsk

² Laboratorium Akustyki Fonicznej, Wydział Elektroniki, Telekomunikacji i Informatyki,
Politechnika Gdańska, ul. Gabriela Narutowicza, 80-233 Gdańsk

Celem pracy jest przygotowanie algorytmu do automatycznego układania kolejności utworów muzycznych i zgrywającego je do postaci jednego, długiego miksu. Algorytm dobiera utwory na podstawie podobieństwa fragmentów końcowych i początkowych utworów. Podobieństwo to jest obliczane za pomocą odległości Euklidesowej pomiędzy wektorami parametrów obliczonymi przez autoenkoder oraz na podstawie analizy skupień (ang. data clustering). Taki sposób ułożenia utworów pozwala na zapewnienie ciągłości listy. Jakość wyników jest weryfikowana za pomocą testów odsłuchowych poprzez porównanie automatycznie generowanych list z listami ułożonymi w sposób losowy.

1. WPROWADZENIE

Wraz ze wzrostem popularności serwisów streamingowych i udostępnianymi listami odtwarzania utworów muzycznych przedstawianymi użytkownikom serwisu wyznaczanie kolejności utworów w tego typu listach jest bardzo istotnym elementem ich funkcjonowania i popularności. Najczęściej stosowane metody tworzenia list utworów opierają się na analizach statystycznych, u podstaw których leżą metody eksploracji dużych zbiorów danych opisujących muzyczne preferencje użytkowników. Szeroko stosowaną metodą łączenia utworów jest mechanizm „filtracji społecznościowej” (ang. *collaborative filtering*). Są to metody, które należą do technik tzw. zbiorowej filtracji oraz metod eksploracji danych. Metody te mogą niestety powodować, że utwory artystów mało znanych nie pojawią się nigdy na wysokim miejscu listy odtwarzania. Alternatywą dla takiego podejścia może być wykorzystanie sztucznych sieci neuronowych. Pozwalają one na przeprowadzenie inteligentnej analizy posiadanej bazy danych utworów muzycznych. Porządek utworów jest określany na przykład na podstawie analizy parametrów utworów obliczonych dla ich początkowych i końcowych fragmentów, które następnie są dopasowywane na podstawie odległości pomiędzy nimi. Jednym z paradygmatów uczenia maszynowego, które może być szczególnie przydatne w tym celu jest uczenie głębokie. Jest ono szczególnie przydatne między innymi w takich obszarach jak rozpoznawanie głosu, obrazu czy też przetwarzanie języka naturalnego. Odbywa się to poprzez wykorzystanie nowych technik uczenia sieci i nowych architektur połączeń sztucznych neuronów, potrafiących między innymi w automatyczny sposób dokonywać ekstrakcji parametrów ze zbioru uczącego podanego na wejście algorytmu. Dlatego głębokie uczenie nadaje się do takich zadań, jakim jest m.in. automatyczne układanie list odtwarzania.

Celem niniejszej pracy jest przedstawienie procesu automatycznego układania list utworów z wykorzystaniem uczenia głębokiego. Weryfikację jakości tworzonych list odtwarzania sprawdzono przeprowadzając testy odsłuchowe, w czasie których badani porównywali listy stworzone przez algorytm z

listami przygotowanymi przez generator liczb losowych. Uzyskane wyniki zweryfikowano za pomocą testów statystycznych.

2. AUTOMATYCZNE UKŁADANIE LIST MUZYCZNYCH

Problem automatycznego tworzenia list odtwarzania jest często podejmowany, m.in. z uwagi na bardzo rozwinięte platformy streamingowe. Bardzo popularną funkcjonalnością jest zapewnienie użytkownikowi specjalnie dla niego przygotowanych list utworów, które są dopasowane do siebie nastrojem, tempem lub zawierają pozycje, które mogą wpasować się w upodobania muzyczne użytkownika, biorąc pod uwagę inne odtwarzane utwory i kontekst przypisanego do nich albumu muzycznego. Problem automatycznego tworzenia list odtwarzania pojawia się w literaturze badawczej głównie ze względu na fakt [7], iż nie znaleziono rozwiązania dającego w pełni satysfakcjonującą efektywność. Głównym powodem takiego stanu rzeczy jest subiektywność postrzegania muzyki. Jednak obecnie listy takie można tworzyć za pomocą uczenia maszynowego (w tym uczenia głębokiego, np. za pomocą sieci neuronowych typu autoenkoder[2]). Dlatego warto ten temat zgłębiać przy zastosowaniu również tego typu narzędzi i algorytmów.

Oord van den A., Dieleman S., Schrauwen B. w swojej pracy [22] wykorzystali bazę danych One Million Songs [4] oraz dwa główne podejścia w obszarze MIR (*Music Information Retrieval*), czyli pozyskanie lokalnych cech sygnału fonicznego oraz wykorzystanie uczenia głębokiego. Obliczono 13 współczynników MFCC (*Mel-Frequency Cepstral Coefficients*, *mfcc*) z sygnału przy segmentacji na okna wielkości 1024 próbek, odpowiadającym 23 ms przy szybkości próbkowania równej 22,05 kHz. Ponadto obliczono różnice pomiędzy parametrami mel-cepstralnymi pierwszego i ostatniego rzędu, uzyskując łącznie 39 parametrów. Następnie wytrenowano słownik 4000 elementów za pomocą algorytmu K-średnich i przypisano wszystkie wektory MFCC do najbliższej średniej. Dla każdego utworu obliczono, ile razy pojawiła się dana średnia. Dane te zapisano do wektora, który wykorzystuje model „*bag-of-words*”. Następnie wykorzystano sieci głębokie, które wytrenowano na trzech sekundach utworu wybranych w sposób losowy. Autorzy przywołanej pracy zaobserwowali, że wiele aspektów dotyczących rekomendacji utworów charakteryzuje się subiektywnym podejściem, ale niektóre cechy mogą być przewidziane na podstawie sygnału fonicznego. Innym rozwiązaniem jest praca z 2017 opisująca algorytm automatycznego DJa [13]. Stworzone rozwiązanie w pełni automatyczny sposób tworzy listy odtwarzania na podstawie narzuconego zbioru utworów. Zapewnia również płynne przejście pomiędzy kolejnymi utworami. Parametry sygnału, wykorzystane do wyboru kolejnego utworu, są ekstrahowane za pomocą metod wykorzystujących sieci głębokie wytrenowane do rozpoznawania gatunku muzycznego. Kolejnym rozwiązaniem, które można znaleźć w literaturze w kontekście automatycznego rozpoznawania muzyki jest wykorzystanie kilku klasyfikatorów, w tym modelu sieci generatywnych z siecią przeciwstawiającą się [6]. Jako narzędzie parametryzacji wykorzystano w tym przypadku sieć neuronową. Stworzone zostały dwa systemy, jeden oparty na sztucznych sieciach neuronowych, gdzie zbiorem danych była baza GTZAN [8][21]) i drugi wykorzystujący sieci splotowe - w tym przypadku zbiorem danych była baza Latin Music Database [14]) Wykorzystano klasyfikator bazujący na lasach losowych (ang. Random Forest) w celu zbadania skuteczności tych dwóch różnych klasyfikacji. Stwierdzono, że lepszą skuteczność rozpoznawania muzyki uzyskano dla rozwiązania wykorzystującego sieci splotowe.

Przytoczone algorytmiczne techniki układania list utworów muzycznych mogą znaleźć swoje zastosowanie przede wszystkim w aplikacjach związanych z rozrywką. Przykładem takiego zastosowania może być radio – zarówno to klasyczne, jak i internetowe. Innym możliwym zastosowaniem jest automatyczne układanie list utworów muzycznych stanowiących podkład dla interaktywnej rozrywki, takiej jak na przykład gry komputerowe. W takim przypadku muzyka musi nie tylko być ułożona w logiczny sposób, ale także odpowiadać i wzmacniać efekt artystyczny tworzony przez grę i jej stronę wizualną oraz narracyjną [19].

3. ZAŁOŻENIA PROJEKTOWE

Zaprojektowanie algorytmu doboru utworów muzycznych wymaga podzielenia tego procesu na kilka etapów. Częstym wyborem jest wykorzystanie algorytmów heurystycznych, w tym - sieci neuronowych. Najpierw konieczne jest zidentyfikowanie zbioru danych, który stanowić będzie podstawę dla trenowania sieci, następnie konieczny jest wybór i opracowanie metod wstępnego przetwarzania danych. Konieczne jest także określenie według jakich reguł sieć taka - zgrubnie - powinna się kierować w procesie układania list. Wybór ten będzie rzutować na dokładny wybór algorytmu, który najlepiej będzie dopasowany właśnie do preferowanego sposobu układania list utworów muzycznych [16].

Wraz ze zwiększeniem wykorzystania sieci neuronowych i uczenia maszynowego w obszarze akustyki muzycznej można znaleźć obecnie wiele baz danych muzycznych, z których większość jest ogólnie dostępna i darmowa. W związku z tym, że projektowany algorytm wykorzystuje do działania nieprzetworzone utwory muzyczne, zrezygnowano z baz dostarczających wektory parametrów, do których zalicza się m.in. bardzo popularna baza The Million Songs Dataset [1]. Dla uzyskania zadowalających efektów treningu, rozważano zbiory danych, które zapewnią dużą ilość danych, które są jednocześnie zróżnicowane. Właśnie dlatego wybrano jedną z pierwszych baz utworów, jakie zostały stworzone, a mianowicie bazę GTZAN [8], [21]. Podzielona jest ona na dziesięć gatunków muzycznych: blues, muzykę poważną, country, disco, hiphop, jazz, metal, pop, reggae i rock. W każdym gatunku przygotowano sto 30-sekundowych fragmentów.

Pozyskany w taki sposób sygnał muzyczny musi być poddany procesowi parametryzacji, czyli wyodrębnienia cech (zwanych także parametrami). Często cechy te mogą one być wykorzystywane np. do klasyfikacji gatunku muzycznego. Dobór parametrów ma znaczący wpływ na efektywność opracowywanego algorytmu. Przykładowo, ważne jest, aby poszczególne składowe wektora nie były ze sobą skorelowane, gdyż może potencjalnie negatywnie wpłynąć to na efekty osiągane z pomocą uczenia maszynowego, np. poprzez obniżenie skuteczności klasyfikacji gatunku muzycznego, czy znajdowanie mniej podobnych do siebie utworów muzycznych [3]. Wektor parametrów może zawierać parametry czasowe, częstotliwościowe, jak również parametry statystyczne czy tzw. dedykowane, bowiem należy pamiętać, że zawartość wektora jest dobierana bezpośrednio do potrzeb algorytmu. Grupa parametrów czasowych jest wyznaczana na podstawie postaci czasowej sygnału, natomiast parametry częstotliwościowe wyznaczane są na podstawie estymacji widma sygnału [10]. Duża część tych parametrów została zdefiniowana w standardzie MPEG 7 [15]. W przypadku parametrów częstotliwościowych bardzo często konieczne jest zdefiniowanie podpasem częstotliwościowych, w których wyznaczany będzie dany parametr. Parametry obliczone na podstawie sygnałów muzycznych mogą posłużyć na przykład do układania utworów z danej bazy w optymalnej kolejności odsłuchu.

Listy takie nie są tworzone tylko i wyłącznie na własne potrzeby użytkowników portali streamingowych [20] czy osób słuchających muzyki. Stanowią one podstawę pracy realizatorów emisji czy redaktorów pracujących w rozgłośniach radiowych, DJów, jak również osób przygotowujących oprawę muzyczną uroczystości czy innych wydarzeń. Listy odtwarzania są wykorzystywane w różnych obszarach, np. w trakcie zakupów w sklepie. Odpowiednia muzyka wpływa na nastrój klienta i może podświadomie spowodować, że kupujący dokona większych zakupów niż zamierzał. Skupiając się głównie na tworzeniu oprawy muzycznej do audycji radiowych, wydarzeń czy w pracy DJa, należy zauważyć, że ważnym elementem, który trzeba wziąć pod uwagę przy tworzeniu listy jest pora dnia, w czasie której będzie ona odtwarzana [2], [5]. Muzyka do audycji radiowych wybierana jest w taki sposób, aby w czasie emisji nocą utwory były dość stonowane, rezygnuje się z mocnych np. metalowych brzmień. W czasie, kiedy najwięcej osób słucha radia – w czasie drogi do pracy, przerwy obiadowej czy w drodze do domu wybiera się muzykę zdecydowanie bardziej dynamiczną i energiczną. Wyjątkami od tych reguł są autorskie audycje, często muzyczne, w czasie których redaktorzy dobierają muzykę zgodnie z wybranym przez siebie tematem. Duża liczba portali i aplikacji streamingowych [20] pozwala na tworzenie list odtwarzania i dzielenie się nimi z wszystkimi użytkownikami. Dostępne są więc playlisty tworzone do odtwarzania w konkretnej sytuacji. W czasie ich tworzenia rezygnuje się z wcześniej przytoczonych

reguł przyjętych w rozgłośniach radiowych, DJów czy osoby tworzące oprawę muzyczną do uroczystości. W przypadku list przygotowanych do odtwarzania w konkretnym celu przy wyborze utworów kieruje się tylko ich dynamiką i nastrojem. Przykładowo lista odtwarzania stworzona do słuchania w czasie treningu będzie skupiała się na utworach szybkich i dynamicznych, często rockowych. Natomiast lista przygotowana w celu wyciszenia i odpoczynku będzie zawierała utwory jazzowe lub klasyczne; często muzykę instrumentalną [2], [5].

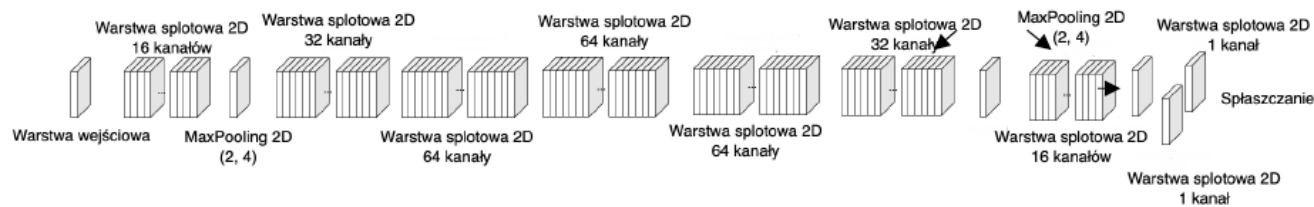
W niniejszej pracy do układania list utworów muzycznych wykorzystana została sieć neuronowa o strukturze autoenkodera. Pozwala ona na nienadzorowaną naukę parametryzacji podanego na jej wejście zbioru wartości. Często w procesie nauki tego zadania sieć jest w stanie zidentyfikować zależności pomiędzy danymi, w tym zidentyfikować podobieństwa pomiędzy przykładami podanymi na jej wejście. Do tego celu szczególnie nadaje się algorytm autoenkodera wariacyjnego, który może być wykorzystywany także jako algorytm generatywny. W prezentowanym algorytmie autoenkoder wykorzystany jest do identyfikowania tych parów utworów, dla których zakończenie pierwszego utworu i zakończenie drugiego utworu są do siebie podobne. Algorytm układania utworów muzycznych wykorzystuje autoenkoder jako metodę parametryzacji utworów muzycznych. Algorytm analizuje z jego pomocą 10-sekundowe fragmenty początków i końców utworów znajdujących się w bazie. Proces dopasowania utworów polega na obliczeniu odległości Euklidesowej między wartościami odpowiadającymi końcowi aktualnie analizowanego utworu, a wartościami początków dostępnych utworów. Efektem działania skryptu jest pliki tekstowy z zapisanymi ścieżkami utworów ułożonymi w obliczonej przez skrypt kolejności oraz plik w formacie m3u, który pozwala na odtworzenie ułożonych utworów zapisanych w kolejności pokrywającej się z plikiem tekstowym w odtwarzaczu multimedialnym.

4. ARCHITEKTURA SIECI NEURONOWEJ (AUTOENKODERA WARIACYJNEGO)

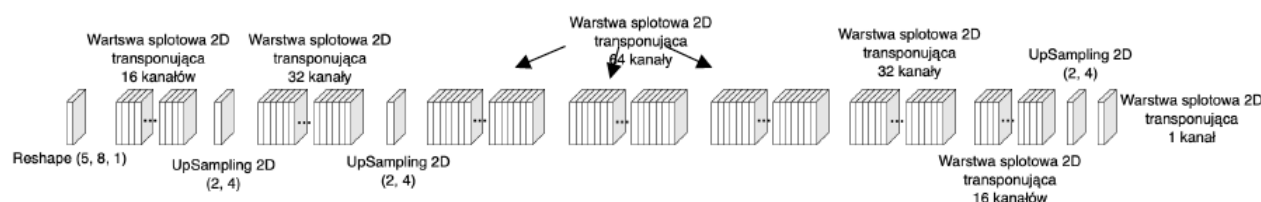
Jak już wcześniej wspomniano, do wyznaczenia kolejności utworów na listach odtwarzania wykorzystano autoenkoder wariacyjny. Jedną z jego cech jest dążenie do umieszczania w przestrzeni reprezentacji podobnych fragmentów muzyki blisko siebie, co jest zgodne z algorytmem układania list utworów muzycznych, który jako zasadę swojego działania wykorzystuje umieszczanie obok siebie podobnych do siebie zakończeń i początków utworów. Ten specyficzny sposób układania punktów w przestrzeni decyzyjnej wynika między innymi z zastosowania dystansu Kullbacka-Leiblera w funkcji straty wykorzystywanej w trakcie treningu sieci. Architektura przygotowanej sieci działa dla konkretnych danych wejściowych, z których część została dobrana tak, aby zgodna była z bazą danych, która stanowi zbiór uczący [8]. Algorytm przetwarza monofoniczne pliki zapisane w formacie wav, dla których szybkość próbkowania wynosi 22 050 Hz. Dane przeznaczone do wprowadzenia na wejście autoenkodera przeliczane są na mfcc-gramy o długości 10 sekund (co odpowiadało długości 512 próbek) i rozdzielczości 40 współczynników mel-cepstralnych, obliczone wartości znormalizowano do zakresu od -1 do 1. Autoenkoder składa się z dwóch części – koder, który dokonuje przekształcenia mfcc-gramu na wektor reprezentacji o długości 40 współczynników i dekodek, który rekonstruuje mfcc-gram na podstawie wynikowych wektorów uzyskiwanych z koder. W trakcie tworzenia architektury sieci sprawdzano jakość dekodowania, jaka została uzyskana na wyjściu autoenkodera. Poniżej na rysunkach 4.1 i 4.2 przedstawiono strukturę, dla której uzyskano najlepsze wyniki.

Drugą ważną cechą autoenkodera wariacyjnego jest fakt, że koder posiada dwa wyjścia. Takie rozwiązanie jest spowodowane tym, że zamiast konkretnych wartości reprezentacji, koder zwraca rozkład prawdopodobieństwa wartości reprezentacji powiązanych z mfcc-gramem na jego wejściu. Koder ten na swoim wyjściu generuje dwa wektory, jeden z nich reprezentuje wartości średnie, a drugi wariancje. Opisują one rozkład wartości składowych wektora losowego, którego realizacja następnie stanowi wejście dekodek. Funkcja straty wykorzystywana w trakcie treningu składa się z kolei z dwóch sumowanych ze sobą elementów. Pierwszym z nich jest wspomniany wcześniej dystans Kullbacka-Leiblera pomiędzy rozkładem występowania punktów kodowanych przez koder a równomiernym rozkładem

punktów co pozwala właśnie na wymuszenie takiego rozkładu punktów w przestrzeni reprezentacji. Drugim składnikiem funkcji straty jest błąd rekonstrukcji mfcc-gramu na wyjściu względem oryginału przekazanego na wejście. Bardziej szczegółowy opis działania autoenkoderów wariacyjnych może być znaleziony w licznych publikacjach literaturowych [4], [13].



Rysunek 4.1 Struktura enkodera



Rysunek 4.2 Struktura dekodera

Ostatnim etapem realizacji algorytmu układającego listy odtwarzania jest skrypt, który wykorzystuje stworzone przez autoenkoder wektory parametrów. Dla każdego utworu generowany jest wektor składający się z 40 wartości będących reprezentacją jego początkowego fragmentu o długości 20 sekund. Analogiczny wektor generowany jest dla końcowego fragmentu utworu. Następnie obliczana jest odległość Euklidesowa pomiędzy końcem a początkiem kolejnego utworu w 40-wymiarowej hiperprzestrzeni utworzonej przez parametry zwracane przez autoenkoder. Kończącym wynikiem działania skryptu są dwa pliki, jeden tekstowy z zapisanymi ścieżkami do utworów, drugi w formacie m3u pozwalający na odtworzenie listy w odtwarzaczu multimedialnym.

5. TESTY SUBIEKTYWNE

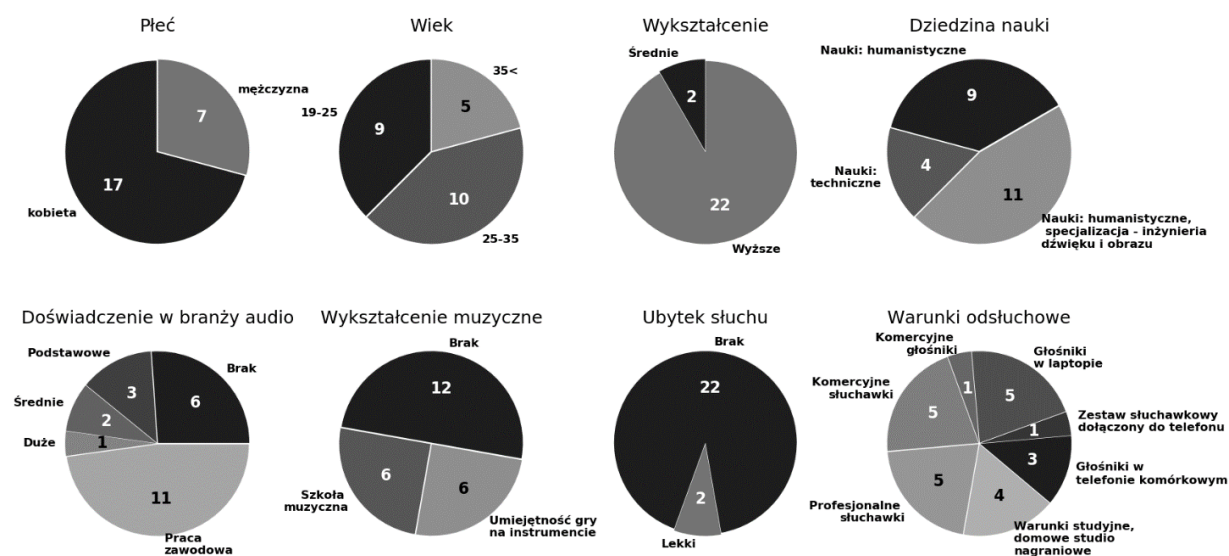
Celem zweryfikowania efektów, jakie uzyskano za pomocą stworzonego skryptu przeprowadzono ankietę internetową, w której respondenci udzielali odpowiedzi po wysłuchaniu dostarczonych przykładów na własnym sprzęcie. Odpowiedzi wraz z podstawowymi informacjami o ankietowanym i wykorzystywanym przez niego sprzęcie były zbierane automatycznie do arkusza kalkulacyjnego. Badania zostały wykonane w kilku grupach odbiorców różniących się płcią, wiekiem, wykształceniem oraz doświadczeniem w branży audio. Możliwe było również rozróżnienie wyników pod względem warunków odsłuchowych. Test odsłuchowy polegał na subiektywnej ocenie stworzonych list odtwarzania. Przygotowano 5 zestawów list, w których jedna stworzona była przez algorytm, a druga wylosowana generatorem liczb losowych. W celu ułatwienia zadania postanowiono wybrać utwory, które są dość popularne lub dobrze reprezentujące dany gatunek muzyczny. W uwagi na fakt, iż celem testu jest sprawdzenie, w jakim stopniu koniec jednego utworu łączy się z początkiem kolejnego, wycięto 15-sekundowe fragmenty z początku i końca każdego utworu i połączono je we właściwych kolejnościach.

Test zaprojektowano w oparciu o zalecenia standardu ITU-R BS 1284-1 [11]. Każdy lista zawiera osiem utworów, a odsłuchanie jednego zestawu (listy ułożonej przez algorytm, drugiej w losowej kolejności) trwa około 8,5 minuty. Oprócz przygotowanych zestawów zapewniono badanym możliwość

odsluchania całych list przez zapewnienie linków do list odtwarzania stworzonych na platformie YouTube. Ankietowani poproszeni byli o porównanie tego, w jakim stopniu dopasowane są do siebie początki i końce utworów. Przedstawiony materiał oceniany był za pomocą zmodyfikowanej skali z zalecenia standardu ITU-R BS 1284-1 [11]. Do wyboru możliwe były następujące opcje:

- -2 utwory na liście I są lepiej ułożone,
- -1 utwory na liście I są trochę lepiej ułożone,
- 0 utwory na obu listach są dobrze ułożone,
- 1 utwory na liście II są trochę lepiej ułożone,
- 2 utwory na liście II są lepiej ułożone.

Postanowiono zmniejszyć stopnie w skali porównawczej podanej w zaleceniu, aby ułatwić badanym proces podejmowania decyzji i możliwie skrócić czas trwania testu, którego odsłuchanie zajmowało około 45 min. Sama długość testu nie była ograniczana i ankietowani mogli odsłuchiwać listy dowolną liczbą razy. Pierwszą listę zawsze stanowiła lista stworzona przy pomocy algorytmu. Respondenci nie posiadali informacji o tym, która lista jest listą stworzoną przez algorytm, a która losowa. Poinformowani byli jednak, że zawsze jedna lista jest stworzona przez wylosowanie utworów generatorem liczb losowych, a druga przez stworzony algorytm, który dobiera utwory, dopasowując ostatnie 10 sekund do pierwszych 10 sekund kolejnego utworu. Wybierając utwory do list odtwarzania, oprócz tego, że w większości wybierano kompozycje uważane za popularne, dobierano je również tak, aby pomimo tego, że lista jest krótka, miała zróżnicowaną dynamikę, tematyka miała możliwość stopniowej zmiany, a utwory nie zawierały się tylko w obrębie jednego gatunku muzycznego lub bardzo do niego zbliżonych. W badaniu łącznie udział wzięły 24 osoby: 6 kobiet i 18 mężczyzn. Szczegóły zebrane na temat badanych przedstawiono na rysunku 5.1.



Rysunek 5.1 Wykresy przedstawiające dane zebrane o badanych, tj. o wykształceniu, doświadczeniu w branży audio, wykształceniu muzycznym, ubytku słuchu i warunkach odsłuchowych.

6. WYNIKI TESTÓW SUBIEKTYWNYCH

Efektom przeprowadzonych ankiet było pozyskanie zbioru 24 odpowiedzi oceniających listy w skali dyskretnej, całkowitoliczbowej od -2 (preferowana jest lista stworzona przez algorytm) do 2 (preferowana jest lista stworzona przez generator liczb losowych). Do analizy statystycznej wykorzystano test Kruskala-Wallisa, oraz test χ^2 . Wszystkie wnioski zostały oparte o założenie, że poziom istotności ma standardową wartość wynoszącą 0,05. Test Kruskala-Wallisa został wybrany z tego względu, że dla

wszystkich rozpatrywanych przypadków nie były spełnione założenia pozwalające na zastosowanie testu ANOVA [20]. Sprawdzenie dokonane zostało testem Levene'a (równość wariancji) i Shapiro-Wilka (na normalność ocen) [20]. Test Shapiro-Wilka był przeprowadzany dla każdego zestawu ocen oddzielnie, więc w jego przypadku zastosowano poprawkę na wielokrotne testowanie (metodą Simesa-Hochberga). W celu implementacji testów wykorzystano język Python i pakiet SciPy [12].

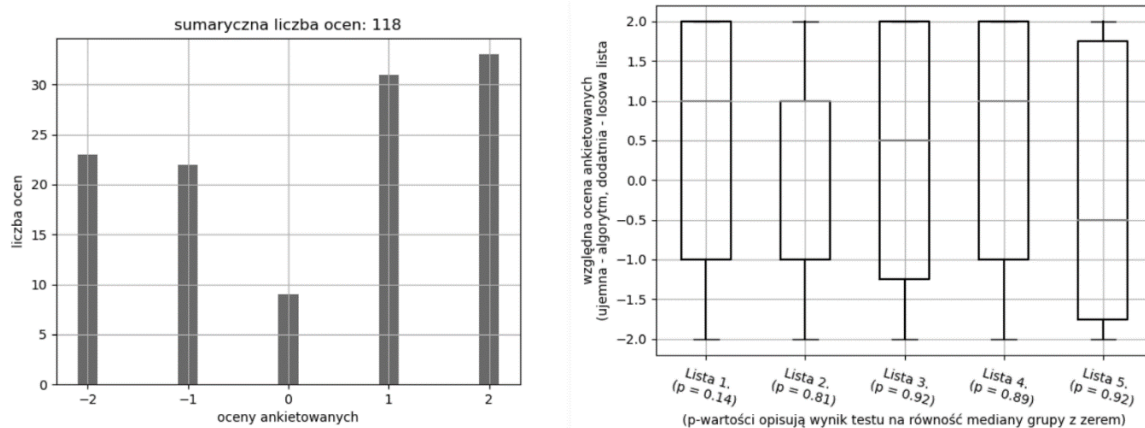
Pierwszą analizę przeprowadzono na całym zbiorze badanych, nie dzieląc ankietowanych zgodnie z ich wykształceniem czy doświadczeniem w branży audio. Analiza przeprowadzona została za pomocą testu Kruskala-Wallisa. Wynikowa p -wartość uzyskana w wyniku testu wyniosła 0,691. Wykazała ona zatem, że badani nie mają preferencji w kontekście czy wolą listy układane przez zaproponowany algorytm, czy listy układane w kolejności losowej. Następnie przeprowadzony został test, mający sprawdzić czy ankietowani słyszeli różnicę pomiędzy algorytmami. W tym celu obliczono test χ^2 oraz test post-hoc, który polegał na wielokrotnym testowaniu poszczególnych par zmiennych za pomocą testu χ^2 . Obliczono w ten sposób macierz p -wartości, która pozwalała na stwierdzenie czy różnice częstości wyboru danej odpowiedzi są istotne statystycznie. Także tutaj zastosowano poprawkę na wielokrotne testowanie metodą Simesa-Hochberga [18]. Wyniki uzyskane dla ogółu badanych przedstawiono w tabeli 6.1 oraz na rysunku 6.1.

Tabela 6.1 Macierz testu post-hoc bazującego na teście χ^2 uzyskana w analizie ogółu badanych. Pogrubioną czcionką oznaczono p -wartości mniejsze od poziomu istotności (0,05).

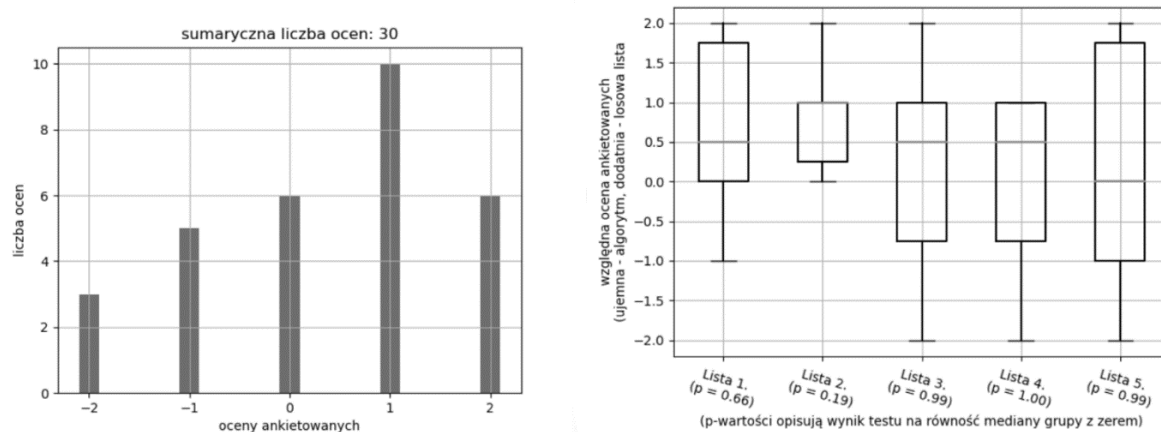
Ocena	-2	-1	0	1	2
-2	-	1	0.267	1	1
-1	1	-	0.352	1	1
0	0,267	0,352	-	0,011	0,005
1	1	1	0,011	-	1
2	1	1	0,005	1	-

Wykres pudełkowy z rysunku 6.1 potwierdza, że ogół ankietowanych nie miał preferencji względem tego, jaki algorytm układa utwory na odsłuchiwanym przez nich liście odtwarzania. Jednak histogram odpowiedzi oraz istotność niektórych częstości wybierania odpowiedzi pomiędzy zerem a pozostałymi ocenami sugeruje, że ankietowani mogli słyszeć różnicę pomiędzy algorytmami. Część ankietowanych w tym przypadku jednak preferowała losowy układ list i dlatego w ogólnym ujęciu żaden algorytm nie jest faworyzowany. Taka obserwacja spowodowała, że przeprowadzone zostały analizy we wszystkich możliwych podgrupach wyznaczonych za pomocą zebranych w ankiecie informacji o osobach ankietowanych.

Jedną z przeprowadzonych w ten sposób analiz wykonano dla osób, które nie posiadają żadnego doświadczenia zawodowego w branży audio. Jej efekty zwizualizowano na rys. 6.2. Test post-hoc bazujący na teście χ^2 wykazał, że żadna z różnic widocznych na rys. 6.2 nie jest istotna statystycznie. Stąd można wnioskować, że ankietowani zaznaczali wszystkie możliwe odpowiedzi z podobną częstością. Można założyć, że brak doświadczenia zawodowego, a tym samym wytrenowania słuchu i prawdopodobnie związanego z tym wypracowania wrażliwości na pewne aspekty związane z percepcją muzyki wiąże się z brakiem postrzegania zmian nastroju utworów, itp., stąd ankietowani nie preferowali żadnego z przedstawionych algorytmów.



Rysunek 6.1 Histogram uzyskany po analizie wyników uzyskanych dla ogółu badanych testem χ^2 oraz zobrazowanie odpowiedzi wykresem pudełkowym.

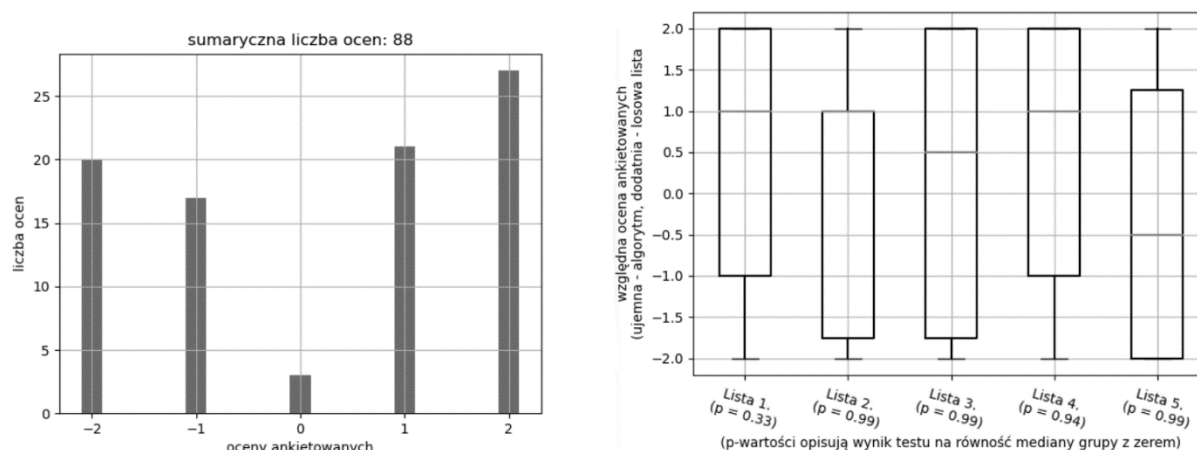


Rysunek 6.2 Histogram uzyskany po analizie wyników w grupie osób bez doświadczenia w branży audio testem χ^2 oraz zobrazowanie odpowiedzi wykresem pudełkowym.

Zbadano również preferencje badanych, którzy deklarowali przynajmniej podstawowe doświadczenie w branży audio. W tym wypadku również nie można było stwierdzić, który sposób układania list odtwarzania jest lepszy na podstawie mediany uzyskanych przez algorytm ocen. Jednakże możliwe było stwierdzenie, że badani słyszą różnice między prezentowanymi listami odtwarzania, ale ich preferencje są różne. Różnice w częstości zaznaczania odpowiedzi 0 względem dowolnej innej odpowiedzi są bowiem istotne statystycznie. Wyniki zobrazowano w tabeli 6.2 oraz na rysunku 6.3.

Tabela 6.2 Macierz testu post-hoc bazującego na teście χ^2 uzyskana w analizie badanych deklarujących przynajmniej podstawowe doświadczenie w branży audio. Pogrubioną czcionką oznaczono p-wartości mniejsze od poziomu istotności.

Ocena	-2	-1	0	1	2
-2	-	1	0,008	1	1
-1	1	-	0,031	1	1
0	0,008	0,0314	-	0,005	0,0003
1	1	1	0,005	-	1
2	1	1	0,0003	1	-



Rysunek 6.3 Histogram uzyskany po analizie wyników w grupie osób deklarujących doświadczenie w branży audio testem χ^2 oraz zobrazowanie odpowiedzi wykresem pudełkowym

6. WNIOSKI I PODSUMOWANIE

Zaprezentowany algorytm układa listy odtwarzania zgodnie z opisanymi wcześniej założeniami projektowymi. Uzyskane efekty zostały przetestowane za pomocą testów odsłuchowych. Przeprowadzone testy subiektywne i ich analiza statystyczna pozwoliła na obserwację, że badani nie mający doświadczenia w branży audio najprawdopodobniej nie słyszą różnic pomiędzy prezentowanymi im listami odtwarzania. Natomiast w grupie osób zainteresowanych technologią audio i pracujących zawodowo w branży audio udało się zaobserwować, że część osób preferuje sposób układania list odtwarzania realizowany przez zaprezentowany algorytm, a druga grupa preferowała listy ułożone poprzez zwykłe losowanie kolejności. Wniosek ten jest możliwy do wysunięcia po analizie częstości wyboru odpowiedzi 0, która wskazuje na brak słyszenia różnic pomiędzy prezentowanymi listami utworów, układanymi przez różne algorytmy. Praktycznym skutkiem tej obserwacji może być na przykład wniosek, że proponowany algorytm może być np. elementem algorytmu układania list w inteligentnym odtwarzaczu muzyki, ale jednocześnie zalecane jest, aby użytkownik takiego odtwarzacza miał kontrolę nad tym czy utwory są układane w kolejności losowej, czy zaproponowanej przez algorytm.

Interesującym kierunkiem badań w przyszłości jest sposób, w jaki badani słuchają muzyki. W tym kontekście, aby podjąć próbę zobiektywizowania tego typu testu odsłuchowego, należałoby wykorzystać utwory i gatunki muzyczne mniej znane, tzn. tylko takie powinny znaleźć się na prezentowanych listach. W dalszych eksperymentach można by zaproponować przygotowanie testu odsłuchowego, w którym zadaniem słuchaczy byłoby ułożenie utworów wg zadanego kryterium dopasowania początków i końców utworów i porównanie tych typowań z listą wygenerowaną przez opracowany algorytm.

BIBLIOGRAFIA

- [1] BERTIN-MAHIEUX, T.; ELLIS, D. P. W.; WHITMAN, B.; LAMERE, P.: The Million Song Dataset. In: *In Proceedings of the 12th International Society for Music Information Retrieval Conference* (2011).
- [2] BŁAŻEJCZYK, W.: Postawy kompozycji muzycznej dla wykonawców - materiały pomocnicze.
- [3] DARST, B.F.; MALECKI, K.C.; ENGELMAN, C.D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet* 19, 65 (2018). [HTTPS://DOI.ORG/10.1186/S12863-018-0633-8](https://doi.org/10.1186/s12863-018-0633-8)
- [4] DIEDERIK P. KINGMA and WELING M. (2019), "An Introduction to Variational Autoencoders", *Foundations and Trends® in Machine Learning*: Vol. 12: No. 4, pp 307-392. <http://dx.doi.org/10.1561/22000000056>
- [5] *DJ top tips*. URL <https://www.djtoptips.com/create-best-dj-set/>. – data dostępu 2020-08-01.

- [6] GÉRON, A.: *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*: O'Reilly Media, 2017.
- [7] GHAS, A.; LOGAN, J.; CHAMBERLIN, D.; SMITH, B. C.: Query by humming (1995), pp. 231–236.
- [8] GTZAN - baza danych. URL <http://marsyas.info/downloads/datasets.html>. - data dostępu 2020-05-15.
- [9] GÓRECKI, T.: Podstawy statystyki z przykładami w R, BTC, 2011.
- [10] HOFFMANN, P.; KOSTEK, B.: Koncepcja korekcji sygnału dźwiękowego z uwzględnieniem charakterystyk częstotliwościowych pomieszczenia oraz gatunku muzycznego. In: *Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej* (2016), Nr. 51, pp. 63–66.
- [11] ITU: BS.1284-1 General methods for the subjective assessment of sound quality. In: *Intern. Telecom. Union Radiocom. Sector* (2003), pp. 1–13.
- [12] JONES, E.; OLIPHANT, T.; PETERSON, P.: *SciPy: Open Source Scientific Tools for Python*. URL <http://www.scipy.org/>. data dostępu 2020-08-15.
- [13] KIM, A.; PARK, S.; PARK, J.; HA, JUNG-WOO; KWON, TAEGYUN; NAM, JUHAN: Automatic DJ Mix Generation Using Highlight Detection. In: 2017.
- [14] *Latin Music Database*. URL <https://sites.google.com/site/carlossillajr/resources/the-latin-music-database-lmd>. data dostępu 2020-05-09.
- [15] MPEG 7. URL <http://mpeg.chiariglione.org/standards/mpeg-7>. Data dostępu 2020-08-16.
- [16] PIETRUSIŃSKA, K. Opracowanie algorytmu do automatycznego układania list odtwarzania utworów muzycznych, Praca dyplomowa magisterska, Katedra Systemów Multimedialnych, Wydział ETI, Politechnika Gdańska (promotor: B. Kostek, konsultant: A. Kurowski) (2020).
- [17] PROKHOROV, V., SHAREGHI, E., LI, Y., PILEHVAR, M. T., COLLIER, N. (2019). On the importance of the Kullback-Leibler divergence term in variational autoencoders for text generation. arXiv preprint arXiv:1909.13668.
- [18] SIMES, R. J., An improved Bonferroni procedure for multiple tests of significance, *Biometrika*, Volume 73, Issue 3, December 1986, Pages 751–754, <https://doi.org/10.1093/biomet/73.3.751>.
- [19] SPORKA, A. J & VALTA, J (2017) Design and implementation of a non-linear symphonic soundtrack of a video game, *New REVIEW of Hypermedia and Multimedia*, 23:4, 229-246, DOI: 10.1080/13614568.2017.1416682.
- [20] *Spotify*. URL <https://www.spotify.com/pl/>. data dostępu 2020-07-18.
- [21] TZANETAKIS, GEORGE; COOK, PERRY: Musical genre classification of audio signals. In: *IEEE Transactions on Speech and Audio Processing* vol. 10 (2002), Nr. 5, pp. 293–302.
- [22] VAN DEN OORD, AARON; DIELEMAN, SANDER; SCHRAUWEN, BENJAMIN: Deep content-based music recommendation. In: BURGESS, C. J. C.; BOTTOU, L.; WELLING, M.; GHAHRAMANI, Z.; WEINBERGER, K. Q. (eds.): *Advances in Neural Information Processing Systems 26*: Curran Associates, Inc., 2013, pp. 2643–2651.

AUTOMATIC CREATION OF PLAYLISTS OF MUSIC TRACKS

The aim of the work is to design an algorithm to automatically arrange the order of musical pieces and thus create one long mix consisting of them. The algorithm selects songs on the basis of the similarity of the ending and opening fragments of the songs. This similarity is calculated using the Euclidean distance between the parameter vectors computed by the autoencoder and on the basis of data clustering. This way of arranging the songs ensures the continuity of the list. The quality of the results is verified employing listening tests by comparing automatically generated lists with those arranged randomly.

INFLUENCE OF MUSIC EDUCATION ON THE GROUPING OF THE ABA-ABA SEQUENCE SOUNDS IN THE GALLOP RHYTHM

Adam ROSIŃSKI¹

¹University of Warmia and Mazury in Olsztyn, Faculty of Arts, Szrajbera 11, 10-007 Olsztyn
adam.rosinski@uwm.edu.pl

This article presents the influence of music education on the grouping of sounds into perception streams with the use of the sounds ABA-ABA, which make up the pattern of the “gallop rhythm.” The experiment carried out in both groups confirms that musicians utilize specific skills for an auditory assessment of the sound, which allows them to perceive the “gallop rhythm,” whereas people of no music education cannot follow the “gallop rhythm,” which is present in the samples of the ABA-ABA pattern, in their perception field.

Psychoacoustic analyses are mostly based on the physical assessment of the organ of hearing, ignoring the issue of music education, which may have a significant impact on the interpretation of the results in the group of people mentioned above. It is of particular importance for the performed experiment to note if the respondents have music education, as such information may bring us closer to determining why people with the same physiological hearing condition differ in the way of processing sounds into perception streams.

1. JOINING SOUNDS IN TONAL STRUCTURES

Grouping is a process with the main concept involving seeking ranges of auditory scenes, joined with respect to a similar type of integrity (e.g. similarity of colour or pitch). Sounds can be grouped in either of two ways: in sequential grouping, which concerns the melody of a music piece, or in simultaneous grouping, which is responsible for the perception of harmony, if similarities or regularities that favour sound integration are found in both cases [1].

When hearing a piece of music, a listener creates an image in their memory, which corresponds to sounds, regardless of whether the piece was a coherent entity or comprised of multiple different parts. Creating a projection based on the music foundation in one’s mind indicates that both the perceptive image of the whole and of individual parts corresponds to a given piece of music. A piece with a hierarchical structure is created when a sound event consisting of parts is perceived as a whole. A descriptive-formative system must be referred to [2] as an aid in confirming that a (melodic or harmonic) sequence of sounds was performed as a specific perceptive object, i.e. it achieved the mental importance.

2. THE ASSUMPTIONS AND OBJECTIVE OF THE EXPERIMENT

The objective of the experiment was to identify the relationship between grouping in perceptive streams and the pitch scale within which the sounds played in the ABA-ABA sequence were distributed. The author’s main assumption in the study was to use sound sequences described in the literature and to compare the results obtained in the two groups under study – musicians and non-musicians. The principal intention of the analysis was to enable referring the experiment results directly to literature data. The study participants were presented with sequences of sound samples of the ABA-ABA pattern, which was usually used in the published studies [3]. Variables included the tempo increasing linearly as a function of the duration of playing the sequence and differences in the pitch scale in which the samples

were played.

3. THE SOUND MATERIAL

Repeated sequences consisting of three sounds at the perfect fifth interval. This interval was played in the great, one-line and three-line octaves. The sound frequencies were established in evenly tempered tune.

- in the great octave: (Tone: A) D = 73.416 Hz, (Tone: B) A = 110.000 Hz, (Tone: A) D = 73.416 Hz,
- in the one-line octave: (Tone: A) d1 = 293.666 Hz, (Tone: B) a1 = 440.000 Hz, (Tone: A) d1 = 293.666 Hz,
- in the three-line octave: (Tone: A) d3 = 1174.700 Hz, (Tone: B) a3 = 1760.000 Hz, (Tone: A) d3 = 1174.700 Hz.

The sounds were played in the following sequence: low pitched sound, high pitched sound, low pitched sound in each series. The time between the beginning of a sound and the beginning of the neighbouring one was taken as the basic time unit in measurements. One sound series in this experiment consisted of three sounds repeated alternately. The sound samples were played in the ABA-ABA pattern, where the dash denotes silence whose duration is equal to the sound duration, as in the experiment conducted by Leo van Noorden [3].

The duration of each sound was computer-controlled and ranged from 400 to 40 ms [4]. In order to avoid crackles when listening to the material, 10 ms of linear sound loudness increase and decrease was added in the sound envelope. The temporal sequence of the sounds was regulated and analysed by the computer program on the WPF .NET platform. The sound series was played 187 times at the most, which gives 561 sounds played in one presentation. The test was repeated three times with each participant in order to eliminate compression artefacts, which could appear in the respondents' responses. The agreed loudness level was approx. 70 phons. The respondents participating in the pilot study did not take part in the test proper [5]. All of the sound samples were generated in the mono system at the sampling frequency of 44.1 kHz, in 16-bit resolution.

4. LISTENERS AND LISTENING SESSIONS

Forty-eight people aged 21 to 27 years took part in the study, including 24 students of or graduates from the Academy of Music in Gdańsk and 24 students or graduates from the Academy of Fine Arts in Gdańsk. The group comprised 40 females and 8 males. The participants in both the groups were chosen at random and they had never taken part in such experiments before.

- Details of the musicians group: the participants had begun their musical training between the age of 5 and 10 years and they had completed the 1st and 2nd degree music school; they had practiced playing a musical instrument without a break for the past 13-22 years. The participants: did not have perfect pitch, played classical music, were all right-handed.
- Details of visual artists group (I): the participants did not learn to play any instrument or sing, nor did they play any music in an amateur environment, they usually listened to popular music in the broad sense of the term, they were all right-handed.
- More details of visual artists group (II): five participants played popular music; the participants had learned to play keyboard, piano, transverse flute and electric guitar at music centres for a period of 4 to 7 years; some of them also learned to sing at culture centres; the participants referred to above were currently playing popular music in amateur music bands affiliated with different culture institutions, with similarity of musical training, e.g. at culture centres, youth centres and music centres; they were all right-handed.

A demonstration series of sounds described in the study by Bregman [6] was played at the beginning of the session in order to familiarise the participants with the phenomenon of joining sounds into one perceptive stream and to explain the task in the experiment.

5. RESULTS

The test revealed that participants with no musical training did not perceive the galloping rhythm formed as a result of internal grouping processes when samples in the ABA-ABA pattern were played. Some non-musicians reported the temporary appearance of a galloping rhythm, but the impression did not persist at a growing sequence playing rate. The published findings [9] show that the galloping rhythm is usually perceived as increasingly fast with an increasing playing rate, whereas the non-musician participants with no musical experience reported that the rhythm became “blurred” and that they did not hear the galloping sequence but rather separate low-pitched and high-pitched sounds. Introducing to the experiment an extra factor (a super-imposed internal figure), which was formed as a result of interpretation by the auditory system of the galloping rhythm, revealed during the listening of the ABA-ABA pattern, poses a great problem to participants with no musical training, regardless of the pitch scale in which the sound samples were played. An analysis of the initial test results took into account the fact that sound diminution is accompanied by spectrum blurring, which is defined as a spectrum of a short tone or a multi-tone acquiring the features of a continuous spectrum, typical of noise. Identification of high-pitched tones becomes much less precise with a blurred spectrum than with longer sounds [7].

Interesting results were obtained in the group of 5 non-musicians, who had learned to play musical instruments as amateurs and performed in amateur popular music bands. The responses from this group indicated that the galloping rhythm was recognised and kept within the perception field at quick rates and at different pitch scales. This information can be regarded only as auxiliary because the group was too small to formulate reliable conclusions concerning generation of perceptive streams which are responsible for formation of the galloping rhythm.

These experiments confirmed that extended musical practice improves the listeners’ competence, and the effects of such training can be confirmed with psychoacoustic tests, provided the training lasted at least 4 years [8].

People with musical training recognised the galloping rhythm in the ABA-ABA sequences at different rates and in different pitch scales, which was consistent with the study assumptions. There was no spectrum blur in this case [7], because long-term, consistent musical training allowed for much more accurate perceptive analysis of the tones played in the test. A comparison of predispositions in participants with and without musical training shows that the former notice many more details even when the differences between the sounds are very small.

6. PROSPECTS FOR FURTHER DEVELOPMENT – CONCLUSIONS

Lack of information on whether the participants of the experiments described in the literature had any musical training is an important factor which makes comparison of the results difficult. Shortage of data probably resulted from the fact that the participants’ musical training was not taken into account and most of them probably were not musicians. However, it is noteworthy that the majority of these analyses were performed in North America and in The Netherlands. Although they may not have any specialist musical training, the American, Canadian and Dutch listeners probably had more musical competence and a better sense of rhythm than Polish non-musicians. The majority of young people in Poland cannot read notes or play an instrument as amateurs. The musical training syllabus in general education schools is often much better developed than in Poland. Having completed the equivalent of a

Polish secondary school of general education, people with no musical training can often play a selected instrument as amateurs, read notes and know the basics of solfege. The term “person with no musical training” acquires a completely new meaning depending on the syllabus in different countries and it means that the level of musical competences varies.

An analysis of the specialist literature cited above shows that routine audiometric tests (checking whether the respondents’ hearing is within the otological norm) are insufficient, because they should be expanded to include qualification criteria, i.e. musical training should be made an ordering factor. The fact that musical training and practice should affect cognitive processes in sound perception should be taken into account in recruiting participants for psychoacoustic studies because identical experiments performed in different groups produce completely different responses [9]. Musical training often has a significant impact on the experiment results and note should be taken of this information as it was not taken into account in planning psychoacoustic experiments.

The results of these analyses show clearly that long-term musical training brings tangible benefits, which means that people with musical training can draw on their knowledge even unconsciously and in tasks unrelated directly to music. Sound interpretation and processing in one’s mind – when sounds come from the external environment – shows that sound receptions by musicians has different characteristics. Grouping and segregation of sounds in the minds of people with musical training differs, as they pay attention to their diverse features, which is not the case in people with no musical training. This study shows that common musical training can bring interesting results, which can be helpful in areas unrelated directly to music, such as learning foreign languages as mentioned above.

REFERENCES

- [1] HUMIEŃKA-JAKUBOWSKA J., *Scena słuchowa muzyki dwudziestowiecznej*, Rhythmos, Poznań 2006, 85.
- [2] BREGMAN A. S., *Auditory scene analysis: the perceptual organization of sound*, The MIT Press, Cambridge, Massachusetts 1990, 165.
- [3] BREGMAN A. S., WOSZCZYK W., *Controlling the perceptual organization of sound: guidelines derived from principles of auditory scene analysis*, W: Audio anecdotes: tools, tips and techniques for digital audio, Vol. 1, GREENEBAUM K., BARZEL R. (red.), A. K. Peters, Natick, 2004, 39-41; BREGMAN A. S., AHAD P. A., CRUM P. A. C., O’REILLY J., Effect of time intervals and tone durations on auditory stream segregation, *Perception & Psychophysics*, No. 62, Vol. 3 (2000), 626-636; SINGH P. G., BREGMAN A. S., Effect of different timbre attributes on the perceptual segregation of complex-tone sequences, *Journal of the Acoustical Society of America*, No. 102, Vol. 4, (1997), 1943-1952; van NOORDEN L. P. A. S., *Temporal coherence in the perception of tones sequences*, Technical University Eindhoven, 1975 (unpublished doctoral dissertation).
- [4] BREGMAN A. S., *Auditory scene analysis: the perceptual organization of sound*, The MIT Press, Cambridge, Massachusetts 1990, 50, 157-158; van NOORDEN L. P. A. S., *Minimum differences of level and frequency for perceptual fission of tone sequences ABAB*, *Journal of Acoustical Society of America*, Vol. 61, No. 4, (1977), 1041; DANNENBRING G. L., BREGMAN A. S., *Effect of Silence between tones on auditory stream segregation*, *Journal of Acoustical Society of America*, Vol. 59, No. 4, (1976), 987; DANNENBRING G. L., BREGMAN A. S., *Stream segregation and the illusion of overlap*, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 2, No. 4, (1976), 546, 568.
- [5] GRUSZCZYŃSKI L. A., *Kwestionariusze w socjologii: budowa narzędzi do badań surveyowych*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 1999, s. 121.
- [6] <http://webpages.mcgill.ca/staff/Group2/abregm1/web/downloadstoc.htm>, [02.05.2020].
- [7] ROGALA T., *Identyfikacja melodii jako metoda badania siły wysokości dźwięku*, *Muzyka*, Vol. 55, 4 (219), Warszawa 2010, s. 65-76.
- [8] BOGUSZ-WITCZAK E., SKRODZKA E., TURKOWSKA H., *Influence of Musical Experience of Blind and Visually Impaired Young Persons on Performance in Selected Auditory Tasks*, *Archives of Acoustics*, Vol. 40, No. 3. (2015), 347.
- [9] SHONLE J. I., HORAN K. E., *Trill threshold revisited*, *Journal of Acoustical Society of America*, Vol. 59, Issue 2, (1976), 469-471; MILLER G. A., HEISE G. A., *The thrill threshold*, *Journal of Acoustical Society of America*, No. 22, (1950), 637-638.

WPLYW PARAMETRÓW KOMPRESJI DYNAMICZNEJ NA SUBIEKTYWĄ GŁOŚNOŚĆ NAGRANIA INSTRUMENTU MUZYCZNEGO

Karol CZESAK¹

Piotr KLECZKOWSKI²

¹Akademia Górniczo Hutnicza im. Stanisława Staszica w Krakowie, 30-059 Kraków,
al. Mickiewicza 30, czesak@agh.edu.pl

²Akademia Górniczo Hutnicza im. Stanisława Staszica w Krakowie, 30-059 Kraków,
al. Mickiewicza 30, kleczkow@agh.edu.pl

W produkcji muzycznej zauważamy zwiększającą się tendencję do zawężania zakresu dynamicznego nagrań muzycznych. Głównym założeniem jest tu zwiększenie całościowej głośności nagrania. W zdecydowanej większości przypadków, pojęcie głośności odnosi się do wartości skutecznej sygnału w kanale transmisji. W rezultacie, materiał często jest obrabiany przy użyciu agresywnej kompresji dynamicznej. Jak pokazuje doświadczenie inżynierów dźwięku – subiektywne odczucie głośności nie jest warunkowane wyłącznie przez stopień kompresji, ale też przez inne jej parametry. Celem tej pracy jest zbadania, które parametry kompresji dynamicznej wywierają istotny wpływ na subiektywnie postrzeganą głośność zróżnicowanego materiału muzycznego. Stwierdzono, że jest możliwe takie dobranie parametrów kompresji dynamicznej, by nagranie sprawiało wrażenie głośniego, bez stosowania nadmiernego stopnia kompresji.

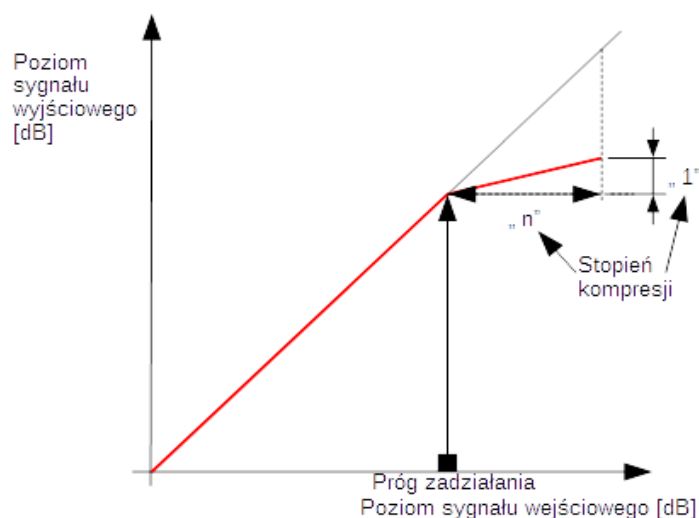
1. WSTĘP

Kompresja dynamiki jest procesem mającym na celu zredukowanie zakresu dynamiki materiału dźwiękowego, poprzez redukcję wzmocnienia sygnału, po przekroczeniu przezeń założonego progu. Pozwala ona zabezpieczyć kanał transmisji przed przesterowaniem, a także zwiększyć komfort odsłuchu materiału. Co więcej – już w latach 50tych, na podstawie obserwacji użytkowników szaf grających, stwierdzono, że materiał o większym poziomie skutecznym jest postrzegany przez odbiorcę jako atrakcyjniejszy. Prowadziło to do sukcesywnego zwiększania poziomu skutecznego nagrań na przestrzeni kolejnych dziesięcioleci. W latach 90tych XX wieku i u progu XXI stulecia, możemy mówić wręcz o „wojnie głośności” [1]. Na ową „Wojnę głośności” postanowiły odpowiedzieć Międzynarodowa Organizacja Telekomunikacyjna (ITU), oraz Europejska Unia Nadawców (EBU), publikując liczne dokumenty, mające na celu uzgodnienie poziomu głośności sygnału w mediach elektronicznych, w rozumieniu jej zakresu dynamicznego oraz wartości skutecznej. W związku z powyższym, zdajemy sobie pytanie: „Czy jest możliwe wyprodukowanie materiału dźwiękowego, spełniającego przyjęte zalecenia techniczne dotyczące głośności, jednak percepcyjnie głośniejszego od innych, sobie podobnych?”. Próba znalezienia odpowiedzi na tak postawione pytanie, okazała się być motywacją do znalezienia zależności między parametrami kompresji dynamicznej, a subiektywnym odczuciem głośności i atrakcyjności materiału, bazującym na podstawie komfortu i poczucia estetyki słuchacza.

2. ROZPATRYWANE PARAMETRY KOMPRESJI DYNAMICZNEJ

Kompresor dynamiki jest narzędziem przekształcającym sygnał w sposób nieliniowy, działającym w dziedzinie czasu. Wspomniane nieliniowości zachodzą jednak w stanach przejściowych, co pozwala nam zaliczyć kompresor do narzędzi liniowych, zmiennych w czasie [2]. Kompresor dynamiczny jest procesorem typu adaptatywnego – wprowadzane przezeń zmiany są zależne od sygnału podawanego na jego wejście, za którym to sygnałem procesor ten niejako „podąża”, w wyniku czego, jego stosowanie pozostaje nie bez wpływu na barwę dźwięku, a niekiedy wprowadza też – mniej lub bardziej słyszalne artefakty. Dzięki tej właściwości – kompresor może być stosowany nie tylko jako zabezpieczenie materiału przed wystąpieniem zniekształceń, ale też jako narzędzie kreacji brzmienia.

Charakterystykę statyczną kompresora definiują parametry: próg zadziałania i stopień kompresji. Próg zadziałania wyrażany jest w skali decybelowej, gdzie poziomem odniesienia jest nominalny poziom sygnału w kanale transmisji. Próg zadziałania parametrem określającym poziom, po przekroczeniu którego przez sygnał podany na wejście kompresora, zaczyna zachodzić tłumienie tegoż sygnału. Stopień kompresji wyrażany jest poprzez iloraz „n:1”, definiujący stosunek „nadwyżki” sygnału wejściowego, przekraczającego wartość progową, do „naddatku” sygnału wyjściowego pozostającego powyżej tejże wartości. Zmienna „n” zawiera się tutaj w przedziale (1: +∞). Charakterystyka statyczna teoretycznego modelu kompresora z zaznaczonymi opisanymi powyżej parametrami przedstawiona została na rysunku 1.



Rys. 1. Charakterystyka statyczna kompresora

Poziom, o jaki sygnał na wyjściu ulega redukcji w stosunku do sygnału na wejściu, określony jest wzorem:

$$S = (X - P) * \left(1 - \frac{1}{n}\right) \quad (1)$$

gdzie:

S – założony poziom redukcji wzmacnienia [dB]

X – poziom sygnału wejściowego [dB]

P – próg zadziałania kompresora [dB]

n – zmienna determinująca stopień kompresji postaci: n:1

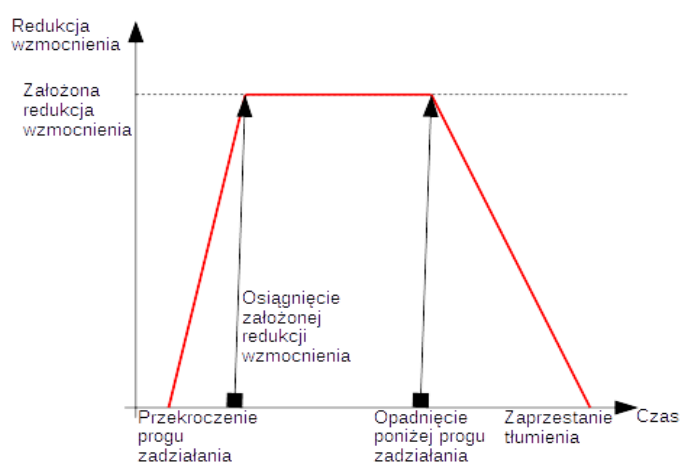
Co pozwala nam określić sygnał wyjściowy Y jako:

$$Y = X - S \quad (2)$$

W związku z powyższym, stopień kompresji przybiera postać:

$$R = \frac{X-P}{Y-P} \quad (3)$$

W celu uniknięcia artefaktów, związanych z rkompresją dynamiki, przebieg redukcji wzmacnienia charakteryzuje się pewną obwiednią czasową. Historycznie, wynikała ona ze swoistej bezwładności układów analogowych implementujących proces kompresji dynamiki. Obecnie jednak, obwiednia ta opisana jest czasem ataku oraz czasem opadania. Obwiednia ta – dla skali decybelowej, przedstawiona została na rysunku 2.



Rys. 2. Obwiednia czasowa tłumienia sygnału

Pamiętajmy jednak, iż kompresor w rzeczywistości pracuje na sygnale napięciowym, mającym charakter liniowy, toteż w rzeczywistości czas ataku przyjmuje postać:

$$t_{at} = 1 - e^{-2,2 \cdot T_s / T_{at}} \quad (4)$$

gdzie:

t_{at} – rzeczywisty czas ataku kompresora [s]

T_s – okres próbkowania [s]

T_{at} – czas ataku w skali decybelowej [s]

Analogiczną postać przyjmuje przeliczenie czasu zwolnienia:

$$t_{op} = 1 - e^{-2,2 \cdot T_s / T_{op}} \quad (5)$$

gdzie:

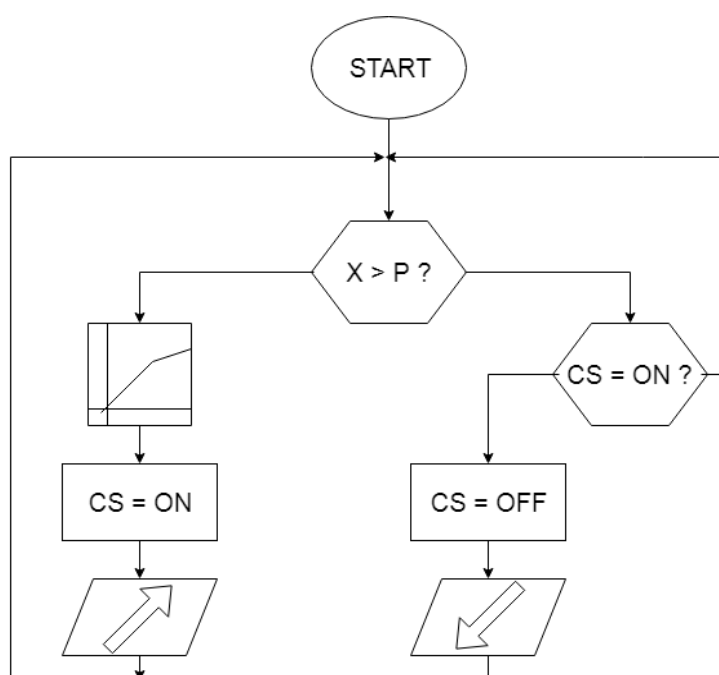
t_{op} – rzeczywisty czas opadania kompresora [s]

T_s – okres próbkowania [s]

T_{op} – czas opadania w skali decybelowej [s]

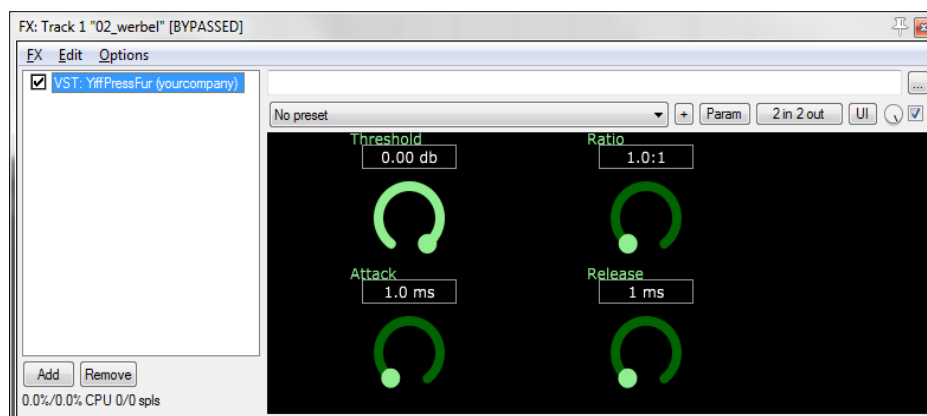
Wprowadzenie regulacji czasu ataku i zwolnienia kompresora nie tylko umożliwia ustrzeżenie się przed potencjalnymi artefaktami i pozwala lepiej dostosować własności kompresji do rodzaju obrabianego materiału, ale też stwarza możliwość zastosowania kompresora w charakterze narzędzia do kształtowania obwiedni amplitudowej sygnału [3]. Przykładem takiego zastosowania może być kontrola transjentów;

Zgodnie z powyższymi rozważaniami, oznaczając stan pracy kompresora zmienną CS, przyjmującą wartości „prawda” lub „fałsz”, działanie kompresora, za pomocą schematu blokowego, przedstawione zostało na rysunku 3.



Rys. 3. Schemat blokowy kompresora dynamiki

W celu uzyskania kompresora, dokładnie spełniającego powyższe założenia, konieczna była jego samodzielna implementacja programowa przez autora. Dla łatwości użytkowania, został on zaimplementowany w postaci wtyczki VST, przy użyciu środowiska JUCE. Widok graficznego Interfejsu użytkownika przedstawia rysunek 4.



Rys. 4. Graficzny interfejs użytkownika kompresora, uruchomionego wewnątrz oprogramowania Cockos Reaper

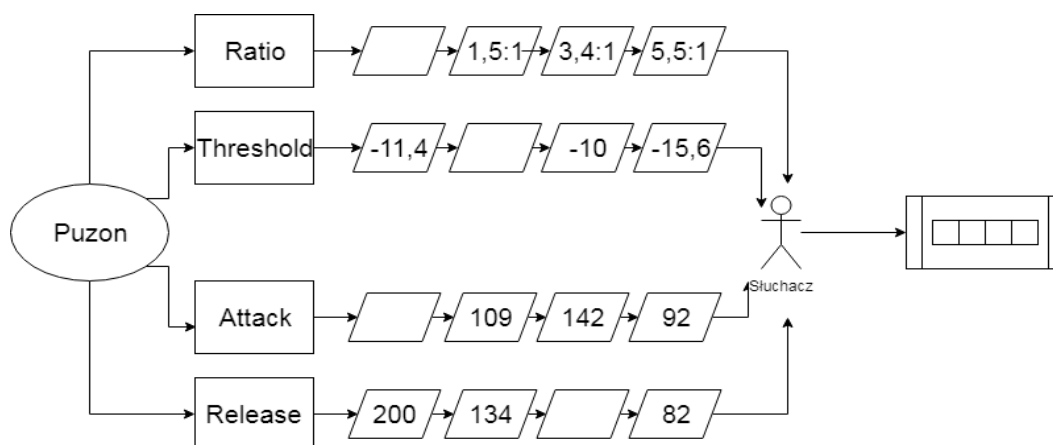
3. DOBÓR PRÓBEK I TESTY ODSŁUCHOWE

Analizę parametrów kompresji na subiektywnie odbieraną głośność zdecydowano się przeprowadzić dla sześciu instrumentów, chętnie poddawanych przez realizatorów dźwięku kompresji dynamicznej. Były to: Gitara basowa, puzon, saksofon, skrzypce, werbel i wokół. Próbkę te pochodziły z wielośladowego nagrania utworu kompozycji Andrzeja Zielińskiego do słów Agnieszki Osieckiej „Nie całuj mnie pierwsza”, w bigbandowej aranżacji Wojciecha Zwierniaka.

Dla każdego z parametrów, dana próbka prezentowana była w testach odsłuchowych czterokrotnie: z trzema różnymi wartościami analizowanego parametru oraz w postaci nieskompresowanej. Wartości parametrów zostały dobrane doświadczalnie. Następnie – wartość skuteczna wszystkich próbek została znormalizowana do poziomu -30LKFS. Każda z próbek trwała od 12 do 15 sekund.

Słuchacz dostawał do odtworzenia zestaw plików, z których każdy zawierał cztery próbki brzmienia instrumentu. Gdy emitowane były próbki ze zmieniającym się konkretnym parametrem kompresji, wszystkie pozostałe parametry pozostawały niezmiennie. Kolejność prezentowania próbek w obrębie danej grupy była określona w sposób pseudolosowy. Zadaniem słuchacza było uszeregowanie – dla każdego z plików – próbek od uważanych przezeń za najcichsze, do uważanych przezeń za najgłośniejsze. Łącznie – słuchacz otrzymał do oceny 24 zestawy próbek. Przebieg testu obrazuje rysunek 5.

Analizie statystycznej zostały poddane wyniki uzyskane z przebadania 15 słuchaczy. Głównie były to osoby młode - Studenci kierunku Inżynieria Akustyczna na Krakowskiej AGH oraz młodzież korzystająca z oferty kulturalnej Świetlicy Wiejskiej w Gierczycach. Wszyscy respondenci pytani byli o wiek oraz doświadczenie w pracy z dźwiękiem lub grze na instrumencie. Dało się zauważyć, że osoby „grające” zgłaszają inne uwagi i spostrzeżenia dotyczące prowadzonego badania, co pozwala wnioskować, że funkcjonują u nich inne mechanizmy percepcji głośności dźwięku. Statystyczna istotność uzyskiwanych wyników oceniana była za pomocą testów Kruskala – Wallisa i wielokrotnego porównania średnich.



Rys. 5. Schemat przebiegu testów odsłuchowych

4. PREZENTACJA UZYSKANYCH WYNIKÓW

W przypadku każdego z badanych parametrów kompresji dynamicznej uzyskano statystycznie istotne wyniki, dla co najmniej dwu instrumentów. Najwięcej – bo aż dla pięciu instrumentów – statystycznie istotnych różnic w percypowanej głośności uzyskano dla progu zadziałania. W przypadku stopnia kompresji -statystycznie istotne różnice pojawiły się dla czterech instrumentów. Dla parametrów określających obwiednię czasową tłumienia sygnału znaleziono po dwa instrumenty ze statystycznie

istotnie różniącymi się postrzeganą głośnością próbkami.

I tak – w przypadku prógu zadziałania, gitara basowa skompresowana z progiem -13,1dBFS okazała się być wyraźnie głośniejsza od skompresowanej z progiem -9,1dBFS., puozn skompresowany z progiem -15,6dBFS okazał się istotnie głośniejszy od puzonu nieskompresowanego. Saksofony zaś – skompresowany z progiem zadziałania -10,8dBFS oraz nieskompresowany okazały się być wyraźnie głośniejsze od saksofonu skompresowanego z progiem zadziałania -12,7dBFS. Wokal, skompresowany z progiem zadziałania -20,8dBFS okazał się być wyraźnie głośniejszy od skompresowanego z progiem zadziałania -16,9dBFS, zaś skrzypce – skompresowane z progiem zadziałania -18dBFS okazały się być głośniejsze od skompresowanych z progiem zadziałania -20,5dBFS. Jak łatwo można zauważyć – w przypadku wokalu i skrzypiec wystąpiło zjawisko przekompresowania.

W przypadku stopnia kompresji, gitara basowa skompresowana ze stopniem 4,2:1 okazała się być istotnie głośniejsza od skompresowanej ze stopniem 2,8:1. Saksofon zaś – skompresowany ze stopniem 3,0:1 okazał się być wyraźnie głośniejszy od saksofonu skompresowanego ze stopniem 1,5:1. Skrzypce – nieskompresowane i skompresowane ze stopniem 3,2:1 okazały się być wyraźnie głośniejsze od skrzypiec skompresowanych ze stopniem 2,1:1 i 1,4:1, zaś wokal, skompresowany ze stopniem 1,9:1 okazał się być głośniejszy od wokalu skompresowanego ze stopniem 1,6:1. Tutaj potwierdziła się sformułowana w założeniach referatu prawidłowość, jednak ciekawe wyniki uzyskane zostały dla skrzypiec – próbka nieskompresowana, jak również najmocniej skompresowna okazały się być głośniejsze od pozostałych.

W przypadku czasu ataku, puzon skompresowany z czzasem ataku 142ms okazał się być głośniejszy od puzonu nieskompresowanego, zaś werbel, skompresowany z czasem ataku 84ms okazał się być wyraźnie głośniejszy od skompresowanego z czasem ataku 32ms. Te - z pozoru absurdalne - wyniki znajdują uzasadnienie w charakterze obwiedni czasowej brzmienia tychże instrumentów w „czystej” postaci.

Dla czasu zwolnienia, puzon nieskompresowany i skompresowany z czasem zwolnienia 82ms, okazał się być wyraźnie głośniejszy od puzonu skompresowanego z czasem zwolnienia 200ms, zaś skrzypce nieskompresowane i skompresowane z czasem zwolnienia 32ms okazały się być istotnie głośniejsze od skrzypiec skompresowanych z czasem zwolnienia 21ms, co by wynikało z pewnego przywyczajenia ucha ludzkiego do nieznieskształconego brzmienia skrzypiec.

4. WNIOSKI I KIERUNKI DALSZYCH DZIAŁAŃ

W ujęciu ogólnym, po przeprowadzonych badaniach możemy sformułować następujące wnioski:

- Z obniżaniem progu zadziałania kompresora, rośnie postrzegana głośność
- Zwiększenie stopnia kompresji powoduje wzrost odbieranej głośności
- Przy instrumentach, o transjentowym charakterze brzmienia, kompresor może stanowić narzędzie kształtowania obwiedni – dłuższy czas ataku powoduje wrażenie głośniejszego nagrania.
- Dobranie zbyt długich czasów zwolnienia powoduje wieksza bezwładność układu i ujemnie wpływa na odbieraną głośność.
- Wartości parametrów kompresji należy dobrać do charakteru obrabianego materiału!

Nie ulega wątpliwości, że badanie miało charakter zgrubny. By dokładnie wyznaczyć wpływ parametrów kompresji dynamicznej, na percepowaną głośność nagrania instrumentu muzycznego, należałoby przeprowadzić badania dla większej ilości wartości pojedynczego parametru, jak również dla znacznie poszerzonego instrumentarium. Dla uzyskania naprawdę rzetelnych wyników, takie posunięcie jest konieczne. To z kolei wymagałoby sporządzenia o wiele większej ilości przykładów oraz podzielenia badań odsłuchowych, zakrojonych na iększą skalę, na części i sekcje. Niemniej – w dalszej perspektywie – jest to jednak możliwe.

Warto rozważyć też analizę wpływu innych, niż parametry kompresji dynamicznej czynników na percepcyjną głośność nagrania instrumentu muzycznego, jak zawartość zniekształceń harmoniczných w sygnale oraz składowych nieharmoniczných, czy szumu.

LITERATURA

- [1] E. DERUTY - „Dynamic Range' & The Loudness War”, <https://www.soundonsound.com/sound-advice/dynamic-range-loudness-war> (dostęp 12.04.2019)
- [2] T. P. ZIELIŃSKI, P. KOROHODA, R. RUMIAN - „Cyfrowe przetwarzanie sygnałów w telekomunikacji”, Wydawnictwo Naukowe PWN SA, Warszawa 2014
- [3] <http://www.studiomastering.net/mastering05.html> (dostęp 28.03.2019 19:32)

EFFECTS OF DYNAMIC COMPRESSION PARAMETERS ON SUBJECTIVE LOUDNESS OF MUSICAL INSTRUMENT RECORDING

There is an increasing trend in music production of decreasing the dynamic range of music recordings. The main purpose is the overall increase of loudness. In most cases, the idea of loudness was referring to the effective value of signal level in a transmission channel. As a result, the material was often treated with aggressive dynamic range compression. The practical experience of audio engineers is, that the subjective impression of loudness is affected not only by the dynamic compression ratio, but by other parameters of compression too. The goal of this work was to investigate, which of the parameters significantly affect the subjective perception of loudness of various music material. It was found that it was possible to set dynamic range compression parameters so that a recording would produce an impression of loudness without excessive dynamic range compression ratio.

POMIAR GŁOŚNOŚCI NAGRAŃ DŹWIĘKOWYCH PRZECHOWYWANYCH NA RÓŻNYCH NOŚNIKACH

Przemysław PLASKOTA¹

Małgorzata GAWLIŃSKA¹

¹Politechnika Wrocławska, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
przemyslaw.plaskota@pwr.wroc.pl

W pracy przedstawiono porównanie wartości parametrów akustycznych związanych z głośnością nagrań dźwiękowych zapisanych na różnych nośnikach i sprawdzono, czy powszechnie panująca opinia odnośnie zwiększającej się głośności utworów jest prawdziwa. Celem pracy jest pomiar parametrów akustycznych nagrań dźwiękowych w zależności od rodzaju nośnika oraz analiza otrzymanych wyników. Zmierzono wartości następujących parametrów: rzeczywistego poziomu szczytowego TPL (True Peak Level), głośności wyrażonej w jednostce LUFS, wartości skutecznej oraz zakresu dynamiki DR (Dynamic Range). Do analizy wykorzystano utwory muzyczne zapisane na płycie analogowej, płycie CD oraz pochodzące z trzech różnych serwisów internetowych (streaming).

WPROWADZENIE

Dokumenty ITU-R [1-4] oraz EBU [5-8] opisują metody pomiaru głośności w mediach takich jak radio, telewizja, czy film oraz określają konkretne wartości, do których należy dążyć. Wytyczne zamieszczone w nich mają pomóc profesjonalistom w tworzeniu produkcji, podczas odtwarzania których nie będzie konieczna regulacja poziomu głośności zarówno w trakcie transmisji jak i przełączania się pomiędzy kanałami. Poziom głośności może nadal różnić się w ramach potrzeby artystycznej lub technicznej programu. W celu uwzględnienia różnorodności głośności konkretnego programu korzysta się z metody normalizacji głośności przy wykorzystaniu średniej wartości głośności. Wyrównywanie głośności dotyczy wszystkich etapów transmisji audio od pozyskania do dystrybucji i przesyłania sygnału.

Jeszcze kilka lat temu do określania poziomu głośności programów korzystano z mierników, które mierzą wartości szczytowe sygnałów PPM (Peak Programme Meter). Z powodu iż mierniki te nie uwzględniają różnic związanych z próbkowaniem sygnału, następuje rozbieżność pomiędzy wartością szczytową PEAK, a rzeczywistą wartością szczytową TPL (True Peak Level), występującą w nadawanym programie. Aby uniknąć wartości, które zostały pominięte podczas pomiaru, a wynikały z niedokładności miernika, został wprowadzony maksymalny poziom PML (Permitted Maximum Level). Pomimo określonych wytycznych, nacisk na głośność programów w komercyjnych stacjach spowodował, że przy wykorzystaniu nowoczesnych mierników i po odpowiednich zabiegach uzyskanie dopuszczalnej wartości stało się możliwe, pomimo wyższej rzeczywistej wartości PEAK programu.

Problem ten rozwiązała Międzynarodowa Unia Telekomunikacyjna (ITU), która stworzyła dokument ITU-R BS. 1770 opisujący algorytm pomiaru głośności i rzeczywistych poziomów szczytowych programów. Dokument ITU-R BS. 1770 opisuje metody pomiaru, a dokument EBU R128 definiuje konkretne wartości normalizacji głośności oraz sposoby bramkowania sygnału, by lepiej dostosować głośność programów zawierających dłuższy okres ciszy lub odizolowaną mowę i nie zaniżać wartości poziomu głośności. EBU w celu ułatwienia twórcom programów określenia głośności wprowadziło trzy parametry: głośność programu, zakres głośności, rzeczywisty poziom szczytowy.

Głośność programu to długoczasowa i uśredniona głośność liczona za czas trwania programu. Parametr występuje w postaci jednej liczby wyrażonej w jednostce LUFS z dokładnością do jednego miejsca po przecinku. Określa ona uśrednioną głośność programu. Pomiar należy wykonywać miernikiem opisanym w dokumencie ITU-R BS. 1770 uzupełnionym o funkcję brakowania. Wartość progowa bramki została ustalona na -8 LU ustawioną względem wartości docelowej – 23 LUFS, a długość bloku powinien wynosić 400 ms. Wartość poziomu do którego należy normalizować sygnał audio to -23,0 LUFS.

Zakres głośności (Loudness Range (LRA)) to parametr opisujący zróżnicowanie wartości głośności programu uwzględniając statyczny jej rozkład pomijając ekstremalne wartości. Przedstawia się go w jednostce LU. Dokumenty ITU oraz EBU nie proponują określonej wartości parametru, ponieważ jest ona uzależniona od progu tolerancji słuchacza, czy rozkładu programów w słuchanej stacji. Parametr ma służyć jako pomoc przy określaniu, czy dany materiał wymaga obróbki dynamiki sygnału.

Rzeczywisty poziom szczytowy TPL (True Peak Level) to maksymalna wartość sygnału w ciągłej dziedzinie czasu. Parametr może osiągać wartość większą niż największa wartość próbki pobieranej w określonych momentach czasu. Dokładność miernika uzależniona jest zatem od częstotliwości nadpróbkowania. Ważne jest pozostawienie 1 dB zapasu poniżej 0 dBFS, by móc uwzględnić potencjalne zanieżenie wskazania o 0,5 dB. Maksymalny dozwolony poziom rzeczywistego szczytu (Maximum Permitted True Peak Level) określony w dokumencie EBU R128 to -1 dBTP, czyli decybel odniesiony do pełnej skali cyfrowej mierzony za pomocą miernika rzeczywistych szczytów.

W pracy wykonano pomiar czterech parametrów: rzeczywistego poziomu szczytowego TPL (True Peak Level), głośności wyrażonej w jednostce LUFS, wartości skutecznej RMS (Root Mean Square), która jest średnią kwadratową amplitudy sygnału z czasu obserwacji, DR, co oznacza różnicę między wartością parametru TPL, a wartością RMS liczonej z 20% najgłośniejszej części utworu.

Do pomiarów wykorzystano pięć różnych nośników, a mianowicie płytę CD, płytę gramofonową oraz trzy serwisy, które zajmują się cyfrową usługą strumieniowania muzyki: Spotify [10], Tidal [11], iTunes [12].

Do pomiarów wybrano dwa albumy. Pierwszy z nich to Maryla Rodowicz „Sing-sing”. Album, który został wydany w 1976 roku przez wytwórnię Pronit. Album został wydany na płycie gramofonowej, a następnie na płycie CD w serii Antologia Maryli Rodowicz (2012-2013) przez wydawnictwo Universal Music Polska.

Drugi album wybrany to pomiarów to Pink Floyd „The Dark Side of the Moon”. Album został wydany w 1973 roku przez wydawnictwo Harvest Records/Capitol Records. Album jest drugim najlepiej sprzedającym się albumem w historii muzyki.

PODSUMOWANIE

Obecnie na rynku istnieje wiele różnych nośników, na których można przechowywać muzykę. Każdy z nich oferuje dźwięk, który charakteryzują parametry o odmiennych wartościach. W przeprowadzonych badaniach można dostrzec doskonale różnicę pomiędzy nośnikami cyfrowymi, a analogowymi. Zakres dynamiki materiału, który się na nich znajdują odbiega znacząco od siebie. Również pomiędzy nośnikami cyfrowymi, takimi jak Spotify, Tidal czy iTunes można zauważyć spore rozbieżności w głośności materiału jak i zakresie dynamiki.

Praca oparta była na pomiarze czterech parametrów akustycznych, a mianowicie wartości TPL (True Peak Level), DR, na podstawie którego określano zakres dynamiki materiału znajdującego się na nośnikach, wartości skutecznej (RMS) oraz głośności materiału mierzonej w nowej jednostce LUFS wykorzystującej algorytm, który pozwala na dostosowanie subiektywnego sposobu percepcji dźwięku przez człowieka do wyników pomiarów.

Materiał, który mierzono był dostępny w postaci pięciu różnych nośników, a mianowicie płyt gramofonowych, płyt CD oraz trzech serwisów zajmujących się strumieniowaniem plików Spotify, Tidal oraz iTunes. Albumy, które poddano pomiarom to Maryla Rodowicz album „Sing sing” oraz Pink Floyd

album „The Dark Side of the Moon”. Jest to przykład ponadczasowych, kultowych albumów, przez co wydawane były na różnych nośnikach i w różnym okresie czasu.

Na podstawie otrzymanych wyników pomiarów można stwierdzić, że serwisy zajmujące się strumieniowaniem plików, choć oferują swoim użytkownikom bardzo podobną usługę to różnią się one pomiędzy sobą materiałem udostępnianym w swoich serwisach. Serwis Spotify dysponuje materiałem, który poddany jest dużej kompresji dynamiki, a głośność utworów jest na granicy przesterowania. Atutem pozostałych serwisów streamingowych jest odmienny format plików udostępnianych przez platformy. Sprawia ono, że pliki udostępniane przez serwisy Tidal oraz iTunes dążą do uzyskania wartości parametrów związanych z głośnością zbliżonych do materiału nagranych na płycie CD.

Pomiary wartości skutecznej RMS oraz głośności wyrażonej w jednostce LUFS pozwoliły na zauważenie rozbieżności wynikających z zastosowania odmiennych parametrów do określenia głośności materiału na nośnikach. Przyczyną tego jest wykorzystany algorytm zaprojektowany przez EBU w celu przybliżenia wyników pomiarów do rzeczywistej wartości głośności percypowanej przez człowieka.

LITERATURA

- [1] ITU-R BS.1770-1. Algorithms to measure audio programme loudness and true-peak audio level, 2009.
- [2] ITU-R BS.1770-2. Algorithms to measure audio programme loudness and true-peak audio level, 2011.
- [3] ITU-R BS.1771-1. Requirements for loudness and true-peak indicating meters, 2012.
- [4] ITU-R BT.2217. Compliance material for Recommendation ITU-R BS.1770, 2011.
- [5] EBU Technical Recommendation R 128 Loudness normalisation and permitted maximum level of audio signals, 2010.
- [6] EBU Tech Doc 3341 Loudness Metering: ‘EBU Mode’ metering to supplement loudness normalisation in accordance with EBU R 128, 2010.
- [7] EBU Tech Doc 3342 Loudness Range: A descriptor to supplement loudness normalisation in accordance with EBU R 128, 2010.
- [8] EBU Tech Doc 3343 Practical Guidelines for Production and Implementation in accordance with EBU R 128, 2010.
- [9] EBU Tech Doc 3344 Practical Guidelines for Distribution of Programmes in accordance with EBU R 128, 2010.
- [10] Spotify Press – About [Online]. Available: <https://press.spotify.com/pl/about/>.
- [11] Tidal Informacje ogólne [Online]. Available: <https://support.tidal.com/hc/pl/categories/200493591-Informacje-Og%C3%B3lne>.
- [12] Audacity 2.2.1 Manual” [Online]. Available: <https://manual.audacityteam.org/>.
- [13] TT DR Offline Meter Software Manual Version 1.1.[Online]. Available: http://www.dynami-crane.de/sites/default/files/DR-Manual-V1_1-English.pdf
- [14] Insight help documentation. [Online]. Available: <https://www.izotope.com/content/dam/izotope/support/help-guides/izotope-insight-help-documentation.pdf>

SOUND RECORDING LOUDNESS DEPENDENCE ON RECORDING MEDIUM

The paper presents a comparison of the values of acoustic parameters related to the loudness of recordings recorded on various media. It was checked the opinion regarding the increasing loudness of songs is true. The purpose of the work is to observe the acoustic parameters of recordings depending on the type of media. The following parameters were measured: the actual peak TPL (True Peak Level), loudness (in LUFS), RMS value and DR (Dynamic Range). For analysis, music tracks recorded on an analogue disc, a CD and from three different internet services (streaming) were used.

OCENA JAKOŚCI DŹWIĘKU SYGNAŁÓW MOWY I MUZYKI NADAWANYCH ZA POMOCĄ RADIOFONII CYFROWEJ DAB+

Maurycy KIN¹

Stefan BRACHMAŃSKI²

¹Politechnika Wrocławska, Wyspiańskiego 27, 50-370 Wrocław
maurycy.kin@pwr.edu.pl

²Politechnika Wrocławska, Wyspiańskiego 27, 50-370 Wrocław
stefan.brachmanski@pwr.edu.pl

W pracy przedstawiono wyniki badań subiektywnych ogólnej oceny jakości sygnałów muzycznych, a także ogólnej jakości mowy, emitowanych w systemie DAB+ na terenie Wrocławia. Badania przeprowadzono dla tradycyjnego, jednonadajnikowego systemu nadawczego oraz trzech nadajników pracujących w warunkach sieci jednoczęstotliwościowej. Do badań wykorzystano przygotowane uprzednio sygnały testowe oraz fragmenty audycji emitowanych przez 7 rozgłośni na terenie Wrocławia. Rejestracji sygnałów muzycznych dokonano w trzech punktach miasta, a sygnałów mowy w dziewięciu. W wyniku przeprowadzonych testów okazało się, że w przypadku nadawania za pomocą sieci jednoczęstotliwościowej wystarczającą jakość dźwięku muzyki zapewnia szybkość bitowa 64 kb/s, natomiast mowy – 48 kb/s, niezależnie od ocenianych sygnałów. Przy konwencjonalnym nadawaniu za pomocą jednego nadajnika uzyskano podobne rezultaty z zastrzeżeniem, że warunki odbioru sygnału radiowego były sprzyjające.

1. WPROWADZENIE

We współczesnym świecie w szybkim tempie rozwija się radiofonia cyfrowa. Na świecie liczba krajów prowadzących regularne i stałe nadawanie w systemie cyfrowym rośnie z każdym rokiem. W Polsce pierwsze regularne emisje DAB+ rozpoczęły się 1 października 2013 roku w dwóch aglomeracjach, a mianowicie warszawskiej i śląskiej. Obecnie w zasięgu emisji cyfrowego radia DAB+ znajduje się 56% ludności Polski [19], a do końca 2021 roku planowane jest pokrycie 63,7% obszaru Polski czyli ok. 81,6% ludności [4]. Niezależnie od ogólnopolskich radiostacji DAB+ rozwijane są lokalne multipleksy.

Zgodnie z zaleceniami Europejskiej Unii Nadawców (*European Broadcasting Union* - EBU) z 2013r., radiofonia cyfrowa powinna objąć swoim zasięgiem zarówno wielkie obszary (cały kraj bądź jego regiony), jak i mniejsze terytoria (np. aglomeracje miejskie). Do tego ostatniego przypadku z uwagi na koszty oraz lokalny charakter przekazywanych treści nadaje się rozwiązanie w postaci sieci jednoczęstotliwościowej. Rozwiązanie takie zastosowano we Wrocławiu: trzy nadajniki sygnału DAB+ rozmieszczono w wierzchołkach trójkąta, co umożliwia pokrycie niemalże całego obszaru miasta. Do kodowania użyto standardu HE-AAC (*High-Efficiency Advanced Audio Coding*) wersja 2 [8], przez co uzyskano możliwość poprawy jakości przy niższych przepływnościach poprzez wykorzystanie przetwarzania SBR (*Spectral Band Replication*) oraz PS (*Parametric Stereo*), umożliwiających zdecydowaną poprawę jakości dźwięku przy szybkościach 64kb/s oraz 48 kb/s [3], [11].

Pomimo znacznego postępu w tworzeniu obiektywnych metod oceny jakości transmisji sygnału mowy i muzyki, nadal jedynym wiarygodnym weryfikatorem jakości są metody wykorzystujące pomiary subiektywne, za pomocą których można wyznaczyć dopuszczalną degradację sygnałów, która nie

powoduje drastycznego spadku jakości dźwięku. Ponieważ w Katedrze Akustyki i Multimediiów przeprowadzono już badania jakości dźwięku nadawanego w systemie DAB+ w latach 2005-2009 [5], [11], przy emisji eksperymentalnej z jednego nadajnika, postanowiono zbadać jakość przekazu sygnałów mowy i muzyki w przypadku obecnego rozwiązania, wykorzystującego sieć jednoczęstotliwościową. Zrealizowana sieć oparta jest na bazie trzech nadajników rozmieszczonych na masztach budynków: Polskiego Radia Wrocław S.A., Instytutu Łączności PiB we Wrocławiu oraz na MPWiK, znajdującym się w bezpośrednim sąsiedztwie terenów Politechniki Wrocławskiej [18].

Z punktu widzenia odbiorcy, skuteczność radiofonii cyfrowej oceniana jest przede wszystkim poprzez jakość odbieranych audycji zarówno muzycznych jak i słownych (*Quality of Experience*). Inne możliwości oferowane w systemie DAB+ mają nieco mniejsze znaczenie.

Celem eksperymentu było:

- zbadanie wpływu szybkości bitowej na jakość sygnałów mowy i muzyki emitowanych w lokalnej radiofonii cyfrowej,
- zbadanie jakości odbieranych audycji słownych oraz muzycznych zarejestrowanych w różnych punktach Wrocławia.

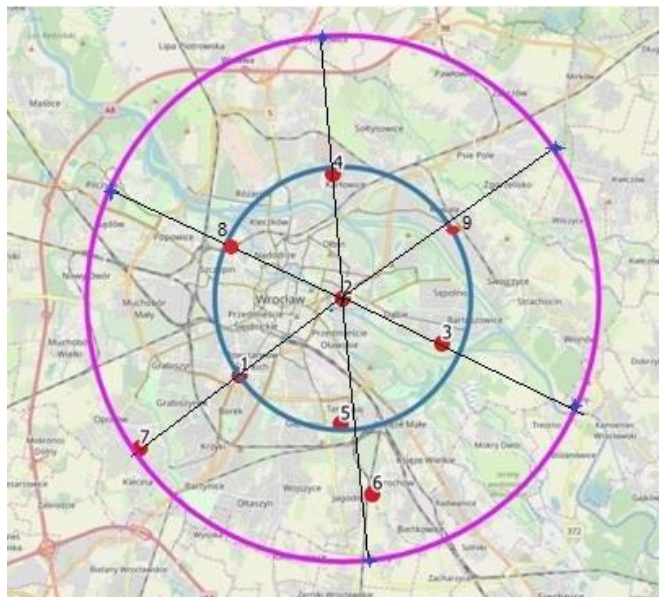
2. METODA BADAŃ

Jakość mowy i muzyki może być oceniana wg pięciostopniowej skali MOS (*Mean Opinion Score*)[10]. Do oceny jakości sygnałów muzycznych można zastosować także procedurę opartą na stu-punktowej skali MUSHRA [9]. W zależności od przyjętego kryterium oceny różny jest materiał testowy: w ocenie jakości mowy stosowane są listy zdaniowe [1], a w przypadku muzyki – odpowiednie ciągi testowe [9]. Według kryterium jakościowego, ocenę wykonuje się najczęściej metodami ACR (*Absolute Category Rating*) lub DCR (*Degradation Category Rating*) [10]. Ze względu na specyfikę odbioru programów radiowych, gdzie najważniejszym aspektem estetycznym jest percepcja materiału dźwiękowego w określonej sytuacji, a wartościowanie oceny polega na ogólnie wyrobionym poczuciu estetycznym [15], do oceny jakości wybrano metodę ACR. Badania przeprowadzono wykorzystując sygnały emitowane przez różne stacje. Ze względu na cel badań, którym była ocena jakości dźwięku emitowanego przez działające rozgłośnie, mających zdefiniowany profil i wykorzystujących konkretne procesory emisyjne, nie wykorzystano specjalnie przygotowanego ciągu sygnałów, a posłużono się materiałem dźwiękowym nadawanym przez konkretne rozgłośnie. W tym celu dokonano nagrań programów radiowych w trzech punktach miasta (teren Politechniki Wrocławskiej, ul. Powstańców Śl. - okolice Sky Tower, oraz okolice Rynku), a następnie z nagrań tych utworzono ciągi testowe, które poddano ocenie subiektywnej.

Rejestracji testowych sygnałów mowy dokonano w dziewięciu punktach Wrocławia, których położenie pokazano na rys. 1. W tym przypadku materiał testowy składał się z 10 list zdaniowych nagranych przez kobietę i mężczyznę. Listy testowe nadawane były z pięcioma szybkościami bitowymi: 32 kb/s, 48 kb/s, 64 kb/s, 96 kb/s oraz 112 kb/s. Jednej stacji radiowej została przypisana jedna szybkość bitowa, co oznacza, że na pięciu kanałach radiowych zamiast normalnych programów nadawano zestawy testowe.

Do nagrań zastosowano rejestratory ZOOM H4n PRO oraz TASCAM DR-100 MKIII, a sygnały nagrano na wyjściu liniowym radioodbiornika DAB Sangean DPR-26. Tak skonstruowany test pozwolił na zbadanie jakości dźwięku samych programów, jak też i zbadanie wpływu miejsca odbioru na jakość dźwięku. Wyselekcjonowane fragmenty przygotowano zgodnie z zaleceniami ITU[10] oraz EBU[4], przy czym czas trwania pojedynczej próbki dźwiękowej sygnałów muzycznych wynosił ok. 20 sekund, przy zachowaniu frazy lub odcinka melodii. Materiał dźwiękowy podzielono na 5 grup: muzyka rozrywkowa, jazz, muzyka klasyczna, rock oraz sygnały mowy. Wykorzystane fragmenty były nadawane z szybkościami bitowymi: 128, 96, 64 oraz 48 kb/s. Jako sygnału referencyjnego użyto dwóch nagrań z płyt CD, które także były prezentowane przez dwie z badanych rozgłośni.

Test skonstruowano w ten sposób, że próbki prezentowano w przypadkowej kolejności po 3 razy, natomiast w jednym ciągu testowym próbka wzorcowa pojawiała się tylko raz. Miało to na celu zbadanie powtarzalności ocen słuchaczy tak dla sygnału wzorcowego, jak i stabilności ocen badanych sygnałów.



Rys. 1. Rozmieszczenie punktów pomiarowych na terenie Wrocławia

W badaniach nad jakością mowy przyjęto szybkość próbkowania wynoszącą 44 100 próbek/s. Wprawdzie sygnał mowy mógłby być próbkowany z szybkością 16 000 próbek/s, lecz w ocenianym cyfrowym radio LokalDAB w procesie przygotowania materiału emisyjnego przyjęto standard próbkowania z szybkością 44 100 próbek/s zarówno dla sygnału muzycznego jak i sygnału mowy. Zmiana szybkości próbkowania sygnału mowy wiązałaby się z problemami organizacyjnymi i technicznymi. Dlatego też w badaniach do emisji przygotowano materiał testowy próbkowany z szybkością 44 100 próbek/s.

Testy odsłuchowe przeprowadzono w studio nagrań Katedry Akustyki i Multimediów Politechniki Wrocławskiej. Pomieszczenie to spełnia zalecenia dotyczące pomieszczeń odsłuchowych, a jego dodatkowym atutem był fakt, że wszyscy uczestnicy badań znali je, co nie wpływało na dekoncentrację słuchaczy związaną z adaptacją do miejsca badań. Każdy słuchacz po wysłuchaniu próbki musiał wystawić odpowiadającą swoim odczuciom ocenę w skali 1 do 5. Badanie odbyło się dla całej grupy, a słuchacze zapisywali swoje odpowiedzi na arkuszu testowym.

3. WYNIKI BADAŃ

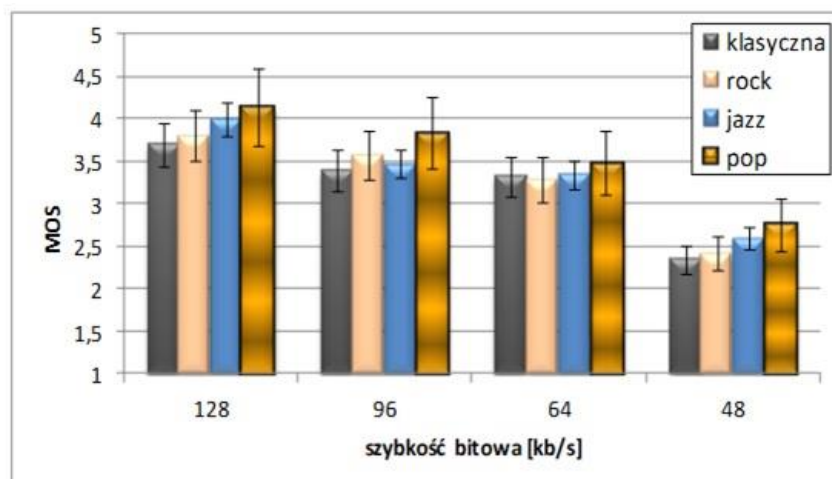
3.1. OCENA JAKOŚCI MUZYKI

Analizę wyników wykonano najpierw globalnie bez podziału na oceniane gatunki muzyki. Dla każdej szybkości bitowej wyznaczono średnią ocenę po wszystkich gatunkach muzyki bez ich rozróżnienia i po wszystkich słuchaczach. Uśredniono wyniki uzyskane dla rozważanych szybkości bitowych, bez podziału na gatunki muzyki.

Wyniki oceny jakości dźwięku sygnałów muzycznych poddano analizie statystycznej: jednorodność ocen całej ekipy sprawdzono za pomocą testu Bartletta jednorodności wielu wariancji na poziomie istotności $\alpha = 0,05$ ($\chi^2 = 4,17 < \chi^2_{kr} = 5,98$), co pozwoliło na uśrednienie ocen wszystkich członków ekipy

odsluchowej. Wykorzystując test ANOVA na poziomie istotności $\alpha = 0,05$ wykazano brak wpływu miejsca odbioru sygnału radiowego ($F = 1,39 < F_{kr} = 3,68$) na ocenę jakości, natomiast wykazano, że gatunek muzyki ma wpływ na ogólną ocenę jakości dźwięku ($F = 6,53 > F_{kr} = 5,16$).

Otrzymane uśrednione wartości subiektywnej oceny jakości MOS badanych fragmentów ze wszystkich punktów pomiarowych wraz z wartościami odchyłeń standardowych przedstawiono na rys. 2. Zwraca uwagę monotoniczne pogorszenie ogólnej jakości wszystkich gatunków muzyki wraz ze zmniejszeniem szybkości bitowej: ocena MOS spada z około 4,0 do ok. 2,5 przy zmniejszeniu przepływności ze 128 kb/s do 48 kb/s. Analizując otrzymane wyniki można stwierdzić, że począwszy od szybkości bitowej wynoszącej 64 kb/s wartość wskaźnika MOS przekracza 4,0, natomiast począwszy od 96 kb/s ocena jest porównywalna z oceną jakości nagrania oryginalnego.



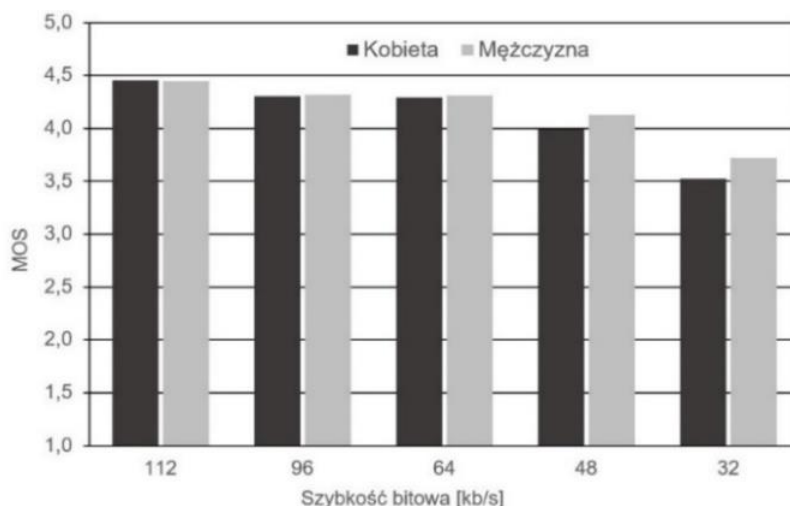
Rys. 2. Wyniki oceny jakości muzyki dla zastosowanych gatunków [13]

Przyjmując, że akceptowalną jakość uzyskuje się dla oceny MOS powyżej 3,0, to można zauważyć, że szybkość bitowa 48 kb/s nie zapewnia wystarczającej jakości dla odbioru muzyki. Badania pokazały, że oceny MOS uzyskane dla wszystkich gatunków muzyki rozrywkowej są wyższe niż dla muzyki klasycznej o ok. 0,3 – 0,6, co jest spowodowane najprawdopodobniej bardziej ścisłymi kanonami i naturalnością dźwięku obowiązującymi dla muzyki poważnej. Dodatkowo stwierdzono, że zależność jakości od przepływności jest monotoniczna, co pozwala oszacować minimalną wartość przepływności, niezbędną do uzyskania jakości dźwięku na założonym poziomie [7]. Warto zaznaczyć także, że do badań wykorzystano typowy dla danego profilu stacji materiał muzyczny, po który sięgają rozgłośnie radiowe, często przetworzony i dostosowany „brzmieniowo” do grupy odbiorców będącej adresatem konkretnego programu lub rozgłośni. Tego typu zabiegi nie pozostają bez wpływu na strukturę sygnału, która powoduje, że materiał dźwiękowy, zwłaszcza muzyczny brzmi możliwie jednakowo [14].

3.2. OCENA JAKOŚCI MOWY

Dla każdego warunku transmisyjnego wyznaczono średnią ocenę jakości MOS i odchylenie standardowe. Następnie sprawdzono warunek 3σ czyli rozrzut ocen jakości mowy. Oceny odbiegające od średniej o wartość 3σ zostały odrzucone, a obliczenia wykonano ponownie bez odrzuconych ocen.

Wykorzystując test t-Studenta, sprawdzono, czy różnice w wartościach średnich ocen słuchaczy są statystycznie istotne, czy nie. Postawiono hipotezę, że różnice są nieistotne. Stwierdzono na poziomie istotności $\alpha = 0,05$, że nie ma podstaw do odrzucenia hipotezy o nieistotności różnic między średnimi poszczególnych słuchaczy. Uśrednione po dziewięciu punktach Wrocławia wyniki pomiarów subiektywnych uzyskanych dla wszystkich szybkości przedstawiono na rys. 3.



Rys. 3. Wyniki oceny jakości mowy uśrednione po wszystkich badanych punktach Wrocławia.

Oceniając wyniki otrzymane dla sygnałów mowy można stwierdzić, że dla przepływności 48 kb/s otrzymano satysfakcjonującą słuchaczy jakość. Dla badanych przepływności, różnice w wartościach oceny w pięciostopniowej skali nie są duże jako, że wartość MOS dla 128 kb/s wynosi 4,42, natomiast dla 48 kb/s – 4,03. Wg. zalecenia ITU-T P.800[10] dobrej jakości transmisji sygnału mowy odpowiada wartość MOS = 4,0.

4. OMÓWIENIE WYNIKÓW

Analizując otrzymane wyniki można stwierdzić, że przy szybkości bitowej wynoszącej 128kb/s wartość wskaźnika MOS jest bliska 4.0, co zapewnia dobrą jakość dźwięków muzycznych nadawanych w badanym systemie, natomiast najmniejsza przepływność gwarantująca dostateczną (akceptowalną) jakość przekazu muzycznego dla wszystkich gatunków muzycznych wynosi 64kb/s. Należy zaznaczyć, że słuchacze oceniali jakość próbki oryginalnej z muzyką rockową podobnie, jak dla sygnałów emitowanych przez DAB+ (MOS 4,1), przy szybkości 128 kb/s, natomiast próbka referencyjna z muzyką klasyczną zawsze oceniana była wyżej (MOS 4,5) w porównaniu do ocen przy przepływności 128 kb/s. Oznacza to, że dla zapewnienia szczególnie wysokiej jakości transmisji z estetycznego punktu widzenia należałoby zastosować większą szybkość bitową. Jest to spowodowane przede wszystkim możliwością utraty szczegółowych informacji dotyczących atmosfery i przestrzennych atrybutów dźwięku, często eksplorowanych właśnie w nagraniach muzyki klasycznej[3], [11]. Zmiany barwy dźwięku, wprowadzane na etapie kodowania, nie są aż tak wyraźne, co gwarantuje wysoką ocenę MOS dla innych gatunków muzyki, ponieważ ten atrybut sceny dźwiękowej jest najbardziej istotnym czynnikiem w subiektywnej ocenie jakości dźwięku [17].

Oceniając wyniki otrzymane dla sygnałów mowy można stwierdzić, że dla przepływności 48 kb/s otrzymano zadowalającą słuchaczy jakość. Dla badanych przepływności, różnice w wartościach oceny MOS nie są duże jako, że wartość MOS dla 128 kb/s wynosi 4,42, natomiast dla 48 kb/s – 4,03. Wg. zalecenia ITU-T P.800 [10] dobrej jakości transmisji sygnału mowy odpowiada wartość MOS = 4,0.

5. PODSUMOWANIE

Podsumowując wyniki badań jakości transmisji sygnałów muzyki i mowy w warunkach rzeczywistej emisji programów w radiofonii cyfrowej DAB+ można przyjąć, że wysoką jakość programów informacyjnych (audycje słowne) uzyskuje się już dla przepływności wynoszącej 48 kb/s, natomiast dla programów muzycznych minimalna wartość przepływności, zapewniająca akceptowalną jakość wynosi 64 kb/s, a w przypadku, gdy szczególnie istotna jest warstwa estetyczna nagrań (jak to ma miejsce dla muzyki klasycznej) przepływność powinna wynosić 96 kb/s lub nawet 128kb/s, co jest zgodne z wynikami badań nad jakością dźwięku kodowanego za pomocą różnych kodeków [16].

Analiza otrzymanych wyników potwierdza wnioski uzyskane w badaniach symulacyjnych opartych na programowym kodowaniu AAC[2], a także na ocenie audycji słownych nadawanych na żywo[13]. Z tych wcześniejszych badań wynika, że już przy szybkości bitowej 48 kb/s uzyskuje się wartość wskaźnika MOS powyżej 4. Należy także wspomnieć, że otrzymane wyniki dla emisji SFN nieznaczająco się różnią od nadawania sygnału DAB+ za pomocą jednego nadajnika [5], [11], [12]. Można zatem stwierdzić, że sposób emisji (sieć jednoczęstotliwościowa vs. jeden nadajnik) nie wpływa istotnie na ocenę jakości dźwięku transmitowanego w systemie DAB+.

LITERATURA

- [1] BRACHMAŃSKI S., Wybrane zagadnienia oceny jakości transmisji sygnału mowy. Wrocław, Oficyna Wydawnicza Politechniki Wrocławskiej, 2015.
- [2] BRACHMAŃSKI S., „Quality evaluation of speech AAC and HE-AAC coding”, Proc. of Joint Conference Acoustics, Ustka, Poland [Danvers MA] IEEE, cop. 2018.
- [3] BRACHMAŃSKI S., KIN M., „Quality evaluation of sound broadcasted via DAB+ system based on a single frequency network”. 144th Conv. of Audio Eng. Society, Milan, Italy. 2018, Conv. paper 10004
- [4] Cyfrowa Polska, „DAB+ postęp cyfryzacji radia w Polsce na tle światowym i europejskim. Analiza sytuacji i rekomendacje”, 2019, https://cyfrowapolska.org/wp-content/uploads/2019/10/Raport_radio_DAB_2019_final.pdf (dostęp 21.03.2020).
- [5] DOBRUCKI A., OSTROWSKI M., BŁASIAK K., KIN M., MALECZEK S., „Badanie jakości dźwięku sygnałów przesyłanych z wykorzystaniem radia cyfrowego DAB+”. Przegląd Telekomunikacyjny. Wiadomości Telekomunikacyjne. 2010, R 83 (6), 2010, 488 – 491.
- [6] EBU, „Listening Conditions for the Assessment of Sound Programme Material”. Technical Recommendation R22-1999, EBU Geneva, Switzerland., 1999.
- [7] GILSKI P., STEFAŃSKI J., „Subjective and objective comparative study of DAB+ broadcast system”. Archives of Acoustics, 2017, 42 (1): 3 – 11.
- [8] HOEGH W., LAUTERBACH T., Digital Audio Broadcasting., 2nd ed. Wiley. England, 2003.
- [9] ITU-R Recommendation BS-1534, „Method for the subjective assessment of intermediate quality level of coding systems”, 2001-2003
- [10] ITU-T Recommendation P.800, „Method for subjective determination of transmission quality”, 1996.
- [11] KIN M., „Subiektywna ocena jakości nagrań muzycznych nadawanych w systemie DAB+”, Przegląd Telekomunikacyjny. Wiadomości Telekomunikacyjne., 2013, R 86 (6): 494 – 497.
- [12] KIN M., „Subjective evaluation of sound quality of musical recording transmitted via DAB+ system”. 134th Convention of Audio Engineering Society, Rome, Italy. 2013, Conv. paper 8874.
- [13] KIN M., BRACHMAŃSKI S., „Badanie jakości dźwięku sygnałów mowy i muzyki emitowanych za pomocą jednoczęstotliwościowej sieci DAB+ we Wrocławiu”, Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne, 2019, 6, 392 – 395.
- [14] MCGREGOR I, CUNNINGHAM S., „Comparative Evaluation Of Radio And Audio Logo Sound Design”. J. Audio Eng. Soc., 2015, 63, (11): 876-887.
- [15] PAWŁOWSKI T., „Subiektywizm”, Wybór pism estetycznych, Universitas, Kraków, 1987, 103 – 140.
- [16] ROGOWSKA A., „Czy słyszymy kompresję?”, Mat. XVI Symp. Inżynierii i Reżyserii Dźwięku ISSET 2015, Warszawa, 140-145.

- [17] WOSZCZYK W., KO D., „*Virtual acoustics for musicians; subjective evaluation of a virtual acoustic system in performance of string quartets.*”, J. Audio Eng. Soc., 2018, 66, (9):712 -723
- [18] ZIELIŃSKI R. J., „*Warunki uzyskania poprawnego odbioru sygnału DAB+ w sieci jedno-częstotliwościowej*”, Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne, 2017, 6, 509 – 512.
- [19] www.worlddab.org/public_document/file/1277/WorldDAB_Infographic_Q2_2019_A4_with_sources_FINAL_ONLINE_ENGLISH_04_03_2020.pdf?1583318859 (dostęp 21.03.2020).

QUALITY EVALUATION OF SPEECH AND MUSICAL SIGNALS BROADCASTED VIA DAB+

Paper presents the results of quality assessment of speech and musical signals transmitted via DAB+. Two DAB+ systems were investigated: the first was conventional one, with single emitter while the second was based on the single frequency network, with three emitters. As a sound material, the parts of various programs broadcasted by 7 radio stations in Wrocław have been used as well as the prepared samples. The programs were recorded in three different places (musical signals) as well as in nine city points (speech) in order to check the differences in qualities in dependence of receiving point and the broadcasting system. The overall quality of musical and speech signals have been examined. It turned out that the acceptable quality for music can be achieved at the bit rate of 64 kbps, or higher. For speech signals the good sound quality can be achieved at a bit rate of 48 kbps.

SUBIEKTYWNA OCENA JAKOŚCI SYGNAŁÓW MOWY I MUZYKI EMITOWANYCH W LOKALNYCH MULTIPLEKSACH RADIOFONII DAB+ W GDAŃSKU I WROCŁAWIU

Przemysław FALKOWSKI-GILSKI¹

Stefan BRACHMAŃSKI²

¹Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki,
Narutowicza 11/12, 80-233 Gdańsk
przemyslaw.falkowski@eti.pg.edu.pl

²Politechnika Wrocławska, Wydział Elektroniki, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
stefan.brachmanski@pwr.edu.pl

Radiofonia cyfrowa DAB+ (Digital Audio Broadcasting plus) dostępna jest dla słuchaczy w Polsce od 2013 roku. Standard ten oferuje szerokie możliwości konfiguracji multipleksów lokalnych, zarówno pod względem ilości jak i jakości nadawanych programów radiowych. Dzięki temu możliwe jest dostosowanie parametrów emitowanych sygnałów, w celu sprostania oczekiwaniom odbiorców końcowych. W przeciwieństwie do radiofonii analogowej FM sygnały audio, pochodzące od różnych nadawców, grupowane są w zbiór, określany mianem ensemble. W pracy przedstawiono wyniki subiektywnych testów oceny jakości programów radiowych, obejmujących sygnały mowy oraz muzyki. Badania przeprowadzono na lokalnych wariantach cyfrowego multipleksu dla Gdańska i Wrocławia. Przedstawiono wyniki ocen tych samych programów radiowych nadających jednakowy materiał dźwiękowy w technice analogowej FM i cyfrowej DAB+ (tzw. simulcast). Wyniki obejmowały zarówno pierwszy multipleks polskiego nadawcy (Gdańsk) jak i pionierski multipleks jednoczęstotliwościowy (Wrocław).

1. WSTĘP

Radiofonia cyfrowa DAB+ (Digital Audio Broadcasting plus) jest jednym z najpopularniejszych standardów radiodifuzji cyfrowej na świecie. Na kontynencie europejskim, a szczególnie w krajach członkowskich Unii Europejskiej, system ten wskazano jako wiodący, określany mianem Paneuropejskim. Co za tym idzie, liczba krajów prowadzących regularną i/lub testową emisję w systemie cyfrowym rośnie każdego roku. Wzrasta też liczba cyfrowych multipleksów oraz oferta programów radiowych dostępnych dla słuchaczy. Przykładowe prace dotyczące projektowania, oceny oraz badania cyfrowych multipleksów oraz usług multimedialnych można znaleźć w [1][2][3][4][5].

W Polsce pierwsze regularne emisje DAB+ rozpoczęto w październiku 2013 roku w aglomeracji warszawskiej oraz śląskiej. Obecnie, stan na rok 2020, w zasięgu cyfrowego multipleksu znajduje się 56% ludności Polski [6]. Z kolei do końca przyszłego 2021 roku planowane jest pokrycie ok. 64% obszaru Polski, a tym samym ok. 82% ludności [7]. Wraz z rozwojem ogólnopolskiej sieci nadajników DAB+ prowadzone są prace z wykorzystaniem lokalnych wariantów multipleksu.

Lokalny multipleks w Gdańsku, przeznaczony dla aglomeracji trójmiejskiej, będący regionalnym wariantem ogólnopolskiego multipleksu, uruchomiono w 2015 roku. Sygnał emitowany jest z jednego nadajnika zlokalizowanego w Chwaszczynie. Zawartość programową oraz konfigurację multipleksu przedstawiono w tabeli 1.

Tabela 1. Zawartość programowa regionalnego multipleksu DAB+ w Gdańsku.

Lp.	Profil stacji radiowej	Szybkość bitowa [kb/s]
1	Publicystyczny 1	112
2	Sztuka	128
3	Publicystyczny 2	112
4	Muzyka popularna	128
5	Informacyjny w jęz. angielskim	64
6	Informacyjny w jęz. polskim	64
7	Muzyka klasyczna	128
8	Dziecięcy	72
9	Regionalny w jęz. polskim	112
10	Regionalny w jęz. angielskim	72
11	Dane	16
12	Journaline	16

W obecnej konfiguracji zbioru (ensemble) dostępnych jest 12 usług, z czego 10 to programy radiowe emitujące sygnały audio (mowa i/lub muzyka), z kolei 2 to programy emitujące wyłącznie dane. Programy emitujące dane (Dane i Journaline) dostępne są z szybkością bitową wynoszącą 16 kb/s. W przypadku programów dźwiękowych, tzw. klasycznych programów radiowych, szybkość bitowa oscyluje na poziomie od 64 do 128 kb/s.

Jednoczęstotliwościowy multipleks we Wrocławiu, będący pionierskim multipleksem w Polsce, uruchomiono w 2018 roku. Sygnał emitowany jest z sieci 3 nadajników, rozmieszczonych na terenie Polskiego Radia Wrocław S.A., Instytutu Łączności PIB we Wrocławiu oraz Politechniki Wrocławskiej. Na potrzeby badań uruchomiono multipleks emitujący 5 programów radiowych, których szybkość bitowa wynosiła od 32 do 112 kb/s, transmitujących sygnały mowy wypowiedane przez kobietę i mężczyznę. Zawartość programową oraz konfigurację multipleksu przedstawiono w tabeli 2.

Tabela 2. Zawartość programowa jednoczęstotliwościowego multipleksu DAB+ we Wrocławiu.

Lp.	Profil stacji radiowej	Szybkość bitowa [kb/s]
1	Mowa (kobieta i mężczyzna)	32
2	Mowa (kobieta i mężczyzna)	48
3	Mowa (kobieta i mężczyzna)	64
4	Mowa (kobieta i mężczyzna)	96
5	Mowa (kobieta i mężczyzna)	112

Z punktu widzenia odbiorcy końcowego, czyli słuchacza, kluczowym parametrem jest jakość dostępnych programów radiowych, a w szczególności audycji słownych jak i muzycznych. W pracy specjalny nacisk położono na ocenę jakości programów w standardzie DAB+ w porównaniu do analogicznej oferty emitowanej w standardzie FM, jak również ocenę jakości treści emitowanych w pionierskiej sieci jednoczęstotliwościowej [8][9].

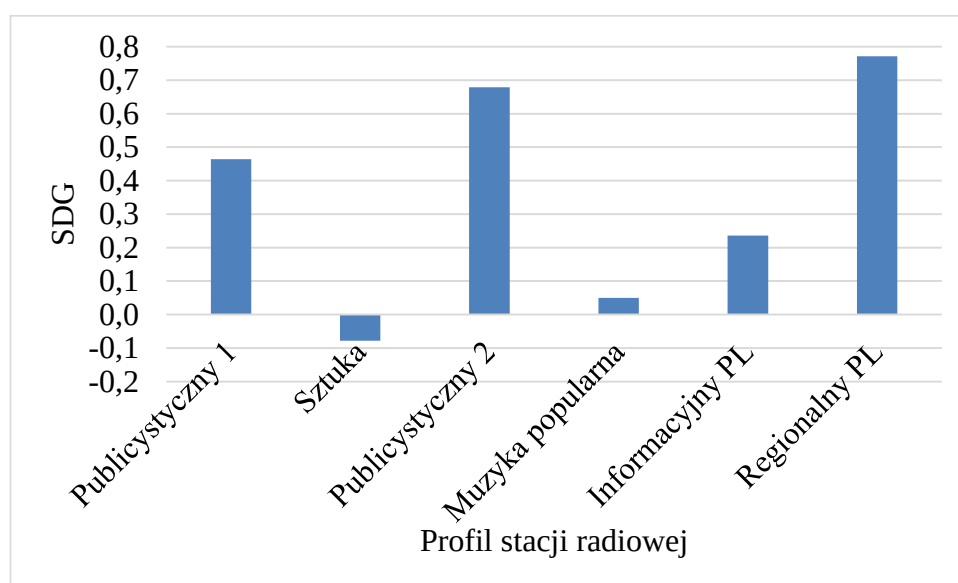
2. BADANIA ODSŁUCHOWE

Badania odsłuchowe mogą być wykonywane w oparciu o kryterium jakościowe i/lub zrozumiałościowe. Testy subiektywne przeprowadzono zgodnie z zaleceniami międzynarodowymi ITU (International Telecommunication Union) oraz EBU (European Broadcast Union), jak również krajowymi

[10][11][12][13], z wykorzystaniem 5-stopniowej skali MOS (Mean Opinion Score) w wariantcie ACR (Absolute Category Rating). W tym podejściu słuchacz dokonuje oceny w odniesieniu do własnego wzorca, utworzonego na podstawie wcześniejszych doświadczeń i preferencji, podając wartość od 1 (jakość zła) do 5 (jakość doskonała).

Badania subiektywne, dotyczące oceny jakości programów radiowych emitowanych jednocześnie w technice analogowej FM oraz cyfrowej DAB+ na terenie Gdańska, obejmowały grupę 45 uczestników. Testy przeprowadzono na zestawie programów nadawanych jednocześnie w obu technikach jak i wszystkie programy dostępne w naziemnym multipleksie.

Obecnie w tzw. simulcasie, spośród 10 regularnych emisji, dostępnych jest 6 programów radiowych. Testy przeprowadzono w środowisku wewnątrzbudynkowym z wykorzystaniem wysokiej klasy odbiornika radiowego oraz słuchawek zamkniętych. Zadaniem uczestników było dokonanie oceny ogólnej jakości programów nadawanych na żywo. Słuchacze najpierw oceniali programy DAB+, przez okres ok. 10-20 sekund, odseparowane 1 sekundową przerwą. Następnie badania powtórzono dla programów FM. Wyniki różnicowe, w formie oceny SDG (Subjective Difference Grade), przedstawiono na rys. 1.



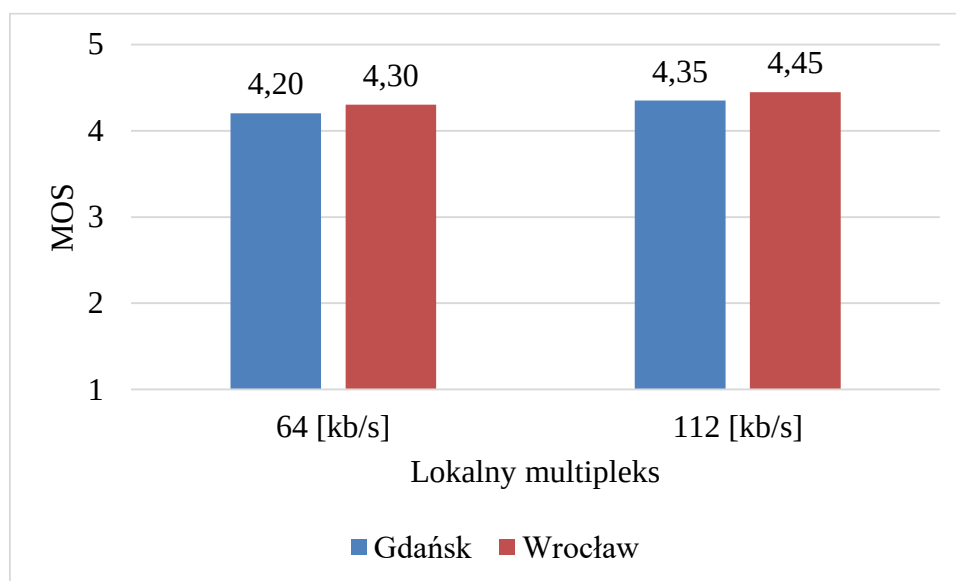
Rys. 1. Różnica w ocenie jakości programów radiowych emitowanych jednocześnie w standardzie DAB+ oraz FM w Gdańsku

Na podstawie uzyskanych wyników można określić, że oceny faworyzują system cyfrowy. Różnice na korzyść standardu DAB+ wynoszą od 0,05 (Muzyka popularna) do 0,77 (Regionalny w jęz. polskim). Jedynie w przypadku pojedynczego programu radiowego (Sztuka) lepiej oceniono standard analogowy FM.

Następnie dokonano uśrednienia wyników wszystkich programów i/lub audycji radiowych nadawanych w standardzie DAB+ na terenie Gdańska, obejmujących sygnały mowy i muzyki, aby porównać je z wynikami otrzymanymi podczas badań odsłuchowych we Wrocławiu.

Badania subiektywne, dotyczące oceny jakości programów radiowych emitujących sygnały mowy w technologii cyfrowej DAB+ na terenie Wrocławia, obejmowały grupę 20 uczestników. Materiał testowy, składający się z 10 list zdaniowych wypowiedzianych przez kobietę i mężczyznę [14], dostępny był w 5 szybkościach bitowych: 32, 48, 64, 96 oraz 112 kb/s. Audycje zostały zarejestrowane w 9 punktach pomiarowych, równomiernie rozlokowanych na terenie Wrocławia, przy pomocy odbiornika wysokiej klasy, do celów późniejszej analizy. Materiał dźwiękowy został poddany ocenie subiektywnej słuchaczy również w 5-stopniowej skali MOS w wariantcie ACR, a następnie uśredniony,

aby porównać otrzymane wyniki z rezultatami badań w Gdańsku. Porównanie dla 2 szybkości bitowych, wynoszących 64 oraz 112 kb/s, związanych z szybkością bitową regularnych audycji radiowych, przedstawiono na rys. 2.



Rys. 2. Porównanie wyników oceny jakości programów radiowych emitujących sygnały mowy w standardzie DAB+ dla multipleksów w Gdańsku i Wrocławiu

Wyniki testów odsłuchowych są niezwykle podobne. Różnice w subiektywnej ocenie słuchaczy, obejmujących sygnały mowy lektorów żeńskich i męskich, wynoszą 0,1 w skali MOS. Warto podkreślić, że ta niewielka różnica występuje zarówno dla sygnałów mowy o szybkości bitowej 64 jak i 112 kb/s. Rezultaty otrzymanych wyników podkreślają słuszność przyjętych podczas prac założeń oraz potwierdzają prawidłowość metodologii zrealizowanych badań. Stanowią także zachętę do prowadzenia dalszych wspólnych prac, których wyniki z pewnością mogą okazać się pomocne dla instytucji rządowych i pozarządowych zainteresowanych procesem cyfryzacji radiofonii naziemnej.

3. PODSUMOWANIE

Niezaprzeczalnie system DAB+ jest wiodącym standardem cyfrowej radiofonii naziemnej. W porównaniu do obecnej na rynku od wielu lat radiofonii analogowej FM, oferuje większe możliwości odnośnie liczby oraz jakości dostępnych programów radiowych. Warto wspomnieć, że w tej chwili wielu nadawców, w szczególności publicznych, emituje te same treści zarówno w standardzie FM jak i DAB+. Fakt ten wprowadza naturalne porównanie jakości nadawanych audycji przez samych słuchaczy.

Szczególną uwagę należy zwrócić na zbieżność wyników uzyskanych w dwóch ośrodkach naukowych, gdańskim oraz wrocławskim. Szybkość bitową wynoszącą 64 kb/s uznać można za górną granicę dla kodowania sygnałów mowy w przypadku audycji radiowych. Wcześniejsze wspólne prace, dotyczące transmisji głosowych komunikatów drogowych w radiofonii DAB+ [15] wykazały, że w przypadku prostych komend głosowych szybkość bitowa może zostać obniżona nawet do poziomu 24 kb/s.

Niezaprzeczalnie proces przejścia z nadawania w technice analogowej FM na cyfrową DAB+ wymaga dalszych badań. Szersza oferta programowa w naziemnej radiofonii z pewnością przyczyniłaby się do wzrostu zainteresowania i świadomości słuchaczy, przyspieszając sam proces.

Otrzymane rezultaty potwierdzają poprawność przyjętych założeń oraz metodologii badań. Stanowią też doskonały materiał oraz punkt wyjścia do dyskusji, szczególnie dla nadawców lokalnych zainteresowanych wejściem na cyfrowy multipleks. Uzyskane wyniki zachęcają także do prowadzenia dalszych prac dotyczących cyfryzacji radiofonii, rozwoju cyfrowego rynku radiowego, a docelowo przejścia z nadawania w technice analogowej FM na cyfrową DAB+.

LITERATURA

- [1] ROSLAN-KUHN K., *Analogowa wyspa – czy cyfrowe radio DAB+*, Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne, 2014, 4, 67–70.
- [2] DOBRUCKI A.B., BRACHMAŃSKI S., KIN M.J., *Objective and subjective evaluation of musical and speech recordings transmitted by DAB+ system*, Vibrations in Physical Systems, 2019, 30, 1–8.
- [3] BRACHMAŃSKI S., KIN M.J., *Quality evaluation of sound broadcasted via DAB+ system based on a single frequency network*, W: Proceedings of the 144th AES Convention, 2018, 1, 1–6.
- [4] GILSKI P., *DAB vs DAB+ radio broadcasting: a subjective comparative study*, Archives of Acoustics, 2017, 42(4), 715–723.
- [5] FALKOWSKI-GILSKI P., *On the consumption of multimedia content using mobile devices: a year to year user case study*, Archives of Acoustics, 2020, 45(2), 321–328.
- [6] WORLD DAB, *WorldDAB infographic Q2 2019*, www.worlddab.org/public_document/file/1277/WorldDAB_Infographic_Q2_2019_A4_with_sources_FINAL_ONLINE_ENGLISH_04_03_2020.pdf?1583318859 (dostęp: 13.09.2020).
- [7] CYFROWA POLSKA, *DAB+ postęp cyfryzacji radia w Polsce na tle światowym i europejskim. Analiza sytuacji i rekomendacje*, https://cyfrowapolska.org/wp-content/uploads/2019/10/Raport_radio_DAB_2019_final.pdf (dostęp: 13.09.2020).
- [8] ZIELIŃSKI R., *Fade analysis in DAB+ SFN network in Wrocław*, W: Proceedings of the International Symposium on Electromagnetic Compatibility, 2019, 1, 106–113.
- [9] ZIELIŃSKI R., *Analysis and comparison of the fade phenomenon in the SFN DAB+ network with two and three transmitters*, International Journal of Electronics and Telecommunications, 2020, 66(1), 85–92.
- [10] ITU Recommendation P.800, *Methods for subjective determination of transmission quality*, ITU, 1996.
- [11] ITU Recommendation BS.1284, *General methods for the subjective assessment of sound quality*, ITU, 2003.
- [12] EBU Technical Recommendation R22, *Listening conditions for the assessment of sound programme material*, EBU, 1999.
- [13] BRACHMAŃSKI S., *Wybrane zagadnienia oceny jakości transmisji sygnału mowy*, Oficyna Wydawnicza Politechniki Wrocławskiej, 2015.
- [14] BRACHMAŃSKI S., *Materiał testowy w pomiarach jakości mowy w radiofonii DAB+*, W: Postępy akustyki, Polskie Towarzystwo Akustyczne, 2017, 251–262.
- [15] FALKOWSKI-GILSKI P., BRACHMAŃSKI S., DOBRUCKI A.B., *Transmisja głosowych komunikatów drogowych w radiofonii cyfrowej DAB+*, Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne, 2019, 6, 334–337.

SUBJECTIVE QUALITY ASSESSMENT OF SPEECH AND MUSIC SIGNALS EMITTED IN LOCAL DAB+ RADIO MULTIPLEXES IN GDAŃSK AND WROCLAW

The DAB+ (Digital Audio Broadcasting plus) radio is available in Poland since 2013. This standard offers a wide range of capabilities when it comes to configuring local multiplexes, including number as well as quality of transmitted radio programs. Thanks to this it is possible to adjust the parameters of emitted signals, in order to meet expectations of end users. Compared to analog FM radio audio signals from different broadcasters are grouped in the so-called ensemble. In this paper we describe results of a subjective quality assessment of radio programs, including speech and music signals. The tests were carried out considering local variants of the digital multiplex in Gdańsk and Wrocław. We describe results of evaluation of equivalent radio programs broadcasting the same audio content in analog FM and digital DAB+ (so-called simulcast). They include both the first Polish multiplex (Gdańsk) as well as the pioneer single-frequency multiplex (Wrocław).

ACOUSTICS COURSE FOR INDUSTRY (ACI)

Marcin ŁUCZAK¹

Karolina JARUSZEWSKA¹

Andrzej DOBRUCKI²

¹KFB Acoustics Sp. z o.o., Mydlana 7, 51-502 Wrocław
k.jaruszevska@kfb-acoustics.com

²Politechnika Wrocławska, Wyspiańskiego 27, 50-370 Wrocław
andrzej.dobrucki@pwr.edu.pl

Temat nadmiernego hałasu oraz konieczności ochrony przed nim staje się coraz częściej obecny w świadomości ludzi, szczególnie tych żyjących w większych miastach i aglomeracjach miejskich, ale nie tylko. Osoby, które do niej pory nie interesowały się zagadnieniami z dziedziny akustyki czy zarządzania hałasem, coraz dotkliwiej odczuwają skutki ekspozycji na nadmierny hałas. Tych skutków jest wiele – są to między innymi zmęczenie i rozdrażnienie, czy zaburzenia snu i koncentracji.

Odpowiadając na potrzebę stworzenia kompleksowego programu edukacyjnego z dziedziny akustyki przemysłowej, firma KFB Acoustics wraz z Partnerami stworzyła kurs ACI – Acoustic Course for Industry (Kurs Akustyki Przemysłowej). Kurs powstał w ramach projektu unijnego Erasmus + i jest skierowany do inżynierów pracujących w dyscyplinach innych niż akustyka oraz dla kadry zarządzającej, podejmującej decyzje dotyczące zarządzania hałasem w przedsiębiorstwie. Celem kursu jest zwiększanie powszechnej świadomości w zakresie hałasu przemysłowego oraz podnoszenie kompetencji i umiejętności praktycznych, pozwalających na lepsze zarządzanie hałasem w miejscu pracy i poza nim. W stworzenie kursu zaangażowało się wielu badaczy z różnych uniwersytetów europejskich, jak i specjalistów z firm zajmujących się akustyką. Partnerami projektu są Politechnika Wrocławska, Uniwersytet KU Leuven, Wydział Inżynierii Elektrycznej i Komputerowej Uniwersytetu w Zagrzebiu oraz zespół techniczny – firma Jazzy.Pro.

Kurs składa się z czterech części, których wspólnym mianownikiem jest problem hałasu i zarządzanie nim w przedsiębiorstwie.

- Pierwsza część kursu obejmuje wprowadzenie do tematu i prezentuje zagadnienia dźwięku oraz hałasu w szerokim kontekście ich występowania w świecie.
- Druga część to przedstawienie ciekawych, a zarazem praktycznych informacji na temat wpływu hałasu na zdrowie i samopoczucie człowieka.
- Trzecia część kursu dotyczy podstawowych terminów i pojęć związanych z hałasem.
- Czwarta, ostatnia część kursu prezentuje case studies, czyli kilka różnych studium przypadku, które ilustrują najczęściej spotykane problemy i wyzwania, stojące przed osobami zarządzającymi hałasem w przedsiębiorstwie. W tym rozdziale omawiane są też najpowszechniejsze strategie radzenia sobie z nadmiernym hałasem w zakładach przemysłowych.

Tym, co odróżnia kurs ACI od innych, wciąż nielicznych materiałów edukacyjnych z zakresu akustyki, jest jego innowacyjny charakter. Materiały do kursu obejmują zestaw różnorodnych próbek dźwiękowych oraz materiałów wizualnych, nakręconych w technologii 360 stopni, dopasowanej do odzwierciedlania w okularach wirtualnej rzeczywistości (VR). Co ważne, nie trzeba posiadać takich okularów,

żeby móc skorzystać ze wszystkich materiałów. Ideą, która przyświecała twórcom kursu, jest zwiększenie świadomości jak największej grupy ludzi na temat otaczającego nas hałasu i dlatego materiały, które powstały w ramach kursu, są łatwo dostępne dla każdego. Wystarczy mieć dostęp do Internetu oraz słuchawki, które można podłączyć do komputera lub laptopa, żeby móc odtwarzać je w zaciszu swojego domu. Aby w pełni móc korzystać z filmów 360 stopni, należy przytrzymać i przeciągać obraz myszką, zarówno w górę i dół, jak i dookoła własnej osi. Co ważne, wraz ze zmianą wyświetlanego obrazu zmienia się też odtwarzany dźwięk.

Twórcy kursu ACI i towarzyszących mu materiałów dołożyli wszelkich starań, aby wiedza przekazywana w ramach kursu była rzetelna i uporządkowana, ale też łatwo przyswajalna dla osób biorących udział w kursie. W związku z tym, materiał w ramach kursu jest zorganizowany w lekcje, które prezentują kluczowe zagadnienia związane z danym blokiem tematycznym. O tempie przechodzenia do następnych lekcji i kolejności analizowanych zagadnień decyduje sam uczestnik kursu. Materiały multimedialne podzielone są na próbki audio i filmy wideo – każdy z nich opatrzony jest odpowiednim opisem tak, aby osoba korzystająca z materiałów mogła umiejscowić wiedzę przekazywaną w ten sposób w odpowiednim kontekście. Istnieje możliwość wielokrotnego odtwarzania materiałów audio i wideo, jeśli jest taka potrzeba. Aby mieć pewność, że materiał został dobrze przyswojony, kurs zakończony jest testem w formie quizu, sprawdzającego wiedzę.

Kurs ACI to nowoczesne, interaktywne narzędzie, pozwalające profesjonalistom w dziedzinach pokrewnych akustyce na uzupełnienie swojej wiedzy z zakresu zarządzania hałasem, szczególnie w kontekście funkcjonowania zakładów przemysłowych, ale nie tylko. Wierzymy, że dzięki przyjaznej dla użytkowników formie przekazywania wiedzy, stworzony kurs będzie w realny sposób przyczynić się do zwiększania społecznej świadomości na temat problemu hałasu i że zachęci on osoby w żaden sposób niezwiązane z akustyką do zdobywania dalszej wiedzy na temat tego problemu.

Kurs dostępny pod adresem: <https://aci.acoucou.org/>