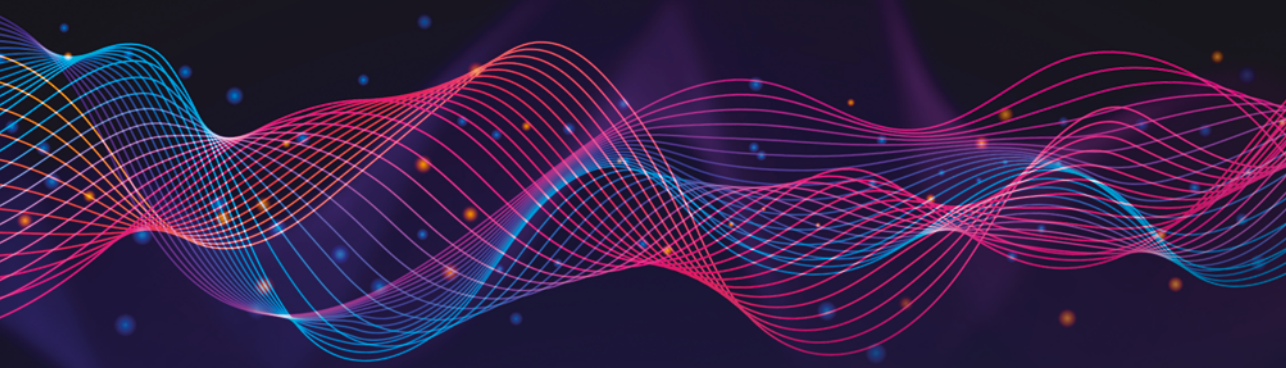
pod redakcją

KRZYSZTOFA J. OPIELIŃSKIEGO

# Postępy badań w inżynierii dźwięku i obrazu

## Nowe trendy i zastosowania technologii dźwięku wielokanałowego oraz badania jakości dźwięku



**Oficyna Wydawnicza Politechniki Wrocławskiej**
Wrocław 2021

# Postępy badań w inżynierii dźwięku i obrazu

## Nowe trendy i zastosowania technologii dźwięku wielokanałowego oraz badania jakości dźwięku

pod redakcją
Krzysztofa J. Opielińskiego

Oficyna Wydawnicza Politechniki Wrocławskiej
Wrocław 2021

Recenzenci

Andrzej CZYŻEWSKI, Andrzej DOBRUCKI, Tadeusz KAMISIŃSKI, Piotr KLECZKOWSKI, Bożena KOSTEK, Piotr KOZŁOWSKI, Ewa ŁUKASIK, Mirosław MEISSNER, Witold MICKIEWICZ, Krzysztof OPIELIŃSKI, Janusz PIECHOWICZ, Przemysław PLASKOTA, Anna PREIS, Aleksander SĘK, Ewa SKRODZKA, Paweł STRUMIŁŁO, Maciej WALCZYŃSKI, Jerzy WICIAK, Jan ŻERA

Opracowanie redakcyjne i korekta

Dorota RAWA

Projekt okładki

Paweł SPALENIAK

Opracowanie typograficzne

Stanisław GANCARZ



OFICYNA WYDAWNICZA POLITECHNIKI WROCŁAWSKIEJ
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
http://www.oficyna.pwr.wroc.pl
e-mail: oficwyd@pwr.wroc.pl

ISBN 978-83-7493-183-0
https://doi.org/10.37190/ido2021

Druk i oprawa: beta-druk, www.betadruk.pl

# Spis treści

Słowo wstępne
Krzysztof Opieliński ..... 5

1. System dźwięku przestrzennego a system wirtualnej akustyki
Piotr Z. Kozłowski ..... 9
2. Produkcja dźwięku immersyjnego. Praktyczne metody i zastosowania dźwięku ambisonicznego wyższego rzędu do tworzenia produkcji audiowizualnych VR/360°
Jan Skorupa, Maciej Głowiak ..... 29
3. Ambisoniczna mapa wybranych miejsc w Trójmieście z obrazem 360°
Cezary Pietrzak, Piotr Odya ..... 47
4. Techniki wielokanałowe wykorzystywane w koncertach i nagraniach muzycznych na odległość
Bartłomiej Mróz, Piotr Odya, Bożena Kostek ..... 67
5. Rozproszony system generowania, edycji i transmisji dźwięku wykorzystujący interfejsy Web Audio API, WebRTC i Web MIDI API
Marcin Walczak, Ewa Łukasik ..... 83
6. System sterowania kolumny z niezależnie sterowanymi głośnikami
Michał Łuczyński ..... 105
7. Praktyczna implementacja przetwornika o stałej szerokości wiązki do projektowania monitorów studyjnych
Tomasz Nowak, Bartłomiej Kruk ..... 123
8. Ocena ogólnej jakości oraz wybranych atrybutów dźwięku sygnałów muzyki i jakości mowy nadawanych za pomocą radiofonii cyfrowej DAB+
Maurycy Kin, Stefan Brachmański ..... 137
9. Subiektywny pomiar jakości sygnałów mowy i muzyki w lokalnych multipleksach radiofonii DAB+ w Gdańsku i Wrocławiu
Przemysław Falkowski-Gilski, Stefan Brachmański ..... 157
10. Poziom głośności nagrań dźwiękowych w zależności od rodzaju nośnika
Przemysław Plaskota, Małgorzata Gawlińska ..... 173
11. Wpływ parametrów kompresji dynamicznej na subiektywną głośność nagrania instrumentu muzycznego
Karol Czesak, Piotr Kleczkowski ..... 185

12. Skuteczność klasyfikacji gatunków muzycznych za pomocą sieci neuronowej w zależności od typu danych wejściowych
Maciej Blaszke, Damian Koszewski, Bożena Kostek ..... 207

13. Automatyczne generowanie kolejności list utworów muzycznych
Kamila Pietrusińska, Adam Kurowski, Bożena Kostek ..... 225

14. Pomiary wskaźnika odbicia dźwięku
Paweł Dziechciński ..... 243

Indeks nazwisk autorów ..... 269

# Słowo wstępne

Słowo „akustyka" pochodzi od starogreckich: *akustós* (*ἀκουστός*) – słyszalny i *akuo* (*ἀκούω*) – słyszę. Akustyka to dział fizyki i techniki zajmujący się zjawiskami związanymi z powstawaniem, propagacją i oddziaływaniem fal akustycznych oraz ich wykorzystaniem w wielu rozmaitych dziedzinach. Ze względu na swoją ogromną różnorodność jest obecnie traktowana jako nauka interdyscyplinarna. Wzmianki o badaniach akustycznych prowadzonych przez chińskich uczonych pojawiają się już ok. 3000 lat p.n.e. Nad postrzeganiem dźwięku przez człowieka zastanawiali się po nich starożytni Grecy uznawani za twórców systemów dźwiękowych: Terpander z Antissy (VII w. p.n.e.) – spartański muzyk i poeta, Pitagoras (VI w. p.n.e.) – sławny matematyk i filozof, oraz Didymos z Aleksandrii (I w. p.n.e.) – gramatyk i filozof. W starożytności zgłębiano i rozwijano wiedzę o oddziaływaniu fal dźwiękowych przydatną głównie w modelowaniu akustyki pomieszczeń. W 20 roku p.n.e. Marcus Vitruvius Polio (Witruwiusz) – rzymski architekt i inżynier, napisał rozprawę, *De architectura*, w której scharakteryzował zjawiska pogłosu, echa i interferencji fal akustycznych. Wiedzę tę wykorzystywali antyczni architekci, na przykład umieszczali w ścianach amfiteatrów rzędy waz (obecnie nazwane by zostały rezonatorami Helmholza) w celu poprawienia właściwości akustycznych. W I wieku n.e. Severinus Boethius (Boecjusz) – rzymski filozof, w traktacie, *De institutione musica* opisał naturę dźwięku przez analogię do fal rozchodzących się po powierzchni wody.

Akustykę klasyczną zapoczątkowali w XVII w. Galileo Galilei (Galileusz) – włoski uczony, oraz pochodzący z Francji Marin Mersenne, którzy określili związek między wysokością dźwięku i częstotliwością drgań. Mersenne jako pierwszy zmierzył prędkość dźwięku w powietrzu. W 1678 roku Robert Hooke – angielski przyrodnik i eksperymentator, sformułował prawo opisujące własności sprężyste ciał stałych: stało się ono podstawą teorii wibracji i elastyczności. Matematyczny opis propagacji dźwięku przedstawił w 1686 r. Isaac Newton – angielski uczony, a sto lat później (w 1787) Ernst Florens Friedrich Chladni – niemiecki fizyk i geolog urodzony we Wrocławiu, który

prowadził doświadczenia z drgającymi strunami i płytami. Zaobserwował między innymi, że piasek rozsypany na powierzchni talerza perkusyjnego gromadzi się w miejscach, gdzie fala stojąca tworzy węzły (tzw. figury Chladniego). W XVIII wieku wielcy uczeni: Leonhard Euler – szwajcarski matematyk i fizyk, Joseph Louis Lagrange – włosko-francuski fizyk i matematyk, oraz Jean Le Rond d'Alembert – francuski intelektualista, opracowali aparat matematyczny do opisu sygnałów dźwiękowych. W 1843 roku Georg Simon Ohm – niemiecki fizyk i matematyk, ustanowił prawo akustyki: zgodnie z nim wszystkie tony muzyczne są funkcjami okresowymi czasu, przy czym ucho jest zdolne rozkładać dźwięki na składowe sinusoidalne. Także w XIX w. Hermann Ludwig Ferdinand von Helmholtz – niemiecki naukowiec, opublikował pracę o wrażeniach słuchowych jako fizjologicznej podstawie dla teorii muzyki. Opisał w niej również specyficzny rezonator nazywany obecnie rezonatorem Helmholza i wykorzystywany do kształtowania akustyki pomieszczeń. Pierwszego zapisu dźwięku dokonał natomiast Édouard-Léon Scott de Martinville – francuski zecer i księgarz. W 1857 roku opatentował on urządzenie wzorowane na błonie bębenkowej i kosteczkach słuchowych nazwane przez niego fonautografem. Składało się ono z tuby i membrany, która drgała pobudzona falami dźwiękowymi, poruszając rylec dotykający do papierowego cylindra. Urządzenie miało służyć do badania właściwości fal dźwiękowych, nie umożliwiało jednak odtwarzania takich nagrań. W 1898 roku Wallace Clement Sabine – amerykański fizyk pracujący na Uniwersytecie Harvarda, opublikował równanie, dzięki któremu możliwe jest szacowanie czasu pogłosu w pomieszczeniu. Sabine jest uznawany za ojca nowoczesnej akustyki architektonicznej.

Niezwykle zasłużoną postacią dla akustyki był również John William Strutt – brytyjski fizyk, laureat Nagrody Nobla, profesor Uniwersytetu w Cambridge, trzeci baron Rayleigh (znany jako Lord Rayleigh), który przeprowadził wiele eksperymentów akustycznych, a także skonstruował urządzenie do pomiaru intensywności dźwięku. Zwieńczeniem jego prac, a zarazem swego rodzaju podsumowaniem dotychczasowej wiedzy, była *Theory of Sound* uznawana za „pomnik" literatury akustycznej.

Na przełomie XIX i XX w. nastąpił dynamiczny rozwój akustyki charakteryzujący się niezliczoną liczbą technicznych zastosowań, w których zaczęła być wykorzystywana na szeroką skalę dotychczasowa wiedza naukowa. W tym okresie Alexander Graham Bell – szkocki naukowiec, wynalazł mikrofon (1876), a następnie połączył go z głośnikiem i w ten sposób skonstruował pierwszy telefon. Thomas Alva Edison natomiast – amerykański przedsiębiorca i wynalazca, udoskonalił telefon Bella przy użyciu cewki indukcyjnej i mikrofonu węglowego oraz skonstruował fonograf (1877): urządzenie służące do zapisu dźwięku za pomocą diamentowej igły rzeźbiącej rowek na cynowej folii

nawiniętej na stalowym walcu napędzanym początkowo korbką, a następnie mechanizmem sprężynowym. Odtwarzanie przebiegało w analogiczny sposób. Funkcję głośnika i mikrofonu pełniła duża lejkowata tuba z metalu. Później folię zastąpiły celuloidowe wałki umożliwiające zapis i powielanie 4-minutowych nagrań. Fonograf został wyparty przez gramofon wynaleziony przez Emila Berlinera – amerykańskiego przemysłowca urodzonego w Niemczech. Gramofon różnił się od fonografu zastosowaniem płaskich płyt wytwarzanych w początkowym okresie z cynku, twardej gumy i szkła, a w późniejszym z szelaku (odmiany żywicy naturalnej pozyskiwanej z wydzieliny owadów). Ideą tego wynalazku była możliwość masowego kopiowania płyt, co umożliwiło wielki rozkwit przemysłu fonograficznego. W 1906 roku Lee De Forest – amerykański radiotechnik i wynalazca, zbudował lampę elektronową: triodę, dzięki czemu z kolei stał się możliwy rozwój radia, telewizji i radaru. Miało to również ogromny wpływ na rozwój komputera. Lee De Forest opracował też metodę optycznego zapisywania dźwięku na taśmie filmowej. Wszystkie jego wynalazki zapoczątkowały powstanie nowych obszarów w akustyce związanych z przekazywaniem sygnałów akustycznych drogą elektryczną.

Kolejne lata to wręcz lawinowy rozwój akustyki skutkujący licznymi zastosowaniami w praktyce, m.in.: wykorzystano ultradźwięki do obrazowania medycznego, opracowano systemy sonarowe, rozwinęła się defektoskopia ultradźwiękowa. Nowe technologie przyczyniły się z kolei do postępu w wielu nowatorskich badaniach akustycznych. Możliwe stało się rejestrowanie, odtwarzanie, syntezowanie, transmitowanie i przekazywanie sygnałów dźwiękowych różnymi sposobami i w różnorodnych warunkach. Po mono-, stereo- i kwadrofonii opracowano zaawansowane technologie dźwięku wielokanałowego umożliwiające słuchaczowi odbiór przestrzennych wrażeń dźwiękowych przez „zanurzenie się" w otaczającym go polu akustycznym (por. np.: ambisonia, dźwięk immersyjny). Dane do tworzenia takich wrażeń wymagają przy tym zaawansowanych obliczeń, nośników i kanałów transmisyjnych o dużych pojemnościach. Istotne jest zatem opracowanie formatów rejestracji dźwięku, dzięki którym będzie można dokonać zapisu na nośnikach cyfrowych w formie zajmującej jak najmniej mniej miejsca i z jak najmniejszą utratą jakości i wierności nagrań.

W niniejszej monografii, będącej kolejnym tomem z cyklu *Postępy badań w inżynierii dźwięku i obrazu*, przedstawiamy Czytelnikom „wycinek" akustyki związany z progresją w zakresie nowych trendów i zastosowań technologii dźwięku wielokanałowego oraz badań jakości dźwięku. W książce zawarto 14 obszernych rozdziałów opracowanych przez polskich akustyków z różnych ośrodków naukowo-badawczych: Katedry Akustyki, Multimediów i Przetwarzania Sygnałów Politechniki Wrocławskiej, Katedry Systemów Multimedialnych oraz Laboratorium Akustyki Fonicznej Politechniki Gdań-

skiej, Katedry Mechaniki i Wibroakustyki Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, Instytutu Informatyki Politechniki Poznańskiej, Poznańskiego Centrum Superkomputerowo-Sieciowego. W przypadku zagadnień dotyczących dźwięku wielokanałowego autorzy skupili uwagę na:

- systemach dźwięku przestrzennego i systemach wirtualnej akustyki wspomaganych przez procesory sygnałowe umożliwiające przestrajanie akustyki pomieszczeń,
- praktycznych metodach i zastosowaniach dźwięku immersyjnego do tworzenia produkcji audiowizualnych,
- rejestracji i opracowaniu interaktywnych środowiskowych map ambisonicznych z towarzyszeniem obrazu,
- produkcji i realizacji zdalnych nagrań muzycznych i koncertów wykorzystujących techniki binauralne i ambisoniczne,
- rozproszonych systemach generowania, edycji i transmisji dźwięku za pomocą specjalizowanych interfejsów programowania wykorzystujących architekturę i protokoły sieci Web,
- kształtowaniu charakterystyki kierunkowości kolumny głośnikowej przez ustawianie odpowiednich poziomów sygnału oraz opóźnień dla poszczególnych kanałów,
- opracowaniu przetwornika o stałej szerokości wiązki do projektowania monitorów studyjnych.

Podjęte przez nich zagadnienia związane z jakością dźwięku objęły z kolei:

- ocenę jakości dźwięku programów radiowych nadawanych cyfrowo w systemie DAB+,
- badania poziomu głośności nagrań dźwiękowych w zależności od rodzaju nośnika,
- badania nad określeniem parametrów kompresji dynamicznej wywierających istotny wpływ na subiektywnie postrzeganą głośność zróżnicowanego materiału muzycznego,
- ocenę jakości wyników automatycznego doboru i klasyfikacji utworów muzycznych,
- badania nad metodyką pomiarów wskaźnika odbicia dźwięku.

Monografia została wydana dzięki staraniom Katedry Akustyki, Multimediów i Przetwarzania Sygnałów Wydziału Elektroniki, Fotoniki i Mikrosystemów Politechniki Wrocławskiej, przy wsparciu Polskiej Sekcji Audio Engineering Society oraz Polskiego Towarzystwa Akustycznego Oddział we Wrocławiu.

Krzysztof J. Opieliński

# 1. System dźwięku przestrzennego a system wirtualnej akustyki

PIOTR Z. KOZŁOWSKI

Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

Nowoczesna sala widowiskowa jest miejscem różnorodnych wydarzeń artystycznych wymagających zupełnie innych warunków akustycznych dla poszczególnych rodzajów widowisk. Sposobem na spełnienie tych różnych oczekiwań jest zainstalowanie systemów umożliwiających zmianę parametrów akustycznych pomieszczenia. Coraz większą popularnością cieszą się tzw. systemy wirtualnej akustyki wykorzystujące rozwiązania elektroakustyczne wspomagane przez zaawansowane DSP służące do znacznego przestrajania akustyki pomieszczenia.

A przy tym w systemach nagłośnienia obiektów widowiskowych coraz bardziej klasyczny system stereofoniczny zostaje zastąpiony wielokanałowymi systemami dźwięku przestrzennego. Czy te dwie opcje można połączyć? Czy wykorzystanie obu rozwiązań wymaga zainstalowania podwojonej liczby urządzeń głośnikowych? Co daje zastosowanie tych dwóch technologii? W rozdziale udzielono odpowiedzi na te i inne pytania dotyczące możliwości systemów wsparcia akustyki (*acoustic enhancement systems*) zwanych też systemami wirtualnej akustyki (*virtual acoustics*) oraz systemów dźwięku przestrzennego (*immersive sound*). System dźwięku przestrzennego będzie tutaj oznaczał system, który oprócz klasycznego już układu urządzeń głośnikowych frontowych (*front channels*) i dookólnych (*surround channels*) zainstalowanych w płaszczyźnie odsłuchowej ma również górne urządzenia (*height channels*), a czasami także dolne urządzenia głośnikowe (*bottom channels*).

# 1.1. Wymagania akustyczne współczesnej wielofunkcyjnej sali widowiskowej

Zarządzający obiektami widowiskowymi są obecnie zainteresowani możliwością organizacji różnorodnych imprez rozrywkowych i kulturalnych w jednej sali. Taka sytuacja wynika z wielu czynników. Jednym z nich jest brak możliwości utrzymywania obiektów widowiskowych jedynie z dotacji publicznych, co jest motorem ekonomicznym do realizowania w obiekcie jak największej liczby różnorodnych imprez przynoszących dochód z wynajmu sali czy sprzedaży biletów. Drugi istotny czynnik to chęć zapewnienia społecznościom lokalnym szerokiej oferty kulturalnej i rozrywkowej szczególnie znaczący w przypadku instytucji, które są jedynymi operatorami tego typu w danym rejonie.

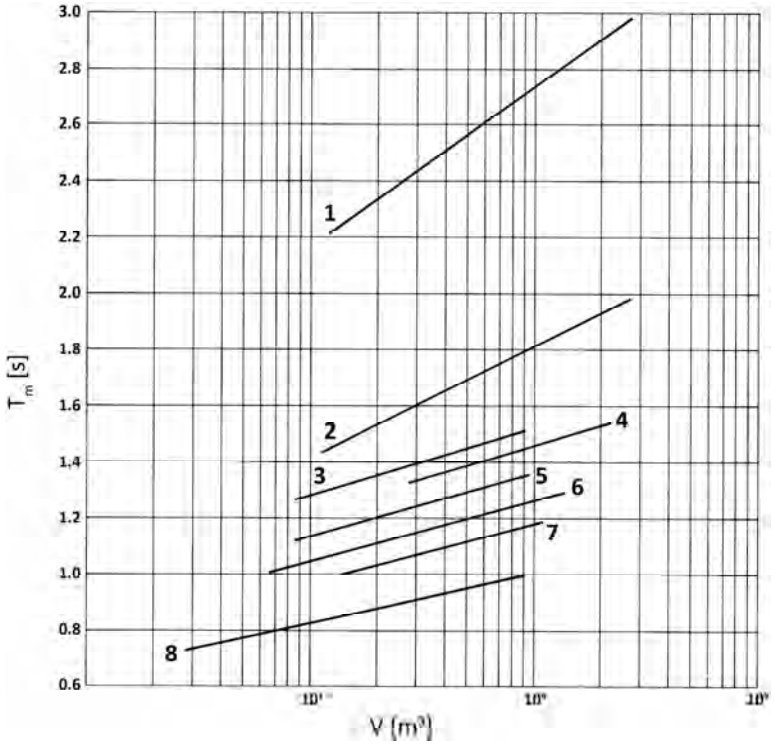
Współczesny rynek koncertowy i impresaryjny najchętniej korzysta w swojej działalności z obiektów przystosowanych do sprawnego realizowania imprez o rożnym charakterze bez konieczności akceptowania zbyt wielu niedogodności realizacyjnych czy kompromisów dotyczących jakości produkowanego widowiska. W głównej mierze dotyczy to jakości doznań dźwiękowych, na co niewątpliwie ma wpływ dopasowanie akustyki do danego typu widowiska.

Mając na uwadze długą historię tworzenia obiektów widowiskowych oraz bogatą literaturę przedmiotu, np. [1, 2, 12], można wskazać na konieczność kreowania odpowiednich warunków akustycznych w zależności od przeznaczenia obiektu (tym samym – charakteru realizowanych wydarzeń artystycznych i społecznych). Najczęściej w szybkiej weryfikacji, czy warunki akustyczne obiektu są odpowiednie do realizacji danego przedsięwzięcia, wykorzystuje się czas pogłosu (*reverberation time*). W tym celu przede wszystkim bierze się pod uwagę wskaźnik $T_m$, czyli wartość średnią czasu pogłosu dla oktaw 500 Hz i 1 kHz. Czasami uwzględnia się również oktawę 2 kHz. Ocena przygotowania sali do realizacji poszczególnych rodzajów imprez powinna być przeprowadzona również na podstawie obserwacji charakterystyki częstotliwościowej czasu pogłosu, dzięki czemu można ocenić nie tylko wartości $T_m$, lecz także parametrów takich jak BR (Bass Ratio) czy TR (Treble Ratio). Typowe wartości czasu pogłosu w przypadku poszczególnych funkcji w zależności od kubatury sali przedstawiono na rys. 1.

Współcześnie wykorzystuje się różne metody przestrajania czasu pogłosu w salach widowiskowych. Znakomita ich większość bazuje na zależności (1) przedstawionej przez Sabine'a, zgodnie z którą czas pogłosu $RT$ jest wprost proporcjonalny do objętości sali $V$, a odwrotnie proporcjonalny do chłonności akustycznej $A$ całego wnętrza.

Chłonność akustyczna $A$ $n$-tej powierzchni/elementu jest tu iloczynem powierzchni $S$ i współczynnika pochłaniania $\alpha$ tego $n$-tego elementu:

$$RT = \frac{0{,}161 \times V}{A} = \frac{0{,}161 \times V}{S_1 \cdot \alpha_1 + S_2 \cdot \alpha_2 + ..., S_n \cdot \alpha_n} \quad [\text{s}] \tag{1}$$




Rys. 1. Wartości czasu pogłosu $T_m$ w zależności od objętości $V$ dla rożnych typów obiektów/widowisk [11]; 1 – organy, chór, 2 – filharmonia, 3 – sala kameralna, 4 – opera, 5 – sala wielofunkcyjna, 6 – teatr dramatyczny, 7 – kino, 8 – sala konferencyjna

W kilku poprzednich pracach poświęconych metodom przestrajania akustyki obiektów widowiskowych (m.in. w publikacjach [9, 10]) przedstawiono bliżej założenia poszczególnych technik przestrajania czasu pogłosu spotykanych w obiektach zrealizowanych na przestrzeni ostatnich 40 lat. W systemach wirtualnej akustyki sygnały odbierane przez zestaw mikrofonów są przetwarzane w procesorach DSP zgodnie z algorytmami opracowanymi przez twórców poszczególnych systemów. W procesorach kreowany jest unikatowy sygnał dla każdego z kilkudziesięciu urządzeń głośnikowych rozmieszczonych równomiernie na ścianach i sufitach sceny oraz widowni [13, 15].

W projektowaniu typowych akustycznych sal koncertowych dużą wagę przywiązuje się do odpowiedniego odbijania i kierowania dźwięku od sufitów i plafonów akustycznych. Istotę tego zagadnienia podkreślono również podczas projektowania i badania jednej z nowszych sal koncertowych Yamaha Ginza Hall w Tokio [14], która jest stosunkowo wąska i wysoka zarazem. W przypadku projektowania systemów wirtualnej akustyki mamy do czynienia z podobną filozofią kreowania i wspierania „odbić fal akustycznych".

Nowoczesna wielofunkcyjna sala widowiskowa to nie tylko możliwość przestrajania akustyki wnętrza do potrzeb danego typu widowiska, lecz także – kreowania przestrzennych obrazów dźwiękowych na obszarze całej widowni. Aby osiągnąć ten cel, obiekty coraz częściej są wyposażane w systemy dźwięku przestrzennego (*immersive sound*). Systemy takie mają za zadanie swobodne przemieszczanie kreowanego obrazu dźwiękowego zarówno w płaszczyźnie poziomej (jest to możliwe w klasycznych już systemach dźwięku dookólnego (*surround systems*) realizowanych m.in. zgodnie z zaleceniami ITU [5]), jak i w płaszczyźnie pionowej (co ma miejsce w naturalnych środowiskach akustycznych, w których codziennie przebywamy, oraz w klasycznych salach koncertowych). Systemy dźwięku przestrzennego wyposażone są w urządzenia głośnikowe rozmieszczone równomiernie na płaszczyznach ograniczających przestrzeń widowni. Sygnały w ramach procesu miksowania są panoramowane do przestrzeni ograniczonej przez tak rozmieszczone urządzenia głośnikowe.

## 1.2. Systemy wirtualnej akustyki na przykładzie czterech różnych obiektów

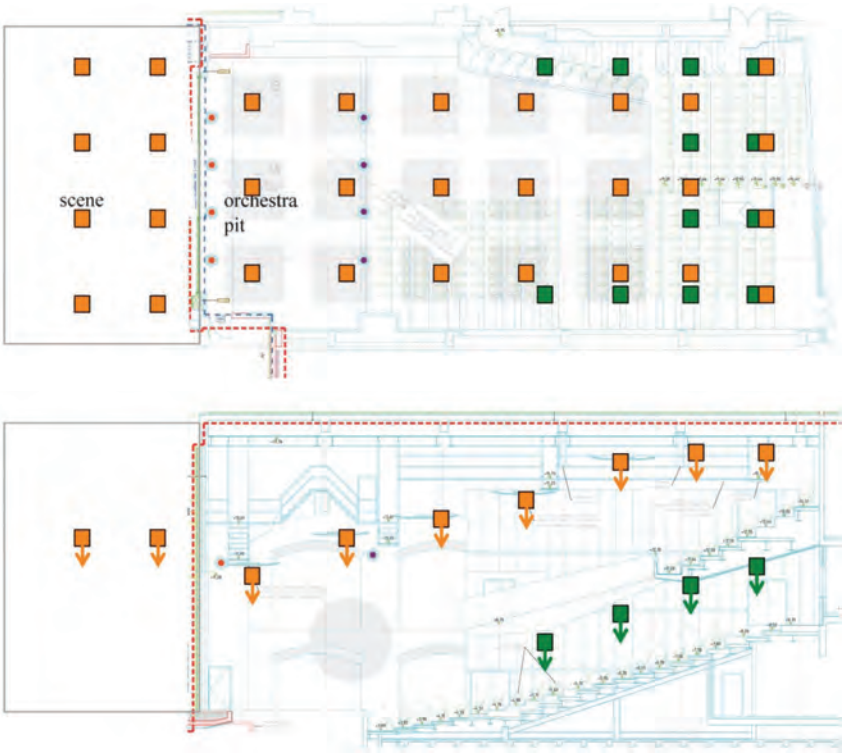
W niniejszym podrozdziale przedstawiono przykłady czterech systemów wirtualnej akustyki zainstalowanych w obiektach o różnym charakterze – operze, filharmonii, kinie, sali widowiskowej. Ze zbioru systemów strojonych i testowanych przez autora tekstu wybrano cztery, żeby pokazać możliwości oraz lokalizację urządzeń głośnikowych systemu wirtualnej akustyki. Jednocześnie prezentowane systemy działają w obiektach, w których z powodzeniem można by wykorzystywać również systemy dźwięku przestrzennego (*immersive systems*) zamiast obecnie zainstalowanych systemów dźwięku dookólnego (*surround systems*).

W przypadku każdego obiektu przedstawiono na rysunkach rozmieszczenie urządzeń głośnikowych systemu wirtualnej akustyki. Zamieszczono również wykresy prezentujące zmierzone charakterystyki częstotliwościowe czasu pogłosu dla różnych ustawień sys-

temu wirtualnej akustyki oraz banerów akustycznych. Oczywiście, systemy wirtualnej akustyki powodują zmianę wartości nie tylko czasu pogłosu, lecz także wielu innych parametrów opisujących akustykę wnętrza, ze względu jednak na ograniczoną objętość niniejszego rozdziału nie mogą zostać przedstawione i przeanalizowane.

### 1.2.1. System wirtualnej akustyki Opery na Zamku w Szczecinie

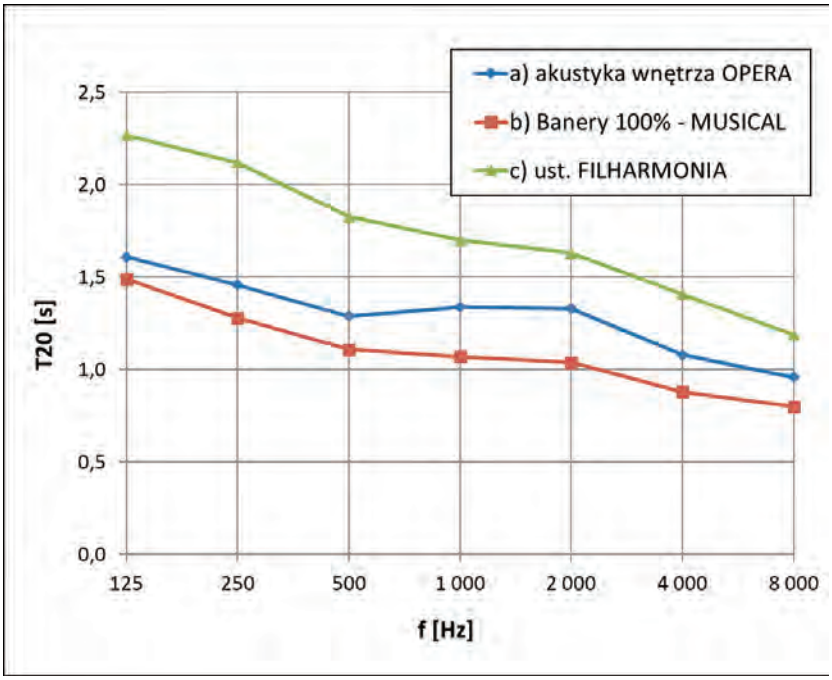
Opera na Zamku w Szczecinie jest typowym teatrem operowym, w którym wystawia się z powodzeniem spektakle operowe w sposób zarówno klasyczny, jak i nowoczesny.




Rys. 2. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w Operze na Zamku w Szczecinie przedstawione na rzucie i przekroju: 8 mikrofonów dookólnych – 4 (●) należące do podsystemu odpowiedzialnego za scenę, 4 (●) należące do podsystemu odpowiedzialnego za widownię, zwieszonych nad proscenium i orkiestronem, 32 głośniki sufitowe (■) rozmieszczone równomiernie nad widownią i sceną, 8 głośników podbalkonowych (■) rozmieszczonych równomiernie w spodzie głównego balkonu, 4 głośniki naścienne (■) rozmieszczone na ścianach widowni pod balkonami bocznymi

W obiekcie realizowane są również musicale, koncerty filharmoniczne, koncerty z wykorzystaniem systemu nagłaśniania. Tak ustawione funkcje spowodowały, że naturalna akustyka sali jest zoptymalizowana ze względu na potrzeby opery. Przy realizacji koncertów bez wykorzystania nagłaśniania (w tym filharmonicznych) czas pogłosu jest wydłużany za pomocą systemu wirtualnej akustyki, a przy realizacji koncertów kameralnych lub działań z wykorzystaniem systemu nagłaśniania czas pogłosu jest skracany za pomocą banerów akustycznych.

Na rysunku 2 – na rzucie i przekroju sali teatralnej Opery na Zamku w Szczecinie przedstawiono rozmieszczenie urządzeń głośnikowych oraz mikrofonów: system wirtualnej akustyki wykorzystuje sumarycznie 44 urządzenia głośnikowe i 8 mikrofonów. Na rysunku 3 przedstawiono zmierzone charakterystyki częstotliwościowe czasu pogłosu w zależności od ustawiania systemu wirtualnej akustyki i banerów akustycznych.

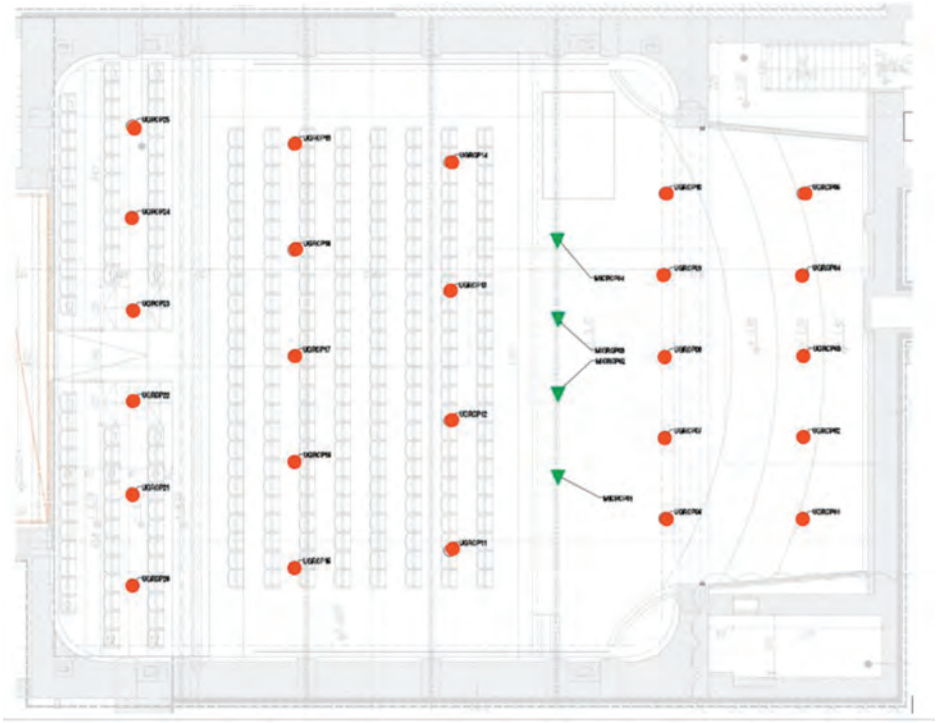



Rys. 3. Wyniki pomiarów czasu pogłosu $T_{20}$ w Operze na Zamku w Szczecinie dla różnych wariantów przestrajania akustyki: akustyka wnętrza OPERA – naturalna akustyka sali, system wirtualnej akustyki wyłączony, Banery 100% MUSICAL – naturalna akustyka sali dodatkowo wytłumiona banerami akustycznymi, system wirtualnej akustyki wyłączony; ust. FILHARMONIA – system wirtualnej akustyki włączony

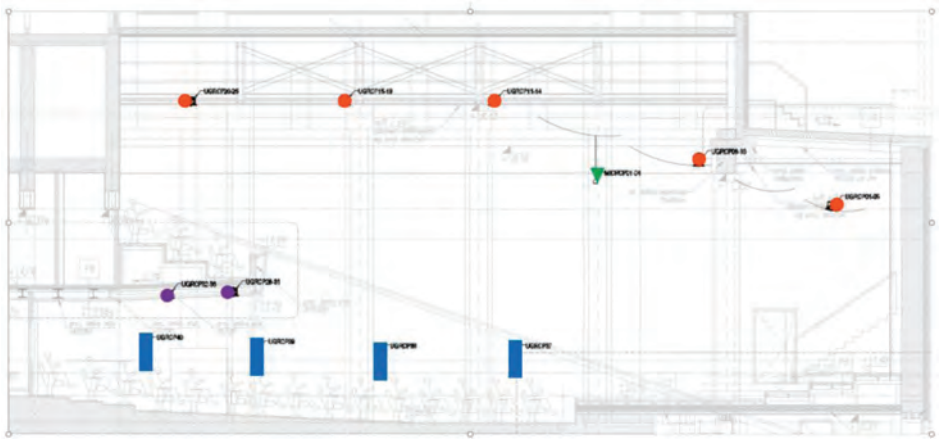
## 1.2.2. System wirtualnej akustyki Filharmonii Sudeckiej w Wałbrzychu

Sala koncertowa Filharmonii Sudeckiej w Wałbrzychu jest wykorzystywana głównie do organizacji prób i koncertów muzyki wykonywanej w sposób klasyczny, tj. bez użycia systemu nagłaśniania. Najczęściej są to koncerty filharmoniczne wymagające dużych wartości czasu pogłosu.

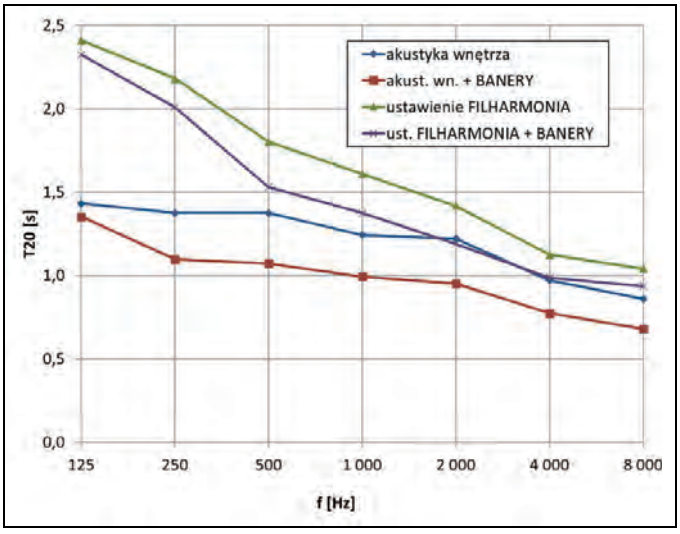
Z powodu ograniczenia kubatury, a zarazem dużej liczby miejsc siedzących na widowni oraz dużego składu orkiestry na estradzie nie jest możliwe osiągnięcie odpowiedniej wartości współczynnika kubaturowego (czyli kubatury pomieszczenia podzielonej przez liczbę osób obecnych we wnętrzu) koniecznej do uzyskania pożądanej w klasycznych



Rys. 4. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w Filharmonii Sudeckiej w Wałbrzychu przedstawione na jej rzucie: 4 mikrofony dookólne (▼) zwieszone nad frontem estrady, 25 głośników sufitowych (●) rozmieszczonych równomiernie nad widownią i estradą



Rys. 5. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w Filharmonii Sudeckiej w Wałbrzychu przedstawione na przekroju: 4 mikrofony dookólne (▼) zwieszone nad frontem estrady, 25 głośników sufitowych (●) rozmieszczonych równomiernie nad widownią i estradą, 11 głośników podbalkonowych (●) rozmieszczonych w spodzie balkonu, 8 naściennych kolumn głośnikowych (▮) równomiernie rozmieszczonych na ścianach bocznych widowni




Rys. 6. Wyniki pomiarów czasu pogłosu $T_{20}$ w Filharmonii Sudeckiej w Wałbrzychu dla różnych wariantów przestrajania akustyki po przebudowie; akustyka wnętrza – naturalna akustyka sali, system wirtualnej akustyki wyłączony, akust. wn. + BANERY – naturalna akustyka sali wraz z rozwiniętymi banerami akustycznymi, system wirtualnej akustyki wyłączony, ustawienie FILHARMONIA – system przestrajania włączony, program dla orkiestr kameralnych i filharmonicznych z maksymalną ustaloną wartością czasu pogłosu, ust. FILHARMONIA + BANERY – system przestrajania włączony, program dla orkiestr kameralnych i filharmonicznych wraz z rozwiniętymi banerami akustycznymi

koncertach filharmonicznych wartości czasu pogłosu, a także kształtu jego charakterystyki częstotliwościowej. Stąd też naturalna akustyka sali jest zoptymalizowana dla koncertów kameralnych. Warunki odpowiednie w przypadku koncertu filharmonicznego uzyskiwane są przy zastosowaniu systemu wirtualnej akustyki wykorzystującego sumarycznie 44 urządzenia głośnikowe. Na rysunku 4 przedstawiono rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki na rzucie, a na rys. 5 – na przekroju.

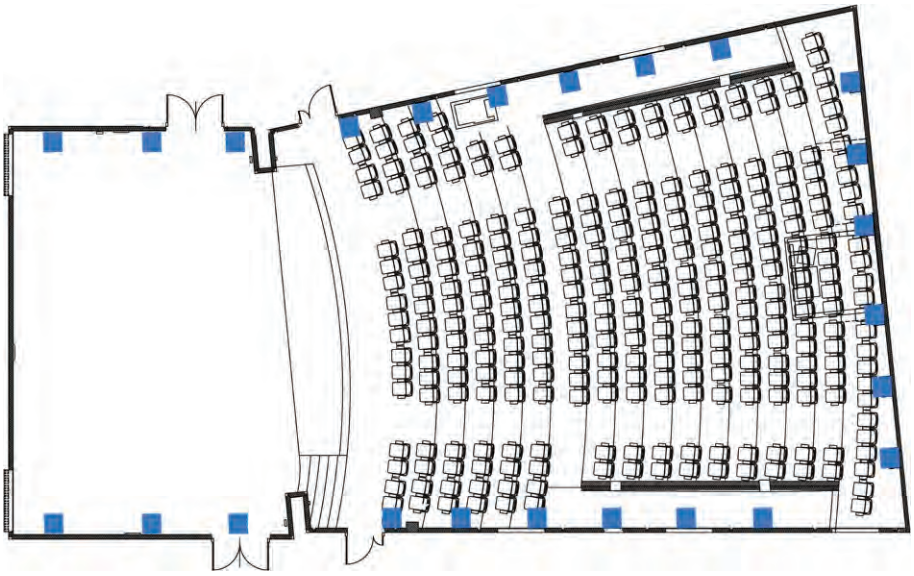
System wirtualnej akustyki umożliwia uzyskanie znacznie większych wartości czasu pogłosu niż te zmierzone. W relacji do rozmiaru sali mieszczącej 400 widzów tak duży pogłos byłby już jednak estetycznie niedopuszczalny, dlatego też ograniczono się do wartości przedstawionych na rys. 6.

## 1.2.3. System wirtualnej akustyki kina Syrena w Wieluniu

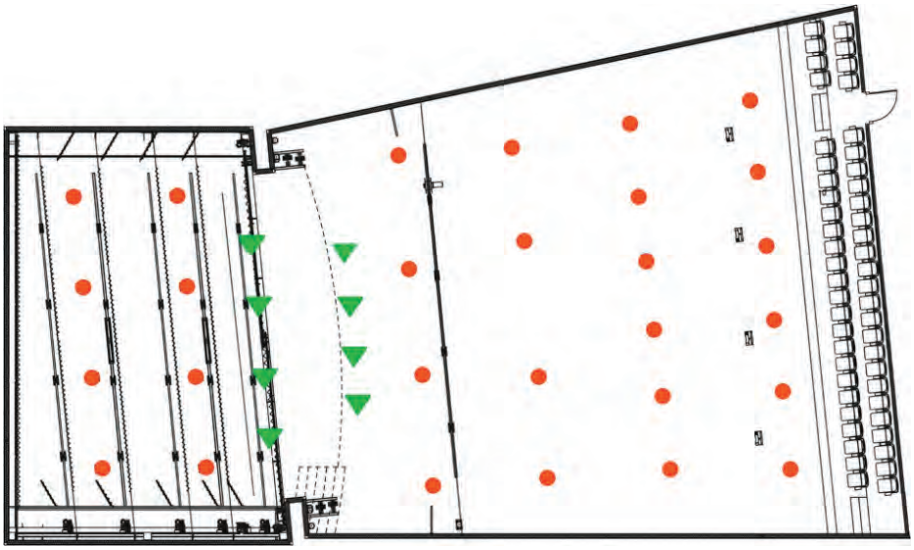
Kino Syrena w Wieluniu jest typową salą kinową, która w praktyce jest używana nie tylko do realizacji seansów kinowych, lecz także służy jako sala lokalnego domu kultury. Jest to jedyny obiekt widowiskowy w powiecie wieluńskim, w praktyce to oznacza, że stanowi salę wielofunkcyjną z wiodącą funkcją kinową. Naturalna akustyka jest zoptymalizowana na potrzeby kina, które wymaga (zgodnie z danymi przedstawionymi na rys. 1) najmniejszych wartości czasu pogłosu ze wszystkich form widowisk. W związku z tym w przypadku innych form muzycznych wspieranych systemem nagłaśniania oraz tych czysto akustycznych (w tym koncertów filharmonicznych czy chóralnych) czas pogłosu jest wydłużany za pomocą systemu wirtualnej akustyki.

Na rysunku 7 przedstawiono na rzucie parteru rozmieszczenie naściennych urządzeń głośnikowych, które równocześnie pełnią funkcję głośników *surround* systemu kinowego, a na rys. 8 – na rzucie piętra rozmieszczenie głośników sufitowych oraz mikrofonów. Rozmieszczenie mikrofonów i urządzeń głośnikowych na przekroju zaprezentowano na rys. 9. Sumarycznie system wirtualnej akustyki kina Syrena w Wieluniu wykorzystuje 8 mikrofonów i 52 urządzenia głośnikowe.

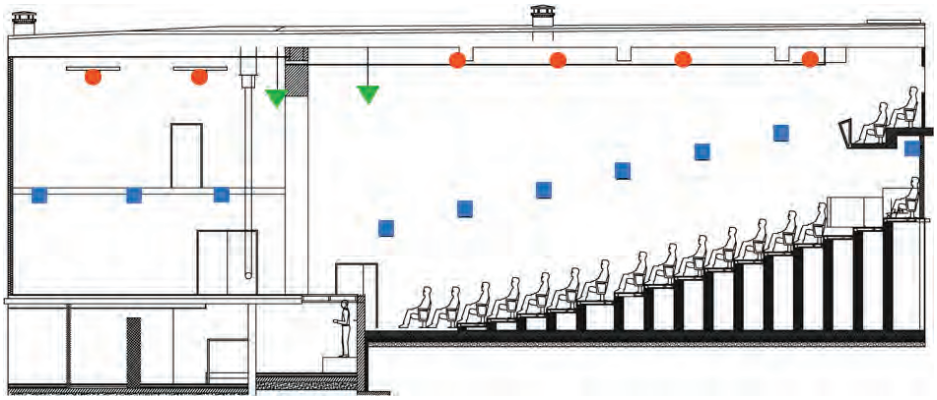
Na podstawie wyników pomiarów przedstawionych na rys. 10 można wnioskować, że system wirtualnej akustyki umożliwia znaczące przestrojenie wartości czasu pogłosu i dostosowanie ich do wymagań poszczególnych typów koncertów nawet wtedy, kiedy naturalna wartość czasu pogłosu dostosowana do wiodącej funkcji kinowej jest prawie czterokrotnie mniejsza od wartości odnoszącej się do występów chóru.



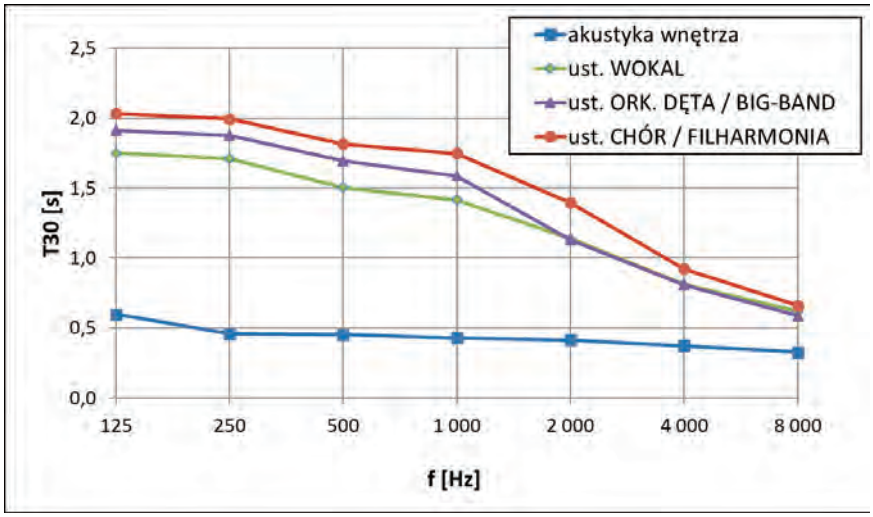
Rys. 7. Rozmieszczenie urządzeń głośnikowych systemu wirtualnej akustyki w kinie Syrena w Wieluniu przedstawione na rzucie parteru: 24 naścienne zestawy głośnikowe (■) rozmieszczone na ścianach widowni i sceny



Rys. 8. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w kinie Syrena w Wieluniu przedstawione na rzucie piętra: 8 mikrofonów (▼) – 4 kardioidalne, 4 dookólne, zwieszonych nad proscenium, 28 głośników sufitowych (●) rozmieszczonych równomiernie nad widownią i sceną



Rys. 9. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w kinie Syrena w Wieluniu przedstawione na przekroju: 8 mikrofonów (▼) – 4 kardioidalne, 4 dookólne, zwieszone nad proscenium, 28 głośników sufitowych (●) rozmieszczonych równomiernie nad widownią i sceną, 24 naścienne zestawy głośnikowe (■) równomiernie rozmieszczone na ścianach widowni i sceny



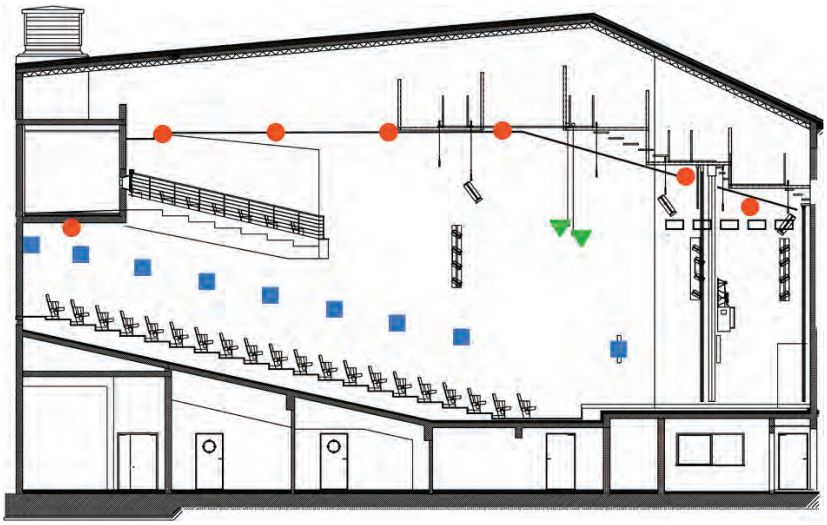

Rys. 10. Wyniki pomiarów czasu pogłosu $T_{30}$ w kinie Syrena w Wieluniu bez publiczności wykonanych w czterech sytuacjach: akustyka wnętrza – naturalna akustyka sali, system przestrajania wyłączony, ust. WOKAL – system przestrajania włączony, program dla zespołów wokalnych z towarzyszeniem zespołu instrumentalnego, ust. ORK. DĘTA / BIG-BAND – system przestrajania włączony, program dla zespołów dętych, big-bandów, ust. CHÓR/FILHARMONIA – system przestrajania włączony, program dla chórów lub orkiestr filharmonicznych z maksymalną ustaloną wartością czasu pogłosu

System wirtualnej akustyki w tym obiekcie ma elementy wspólne z systemem dźwięku dookólnego (procesory, wzmacniacze mocy, urządzenia głośnikowe kanałów *surround*). Ta realizacja świadczy o tym, że zintegrowanie systemu wirtualnej akustyki z systemem dźwięku przestrzennego (*immersive*) jest jak najbardziej możliwe w omawianym przypadku.

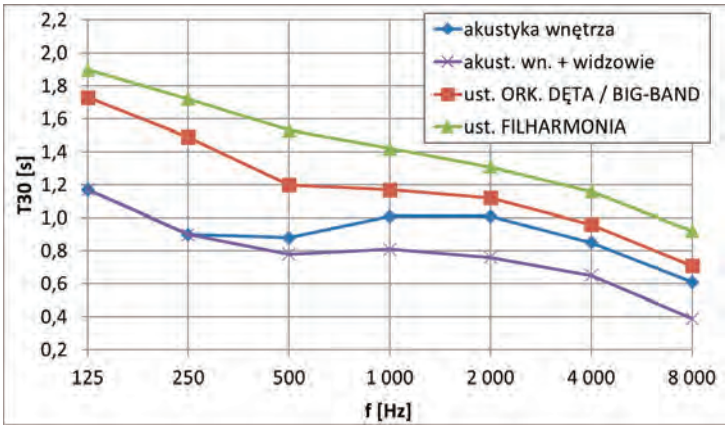
## 1.2.4. System wirtualnej akustyki Miejskiej Szkoły Artystycznej w Mińsku Mazowieckim

Sala koncertowa Miejskiej Szkoły Artystycznej w Mińsku Mazowieckim jest wielofunkcyjną salą koncertową, w której realizowane są zarówno koncerty muzyki wykonywanej akustycznie bez wykorzystania nagłośnienia, jak i imprezy z nagłośnieniem, a także projekcje kinowe. Rozmieszczenie elementów systemu wirtualnej akustyki przedstawiono na rzucie (rys. 11) i przekroju (rys. 12).

Rys. 11. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w sali koncertowej Miejskiej Szkoły Artystycznej w Mińsku Mazowieckim przedstawione na rzucie:
8 mikrofonów (▼) – 4 kardioidalne, 4 dookólne, zwieszonych nad frontem estrady,
30 głośników sufitowych (●) rozmieszczonych równomiernie nad widownią i estradą,
24 naścienne zestawy głośnikowe (■) rozmieszczone na ścianach widowni



Rys. 12. Rozmieszczenie urządzeń głośnikowych i mikrofonów systemu wirtualnej akustyki w sali koncertowej Miejskiej Szkoły Artystycznej w Mińsku Mazowieckim przedstawione na przekroju: 8 mikrofonów (▼) – 4 kardioidalne, 4 dookólne, zwieszonych nad frontem estrady, 30 głośników sufitowych (●) rozmieszczonych równomiernie nad widownią i estradą, 24 naścienne zestawy głośnikowe (■) rozmieszczone na ścianach widowni




Rys. 13. Wyniki pomiarów czasu pogłosu $T_{30}$ w sali koncertowej Miejskiej Szkoły Artystycznej w Mińsku Mazowieckim wykonanych w czterech sytuacjach: akustyka wnętrza – naturalna akustyka sali bez publiczności, system przestrajania wyłączony, akust. wn. + widzowie – naturalna akustyka sali z publicznością, system przestrajania wyłączony, ust. ORK. DĘTA / BIG-BAND – system przestrajania włączony, program dla zespołów dętych, big-bandów, ust. FILHARMONIA – system przestrajania włączony, program dla orkiestr kameralnych i filharmonicznych z maksymalną ustaloną wartością czasu pogłosu

Naturalna akustyka sali koncertowej Miejskiej Szkoły Artystycznej w Mińsku Mazowieckim została zoptymalizowana na potrzeby imprez z wykorzystaniem nagłośnienia. Po analizie wyników pomiarów przedstawionych na rys. 13 odnoszących się do koncertów akustycznych (np. chóru, big-bandu, orkiestry dętej czy orkiestry filharmonicznej) można przyjąć, że wydłużenie czasu pogłosu jest możliwe za pomocą systemu wirtualnej akustyki sumarycznie wykorzystującej 54 urządzenia głośnikowe i 8 mikrofonów.

System wirtualnej akustyki współdzieli część urządzeń (procesor sygnałowy, wzmacniacze mocy, ścienne urządzenia głośnikowe kanałów *surround*) z systemem dźwięku kinowego. Zarówno struktura systemu, jak i rozmieszczenie urządzeń głośnikowych wskazuje, że również głośniki sufitowe mogą być współdzielone z systemem nagłaśniania, co wprost prowadzi do możliwości zrealizowania systemu dźwięku przestrzennego na bazie składników systemu wirtualnej akustyki.

## 1.3. Systemy dźwięku przestrzennego

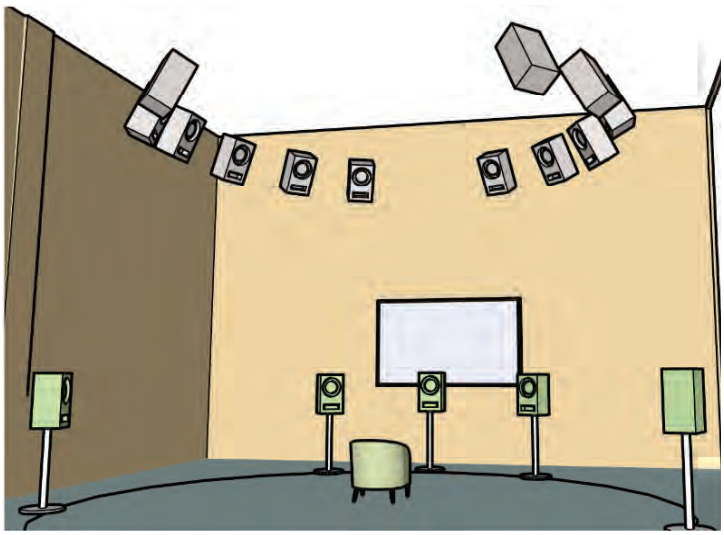
W ciągu ostatnich kilkunastu lat można zaobserwować dynamiczny rozwój systemów dźwięku przestrzennego (*immersive sound*), które w odróżnieniu od klasycznych systemów stereo 2.0 czy *surround* 5.1 umożliwiają budowanie obrazu dźwiękowego nie tylko w płaszczyźnie poziomej, lecz także w płaszczyźnie pionowej, co wynikowo daje możliwość pozycjonowania źródeł pozornych w przestrzeni 3D. Kluczową rolę w tej operacji odgrywają urządzenia głośnikowe znajdujące się w płaszczyźnie powyżej płaszczyzny odsłuchowej zwane górnymi kanałami (*height channels*). W systemach tych wykorzystuje się czasami również urządzenia głośnikowe instalowane na podłodze, czyli poniżej płaszczyzny odsłuchowej, zwane kanałami podłogowymi (*bottom channels*).

Mamy zatem do czynienia z trzema wysokościami/warstwami urządzeń głośnikowych: górną (*height*/***U****pper layer*), środkową na wysokości uszu słuchaczy (***M****iddle layer*) i podłogową (***B****ottom layer*). Konfiguracje systemów określa się przez podanie liczby kanałów/ urządzeń głośnikowych w poszczególnych warstwach w układzie (U + M + B.S), co oznacza U kanałów w górnej warstwie, M kanałów w warstwie środkowej, B kanałów podłogowych oraz S *subwooferów*.

Wielu autorów na podstawie przeprowadzonych badań wskazało, że w kreowaniu poczucia przestrzenności dźwięku efektywniejsze jest rozbudowanie systemu o głośni-

ki w górnej warstwie niż zwiększenie liczby głośników w warstwie środkowej ponad liczbę 5 znaną z klasycznego układu *surronud* 5.1 [6].

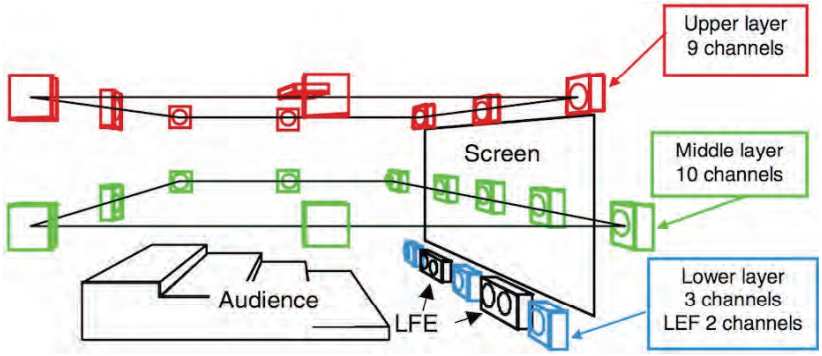
Z badań dotyczących doboru liczby oraz rozmieszczenia głośników w górnej warstwie [7] przeprowadzonych przy wykorzystaniu systemu 17.0 (12 + 5 + 0.0) przedstawionego na rys. 14 wynika, że już przy wykorzystaniu zaledwie czterech urządzeń głośnikowych w górnej warstwie możliwe jest uzyskanie wrażenia dźwięku przestrzennego. Kwestią do ustalenia pozostaje określenie, jakie powinno być położenie tych czterech źródeł w górnej warstwie, ponieważ każda z analizowanych propozycji kąta odchylenia od osi odsłuchu ma swoje wady i zalety.



Rys. 14. Rozmieszczenie urządzeń głośnikowych systemu dźwięku przestrzennego 17.0 przedstawione na widoku aksonometrycznym: 5 zestawów głośnikowych horyzontalnych rozmieszczonych zgodnie ze standardem ITU-R BS.775.2 [5] oraz 12 zestawów głośnikowych sufitowych nachylonych względem płaszczyzny odsłuchowej o 30° umieszonych na lewo i prawo symetrycznie względem osi odsłuchowej pod kątami ±30°, ±50°, ±70°, ±90°, ±110°, ±130° [8]

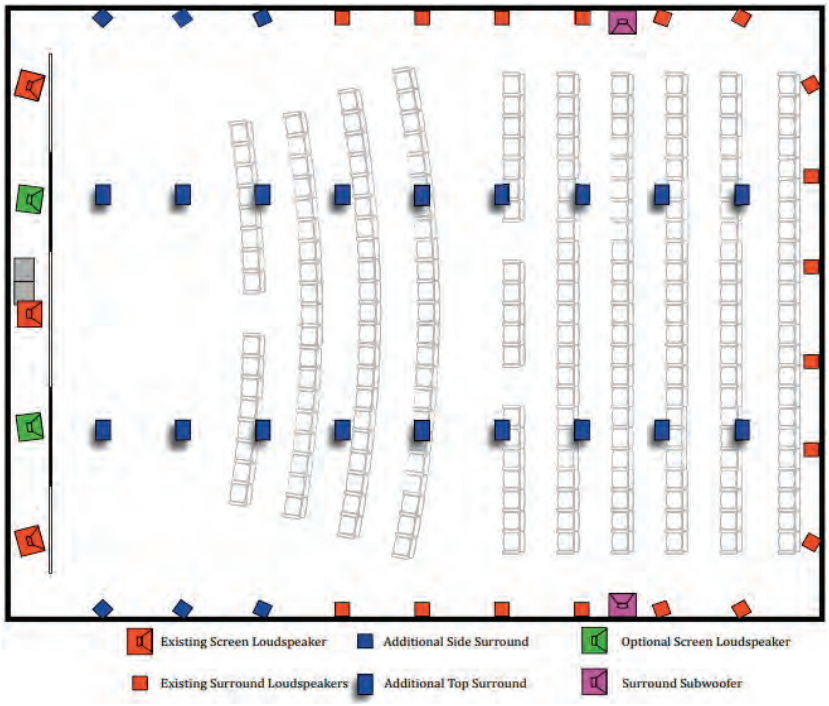
Na potrzeby prezentowania materiału wideo *ultra-high definition* (UHD) o rozdzielczości 4320 px × 7 680 px i kącie obserwacji ponad 100° w płaszczyźnie poziomej poszukiwano systemu, który mógłby zapewnić przestrzenne i naturalne trójwymiarowe pole dźwiękowe. Tak powstał największy ustandaryzowany format oparty na niezależnych kanałach głośnikowych znany jako NHK 22.2 (9 + 10 + 3.2) – jego konfiguracja jest przedstawiona na rys. 15.

Na podstawie opublikowanych badań należy przyjąć, że system zapewnia (zwłaszcza w porównaniu do klasycznych systemów stereo 2.0 czy surround 5.1) możliwość swobodnej lokalizacji źródeł w przestrzeni 3D przy zachowaniu naturalnego brzmienia oraz jednoczesnym wykluczeniu niedoskonałości zgłaszanych podczas badań subiektywnych systemów przestrzennych wykorzystujących mniejsze struktury głośnikowe [4, 6].




Rys. 15. Rozmieszczenie urządzeń głośnikowych systemu dźwięku przestrzennego NHK 22.2 (9 + 10 + 3.2) przedstawione na widoku aksonometrycznym: 9 zestawów głośnikowych górnych (*upper layer*) – 8 nachylonych względem płaszczyzny odsłuchowej o kąt 30–45° umieszonych na lewo i prawo symetrycznie względem osi odsłuchowej pod kątami 0°, ±45–60°, ±90°, ±110–135°, 180° oraz zestaw głośnikowy umieszczony dokładnie nad pozycją odsłuchową, 10 zestawów głośnikowych umieszczonych na wysokości płaszczyzny odsłuchowej (*middle layer*) umieszczonych pod kątami względem osi odsłuchowej 0°, ±22,5–30°, ±45–60°, ±90°, ±110–135°, 180°, 3 zestawy głośnikowe dolne (*bottom layer*) nachylone względem płaszczyzny odsłuchowej o kąt 15–25° umieszone na lewo i prawo symetrycznie względem osi odsłuchowej pod kątami 0°, ±30–45°; 2 *subwoofery* zlokalizowane na podłodze i odchylone od osi odsłuchowej o kąt ±30–90°[4]

Systemem dźwięku przestrzennego komercyjnie wykorzystywanym w przemyśle kinowym jest Dolby Atmos. Rozmieszczenie głośników zostało w tym przypadku rozbudowane względem poprzednio wykorzystywanych układów (np. DTS) o dwa rzędy urządzeń głośników równomiernie rozmieszczonych na suficie oraz o dodatkowe głośniki na ścianach bocznych dochodzące do ekranu (te dodatkowe kanały głośnikowe na rys. 16 przedstawiono w postaci niebieskich prostokątów). Dzięki rozszerzeniu idei systemu opartego na kanałach dźwiękowych o obiekty dźwiękowe uzyskano system hybrydowy zapewniający poprawną lokalizację źródeł dźwiękowych w przestrzeni 3D praktycznie bez większego wpływu położenia słuchacza na widowni.




Rys. 16. Rozmieszczenie urządzeń głośnikowych systemu hybrydowego Dolby Atmos (wykorzystującego równocześnie ideę obiektów dźwiękowych i niezależnych kanałów dźwiękowych); na niebiesko zaznaczono urządzenia głośnikowe dodatkowe względem klasycznego systemu dźwięku kinowego *surround* [3]

## 1.4. Podsumowanie

Na podstawie analizy przedstawionych głównych cech systemów wirtualnej akustyki i podanych przykładów ich realizacji, a także struktur głośnikowych systemów dźwięku przestrzennego (tab. 1) można wnioskować o możliwości spójnego zrealizowania tych systemów w salach wielofunkcyjnych przy wykorzystaniu wspólnych komponentów, takich jak: urządzenia głośnikowe, wzmacniacze mocy, procesory DSP.

Mimo że systemy te służą do realizacji odmiennych celów, duże podobieństwo zarówno ich struktur, jak i zasad rozmieszczenia urządzeń głośnikowych umożliwia jednak wprost ich budowanie przy wykorzystaniu tych samych struktur – wzmacniacze mocy / okablowanie / urządzenia głośnikowe. Również moduły sterowania realizowa-

ne za pomocą DSP można w dużej mierze stosować wspólnie przez systemy wirtualnej akustyki i systemy dźwięku przestrzennego.

Tabela 1. Porównanie cech strukturalnych systemów wirtualnej akustyki i systemów dźwięku przestrzennego

| Cecha strukturalna | System wirtualnej akustyki | System dźwięku przestrzennego |
|---|---|---|
| DSP | √ | √ |
| Liczba urządzeń głośnikowych | ≥50 ... 100 ... 150 | ≥17 ... 22 ... 70 |
| Niezależne kanały wzmacniaczy mocy | √ | √ |
| Głośniki sufitowe | √ | √ |
| Głośniki ścienne | √ | √ |
| Niezależne kanały DSP dla każdego kanału głośnikowego | √ | √ |

Należy zauważyć, że w przypadku systemów dźwięku przestrzennego oczekuje się od urządzeń głośnikowych możliwości uzyskania zdecydowanie większych wartości poziomu ciśnienia akustycznego, niż ma to miejsce dla systemów wirtualnej akustyki. Wymagania względem kątów zasięgu są podobne dla obu systemów. Można zatem przyjąć, że system spełniający wymagania strukturalne stawiane wirtualnej akustyce oraz zapewniający poziomy potrzebne do kreacji dźwięku przestrzennego będzie w stanie w pełni spełniać oczekiwania obu typów analizowanych w niniejszym rozdziale systemów.

Wspólne wykorzystywanie systemu dźwięku przestrzennego wraz z systemem wirtualnej akustyki przynosi również mniej oczywiste korzyści poza tymi czysto oszczędnościowymi wymienionymi już wcześniej. System wirtualnej akustyki może być traktowany z poziomu obsługi systemu nagłaśniania jako bardzo zaawansowany procesor pogłosowy zapewniający możliwość korzystania z naturalnie brzmiącego pogłosu dostosowywanego swoim charakterem do sali widowiskowej, w której jest zainstalowany, lub pogłosu „dowolnego innego wnętrza, który chce wykorzystać realizator dźwięku".

**Słowa kluczowe:** akustyka wnętrza, wirtualna akustyka, dźwięk przestrzenny, system elektroakustyczny, system wielokanałowy, przestrajanie akustyki, zmienna akustyka, system wsparcia akustyki, sala wielofunkcyjna.

# Bibliografia


[1] Barron M., *Auditorium Acoustics and Architectural Design*, Spon Press, London 2010.

[2] Beranek L., *Concert Halls and Opera Houses*, Springer Science + Business Media, New York 2004.

[3] DOLBY ATMOS WHITE PAPER, Dolby 2012.

[4] Hamasaki, K., Hiyama K., OKUMURA R., *The 22.2 Multichannel Sound System and Its Application*, Proc. Audio Engineering Society 118th Int. Conv, AES, preprint 6406, Barcelona 2005.

[5] ITU-R, Recommendation BS.775-2, MultiChannel Stereophonic Sound System with or without Accompanying Picture, Int. Telecommunications Union Radiocommunication Assembly, Geneva 1992–2004.

[6] Kim S., *Height Channels*, [w:] *Immersive Sound – The Art and Science of Binaural and Multi-Channel Audio*, A. Roginska, P. Geluso (red.), Routledge 2017.

[7] Kim S., Indelicato M.J., Imamura H., Miyazaki H., *Height loudspeaker position and its influence on listeners' hedonic responses*, Audio Engineering Society, Conference on Sound Field Control, AES, Guildford, UK, 2016.

[8] Kim S., King R., Kamekawa T., *A CrossCultural Comparison of Salient Perceptual Characteristics of Height Channels for a Virtual Auditory Environment*, „Virtual Reality" 2015, 19(3), s. 149–160.

[9] Kozłowski P.Z., *How to Adjust Room Acoustics to Multifunctional Use at Music Venues*, Proc. of 2018 Joint Conference ACOUSTICS, Polish Acoustical Society, Ustka 2018.

[10] Kozłowski P.Z., *How to Prepare Typical Cinema Theatre to Become Multipurpose Music Venue*, Proc. of 146th Audio Engineering Society Convention, paper 10188, Dublin 2019.

[11] Mapp P., *Audio System Designer. Technical Reference*, Klark-Teknik, Chapman Partnership, Cheltenham 1985.

[12] Mehta M., Johnson J., Rocafort J., *Architectural Acoustics Principles and Design*, Prentice Hall, Hoboken, NJ, 1998.

[13] Miyazaki H. et al., *Active Field Control (AFC) – Reverberation Enhancement System Using Acoustical Feedback Control*, Proc. of AES 115th Convention, paper 5861, New York, NY, 2003.

[14] Miyazaki H., Yamashita S., Shimizu Y., Shiba Y., Tanaka A., *The Acoustical Design of the New Yamaha Hall*, Proc. of the International Symposium on Room Acoustics, ISRA 2010, Melbourne 2010.

[15] Watanabe T., Ikeda M., *Various Applications of Active Field Control*, Proc. of AES 134th Convention, paper 8859, Rome, Italy, 2013.

# 2. Produkcja dźwięku immersyjnego. Praktyczne metody i zastosowania dźwięku ambisonicznego wyższego rzędu do tworzenia produkcji audiowizualnych VR/360°

JAN SKORUPA, MACIEJ GŁOWIAK

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

Systemy wirtualnej i poszerzonej rzeczywistości stają się coraz bardziej popularne. Rosnące wymagania konsumentów odnośnie do treści i jakości treści mają ogromny wpływ na rozwój zarówno w obszarze immersyjnego wideo i przemysłu gier, jak i w dziedzinie systemów dźwięku przestrzennego. Tradycyjne systemy stereo czy *surround* (np. 5.1, 7.1) nie są jednak wystarczające do zaspokojenia potrzeb współczesnego odbiorcy. Aby sprostać wymaganiom, konieczne jest rozszerzenie tradycyjnie pojętej produkcji w obszarze dźwięku przestrzennego. W niniejszym rozdziale podjęto temat produkcji dźwięku przestrzennego opartego na ambisonii wyższego rzędu (ang. High Order Ambisonics), w tym nagrywanie i postprodukcję dźwięku towarzyszącego materiałom wideo o wysokiej rozdzielczości VR/360°. Omówiono również budowę i konfigurację instalacji ambisonicznej z wykorzystaniem doświadczenia dotyczącego 24-kanałowego systemu odsłuchowego zbudowanego 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 przygotował Dział Nowych Mediów PCSS w ramach projektu badawczego Immersify finansowanego z programu Horizon 2020 w latach 2017–2020.

## 2.1. Wstęp

Historia dźwięku przestrzennego jest od samego początku ściśle związana z przemysłem rozrywkowym. Już w latach 40. XX w. można było 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, że kierunek rozwoju tej dziedziny inżynierii dźwięku jest warunkowany przez nowe rozwiązania wizyjne. Obecnie rynek rozrywki napędzany wymaganiami użytkowników szuka nowych rozwiązań – tym samym zauważyć można odchodzenie od tradycyjnego obrazu dwuwymiarowego na korzyść systemów immersyjnych. Obserwowane jest również wkraczanie w erę wirtualnej oraz rozszerzonej rzeczywistości zapewniającej widzom nowe doznania audiowizualne – wymaga to jednak szczególnego i innowacyjnego podejścia do kwestii dźwięku otaczającego.

Powszechnie stosowane systemy *surround*, takie jak: 5.1, 7.1, 7.4.2, tylko nieznacznie różnią się od tych opracowywanych jeszcze w 2. poł. XX w. Przede wszystkim oparte są na głośnikach umieszczonych wokół słuchacza, z których każdy ma specjalny identyfikator i stałe miejsce w przestrzeni (np. *Center*, *Front-Left*, *Front-Right*). Systemy *surround* są znacząco ograniczone – szczególnie w połączeniu z technologią VR – 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 też ponownego zmiksowania niektórych kanałów, co utrudnia adaptację raz przygotowanego dźwięku do różnych rodzajów instalacji dźwiękowych.

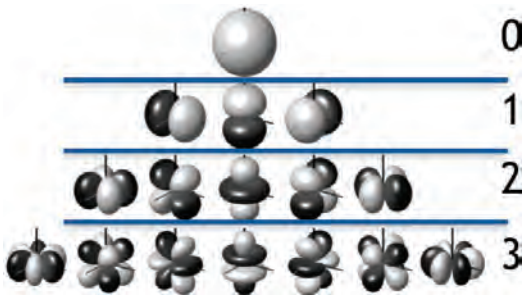
Niedoskonałością takiego podejścia jest także stosunkowo niska rozdzielczość przestrzenna oraz mały obszar tzw. *sweet-spot*, czyli miejsca o najlepszych parametrach odsłuchowych dźwięku wielokanałowego [8]. Istnieje jednak inny sposób podejścia do dźwięku przestrzennego opracowany w latach 70. przez brytyjskiego akustyka i matematyka M. Gerzona – nazwany przez niego ambisonią [12]. Warto wspomnieć w tym miejscu, że proponowane przez niego rozwiązanie miało być alternatywą do wcześniejszego komercyjnego systemu dźwięku przestrzennego, tj. kwadrofonii. Według Gerzona kwadrofonia zapewniała niską rozdzielczość przestrzenną, a zestawu nie można było rozszerzyć o większą liczbę głośników. Największy problem stanowiła jednak „niesta-

bilność" pozycji źródła fantomowego w odniesieniu do pozycji słuchacza, nawet jego mały ruch subiektywnie zaburzał bowiem obraz przestrzenny wytworzony przez system [12].

## 2.2. Ambisonia

### 2.2.1. Podstawowe pojęcia związane z ambisonią

Ambisonia jest koncepcją reprodukcji dźwięku przestrzennego opartą na dekonstrukcji pola akustycznego na podstawie właściwości kierunkowych harmonik sferycznych [5, s. 16–20], w której w przeciwieństwie do systemów wielokanałowych typu channel-based audio poszczególne ścieżki nie przenoszą informacji o sygnale odpowiadającym jednemu tylko głośnikowi, a w zamian reprezentują właściwości całego pola akustycznego. W praktyce jest to możliwe dzięki enkodowaniu oraz dekodowaniu stosowanych formatów dźwiękowych [11, 15]. Docelowo istnieją dwa formaty, dzięki którym dokonuje się zapisu właściwości akustycznych: A-format [10] pozyskiwany bezpośrednio z mikrofonu typu soundfield [23] i B-format [5, s. 17] pozyskiwany przy użyciu enkoderów lub specjalnych ustawień mikrofonów pojemnościowych [24]. Zgodnie z teorią harmonik sferycznych w ambisonii wyróżniamy tzw. rzędy ambisoniczne charakteryzujące wprost rozdzielczość przestrzenną docelowego nagrania. Na przykład pierwszy rząd ambisonii (First Order Ambisonics; FOA) [15] odpowiada nagraniom zrealizowanym za pomocą jednego mikrofonu omnikierunkowego oraz trzech mikrofonów o charakterystyce kierunkowości ósemkowej ustawionych odpowiednio zgodnie z osiami *X*, *Y*, *Z* [25]. Jeśli przełożyć to na B-format, otrzymuje się




Rys. 1. Graficzna reprezentacja pierwszych trzech rzędów harmonik sferycznych

4-kanałowe nagranie, w którym pierwszy komponent definiuje amplitudę nagrania, a pozostałe trzy kanały właściwości kierunkowe. W przypadku ambisonii wyższego rzędu (High Order Ambisonics; HOA) [2, 25] schemat ten zostaje zachowany. Liczba kanałów B-formatu dla wybranego rzędu definiuje się wzorami – dla ambisonii 2D: $2n + 1$, a dla ambisonii 3D (peryfonalnej): $(n + 1)^2$ [12, 4]. W celu zobrazowania zagadnienia graficzną reprezentację pierwszych trzech rzędów harmonik sferycznych przedstawiono na rys. 1.

## 2.2.2. Enkodowanie dźwięku ambisonicznego

Pliki B-Format przechowujące nagrania ambisoniczne różną się od siebie w zależności od kilku parametrów. Pierwszy z nich dotyczy sposobu numeracji poszczególnych komponentów ambisonicznych. W praktyce stosowane są tutaj dwa formaty: FuMa (Furse-Malham) lub ACN (Ambisonics Channel Number) [1, 4] – gdy pominięte zostaną różnice w numeracji kanałów, parametr ten nie wpłynie bezpośrednio na finalną jakość nagrania.

Kolejny parametr dotyczy rodzaju normalizacji poszczególnych kanałów. Wyróżnia się tutaj następujące możliwości:

- **MaxN** – normalizuje każdy pojedynczy komponent, aby nigdy nie przekraczał wzmocnienia 1,0 dla spanoramowanego źródła monofonicznego; parametr używany jest w schemacie FuMa [18].
- **N3D** – zapewnia równą moc kodowanych komponentów w przypadku idealnie rozproszonego pola 3D [4].
- **SN3D** – dzięki niemu żaden komponent nigdy nie przekroczy wartości szczytowej komponentu zerowego rzędu dla źródeł jednopunktowych [4, 18].

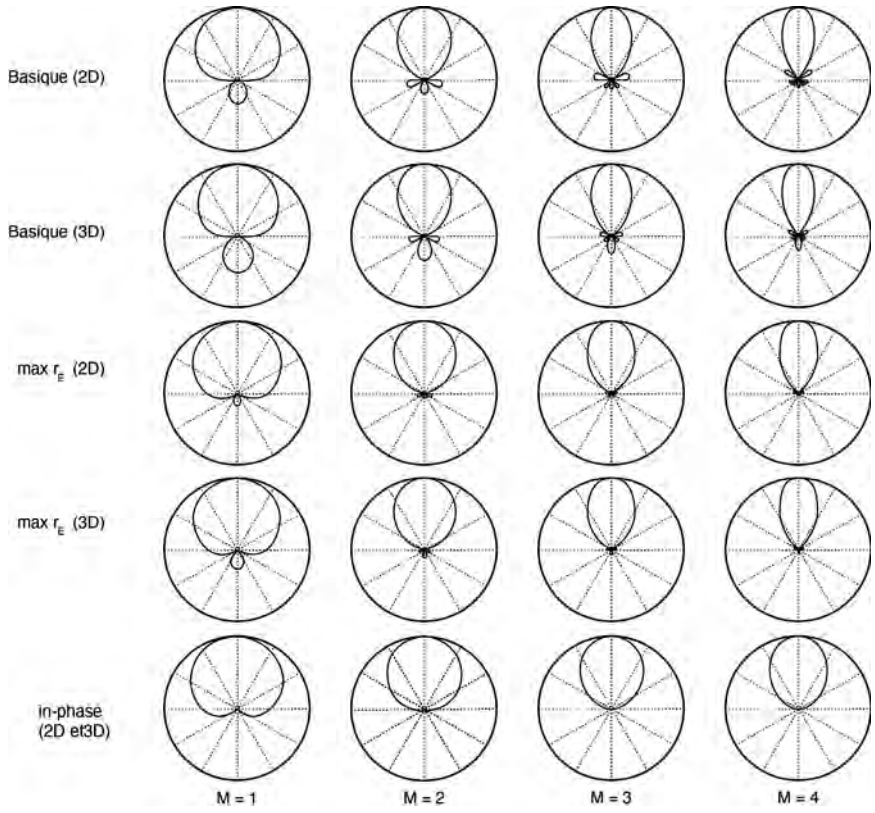
Aktualnie najszerzej stosowanym standardowym formatem ambisonicznym jest Ambix, czyli B-format o numeracji kanałów ACN oraz normalizacji SN3D [11].

## 2.2.3. Dekodowanie dźwięku ambisonicznego

Każdy materiał ambisoniczny zapisany w B-formacie w celu poprawnego odtworzenia musi zostać zdekodowany do docelowej instalacji głośnikowej. W praktyce każde nagranie może zostać zdekodowane do dowolnej instalacji, lecz zgodnie z teorią, aby zachować pełne właściwości enkodowanego pola dla poszczególnych rzędów ambisonii, wymagana jest minimalna liczba głośników równa liczbie komponentów

ambisonicznych B-formatu. Dodatkowymi dwoma parametrami wpływającymi na efekt finalny dekodowania jest Order Weighting oraz docelowa metoda dekodowania.

Order Weighting to parametr o decydującym wpływie na szerokość wstęg głównych oraz promieniowanie boczne modeli harmonik sferycznych reprezentujących dekodowane pole akustyczne. Popularnie stosuje się cztery różne typy poprawek Order Weighting: Basic, Max rE, In-phase, Dual Band [5, s. 20–26; 8], a poszczególne ustawienia mają bezpośredni wpływ na: szerokość obszaru *sweet-spot* czy lepszą rekonstrukcję pola w dziedzinie częstotliwości.




Rys. 2. Charakterystyki kierunkowe dla poszczególnych rodzajów Order Weighting oraz rzędów ambisonii [1]

Poprawka Basic charakteryzuje się dobrą lokalizacją źródła dźwięku w zakresie niskich częstotliwości – poniżej 700 Hz [5, s. 20–26; 17], a Max rE umożliwia lepszą lokalizację dźwięków powyżej 700 Hz oraz ze względu na zmniejszone promieniowanie boczne zapewnia większy obszar *sweet-spot* [5, s. 20–26; 8; 17]. In-phase to natomiast poprawka ade-

kwatna dla systemów ambisonicznych w dużych pomieszczeniach – dzięki niej ze względu na eliminacje wstęg bocznych, a także znaczne poszerzenie wstęgi głównej można wygenerować największy obszar *sweet-spot* – odbywa się to jednak kosztem obniżenia zdolności lokalizacyjnej [5, s. 20–26; 8; 17]. Dodatkowo poprawka ta ujednolica fazę we wszystkich głośnikach i jednocześnie niweluje negatywne interferencje mogące wynikać z dogłośnienia większych przestrzeni ze zbyt dużą pogłosowością.

Dual Band [24] z kolei jest połączeniem tych dwóch opisanych poprawek: Basic – dobrej dla rekonstrukcji częstotliwości niskich, oraz Max rE – dobrej dla rekonstrukcji wyższych składowych. Rodzaj stosowanej poprawki dekodera może znacząco wpłynąć na parametry jakościowe odsłuchiwanego materiału.

Ostatnim elementem wpływającym na jakość dekodowanego materiału jest rodzaj strategii dekodowania (*decoding strategy*). Popularnie stosowane są cztery metody: SAD (Simple Ambisonic Decoder) i MMD (Mode Matching Decoder) wykorzystywane w instalacjach z regularnym ustawieniem głośników oraz EPAD (Energy Presrving Ambisonic Decoder) i AllRAD (All Round Ambisonic Decoder) przeznaczone do nieregularnych instalacji głośnikowych z różnymi odległościami głośników od środka układu [8, 26].

## 2.3. Wpływ parametrów ambisonii na jakość projekcji

W przypadku ambisonii osiągnięcie jak najlepszych efektów rekonstrukcji pola akustycznego związane jest przede wszystkim z parametrami oceny jakości nagrania. Znajomość zarówno podstawowych założeń teoretycznych różnych metod enkodowania i dekodowania, jak i budowy odpowiedniego systemu umożliwia stworzenie dobrego obrazu przestrzennego. W tej dziedzinie szczególnie istotne są trzy parametry [8]:

1. Lokalizacja źródła rozumiana jako wielkość błędu percepcyjnego względem panoramowanego źródła.
2. Stała szerokość panoramowanego źródła bez względu na azymut i elewację.
3. Homogeniczność barwna, czyli niezmienna gęstość widmowa niezależna od pozycji panoramowanego źródła.

Na lokalizacje źródła wpływ mają przede wszystkim: rząd ambisoniczny, pozycja słuchacza względem obszaru *sweet-spot* oraz Order Weighting [3, 8]. Rząd ambisoniczny w tym przypadku ma decydujące znaczenie w osiągnięciu odpowiedniej lokalizacji dźwięku. Warto podkreślić, że jest to ściśle związane z liczbą głośników docelowej insta-

lacji. Zwiększanie rzędu ambisonicznego bez zwiększania liczby głośników nie usprawnia warunków odsłuchowych. Dobre efekty można uzyskać przy ambisonii piątego rzędu, w której błąd lokalizacji horyzontalnie oraz wertykalnie jest bliski 7° [8]. Dodatkowa zmiana parametrów Order Weighting może ten efekt albo poprawić, albo pogorszyć.

Innym istotnym czynnikiem wpływającym na subiektywną percepcję materiału jest także niezmienność szerokości panoramowanego źródła podczas jego ruchu. Szczególnie problematyczne jest tu pokrywanie się ruchu źródła z rzeczywistą pozycją głośników – słuchacz może mieć subiektywne wrażenie poszerzenia się wirtualnego źródła. Jeśli zwiększy się liczbę głośników oraz rząd ambisonii, efekt ten można znacząco wyeliminować [8]. Odnośnie do homogeniczności barwnej najważniejszym elementem jest korekcja rzędu, w której największą neutralność zapewnia poprawka Max-rE [8, 17].

Zaadaptowanie wymienionych elementów do procesu produkcji dźwięku ambisonicznego jest kluczowe w osiągnięciu dobrych parametrów przestrzennych, ponieważ umożliwiają one swobodną pracę postprodukcyjną z podkreśleniem walorów przestrzennych przygotowywanego materiału.

## 2.4. Nagrania ambisoniczne

### 2.4.1. Metody pozyskiwania dźwięku ambisonicznego

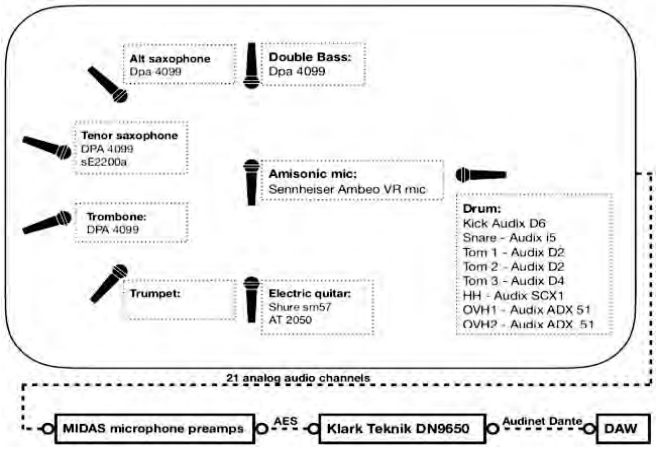
Aktualnie możliwe jest zastosowanie wielu zarówno sprzętowych, jak i programowych rozwiązań w celu uzyskania nagrań ambisoniczych – poczynając od tych klasycznych opracowanych jeszcze przez M. Gerzona. Jest nimi m.in. zastosowanie mikrofonu typu *soundfield* rejestrującego zmiany pola akustycznego za pomocą czterech kardioidalnych kapsuł, w których wyjściowym formatem nagrania był A-format wymagający jednak dalszego dekodowania. Kolejną metodą stosowaną przez Gerzona było wykorzystanie rozszerzonego do nagrań przestrzennych koincydentalnego ustawienia mikrofonów o charakterystyce kierunkowości ósemkowej typu *blumlein* [25]. Obecnie możliwe jest zastosowanie nowszej generacji mikrofonów ambisonicznych również opartych na pionierskich próbach Gerzona rozwijających metody pozyskiwania dźwięku ambisonicznego w dziedzinie rozdzielczości przestrzennej [16]. Możliwe jest także zastosowanie software'owych enkoderów i przeniesienie ścieżek nie tylko monofonicznych, lecz także stereofonicznych do dziedziny B-formatu oraz stworzenie wirtualnej przestrzeni akustycznej. Praktyczne realizacje nagrań ambisonicznych w różnych warunkach i przy zastosowaniu różnych podejść zostaną przedstawione w dalszej części rozdziału.

## 2.4.2. Nagranie septetu jazzowego

Nagrania Septetu Jazzowego zrealizowano w dobrze zaadaptowanym akustycznie studiu Poznańskiego Centrum Superkomputerowo-Sieciowego. Przedmiotem produkcji były dwa utwory zespołu Anomalia, w składzie: kontrabas – Piotr Cienkowski, perkusja – Stanisław Aleksandrowicz, gitara elektryczna – F. Szulgit, saksofon altowy – K. Kuśmierek, saksofon tenorowy – K. Krupa, puzon – A. Kurek, trąbka – P. Rynkiewicz. Głównym celem nagrań, co trzeba zaznaczyć, było przygotowanie materiału immersyjnego na podstawie wideo sferycznego z towarzyszącym mu dźwiękiem ambisonicznym. Oczekiwanym finałem miała być produkcja VR, w której dźwięk poszczególnych instrumentów będzie niejako „podążać" za ruchami głowy widza.

Żeby zachować jak największą realistyczność docelowego materiału zgodnie z założeniem produkcyjnym, zarówno nagrania audio, jak i wideo odbywały się w tym samym czasie – wszyscy muzycy zostali ustawieni w okręgu w równych odstępach od siebie, a sekcja dęta oraz rytmiczna (kontrabas, perkusja) nie zostały rozdzielone. Przyjęto, że takie ustawienie umożliwi wypełnienie całej dookólnej przestrzeni dźwiękowej [22].

Ponieważ ambisonia jest sposobem rekonstrukcji pola akustycznego, założono ponadto, że dobrym rozwiązaniem będzie nieeliminowanie przesłuchów w poszczególnych mikrofonach, a w całości nagrania jednoczesne wykorzystanie mikrofonu ambisonicznego, klasycznych mikrofonów pojemnościowych oraz mikrofonów dynamicznych. Mikrofon ambisoniczny Sennheiser Ambeo [21] znajdujący się w środku okręgu nagrywał ogólny plan dźwiękowy – stanowił podstawę przestrzenną nagrań i definiował ustawienie poszczególnych instrumentów, a także odwzorowywał warunki akustyczne pomieszczenia. Ponieważ nagrania ambisoniczne pierwszego rzędu charakteryzują się dość niską rozdzielczością przestrzenną, zastosowano dodatkowe mikrofony, które na etapie postprodukcji umożliwiły zbudowanie dużo większej przestrzenności, a ponadto uzyskanie dokładniejszych obrazów dźwiękowych poszczególnych instrumentów. Do tego celu wykorzystano mikrofony DPA 4099 [7] w przypadku kontrabasu, puzonu, saksofonu tenorowego i sopranowego. Saksofon tenorowy został dodatkowo zarejestrowany za pomocą mikrofonu SE2200A [20], a kontrabas przez liniowe wyjście ze wzmacniacza instrumentalnego. Do rejestracji perkusji wykorzystano mikrofony firmy Audix. Nagranie zrealizowano przy użyciu przedwzmacniaczy mikrofonowych firmy Midas i z zastosowaniem programu Reaper [6]. Szczegółowy techniczny plan nagrań zespołu Anomalia w studio PCSS przedstawiono na rys. 3.




Rys. 3. Techniczny plan nagrań septetu jazzowego

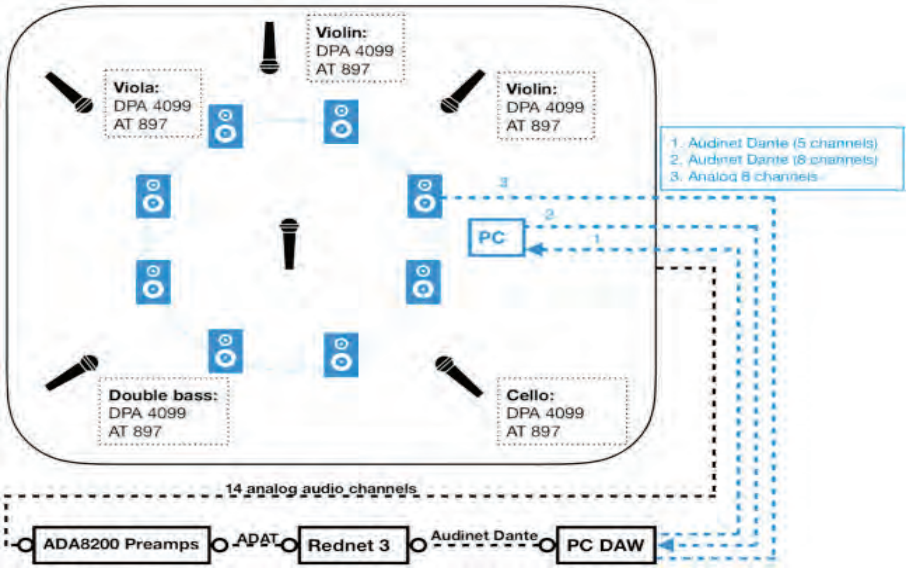
## 2.4.3. Nagranie kwintetu smyczkowego

Nieco inne podejście zastosowano przy realizacji nagrań kwintetu smyczkowego. W tym przypadku założono, że w celu wzbogacenia treści wizualnej nagranie zostanie wykonane poza studiem, w ciekawej scenerii dachu PCSS. Wybór ten podyktowany był zastosowaniem prototypowego systemu wielokamerowego opracowanego w PCSS i służącego do realizacji nagrań sferycznych wysokiej rozdzielczości. Kamera wymagała szerokiego planu – w tej sytuacji panoramy Poznania – jednocześnie możliwe było przetestowanie nagrań ambisonicznych poza zamkniętym studiem.

Przy wcześniejszym nagraniu septetu dźwiękowego zauważono pewną niedoskonałość, a mianowicie brak wypełnienia pełnej przestrzeni dźwiękowej. W materiale brakowało elementów dźwiękowych wypełniających sferę w elewacji powyżej 45°. W przypadku kwintetu smyczkowego ten brak postanowiono uzupełnić tłem akustycznym dachu budynku, na którym słuchać odgłosy miasta. Nietypowy był również sam utwór – kompozytor bowiem w czasie rzeczywistym przetwarzał elektronicznie dźwięk poszczególnych instrumentów, a efekt odtwarzany był za pomocą ośmiu głośników ustawionych oktogonalnie.

Podobnie jak w przypadku nagrań septetu jazzowego wykorzystano zarówno mikrofon ambisoniczny, jak i zestaw mikrofonów pojemnościowych. Tym razem celem była rejestracja trzech różnych obrazów akustycznych, przy czym każdy kolejny był coraz węższy względem pojedynczego instrumentu. Pierwszy stanowił nagranie z mikrofonu ambisonicznego, którym zarejestrowany został ogólny plan dźwiękowy całego

nagrania wraz ze wszystkimi elementami i tłem akustycznym. Drugim – były mikrofony Audio Technica AT897 ustawione w odległości ok. 50 cm od poszczególnych instrumentów, a trzecim – mikrofony instrumentalne DPA 4099 umieszczone zaledwie ok. 5 cm od samego instrumentu. Plan ten zapewniał największą precyzyjność w rejestracji najmniejszych szczegółów dźwiękowych, jak np. szum spowodowany ruchem smyczka po strunie. Oddzielnym torem zarejestrowano bezpośrednio osiem ścieżek wyjściowych z komputera kompozytora wykonującego partię elektroniczną. Nagranie zrealizowano za pomocą oprogramowania Reaper i protokołu Dante przy użyciu Focusrite Rednet połączonego za pomocą protokołu ADAT z czterema przedwzmacniaczami mikrofonowymi Behringer ADA 8200. Techniczny plan nagrań kwintetu smyczkowego przedstawiono na rys. 4.




Rys. 4. Techniczny plan nagrań kwintetu smyczkowego

## 2.5. Postprodukcja dźwięku ambisonicznego

### 2.5.1. Postprodukcja

Postprodukcja dźwięku ambisonicznego opiera się na podobnym schemacie jak w przypadku pracy z materiałem stereofonicznym. Różnice w dziedzinie miksu ambiso-

nicznego wynikają przede wszystkim z przyczyn technicznych wynikających z wielokanałowego B-formatu. Wiąże się z tym nietypowy, dookólny układ głośników w studio odsłuchowym. Poza tym w pracy z ambisonią szczególnie złożone jest zagadnienie kreowania przestrzeni akustycznej. Przede wszystkim, dotyczy to całego procesu panoramowania źródeł dźwięku realizującego się zarówno w przestrzeni dookólnej, jak i za pomocą sztucznych pogłosów czy efektów typu delay, które odgrywają w tym przypadku szczególnie istotną rolę. Budowanie przestrzeni nie ogranicza się jednak wyłącznie do tych dwóch zabiegów. Przy korzystaniu nie tylko z nagrań ambisonicznych, lecz także monofonicznych oprócz etapu enkodowania oraz budowania przestrzeni zabiegi takie jak korekcja czy kompresja są również bardzo użyteczne w poprawie subiektywnego odczucia przestrzeni.

Pierwszym etapem pracy związanej z postprodukcją nagrań ambisonicznych jest zdefiniowanie podstawowych aspektów ambisonii, czyli m.in. wielkości rzędu, sposobu numeracji kanałów czy rodzaju normalizacji poszczególnych komponentów B-formatu. Wszystkie te parametry w każdym stosowanym enkoderze i dekoderze muszą być zgodne. Za ich pomocą definiowana jest rozdzielczość przestrzenna edytowanego materiału. W kolejnym etapie, aby stworzyć jak najlepsze warunki odsłuchowe, należy określić parametry wykorzystywanego dekodera, a następnie zdefiniować parametry Order Weighting oraz wybór strategii dekodowania adekwatnej do systemu odsłuchowego (liczba głośników, ich rozmieszczenie, promień instalacji, warunki akustyczne pomieszczenia odsłuchowego). W opisanych wcześniej nagraniach dźwięku ambisonicznego w PCSS w obydwu przypadkach nagrania zostały przygotowane w ambisonii pierwszego, trzeciego, piątego oraz siódmego rzędu w formacie Ambix. Materiały dekodowano z wykorzystaniem IEM AllRAD Decodera przy poprawce Max-rE oraz strategii AllRAD [26]. Do postprodukcji materiału zastosowano zestaw wtyczek VST IEM plug-in Suite [13] i oprogramowanie Reaper.

## 2.5.2. Panoramowanie i enkodowanie

Docelowy miks materiałów ambisonicznych rozpoczęto od zbudowania podstawowej sceny dźwiękowej opartej na nagraniach z mikrofonu ambisonicznego i wstępnie zsynchronizowano przestrzennie ścieżkę audio z wycinkiem nagrania sferycznego. Przy użyciu kompresji wyrównano dynamicznie cały obraz. Kompresja w domenie ambisonicznej umożliwia korekcję dynamiczną dowolnych wycinków przestrzeni dźwiękowej. Wyrównanie dynamiczne nagrania percepcyjnie wiąże się z utrzymaniem stałej szerokości źródeł, a defekt ten szczególnie zaburza jednolitość obrazu podczas

ewentualnej rotacji nagrań (binauralna projekcja połączona z VR). Na tym etapie nie stosowano jeszcze żadnej korekcji barwowej. Ponieważ nagrania zrealizowane przy użyciu mikrofonu typu *soundfield* charakteryzują się dość niską rozdzielczością przestrzenną, żeby poprawić ten defekt panoramowano poszczególne monofoniczne nagrania instrumentów z zastosowaniem enkoderów i multienkoderów – i dopasowywano je przestrzennie do nagrania ambisonicznego oraz ujęć z kamery sferycznej. Poszczególne monofoniczne nagrania zostały wcześniej poddane wstępnej obróbce w celu odfiltrowania niepożądanych nieczystości dźwięku.

### 2.5.3. Equalizacja

Equalizacja w produkcji ambisonicznej oprócz ujednolicenia miksu czy usunięcia interferujących fal o konkretnych częstotliwościach umożliwia dopasowanie do siebie poszczególnych planów dźwiękowych wynikających z różnych ustawień dodatkowych mikrofonów. Za pomocą korekcji oraz odpowiedniego balansu głośności można zbudować szczegółowy i przestrzenny obraz poszczególnych instrumentów. Na przykład w przypadku altówki za pomocą nagrań realizowanych mikrofonem DPA 4099 z zastosowaniem filtra górnoprzepustowego o nachyleniu 12 dB/oct odfiltrowano częstotliwości poniżej 400 Hz. Częstotliwości poniżej 400 Hz zostały pozyskane z mikrofonu AudioTechnika AT897, dodatkowo w przypadku częstotliwości 3200 Hz zrealizowano korekcje typu *peak* o parametrze *Q* równym 0,7 i wzmocnieniu –7.5 dB. Przy dodatkowym balansie głośności wymienionych dwóch ścieżek pozyskano klarowny i przestrzenny obraz instrumentu. Technika nakładania na siebie trzech planów dźwiękowych, z których każdy był coraz bardziej szczegółowy, umożliwiła również zbudowanie przestrzenności nie tylko w dziedzinie azymutu i elewacji, lecz także subiektywnego poczucia odległości instrumentu. Azymut enkodowanych nagrań z mikrofonów Audiotechnika oraz DPA został ustawiony na takie same wartości. W elewacji poszczególne ścieżki rozsunięto od siebie w przestrzeni o ok. 5°. Zabieg ten subiektywnie rozszerzył szerokość źródła instrumentu. Podobnym przekształceniom poddano nagrania pozostałych instrumentów w obydwu produkcjach. Ze względu na dużą przestrzeń panoramiczną nagrania ambisonicznego w przeciwieństwie do produkcji stereofonicznej dużo mniejszym problemem są interferujące ze sobą fale w poszczególne pasmach częstotliwości. Szerokie rozstawienie w panoramie poszczególnych źródeł dźwięku znacząco ułatwia ten etap miksu. Dodatkowo przy użyciu narzędzia Directivity Shaper [13] w prosty sposób można odseparować panoramowanie poszczególnych pasm źródła. Za jego pomocą możliwa

jest kontrola azymutu i elewacji oraz wielkości rzędu dekodowania dowolnie zdefiniowanego pasma częstotliwości. Dzięki kontroli tych parametrów można zbliżyć miks do naturalnych warunków odsłuchowych, a poszerzenie obrazu niższych częstotliwości oraz zawężenie wyższych percepcyjnie nada przy tym nagraniu większej naturalności.

### 2.5.4. Kompresja

Kompresja w postprodukcji ambisonicznej ma dość istotne znaczenie w ujednoliceniu dynamicznym zbudowanej przestrzeni. W odróżnieniu od narzędzi stosowanych w domenie stereofonicznej kompresja ambisoniczna dodatkowo umożliwia kontrolę dynamiczną pojedynczych wycinków całego obrazu przez nakładanie masek o definiowalnej szerokości. Zabieg ten był szczególnie pomocny w miksie ścieżek zawierających nagrania perkusji oraz instrumentów smyczkowych grających techniką pizzicato.

### 2.5.5. Przestrzenność

Docelowego efektu przestrzennego, oprócz wymienionych etapów postprodukcji, nie udałoby się pozyskać bez dodatkowych pogłosów czy efektów opóźniających. Pogłosowość jest zjawiskiem towarzyszącym praktycznie w każdych warunkach odsłuchowych. Choć nagranie ambisoniczne zostanie przygotowane w sposób bardzo selektywny i poprawny, bez pogłosu będzie ono brzmiało nienaturalnie i sztucznie. Ponadto dzięki zastosowaniu pogłosu można również „zasłonić" pewne niedoskonałości wynikające z wcześniejszego dopasowywania poszczególnych obrazów dźwiękowych pozyskiwanych z różnych mikrofonów. Analogicznie do produkcji stereofonicznych skuteczną metodą okazało się zastosowanie oddzielnych ścieżek efektowych, na które wysyłano poszczególne instrumenty w odpowiednich proporcjach. Efekt opóźniający z bardzo krótkim parametrem czasu umożliwił dość skuteczną imitację wczesnych odbić, a balans między wczesnymi odbiciami a parametrem *fade-in* we wtyczce FDN reverb [13] – osiągnięcie dobrych efektów imitacji przestrzeni akustycznej.

Innym narzędziem do zbudowania naturalnie brzmiącej wirtualnej przestrzeni dźwiękowej był room encoder, który można zastosować w ścieżce efektów zamiast tradycyjnego enkodera B-formatu. Wtyczka ta umożliwia imitację warunków aku-

stycznych pomieszczenia o zdefiniowanej kubaturze przy kontroli liczby wczesnych odbić generowanych oraz ich korekcję barwową. Narzędzie okazało się szczególnie użyteczne przy produkcjach związanych z udźwiękowieniem.

Etap tworzenia wirtualnej przestrzeni ujawnił się jako szczególnie istotny w przypadku postprodukcji nagrań do domeny binauralnej. Medium to ze względu na ograniczenia w dziedzinie dobrych parametrów lokalizacji źródła wymaga szczególnego podejścia w kontekście wykorzystania wczesnych odbić, pogłosu czy room enkoderów. Materiał bezpośrednio przygotowany do odsłuchu wielogłośnikowego wymaga poprawki przy translacji do domeny binauralnej.

## 2.6. Instalacja ambisoniczna PSNC

Ponieważ nie istnieją aktualnie żadne ustandaryzowane zestawy głośnikowe do odtwarzania produkcji ambisonicznych, w Poznańskim Centrum Superkomputerowo-Sieciowym zdecydowano się na budowę własnej instalacji 24.1 (z możliwością powiększenia do 25.2). Dobór i umieszczenie głośników były przedmiotem wcześniejszej analizy – w obecnym kształcie stwarzają możliwość elastycznej regulacji wysokości i średnicy, a także pełną dowolność w umieszczaniu głośników w celu testowania najlepszych konfiguracji sprzętowych. Dodatkowym celem przy tym, co ważne, było zbudowanie przenośnego zestawu, który byłby stosunkowo łatwy w transporcie i instalacji.

Rys. 5. Instalacja ambisoniczna laboratorium Działu Nowych Mediów Poznańskiego Centrum Superkomputerowo-Sieciowego

Omawiane produkcje ambisoniczne opisane w niniejszym rozdziale zostały zrealizowane przy użyciu 24-głośnikowej instalacji ambisonicznej zainstalowanej w laboratorium Działu Nowych Mediów. W konstrukcji wykorzystano monitory studyjne bliskiego pola Genelec 8010A umiejscowione oktogonalnie w trzech pierścieniach o różnych wysokościach. Instalacja ta umożliwia pracę z pełną ambisonią 3D aż do 4. rzędu [22]. Wszystkie głośniki tworzą pełną sferę ze źródłami dźwięku umieszczonymi powyżej i poniżej głowy słuchacza.

Przestrzeń akustyczna opiera się na interfejsie audio Focusrite RedNet 3 połączonym przez protokół ADAT z czterema przedwzmacniaczami Behringer ADA8200. Interfejs jest również wyposażony w protokół Dante podłączonym do karty Focusrite RedNet PCIe za pomocą sieci. Połączenie to zapewnia do 32 fizycznych wyjść audio o częstotliwości próbkowania do 48 kHz.

## 2.7. Podsumowanie

Choć teoretyczne podstawy dźwięku ambisonicznego zostały opracowane pół wieku temu, technologia ta ciągle nie weszła do głównego nurtu zastosowań i rozwiązań proponowanych przez producentów, którzy decydują się raczej na skomplikowane systemy wielokanałowe oferujące, co prawda, równie doskonałą jakość dźwięku, ale wymagające idealnego odtworzenia przewidzianych przez producenta warunków akustycznych związanych zarówno z liczbą, jak i lokalizacją głośników. Dźwięk ambisoniczny nadal traktowany jest bardziej jako ciekawostka lub rozwiązanie niszowe niż technologia możliwa do zastosowania komercyjnego. Obecnie większość praktycznych zastosowań ma związek z aplikacjami VR, w których jednak najczęściej stosuje się zaledwie ambisonię 1. rzędu.

Dzięki pracom prowadzonym w laboratorium Działu Nowych Mediów PCSS oraz projektowi Immersify udało się z jednej strony na przedstawionych wcześniej przykładach pokazać, że stosunkowo łatwo można produkować dźwięk ambisoniczny wyższego rzędu, a z drugiej, że dźwięk ten jest doskonałym uzupełnieniem aplikacji immersyjnych.

Co istotne, PCSS aktywnie włączył się również w promocję rozwiązań związanych z ambisonią. Efekty nagrań ambisonicznych 7. rzędu były prezentowane m.in. w ramach konferencji TNC2019 w Tallinie, NPAPWS19 w Pradze oraz podczas Ars Electronica Festival 2019 [19]. Dodatkowo rozpoczęto prace nad subiektywnymi i obiektywnymi testami jakościowymi nagrań ambisonicznych oraz 24-głośnikowej instalacji zbudowanej w PCSS, w kolejnych pracach implementacyjnych uwaga została skupiona na stru-

mieniowaniu dźwięku w B-Formacie w czasie rzeczywistym i odtwarzaniu go w dowolnej konstelacji głośnikowej.

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) [19].

**Słowa kluczowe:** ambisonia, ambisonia wyższego rzędu, inżynieria dźwięku, produkcja dźwięku ambisonicznego, nagrania ambisoniczne, dźwięk immersyjny.

## Bibliografia

[1] Ambisonics Component Ordering; https://www.audiokinetic.com/library/edge/?source=Help&id=ambisonics_ channel_ordering [dostęp: 12.11.2020].
[2] Arteaga D., *Introduction to Ambisonics*, [w:] *Introduction to higher order Ambisonics (HOA)*, Escola Superior Politècnica University Pompeu Fabra, 2018, s. 19–22.
[3] Bertet S., Daniel J., Parizet E., Warusfer L., *Inverstigation on Localisation Accuracy for First and Higher Order Ambisonics Repduced Sound Sources*, 2013; https://hal.archives-ouvertes.fr/hal-00848764/document [dostęp: 12.11.2020].
[4] Chapman M., Ritsch W., Musil T., Zmölnig D., Pomberger H., Zotter F., Sontacchi A., *A Standard for Interchange of Ambisonic Signal Sets Including a File Standard with Metadata*, Ambisonics Symposium, Graz 2009.
[5] Corcuera A., *A real time encoding tool for Higher Order Ambisonics*, [w:] A. Corcuera, *Ambisonics*, Barcelona, University Pompeu Fabra, 2014,
[6] DAW Reaper; https://www.reaper.fm [dostęp: 12.11.2020].
[7] Dpa 4099; https://www.dpamicrophones.com/DPA/media/DPA-Manual/dvote-4099-Instrument-Microphones -Users-Manual.pdf?ext=.pdf [dostęp 12.11.2020].
[8] Frank M., *How to make ambisonics sound good*, 2014; https://www.researchgate.net/publication/315797065_How_to_make_Ambisonics_sound_good [dostęp: 28.09.2020].
[9] Frank M., Zotter F., Sontacchi A., *Producing 3D Audio in Ambisonics*, AES 57th International Conference, Hollywood 2015.
[10] Gerzon M., *Ambisonics*, cz. 1: *General system description*, „Studio Sound" 1975, Vol. 17, No. 8, s. 20–22; https://www.michaelgerzonphotos.org.uk/articles/Ambisonics%201.pdf [dostęp: 12.11.2020].
[11] Gerzon M., *Ambisonics*, cz. 2: *Studio techniques*, „Studio Sound" 1975, Vol. 17, No. 8, s. 24–26, 28; https://www.michaelgerzonphotos.org.uk/articles/Ambisonics%201.pdf [dostęp: 12.11.2020].
[12] Gerzon M., *What is wrong with quadraphonics*, 1974; https://www.michaelgerzonphotos.org.uk/ambisonics.html [dostęp: 28.09.2020].
[13] IEM Plug-in Suite; https://plugins.iem.at [dostęp: 12.11.2020].
[14] Malhman D., *Higher order Ambisonic system,*[w:] *Space in Music – Music in Space*, University of York, 2003; https://www.yumpu.com/en/document/read/10492846/higher-order-ambisonic-systems-university-of-york [dostęp: 12.11.2020].

[15] Malhm A., David G., Myatt A., *3-D Sound Spatialization Using Ambisonic Techniques*, „Computer Music Journal" 1995, Vol. 19, No. 4, s. 58–70.

[16] Moreau S., Daniel J., Bertet S., *3D Sound field recording with higher order Ambisonics – objective measurements and validation of a 4th order spherical microphone*, AES 120th Convention, Paris 2006.

[17] Mutillo D., Fazi F., Shin M., *Evaluation of ambisonics decoding methods with experimental measurements*, EAA Joint Symposium on Auralization and Ambisonics, Berlin 2014.

[18] Nachbar C., Zotter F., Deleflie E., Sontacchi A., *Ambix – A Suggested Ambisonics Format*, Ambisonics Symposium, Lexington, 2011.

[19] Projekt Immersify; www.immersify.eu [dostęp: 12.11.2020].

[20] sE2200; https://www.seelectronics.com/se2200-microphone [dostęp: 12.11.2020].

[21] Sennheiser AMBEO VR Mic; https://sennheiser.pl/o/ambeo-vr-mic [dostęp: 12.11.2020].

[22] Skorupa J., Głowiak M., *Content production guidelines: Ambisonic recordings and postproduction*, Immersify, 2020.

[23] Thornton S., *Ambisonics: Quadraphonics, as at present widely conceived, is a dead end*; https://www.michaelgerzonphotos.org.uk/ambisonics.html [dostęp 12.11.2020].

[24] Zinemanas P., Rocamora M., Jure L., *Improving Csound's Ambisonics decoders*, ICSC 5th International CSound Conference, Cagli 2019.

[25] Zotter F., Frank M., *Ambisonics A practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality*, Springer, Cham 2019, s. 2–22.

[26] Zotter F., Frank M., Hannes P., *Comparasion of energy-preserving and all-round Ambisonic decoders*, University of Music and Preforming Arts, Graz.

# 3. Ambisoniczna mapa wybranych miejsc w Trójmieście z obrazem 360°

CEZARY PIETRZAK, PIOTR ODYA

Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki,
Katedra Systemów Multimedialnych, ul. Gabriela Narutowicza 11–12, 80-233 Gdańsk

W projekcie, który zostanie opisany w niniejszym rozdziale, założonym celem było stworzenie ambisonicznej mapy Trójmiasta w formie aplikacji internetowej. Materiały wideo w technologii 360° z dźwiękiem w postaci sygnału ambisonicznego zostały zarejestrowane w wybranych lokalizacjach uznanych za charakterystyczne dla tej aglomeracji. Celem badawczym projektu było porównanie dostępnych algorytmów miksowania sygnałów ambisonicznych przez przeprowadzenie testów odsłuchowych. Wykonano test porównań parami, aby uzyskać informacje o preferencjach odnośnie do 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 z analizy wyników obu testów posłużyły jako wskazówki przy postprodukcji nagrań. Otrzymane materiały zostały zamieszczone na interaktywnej mapie w aplikacji internetowej.

## 3.1. Wprowadzenie

Pojęcie systemu ambisonicznego po raz pierwszy pojawiło się w latach 70. XX w. W roku 1974 w magazynie „Studio Sound" ukazał się artykuł, *What's wrong with quadrophonics?*, w którym M. Gerzon wskazał wady popularnego wówczas systemu kwadrofonicznego – wśród nich m.in. problemy z odtwarzaniem lokalizacji poszcze-

gólnych źródeł dźwięku, stosunkowo mała powierzchnia dobrego odsłuchu czy podatność na zmiany tego obszaru, który był silnie uzależniony od pozycji słuchacza. Gerzon przedstawił również alternatywną koncepcję dla kwadrofonii – dziś znaną jako system ambisoniczny. Mimo 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].

Postępujący rozwój technologiczny i powszechna dostępność materiałów audiowizualnych w Internecie sprawiły jednak, że nagrania ambisoniczne znalazły swoje miejsce w wydawałoby się dostatecznie złożonym i różnorodnym globalnym zbiorze multimediów. 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 do wspieranych formatów oraz wymaganych dodatkowych metadanych [3, 16].

## 3.2. Realizacja nagrań

Wybór urządzeń do przeprowadzenia nagrań miał związek z 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 (rys. 1). Oba urządzenia w czasie nagrań mocowano w jednej osi na dwustronnym uchwycie – kamera była skierowana do góry, a mikrofon do podłoża, sam uchwyt zamontowano natomiast na statywie.

W kamerze ustawiono rozdzielczość 5,7 K 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 w pełnym słońcu [9]. W przypadku sygnałów audio formatem wyjściowym był B-format przy częstotliwości próbkowania 96 kHz i rozdzielczości 24 b. Automatyczną konwersję rejestrowanego nagrania do odpowiedniej postaci uwzględniającej orientację mikrofonu osiągnięto dzięki wbudowanemu w mikrofon żyroskopowi. Podczas nagrań korzystano również z możliwości monitorowania sygnału w czasie rzeczywistym przez wyjście słuchawkowe urządzenia, co pozwoliło na kalibrowanie sygnału [21].

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

Materiały zarejestrowano w 12 wybranych miejscach uznawanych 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 16 zestawów nagrań dźwiękowych oraz wideo zarejestrowanych we wszystkich wybranych wcześniej lokalizacjach.

## 3.3. 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 (łącznie cztery kanały) oraz ambisonia pierwszego rzędu z dźwiękiem stereo (łącznie sześć kanałów) charakteryzujący się stałą pozycją ścieżki stereo, bez względu na ruch

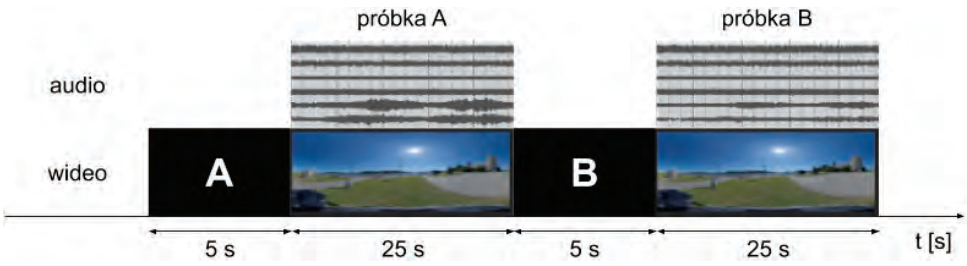
głowy słuchacza w danej chwili. Z uwagi na tę możliwość zdecydowano się na przeprowadzenie testu odsłuchowego mającego na celu sprawdzenie preferencji użytkowników odnośnie do występowania podkładu stereo oraz poziomu jego głośności w materiale 360° z dźwiękiem przestrzennym.

W niektórych sytuacjach nagrania krajobrazu dźwiękowego mogą charakteryzować się stosunkowo niewielką intensywnością dźwięków. Może to spowodować szybkie znużenie potencjalnego słuchacza. Zaproponowanym rozwiązaniem było dodanie do nagrań muzyki o charakterze relaksacyjnym i właściwie dobranym poziomie. Tego typu podkład dźwiękowy nie powinien rozpraszać słuchacza i jednocześnie eliminuje problem braku przykuwających uwagę dźwięków.

Przyjętą formułą testu była metoda porównań parami ze względu na prostotę testu oraz jego przystępność dla potencjalnych uczestników. Zwykle ten typ testu wykorzystywany jest do prostego porównywania dwóch różnych sygnałów dźwiękowych w celu określenia, która próbka jest lepszej jakości. Na potrzeby przygotowanego testu zmieniono jednak to podejście – uczestnikom zadano pytanie o preferowaną wersję spośród prezentowanej pary próbek. Dodatkowo zastosowano również drugą część z zamienioną kolejnością próbek w celu zweryfikowania ocen testujących, założono bowiem, że uczestnikami testu będą osoby niebędące ekspertami. Liczność grupy testowej określono na 20 osób [2].

Hipotezą postawioną w teście było założenie, że potencjalny odbiorca nagrania ambisonicznego zarejestrowanego w przestrzeni publicznej preferuje wersję pozbawioną dodatkowego podkładu muzycznego [16].

Zdecydowano się na przygotowanie sześciu próbek audiowizualnych, z których utworzono osiem par testowych, w tym cztery ze zamienioną kolejnością. Każda para prezentowana była w takim układzie, jak przedstawiony na rys. 2. Przed każdym z sygnałów z danego zestawu zamieszczano fragment ciszy – na planszy pojawiał się wówczas komunikat o następującej po niej próbce.




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

Z zarejestrowanych wcześniej nagrań wybrano losowo dwa miejsca. Dla każdego z nich stworzono po trzy próbki testowe o różnych poziomach podkładu muzycznego oznaczone literami A, B, C oraz numerem miejsca (1 lub 2). Podkład stereo stanowił utwór *The Blue Pearl* autorstwa Jesse'ego Gallaghera zapisany w formacie WAV. Informacje o poszczególnych próbkach oraz dobranych parach przedstawiono w tabelach 1 i 2.

Tabela 1. Próbki 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 | |

Tabela 2. Pary próbek w teście porównań parami

| Numer 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 specjalistycznej platformy przeznaczonej do prowadzania oceny subiektywnej materiałów 360° test został utworzony za pomocą aplikacji Google Forms umożliwiającej stworzenie formularza mogącego zawierać materiały wideo zamieszczone wcześniej w serwisie YouTube. Test podzielono na dziesięć sekcji. Pierwsza z nich, traktowana jako sekcja wstępna, zawierała opis przebiegu testu i dodatkowe informacje dotyczące konieczności odsłuchu w słuchawkach, obracania obrazu 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, a także 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 min. Między obiema częściami nie stosowano przerwy.

W teście udział wzięło łącznie 21 osób w wieku 18–51 lat. Z odpowiedzi udzielonych przez respondentów w ankiecie wiadomo, że jedynie sześciu z nich miało wcześniej doświadczenia w realizacji materiałów audiowizualnych, a w tym tylko jedna osoba przy rejestracji wideo w technologii 360°. Można zatem uznać, że grupy testowej nie stanowili eksperci. 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.

Na podstawie oceny stabilności odpowiedzi każdego z uczestników z dalszej analizy wykluczono wyniki osób, których odpowiedzi były zgodne 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 uwzględnionych w teście par można wyodrębnić dwa ich rodzaje: pary, w których należało dokonać wyboru między próbką bez podkładu muzycznego a próbką z podkładem (1., 2., 4., 6., 7., 8.), oraz pary zawierające obie próbki z podkładem muzycznym, ale o różnym poziomie głośności (3., 5.). W pierwszym z wymienionych rodzajów par uczestnicy wybierali wersję bez podkładu średnio w 14 przypadkach na 21. W takim samym stosunku wybór badanych padał na próbki zawierające cichszą wersję podkładu dla par, w których nie występowała próbka bez podkładu.

Tabela 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 |
| 13/21 | 8/21 | 14/21 | 7/21 | 7/21 | 14/21 | 7/21 | 14/21 | 13/21 | 8/21 | 16/21 | 5/21 | 7/21 | 14/21 | 9/21 | 12/21 |

Na podstawie otrzymanych danych wyznaczono statystykę chi-kwadrat umożliwiającą porównanie wyników obu części testu. Dzięki niej można określić, czy istnieje istotna różnica między odpowiedziami uczestników w obu częściach testu. Wartość chi-kwadrat obliczono zgodnie ze wzorem [10]:

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

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,

$n_j$ – suma wyborów $j$-tego obiektu we wszystkich częściach testu,

$n$ – suma wszystkich wyborów podczas testu.

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

Statystyczną zgodność wyników określono przez porównanie obliczonej wartości chi-kwadrat z wartością krytyczną, która wyniosła 11,0705. Odczytano ją z tablic rozkładu tej statystyki dla wyznaczonej liczby stopni swobody oraz obranego poziomu istotności równego 0,05. Obliczona wartość statystyki chi-kwadrat nie przekracza wartości krytycznej, można zatem potwierdzić hipotezę o zgodności wyników. Następnie określono istotności różnic w wynikach dla poszczególnych par w obu turach na podstawie wzoru [10]:

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

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 $\alpha = 0{,}05$ 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 „–”). 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].

Tabela 4. Istotności różnic między 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 stanowi potwierdzenie założonej wcześniej hipotezy – czyli można przyjąć, że docelowy odbiorca nagrania ambisonicznego zrealizowanego w przestrzeni publicznej będzie preferował wersję „naturalną" bez podkładu muzycznego.

## 3.4. Test parametryczny

Dostępnych jest wiele wtyczek przeznaczonych do postprodukcji sygnałów wielokanałowych zarówno kompatybilnych z wieloma formatami wielokanałowymi, jak i przeznaczonych wyłącznie do nagrań ambisonicznych. Dlatego zdecydowano się na przeprowadzenie testu parametrycznego, dzięki któremu pozyskane zostaną wytyczne dotyczące wartości parametrów wybranych wtyczek oraz ich potencjalnej przydatności podczas obróbki materiału w procesie tworzenia docelowej bazy nagrań. Chodziło o przetestowanie z udziałem słuchaczy wpływu różnych ustawień na odbiór materiału.

Podczas wyboru testowanych wtyczek kierowano się charakterem zarejestrowanych nagrań, a także potencjalną przydatnością wtyczek podczas procesu miksowania. Z tego powodu spośród grupy testowanych wtyczek wykluczono takie procesory efektowe, jak pogłos czy echo. W przypadku nagrań realizowanych w przestrzeni publicznej, które zarejestrowane zostały bezpośrednio w formacie ambisonicznym, dodatkowy pogłos lub echo są efektami niepożądanymi [1].

Ponieważ 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 i kompresora wielopasmowego w zakresie niskich częstotliwości. Celem było sprawdzenie wpływu zminimalizowania składowych odpowiadających powiewom wiatru oraz usunięcia niepożądanych składowych o wysokich częstotliwościach na odczuwaną jakość

nagrań. Przetestowano również działanie obu tych wtyczek przy wzmacnianiu charakterystycznych dźwięków otoczenia znajdujących się w zakresie średnich częstotliwości. W ramach eksperymentu postanowiono także sprawdzić, jak na percypowaną jakość nagrania wpłynie wyrównanie poziomów dla kanałów w sygnale ambisonicznym. Dlatego przetestowano dwie wtyczki wyrównujące wzmocnienia kanałów ortogonalnych względem kanału dookólnego. Ponadto, ponieważ nagrania realizowano również w miejscach, w których znajdował się dźwięk charakterystyczny dla danego otoczenia, przetestowano wpływ działania wtyczek wzmacniających dany kierunek lub przestrzeń w nagraniu.

Analogicznie do przypadku testu porównań parami z uwagi na docelowe przeznaczenie materiałów założono, że w teście będą uczestniczyć osoby niebędące ekspertami. Minimalną liczebność grupy testowej określono na 35 osób. Formuła testu została zdefiniowana zgodnie z wymaganiami stosowanymi w podejściu typu MUSHRA [14]. Na ich podstawie przyjęto długość próbek testowych równą 10 s. Zadaniem uczestników testu była subiektywna ocena jakości każdej z prezentowanych próbek w skali 1–10 (przyjęto 10 za najwyższą możliwą ocenę – najwyższą jakość).

Test podzielono na 10 sekcji zawierających próbki przetworzone za pomocą wybranej wtyczki i danego parametru. Zestawienie parametrów ocenianych dla poszczególnych wtyczek oraz przypisanych im sekcji przedstawiono w tab. 5. Sekcja 1. i sekcje 3.–10. zawierały po cztery próbki testowe, a sekcja 2. – pięć, czyli łącznie to 41 próbek testowych. Mając na uwadze wyniki testu porównań parami, zrezygnowano z dodatkowego podkładu muzycznego. Funkcję „kotwicy" w każdym z prezentowanych zestawów próbek pełnił sygnał niepoddany działaniu wtyczki w badanym aspekcie.

Próbki testowe utworzono z nagrań zrealizowanych w następujących miejscach:

- Gdynia – Skwer Kościuszki: sekcje 1.–5.,
- Gdańsk – Długi Targ: sekcje 6., 7.,
- Gdańsk – Park Oliwski (Potok Oliwski): sekcje 8.–10.

Do stworzenia formularza testowego ponownie posłużyła platforma Google Forms. 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 osobne okno odtwarzacza YouTube i pytanie o jej ocenę. Każda sekcja zawierała jednak informacje o możliwości wielokrotnego odsłuchiwania próbek w dowolnej kolejności, żeby zachować zasady przebiegu testu zgodnie z normą ITU-R BS.1534. Na końcu każdej sekcji dodano również pole tekstowe prze-

Tabela 5. Właściwości wybranych wtyczek zbadane w ramach testu parametrycznego

| Numer sekcji | Nazwa wtyczki | Rodzaj wtyczki | Badany aspekt | Testowany parametr |
|---|---|---|---|---|
| 1. | MultiEQ (IEM Plug-in Suite) | korektor wielopasmowy | korekcja barwy w zakresie niskich częstotliwości | częstotliwość odcięcia filtru górnoprzepustowego: 60 Hz, 120 Hz, 180 Hz |
| 2. | | | korekcja barwy w zakresie wysokich częstotliwości | poziom tłumienia górnozakresowego korektora półkowego (–6 dB, –12 dB) lub filtru pasmowo-zaporowego ustawionego w zakresie wysokich częstotliwości (–3 dB, –6 dB) |
| 3. | | | wzmocnienie w zakresie środkowych częstotliwości | poziom wzmocnienia filtru pasmowo-zaporowego ustawionego w środkowym zakresie: 2 dB, 4 dB, 6 dB |
| 4. | MultiBand Compressor (IEM Plug-in Suite) | procesor dynamiki | kompresja w zakresie niskich częstotliwości | stopień kompresji – 2 : 1, 4 : 1, 8 : 1 |
| 5. | | | kompresja w zakresie środkowych częstotliwości | stopień kompresji – 2 : 1, 4 : 1, 8 : 1 |
| 6. | OmniCompressor (IEM Plug-in Suite) | | wyrównanie wzmocnienia dla kanałów ortogonalnych względem kanału dookólnego | stopień kompresji – 2 : 1, 4 : 1, 8 : 1 |
| 7. | FB360 Spatializer (Facebook Spatial Workstation) | | | parametr *Envelopment* – 0,10, 0,27, 0,42 |
| 8. | Directional Compressor (IEM Plug-in Suite) | wtyczki uwydatniające dane kierunki lub przestrzenie w obrazie dźwiękowym | uwydatnienie jednego kierunku lub przestrzeni w nagraniu względem innych | stopień kompresji – 2 : 1, 4 : 1, 8 : 1 |
| 9. | Transform Dominance (ATK for Reaper) | | | wzmocnienie – 2 dB, 4 dB, 6 dB |
| 10. | Transform FocusPressPushZoom (ATK for Reaper) | | | stopień przekształcenia– 20, 30, 40 |

znaczone do zgłaszania uwag dotyczących danej części testu. Podobnie jak w formularzu przygotowanym dla testu porównań parami na jego początku zamieszczono sekcję wstępną zawierającą informacje dotyczące przebiegu testu oraz dodatkowe informacje dotyczące konieczności odsłuchu próbek przy użyciu słuchawek oraz ograniczeń platformy Google Forms w kontekście odtwarzania materiałów testowych [14].

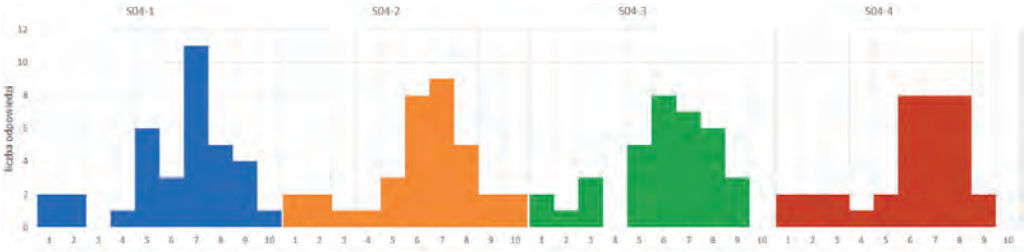
W teście udział wzięło łącznie 35 osób – spełniono zatem założenie dotyczące liczebności grupy testowej. Pozostałe warunki zostały podzielone na kilka grup. W początkowym etapie przy użyciu miar statystycznych i na podstawie przyjętych wobec nich założeń wyodrębniono grupę ekspertów. Następnie przeprowadzono test Kruskala–Wallisa z uwzględnieniem wyników ekspertów. Przyjęto w nim domyślną hipotezę zerową, w której zakłada się równość między średnimi rangami lub medianami dwóch serii danych. Ostatnim etapem przed sformułowaniem wniosków była analiza odpowiedzi grupy eksperckiej przy użyciu skryptu w języku R.

Zgodnie z przedstawionym opisem najpierw wyłoniono spośród wszystkich uczestników testu grupę ekspertów. Dokonano tego na podstawie wariancji obliczonej dla każdego uczestnika w kontekście całego testu. Następnie 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.

Mimo przyjętej metody wyłaniania ekspertów obliczono również wariancje odpowiedzi każdego z uczestników w pojedynczych sekcjach testowych. Uzyskane w ten sposób statystyki jednoznacznie świadczyły o właściwym sposobie doboru ekspertów, ponieważ wariancje ocen osób wykluczonych z dalszej analizy osiągały wartości bliskie 0, co oznacza brak zmian zauważanych przez tych uczestników między próbkami w poszczególnych sekcjach testu.

W kolejnym etapie dla otrzymanej grupy ekspertów wygenerowano histogramy rozkładu ocen poszczególnych próbek testowych oraz obliczono dla nich statystyki opisowe. Na rysunku 3 przedstawiono przykładowe histogramy otrzymane dla sekcji testowej nr 4. 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 4.–6. i 9.: w granicach 0,7–0,9, co świadczy o symetryczności rozkładu ocen względem jego środka [8].

W analizie poziomów wariancji – będącej miarą rozproszenia wyników względem wartości średniej, najwyższe wartości zaobserwowano w przypadku tych 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.




Rys. 3. Histogramy rozkładu odpowiedzi otrzymane dla 4. sekcji testowej

Po wyodrębnieniu grupy ekspertów dla ocen próbek testowych znajdujących się w pierwszych siedmiu sekcjach testowych wartość wariancji wzrosła. W tych sekcjach jednak, w których badano wtyczki uwydatniające wybrane kierunki w nagraniu, miara rozproszenia nieznacznie spadła w ośmiu przypadkach na dwanaście badanych. Takie zależności między wartościami statystyki wariancji otrzymanymi dla obu grup stanowią potwierdzenie prawidłowego doboru ekspertów [8].

W następnym etapie przeprowadzono test Kruskala–Wallisa z uwzględnieniem wyników ekspertów – wynikało to z występowania miar porządkowych oraz z tego, że otrzymane rozkłady odpowiedzi nie były rozkładami normalnymi. W hipotezie zerowej w tym teście przyjęto założenie, że średnie rangi lub mediany dwóch serii danych są takie same. Przeciwny przypadek stanowił hipotezę alternatywną. Hipotezą do przeprowadzanego testu była domyślna hipoteza zerowa. W pierwszej kolejności obliczono wartość statystyki $T$ dla całego zbioru odpowiedzi według następującego wzoru [15]:

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

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 przez 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 wynoszący 55,7585. Oznacza to, że 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 (tab. 6). 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 (tab. 7) 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.

Tabela 6. Podział sekcji na grupy testowe

| Numer grupy | Sekcje | Cel przetwarzania |
|---|---|---|
| 1. | 1., 2., 4. | zminimalizowanie niepożądanych składowych w zakresie niskich i wysokich częstotliwości |
| 2. | 3., 5. | wzmocnienie sygnału w zakresie średnich częstotliwości |
| 3. | 6., 7. | wyrównanie poziomów składowych sygnału ambisonicznego |
| 4. | 8.–10. | uwydatnienie danej przestrzeni lub kierunku w nagraniu |

Tabela 7. Zestawienie statystyk dla poszczególnych grup testowych

| Numer grupy | Statystyka T | Liczba stopni swobody | Wartość krytyczna chi-kwadrat |
|---|---|---|---|
| 1. | 9,556 | 12 | 21,0261 |
| 2. | 5,471 | 7 | 14,0671 |
| 3. | 5,416 | 7 | 14,0671 |
| 4. | 7,644 | 11 | 19,6752 |

Kolejne kroki analizy przeprowadzono przy użyciu skryptu w języku R oraz pakietu fBasics zawierającego zbiór funkcji do analizy danych – w tym testu Kruskala–Wallisa [20]. Serie odpowiedzi wyłonionych ekspertów stanowiące dane wejściowe zapisano w pliku CSV. Po podaniu kolejnych wektorów odpowiedzi do funkcji w kombinacji „każdy z każdym" otrzymano macierz statystyki chi-kwadrat oraz macierz *p*-wartości dla całego testu.

Następnie z otrzymanych macierzy wyodrębniono zbiory wartości statystyk dla każdej z grup. Uzyskane wartości statystyki chi-kwadrat dla wszystkich badanych próbek w kolejnych grupach są mniejsze od wyznaczonej dla każdej z nich wartości krytycznej – świadczy to o istotności różnic między udzielanymi odpowiedziami. Jeśli przeanalizuje się 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, tym razem jednak pod uwagę wzięto jedynie oceny powyżej średniej arytmetycznej dla danego obiektu testowego ocenianego przez grupę wyznaczonych ekspertów. Na podstawie wyników dla każdej sekcji wyznaczono najwyżej oceniane próbki oraz odpowiadające im wartości testowanych parametrów (tab. 8).

Tabela 8. Preferowane próbki oraz przypisane do nich wartości testowanych parametrów

| Numer sekcji | Numer próbki | Testowany parametr | Wartość parametru |
|---|---|---|---|
| 1. | 1. lub 3. | częstotliwość odcięcia filtru górnoprzepustowego | bez lub 120 Hz |
| 2. | 3. | poziom tłumienia górnozakresowego korektora półkowego lub filtru pasmowo-zaporowego ustawionego w górnym paśmie | korektor półkowy: –12 dB |
| 3. | 1. lub 4. | poziom wzmocnienia filtru pasmowo-zaporowego ustawionego w środkowym paśmie | bez lub 4 dB |
| 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 | 2 dB |
| 10. | 1. | stopień przekształcenia | bez |

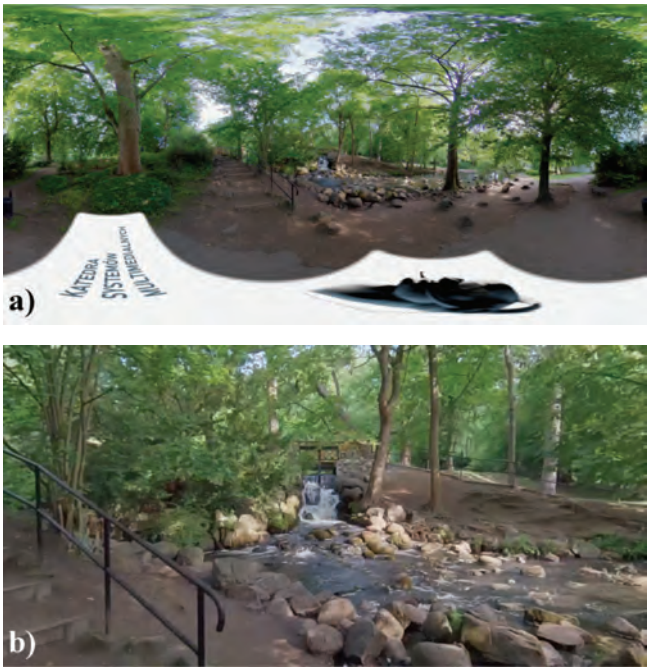
Wyniki analizy przeprowadzonego testu są w wielu aspektach wyraźnie niejednoznaczne. Przyczyn tego należałoby szukać w kilku czynnikach. Pierwszym z nich jest niewątpliwie nieodpowiedni dobór uczestników, o czym świadczą m.in. wartości wariancji uzyskane przez uczestników w poszczególnych sekcjach. W przypadku przeprowadzania testu, w którym oceniane próbki przetworzone zostały w stopniu

średnim lub znacznym, rekomenduje się dobranie grupy testowej składającej się z tzw. ekspertów. Z uwagi jednak na ograniczenia występujące podczas realizacji omawianych eksperymentów (pandemia COVID-19) możliwe było zorganizowanie testu wyłącznie w formie zdalnej, co powodowało dodatkowe komplikacje w postaci braku kontroli nad przebiegiem sesji odsłuchowych i dysproporcji w odbiorze próbek testowych ze względu na zróżnicowany sprzęt, z jakiego korzystali uczestnicy. W związku z tym wnioski otrzymane podczas przeprowadzonej analizy wyników należy traktować raczej jako punkt wyjścia do dalszych badań związanych z przetwarzaniem nagrań ambisonicznych niż jako konkretne wskazówki dotyczące tego zagadnienia [14].

## 3.5. Stworzenie bazy nagrań oraz aplikacji

W każdym miejscu, w którym zarejestrowano materiały, otrzymano zestaw dwóch nagrań – wideo i 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, przeznaczona do użycia wyłącznie we współpracy z oprogramowaniem Adobe Premiere Pro (począwszy od wersji 2018). Dzięki niej możliwa jest obsługa oraz edycja plików w formacie INSV. Ze względów technicznych 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. Znajdujące się w kadrze statyw i mikrofon natomiast zasłonięto maską z logiem Katedry Systemów Multimedialnych nałożoną na obraz. Materiały wideo wyeksportowano bez ścieżki dźwiękowej przy użyciu kodeka h.264 z przepływnością 12 Mbit/s. Pliki zapisano w formie odwzorowania walcowego równoodległościowego (wymaganego przez platformę YouTube). Na rysunku 4 przedstawiono porównanie kadrów przykładowego materiału wideo z odwzorowaniem walcowym równoodległościowym oraz w postaci prezentowanej przez odtwarzacze wspierające wideo sferyczne.




Rys. 4. Porównanie sposobów prezentacji sferycznych materiałów wideo:
a) odwzorowanie walcowe równoodległościowe,
b) sposób prezentacji w odtwarzaczu wspierającym wideo sferyczne

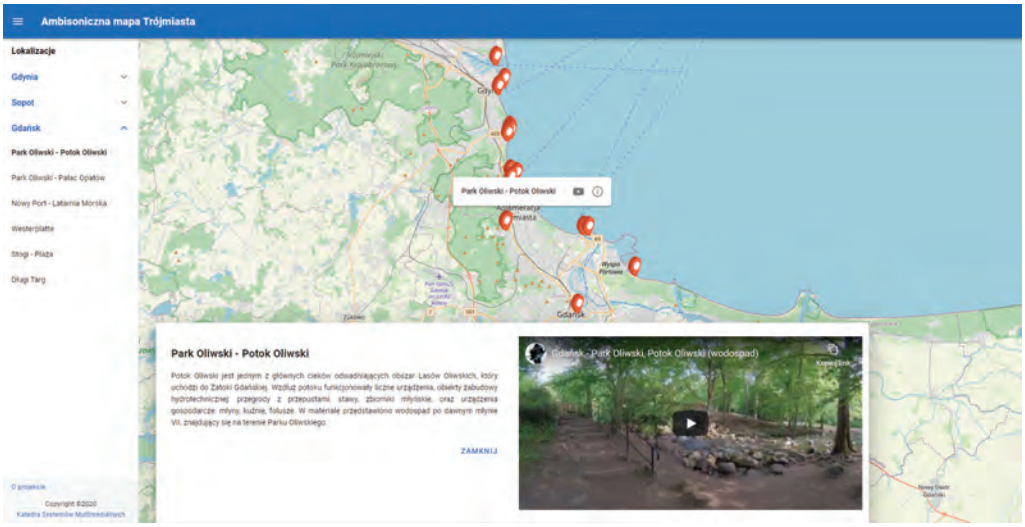
Do postprodukcji nagrań audio wybrano cyfrową stację roboczą Reaper w wersji 6.02 umożliwiającej miksowanie oraz eksportowanie sygnałów wielokanałowych [13]. W każdym projekcie, w którym przeprowadzono obróbkę dźwięku, ustawione zostały parametry tożsame z parametrami zarejestrowanych sygnałów, czyli częstotliwość próbkowania – 96 kHz i rozdzielczość bitową – 24 b.

Po uwzględnieniu wyników testu porównań parami zrezygnowano z dodatkowego podkładu stereo, routing kanałów został więc ustawiony tak, aby zarówno dla kanału głównego, jak i pozostałych sygnał wyjściowy był w postaci czterokanałowej. Z uwagi na to, że nagrania zostały zarejestrowane w ambisonicznym B-formacie, nie było konieczne użycie enkodera. Monitorowanie sygnału przez 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 barwy dla niskich i wysokich częstotliwości – sugerowano się przy tym wariantami najlepiej ocenianymi przez ekspertów. Pliki dźwiękowe zostały wyeksportowane w formacie WAV dla ustawień projektu.

Przygotowane w ten sposób pliki audio oraz wideo połączono i zapisano w kontenerze MOV za pomocą kodeka FFMPEG. Następnie przy użyciu narzędzia Spatial Media Metadata Injector dodano również wymagane metadane zgodnie z wymaganiami dotyczącymi 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 i VueLayers [17–19]. Ostateczny wygląd aplikacji z otwartym filmem pokazującym okolice wodospadu w Parku Oliwskim, przedstawiono na rys. 5. Aplikacja została umieszczona na serwerze Katedry Systemów Multimedialnych pod adresem: https://multimed.org:8100/. Link do mapy znajduje się także na stronie głównej Katedry: https://www.multimed.org/.




Rys. 5. Aplikacja internetowa z interaktywną mapą dostępna pod adresem: https://multimed.org:8100/

## 3.6. Podsumowanie

W opisanym projekcie zrealizowano serię materiałów audiowizualnych w wybranych lokalizacjach 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 Trój-

mieście. Można wziąć również pod uwagę realizację nagrań o różnych porach dnia czy w różnych porach roku.

Na podstawie przeprowadzonego testu porównań parami można wnioskować, że docelowi odbiorcy materiałów zdecydowanie preferują nagrania pozbawione dodatkowego podkładu muzycznego. W drugim teście sprawdzano natomiast, jak sposób oraz stopień przetwarzania nagrania ambisonicznego wpływa na jego odbiór. Niestety wnioski pochodzące z analizy wyników tego testu 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. Uzyskane wyniki testów stanowią punkt wyjścia do dalszych badań dotyczących sposobu przygotowywania nagrań z wideo zapisanym w technice 360° i dźwiękiem ambisonicznym.

**Słowa kluczowe:** ambisonia, wideo 360°, przetwarzanie dźwięku, rejestracja dźwięku i obrazu, *soundscape*.

## Bibliografia

[1] 5 Tips for Mixing Ambisonics | 360° | VR | Spatial Audio | Part 5/7 | Berklee Online; https://www.youtube.com/watch?v=1dvK4ojh2b0 [dostęp: 5.10.2020].
[2] Anstendig M.B., *AB Testing. A misapplication of visual criteria in audio*, 2006; http://www.anstendig.org/ABTesting.html [dostęp: 5.10.2020].
[3] Facebook 360 Video; https://facebook360.fb.com/ [dostęp: 5.10.2020].
[4] FFmpeg Codecs Documentation; https://www.ffmpeg.org/ffmpeg-codecs.html [dostęp: 5.10.2020].
[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, [dostęp: 5.10.2020].
[7] google/spatial-media: Specifications and tools for 360º video and spatial audio; https://github.com/google/spatial-media [dostęp: 5.10.2020].
[8] Heumann C., Schomaker M., Shalabh, *Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R*, Springer International Publishing, Cham, Switzerland, 2016.
[9] Insta360 ONE X – Own the moment; https://www.insta360.com/product/insta360-onex [dostęp: 5.10.2020].
[10] Kostek B., [wykład z przedmiotu: technologia nagrań – testy subiektywne, Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Katedra Systemów Multimedialnych; https://multimed.org/student/tn/testy_subiektywne.pdf [dostęp: 5.10.2020].
[11] Michael Gerzon Audio Pioneer – Ambisonics; https://www.michaelgerzonphotos.org.uk/ambisonics.html [dostęp: 5.10.2020].
[12] Profesjonalny edytor wideo | Adobe Premiere Pro; https://www.adobe.com/pl/products/premiere.html [dostęp: 5.10.2020].

[13] REAPER | Audio Production Without Limits; https://www.reaper.fm/ [dostęp: 5.10.2020].
[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 [dostęp: 5.10.2020].
[15] Sprent P., Smeeton N.C., *Applied Nonparametric Statistical Methods, Third Edition*, Champan & Hall/CRC, Boca Raton, Florida, USA, 2001.
[16] *Używanie dźwięku przestrzennego w filmach 360° i rzeczywistości wirtualnej (VR)*; https://support.google.com/youtube/answer/6395969 [dostęp: 5.10.2020].
[17] Vue.js; https://vuejs.org/ [dostęp: 5.10.2020].
[18] Vuetify; https://vuetifyjs.com/en/ [dostęp: 5.10.2020].
[19] VueLayers: Homepage; https://vuelayers.github.io/ [dostęp: 5.10.2020].
[20] Wuertz D., Setz T., Chalabi Y., Maechler M., *fBasics: Rmetrics – Markets and Basic Statistics*; https://cran.r-project.org/package=fBasics [dostęp: 5.10.2020].
[21] ZOOM H3-VR Handy Recorder | Zoom; https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder [dostęp: 5.10.2020].

# 4. Techniki wielokanałowe wykorzystywane w koncertach i nagraniach muzycznych na odległość

Bartłomiej Mróz[1, 2], Piotr Odya[1],
Bożena Kostek[2]

[1] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki,
Katedra Systemów Multimedialnych,
ul. Gabriela Narutowicza, 80-233 Gdańsk
[2] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki,
Laboratorium Akustyki Fonicznej,
ul. Gabriela Narutowicza, 80-233 Gdańsk

W czasie pandemii koronawirusa COVID-19 nowego znaczenia nabrały możliwości transmisji dźwięku z obrazem – zwłaszcza do pracy zdalnej, która w przypadku muzyków jest szczególnym wyzwaniem zarówno w kontekście wspólnych ćwiczeń i prób, jak i koncertów. Wynikła konieczność wielośródłowego połączenia ujawniła potrzebę uprzestrzennienia dźwięku w celu łatwiejszej lokalizacji źródeł dźwięku. Tworzenie zdalnych nagrań muzycznych stało się obecnie niepowtarzalną okazją do produkcji wielokanałowych, przestrzennych, wykorzystujących techniki ambisoniczne i binauralne. Techniki te umożliwiają stworzenie nowych, immersyjnych doznań dla słuchaczy. W rozdziale przedstawiono zrealizowane nagrania ambisoniczne Akademickiego Chóru Politechniki Gdańskiej. Zawarto opis związany z warsztatem realizatora dźwięku i obrazu oraz omówiono problemy związane z synchronizacją dźwięku. W pierwszej kolejności przedstawiono krótko podstawy teoretyczne ambisonii. Podano również plan dalszych prac, które będą stanowić rozwinięcie wykonanych nagrań w kontekście ich oceny.

# 4.1. Wprowadzenie

W ciągu ostatnich lat postęp technologiczny umożliwił szerokie zastosowanie urządzeń do pracy zdalnej w czasie rzeczywistym. Tym niemniej, zdalna praca twórcza i artystyczna, a w szczególności wykonawstwo muzyczne nadal pozostają wyzwaniem. Jednokierunkowe połączenia, dzięki którym wydarzenia artystyczne są transmitowane w czasie rzeczywistym, stały się dostępne przy użyciu sieci zdolnych obsługiwać dużą przepustowość, trasowanie (*routing*) pakietów o małych opóźnieniach i gwarantowaną jakość usług (Quality of Service; QoS) [10, 13, 14]. Można je określić jako jednokierunkowe, umożliwiające artystom zdalne połączenie jedynie między miejscem wykonania a publicznością, a nie między sobą. Dużo trudniej jest, gdy wykonawcy w dwóch lub więcej odległych lokalizacjach próbują razem wykonać ustaloną kompozycję lub improwizację w czasie rzeczywistym. Weinberg [53] określa ten sposób wykonywania muzyki podejściem „pomostowym" (*bridge approach*). W takich przypadkach wiadomo, że nieuniknione opóźnienie spowodowane fizycznym czasem tranzytu pakietów sieciowych ma wpływ na wydajność [4, 6, 12, 26, 37, 39]. W związku z tym próbowano uwzględnić te opóźnienia przez projekt lub kompozycję [7, 9, 28].

Warto nadmienić, że pierwsze tego typu realizacje zastosowano w sieciach akademickich, a eksperymentalny koncert z artystami (Amfc Vocal Consort, Schola Cantorum Gedanensis i Capella Cracoviensis) z różnych miast w Polsce miał miejsce już w 2001 r. [36]. Z okazji jubileuszu 10-lecia Internetu w Polsce Naukowa i Akademicka Sieć Komputerowa (NASK) oraz Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW (ICM) przy udziale Akademii Muzycznej im. Fryderyka Chopina zorganizowały 14 września 2001 r. eksperymentalny koncert internetowy wykonywany w trzech miastach (Warszawa, Gdańsk, Kraków) przy wykorzystaniu łączności internetowej. W ramach koncertu zostało przedstawione prawykonanie utworu Stanisława Moryto, *Ad laudes* napisanego specjalnie na 10-lecie Internetu, ale jednocześnie upamiętniające ofiary zamachów terrorystycznych z 11 września 2001 r. Dźwięk i obraz były przesyłane za pośrednictwem Internetu przez sieć POL-34. Muzycy widzieli się wzajemnie na monitorach, a zgromadzeni w tych trzech miastach słuchacze oglądali koncert na monitorach plazmowych. Równolegle transmisja odbywała się przez wszystkie ośrodki regionalne TVP3. Kierownictwo artystyczne przyjął dyrygent R. Zimak [36]. Trzeba wspomnieć również o warstwie technicznej połączenia sieciowego. Zestawione kanały miały przepustowość do 15 Mb/s. Kodowanie sygnału wizyj-

nego odbywało się w standardzie MJPEG, co zapewniło transmisji minimalne opóźnienie – najważniejsze ze względu na powodzenie eksperymentu. Realnie uzyskano opóźnienia ok. 180 ms między miastami w jedną stronę, z czego ok. 70 ms zajmowało kodowanie i dekodowanie sygnału.

W transmisjach strumieniowych jednak 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, 8, 17, 18, 35, 54]. Należy nadmienić o platformie AltSpaceVR [2], w której można przeprowadzać spotkania telekonferencyjne, pokazy, prezentacje, zajęcia czy spotkania towarzyskie; dźwięk wszystkich obiektów w platformie podlega auralizacji i binauralizacji.

W przypadku tzw. sieciowych/zdalnych wykonań muzycznych (Networked Music Performance; NMP) [48] już wcześniej zostały podjęte udane próby transmitowania dźwięku wielokanałowego [47, 50, 55, 58]. Przeprowadzono również transmisje z wykorzystaniem ambisonii [27]. Ambisonię można traktować jako format transmisyjny umożliwiający tworzenie przestrzennego dźwięku 3D. Opiera się na reprezentacji pola dźwiękowego przez rozłożenie go na podstawowe funkcje ortonormalne – zwane harmonicznymi sferycznymi. Taka reprezentacja umożliwia 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 r. był z pewnością jednym z pionierskich wydarzeń tego typu [23, 44, 49]. 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). Sala koncertowa była przygotowana do rejestracji 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 Meolą w lipcu 2016 r. obejmującego przestrzenne efekty w czasie rzeczywistym i transmisję do innego wnętrza [23].

Wszystkie te prezentacje miały jednak charakter eksperymentalny. Aby stały się standardem nagrań czy transmisji, muszą zostać zaimplementowane w powszechnie używanych platformach. Na przykład 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 [56]. Kolejna platforma – Face-

book wspiera transmisję na żywo z obrazem 360° oraz ambisonią 1. rzędu [56], przesłane gotowe nagranie natomiast może zawierać dźwięk ambisoniczny 2. rzędu [20, 21]. Powstała też zupełnie nowa platforma ukierunkowana na prezentację treści z obrazem dookólnym oraz dźwiękiem zmiksowanym w ambisonii wyższych rzędów – HOAST [16]. 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 znajdującej się w odległych salach koncertowych. Przeszkodą dla pełnego NMP nadal są opóźnienia rzędu 25 ms, w szczególności przy wykorzystaniu popularnych, konsumenckich łącz jak ADSL. Dopiero technologia 5G daje nadzieję na pokonanie tych trudności [11].

Z tego względu zdalne wykonania muzyczne obecne są w postaci nagrań tworzonych w postprodukcji na podstawie nagrań poszczególnych partii utworów przesłanych przez ich wykonawców. Najbardziej znane są dokonania kompozytora E. Whitacre'a, który od 2009 r. organizuje koncerty tzw. wirtualnego chóru [19]. Powstało sześć edycji tych sesji, a w każdej z nich uczestniczyły tysiące osób z całego świata. Zwłaszcza ostatnia – 6. edycja spotkała się ze zdecydowanie większym zainteresowaniem, co wynikało z tego, że odbyła się wiosną 2020 r., kiedy COVID-19 został uznany za światową pandemię. W tym czasie muzycy szczególnie intensywnie zaczęli szukać alternatywnych, zdalnych metod wykonawczych; zapotrzebowanie na takie rozwiązania może potwierdzić to, że 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 [1, 24, 29, 34, 52]. Jedną z nich jest utwór Giovanniego Pierluigi da Palestriny, *Sicut Cervus* wykonany przez Akademicki Chór Politechniki Gdańskiej w pionierskim nagraniu, *Wirtualna katedra* [51]. Omówienie produkcji tego nagrania zostanie przedstawione w następnych podrozdziałach.

W niniejszym rozdziale przedstawiono pokrótce przegląd realizacji koncertów na odległość i trudności z tym związane. Odniesiono się również do podstaw teoretycznych ambisonii będącej jednym z formatów reprodukcji dźwięku, w którym nacisk położony jest na immersję słuchacza w scenie dźwiękowej. W sposób szczegółowy pokazano proces realizacji nagrania koncertu Akademickiego Chóru Politechniki Gdańskiej, począwszy od etapu przygotowań, przez nagranie muzyków aż do postprodukcji. Zawarto opis związany z warsztatem realizatora dźwięku i obrazu oraz omówiono problemy dotyczące synchronizacji dźwięku. W Podsumowaniu przedstawiono dalsze plany odnośnie do realizacji koncertów Akademickiego Chóru Politechniki

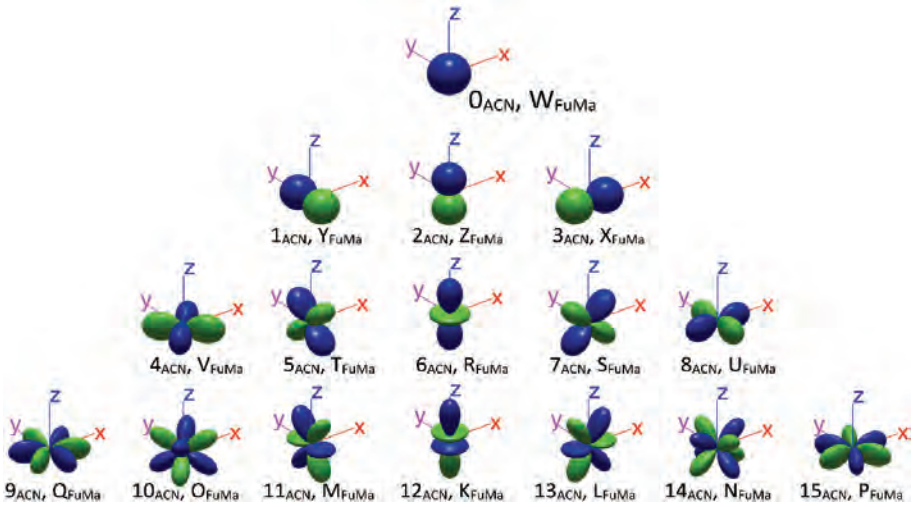
Gdańskiej, w szczególności – do zaprojektowania testów subiektywnych umożliwiających ocenę zrealizowanych nagrań.

## 4.2. Ambisonia – podstawy

Ambisonia została zaproponowana przez M. Gerzona, P. Felgetta i G. Bartona na początku lat 70. na Uniwersytetach w Oxfordzie i Surrey [25]. Ambisonia odwołuje się do „peryfonii" (*periphony*), czyli techniki odtwarzania dźwięku zarówno w pionie, jak i w poziomie wokół słuchacza. Zgodnie z definicją: ambisonia to pełnosferyczny format dźwięku przestrzennego odnoszący się do sceny dźwiękowej reprezentowanej przez zbiór tzw. harmonik sferycznych. W przeciwieństwie do konwencjonalnych formatów dźwięku stereofonicznego i przestrzennego (opierających się na zasadzie przesyłania sygnałów dźwiękowych do konkretnych głośników) ambisonia przechwytuje pełną informację o kierunkowości dla każdej fali dźwiękowej, która jest rejestrowana przez mikrofon – również w płaszczyźnie wertykalnej [46]. W ambisonii istotne jest rozróżnienie między tzw. A-Formatem oraz B-Formatem. A-Format jest zapisem sygnału właściwym dla danego mikrofonu – to bezpośredni zapis z każdej kapsuły mikrofonowej. W tej postaci nie jest on jeszcze użyteczny, każdy producent zapewnia jednak metody konwersji dla swojego mikrofonu do domeny ambisonicznej, czyli do B-Formatu. W B-Formacie wyróżnia się dwa sposoby numeracji kanałów: FuMa (Furse-Malham) oraz ACN (Ambisonic Channel Number). W numeracji FuMa stosuje się oznaczenia literowe, a w numeracji ACN – liczbowe; ponadto kanały te są podawane w innej kolejności. Przykładowo: 1. rząd ambisonii ma cztery kanały zapisywane w notacji FuMa WXYZ (W – wszechkierunkowy, tzw. omni, X – przód–tył, Y – lewy–prawy, Z – góra–dół). W notacji ACN będą to odpowiednio kanały: 0, 3, 1, 2. Liczba kanałów w ambisonii 2D równa jest 2N + 1, w ambisonii 3D natomiast – odpowiednio: $(N+1)^2$.

Na rysunku 1 przedstawiono ilustrację szesnastu (czyli do 3. rzędu) harmonik sferycznych razem z numeracją kanałów ambisonicznych FuMa i ACN.

Warto też wspomnieć o formatach zapisu, począwszy od historycznego UHK do aktualnych: AMB – Microsoft Wave Format Extensible (WAVE-EX), zapis B-Formatu ze współczynnikami wagowymi FuMa. Główną wadą tego formatu pliku bazującego na opracowanym przez Microsoft kontenerze WAV jest ograniczenie wielkości pliku do 4 GB. Ambisonic Exchange (AmbiX) wykorzystuje z kolei jako kontener Apple'owski .caf (Core Audio Format).




Rys. 1. Ilustracja harmonik sferycznych wraz z numeracją FuMa i ACN odpowiadającą kanałom w ambisonii

Kanały są poddane normalizacji w zależności od formatu zapisu sygnałów:

- MaxN – normalizuje każdy pojedynczy komponent, aby nie przekraczał wzmocnienia 1,0 dla panoramowanego źródła monofonicznego, używany w FuMa;
- N3D (podobnie jak SN3D) – ortonormalna podstawa dekompozycji 3D. Zapewnia równą moc kodowanych komponentów w przypadku idealnie rozproszonego pola 3D;
- SN3D – (w kolejności kanałów ACN) szeroko stosowana. W przeciwieństwie do N3D żaden komponent nigdy nie przekroczy wartości szczytowej komponentu 0. rzędu dla źródeł jednopunktowych. Ten schemat został przyjęty w formacie kodowania AmbiX i jest powszechnie stosowany.

Dekodowanie obejmuje natomiast algorytmy i narzędzia, takie jak: AllRAD [58] / AllRAD2 [60], Harpex [5], DirAC [42, 43]. AllRAP (All-Round Ambisonic Panning) jest to algorytm dla dowolnych aranżacji głośników mający na celu stworzenie źródeł pozornych o stabilnej głośności i regulowanej szerokości. Metoda AllRAD (All-Round Ambisonic Decoding) pasuje do koncepcji formatu ambisonicznego. Konwencjonalne dekodowanie ambisonii jest proste tylko przy optymalnych ustawieniach głośników, dla których uzyskuje się niezależną od kierunku energię i rozproszenie energii, szacowaną głośność i szerokość źródła pozornego. Algorytm AllRAP/AllRAD jest nadal prosty, ale bardziej wszechstronny i wykorzystuje kombinację wirtualnego optymalnego ustawienia głośników z funkcją VBAP (Vector-Base Amplitude Panning). Dekoder

Harpex (High Angular Resolution Planewave Expansion) jest z kolei narzędziem, które łączy przestrzenną ostrość metod parametrycznych z fizyczną poprawnością dekodowania liniowego bez wprowadzania słyszalnych artefaktów. Algorytm DirAC (Directional Audio Coding) wykorzystuje metodę SIRR (Spatial Impulse Response Rendering). W SIRR analizowane w pasmach częstotliwości są odpowiedzi impulsowe pomieszczenia, ich rozproszenie oraz zależny od czasu kierunek nadejścia. Na podstawie danych analitycznych syntetyzowana jest odpowiedź wielokanałowa nadająca się do reprodukcji z dowolną wybraną konfiguracją głośników *surround*. Trzeba też dodać, że specjalnym przykładem dekodowania ambisonii jest użycie funkcji HRTF, aby zdekodować sygnał ambisoniczny do formatu binauralnego [57].

## 4.3. Realizacja zdalnego nagrania muzycznego

Ważnym odniesieniem do realizacji dźwięku w kontekście zarówno strumieniowania koncertu, jak i możliwości reprodukcji dźwięku immersyjnego, czyli wywołania u widza/słuchacza efektu „zanurzenia w dźwięku", jest rozdział autorstwa S. Meltzera, A. Murtaza oraz G. Pietrzyka, *Standard MPEG-H 3D Audio i jego zastosowania w telewizji cyfrowej* [41] (zawarty w książce, *Postępy badań w inżynierii dźwięku i obrazu. Nowe trendy i zastosowania technologii multimedialnych*, B. Kostek (red.)) [38], w którym przywołują podstawy przesyłania dźwięku immersyjnego, jakim jest użycie standardu MPEG-H Audio wykorzystującego ambisonię wyższego rzędu (Higher Order Ambisonics; HOA) jako optymalne wejście do kodeka dźwięku 3D. Technika ambisoniczna została zaprojektowana w taki sposób, aby nagrywać i później reprodukować pełne pole dźwiękowe za pomocą obiektów dźwiękowych. Wymaga to przesyłania informacji o położeniu obiektu w przestrzeni i czasie w postaci metaopisu. Metadane umożliwiają poprawne renderowanie obiektów po stronie odtwórczej [41].

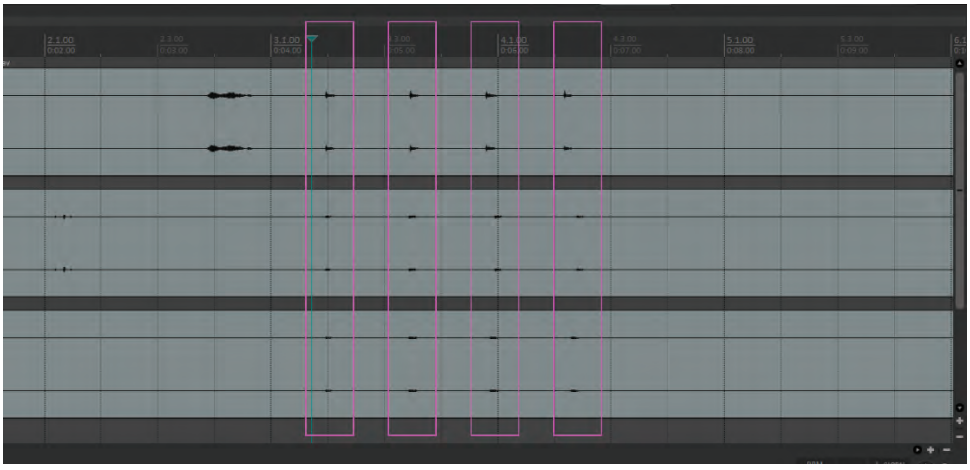
W niniejszym podrozdziale przedstawiony zostanie etap realizacji nagrania koncertu, począwszy od etapu przygotowania po postprodukcję nagrania.

### 4.3.1. Przygotowanie audiowizualnego nagrania zdalnego

Pierwszym krokiem w nagraniu audiowizualnym jest omówienie materiału muzycznego z wykonawcami (chórzystami): przedstawienie interpretacji muzycznej materiału przez dyrygenta lub lidera zespołu, wskazanie miejsc w partyturze (ewentualnie

ich zaznaczenie) wymagających szczególnej uwagi itd. 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 zapewnienie repozytorium do nadsyłania plików z nagraniami przez uczestników projektu. To istotne ze względu na organizację pracy – w ten sposób uczestnicy nagrania nie będą musieli we własnym zakresie szukać sposobu nadsyłania nagrań, co znacząco usprawni komunikację w projekcie nagrania.

Następnym krokiem jest stworzenie „nagrania wzorcowego" – na jego podstawie 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 (lub innym instrumencie, chociaż pianino klasyczne najczęściej towarzyszy chóralnym próbom – warstwę dźwiękową może wykonać też akompaniator). Potem do skonstruowanego w ten sposób materiału – fonia + wizja = dyrygent + akompaniament własne nagrania audiowizualne tworzą muzycy reprezentujące swoje partie. Na przykład dla chóru są to: sopran, alt, tenor i bas, a dla kwintetu smyczkowego: I skrzypce, II skrzypce, altówka, wiolonczela, kontrabas. Ważne jest, aby te nagrania były nieco bardziej ekspresyjne niż typowo – w tym celu odtwórcy takiego „wzorca" dostaną dodatkową warstwę informacji, z której będą mogli odtworzyć wspólną interpretację utworu. Poza tym niezwykle pomocne może być ustalenie z wykonawcami, żeby przed wykonaniem swojej partii w umówiony sposób wystukali 1–2 takty tempa (np. 1–2 takty przed rozpoczęciem utworu). Umożliwi to kolejnym odtwórcom „wzorca" wczuć się w rytm oraz przygotować do nagrania. Po-

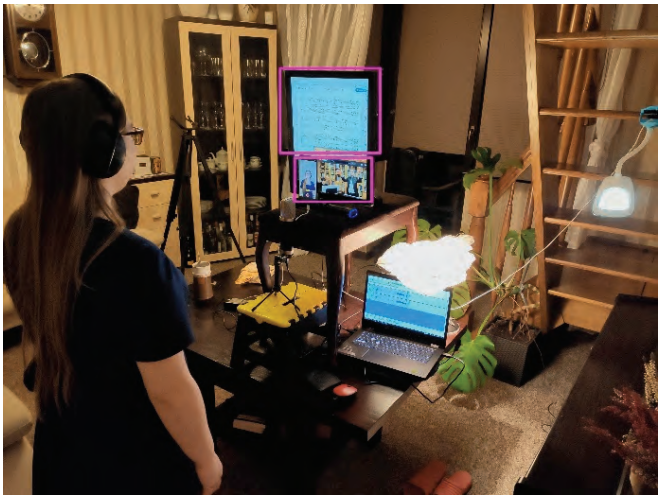


Rys. 2. Wstępna synchronizacja ścieżek do czterech klaśnięć z początku nagrania

nadto w postprodukcji to dobry punkt odniesienia do synchronizacji ścieżek – co więcej dzięki temu będzie odpowiednie miejsce przed rozpoczęciem wykonania do edycji w postprodukcji, uniknie się zatem sytuacji, w której dane nagranie jest za krótkie, ponieważ ktoś na przykład nie był jeszcze gotowy do wykonania swojej partii. Można również skorzystać z metronomu, z doświadczeń autorów jednak wynika, że jest to zbędne rozproszenie uwagi wykonawcy ze względu na obsługę dodatkowego urządzenia. Poza tym w przedstawianych nagraniach metronom byłby zapewne zrealizowany za pomocą telefonu zwykle wykorzystywanego w nagraniu w inny sposób (np. jako urządzenie do odtworzenia wzorca lub do przeprowadzenia nagrania). Na rysunku 2 przedstawiono obrazowo ideę takiego podejścia.

### 4.3.2. Nagrania muzyków

Przygotowane nagrania wzorcowe powinny stanowić bazę, do której dogrywać swoje partie będą pozostali członkowie zespołu. Bardzo wygodne jest umieszczenie wzorców w platformie YouTube jako niepublicznych nagrań, 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. Na ilustracji z rys. 3 zaprezentowano sytuację nagraniową w opisanych warunkach. Nagrania nadesłane



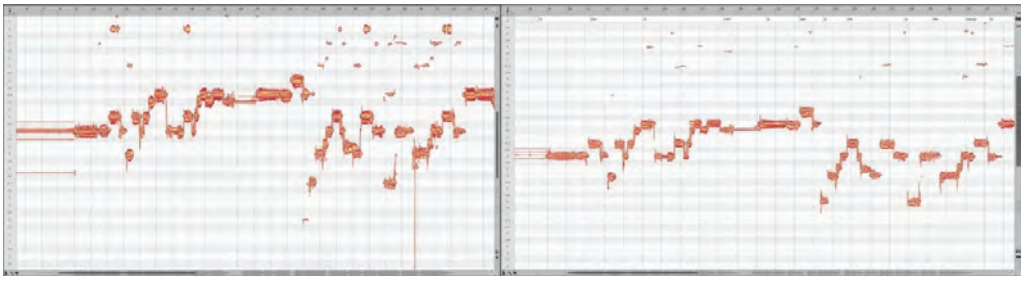
Rys. 3. Przykładowe ustawienie nagrania w warunkach domowych; na ekranie widoczne są: zapis nutowy, wzorzec, a także osobne urządzenie (laptop) rejestrujące dźwięk

przez muzyków powinny być zweryfikowane przez dyrygenta i/lub producenta projektu pod względem zarówno artystycznym, jak i technicznym – w szczególności ze względu na estetykę kadru oraz wierność odtworzenia wzorca, zwłaszcza od strony rytmicznej. Im wierniejsze wykonanie, tym mniej korekt trzeba zastosować w późniejszym etapie procesu wydania nagrania.

### 4.3.3. Postprodukcja warstwy dźwiękowej

W celu osiągnięcia najlepszego efektu konieczne są ręczne poprawki każdej z nadesłanych ścieżek dźwiękowych. Niestety większość muzyków nie ma w domu warunków do profesjonalnej realizacji nagrań. W przypadku zespołu amatorskiego, jakim jest chór akademicki, większość uczestników wykonała swoje nagrania z użyciem telefonu komórkowego – dlatego najczęściej wymagało ono usunięcia szumów. Można tego dokonać za pomocą bramek szumów oraz filtrów FIR.

Innym zagadnieniem jest korekta samej warstwy melodyczno-rytmicznej. O ile nierówności intonacyjne w dużych grupach wykonawczych mogą zostać przesłonięte przez pozostałe bardziej poprawne wykonania, o tyle nierówności rytmiczne są dużo bardziej zauważalne i osłabiają dobre wrażenie odbioru produkcji. Żeby usunąć te problemy, wykorzystano dostępne wtyczki VST, m.in. wbudowaną w cyfrową stację roboczą (Digital Audio Workstation; DAW) Reaper [45] wtyczkę ReaTune służącą do korekty linii melodycznych. Bardzo dobrym narzędziem jest też wtyczka Melodyne Studio [40] – to niezwykle rozbudowane narzędzie umożliwiające korygowanie nie tylko wysokości czy czasu trwania dźwięku, lecz także mające algorytmy do edycji formantów, czasu ataku poszczególnych dźwięków czy przejść między dźwiękami. Ilustrację pracy w tym środowisku pokazano na rys. 4.



Rys. 4. Nagranie przed korektą i po korekcie w programie Melodyne Studio (partia altowa)

Duże grupy wykonawcze najczęściej nagrywa się w dużych salach koncertowych, a nie indywidualnie, każdego z osobna. W ten sposób zrealizowane nagranie umożliwia potraktowanie ścieżek obiektowo i uprzestrzennienie ich w ambisonii. Ważne jest, aby zadbać o zapewnienie możliwie najbardziej wiernej akustyki zarówno wirtualnego pomieszczenia, jak i renderowanych „obiektów", tj. chórzystów. W tym celu można nadać wzorzec kierunkowości dla każdego obiektowego źródła. Ponadto wirtualną akustykę można modelować za pomocą odbić pozornych, a także przez dodanie pogłosu rozproszonego. Zdecydowano się wykorzystać wtyczki VST z IEM Plug-in Suite [33], a w szczególności DirectivityShaper [30], RoomEncoder [32] oraz FDNReverb [31]. Na rysunku 5 przedstawiono przykładowe ustawienie tych trzech wtyczek VST.


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

## 4.3.4. Postprodukcja warstwy wizualnej

Przy produkcji wideo klasycznego chóru wirtualnego dużo uwagi wymaga zaplanowanie wszystkich ujęć, a także przejść między nimi. W omawianym nagraniu zdecydowano się na statyczny obraz 360°, na którym „zawieszone" będą nagrania chórzystów. W całym nagraniu widoczne są te same osoby, gdyż przejścia między nimi byłyby nieestetyczne – ponadto w nagraniu 360° twórca nie ma kontroli nad tym, gdzie w danym momencie patrzy widz. Dlatego obraz w całej sferze został niezmienny.

Zdecydowano się natomiast na stworzenie dopracowanego modelu, który posłużył do wygenerowania tła. Postanowiono wykorzystać 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 podkład wizualny mógł posłużyć jako tło do ścieżek obrazu w programie do montażu nieliniowego – DaVinci Resolve [15]. Na rysunku 6 zaprezentowano układ chórzystów rozmieszczonych na tle opracowanej grafiki.

Tak stworzony model można następnie połączyć z dźwiękiem w B-formacie. W przypadku ambisonii pierwszego rządu jest to 4-kanałowa ścieżka audio. W tym celu posłużyć może program FB360Encoder [22] 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 zawartym w pliku metadanym). Po przetworzeniu przez platformę jest on gotowy do odtwarzania.

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

## 4.4. Podsumowanie

W niniejszym rozdziale przedstawiono i krótko omówiono dotychczasowe dokonania w dziedzinie zdalnych połączeń audio-wizualnych. Przedstawiono trudności zwią-

zane 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. Mimo że istniejące platformy zdalnego połączenia nie zapewniają pełnej możliwości realizacji nagrań muzycznych, jakie dają obecnie technologie immersyjne, nagranie to prezentuje wysoki poziom artystyczny – uzyskało szereg pozytywnych opinii i komentarzy nie tylko od dyrygenta, chórzystów i słuchaczy Akademickiego Chóru Politechniki Gdańskiej, lecz także od jurorów wirtualnego festiwalu chóralnego: Bandung Choral Society World Virtual Choir Festival 2021 w Indonezji, w ramach którego nagranie zdobyło złotą nagrodę za: „[...] wybitne klasyczne wykonanie". A to oznacza, że zapewnienie słuchaczom nowych doznań artystycznych wynikających z immersji w wirtualnej rzeczywistości może podkreślać walory chórów wirtualnych, a zwłaszcza ich innowacyjny charakter i nowe możliwości twórcze mogące stać się wartością dodaną kultury XXI w. Niewątpliwie zarysowane techniki wymagają zapewnienia starannego procesu postprodukcji przygotowanego nagrania.

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

**Słowa kluczowe:** nagrania na odległość, techniki wielokanałowe, ambisonia.

## Bibliografia

[1] *A socially-distanced, 360 performance of Puccini's Turandot* (Royal Opera House Chorus and Orchestra); https://youtu.be/VwOpNf8eHeY [dostęp: 1.06.2021].

[2] AltSpaceVR; https://altvr.com/ [dostęp: 1.06.2021].

[3] Aoki S., Cohen M., Koizumi N., *Design and control of shared conferencing environments for audio telecommunication using individually measured HRTFs*, „Presence: Teleoperators and Virtual Environments" 1994, 3(1), s. 60–72.

[4] Bartlette C., Headlam D., Bocko M., Velikic G., *Effect of network latency on interactive musical performance* „Music Perception" 2006, 24(1), s. 49–62.

[5] Berge S., Barrett N., *High angular resolution planewave expansion*, Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, 6–7 maja 2010.

[6] 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, 4–6 czerwca 2009, s. 135–140.

[7] 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, 6–10 czerwca 2007, s. 142–147.

[8] Buxton W., *Telepresence: integrating shared task and person spaces*, Proceedings of Graphics Interface '92, 11–15 maja 1992, s. 123–129.
[9] 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, 10–12 września 2008.
[10] Caceress J.-P., Chafe C., *JackTrip/SoundWIRE meets server farm*, „Computer Music Journal" 2010, 34(3), s. 29–34.
[11] 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, 6–10 lipca 2020, DOI: 10.1109/ICMEW46912.2020.9105995, s. 1–6.
[12] Chafe C., Caceres J.P., Gurevich M., *Effect of temporal separation on synchronization in rhythmic performance*, „Perception" 2010, 39(7), s. 982–992.
[13] Chafe C., Wilson S., Leistikow R., Chisholm D., Scavone G., *A simplified approach to high quality music and sound over IP*, Conference on Digital Audio Effects, 7–9 grudnia 2000, s. 159–164.
[14] Chafe C., *Tapping into the internet as an acoustical/musical medium*, „Contemporary Music Review" 2009, 28(4), s. 413–420.
[15] *DaVinci Resolve*; https://www.blackmagicdesign.com/products/davinciresolve/ [dostęp: 1.06.2021].
[16] 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, 25–28 maja 2020.
[17] Durlach N.I., Shinn-Cunningham B.G., Held R.M., *Supernormal auditory localization*, „Presence: Teleoperators and Virtual Environments" 1993, 2(2), s. 89–103.
[18] Durlach N., *Auditory localization in teleoperator and virtual environment systems: ideas, issues, and problems*, „Perception" 1991, 20(4), s. 543–554.
[19] Eric Whitacre's Virtual Choir; https://ericwhitacre.com/the-virtual-choir/about [dostęp: 1.06.2021].
[20] Facebook 360 spatial workstation – Creating Videos with Spatial Audio for Facebook 360; https://facebookincubator.github.io/facebook-360-spatial-workstation/KB/CreatingVideosSpatialAudioFacebo- ok360.html [dostęp: 1.06.2021].
[21] Facebook 360 spatial workstation – Using an Ambisonic Microphone With Your Live 360 Video on Facebook; https://facebookincubator.github.io/facebook-360-spatial-workstation/KB/Using AnAmbisonic -Microphone.html [dostęp: 1.06.2021].
[22] Facebook 360 Spatial Workstation; https://facebook360.fb.com/spatial-workstation/ [dostęp: 1.06.2021].
[23] Frank M., Sontacchi A., *Case Study on Ambisonics for Multi-Venue and Multi-Target Concerts and Broad – casts*, „J. Audio Eng. Soc." 2017, 65(9), DOI: 10.17743/jaes.2017.0026, s. 749–756.
[24] Georgia Symphony Chorus, *Georgia On My Mind – 360° Virtual Choir with adaptive audio in 8K*; https://youtube/BrXZ63nOUhU [dostęp: 1.06.2021].
[25] Gerzon M.A., *What's wrong with Quadraphonics*, „Studio Sound" 1974, 16(12), 50/5.
[26] Gu X., Dick M., Kurtisi Z., Noyer U., Wolf L., *Network-centric music performance: practice and experiments*, „IEEE Communications Magazine" 2005, 43(6), s. 86–93.
[27] Gurevich M., Donohoe D., Bertet S., *Ambisonic spatialization for networked music performance*, 17th International Conference on Auditory Display, 20–23 czerwca 2011.
[28] Gurevich M., *JamSpace: a networked real-time collaborative music environment'*, CHI'06 Extended Abstracts on Human Factors in Computing Systems, 2006, s. 821–826.
[29] *I(solace)ion (Juliana Kay & Exaudi) | 360° – Exaudi*. https://youtu.be/HkiIUeuugk8 [dostęp: 1.06.2021].

[30] IEM Plug-in Suite – DirectivityShaper; https://plugins.iem.at/docs/directivityshaper/ [dostęp: 1.06.2021].
[31] IEM Plug-in Suite – *FDNReverb*; https://plugins.iem.at/docs/plugindescriptions/#fdnreverb [dostęp: 1.06.2021].
[32] IEM Plug-in Suite – *RoomEncoder*. https://plugins.iem.at/docs/plugindescriptions/#roomencoder [dostęp: 1.06.2021].
[33] IEM Plug-in Suite. https://plugins.iem.at/ [dostęp: 01.06.2021].
[34] BACH J.S., *Koncert na dwoje skrzypiec BWV 1043 [360°]*; https://youtu.be/mQXNneuRG3s [dostęp: 1.06.2021].
[35] Jouppi N.P., Pan M.J., *Mutually-immersive audio telepresence*, Proceedings of the 113th Audio Engineering Society Convention, 5–8 października 2002.
[36] *Jubileusz 10-lecia Internetu w Polsce*, [koncert internetowy]; http://www.internet10.pl/koncert.html [dostęp: 1.06.2021].
[37] Kapur A., Wang G., Davidson P., Cook P.R., *Interactive network performance: A dream worth dreaming?*, „Organised Sound" 2005, 10(3), s. 209–219.
[38] Kostek B., *Postępy badań w inżynierii dźwięku i obrazu. Nowe trendy i zastosowania technologii multimedialnych*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2019.
[39] 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, 25–26 czerwca 2001, s. 157–166.
[40] Melodyne Studio; https://www.celemony.com/en/melodyne/what-is-melodyne [dostęp: 1.06.2021].
[41] Meltzer S., Murtaza A., Pietrzyk G., *Standard MPEG-H 3D Audio i jego zastosowania w telewizji cyfrowej*, [w:] *Postępy badań w inżynierii dźwięku i obrazu. Nowe trendy i zastosowania technologii multimedialnych*, B. Kostek (red.), Akademicka Oficyna Wydawnicza EXIT, Warszawa 2019, s. 16–44.
[42] Murillo D., Fazi F., Shin M., *Evaluation of Ambisonics decoding methods with experimental measurements*, 2014 Mar 18, DOI: 10.14279/depositonce-4103.
[43] Pulkki V., Merimaa J., *Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests*, „Journal of the Audio Engineering Society" 2006, February, 54(1/2), s. 3–20.
[44] PURE Ambisonics Concert & the Night of Ambisonics; https://ambisonics.iem.at/icsa2015/pure-ambisonics-concert [dostęp: 1.06.2021].
[45] Reaper; https://www.reaper.fm/ [dostęp: 1.06.2021].
[46] RØDE Blog – The Beginner's Guide To Ambisonics; https://www.rode.com/blog/all/what-is-ambisonics [dostęp: 1.06.2021].
[47] Rosen T., *Is it live or is it Internet2?: Miro quartet shows how technology may change the future of live performances*, 2004; http: //www.utexas.edu/features/archive/2004/internet.html [dostęp: 1.06.2021].
[48] Rottondi C., Chafe C., Allocchio C., Sarti A., *An Overview on Networked Music Performance Technologies*, „IEEE Access" 2016, 4, DOI: 10.1109/ACCESS.2016.2628440, s. 8823–8843.
[49] Rudrich D., Zotter F., Frank M., *Efficient Spatial Ambisonic Effects for Live Audio*, Proceedings of 29th Tonmeistertagung – VDT International Convention, 17–20 listopada 2016.
[50] Sawchuk A., Chew E., Zimmermann R., Papadopoulos C., Kyriakakis C., *From Remote Media Immersion to Distributed Immersive Performance*, 2003, DOI: 10.1145/982484.982506.
[51] *Sicut Cervus – Wirtualna Katedra #ZostańWDomu #ŚpiewajWDomu [4k 360°]*; https://youtube/4dwSRNxUrlU [dostęp: 1.06.2021].
[52] *Socially Distant Orchestra plays "Jupiter" in 360°*. https://youtu.be/eiouj6HkjfA [dostęp: 1.06.2021].

[53] Weinberg G., *Interconnected musical networks: Toward a theoretical framework*, „Computer Music Journal" 2005, 29(2), s. 23–39.

[54] 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, 27 kwietnia–2 maja 1991, s. 351–359.

[55] Xu A., Woszczyk W., Settel 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." 2000, 48(7–8), s. 627–641.

[56] YouTube help – Use spatial audio in 360-degree and VR videos; https://support.google.com/youtube/answer/6395969 [dostęp: 1.06.2021].

[57] Wiggins B., Paterson-Stephens I., Schillebeeckx P., *The analysis of multi-channel sound reproduction algorithms using HRTF data*, 19th International AES Surround Sound Convention, Schloss Elmau, Germany, 21–24 czerwca 2001; http://www.aes.org/elib/browse.cfm?elib=10112, s. 111–123.

[58] Zimmermann R., Chew E., Ay S.A., Pawar M., *Distributed musical performances: Architecture and stream management*. „ACM Transactions on Multimedia Computing Communications and Applications" 2008, 4(2), DOI: 10.1145/1352012.1352018, s. 1–23.

[59] Zotter F., Frank M., *All-round Ambisonic panning and decoding*, „J. Audio Eng. Soc." 2012, 60(10), s. 807–820.

[60] Zotter F., Frank M., *Ambisonic decoding with panning-invariant loudness on small layouts (allrad2)*, 144th Audio Engineering Society Convention, 24–26 maja 2018.

# 5. Rozproszony system generowania, edycji i transmisji dźwięku wykorzystujący interfejsy Web Audio API, WEBRTC i Web MIDI API

MARCIN WALCZAK, EWA ŁUKASIK

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

W rozdziale opisano zagadnienia dotyczące generowania, edycji i transmisji dźwięku z uwzględnieniem specjalizowanych interfejsów programowania aplikacji API opartych na architekturze i protokołach sieci Web. Zaprezentowano implementację aplikacji internetowej z wykorzystaniem interfejsów Web Audio API, WebRTC i Web MIDI API. Celem, jaki sobie założono, była próba realizacji systemu, który z poziomu przeglądarki internetowej zapewniłby możliwość komunikacji użytkowników za pomocą dźwięku w czasie rzeczywistym z minimalnym opóźnieniem. Aplikację przedstawiono w kontekście innych systemów umożliwiających komunikację audio z małym opóźnieniem, stosowanych przez muzyków do wspólnego muzykowania i przeprowadzania zdalnych prób.

## 5.1. Wprowadzenie

Komunikacja międzyludzka i wymiana treści multimedialnych za pośrednictwem Internetu stały się powszechne i ogólnie akceptowane. Ostatnie miesiące koniecznej izolacji związanej z rozprzestrzenianiem się choroby COVID-19 sprawiły, że stosowanie środków komunikacji internetowej było nieodzowne we wszystkich dziedzinach życia. Bezpłatne komunikatory internetowe, np. Skype, Facebook Messenger, WhatsApp, Signal, Google Hangout, stosowane są do komunikacji prywatnej. Ze względu na cele

profesjonalne do listy tej dopisać trzeba Microsoft Teams, Zoom, Webex czy otwarte systemy oparte na BigBlueButton [2]. Istnieją również darmowe komunikatory głosowe przeznaczone dla komunikujących się ze sobą uczestników internetowych gier komputerowych, np. Discord (popularny również poza tą grupą użytkowników). Wydaje się, że w niektórych branżach takie formy komunikacji pozostaną nawet po ustaniu choroby.

Pierwszy na świecie komunikator internetowy (istniejący do dziś), ICQ, został utworzony przez izraelską firmę Mirabilis w 1996 r. W świecie muzycznym pomysły na współpracę zdalną pojawiały się jeszcze wcześniej – już w latach 70. ubiegłego wieku, choć początkowo były to próby rozproszonego wspólnego komponowania. Rozwój Internetu w XXI w. przyniósł wiele nowych możliwości, m.in. działania nazwane Networked Music Performance (sieciowe wykonywanie muzyki) [12] i w tym celu wykorzystywane systemy sieciowe umożliwiające wspólne muzykowanie, organizowanie *jam sessions*, improwizowanie oraz odbywanie prób artystycznych. Systemy te głównie oparte były na specjalnych serwerach komunikacyjnych. Satysfakcjonujące wspólne muzykowanie wymaga od systemu zachowania wielu ścisłych wymagań technicznych odnośnie do takich parametrów, jak latencja (opóźnienie) czy *jitter* (fluktuacja sygnału). Przyjęło się, że całkowite opóźnienia nie powinny przekraczać 20 ms, co odpowiada odległości fizycznej równej ok. 8 m (choć muzycy zauważają opóźnienie już poniżej 10 ms). Innym problemem jest brak synchronizacji zegarów dla strumieni pochodzących z różnych lokalizacji. Szczegółową analizę przyczyn i skutków całkowitego opóźnienia pojawiającego się w rozproszonych systemach sieciowego wykonywania muzyki przeprowadzono w pracy [12].

Internet nie został pierwotnie opracowany do przesyłania ruchu w czasie rzeczywistym. Utrata pakietów i opóźnienia wpływają negatywnie na jakość przesyłanego sygnału audio. Usługi wideokonferencji rozwiązują problemy związane z utratą pakietów dzięki zastosowaniu długich ramek audio, dużych buforów sieciowych i zachowaniu zasady retransmisji pakietów. W konsekwencji powoduje to dodatkowe opóźnienia – ok. kilkaset milisekund. Nie stanowią one problemu w kontekście komunikacji głosowej, w przypadku wymiany sygnałów muzycznych w czasie rzeczywistym jest to jednak znaczne ograniczenie. Trudności, które należało rozwiązać w celu realizacji systemów Networked Music Performance przedstawił Carot w swojej rozprawie doktorskiej [5].

W efekcie wieloletnich prac powstało kilka systemów wykorzystywanych przez muzyków do prowadzenia zdalnych prób i przygotowania nagrań z wykorzystaniem sieci Internet. Są to zarówno systemy otwarte, np. Jamulus [8], SoundJack [16], LoLa [9], jak i komercyjne, np. Artsmesh [1].

Mimo wspomnianych wcześniej zaawansowanych rozwiązań przetwarzanie audio nie było przez długi czas wspierane w przeglądarkach internetowych. Wprowadzenie elementu <audio> w HTML5 umożliwiało strumieniowe odtwarzanie dźwięku, ale bez możliwości budowy bardziej skomplikowanych aplikacji audio i bez obsługi audio w zaawansowanych grach internetowych oraz aplikacjach interaktywnych. Nie zapewniało miksowania, przetwarzania i filtrowania na równi z nowoczesnymi rozwiązaniami dostępnymi w komputerach stacjonarnych. To zapewnił dopiero specjalizowany sieciowy interfejs programowania aplikacji, w którym wykorzystuje się architekturę i protokoły sieci Web (w szczególności protokół HTTP) do komunikacji między aplikacjami znajdującymi się na oddzielnych urządzeniach w sieci – Web Audio API [19]. Pierwsza jego wersja pojawiła się w 2011 r., a prace nad nim trwają do dziś. 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, np. Tone.js [17], umożliwiających tworzenie interaktywnej muzyki w przeglądarce. Dodatkowo dzięki wspieraniu przez przeglądarki innego interfejsu, czyli Web MIDI API, jest możliwa bezpośrednia obsługa urządzeń MIDI.

Pełne wykorzystanie Web Audio API jest możliwe w połączeniu z technologią WebRTC – dzięki niej da się zbudować aplikacje internetowe zapewniające komunikację w czasie rzeczywistym. Stanowi to duże ułatwienie dla użytkowników, którzy nie muszą instalować żadnych komponentów komunikacyjnych na swoich komputerach. W niniejszym rozdziale opisano prototypowe wdrożenie i testy przykładowego systemu wykorzystującego Web Audio API, WebRTC i Web MIDI do tworzenia i przesyłania strumieni audio z wielu źródeł, a także oparty na WebRTC system biurowy do komunikacji w zasięgu pola widzenia [4].

## 5.2. Systemy rozproszonego wykonywania muzyki

W pierwszych 20 latach XXI w. powstało wiele systemów NMP (Networked Music Performance) przeznaczonych do rozproszonego wykonywania muzyki za pośrednictwem sieci Internet w czasie rzeczywistym. Ich charakterystyka została przedstawiona w przejrzystej formie w publikacji [12]. Są wśród nich systemy zapewniające jedynie transport audio (Audioscape, Jamulus, JackTrip), systemy łączące audio i wideo (Gigaport, Musinet, Soundjack, LOLA, Diamouses) i jeszcze te umożliwiające operowanie sygnałami MIDI. Systemy różnią się architekturą. Niektóre z nich

wspierają architekturę klient–serwer, inne – *peer-to-peer* (SoundJack, LOLA, Wemust, Jamberry), a niektóre – obie technologie (Diamouses, JackTrip, Musinet). W architekturze *peer-to-peer* wszystkie urządzenia są równe w hierarchii – każde urządzenie może przesyłać dane do każdego innego na tych samych prawach. W architekturze klient–serwer każdy z użytkowników przesyła swój strumień danych do centralnego serwera łączącego je wszystkie w jeden i wysyła zwrotnie połączony sygnał do każdego uczestnika. Serwer transmituje *n* strumieni, a każdy klient tylko jeden – swój. Przy takiej konfiguracji pojawia się większe opóźnienie. Czasu opóźnienia nie można jednak określić jednoznacznie, ponieważ zależy od wielu zmiennych czynników. Należą do nich między innymi:

- czas przebycia dźwięku od źródła do mikrofonu,
- czas przetwarzania blokowego w karcie dźwiękowej i w komputerze (po stronie odbiorczej i po stronie nadawczej),
- opóźnienie spowodowane pakietyzacją danych,
- konfiguracja sieci, np. – topologia sieci, medium transmisji (sieć przewodowa lub bezprzewodowa), rodzaj i liczba węzłów sieciowych,
- opóźnienie sieci,
- naprawa zagubionych pakietów i innych błędów transmisji,
- interpolacja danych.

Innym ważnym aspektem jest jakość dźwięku związana ze sposobem jego kodowania. Systemy wspierają różne ich typy – od surowych danych (nieskompresowanych), przez skompresowane bezstratnie (np. FLAC), aż do skompresowanych stratnie (np. MP3). Sposób reprezentacji danych dodatkowo wpływa na percepcję dźwięku przez użytkowników.

Ze względu na wielość czynników składających się na jakość transmitowanego dźwięku i konieczność dopasowywania parametrów transmisji systemy zdalnego muzykowania rzadko są łatwe w stosowaniu. Nawet jeśli profesjonaliści wykorzystują te systemy na co dzień, to dla zwykłych użytkowników wciąż jeszcze ich stosowanie jest z wielu względów trudne. Do najpopularniejszych systemów Networked Music Performance należą:

1. JackTrip [6] – system otwarty. Prace nad nim rozpoczęto w pierwszych latach XXI w. w CCRMA na Uniwersytecie w Stanfordzie. Służył wielu zespołom w zdalnie prowadzonej działalności artystycznej. Transmituje dane nieskompresowane z niską latencją, lecz jest dość skomplikowany w obsłudze. Okres pandemii sprawił, że system został znacznie rozwinięty pod kątem łatwości użycia i dalszej redukcji opóźnienia umożliwiającej synchronizację rytmiczną. Jest uzupełniany specjalnymi rozwiązaniami wykorzystującymi Raspberry Pi oraz WebRTC. JackTrip wydaje się być najprężniej rozwijanym systemem NMP.

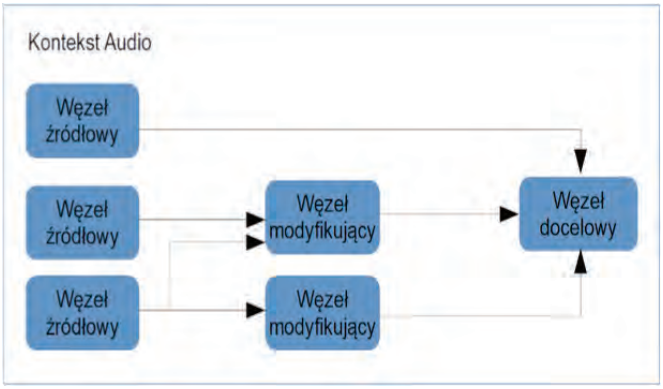
2. Sonobus [14] – najnowszy ze wszystkich systemów, powstał w 2020 r. Jego twórcą jest J. Chappell (Sonosaurus LLC). Stanowi rozwiązanie typu *peet-to-peer* i charakteryzuje się bardzo małą latencją, dużą elastycznością działania (dostosowuje swoje parametry do warunków sieci) i przyjaznym interfejsem. Przesyła dane nieskompresowane. Razem z systemem JackTrip jest najbardziej zaawansowanym rozwiązaniem do zdalnego muzykowania, nauczania i przeprowadzania zdalnych prób.
3. Jamulus [8] – system otwarty, wprowadzony do użytkowania w 2006 r. przez V. Fishera (Monachium). Ostatnie wydanie – grudzień 2020. Zbudowany w architekturze klient–serwer. Wykorzystuje publiczne serwery. Ma szerokie grono zwolenników, którzy dzięki niemu wspólnie muzykują. Co sobotę odbywa się takie zdalne muzykowanie pod nazwą WorldJam.
4. SoundJack [16] – system autorstwa A. Carota [5], powstał w 2006 r. Ma przyjazny interfejs, ale jest dosyć trudny w instalacji. Wykorzystuje najnowocześniejsze media transmisyjne, w tym technologię 5G. Nadal rozwijany. Ostatnie wydanie – listopad 2020). Cieszy się popularnością – sprawdza się jako system do prób, lekcji i wspólnego muzykowania.
5. JamKazam [7] – wprowadzony na rynek w 2014 r. Autorzy podają informację, że jest stale rozbudowywany. Synchronizacja przebiega dzięki wbudowanemu metronomowi. Żeby poprawnie działać, wymaga sprzętowego interfejsu audio, co może stanowić przeszkodę dla niektórych użytkowników. Polecany w edukacji, do przeprowadzania prób, transmisji czy zabawy z ulubionymi utworami muzycznymi.
6. Artsmesh [1] – system otwarty, powstały w 2008 r., rozwijany przez zespół kierowany przez K. Fieldsa, który rozpoczynał pracę nad systemem. Umożliwia wspólne muzykowanie muzyków będących w różnych zakątkach świata (*peer-to-peer*, audio/wideo i transmisja). Obecnie rozwijany w Chinach.
7. LOLA [9] – wysokiej jakości zintegrowana platforma transmisji i interakcji audio/wideo wymagająca doskonałej sieci. Praca nad systemem rozpoczęła się w 2005 r. w Conservatorio di Musica Giuseppe Tartini w Pizie. Pierwsza demonstracja odbyła się w 2010 r., a wersję beta wydano w 2019 r. Jako jedyna umożliwiała przekaz wideo połączony z audio. Dziś wydaje się być zdominowana przez inne rozwijające się systemy.

Należy zaznaczyć, że obecnie przy większym zapotrzebowaniu na systemy rozproszonego wspólnego muzykowania sytuacja na rynku zmienia się dynamicznie – powstają nowe systemy, coraz łatwiejsze w obsłudze, często wykorzystujące nowe możliwości przeglądarek oparte na Web Audio API, WebRTC i Web MIDI.

## 5.3. Web Audio API oraz Web MIDI API

Web Audio API jest JavaScript-owym interfejsem API wysokiego poziomu do przetwarzania i syntezy dźwięku w aplikacjach internetowych. Umożliwia generowanie, przekazywanie, modyfikację, analizę i podgląd wielu strumieni audio wraz z kontrolą czasu rzeczywistego z poziomu przeglądarki. Podstawowym paradygmatem tej technologii jest graf routingu audio nazywany również kontekstem audio (*AudioContext*), w którym wiele obiektów *AudioNode* (węzłów audio) jest połączonych ze sobą, aby zapewnić wygenerowanie końcowej wersji dźwięku (renderowanie). Modularna budowa zapewnia interfejsowi elastyczność i umożliwia tworzenie złożonych operacji na strumieniach audio z użyciem efektów dynamicznych. Dzięki Web Audio API jest możliwe generowanie i przetwarzanie sygnału audio w czasie rzeczywistym. Graf definiuje przepływ strumieni audio (*audio streams*) od węzłów źródłowych, przez węzły modyfikujące (wykorzystujące algorytmy cyfrowego przetwarzania sygnałów), do węzłów docelowych (np. głośników) [15, 19]. Przykład grafu kontekstu audio przedstawiono na rys. 1. Możliwe jest tworzenie wielu osobnych kontekstów audio za pomocą wspomnianego wcześniej konstruktora obiektu *AudioContext*.

Wyjście każdego węzła audio może zostać połączone z wejściem dowolnej liczby węzłów za pomocą metody *connect()*. Możliwe jest również odłączenie wyjścia węzła od wybranych lub wszystkich połączonych węzłów metodą *disconnect()*. Tego typu rozwiązanie zapewnia pełną konfigurowalność grafu przetwarzania strumienia audio w czasie rzeczywistym.




Rys. 1. Przykładowy graf kontekstu audio; oprac. własne

Web Audio API umożliwia zdefiniowanie dla kontekstu audio częstotliwości próbkowania w zakresie 8–96 kHz. Domyślna wartość częstotliwości próbkowania jest zależna od urządzenia, w większości przypadków jednak wynosi ona 44,1 kHz lub 48 kHz.

Praca nad dźwiękiem, w szczególności w aplikacjach muzycznych, wymaga pomiaru czasu o wysokiej precyzji. Domyślny zegar silnika JavaScript cechuje się niską dokładnością, ponieważ pojawia się w tym samym wątku, co pozostałe operacje silnika jak renderowanie, wywołania zwrotne czy usuwanie niepotrzebnych już obiektów (*garbage collection*), które mogą powodować trudne do przewidzenia opóźnienia. Z tego względu Web Audio API ma swój własny zegar w osobnym wątku. Zegar Web Audio API odsłania zegar sprzętowy urządzenia i tym samym zapewnia wysoką precyzję pomiaru czasu. Każdy kontekst audio ma swój własny zegar odmierzający czas od utworzenia kontekstu. Aktualna wartość czasu kontekstu jest dostępna w sekundach za pomocą zmiennej *AudioContext.currentTime*. Dzięki temu jest możliwa synchronizacja z wysoką dokładnością operacji na poszczególnych węzłach wewnątrz kontekstu audio.

Na kontekst audio składają się następujące węzły: źródłowe, modyfikujące i docelowe – mogą one być konfigurowane w różnoraki sposób.

**Węzły źródłowe** (*source nodes*) odpowiadają za generowanie lub odtwarzanie strumienia audio wewnątrz kontekstu audio. Nie mają wejść dla innych węzłów, mają jedno wyście, które może zostać podłączone do wielu węzłów modyfikujących lub wyjściowych. Węzłami źródłowymi są:

- węzeł oscylatora (*OscillatorNode*) – odpowiada za generowanie sygnału audio o określonym kształcie i częstotliwości. Jest węzłem jednorazowego wykorzystania – zatrzymanie jego działania powoduje, że nie można go już ponownie uruchomić;
- węzeł bufora audio (*AudioBuffer node*) – jego zadaniem jest odtwarzanie dźwięku przechowywanego w pamięci w postaci bufora audio. To węzeł jednorazowego wykorzystania – zakończenie odtwarzania dźwięku powoduje, że nie można go już ponownie uruchomić. Bufor audio stanowi reprezentację zasobu audio przechowywanego w pamięci, utworzonego z pliku audio za pomocą metody *AudioContext.createBuffer*() lub *AudioContext.decodeAudioData*(). Umożliwia przekazanie zasobu do węzła *AudioBufferSourceNode* w celu jego odtworzenia;
- węzeł elementu audio (*AudioNode*) – jest węzłem reprezentującym elementy mediów HTML5, takie jak <audio> i <video>, jako źródła sygnału audio;
- węzeł strumienia audio (*MediaStreamAudioSourceNode*) – to węzeł źródłowy otrzymujący strumień audio z WebRTC API; źródłem sygnału audio może być

zarówno urządzenie wejścia (np. mikrofon), jak i zdalne połączenie *peer-to-peer*.

**Węzły modyfikujące** (*modification nodes*) definiują i przetwarzają cyfrowo przechodzące przez nie strumienie audio. Mają i wejście, i wyście, do którego może być podłączonych wiele różnych węzłów, np.:

- węzeł wzmocnienia (*GainNode)* – kontroluje wzmocnienie przechodzących przez niego strumieni audio, ma tylko jeden parametr *gain*,
- węzeł opóźnienia (*DelayNode*) – opóźnia przechodzące przez niego strumienie audio,
- węzeł filtru rekursywnego drugiego rzędu (*BiquadFilterNode*),
- węzeł panoramy stereo (*StereoPannerNode*) – zapewnia kontrolę nad balansem między lewym i prawym kanałem stereo,
- węzeł panoramy przestrzennej (*PannerNode*),
- węzeł rozdzielania kanałów (*ChannelSplitterNode*) – rozdziela źródło sygnału audio na pojedyncze kanały monofoniczne,
- węzeł łączenia kanałów (*ChannelMergerNode*) – łączy wiele źródeł sygnału audio mono, stąd ma wiele wejść i jedno wyjście,
- węzeł splotu (*ConvolverNode*) – wykonuje splot liniowy na strumieniu audio w buforze i zadanych współczynnikach,
- węzeł kompresji dynamicznej (*DynamicsCompressorNode*) – jest odpowiedzialny za efekt kompresji amplitudy,
- węzeł kształtowania fali (*WaveShaperNode*) – stanowi nieliniowy generator zniekształceń.

**Węzeł analizy** (*AnalyserNode*) – definiuje elementy analizy czasowej i czasowo-częstotliwościowej (FFT) przechodzącego przez niego strumienia audio bez jego modyfikacji.

**Węzły wyjścia** (*destination nodes*) stanowią wyjście dla ścieżek dźwiękowych poza kontekst audio. Węzeł *AudioDestinationNode* stanowi element końcowy grafu kontekstu audio. Zwykle elementem docelowym węzła jest wyjście audio urządzenia (np. głośniki). Każdy kontekst audio ma tego typu węzeł dostępny jako parametr *AudioContext.destination*. Węzeł wyjścia strumienia audio *MediaStreamDestinationNode* stanowi element pośredni między kontekstem audio Web Audio API a WebRTC API. Zawiera on parametr *stream* będący obiektem *MediaStream*, który można dołączyć jako ścieżkę do utworzonego przez WebRTC połączenia *peer-to-peer* [13].

Jak widać, Web Audio API umożliwia generowanie, przekazywanie, modyfikację, analizę i podgląd wielu strumieni audio wraz z kontrolą czasu rzeczywistego z pozio-

mu przeglądarki. Podstawowe funkcje, jakie spełnia Web Audio API, to: 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, łączenie strumieni, przetwarzanie źródeł sygnału audio z obiektu multimedialnego <audio> lub <video>, przetwarzanie strumienia audio generowanego na żywo, przetwarzanie sygnału audio odebranego od zdalnego użytkownika i wysyłanie przetworzonego strumienia, obsługa dźwięku przestrzennego, efektywna analiza muzyki w dziedzinie czasu i częstotliwości w czasie rzeczywistym, wydajna filtracja za pomocą popularnych filtrów, kształtowanie fali i inne efekty nieliniowe. Specyfikacja Web Audio API jest wciąż jeszcze w fazie roboczej W3C, stąd zestaw funkcji stale ulega zmianom.

Zmiany dotyczą również innych mechanizmów przeglądarkowych obsługujących sygnał audio. Otóż wprowadzany stopniowo w przeglądarkach AudioWorklet stanowi uproszczoną formę *workera* [23] i umożliwia uruchamianie własnych skryptów przetwarzania strumieni audio. Został on opracowany w celu zastąpienia przestarzałego węzła *ScriptProcessorNode* [3], który uruchamia skrypt przetwarzania sygnału audio w głównym wątku silnika JavaScript. AudioWorklet wykonuje skrypt w wątku renderowania dźwięku, razem z kontekstem audio Web Audio API. Redukuje to znacząco opóźnienia w przetwarzaniu dźwięku w porównaniu do poprzednich rozwiązań. AudioWorklet jest wspierany przez przeglądarkę Google Chrome od 2018 r., pozostałe nowoczesne przeglądarki natomiast wdrożyły wsparcie dla niego w 2020 r., z wyjątkiem przeglądarki Safari [23].

Web Audio API umożliwia również integrację z WebRTC (Web Real-Time Communication) – technologią, dzięki której aplikacje internetowe mogą przechwytywać i przesyłać strumień audio i wideo oraz inne dane bezpośrednio między urządzeniami metodą *peer-to-peer* w czasie rzeczywistym, co daje podstawy do budowy systemów rozproszonego przesyłania strumieni audio.

Protokół MIDI (Musical Interface for Digital Instruments) jest powszechnie stosowanym w przemyśle muzycznym protokołem umożliwiającym komunikację między kompatybilnymi urządzeniami muzycznymi. Celem Web MIDI API jest komunikacja między przeglądarką a urządzeniami MIDI – zarówno wejścia, jak i wyjścia [21]. Dostęp do urządzeń MIDI można uzyskać za pomocą metody *navigator.requestMIDIAccess()* zwracającej obietnicę, która w przypadku sukcesu zwraca obiekt MIDIAccess zawierający

referencje do wejść i wyjść urządzeń MIDI. Web MIDI API wymaga zgody użytkownika na uzyskanie połączenia z urządzeniami MIDI.

## 5.4. WebRTC

WebRTC (Web Real-Time Communication) jest otwartym standardem umożliwiającym komunikację w sieci internetowej w czasie rzeczywistym bez użycia wtyczek programowych z wykorzystaniem interfejsów JavaScript API. Jego działanie opiera się na połączeniu *peer-to-peer* pozwalającym na bezpośrednią komunikację między urządzeniami sieciowymi. Dzięki temu możliwe jest wykonywanie połączeń głosowych, wideokonferencji oraz przesyłanie plików za pomocą przeglądarki. WebRTC obejmuje trzy zasadnicze API [18]:

1. MediaStream API zapewniający dostęp do strumieni danych audiowizualnych z wielu źródeł, jak np. mikrofony i kamery, zarówno lokalnych, jak i zdalnych dostępnych za pomocą *RTCPeerConnection*.
2. RTCPeerConnection do tworzenia, zarządzania i kończenia połączeń *peer-to-peer*.
3. RTCDataChannel obsługujący komunikację *peer-to-peer* danych niemedialnych.

**Architektura WebRTC** jest oparta na trapezoidzie (SIP Session Initiation Protocol) [10]. 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, m.in. protokołem HTTP lub WebSocket. Po nawiązaniu połączenia dalsza komunikacja między przeglądarkami następuje według modelu *peer-to-peer*, bez udziału serwerów.

Po stronie przeglądarki aplikacje internetowe oparte na technologii WebRTC mają do swojej dyspozycji WebRTC API. Stwarza to możliwość wykorzystania w czasie rzeczywistym funkcji przeglądarki internetowej.

Aplikacje internetowe korzystające z WebRTC używają serwerów pośredniczących do sygnalizacji. Aby zapewnić komunikację między użytkownikami *RTCPeerConnection*, dąży się do bezpośredniego połączenia między klientami (*peer-to-peer*). Wiele urządzeń jednak komunikuje się z siecią Internet przez NAT (Network Address Translator)oraz ma oprogramowanie antywirusowe i zapory ogniowe mogące blokować niektóre porty lub protokoły [16]. W takiej sytuacji wykorzystuje się serwery ICE

(Interactive Connectivity Establishment) – ich adresy URL zostają przekazane do obiektu *RTCPeerConnection.*

ICE stara się znaleźć optymalną ścieżkę połączenia między urządzeniami. Najpierw podejmuje się próbę ustanowienia połączenia dzięki wykorzystaniu adresów hostów pozyskanych z systemów operacyjnych i kart sieciowych urządzeń. Jeśli okaże się, że nie jest to możliwe (ponieważ na przykład urządzenia znajdują się w sieci NAT), ICE ponawia próbę z wykorzystaniem zewnętrznych adresów uzyskanych od serwera STUN (Session Traversal Utilities for NAT). W przypadku, gdy to też zawiedzie, ruch sieciowy zostaje obsłużony przez serwer TURN (Traversal Using Relays around NAT).

Zadaniem serwerów STUN jest przesyłanie publicznych adresów IP i portów aplikacji, które wysyłają do niego żądanie. Sieć wykorzystująca NAT przypisuje urządzeniom znajdującym się w niej prywatne adresy IP, jakich nie można wykorzystać poza siecią. Z tego powodu w przypadku aplikacji uruchomionych na urządzeniach wewnątrz takiej sieci, WebRTC API musi wysłać żądanie do serwera STUN w celu uzyskania informacji o jej publicznym adresie IP i porcie. Na rysunku 3 przedstawiono umiejscowienie serwerów STUN w strukturze trapezoidu SIP WebRTC.

Jeśli nie jest możliwe nawiązanie połączenia typu *peer-to-peer*, *RTCPeerConnection* może wykorzystać do pośredniczenia w wymianie danych serwer TURN. Serwery TURN służą do przesyłania danych między urządzeniami znajdującymi się za zaporą sieciową lub serwerem *proxy*. Ze względu na to, że muszą przesyłać dużo danych między klientami, wymagają one wydajnych serwerów o szerokopasmowym połączeniu sieciowym. Implementacja serwerów TURN w strukturze trapezoidu SIP przedstawiono na rys. 3.

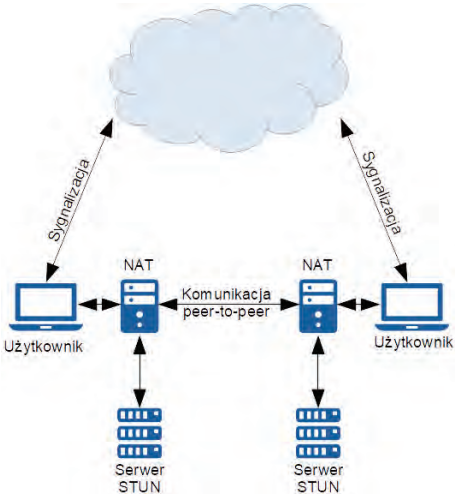
## 5.5. Realizacje praktyczne

### 5.5.1. Testowanie wbudowanych mechanizmów przeglądarkowych do wielokanałowej komunikacji za pomocą strumieni audio

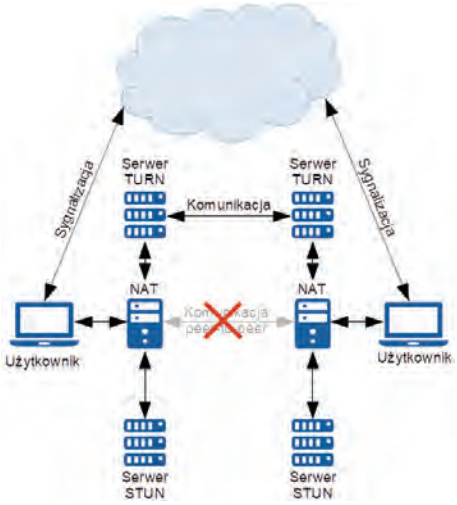
#### Opis projektu

Celem opisywanego projektu była budowa i test prototypowego systemu umożliwiającego generowanie i modyfikację przez użytkownika ścieżek dźwiękowych z wykorzystaniem przeglądarki internetowej zarówno dla strumieni wejściowych, jak i wyjściowych. Strumienie audio mogą podlegać konfiguracji oraz modyfikacji w czasie

rzeczywistym przez filtry i inne elementy przekształcające strumień audio. Możliwa jest również rejestracja sygnałów audio pochodzących z urządzeń wejścia (jak np. mikrofon) i generowanych przez aplikację strumieni audio. Zarejestrowane w ten sposób ścieżki dźwiękowe można odtworzyć w aplikacji.

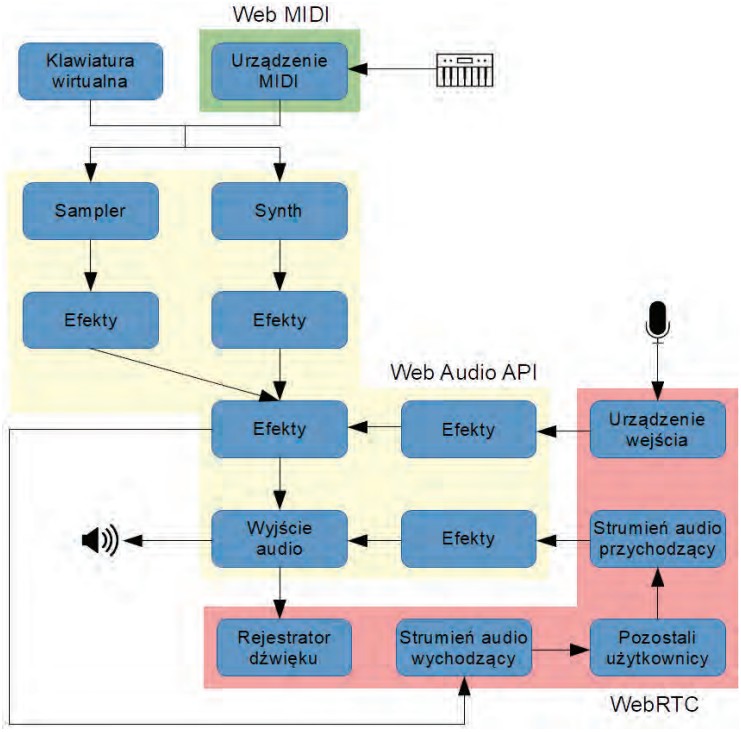



Rys. 2. Wykorzystanie serwerów STUN
do uzyskania publicznego adresu IP; oprac. własne




Rys. 3. Wykorzystanie serwerów TURN do komunikacji
w przypadku niepowodzenia
przy nawiązaniu połączenia *peer-to-peer*; oprac. własne

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 między użytkownikami. Użytkownik może tworzyć i zarządzać wirtualnymi pokojami pozwalającymi innym użytkownikom na dołączenie do nich. Osoby korzystające z tego samego pokoju mogą komunikować się między sobą dzięki połączeniu *peer-to-peer* [11]. Schemat przepływu strumieni audio wewnątrz aplikacji wraz z oznaczeniem elementów wykorzystujących technologie Web Audio API [19], WebRTC [18] i Web MIDI API [21] przedstawiono na rys. 4.



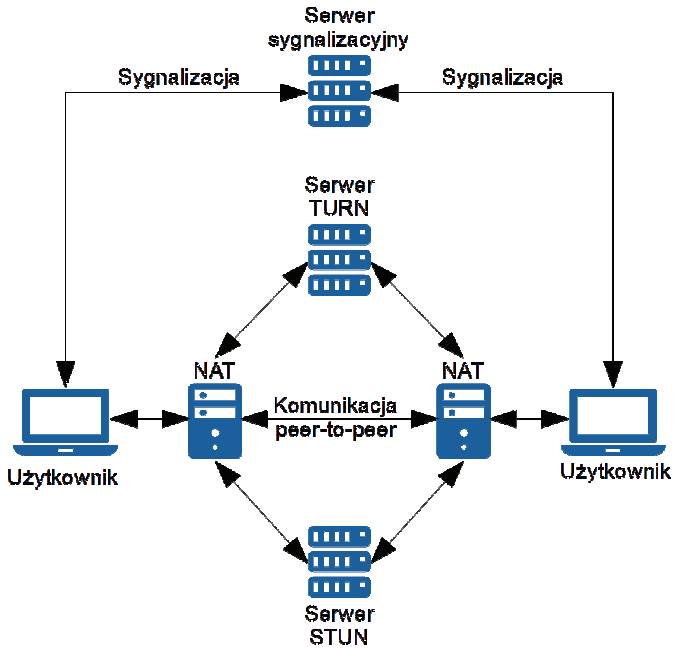

Rys. 4. Przykładowy schemat przepływu strumieni audio wewnątrz aplikacji wraz z oznaczeniem elementów wykorzystujących technologie Web Audio API, WebRTC i Web MIDI; oprac. własne

Podstawowym zadaniem serwera jest pośrednictwo przy komunikacji między klientami aplikacji. Nawiązanie połączenia *peer-to-peer* za pomocą technologii WebRTC wy-

maga 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 wówczas, gdy co najmniej jeden z nich znajduje się wewnątrz sieci NAT.

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

Wtedy, kiedy nie jest możliwe nawiązanie połączenia *peer-to-peer* ze względu na zabezpieczenia sieci, umożliwiono zastosowanie serwera TURN (rys. 5). Został on wdrożony jako zastępcza forma komunikacji w sytuacji blokowania przez sieć połączenia *peer-to-peer* z innym klientem aplikacji. Działa on wówczas jako pośrednik, czyli przesyła dane między użytkownikami. W sytuacji, w której strumień danych przechodzi przez serwer TURN zamiast bezpośrednio przez połączenie *peer-to-peer*, opóźnienia mogą ulec zwiększeniu.




Rys. 5. Architektura SIP zbudowanej aplikacji; oprac. własne

Warstwa dźwiękowa aplikacji została opracowana za pomocą Tone.js [17]. Jest to zestaw narzędzi bazujący na Web Audio API, który umożliwia generowanie, odtwarzanie i przetwarzanie dźwięku. Zawiera szereg wysokopoziomowych narzędzi rozszerzających możliwości technologii, na jakiej został oparty.

Do rejestrowania i eksportu ścieżek dźwiękowych do wybranych formatów plików audio użyto WebAudioRecorder.js [20]. Jest to biblioteka umożliwiająca rejestrowanie przechwytywanego sygnału audio, zawierająca zestaw narzędzi kodujących ścieżkę dźwiękową w formacie WAV, Ogg Vorbis lub MP3.

Komunikacja z urządzeniami MIDI została uzyskana za pośrednictwem biblioteki WebMIDI.js [22]. Dzięki niej i zastosowaniu technologii WebMIDI jest możliwa komunikacja między aplikacją internetową a urządzeniami MIDI. Dodatkowo stanowi narzędzie tłumaczące komunikaty MIDI na inny, bardziej dostępny format.

## Testowanie systemu

Przeprowadzono szereg testów aplikacji metodą białej skrzynki – czyli za pomocą zestawu testów mających za zadanie zbadanie przepływu sterowania i danych oraz poprawności działania wybranych fragmentów kodu aplikacji. Skupiono się w nich na sprawdzeniu działania poszczególnych elementów aplikacji na wybranych przeglądarkach wspierających Web Audio API i WebRTC.

Testy przeprowadzono na następujących 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.

W dalszej części rozdziału zostanie opisanych osiem przeprowadzonych testów według schematu: tytuł testu, cel, opis testu, spodziewany rezultat i faktyczny rezultat testu.

### 1. Nawiązywanie i zamykanie połączeń WebRTC między użytkownikami

**Cel testu:** sprawdzenie możliwości zarządzania połączeniami WebRTC przez aplikację.

**Opis:** w teście bierze udział co najmniej dwóch użytkowników korzystających z osobnych urządzeń. Jeden z nich tworzy pokój wewnątrz aplikacji. Następnie pozostali podejmują próbę dołączenia do pokoju i nawiązania dwustronnych połączeń WebRTC z pozostałymi uczestnikami. Potem użytkownicy podejmują próbę opuszczenia pokoju łącznie z jego twórcą.

Test przeprowadzono w następujących przypadkach:

- urządzenia korzystają z tej samej sieci lokalnej,
- urządzenia nie korzystają z tej samej sieci lokalnej,
- co najmniej jedno z urządzeń korzysta z sieci lokalnej NAT,
- co najmniej jedno z urządzeń korzysta z sieci lokalnej NAT o ustawieniach zapory sieciowej blokujących ustanowienie połączenia *peer-to-peer*.

**Spodziewany rezultat:** zostaje utworzony pokój widoczny dla pozostałych użytkowników. Użytkownicy pomyślnie dołączają do pokoju i nawiązują dwustronne połączenia WebRTC między sobą. Jeśli jedno z urządzeń korzysta z sieci NAT o ustawieniach zapory sieciowej blokujących ustanowienie połączenia *peer-to-peer*, aplikacja powinna ustanowić połączenie za pośrednictwem serwera TURN. Gdy użytkownik opuszcza pokój, wszystkie jego połączenia WebRTC z pozostałymi użytkownikami zostają zamknięte. A gdy twórca pokoju go opuszcza, pokój zostaje zamknięty, opuszczają go wszyscy użytkownicy i wszystkie połączenia WebRTC zostają zamknięte.

**Rezultat:** zgodny ze spodziewanym.

**2. Przesył strumieni audio między użytkownikami**

**Cel testu:** sprawdzenie przesyłu strumieni audio między użytkownikami pokoju.

**Opis:** w teście bierze udział co najmniej dwóch użytkowników korzystających z osobnych urządzeń. Użytkownik będący członkiem aktywnego pokoju generuje strumień audio za pomocą dostępnych wewnątrz aplikacji narzędzi.

**Spodziewany rezultat:** wygenerowany strumień audio zostaje przesłany do pozostałych członków pokoju i odtworzony przez wyjście audio ich urządzeń.

**Rezultat:** zgodny ze spodziewanym w przypadku przeglądarek Google Chrome, Mozilla Firefox, Opera i Safari. W Microsoft Edge – aplikacja nie mogła przesłać wygenerowanych strumieni audio. Było to spowodowane brakiem w implementacji Web Audio API dla tej przeglądarki węzła *AudioContext.MediaStreamAudioDestinationNode* lub jego odpowiednika (w późniejszych wersjach przeglądarki stosujących silnik Blink problem ten został wyeliminowany).

**3. Przesył danych niemedialnych między użytkownikami**

**Cel testu:** sprawdzenie przesyłu danych niemedialnych między użytkownikami pokoju.

**Opis:** w teście bierze udział co najmniej dwóch użytkowników korzystających z osobnych urządzeń. Użytkownik będący członkiem aktywnego pokoju wprowadza wiado-

mość tekstową do okna czatu aplikacji, a następnie próbuje ją wysłać do pozostałych uczestników pokoju.

**Spodziewany rezultat:** wiadomość zostaje wysłana do pozostałych uczestników pokoju i wyświetlona w oknie czatu.

**Rezultat:** zgodny ze spodziewanym.

**4. Generowanie i przetwarzanie strumieni audio przez użytkowników**

**Cel testu:** sprawdzenie możliwości generowania i modyfikacji strumieni audio przez aplikację.

**Opis:** użytkownik tworzy nowe elementy generujące strumienie audio (syntezatory i samplery (widoczne na rys. 4) i dokonuje próby wygenerowania za ich pomocą strumieni audio. Następnie dokonuje próby zmiany właściwości generowanych strumieni audio przez dodanie elementów je modyfikujących (efekty widoczne na rys. 4).

**Spodziewany rezultat:** zostają utworzone elementy generujące strumienie audio. Generowane przez nie strumienie audio ulegają zmianie w zależności zarówno od ich konfiguracji, jak i zastosowanych efektów.

**Rezultat:** zgodny ze spodziewanym.

**5. Przechwytywanie sygnałów audio pochodzących z urządzeń wejścia urządzenia**

**Cel testu:** sprawdzenie możliwości przechwytywania sygnałów audio pochodzących z urządzeń wejścia urządzenia.

**Opis:** użytkownik wywołuje żądanie dostępu do urządzenia wejścia audio.

**Spodziewany rezultat:** aplikacja przechwytuje sygnał audio pochodzący z wybranego urządzenia wejścia. Sygnał zostaje dodany do kontekstu audio aplikacji.

**Rezultat:** zgodny ze spodziewanym.

**6. Rejestracja i eksport wygenerowanych strumieni audio**

**Cel testu:** sprawdzenie możliwości rejestrowania wygenerowanego strumienia audio i jego eksportu do wybranych formatów plików audio.

**Opis:** użytkownik uruchamia rejestrację ścieżek dźwiękowych. Po zakończeniu rejestracji dokonuje próby odtworzenia zarejestrowanej ścieżki. Następnie dokonuje próby eksportu ścieżki do formatu audio WAV, Ogg Vorbis i MP3.

**Spodziewany rezultat:** udostępniona zostaje zarejestrowana ścieżka dźwiękowa w wybranym formacie pliku.

**Rezultat:** zgodny ze spodziewanym.

**7. Obsługa urządzeń MIDI**

**Cel testu:** sprawdzenie możliwości wykorzystania urządzenia MIDI do sterowania działaniem aplikacji.

**Opis:** użytkownik żąda dostępu do urządzenia MIDI.

**Spodziewany rezultat:** aplikacja dokonuje połączenia z urządzeniem MIDI, które umożliwia sterowanie aplikacją. W przypadku braku wsparcia dla urządzeń MIDI użytkownik powinien zostać o tym poinformowany.

**Rezultat:** zgodny ze spodziewanym. Przeglądarki Google Chrome, Opera i Safari umożliwiały komunikację z urządzeniami MIDI, a w Mozilla Firefox i Microsoft Edge poprawnie wykryto brak wsparcia dla Web MIDI API i wyświetlono odpowiedni komunikat.

**8. Stabilność połączeń WebRTC**

**Cel testu:** sprawdzenie maksymalnej liczby jednoczesnych stabilnych połączeń *peer-to-peer* między użytkownikami.

**Opis:** w teście bierze udział co najmniej dwóch użytkowników korzystających z osobnych urządzeń. Jeden z użytkowników tworzy pokój wewnątrz aplikacji. Następnie pozostali stopniowo podejmują próbę dołączenia do pokoju i nawiązania dwustronnych połączeń WebRTC z pozostałymi uczestnikami pokoju.

**Spodziewany rezultat:** połączenia WebRTC między użytkownikami tracą stabilność przy co najmniej dziesięciu jednoczesnych połączeniach.

**Rezultat:** testy wykazały utratę stabilności połączeń WebRTC przy 8 (do 12) jednoczesnych połączeniach, co skutkowało ich utratą.

Testy aplikacji w sieci wykazały również, 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.

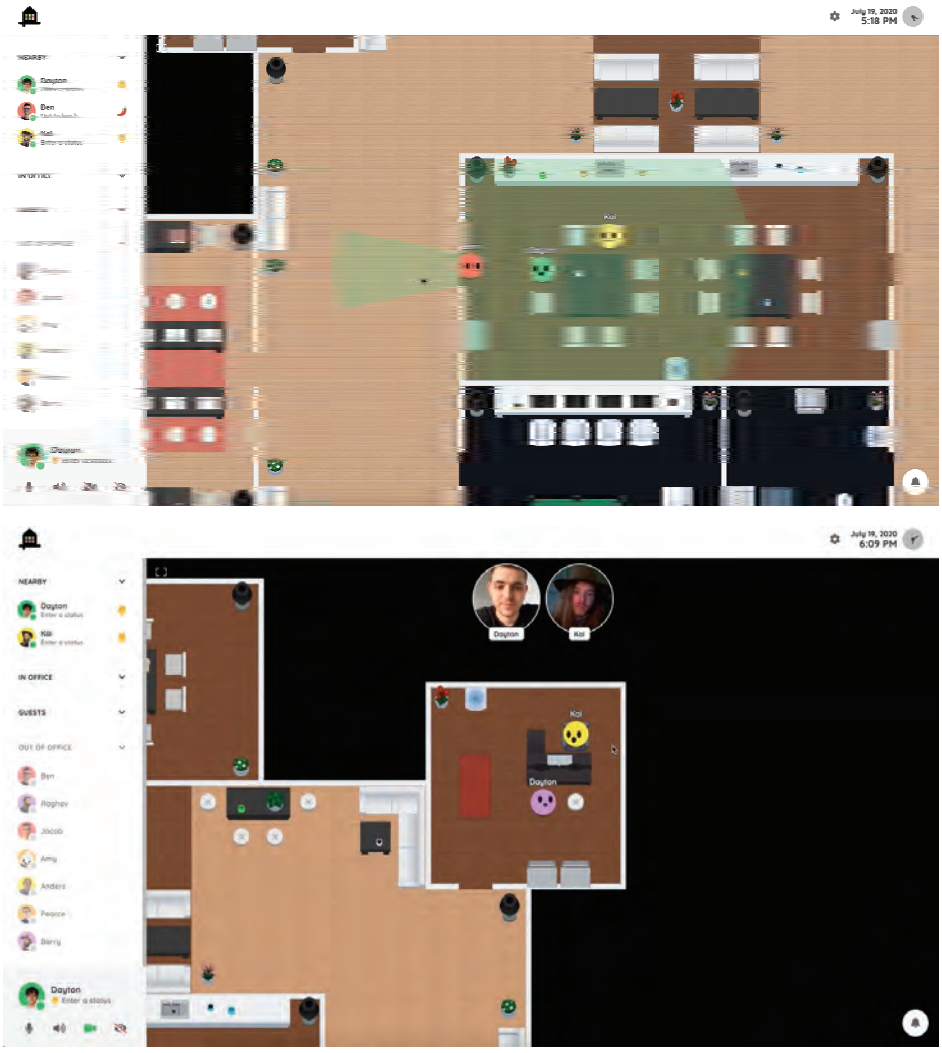
Na podstawie przeprowadzonych badań można założyć, że jest możliwe zbudowanie sieciowego systemu komunikacji dźwiękowej opartej na mechanizmach przeglądarkowych, ale żeby spełniały one warunek minimalnej latencji, realizacja wymaga dostępu do sieci bez ograniczeń.

### 5.5.2. Branch – System komunikacji w zasięgu pola widzenia

Doświadczenie zdobyte przez M. Walczaka (jednego z autorów tego rozdziału) przy realizacji przedstawionego w podrozdz. 5.1 rozwiązania stało się podstawą do podjęcia prac w ramach komercyjnego, międzynarodowego projektu Branch [4]. Rozwiązanie stanowi wirtualną przestrzeń biurową dostępną z poziomu przeglądarki internetowej, umożliwiającą komunikację audiowizualną między użytkownikami z wykorzystaniem technologii WebRTC.

Cechą wyróżniającą projekt jest sposób, w jaki rozwiązano komunikację między użytkownikami. Aplikacja pozwala na dynamiczne nawiązywanie dwustronnych połączeń *peer-to-peer* 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 – dynamiczne tworzenie grupowych spotkań w sposób odpowiadający temu, jak przebiegają w rzeczywistym życiu. Wrażenie rzeczywistej interakcji między uczestnikami grupy jest potęgowane przez uzależnienie głośności sygnału audio rozmówców od odległości między nimi. Dodatkową zaletą jest ograniczenie aktywnych połączeń WebRTC, co redukuje generowane przez nie obciążenie łącza sieciowego.

Testy z udziałem docelowych użytkowników wykazały, że wynikające z negocjacji połączenia WebRTC opóźnienie między wejściem w pole widzenia a nawiązaniem połączenia wpływało negatywnie na jakość użytkowania aplikacji. W zależności od takich czynników, jak jakość połączenia sieciowego, rodzaje sieci czy odległości między użytkownikami – czas negocjacji i nawiązywania połączenia *peer-to-peer* wahał się szacunkowo od 1 s przy połączeniu lokalnym do nawet 5 s w przypadku połączenia międzykontynentalnego. Opóźnienia były szczególnie odczuwalne w sytuacji nagłego nawiązania „kontaktu wzrokowego" z bliskiej odległości, np. przy „wyjściu zza rogu". Z tego względu w późniejszych wersjach aplikacji sposób zarządzania połączeniami uległ modyfikacji – zostało uzależnione od odległości między użytkownikami. Pole widzenia natomiast stało się odpowiedzialne za zarządzanie przesyłem strumieni audio i wideo. Ponieważ promień zasięgu utrzymywania połączenia jest większy od promienia pola widzenia, połączenia zostają nawiązywane i utrzymywane w tle. Zapewniają tym samym płynną komunikację między użytkownikami. Na rysunku 6 przedstawiono zrzuty ekranów różnych instancji aplikacji Branch.




Rys. 6. Przykładowe zrzuty ekranu przedstawiające formy komunikacji w systemie Branch [4]

## 5.6. Podsumowanie

W rozdziale opisano zagadnienia dotyczące generowania, edycji i transmisji dźwięku z wykorzystaniem specjalizowanych interfejsów programowania aplikacji API opartych na architekturze i protokołach sieci Web. Na tle rozwijających się systemów Networked Music Performance, NMP, służących do zdalnego wspólnego muzykowania

za pomocą sieci Internet, zapewniających minimalne opóźnienie. Zaprezentowano implementację aplikacji internetowej opartej na Web Audio API, WebRTC i Web MIDI API. Celem była próba realizacji systemu, który z poziomu przeglądarki internetowej zapewniłby możliwość komunikacji użytkowników za pomocą dźwięku w czasie rzeczywistym z minimalnym opóźnieniem. Dzięki aplikacji można ustanowić połączenie *peer-to-peer* między wieloma użytkownikami za pomocą technologii WebRTC. Wdrożono możliwość generowania, odtwarzania, modyfikacji, przechwytywania oraz przesyłania ścieżek audio drogą sieciową. Mimo że aplikacja została opracowana jako demonstracja możliwości opisanych technologii, ma potencjał rozwojowy. Na podstawie już zaimplementowanych technologii możliwe jest wdrożenie nowych funkcji z zakresu generowania i przetwarzania sygnału audio, komunikacji między użytkownikami, a także rozwinięcia już istniejących funkcji. Aplikacja może zostać rozbudowana m.in. o sekwencjonowanie ścieżek dźwiękowych, graficzną reprezentację cech strumieni audio w czasie rzeczywistym czy transmisję strumieni audio między wieloma użytkownikami za pomocą serwera mediów. Wtedy należy przeprowadzić badania jakości transmisji, w tym opóźnień systemowych.

Doświadczenie zdobyte przez jednego z autorów (M.W.) przy realizacji przedstawionego projektu stało się podstawą do podjęcia prac w ramach komercyjnego międzynarodowego projektu Branch [4] udostępniającego wirtualną przestrzeń biurową, w której konwersacja odbywa się w zasięgu pola widzenia.

Omówione technologie nie są jeszcze w pełni wspierane przez wszystkie nowoczesne przeglądarki internetowe. Ich standardy i rozwiązania wciąż rozwijają się i są uzupełniane o nowe funkcje i możliwości. Przykładem może być AudioWorklet, który przy współpracy z Web Audio API umożliwia przyspieszenie transmisji.

**Słowa kluczowe:** Networked Music Performance, Web Audio API, WebRTC API, strumieniowanie multimediów, *peer-to-peer*, komunikacja w czasie rzeczywistym.

## Bibliografia

[1] ARTSMESH; https://www.artsmesh.com/ [dostęp: 19.07.2021].

[2] BIGBLUEBUTTON; https://bigbluebutton.org [dostęp: 19.07.2021].

[3] *Background audio processing using AudioWorklet*; https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API/Using_AudioWorklet [dostęp: 19.07.2021].

[4] BRANCH; https://branch.gg/ [dostęp: 19.07.2021].

[5] Carôt A., *Musical Telepresence – A Comprehensive Analysis Towards New Cognitive and Technical Approaches*, PhD Thesis, Lübeck 2009.

[6] JackTrip; https://ccrma.stanford.edu/software/jacktrip/ [dostęp: 19.07.2021].
[7] JamKazam; https://jamkazam.com/ [dostęp: 19.07.2021].
[8] Jamulus; http://llcon.sourceforge.net/ [dostęp: 19.07.2021].
[9] LOLA; https://lola.conts.it/ [dostęp: 19.07.2021].
[10] Loreto S., Romano S.P., *Real-Time Communication with WebRTC: Peer-to-Peer in the Browser*, O'Reilly Media, Inc., Sebastopol 2014.
[11] 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.
[12] Rottondi C., Chafe Ch., Allocchio C., Sarti A., *An Overview on Networked Music Performace Technologies*, IEEE Access 4: 8823-8843 2016.
[13] Sergiienko A., *WebRTC Blueprints*, Packt Publishing, Birmingham 2014.
[14] SonoBus; https://sonobus.net/ [dostęp: 19.07.2021].
[15] Smus B., *Web audio API*, O'Reilly Media, Inc, Sebastopol 2013.
[16] SOUNDJACK; https://www.soundjack.eu/ [dostęp: 19.07.2021].
[17] Tone.js; https://tonejs.github.io/ [dostęp:19.07.2021].
[18] *WebRTC, Real Time Communication for the Web*; https://webrtc.org [dostęp: 19.07.2021].
[19] *Web Audio API. W3C Candidate Recommendation, 11 June 2020*; https://www.w3.org/TR/webaudio/ [dostęp: 19.07.2021].
[20] *WebAudioRecorder.js*.; https://github.com/higuma/web-audio-recorder-js [dostęp: 19.07.2021].
[21] *Web MIDI API. W3C Working Draft*, March 2015; https://www.w3.org/TR/2015/WD-webmidi-20150317 [dostęp: 19.07.2021].
[22] *WebMidi.js*; https://github.com/djipco/webmidi [dostęp: 19.07.2021].
[23] *Worklet*; https://developer.mozilla.org/en-US/docs/Web/API/Worklet [dostęp: 19.07.2021].

# 6. System sterowania kolumny z niezależnie sterowanymi głośnikami

MICHAŁ ŁUCZYŃSKI

Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

W rozdziale zaprezentowano propozycję sterowania kolumny głośnikowej wyposażonej w niezależnie sterowane przetworniki. Jest ona elementem stanowiska dydaktycznego przeznaczonego do demonstracji zasady działania sterowanych cyfrowo kolumn głośnikowych. Stanowisko składa się z kolumny wyposażonej w osiem przetworników ze wzmacniaczami, zestawu mikrofonów pomiarowych, zewnętrznej karty dźwiękowej i oprogramowania. Do sterowania można wykorzystać zarówno komputerowy system edycji dźwięku (ang. Digital Audio Workstation; DAW) obsługujący pracę z wieloma wyjściami, jak i środowisko programistyczne przeznaczone do przetwarzania sygnałów (np. PureData i MaxMSP). Przez ustawianie odpowiednich poziomów sygnału oraz opóźnień dla poszczególnych kanałów można dokonywać zmiany kierunkowości źródła. Analiza zmian kierunkowości może być przeprowadzana słuchowo lub w trakcie obserwacji zmian poziomów ciśnienia akustycznego rejestrowanego przez mikrofony pomiarowe. W pracy przedstawiono konfigurację systemu z użyciem programu DAW i specjalnego oprogramowania oraz przykładowe wyniki symulacji i pomiarów.

## 6.1. Wprowadzenie

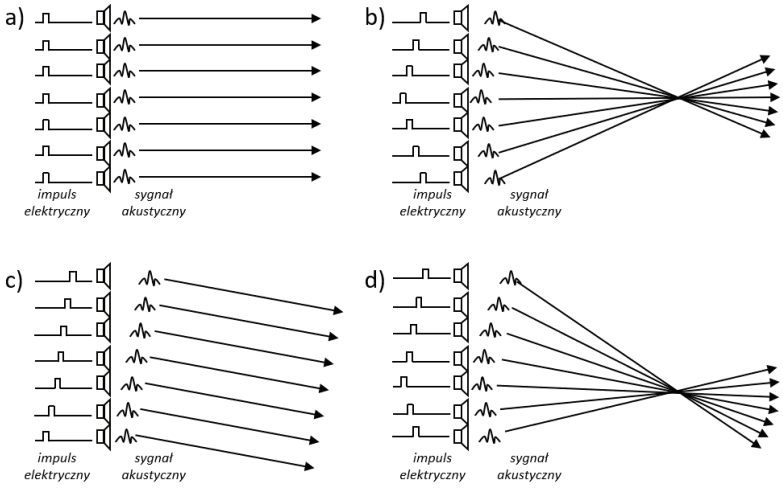
Kolumna głośnikowa jest specyficznym rozwiązaniem wśród urządzeń głośnikowych i o bardzo nietypowych właściwościach. Jednocześnie termin często jest mylnie

przypisywany zestawom głośnikowym. W literaturze przedmiotu stanowiącej podstawowe źródło wiedzy w dziedzinie [1, 2] poza zdefiniowaniem terminu: „kolumna głośnikowa" przedstawiony jest opis matematyczny umożliwiający obliczenie ich parametrów akustycznych. Kolumnę głośnikową tworzy kilka lub kilkanaście jednakowych głośników, które są umieszczone w jednej linii i najczęściej 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 zależy ona od częstotliwości, a dokładnie od stosunku odległości między osiami przetworników a długością fali oraz liczby przetworników tworzących kolumnę. Na charakterystykę kierunkowości kolumny można wpłynąć dodatkowo przez indywidualne sterowanie każdym z przetworników. Sygnały sterujące różnią się wzmocnieniem i biegunowością (przesunięcia fazowe). Głównym zadaniem jest sztuczne uzyskanie efektu pochylenia kolumny głośnikowej. Propozycję kolumny głośnikowej ze zmiennym kątem pokrycia zaprezentowano w 1962 r. [3]. W konsekwencji podjętych badań zmniejszano sprzężenia akustyczne wówczas, gdy kolumna charakteryzowała się większą kierunkowością. Ponadto zwrócono uwagę, że są to rozwiązania kosztowne. W latach 80. XX wieku przedstawiono analizy zastosowania kolumn głośnikowych w celu nagłośnienia w kościołach [4]. Celem była poprawa zrozumiałości mowy. Rozwój technologii w zakresie procesorów sygnałowych umożliwił stosowanie kolumn sterowanych cyfrowo. Przykładem takiego zastosowania jest system nagłośnienia na hali odlotów na lotnisku [5]. Dzięki zastosowaniu zmiany kierunku propagacji uzyskano w przybliżeniu stały stosunek sygnału bezpośredniego do pogłosowego w zasięgu do 30 m. Obecnie większość produktów komercyjnych zapewnia uzyskanie zbliżonych parametrów akustycznych [6]. Poza tym za pomocą metody DGRC (Digital and Gometric Radiation Control) możliwe jest przypisanie dużej liczby głośników kolumny głośnikowej do ograniczonej liczby kanałów elektronicznych [7, 8].

Oczywistą wadą urządzeń głośnikowych opartych na jednym rodzaju przetwornika jest ograniczenie pasma częstotliwości wynikające z trudności w efektywnym pobudzeniu całego pasma akustycznego przez pojedynczy głośnik. Wśród współczesnych rozwiązań jest zastosowanie dodatkowego głośnika niskotonowego [9] mogącego stanowić jedno modułowe urządzenie. Dzięki temu możliwe jest uzyskanie większego użytecznego zakresu częstotliwości. Należy jednak pamiętać, że takie rozwiązanie wykazuje cechy kolumny tylko w takim zakresie częstotliwości, w jakim pracują głośniki tworzące szereg jednakowych przetworników. Mimo wszystko urządzenia składa-

jące się z wielu jednakowych przetworników (czyli listwy dźwiękowe; ang. *soundbar*), w których można niezależnie sterować pracą każdego z nich, są coraz bardziej popularne. W tym przypadku stosowane są dużo bardziej skomplikowane algorytmy przetwarzania sygnałów niż w typowych kolumnach głośnikowych – sterowanie odbywa się tu głównie w celu zmiany kierunku promieniowania dźwięku.

Należy również wspomnieć o tym, że kształtowanie sygnału za pomocą odpowiedniego opóźniania i wzmacniania sygnałów sterujących wykorzystywane jest również w innych dziedzinach akustyki. W technice ultradźwiękowej stosuje się takie przetwarzanie w celu uzyskania odpowiedniego ogniskowania wiązki ultradźwiękowej [10] – testowanie takich rozwiązań w zakresie częstotliwości słyszalnych może pomóc zrozumieć omawiane zagadnienie studentom. Na rysunku 1 przedstawiono w sposób schematyczny cztery rodzaje kształtowania wiązki ultradźwiękowej [11]: wiązkę synchroniczną, wiązkę skoncentrowaną, wiązkę sterowaną oraz wiązkę skoncentrowaną i sterowaną.




Rys. 1. Schematyczne przedstawienie kształtowania wiązki: a) wiązka synchroniczna, b) wiązka skoncentrowana, c) wiązka sterowana, d) wiązka skoncentrowana i sterowana

Urządzenie składające się z wielu jednakowych przetworników, które w łatwy sposób mogą być zasilane niezależnymi sygnałami sterującymi, może być z powodzeniem stosowane w dydaktyce. Podstawowe zastosowania związane są z parametrami kolumn głośnikowych, czyli wpływem jej długości na kierunkowość i zmiany kierunku propagacji. Wykorzystanie omawianego dalej stanowiska umożliwi wykonywanie pomiarów i realizowanie testów odsłuchowych dla kolumny o różnej długości. Za pomocą wycisza-

nia odpowiednich sygnałów sterujących można wyciszać poszczególne głośniki. Dzięki temu jest możliwa zarówno zmiana długości kolumny skutkiem wyłączenia źródeł skrajnych, jak i porównanie kolumn głośnikowych zbudowanych z tej samej liczby przetworników, ale rozmieszczonych w różnych odległościach. Zrealizowanie takiego porównania sprowadza się do podania sygnałów sterujących na przykład do co drugiego przetwornika. Jak już zostało wcześniej wspomniane, za pomocą ustawienia odpowiednich opóźnień i wzmocnień można spowodować zmianę kierunku propagacji dźwięku. Studenci mogą mieć okazję do zaznajomienia się, jakie rzędy wartości umożliwiają uzyskanie takiego sterowania. Ta wiedza jest przydatna w kontekście sterowania liniowymi systemami nagłaśniania. Dzięki przedstawionemu układowi przetworników można również zaprezentować zasadę działania ogniskowania fali ultradźwiękowej realizowanej dla wielu przetworników. Analogia opierająca się na przeniesieniu rozpatrywanego zjawiska do pasma akustycznego pomoże w zrozumieniu zagadnienia.

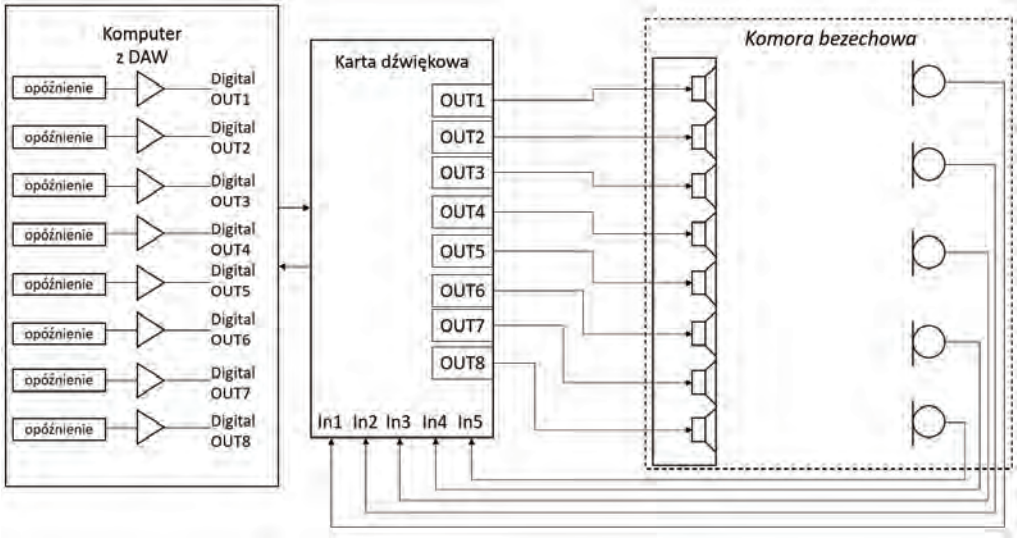
Celem tej pracy jest przedstawienie propozycji nietrudnego w budowie i obsłudze stanowiska laboratoryjnego, które umożliwi zapoznanie się z układami kolumn głośnikowych w praktyce. Dzięki temu rozwiązaniu można nie tylko załączać różną liczbę przetworników i niezależnie sterować poszczególnymi głośnikami, lecz także w prosty sposób rozbudować konstrukcję o kolejne moduły. Konstrukcja i system sterowania powstały przede wszystkim na potrzeby dydaktyki. Kolejne moduły można rozbudowywać oraz tworzyć i dłuższe kolumny głośnikowe, i dwuwymiarowe matryce głośnikowe.

## 6.2. Opis stanowiska

Podstawowe założenia projektowe dotyczące utworzenia stanowiska laboratoryjnego to możliwie prosta i tania konstrukcja nieskomplikowana w obsłudze z przeznaczeniem dla studentów kształcących się na kierunkach związanych z akustyką. W dalszej części rozdziału przedstawiono koncepcję budowy stanowiska wraz z przykładowymi symulacjami i propozycją pomiarów i eksperymentów, które można przeprowadzić z jego użyciem.

System składa się z kolumny głośnikowej wyposażonej w osiem przetworników 2,5-calowych i osiem wzmacniaczy 10-watowych, 8-kanałową zewnętrzną kartę dźwiękową, komputer z oprogramowaniem DAW oraz okablowanie. Konfiguracja systemu polega na przygotowaniu ośmiu ścieżek w programie DAW i przypisaniu ich do odpo-

wiednich wyjść. Następnie wyjścia sygnałowe należy podłączyć do wejść wzmacniaczy zasilających przetworniki. Schemat blokowy proponowanego systemu przedstawiono na rys. 2.




Rys. 2. Schemat blokowy systemu sterowania cyfrową kolumną głośnikową

Zastosowane wzmacniacze to gotowe układy scalone zasilane napięciem stałym. Wszystkie przetworniki znajdują się w jednej obudowie. Odległość między osiami przetworników jest stała i wynosi 7,5 cm. Całkowity wymiar kolumny to 60 cm. Odległość między osiami skrajnych przetworników a krawędzią obudowy stanowi połowę odległości między osiami przetworników. Układ można rozbudować przez dołożenie dodatkowego urządzenia składającego się z kolejnych ośmiu przetworników – musi pozostawać przy tym stała odległość między osiami przetworników. W ten sposób można utworzyć dwuwymiarową macierz głośnikową. A w przypadku zestawienia ze sobą $n$ kolumn głośnikowych – macierz o wymiarach $n \times 8$.

Karta dźwiękowa wykorzystana do systemu to U-PHORIA UMC1820 [12] – interfejs mający osiem wejść typu combo (wejścia mikrofonowe/liniowe), dziesięć wyjść liniowych i dwa wyjścia słuchawkowe. Do tego jest możliwość rozszerzenia liczby wejść i wyjść za pomocą protokołu ADAT. Interfejs jest podłączany do komputera z wykorzystaniem przewodu USB. Używanym sterownikiem jest ASIO [13], a wybrany program do obróbki dźwięku to Reaper [14]. Propozycja specjalnego oprogramowania została przygotowana w środowisku programistycznym Pure Data [15].

## 6.3. System sterowania

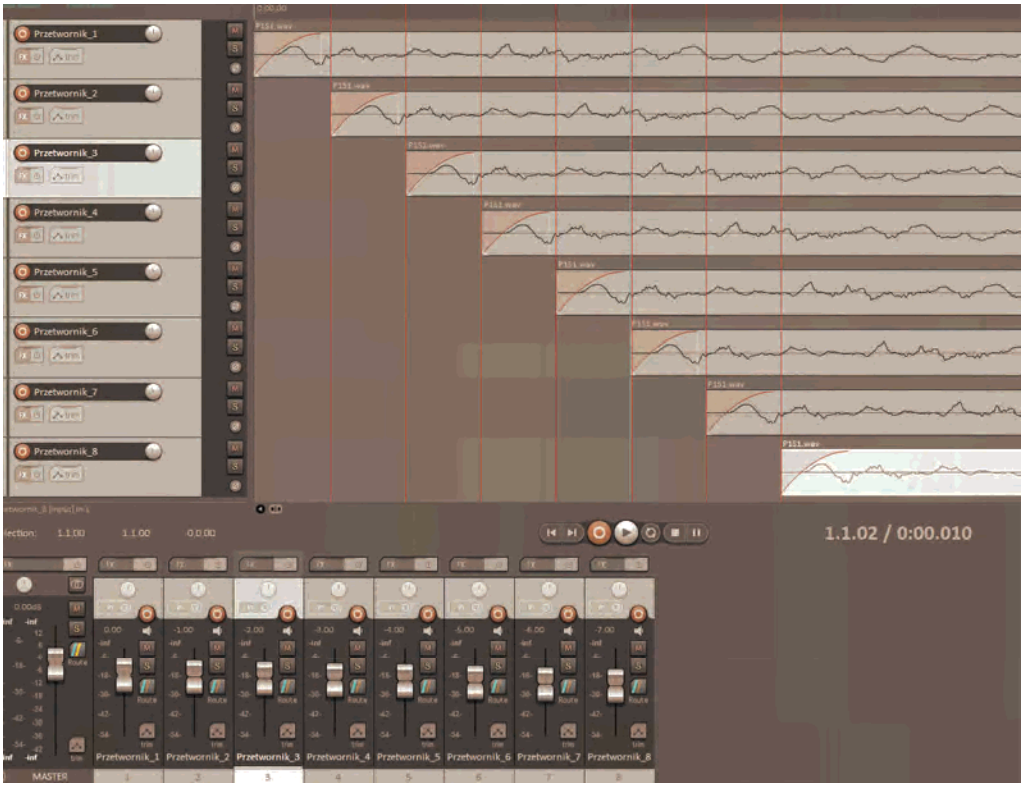
Sterowanie kolumną odbywa się z użyciem specjalnie przygotowanych sygnałów – konkretnych dla każdego z przetworników w kolumnie. Taki system powinien w swojej podstawowej formie umożliwiać realizację opóźnień i modyfikację wzmocnienia sygnału. Właśnie te dwa parametry są kluczowe do uzyskania odpowiedniej kierunkowości źródła w takim stopniu, w jakim będzie możliwe spełnienie wymagań związanych z realizacją zadań dydaktycznych. Następnie rozważono wprowadzanie płynnej zmiany tych parametrów. Tym sposobem można uzyskać efekt zmiany charakterystyki kierunkowości źródła w trakcie odsłuchu. A to podstawa do uzyskania efektu skoncentrowania wiązki dźwiękowej w punkcie położenia pozycji odsłuchowej. Ponadto system sterowania powinien stwarzać możliwość realizacji innego typu efektów takich jak filtracja częstotliwościowa.

### 6.3.1. Oprogramowanie DAW

Pracę należy rozpocząć od otworzenia nowej sesji, wybrania odpowiednich parametrów sesji, utworzenia tylu ścieżek, ile zaplanowano wejść i wyjść sygnałowych i konfiguracji połączeń wejść/wyjść. Każdemu przetwornikowi odpowiada osobna ścieżka. Takie rozwiązanie stwarza możliwość prostego ustawienia wzmocnień za pomocą suwaków odpowiadającym ścieżkom oraz użycie dodatkowych procesorów efektowych niezależnie stosowanych do każdego sygnału sterującego przetwornik. Opóźnienie może być ustawiane dzięki ręcznemu przesunięciu fragmentu materiału na ścieżce. A porównanie konfiguracji opóźnień realizujących różne kierunkowości kolumny – przez wstawienie w pliku sesji kolejnych sygnałów zasilających jeden po drugim. Analizę zmian poziomu ciśnienia akustycznego w zależności od konfiguracji wzmocnień i opóźnień sygnałów można przeprowadzić dzięki rejestracji sygnału za pomocą mikrofonów. Należy wtedy skorzystać z dodatkowych ścieżek i skonfigurować połączenia z odpowiednimi kanałami wejściowymi zewnętrznej karty dźwiękowej.

Na rysunku 3 przedstawiono przykładową konfigurację sygnałów sterujących przetworniki. Każdy kolejny przetwornik jest w tym przypadku zasilany sygnałem o poziomie 1 dB mniejszym, a opóźnienie na każdym z przetworniku jest o 2 ms większe.

Zaletą realizacji opóźnienia na skutek odpowiedniego ustawienia odtwarzanego regionu na osi czasu jest łatwe do wykonania i natychmiastowo widać zadane opóźnienie. Metoda ta jednak jest nieprecyzyjna i mało powtarzalna.



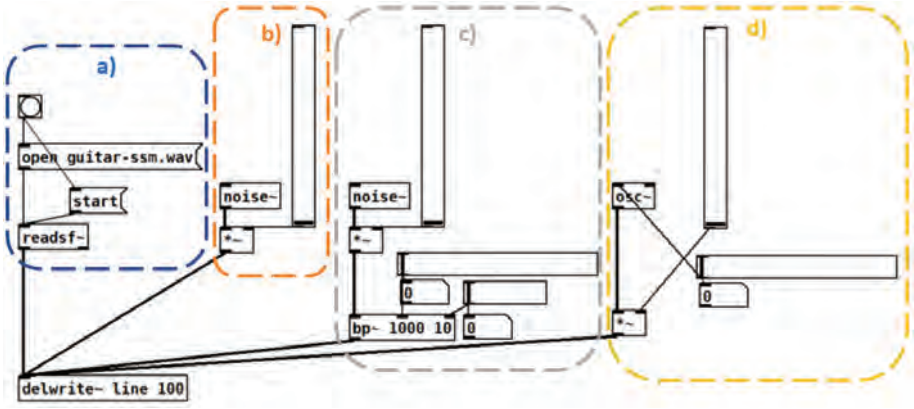

Rys. 3. Przykładowa konfiguracja opóźnień i wzmocnień z poziomu programu DAW

Drugim rozwiązaniem ustawienia opóźnień jest zastosowanie odpowiednich wtyczek VST. Na przykład można użyć wtyczki Time Adjustment Delay, która jest dostępna z podstawową wersją programu Reaper. Za jej pomocą wartość opóźnienia można ustawić suwakiem w zakresie –1000–1000 ms. W kolumnach głośnikowych stosuje się mniejsze wartości – maksymalnie rzędu pojedynczych milisekund. W celu zapewnienia późniejszej wygody wtyczkę można zmodyfikować. W linii kodu, która definiuje działanie suwaka, należy zmienić fragment: „[…] **slider1:0<-1000,1000,1>Delay Amount (ms) na slider1:0 < 0,10,0.1>Delay Amount (ms)**". W efekcie suwak będzie pracował od 0 ms do 10 ms z krokiem co 0,1 ms. Jeżeli na jakimś etapie pracy będzie potrzebny inny zakres lub rozdzielczość, należy rozważyć ponowną modyfikację lub wpisanie wartości ręcznie. Zastosowanie tej metody konfiguracji opóźnień dla poszczególnych ścieżek reprezentujących przetworniki skutkuje nie tylko możliwością bardziej precyzyjnego ustawiania wartości opóźnień, lecz także płynniejszego ich modyfikowania – a jest to możliwe dzięki zastosowaniu automatyki.

### 6.3.2. Środowisko Pure Data

Druga propozycja systemu sterowania bazuje na zastosowaniu graficznego środowiska programowania Pure Data przeznaczonego do przetwarzania sygnałów. Pure Data jest darmowym środowiskiem dostępnym na komputerach z systemami operacyjnymi Windows, MacOS i Linux. Funkcje i zmienne są reprezentowane przez obiekty graficzne, które są ze sobą łączone. Każdy obiekt wykonuje określone działanie od prostych operacji matematycznych, przez generację sygnałów (tony proste, szum) do transformacji FFT czy procesorów pogłosowych. Utworzenie prostych programów nie wymaga większej wiedzy programistycznej, dzięki czemu środowisko Pure Data bardzo dobrze sprawdza się jako narzędzie dydaktyczne i prezentacyjne różnego typu algorytmów przetwarzania sygnałów audio.

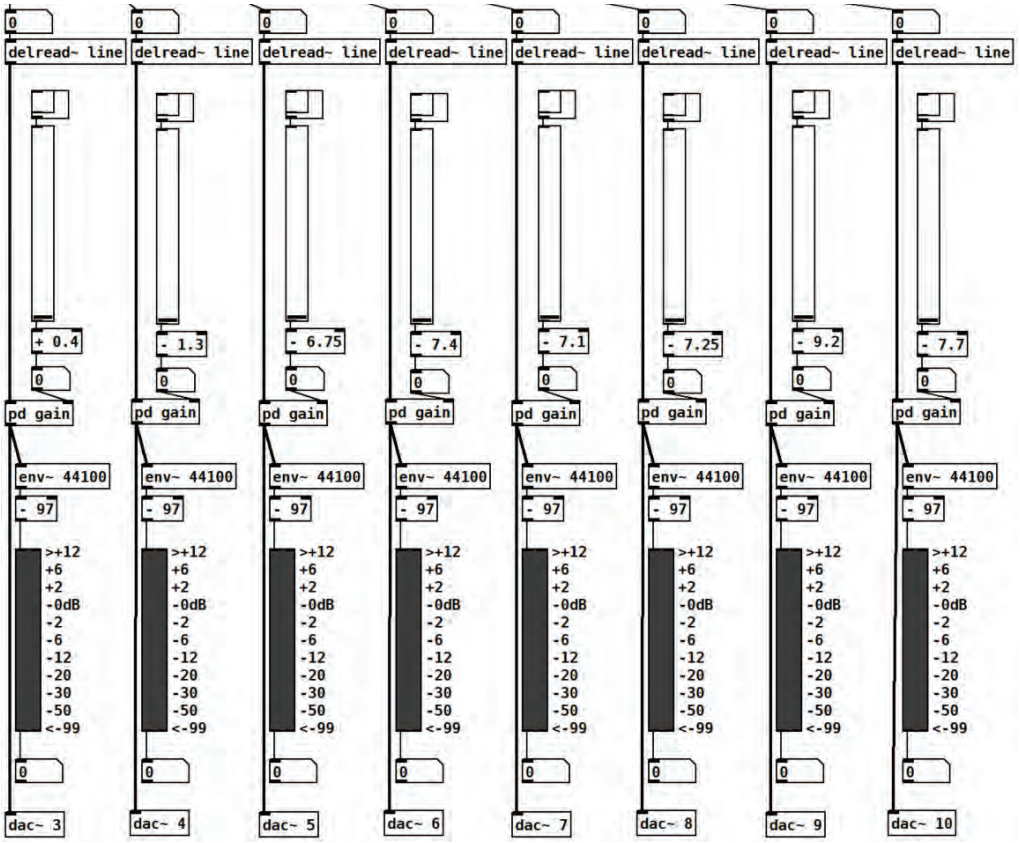
W ramach systemu sterowania przygotowano kilka różnych typów sygnałów wejściowych. Na wyjściu każdego generatora sygnału podłączono regulację wzmocnienia. Jako sygnał wejściowy można wybrać dźwięk z pliku .wav, szum biały, szum biały filtrowany filtrem pasmowo-przepustowym o definiowanej częstotliwości środkowej i dobroci oraz ton prosty. Sygnały można łączyć dzięki odpowiedniej regulacji poziomu wzmocnienia. Kod programu przedstawiający sygnały wejściowe przedstawiono na rys. 4.




Rys. 4. Kod programu do generowania sygnałów wejściowych podawanych na linię opóźniającą: a) odczyt z pliku .wav, b) szum biały, c) filtrowany szum biały o modyfikowanej dobroci i częstotliwości środkowej filtru, d) ton prosty o modyfikowanej częstotliwości

Wyjście sygnałowe każdego typu sygnału pobudzającego podłączono do funkcji realizującej linię opóźniającą (*delwrite~*). Jest to funkcja przechowująca w pamięci liczbę

próbek zależną od zdefiniowanej w preferencjach programu szybkości próbkowania oraz czasu wyrażonego w milisekundach. W przykładzie przedstawionym na rys. 5 ustawiono długości linii opóźniającej na 100 ms, co przy częstotliwości próbkowania 44 100 próbek na sekundę odpowiada liczbie 4410 próbek. Następnie sygnał jest odczytywany z linii opóźniającej tyle razy, ile zdefiniowano kanałów wyjściowych. W tym celu zastosowano funkcję *delread~*. Następnie wyjście funkcji pobierającej sygnał z linii opóźniającej przekazane jest na tor realizujący wzmocnienie sygnału – tor wyposażony jest w tłumik, możliwość skokowego wyciszania sygnału i pomiar wysterowania sygnału. W prezentowanym przykładzie wyznaczana jest wartość RMS sygnału ze stałą czasową SLOW. Można ją modyfikować przez zmianę argumentu funkcji *env~*. Na koniec sygnał jest wysyłany na wyjście karty dźwiękowej. Służy do tego funkcja *dac~*, której argumentem jest numer kanału wyjściowego zewnętrznej karty dźwiękowej. W prezen-

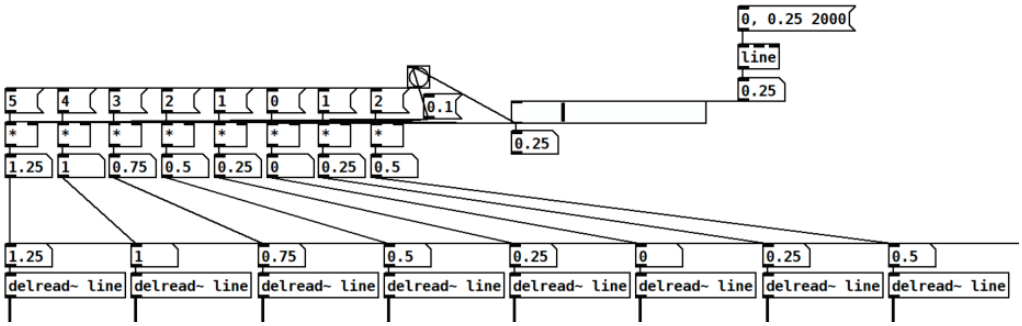



Rys. 5. Kod programu przedstawiający tory przetwarzania sygnałów sterujących

towanym przykładzie zastosowano wyjścia o numerach 3–10. Kod programu przedstawiający poszczególne tory sygnałów wyjściowych przedstawiono na rys. 5.

Zamierzoną konfigurację opóźnień można regulować przez doprowadzenie odpowiednich wartości do funkcji odczytujących sygnał z linii opóźniających (*delread~*). Przykład funkcji automatycznie przeliczającej wartości opóźnienia do każdego przetwornika w celu uzyskania koncentracji wiązki poza osią kolumny przedstawiono na rys. 6 – w tym przypadku wykorzystano suwak umożliwiający zmianę wartości opóźnień w sposób płynny. Ponadto dzięki funkcji *line* umożliwiającej płynną zmianę wartości na jej wyjściu można automatycznie sterować wartością na suwaku. W zaprezentowanym przykładzie wartość ustawiana przez suwak zmieniana jest 0–0,25 ms w czasie 2 s.

Zmiana wzmocnienia sygnału może być realizowana za pomocą zmiany ustawień suwaków przedstawionych na rys. 5. Wartości suwaków mogą być automatycznie sterowane w sposób analogiczny do przedstawionego w przykładzie sterowania opóźnieniem. Automatyczne sterowanie zarówno opóźnieniem, jak i wzmocnieniem można również zrealizować niezależnie dla każdego przetwornika lub grup przetworników. W ten sposób można zrealizować zaplanowane scenariusze zmian kierunkowości źródła.




Rys. 6. Automatyczne przeliczanie wartości opóźnienia sygnałów sterujących

## 6.4. Symulacje

Skutecznym wspomaganiem systemu sterowania może być przeprowadzanie symulacji rozkładu poziomu ciśnienia akustycznego generowanego przez analizowaną kolumnę głośnikową. Takie symulacje z jednej strony pomagają dobrać odpowiednie

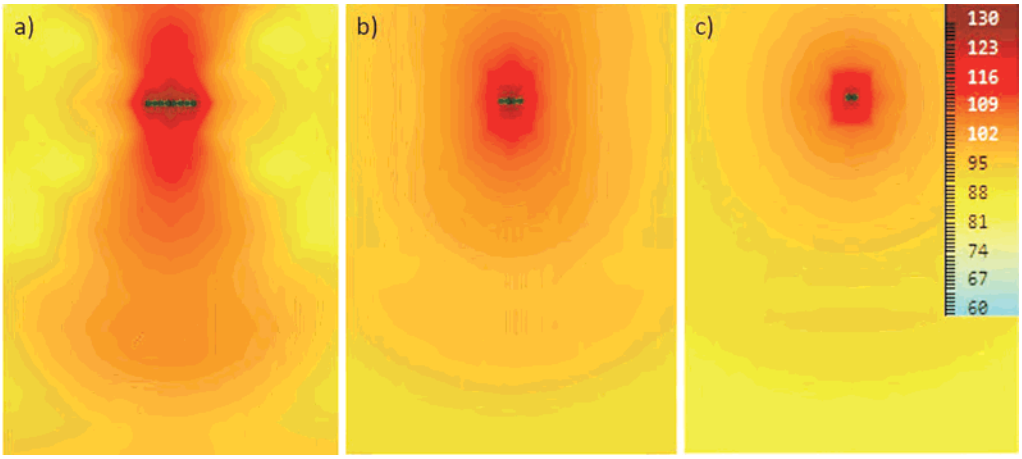
wartości opóźnień i wzmocnień dla poszczególnych przetworników w celu uzyskania założonego efektu, a z drugiej wykonać analizę działania kolumny w innych warunkach akustycznych niż te dostępne w ramach ćwiczeń laboratoryjnych lub w przypadku innej konstrukcji kolumny głośnikowej (np. przy innej odległości między głośnikami). Symulacje stwarzają możliwość znaczącego rozszerzenia możliwości dydaktycznych. Tym bardziej należy dążyć do tego, aby ich przeprowadzenie było jak najprostsze.

W ramach pracy wybrano środowisko symulacyjne Acoustic Boundary Element Calculator (ABEC) [16]. Jest to środowisko obliczeniowe wykorzystujące Metodę Elementów Brzegowych i umożliwiające bardzo dokładne modelowanie pola akustycznego. Zastosowanie znajduje w obliczeniach związanych z urządzeniami głośnikowymi, propagacji od źródeł hałasu, a także w obliczeniach związanych z analizą rozkładu pola akustycznego w pomieszczeniach w zakresie małych częstotliwości oraz wszędzie tam, gdzie należy uwzględniać zjawiska falowe w powietrzu.

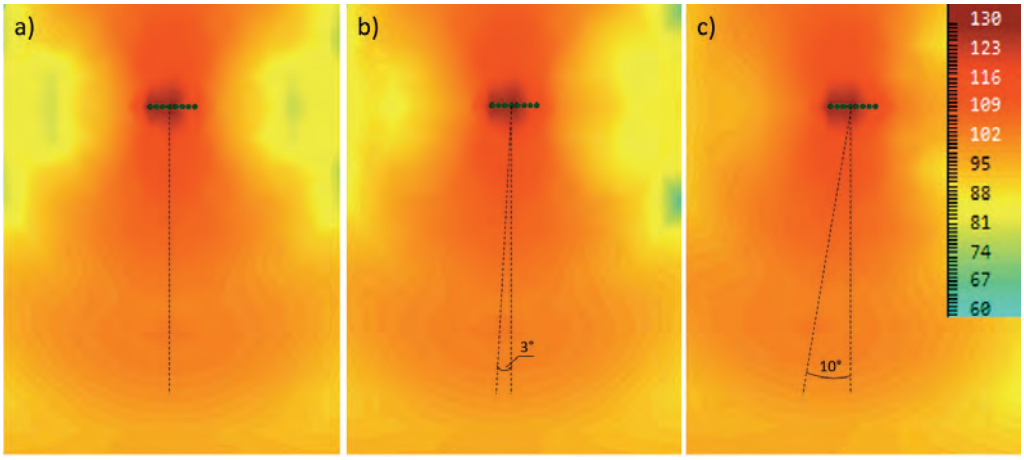
W przedstawionych przykładach symulację przeprowadzono w polu swobodnym o zdefiniowanym polu obliczeniowym mającym wymiar 3 m × 4 m. Nie wykonywano modelowania charakterystyki promieniowania głośników użytych w konstrukcji kolumny głośnikowej. Przyjęto, że źródła dźwięku są źródłami punktowymi. Użytym sygnałem był ton prosty o częstotliwości 1 kHz. Symulację wykonano dla różnych długości kolumny głośnikowej, różnych konfiguracji parametrów źródeł dźwięku i dla kilku częstotliwości. Na rysunku 7 przedstawiono symulacje rozkładu poziomu ciśnienia akustycznego w przypadku trzech różnych długości kolumny głośnikowej, a na rys. 8 – symulację przeprowadzoną, gdy poszczególne źródła są sterowane sygnałami z różnymi opóźnieniami. W symulacji wykorzystano siedem przetworników. Opóźnienia różnicowano względem przetwornika środkowego. Celem symulacji powinno być znalezienie takich wartości granicznych opóźnień, wzmocnień i częstotliwości, przy których kolumna wykazuje pożądane właściwości.

Można zaobserwować, że wraz ze zwiększeniem liczby przetworników składających się na kolumnę głośnikową zwiększa się jej kierunkowość. Ponadto wyniki symulacji przedstawione na rys. 8 potwierdzają uzyskanie odchylenia kierunku propagacji dźwięku względem osi kolumny.

Efektem symulacji może być zarówno rozkład poziomu ciśnienia akustycznego (przedstawiony w tym rozdziale), jak i poziom ciśnienia akustycznego w wyznaczonych punktach pomiarowych. A to może stanowić podstawę do bezpośredniego porównania wartości uzyskanych w symulacjach oraz wartości otrzymanych drogą pomiarową.




Rys. 7. Wyniki symulacji rozkładu poziomu ciśnienia akustycznego dla kolumny o różnej długości: a) osiem przetworników, b) cztery przetworniki, c) dwa przetworniki; skala podana w dB

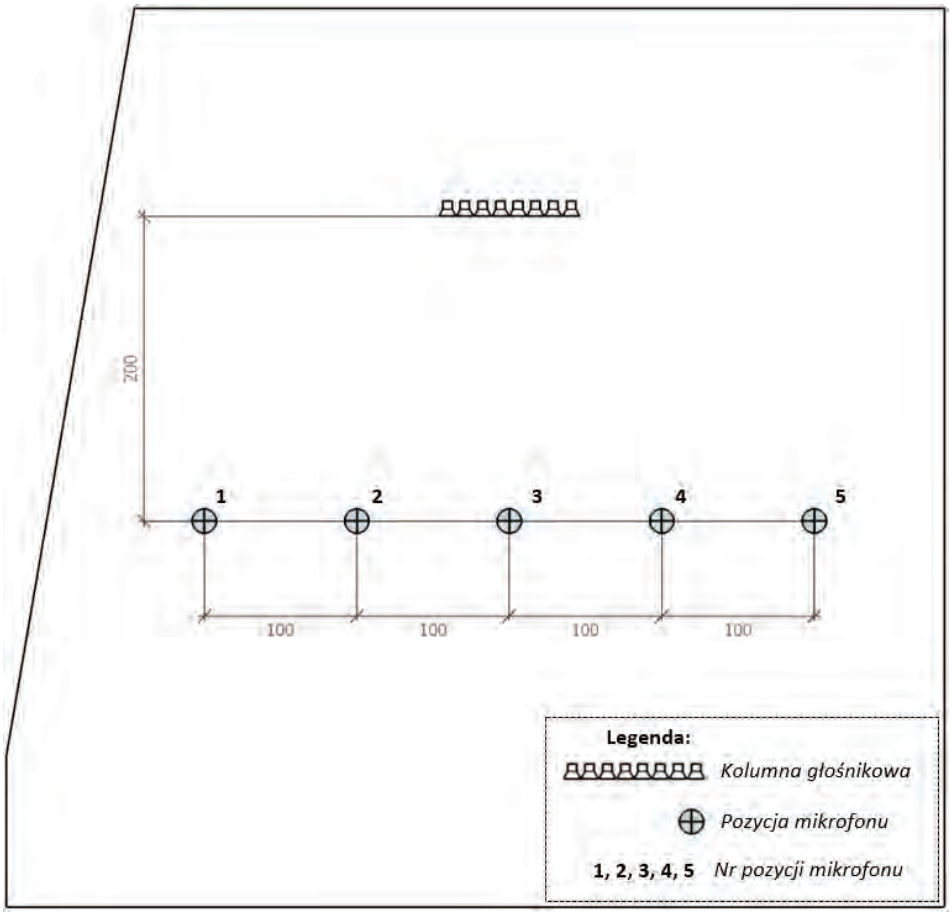



Rys. 8. Wyniki symulacji rozkładu poziomu ciśnienia akustycznego dla różnych opóźnień sygnałów sterujących: a) bez opóźnień, b) zmiana opóźnienia o 0,01 ms, c) zmiana opóźnienia o 0,03 ms; skala podana w dB

## 6.5. Pomiary

W ramach pracy przeprowadzono pomiary poziomu ciśnienia akustycznego pochodzącego od kolumny głośnikowej dla różnych konfiguracji parametrów sygnałów

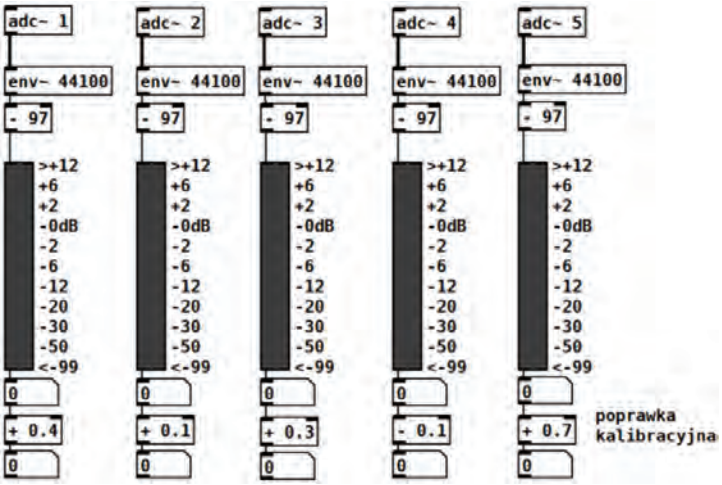
sterujących. Wykonano je w wyznaczonych punktach pomiarowych za pomocą mikrofonów pomiarowych ECM999 znajdujących się na wysokości osi membran głośników. Rozkład punktów pomiarowych przedstawiono na rys. 9. Do tego celu zastosowano system sterowania przygotowany w środowisku Pure Data. Sygnał sterujący to szum oktawowy o częstotliwości środkowej 1 kHz. Pomiary zostały przeprowadzone w pomieszczeniu zaadaptowanym akustycznie, w którym czas pogłosu dla pasma oktawowego 500 Hz wynosi 0,5 s. Chodziło o weryfikację, czy jest możliwe zaobserwowanie spodziewanych zjawisk w polu akustycznym, które nie jest polem swobodnym. Takie pomiary umożliwiają omówienie na zajęciach dydaktycznych zagadnienia dotyczącego zastosowania takich urządzeń w pomieszczeniach zwyczajnych.




Rys. 9. Rozmieszczenie punktów pomiarowych w badanym pomieszczeniu – rzut z góry (wymiary podane w centymetrach)

Przed przystąpieniem do pomiarów należy wykonać kalibrację torów elektroakustycznych – i tych służących do generacji sygnału, i tych do jego pomiaru. Niekontrolowane różnice poziomów ciśnienia akustycznego generowanych przez poszczególne przetworniki mogą doprowadzić do braku uzyskania efektu pożądanej charakterystyki kierunkowej kolumny głośnikowej. Kalibracja torów pomiarowych umożliwia odczytanie poziomu ciśnienia akustycznego generowanego przez urządzenia w zakresie akceptowalnej niepewności pomiarowej.

Kalibracja torów pomiarowych (mikrofonów, przedwzmacniacza, systemu akwizycji danych) może się odbyć przy użyciu kalibratora dźwięku lub metodą porównawczą – przez generowanie dźwięku za pomocą jednego przetwornika i umieszczenie mikrofonów w bliskiej lokalizacji względem siebie. Przy spełnieniu odpowiednich warunków można założyć, że z pewnym przybliżeniem parametry pola akustycznego oddziałującego na mikrofony są dla każdego z nich jednakowe. Podczas kalibracji zarówno metodą porównawczą, jak i z zastosowaniem kalibratora akustycznego należy uzyskać jednakowy poziom sygnału na kanałach wejściowych systemu sterowania. Do kontroli poziomów sygnału wejściowego służy kod programu przedstawiony na rys. 10.



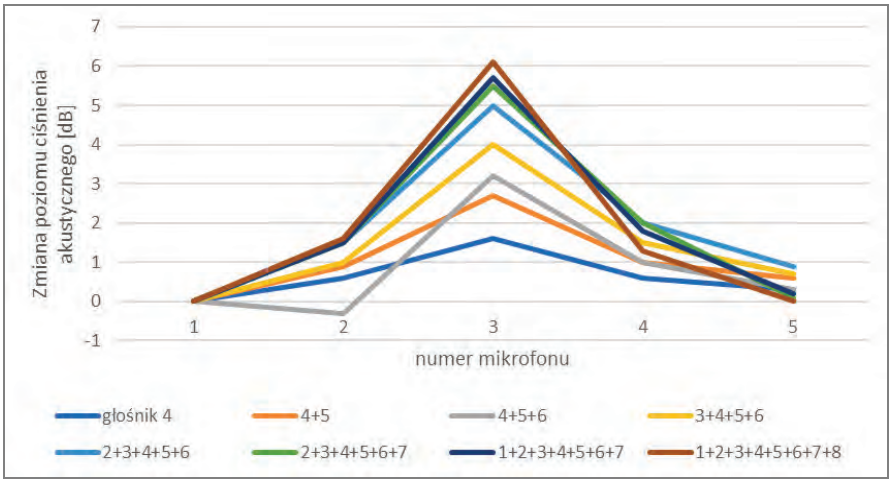

Rys. 10. Kod programu przedstawiający bloki przetwarzania sygnału wejściowego (pomiarowego)

Numer wejścia zewnętrznej karty dźwiękowej jest argumentem funkcji *adc~*. Następnie obliczana jest wartość skuteczna sygnału – w opisywanym przykładzie realizowana ze stałą czasową SLOW. Wartość sygnału przedstawi się graficznie za pomocą miernika wysterowania oraz liczbowo. Zgrubna nastawa czułości wejścia może być

zrealizowana dzięki ustawieniu poziomu wzmocnienia przedwzmacniacza mikrofonowego. W celu dokładnego wysterowania stosowana jest poprawka kalibracyjna (jak na rysunku). Nie ma potrzeby, aby zmierzone wartości odpowiadały faktycznym wartościom zgodnym z jednostką dB (SPL). W przypadku tych pomiarów interesująca jest różnica poziomów zarejestrowanych przez poszczególne mikrofony.

Kalibrację sygnałów wyjściowych, tj. sygnałów sterujących przetworniki, wykonuje się przez generowanie sygnału poszczególnych głośników i wykonanie pomiaru poziomu za pomocą jednego mikrofonu. Mikrofon powinien być umieszczony w takim miejscu, w jakim różnica poziomu ciśnienia akustycznego pochodzącego od każdego ze źródeł w punkcie pomiaru (wynikająca z różnicy odległości między źródłami a mikrofonem) będzie pomijalnie mała. Funkcja odpowiedzialna za poprawkę kalibracyjną poziomów sygnału sterującego jest realizowana niezależnie dla każdego z kanałów i znajduje się w bloku przetwarzania sygnałów sterujących za suwakiem służącym do zmiany poziomu sygnału.

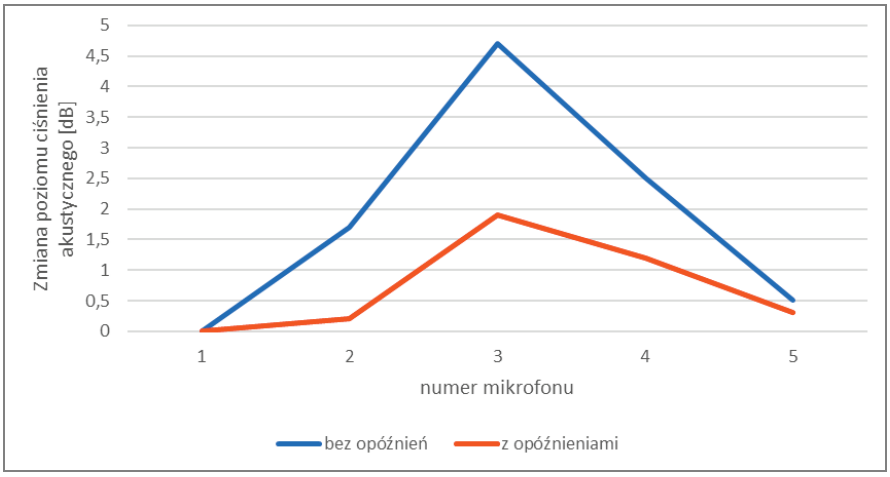
Po przeprowadzeniu kalibracji przystąpiono do wykonania pomiaru. Zmierzone wartości przeliczono tak, aby poziom 0 dB odpowiadał najmniejszemu zmierzonemu poziomowi dla danej konfiguracji sygnałów sterujących. W pierwszej części wykonano serię ośmiu pomiarów. Pierwszy pomiar został przeprowadzony przy włączonym jednym głośniku, następnie zwiększano liczbę pracujących głośników o jeden. Wyniki pomiarów w pięciu punktach poziomów znormalizowanych względem poziomu dla mikrofonu nr 1 przedstawiono na rys. 11.




Rys. 11. Wyniki pomiarów w przypadku zwiększanej długości kolumny; poziom znormalizowany względem poziomu dla mikrofonu nr 1

Z analizy rys. 11 wynika, że ze wzrostem długości kolumny głośnikowej (zwiększeniem liczby aktywnych przetworników) zwiększa się różnica poziomu zmierzonego na mikrofonie środkowym (nr 3) i poziomów zmierzonych na mikrofonach skrajnych. Potwierdza to możliwość zwiększenia kierunkowości źródła za pomocą przedstawionego systemu sterowania. Efekt został zaobserwowany, mimo że pomiary przeprowadzono w pomieszczeniu, które nie zapewnia pola swobodnego.

Na rysunku 12 przedstawiono porównanie dwóch wariantów z wykorzystaniem ośmiu przetworników. W wariancie pierwszym nie zastosowano żadnych opóźnień. W drugim – zmianę opóźnienia względem przetwornika skrajnego. Można zaobserwować, że dzięki zastosowaniu opóźnień względem jednego ze skrajnych przetworników uzyskano lekkie odchylenie kierunkowości źródła. Objawia się to zwiększeniem poziomu na mikrofonie nr 4 i zmniejszeniu poziomu na mikrofonie nr 3 i nr 2.




Rys. 12. Wyniki pomiarów bez zastosowanego opóźnienia i z zastosowanym opóźnieniem; poziom znormalizowany względem poziomu dla mikrofonu nr 1 (numery mikrofonów zgodnie z rys. 10)

## 6.6. Podsumowanie

W rozdziale przedstawiono koncepcję systemu sterowania kolumną głośnikową z możliwością niezależnego zasilania każdego z przetworników. System jest przeznaczony do sterowania kolumną głośnikową – bazuje na typowej wielokanałowej zewnętrznej karcie dźwiękowej i ogólnie dostępnym oprogramowaniu. Stanowisko i system stero-

wania znajduje szerokie zastosowanie w dydaktyce. Dzięki niemu studenci kierunków związanych z akustyką mogą w praktyce zapoznać się z linią źródeł punktowych, kierunkowością kolumny głośnikowej oraz z niezależnym sterowaniem przetwornikami. Zaproponowane stanowisko umożliwia badanie zjawisk wykorzystywanych w sterowaniu urządzeń typu Soundbar. Kolumna głośnikowa i system sterowania są przygotowane w taki sposób, żeby można było je z łatwością rozbudować. Oznacza to możliwość zarówno zwiększenia długości kolumny, jak i zbudowania dwuwymiarowej matrycy głośnikowej umożliwiającej odtworzenie czoła fali.

Przedstawiono również symulacje rozkładu poziomu ciśnienia akustycznego pochodzącego od kolumny składającej się z wielu źródeł punktowych. Otrzymane wyniki odnosiły się nie tylko do różnej długości kolumny, lecz także do różnej konfiguracji opóźnień sygnałów generowanych przez te źródła. Ponadto zaprezentowano wyniki pomiarów w ustalonej odległości od źródła. Pomiarów dokonano w pomieszczeniu odsłuchowym niezapewniającym warunków pola swobodnego – mimo to w ich ramach udało się zaobserwować efekty zmiany kierunku propagacji dla wartości opóźnień zbliżonych do tych uzyskanych w symulacjach. Wartości poziomów różniły się jednak, co wynika z różnic w parametrach pola akustycznego zdefiniowanego w symulacjach oraz – w warunkach rzeczywistych. Potwierdza to możliwość stosowania opisanej konstrukcji wraz z systemem w dydaktyce – nawet w przypadku braku dostępności komory bezpogłosowej.

Zaprezentowane urządzenie i system sterowania można również wykorzystać w ćwiczeniach laboratoryjnych dotyczących fundamentalnych zjawisk akustycznych. Na przykład jednym z problemów urządzeń głośnikowych składających się z więcej niż jednego przetwornika jest wzajemne wpływanie ich na siebie. Zjawisko nazywane wzajemną impedancją promieniowania zależy m.in. od odległości między przetwornikami i częstotliwością. Jeśli zastosuje się proponowany system, będzie można porównywać wartości obliczone teoretycznie z wartościami zmierzonymi, czy badać wpływ filtracji sygnałów szerokopasmowych podawanych na odpowiednie przetworniki na pracę urządzeń dwudrożnych. Kolumnę głośnikową można ponadto wykorzystać do badania linii źródeł punktowych i jako źródła koherentne, i niekoherentne. W obu przypadkach można porównywać wartości teoretyczne spadku poziomu ciśnienia akustycznego w funkcji odległości od źródła z wartościami zmierzonymi. Analizę można przeprowadzać dla różnej konfiguracji źródeł punktowych, czyli różnej odległości między źródłami i różnej długości linii.

**Słowa kluczowe:** cyfrowa kolumna głośnika, system sterowania, kierunkowość, metoda elementów brzegowych.

# Bibliografia


[1] Żyszkowski Z., *Podstawy Elektroakustyki*, Wydawnictwo Naukowo-Techniczne, Warszawa 1965.

[2] Dobrucki A,. *Przetworniki Elektroakustyczne*, Wydawnictwo Naukowo-Techniczne, Warszawa 2007.

[3] Novak J.F., *A Column Loudspeaker with Controlled Coverage Angle*, Paper 260, 1962, October.

[4] Smith A.P., Fagen D.G., *Applications of Columnar Loudspeaker Systems in Reverberant Halls of Worship*, Paper 2153, 1984, October.

[5] van der Werff J., *Design and Implementation of a Sound Column with Exceptional Properties*, Paper 3835, 1994, February.

[6] Feistel S., Goertz A., *Digitally Steered Columns: Comparison of Different Products by Measurement and Simulation*, Paper 8935, 2013, October.

[7] Meynial X., *DGRC Arrays: A Synthesis of Geometrical and Electronic Loudspeaker Arrays*, Paper 6786, 2006, May.

[8] Meynial X., Grégoire G., *Design of a Passive DGRC Column Loudspeaker with Wave Front Synthesis*, Paper 8353, 2011, May.

[9] PIETSCHMANN A., Active Loudspeaker Column System, patent nr US20170171649A1; https://patents.google.com/patent/US20170171649A1/en [dostęp: 26.03.2021].

[10] Shuki V., *Systems and methods for testing and calibrating a focused ultrasound transducer array*, patent nr US6543272B1; https://patents.google.com/patent/US6543272B1/en [dostęp: 26.03.2021].

[11] Rupitsch S.J., *Piezoelectric Ultrasonic Transducers*, [w:] *Piezoelectric Sensors and Actuators. Topics in Mining, Metallurgy and Materials Engineering*, Springer, Berlin–Heidelberg 2019, DOI: 10.1007/978-3-662-57534-5_7.

[12] U-PHORIA UMC1820 Quick Start Guide; https://www.bhphotovideo.com/lit_files/155647.pdf [dostęp: 26.03.2021].

[13] ASIO4ALL – Universal ASIO Driver For WDM; http://www.asio4all.org/ [dostęp: 26.03.2021].

[14] REAPER DIGITAL AUDIO WORKSTATION; https://www.reaper.fm/ [dostęp: 26.03.2021].

[15] PURE DATA – Pd Community Site; https://puredata.info/ [dostęp: 26.03.2021].

[16] Panzer J., *Coupled lumped and boundary element simulation for electro-acoustics*, „Acoustics" 2012, April, ffhal-00811256f.

# 7. Praktyczna implementacja przetwornika o stałej szerokości wiązki do projektowania monitorów studyjnych

TOMASZ NOWAK, BARTŁOMIEJ KRUK

Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

Pierwsze badania nad przetwornikiem o stałej szerokości wiązki (Constant Beamwidth Transducer; CBT) sięgają lat 80. XX w., dotyczyły one wówczas wojskowych badań nad przetwornikami podwodnymi [1]. Później Don Keele zastosował tę teorię do głośników pracujących w paśmie słyszalnym; otrzymane wyniki wielokrotnie prezentował w publikacjach Audio Engineering Society.

W niniejszym rozdziale przedstawiono zalety wykorzystania CBT do budowy monitorów studyjnych. Podobnie jak w przypadku źródeł liniowych tą samą korzyścią jest na przykład to, że dzięki nim można zmniejszyć spadek poziomu ciśnienia akustycznego przez podwojenie odległości od źródła o 3 dB w porównaniu do 6 dB spadku dla źródła punktowego. A drugą – uzyskanie większego obszaru odsłuchowego z bardziej równomiernym rozkładem poziomu ciśnienia akustycznego w porównaniu do źródła punktowego.

## 7.1. Wprowadzenie

### 7.1.1. Początki CBT

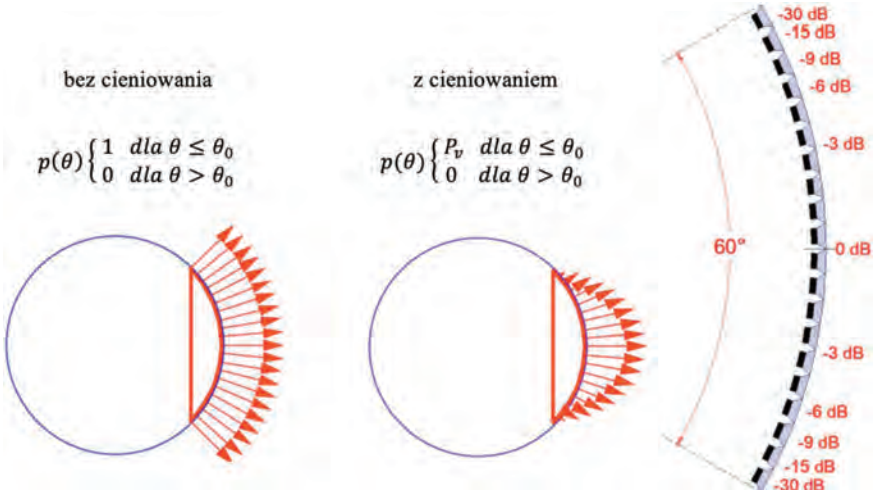
W badaniach dotyczących CBT pod uwagę brany jest przetwornik o zakrzywionej powierzchni w formie kulistej obudowy z zastosowanym cieniowaniem Legendre'a

niezależnym od częstotliwości. Cieniowanie jest konieczne do uzyskania szerokopasmowej, stałej kierunkowości zapewniającej stały kąt promieniowania w funkcji częstotliwości – praktycznie bez bocznych listków.

Keele teorią tą rozszerzoną o wykorzystanie matryc głośnikowych w kształcie łuku kołowego zajmuje się od 2000 r. [2–4, 6].

## 7.1.2. Koncepcja działania

W teorii CBT zakłada się, że każdy przetwornik w matrycy powinien generować różne poziomy ciśnienia akustycznego, które są zgodne z ciągłą funkcją cieniowania Legendre'a [1]. Funkcja cieniowania Legendre'a stopniowo zmniejsza poziomy każdego z przetworników jak w przykładzie przedstawionym na rys. 1 – dwa najniższe głośniki w matrycy uzyskują pełną moc, w dwóch następnych poziomy zredukowano o 3 dB, a w najwyższym przetworniku aż o 6 dB. Stopniowe zmiany wzmocnienia zrealizowano pasywnie za pomocą dzielników napięcia oraz przez zmianę impedancji przetwornika (kombinacja połączeń szeregowo-równoległych).




Rys. 1. Schemat zestawu głośnikowego CBT o łuku kołowym 60° z cieniowaniem Legendre'a. Zależne od kąta poziomy cieniowania są minimalne w środku matrycy i zwiększają się w kierunku zewnętrznych krawędzi matrycy [5]

Dużą zaletą cieniowania jest to, że można zrealizować je w technice analogowej na elementach R, L, C, czyli bez cyfrowego przetwarzania sygnałów czy filtrów opartych

na wzmacniaczach operacyjnych. Umożliwia to zbudowanie urządzenia działającego na pojedynczym kanale wzmacniacza mocy.

### 7.1.3. Zalety w stosunku do źródła punktowego

Zmianę poziomu ciśnienia akustycznego z odległością dla źródła punktowego (fala sferyczna) można wyrazić za pomocą wzoru:

$$SPL_2 = SPL_1 - 20 \cdot \log \frac{R_2}{R_1} \tag{1}$$

gdzie:

$SPL_1$ – poziom ciśnienia akustycznego w punkcie 1,

$SPL_2$ – poziom ciśnienia akustycznego w punkcie 2,

$R_1$ – odległość od źródła dźwięku do punktu 1,

$R_2$ – odległość od źródła dźwięku do punktu 2.

Zmianę poziomu ciśnienia akustycznego z odległością dla źródła liniowego (fala cylindryczna) opisuje się natomiast zgodnie ze wzorem:

$$SPL_2 = SPL_1 - 10 \cdot \log \frac{R_2}{R_1} \tag{2}$$

Ze wzorów (1) i (2) wynika spadek poziomu amplitudy ciśnienia akustycznego fali 3 dB na każde podwojenie odległości od źródła liniowego (w odległości krytycznej) – w porównaniu do 6 dB w źródle punktowym. W przypadku zakrzywionej matrycy i konstrukcji CBT nawet przy odległości większej niż krytyczna pozorne centrum akustyczne znajduje się dalej od przedniej przegrody urządzenia, co skutkuje większym natężeniem dźwięku w porównaniu do źródła punktowego umieszczonego w miejscu przedniej części CBT.

Odległość krytyczna to przybliżona granica między polem bliskim i polem dalekim, w której urządzenie przestaje zachowywać się jak źródło liniowe lub łukowe, a zaczyna – jak źródło punktowe [7]. Opisuje ją wzór:

$$r_c = \frac{2L^2}{\lambda} - \frac{\lambda}{8} \tag{3}$$

gdzie:

$L$ – długość łuku,

$r_c$ – odległość krytyczna,

$\lambda$ – długość fali.

Dobrą aproksymacją jest następujące uproszczenie (zostanie ono użyte w następnej części pracy):

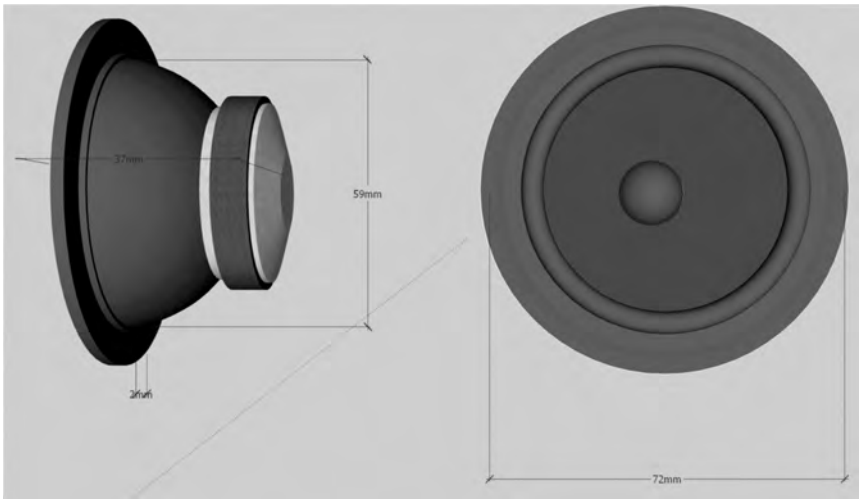
$$r_c \approx\sim 0{,}006L^2 \tag{4}$$

# 7.2. Wytyczne i ograniczenia w projektowaniu kompaktowego urządzenia typu CBT

## 7.2.1. Wybór przetwornika

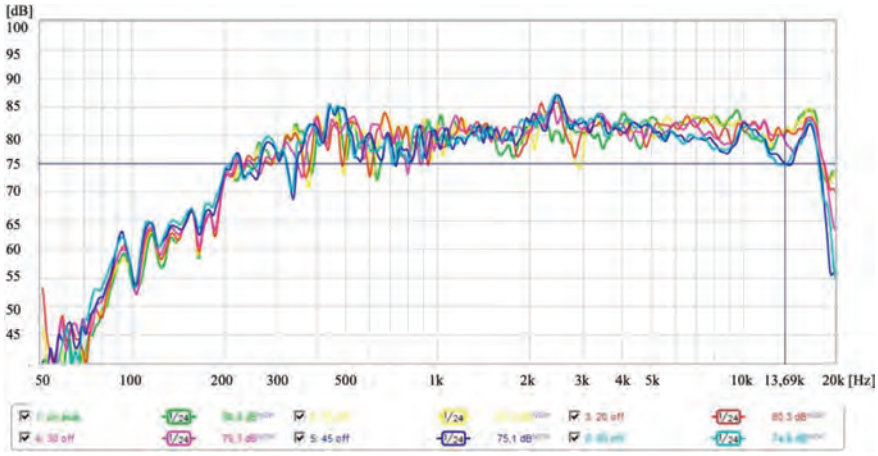
Urządzenie głośnikowe jest podzielone na sekcję niskotonową oraz wiązkę przetworników CBT – pierwsza z nich ze względu na kompaktowy format typowy dla monitorów studyjnych. Sekcja ta składa się z przetwornika o dużym przemieszczeniu membrany (o średnicy 125 mm) umieszczonego w 8-litrowej komorze z rezonatorem Helmholtza dostrojonym do częstotliwości 45 Hz. Druga sekcja to pięć przetworników szerokopasmowych (o średnicy 72 mm) i zmierzonym paśmie przenoszenia 200 Hz–18,5 kHz (–6 dB) i nominalnej dyspersji 90° (–6 dB). Wymiary geometryczne przetworników użytych do zaprojektowania monitora CBT przedstawiono na rys. 2.

Średnica przetworników pełnopasmowych określa minimalny akustyczny odstęp między środkami na przedniej ściance. Rozmiar membrany ma również wpływ na




Rys. 2. Wymiary geometryczne przetworników użytych do projektowania CBT

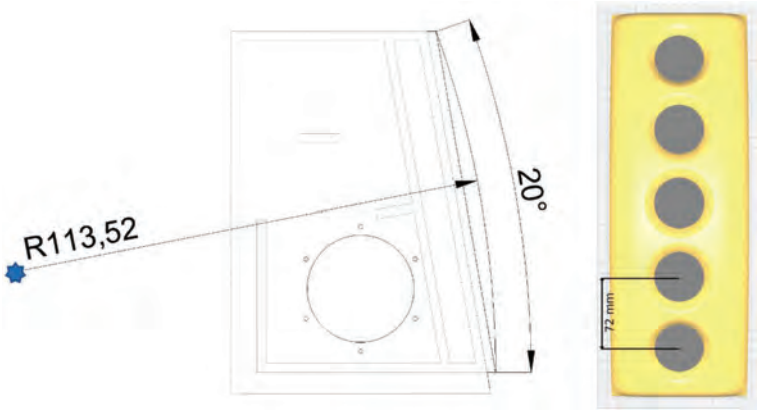
ograniczenie dolnej częstotliwości oraz zawężenie szerokości wiązki wysokich częstotliwości. Wybór przetwornika szerokopasmowego był kompromisem między wymienionymi parametrami.




Rys. 3. Charakterystyka częstotliwościowa przetwornika
użytego do projektowania CBT pod różnymi kątami (od osi symetrii do kąta 60° poza nią)

## 7.2.2. Położenie pozornego źródła akustycznego

Obudowa głośnika została zaprojektowana w taki sposób, aby przetworniki szerokopasmowe znajdowały się w jednej linii ze sobą na zakrzywionej przegrodzie przedniej.

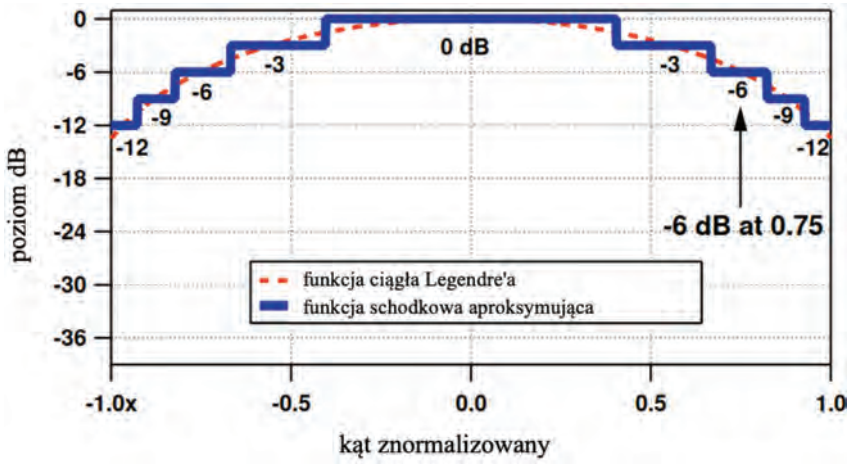



Rys. 4. Schemat obudowy urządzenia głośnikowego i rzut pionowy
projektowanego frontu kształtującego urządzenie CBT

Styczna tej krzywizny jest prostopadła do płaszczyzny podłoża po uwzględnieniu odbicia fali od tej płaszczyzny, a oś akustyczna jest równoległa do powierzchni, na której ustawione jest urządzenie. Promień przegrody przedniej wynosi 1,1352 m – to odległość determinująca położenie źródła pozornego znajdującego się za urządzeniem głośnikowym (rys. 4).

### 7.2.3. Koncepcja filtra elektrycznego

Cieniowanie Legendre'a zastosowano w celu poprawy kierunkowości pionowej w funkcji częstotliwości (rys. 5). Bez cieniowania przy dużych wartościach kąta odchylenia od osi pojawiają się tzw. listki boczne.




Rys. 5. Cieniowanie Legendre'a – aproksymacja schodkowa (kąt w radianach) [2]

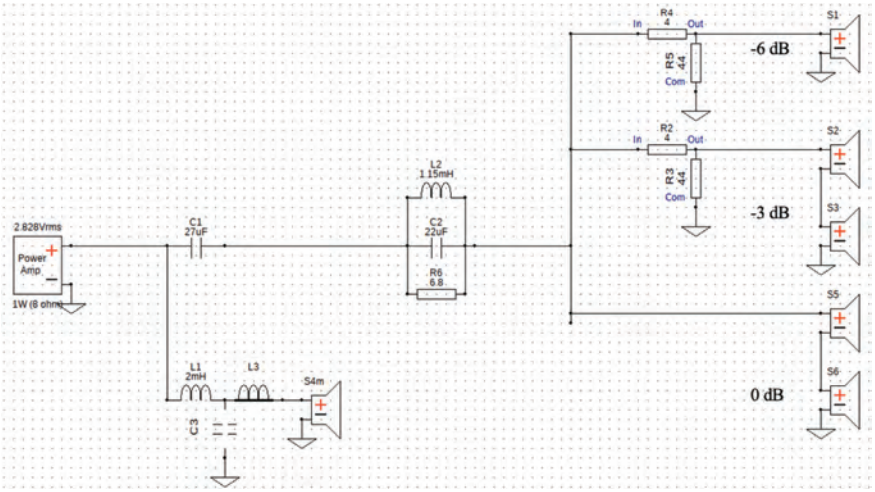
Układ pięciu głośników został podzielony na trzy sekcje: dwa przetworniki dolne połączone w szeregu (0 dB), dwa przetworniki środkowe również w szeregu z dzielnikiem napięcia (–3 dB), jeden przetwornik górny połączony z dzielnikiem napięcia (–6 dB).

Istnieje szereg parametrów technicznych, które trzeba odpowiednio zaprojektować i utworzyć monitory studyjne. Jednym z istotniejszych jest impedancja głośników zapewniająca odpowiednie parametry pracy pozostałych elementów toru fonicznego. Wykorzystane przetworniki mają impedancję znamionową o wartości 4 Ω. Matryca głośników powinna mieć impedancję wypadkową o wartości również co najmniej 4 Ω. Niższa wartość może być problematycznym obciążeniem dla popularnych wzmacniaczy mocy. Impedancja wejściowa dzielników napięcia została tak dobrana, aby można

było uzyskać impedancję wypadkową o wartości 4 Ω (rys. 6). Wartości elementów natomiast – eksperymentalnie (pomiarami), by uwzględnić różnice między przetwornikami: w przebiegach impedancji w funkcji częstotliwości (do 0,5 Ω) oraz w ich skuteczności napięciowej (do 1,5 dB).

Tłumienie zostało zrealizowane za pomocą dwóch dzielników o tych samych wartościach elementów. Jeden z nich tłumił parę szeregowo połączonych przetworników, a drugi pojedynczy przetwornik.

Połączenie szeregowe jednakowych głośników skutkuje zyskiem sprawności ze względu na sumowanie poziomów ciśnienia akustycznego oraz jednocześnie dwukrotnie mniejszym napięciem na każdym z przetworników. W rezultacie dwa przetworniki połączone szeregowo generują taki sam poziom ciśnienia akustycznego jak pojedynczy przetwornik – przy jednakowym napięciu wejściowym w obu przypadkach.

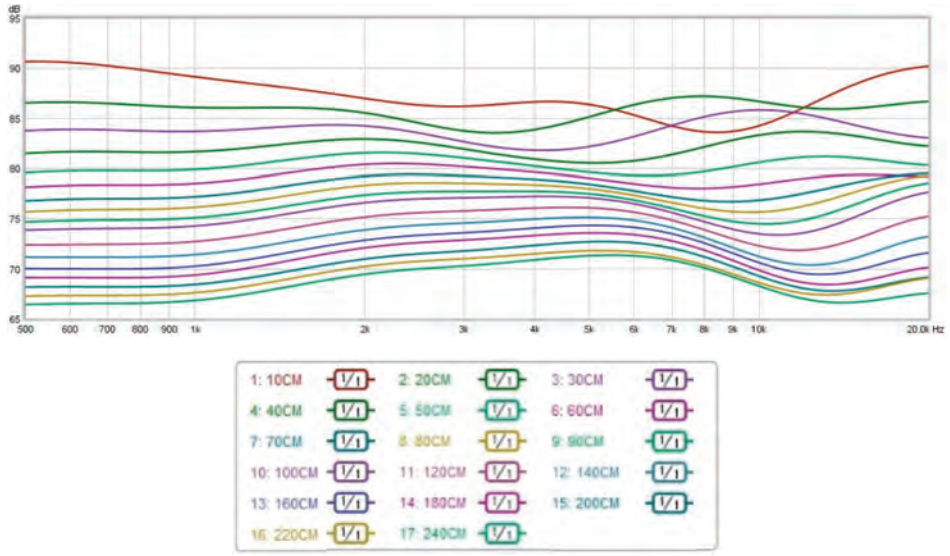



Rys. 6. Schemat elektryczny realizacji cieniowania

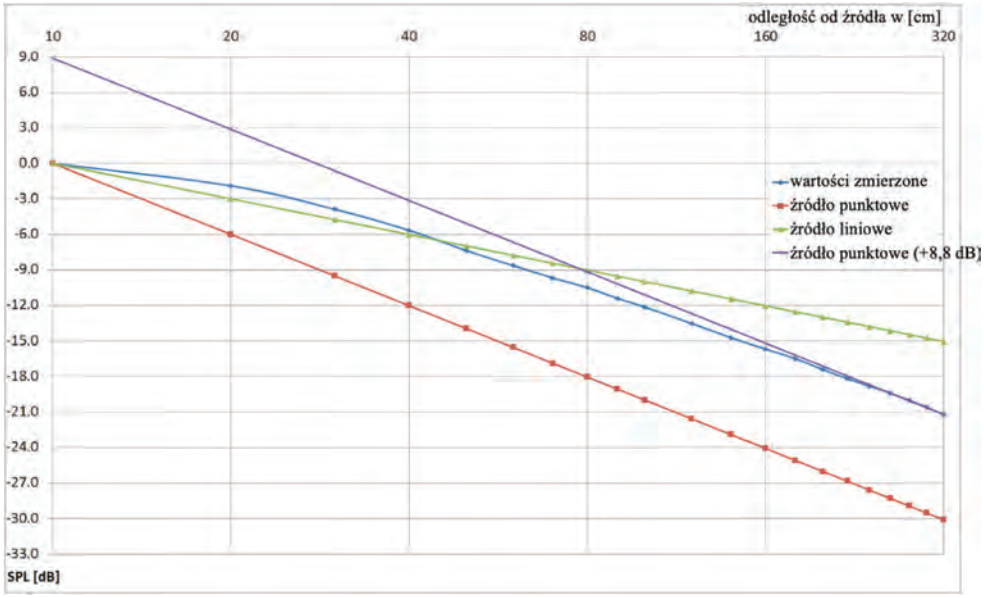
## 7.3. Pomiary prototypu

Pomiary charakterystyki częstotliwościowej prototypu monitora studyjnego z przetwornikami CBT wykonano na płaszczyźnie podłoża (rys. 7) w odległościach 10–320 cm. Dane przedstawiono w postaci średniego SPL (w paśmie 500 Hz–20 kHz) w funkcji odległości względem poziomu zmierzonego w odległości 10 cm od urządzenia (rys. 8).

Zmierzone wartości porównano z teoretycznymi spadkami dla źródła liniowego i źródła punktowego.



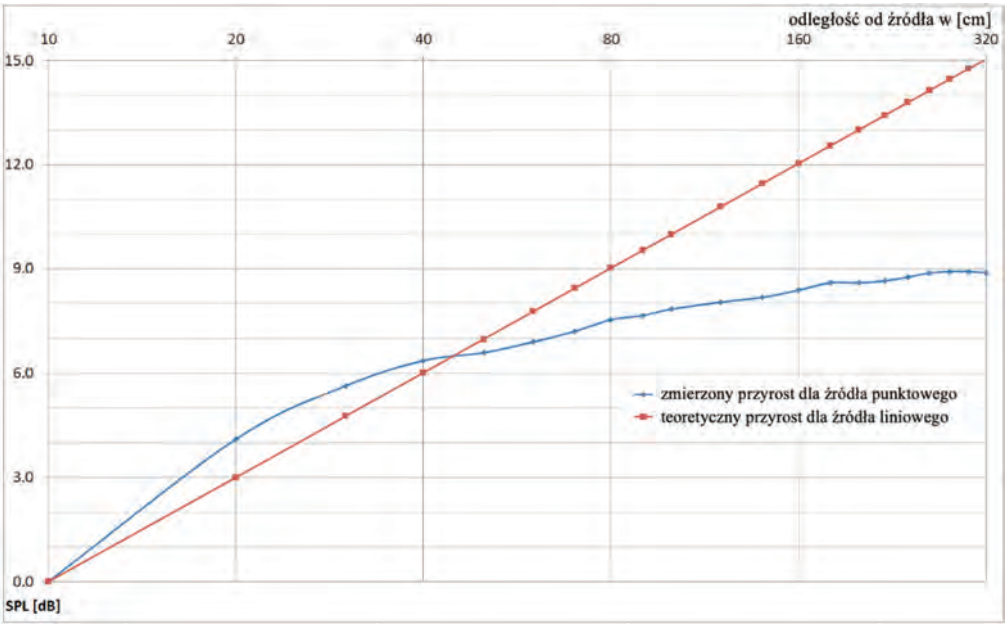

Rys. 7. Kolejne pomiary charakterystyki częstotliwościowej w zależności od odległości od urządzenia




Rys. 8. Zmierzony SPL w funkcji odległości
w porównaniu z teoretycznym zachowaniem źródła liniowego i punktowego

Nachylenie zmierzonego zbocza jest mniejsze niż nachylenie zbocza źródła liniowego w polu bliskim przy odległości do 40 cm, a następnie powoli przechodzi w równoległe do nachylenia zbocza źródła punktowego przy odległości 320 cm. Tłumienie w funkcji odległości jest między wartościami teoretycznymi źródła liniowego i punktowego w polu średnim.

W celu bardziej przejrzystej prezentacji dodatkowa krzywa tłumienia źródła punktowego została podniesiona o 8,8 dB, żeby dopasować SPL zmierzonego urządzenia w odległości 320 cm. Wskazuje to na przewagę CBT nad źródłem punktowym, który charakteryzuje się większą różnicą poziomów ciśnienia akustycznego w funkcji odległości i mniejszą skutecznością w warunkach pola dalekiego. Przewagę CBT można przedstawić jako przyrost SPL w stosunku do źródła punktowego w funkcji odległości (rys. 9) i porównać z tak samo obliczonym przyrostem źródła liniowego.




Rys. 9. Zmierzony przyrost SPL w funkcji odległości dla źródła CBT oraz teoretyczny przyrost źródła liniowego

## Pozorne położenie źródła dźwięku

Wzór na spadek poziomu ciśnienia akustycznego dla źródła punktowego (1) można zmodyfikować przez wprowadzenie promienia pozornego położenia źródła, który jest

dodatkową odległością z perspektywy słuchacza:

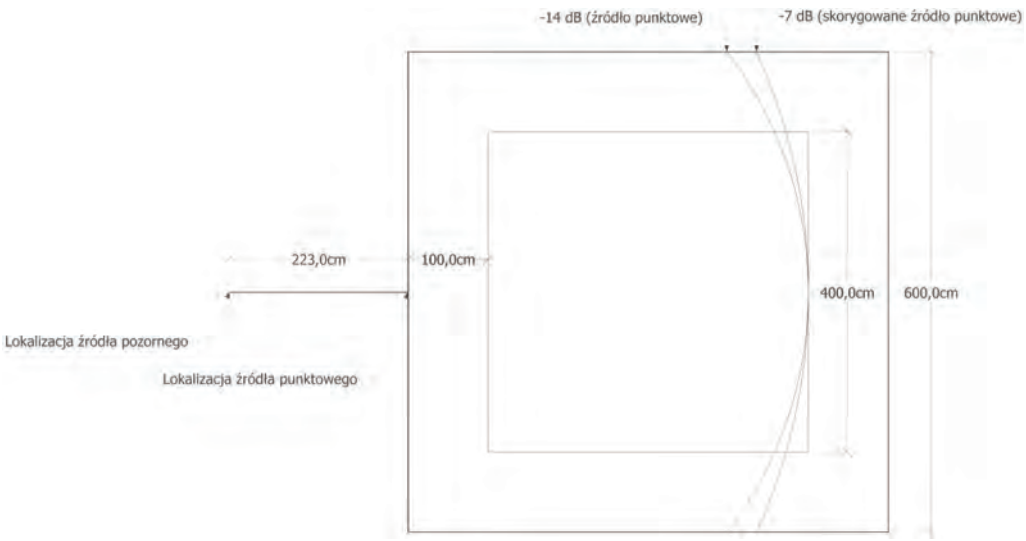
$$SPL_2 = SPL_1 - 20 \cdot \log \frac{R_2 + r'}{R_1 + r'} \tag{5}$$

gdzie:

$r'$ – odległość pozornego źródła punktowego od urządzenia rzeczywistego.

Zmienna $r'$ wyprowadzona ze wzoru (5) przyjmie postać:

$$r' = \frac{R_1 \cdot 10^{\frac{SPL_2 - SPL_1}{20}}}{1 - 10^{\frac{SPL_2 - SPL_1}{20}}} \tag{6}$$

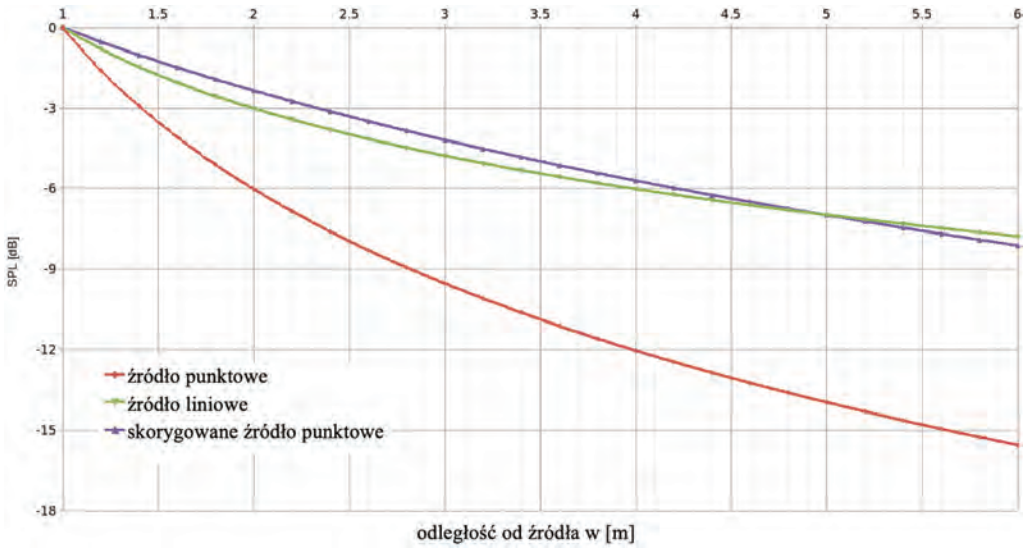



Rys. 10. Przykład omawianego problemu; rzut pomieszczenia z góry

Zarysowany problem można przedstawić (por. rys. 10) na przykładzie kwadratowego pokoju o wymiarach 6 m × 6 m. Miejsce odsłuchowe znajduje się na środku tego pomieszczenia, a granica docelowej przestrzeni odsłuchowej jest oddalona o 1 m od ścian pomieszczenia. Źródło dźwięku umieszczono na ścianie frontowej. Założono jednakowy poziom ciśnienia akustycznego w odległości 1 m od źródła punktowego, liniowego i skorygowanego źródła punktowego. W odległości 5 m źródło liniowe wykazuje spadek SPL o 7 dB. Żeby wyznaczyć taką odległość źródła pozornego ($r'$), by

spadek SPL źródła liniowego był równy, należy zastosować równanie (6) umożliwiające obliczenie tej wartości (rys. 11):

$$r' = \frac{1 \cdot 10^{\frac{-7-0}{20}}}{1-10^{\frac{-7-0}{20}}} \Rightarrow r' = 2{,}23 \text{ m} \tag{7}$$




Rys. 11. SPL w funkcji odległości – dostosowane źródło punktowe przy spadku 7 dB w odległości 5 m: $r' = 2{,}23$ m

W przypadku każdego źródła w kształcie łuku istnieje odległość krytyczna określona przez zależność długości łuku i długości fali – przybliżoną odległość można wyznaczyć na podstawie wzoru na odległość krytyczną dla matrycy liniowej (4).

W podanym przykładzie przy założeniu najkrótszej długości fali – 0,034 m (10 kHz) długość źródła łuku powinna wynosić co najmniej:

$$L \approx \sqrt{\frac{rc}{0{,}006 \cdot f}} \tag{8}$$

$L_h \approx 0{,}29$ m (odległość krytyczna dla górnej granicy pasma przenoszenia).

Przy przyjęciu dolnej granicy częstotliwości 100 Hz:

$L_l \approx 2{,}9$ m (odległość krytyczna dla dolnej granicy pasma przenoszenia).

## 7.4. Podsumowanie

Biorąc pod uwagę wyniki przedstawione w niniejszym rozdziale, można wnioskować, że niewielki, zakrzywiony zestaw głośnikowy zapewnia wymierną przewagę nad głośnikiem punktowym w zakresie równomierności poziomu ciśnienia akustycznego w funkcji odległości. Skutkiem tego są znacznie mniejsze różnice poziomu ciśnienia akustycznego w funkcji odległości od urządzenia (rozkład natężenia dźwięku w obszarze odsłuchowym ma mniejsze odchylenie od średniej w przypadku urządzenia CBT niż w przypadku źródła punktowego). Potencjalnie może to prowadzić do zwiększenia użytecznej przestrzeni odsłuchowej w obszarze odsłuchu w przypadku urządzeń wykorzystujących CBT. Zmniejszony poziom ciśnienia akustycznego między pozycjami odsłuchowymi w polu bliskim i dalekim może zapewnić elastyczność w ustawieniu systemów odsłuchowych w studio.

Planowane są dalsze pomiary charakterystyki kierunkowości oraz optymalizacja urządzeń.

**Słowa kluczowe:** przetwornik o stałej szerokości wiązki, źródło punktowe, źródło liniowe, monitory studyjne.

## Bibliografia

[1] Buren A.L. van, Rogers P.H., *New Approach to a Constant Beamwidth Transducer*, „J. Acous. Soc. Am." 1978, July, Vol. 64, No. 1, s. 38–43.

[2] Keele Jr D.B., *The application of broadband constant beamwidth transducer (CBT) theory to loudspeaker arrays*, AES E-Library, paper number: 5216, Audio Engineering Society Convention 109, Audio Engineering Society, 1 September 2000.

[3] Keele Jr D.B., *Practical implementation of constant beamwidth transducer (CBT) loud- speaker circular-arc line arrays*, AES E-Library, paper number: 5863, Audio Engineering Society Convention 115, Audio Engineering Society, 1 October 2003.

[4] Keele Jr D.B., Button D.J., *Ground-plane constant beamwidth transducer (CBT) loud-speaker circular-arc line arrays*, AES E-Library, paper number: 6594, Audio Engineering Society Convention 119, Audio Engineering Society, 1 October 2005.

[5] Keele Jr D.B, *Design of Free-Standing Constant Beamwidth Transducer (CBT) Loudspeaker Line Arrays for Sound Reinforcement*, AES E-Library, paper number: 9624, Audio Engineering Society Convention 141, Audio Engineering Society, 20 September 2016.

[6] Keele Jr D.B., *A Ground Plane Measurement Comparison Between Two Floor-Standing Loudspeaker Systems: A Conventional Three-Way Studio Monitor vs. A Ground-Plane Constant Beamwidth Trans-*

*ducer (CBT) Line Array*, AES E-Library, eBrief: 279, Audio Engineering Society Convention 141, Audio Engineering Society, 20 September 2016.

[7] Ureda S.M., *Pressure Response of Line Sources*, AES E-Library, paper number: 5649, Audio Engineering Society Convention 113, Audio Engineering Society, 1 October 2002.

# 8. Ocena ogólnej jakości oraz wybranych atrybutów dźwięku sygnałów muzyki i jakości mowy nadawanych za pomocą radiofonii cyfrowej DAB+

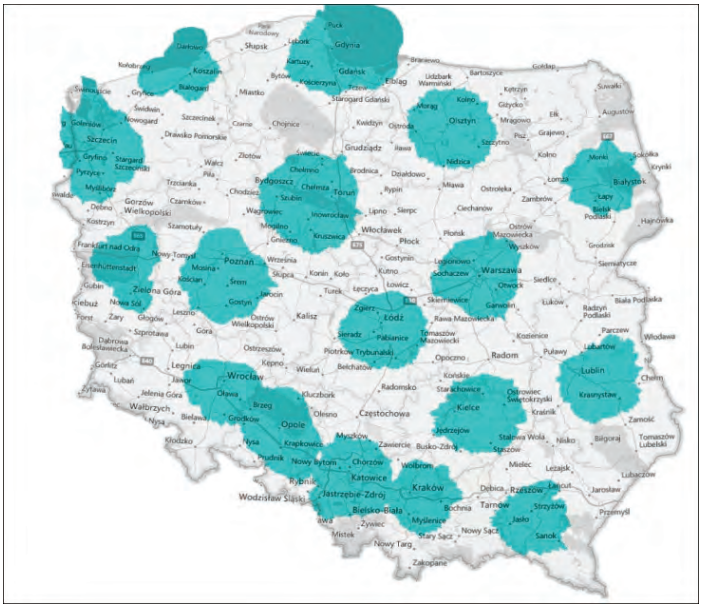
MAURYCY KIN, STEFAN BRACHMAŃSKI

Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

W niniejszym rozdziale 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 siedem 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, a mowy – 48 kb/s, niezależnie od ocenianych sygnałów. Przy konwencjonalnym nadawaniu z wykorzystaniem jednego nadajnika uzyskano podobne rezultaty z zastrzeżeniem, że warunki odbioru sygnału radiowego były sprzyjające, gdyż niekiedy występowały przerwy w sygnale.

# 8.1. Wprowadzenie

Obecnie 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 r. w dwóch aglomeracjach – warszawskiej i śląskiej. Aktualnie w zasięgu emisji cyfrowego radia DAB+ znajduje się 56% ludności Polski [30], a do końca 2021 r. planowane jest pokrycie 63,7% obszaru Polski, czyli zasięg obejmie ok. 81,6% ludności [5]. Rejony odbioru sygnałów DAB+ w Polsce przedstawiono na rys. 1.




Rys. 1. Mapa zasięgu cyfrowej radiofonii DAB+ w Polsce [31]

Niezależnie od ogólnopolskich radiostacji DAB+ rozwijane są lokalne multipleksy. Zgodnie z zaleceniami Europejskiej Unii Nadawców (European Broadcasting Union; EBU) z 2013 r. radiofonia cyfrowa powinna objąć swoim zasięgiem zarówno wielkie obszary (cały kraj bądź jego regiony), jak i mniejsze terytoria (np. aglomeracje miejskie), w których przypadku ze względu na koszty oraz lokalny charakter przekazywanych treści nadaje się rozwiązanie w postaci sieci jednoczęstotliwościowej. Zastosowano je we Wrocławiu: trzy nadajniki sygnału DAB+ rozmieszczono w wierzchołkach trójkąta – umożliwia to pokrycie niemalże całego obszaru miasta. Do kodowania został wyko-

rzystany standard HE-AAC (High-Efficiency Advanced Audio Coding) wersja 2 [11], dzięki czemu uzyskano możliwość poprawy jakości przy niższych szybkościach bitowych przez wykorzystanie przetwarzania technik: SBR (Spectral Band Replication) oraz PS (Parametric Stereo)skutkujących wyraźną poprawę jakości dźwięku przy szybkościach wynoszących 64 kb/s i 48 kb/s [4, 9, 14].

Rozwój technologii cyfrowych dostarczył nowych możliwości przetwarzania sygnałów fonicznych. Najważniejszą zaletą sygnałów cyfrowych jest łatwość ich przechowywania, przesyłania i przetwarzania. Obecnie jednak wciąż istnieją ograniczenia związane z magazynowaniem oraz przepustowością kanałów transmisyjnych. W wielu przypadkach jest wskazane, aby sygnały audio były kodowane jak najwydajniej. Realizacja tego zadania polega przede wszystkim na procesie redukcji bitów. Najprostszym rozwiązaniem wydaje się być zastosowanie niższej częstotliwości próbkowania, a także mniejszej rozdzielczości bitowej – ten zabieg jednak skutkuje znaczną degradacją jakości sygnału. Konieczne stało się zatem wynalezienie innych metod umożliwiających zmniejszenie liczby informacji. W telekomunikacji dość powszechnie stosuje się kwantyzację nieliniową. Nie można niestety dzięki niej uzyskać wystarczającej jakości dla bardziej złożonych sygnałów fonicznych (np. muzyki). Najbardziej popularną metodą znajdującą zastosowanie dla wszystkich rodzajów sygnałów jest tzw. kompresja danych. Stopień kompresji określający stosunek liczby danych przed kompresją do liczby danych po kompresji zależy od konkretnego algorytmu. W przybliżeniu w przypadku metod kompresji bezstratnej uzyskuje się ok. 40–60% redukcji zawartości, kompresja stratna może natomiast powodować redukcję nawet do kilku procent objętości danych pierwotnych. Aktualnie dostępna jest duża liczba algorytmów redukcji bitów wykorzystujących zarówno kompresję stratną, jak i bezstratną. Poszczególne kodeki są zoptymalizowane pod kątem różnych zastosowań, tj. radiofonia cyfrowa, telewizja cyfrowa, systemy VoiP, media internetowe, kinematografia, systemy kina domowego czy pobieranie i przechowywanie danych na odtwarzaczach osobistych.

Mimo znacznego postępu w tworzeniu obiektywnych metod oceny jakości transmisji sygnału mowy i muzyki nadal jedynym wiarygodnym weryfikatorem jakości są metody wykorzystujące pomiary subiektywne [24, 13], za pomocą których można wyznaczyć dopuszczalną degradację sygnałów niepowodującą drastycznego spadku jakości dźwięku. Ponieważ w Katedrze Akustyki i Multimediów przeprowadzano wcześniej (2005–2009) badania jakości dźwięku nadawanego w systemie DAB+ [6, 14] przy emisji eksperymentalnej z jednego nadajnika, postanowiono zbadać jakość przekazu sygnałów mowy i muzyki w rozwiązaniu wykorzystującym sieć jednoczęstotliwościową. Zrealizowana sieć oparta jest na bazie trzech nadajników rozmieszczonych na masz-

tach budynków: Polskiego Radia Wrocław S.A., Instytutu Łączności PiB we Wrocławiu oraz na terenie MPWiK, znajdującym się w bezpośrednim sąsiedztwie terenów kampusu Politechniki Wrocławskiej [18, 28].

Z punktu widzenia odbiorcy skuteczność radiofonii cyfrowej oceniana jest przede wszystkim ze względu na jakość odbieranych audycji – muzycznych i słownych (Quality of Experience). Inne możliwości oferowane w systemie DAB+ mają nieco mniejsze znaczenie.

Celem eksperymentu było zbadanie:

- wpływu szybkości bitowej na ogólną jakość sygnałów mowy i muzyki emitowanych w lokalnej radiofonii cyfrowej,
- wpływu szybkości bitowej na ocenę wybranych atrybutów wrażenia słuchowego nagrań muzycznych,
- jakości odbieranych audycji słownych oraz muzycznych zarejestrowanych w różnych punktach Wrocławia.

## 8.2. Metoda badań

Jakość mowy i muzyki może być oceniana wg 5-stopniowej skali MOS (Mean Opinion Score) [13]. Do oceny jakości sygnałów muzycznych można zastosować także procedurę opartą na 100-punktowej skali MUSHRA [12]. 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. Zgodnie z kryterium jakościowym ocenę wykonuje się najczęściej metodami ACR (Absolute Category Rating) lub DCR (Degradation Category Rating) zalecanymi przez ITU [13]. Ze względu na specyfikę odbioru programów radiowych najważniejszym aspektem estetycznym jest percepcja materiału dźwiękowego w określonej sytuacji bez porównania z wzorcem. Ponieważ wartościowanie oceny polega na ogólnie wyrobionym poczuciu estetycznym [19], do oceny jakości wybrano metodę ACR. Badania przeprowadzono z wykorzystaniem sygnałów emitowanych przez różne stacje. Mając na uwadze cel badań, czyli ocenę jakości dźwięku emitowanego przez rozgłośnie już działające posłużono się materiałem dźwiękowym słownym i muzycznym nadawanym przez te konkretne rozgłośnie, a nie przygotowanym specjalnie ciągiem sygnałów. To oznaczało wykonanie nagrań programów radiowych w różnych punktach miasta, następnie utworzenie z nich ciągów testowych, które poddane zostały ocenie subiektywnej. W przypadku sygnałów muzycznych ocenie podlegały zarówno ogólna jakość dźwięku, jak i na-

stępujące atrybuty wrażenia słuchowego: barwa dźwięku (rozumiana jako naturalne brzmienie źródeł dźwięku niezależne od zmiany struktury widmowej wynikającej z procesu przetwarzania), wrażenie perspektywy (zdolność do różnicowania planów dźwiękowych przez określenie dystansu między słuchaczem a pozornym źródłem dźwięku) oraz ostrość lokalizacji źródeł w panoramie (czyli precyzja lokalizacji pozornych źródeł dźwięku).

Rejestrację testowych sygnałów muzycznych zrealizowano w trzech punktach miasta następująco zaznaczonych na rys. 2: 1 – ul. Powstańców Śl. (okolice budynku Sky Tower), 2 – teren Politechniki Wrocławskiej, 8 – okolice Rynku (rys. 2). Rejestrowane audycje były nadawane z szybkościami bitowymi: 48 kb/s, 64 kb/s, 96 kb/s i 128 kb/s, przy czym liczba zarejestrowanych próbek dźwiękowych była taka sama dla każdej z badanych przepływności, co pozwoliło na wykorzystanie w testach ponad 200 próbek sygnałów muzycznych z różnych gatunków.


Rys. 2. Rozmieszczenie punktów pomiarowych na terenie miasta Wrocławia

Sygnały mowy zarejestrowano w dziewięciu punktach Wrocławia (patrz rys. 2). W tym przypadku materiał testowy składał się z dziesięciu 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, 112 kb/s. Każdej 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 umożliwił zbadanie nie tylko jakości dźwięku samych programów, lecz także wpływu miejsca odbioru na jakość dźwięku. Wyselekcjonowane fragmenty przygotowano zgodnie z zaleceniami podanymi w normach ITU [13] oraz EBU [7], czas trwania pojedynczej próbki dźwiękowej sygnałów muzycznych wynosił ok. 20 s przy zachowaniu frazy lub odcinka melodii. Materiał dźwiękowy podzielono na pięć grup: muzyka rozrywkowa, jazz, muzyka klasyczna, rock, sygnały mowy. Sygnałem referencyjnym do sygnałów muzycznych były dwa nagrania z płyt CD – z muzyką klasyczną i z muzyką rockową prezentowane również przez dwie z badanych rozgłośni.

Test skonstruowano następująco: próbki prezentowano w przypadkowej kolejności po trzy razy, w jednym ciągu testowym próbka wzorcowa pojawiała się natomiast tylko raz. Miało to na celu zbadanie powtarzalności ocen słuchaczy dotyczących zarówno sygnału wzorcowego, jak i stabilności ocen badanych sygnałów.

W badaniach nad jakością sygnałów mowy przyjęto szybkość próbkowania: 44 100 próbek/s. Sygnał mowy mógłby być wprawdzie próbkowany z szybkością 16 000 próbek/s [3], ale w ocenianym cyfrowym radiu LokalDAB w procesie przygotowania materiału emisyjnego przyjęto standard próbkowania z szybkością 44 100 próbek/s i dla sygnału muzycznego, i dla 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.

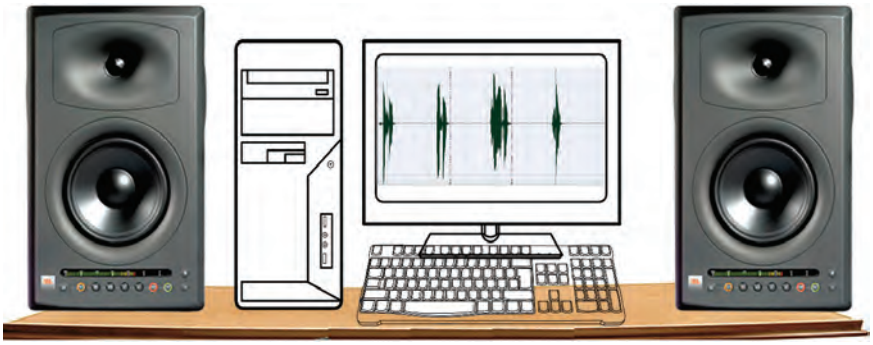
Badanie jakości mowy i muzyki przeprowadzono metodą skalowania absolutnego ACR, bez porównania jakości badanej próbki z wzorcem [12, 13]. W badaniach dotyczących oceny muzyki wzięło udział 11 osób w wieku 26–33 lat (8 mężczyzn i 3 kobiety) z otologicznie zdrowym słuchem (ubytki nie większe niż 5 dB) i bez odnotowanych wcześniej schorzeń układu słuchowego. Wszystkie miały doświadczenie w badaniach odsłuchowych (w tym badaniach nad jakością dźwięku emitowanego przez DAB+) i znały cel eksperymentu. Osiem osób pracowało czynnie jako realizatorzy dźwięku w rozgłośniach radiowych lub na potrzeby fonografii, pozostałe trzy były związane z branżą dźwiękową. Pojedynczy ciąg testowy trwał ok. 20 min, potem następowała przerwa 5-minutowa. Całkowity czas badania to ok. 2 godz.

Testy odsłuchowe przeprowadzono w studiu nagrań Katedry Akustyki i Multimediów Politechniki Wrocławskiej – pomieszczeniu spełniającym zalecenia dotyczące pomieszczeń odsłuchowych. Jego dodatkowym atutem było to, że wszyscy uczestnicy badań je znali, co nie wpływało na ich dekoncentrację związaną z ewentualną adaptacją do miejsca badań. Każdy po wysłuchaniu ciągu uczącego, a w dalszej kolejności – próbki musiał wystawić odpowiadającą swoim odczuciom ocenę w skali 1–5. Badanie odbyło się jednocześnie dla całej grupy, a słuchacze zapisywali swoje odpowiedzi na arkuszu testowym.

Subiektywną ocenę jakości mowy wystawiła ekipa odsłuchowa złożona z innych osób niż w badaniach z wykorzystaniem muzyki – stanowiło ją 30 osób wybranych zgodnie z zaleceniami ITU-T P.800 [13]. Wiek słuchaczy mieścił się w przedziale 18–25 lat, byli to głównie studenci Politechniki Wrocławskiej. W tabeli 1 podano zestawienie danych grupy odsłuchowej.

Tabela 1. Parametry grupy odsłuchowej

| Parametry | Grupa |
|---|---|
| Miejsce | Pracownia AiPSA |
| Liczebność | 30 |
| Wiek | 18–25 |
| Płeć | kobiety – 10, mężczyźni – 20 |



Rys. 3. Stanowisko do prezentacji list zdaniowych

Odsłuch wykonano w pomieszczeniu Pracowni Analizy i Przetwarzania Sygnałów Akustycznych spełniającym wymogi zalecenia ITU-T P.800 [13] odnośnie do poziomu szumów pomieszczenia, a także czasu pogłosu. Sygnały testowe (zdania) prezentowane

były słuchaczom za pomocą zestawu głośnikowego o pasmie przenoszenia 50 Hz–20 kHz ±1 dB (rys. 3). Procedura pomiarowa była taka jak w ocenie muzyki. Słuchacze podawali swoją ocenę jakości mowy po usłyszeniu grupy zdań na przygotowanych wcześniej arkuszach ocen – przed przystąpieniem do właściwych pomiarów byli oni informowani o zasadach oceniania, a także odbyli jedną sesję treningową.

## 8.3. Wyniki badań

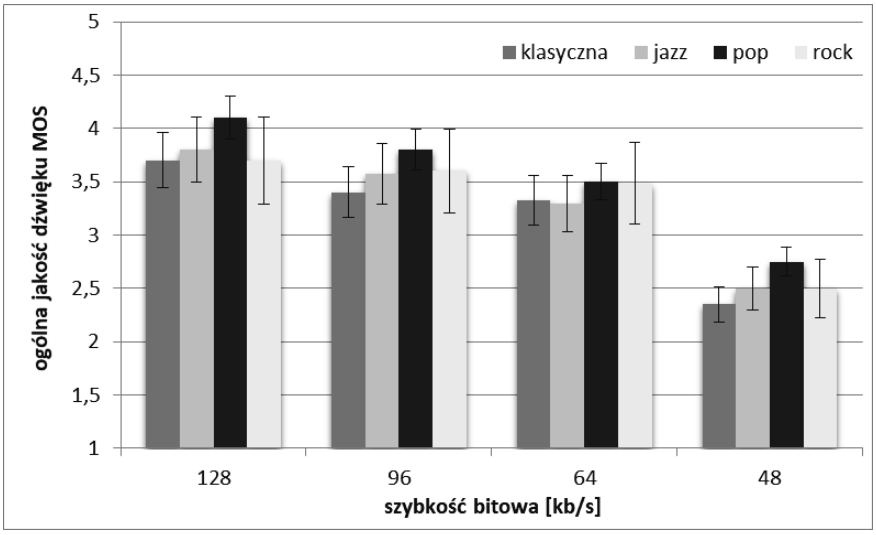
### 8.3.1. Ogólna ocena jakości muzyki

Analizę wyników wykonano najpierw globalnie bez podziału na oceniane gatunki muzyki. W przypadku każdej szybkości bitowej wyznaczono średnią ocenę dla wszystkich gatunków muzyki (bez ich rozróżnienia) i wszystkich słuchaczy. Uśredniono wyniki uzyskane dla poszczególnych szybkości bitowych i 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 $\alpha = 0{,}05$ ($\chi^2 = 4{,}22 < \chi^2_{kr} = 5{,}98$), co pozwoliło na uśrednienie ocen wszystkich członków ekipy odsłuchowej. Za pomocą testu ANOVA na poziomie istotności $\alpha = 0{,}05$ wykazano brak wpływu miejsca odbioru sygnału radiowego ($F = 1{,}39 < F_{kr} = 3{,}68$) na ocenę jakości. Okazało się natomiast, że gatunek muzyki ma wpływ na ogólna ocenę jakości dźwięku ($F = 6{,}53 > F_{kr} = 5{,}16$).

Otrzymane uśrednione wartości subiektywnej oceny jakości MOS badanych fragmentów ze wszystkich trzech punktów pomiarowych wraz z wartościami odchyleń standardowych przedstawiono na rys. 4. 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. Na podstawie analizy otrzymanych wyników można stwierdzić, że począwszy od szybkości bitowej 64 kb/s, wartość wskaźnika MOS przekracza 3,0, a począwszy od 96 kb/s – ocena jest porównywalna z oceną jakości nagrań oryginalnych zamieszczonych na płytach CD.

Jeśli przyjmie się, że akceptowalna jakość jest uzyskiwana w przypadku oceny MOS powyżej 3,0, to można zauważyć, że szybkość bitowa 48 kb/s nie zapewnia żadnemu rodzajowi muzyki wystarczającej jakości odbioru. W badaniach wykazano, że oceny




Rys. 4. Wyniki oceny jakości muzyki dla zastosowanych gatunków

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 odnośnie do dynamiki poszczególnych grup instrumentów, a także naturalności dźwięku obowiązującymi w przypadku muzyki poważnej oraz jazzu wykonywanego na instrumentach akustycznych. Dodatkowo stwierdzono, że dla wszystkich badanych gatunków muzycznych zależności jakości od przepływności są monotoniczne, co umożliwia oszacowanie minimalnej wartości przepływności niezbędnej do uzyskania jakości dźwięku na założonym przez określonego nadawcę poziomie [10]. 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 w jednakowy sposób [16, 17]. Często też słuchacze poznają dany materiał dźwiękowy w czasie słuchania określonej stacji radiowej.

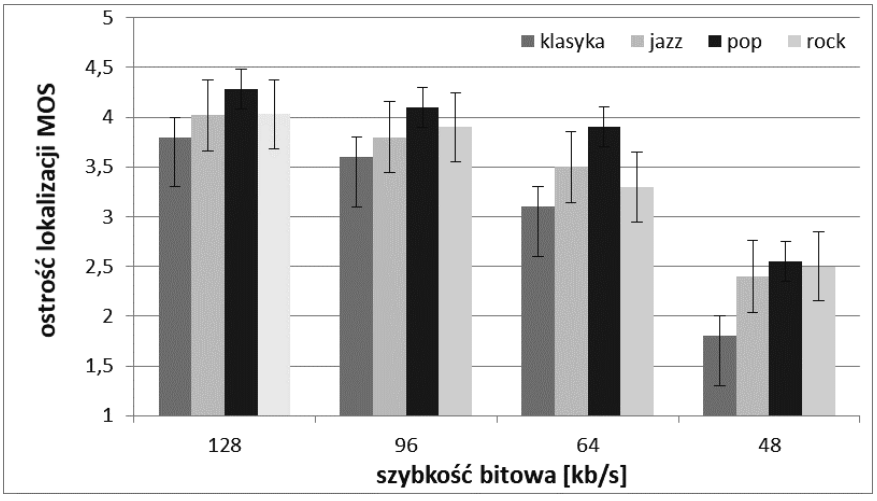
Reasumując: dla szybkości bitowej większej lub równej 64 kb/s uzyskano zadowalającą ogólną jakość muzyki przesyłanej za pomocą jednoczęstotliwościowej sieci nadawczej DAB+ – to umożliwia efektywną redukcję danych przesyłanych w ramach określonego programu radiowego, a w konsekwencji całego pakietu programów adresowanych do konkretnych odbiorców.

## 8.3.2. Ocena wybranych atrybutów dźwięku sygnałów muzycznych

W celu weryfikacji istotności statystycznej wpływu szybkości bitowej na percepcję poszczególnych atrybutów dźwięku różnych gatunków muzyki przeprowadzono test jednorodności wielu wariancji (test Bartletta) z hipotezami zerowymi o braku wpływu tych czynników na zauważalność zmian poszczególnych atrybutów. W przypadku poziomu istotności $\alpha = 0{,}05$ dla każdego z przypadków uzyskano nierówność $\chi^2 > \chi^2_\alpha$ – wszystkie hipotezy zerowe zostały zatem odrzucone, co dowodzi, że zarówno szybkość bitowa, jak i gatunek transmitowanej muzyki wpływają na percepcję zmian lokalizacji źródeł pozornych w panoramie, perspektywy oraz barwy dźwięku sygnału fonicznego.

### Ostrość lokalizacji źródeł pozornych

Na rysunku 5 przedstawiono wyniki oceny ostrości lokalizacji źródeł pozornych w nagraniach muzycznych badanych gatunków. Podobnie jak w przypadku ogólnej oceny jakości do badań wykorzystano metodę ACR ze skalą 5–1, gdzie: 5 – to doskonałe odczucie lokalizacji źródeł dźwięku w nagraniu stereofonicznym, a 1 – to bardzo złe (nie do przyjęcia). Na podstawie otrzymanych wyników można stwierdzić, że ostrość

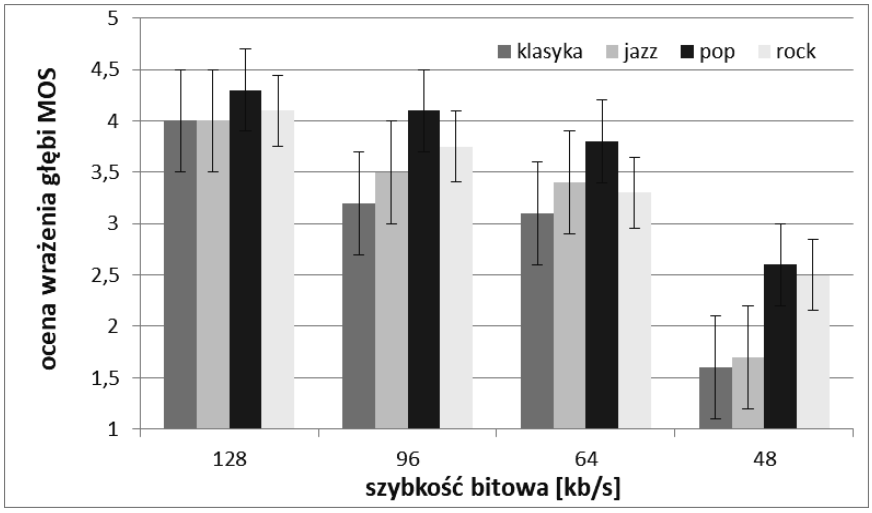



Rys. 5. Wyniki oceny ostrości lokalizacji źródeł pozornych w nagraniach muzycznych dla badanych gatunków

lokalizacji źródeł pozornych jest oceniana na zbliżonym poziomie w przypadku: przepływności 64 kb/s i jest większa dla muzyki pop oraz jazzu (różnice wynoszą 0,5 MOS), dla muzyki klasycznej oraz rockowej natomiast wpływ szybkości bitowej na ocenę lokalizacji jest większy (różnice w ocenie wynoszą ok. 0,8 MOS, przy dwukrotnym zmniejszeniu szybkości bitowej ze 128 kb/s do 64 kb/s). W przypadku szybkości bitowej 48 kb/s dla żadnego z badanych gatunków nie odnotowano oceny wyższej niż 3,0 MOS, co oznacza, że przy tej szybkości transmisji ostrość lokalizacji źródeł dźwięku w panoramie jest nieakceptowana. Można więc sformułować stwierdzenie, że transmisja sygnałów muzycznych przy tej szybkości bitowej nie zapewnia odpowiedniego poziomu wrażeń, dlatego wartość 48 kb/s nie może być stosowana do transmisji audycji, w których prezentowane są nagrania cechujące się istotnymi szczegółami w panoramie, np. nagrania muzyki klasycznej czy też słuchowiska radiowe.

## Ocena wrażenia perspektywy

Kolejnym badanym atrybutem wrażenia słuchowego było wrażenie perspektywy nagrania. Wyniki oceny zamieszczono na rys. 6. Jak można zauważyć, tendencje oceny są podobne do tych, jakie uzyskano podczas badań nad ostrością lokalizacji źródeł dźwięku w panoramie. Największą degradację wrażenia głębi wraz ze zmniejszeniem szybkości bitowej (ze 128 kb/s do 64 kb/s) odnotowano dla muzyki klasycznej (o 0,9 MOS), a najmniejszą – dla muzyki pop i rock (o ok. 0,6 MOS).

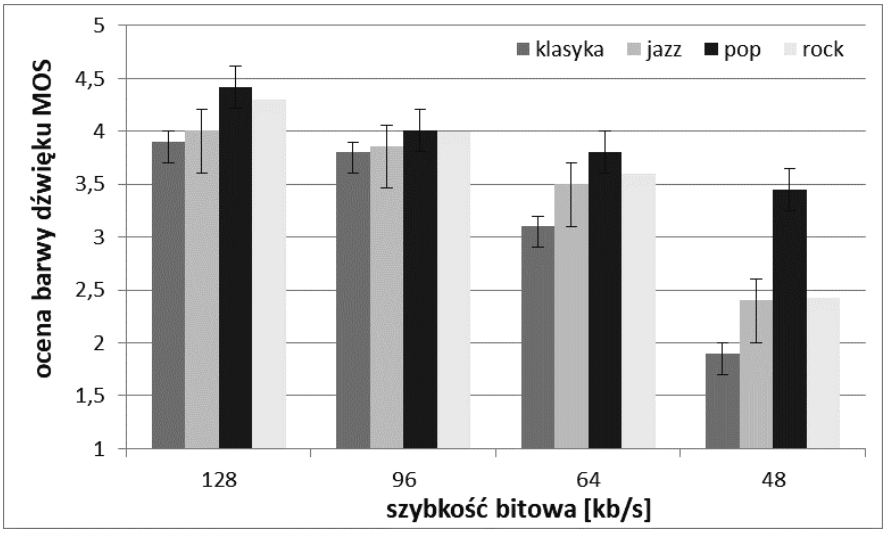



Rys. 6. Wyniki oceny wrażenia perspektywy
w nagraniach muzycznych dla badanych gatunków

Podobnie jak w przypadku lokalizacji źródeł pozornych szybkość bitowa 48 kb/s nie zapewnia wystarczającej jakości transmisji, jeśli chodzi o ocenę wrażenia głębi obrazu dźwiękowego dla każdego z badanych gatunków muzycznych, gdyż dla wszystkich badanych gatunków ocena tego atrybutu nie przekroczyła wartości 3,0 MOS. Na podstawie otrzymanych wyników można stwierdzić, że transmisja sygnałów muzycznych przy szybkości bitowej 48 kb/s nie zapewnia odpowiedniego poziomu wrażenia perspektywy nagrań muzycznych niezależnie od ocenianego gatunku.

## Ocena barwy dźwięku

Na rysunku 7 przedstawiono wyniki oceny barwy dźwięku w zależności od badanych wartości szybkości bitowej dla tych samych sygnałów i gatunków muzycznych, co poprzednio. Analogicznie do obu ocenianych atrybutów przestrzennych największą degradację przy zmniejszaniu szybkości transmisji ze 128 kb/s do 64 kb/s odnotowano dla muzyki klasycznej (spadek o 0,8 MOS), najmniejsze pogorszenie w ocenie barwy dźwięku odnotowano natomiast dla muzyki pop (spadek oceny o 0,5 MOS). Muzyka klasyczna, a także jazz są przykładami sygnałów o szerokim zakresie dynamiki – zwiększa to prawdopodobieństwo zauważenia różnicy w brzmieniu, zwłaszcza przy zmianie dynamiki. Muzyka pop na ogół nie operuje zmianami w tak szerokim zakresie, jak klasyczna, różnice w barwie nie są zatem aż tak dobrze percypowane. Należy także




Rys. 7. Wyniki oceny barwy dźwięku nagrań muzycznych
w przypadku zastosowanych gatunków w zależności od szybkości bitowej

zwrócić uwagę na znaczne zmiany w ocenie barwy dźwięku, jakie odnotowano dla próbek rockowych, sięgające 0,8 MOS, czyli podobnie jak dla muzyki klasycznej. Można to tłumaczyć tym, że muzyka rockowa charakteryzuje się szerokim widmem, dzięki czemu łatwiej jest zauważyć zmiany zachodzące w dziedzinie częstotliwości. Najczęstszym powodem wystawiania niskich ocen przy najniższej wartości przepływności dla rocka według uczestników testu było uwydatnienie małych częstotliwości i nienaturalność brzmienia najwyższych składowych.

W przypadku szybkości bitowej 48 kb/s odnotowano wartości ocen większe od 3,0 MOS jedynie dla muzyki pop (3,48 MOS) – z ocen pozostałych badanych gatunków wynika że transmisja przy tej wartości przepływności nie zapewnia odpowiedniego poziomu estetycznego w zakresie jakości barwy dźwięku transmitowanych sygnałów.

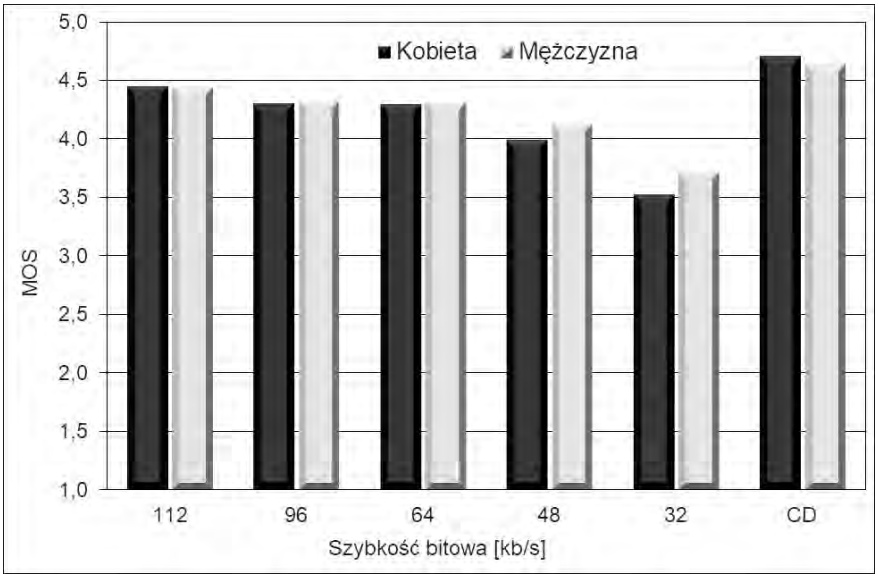
### 8.3.3. Ocena jakości mowy

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

Za pomocą testu *t*-Studenta sprawdzono, czy różnice w wartościach średnich ocen słuchaczy są statystycznie istotne: na poziomie istotności $\alpha = 0{,}05$ okazało się, że grupa odsłuchowa jest jednorodna ($t = 1{,}67 < t_{kr} = 2{,}06$) – umożliwiło to uśrednienie ocen wszystkich słuchaczy. Analogicznie do sygnałów muzycznych wykonanych z zastosowaniem testu ANOVA nie stwierdzono wpływu miejsca odbioru programu na ogólną jakość dźwięku ($F = 0{,}87 < F_{kr} = 3{,}68$). Uśrednione na podstawie pomiarów dokonanych w dziewięciu punktach Wrocławia wyniki oceny subiektywnej uzyskane dla wszystkich szybkości bitowych przedstawiono na rys. 8. Miarą odniesienia były wyniki uzyskane w przypadku list zdaniowych o jakości płyty CD (częstotliwość próbkowania – 44,1 kHz, rozdzielczość – 16 b) wypowiadanych przez kobietę i mężczyznę. Wartość oceny MOS dla jakości CD to: 4,71 – głos żeński, 4,64 – głos męski.

Po przeanalizowaniu wyników otrzymanych dla sygnałów mowy można stwierdzić, że przy przepływności 48 kb/s otrzymano satysfakcjonującą słuchaczy jakość. Dla badanych przepływności różnice w wartościach oceny w 5-stopniowej skali nie są duże: wartość MOS dla 112 kb/s wynosi 4,42, a dla 48 kb/s – 4,03, choć jest to spadek statystycznie istotny ($F = 4{,}52 > F_{kr} = 4{,}08$).

Zgodnie z zaleceniem zawartym w normie ITU-T P.800 0 dobrej jakości transmisji sygnału mowy odpowiada wartość MOS = 4,0. Można zatem stwierdzić, że dla wszystkich badanych szybkości bitowych uzyskano dobrą jakość transmisji sygnału mowy.




Rys. 8. Wyniki oceny jakości mowy uśrednione ze wszystkich badanych punktów Wrocławia

## 8.4. Omówienie wyników

Uzyskane wyniki dotyczące oceny jakości sygnałów muzycznych stanowią podstawę, by stwierdzić, że przy szybkości bitowej wynoszącej 128 kb/s wartość wskaźnika MOS jest bliska 4,0, co zapewnia dobrą jakość dźwięków muzycznych nadawanych w badanym systemie, najmniejsza przepływność gwarantująca dostateczną (akceptowalną) jakość przekazu muzycznego dla wszystkich gatunków muzycznych wynosi natomiast 64 kb/s. Należy zaznaczyć, że słuchacze oceniali jakość próbki oryginalnej z muzyką rockową podobnie jak w przypadku sygnałów emitowanych przez DAB+ (MOS 4,1), przy szybkości 128 kb/s, a próbka referencyjna z muzyką klasyczną zawsze oceniana była wyżej (MOS 4,5) w porównaniu do ocen przy przepływności 128 kb/s. Oznacza to, że w celu zapewnienia szczególnie wysokiej jakości transmisji z estetycz-

nego 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 [4, 14]. Zmiany barwy dźwięku wprowadzane na etapie kodowania nie są aż tak wyraźne i to gwarantuje wysoką ocenę MOS dla muzyki popularnej, ponieważ ten atrybut sceny dźwiękowej jest najbardziej istotnym czynnikiem w subiektywnej ocenie jakości dźwięku [27].

Jeśli wziąć pod uwagę degradacje ocen atrybutów przestrzennych: panoramy i perspektywy, to ich przyczyną może być redukcja tych składowych, które nie są słyszalne oddzielnie, a mają wpływ na połączenia pozornych źródeł dźwięku w przestrzeni. Wraz z ciągłością przestrzenną w płaszczyźnie obserwacji (poziomej lub pionowej) tworzy się swoiste kontinuum w przestrzeni dźwiękowej, gdzie zacierają się wyraźne różnice między poszczególnymi źródłami, a ich rozmiary nie są ściśle określone. Można więc to zgeneralizować na całość konstrukcji dźwiękowej, podobnie jak w przypadku przestrzeni wizualnej – jeżeli z pojedynczych form (części) utworzona jest spoista całość, to dodanie lub usunięcie elementów tworzących tę spoistość jest zawsze wyraźnie spostrzegane [27, 29]. Dlatego też zasadne wydaje się zachowanie w procesie redukcji danych lub wykreowanie na etapie produkcji nagrania swoistej aury dźwiękowej mogącej pomóc w kształtowaniu pewnych wrażeń zmysłowych na zasadzie tła, na którym można by umieścić wszystkie zdarzenia dźwiękowe wywołujące określone wrażenia. Mogłoby to zmniejszyć zależność degradacji atrybutów przestrzennych ocenianego dźwięku poddanego tego rodzaju kompresji.

Znaczne pogorszenie atrybutów przestrzennych nagrania obserwowane w przypadku muzyki klasycznej oraz jazzowej (zwłaszcza w jej odmianie akustycznej) wydaje się mieć związek ze sposobem kreowania perspektywy w nagraniach tych gatunków [23]. Reżyserzy dźwięku często stosują wielowarstwowe ujęcia mikrofonowe, co powoduje, że wrażenie głębi odpowiada wyobrażeniu o wielkości pomieszczenia, w którym dokonano nagrania. Sygnały te mogą jednak zostać zredukowane w procesie kodowania percepcyjnego przy użyciu kompresji stratnej [24] – wrażenie głębi jako czynnik o mniejszej wadze semantycznej może przez to ulec zmianie. W takim razie można zaryzykować stwierdzenie, że kompresja nagrań muzyki poważnej objawia się najczęściej zmianą wyobrażenia o pomieszczeniu, gdzie nagranie zostało dokonane. W innych gatunkach muzycznych wrażenie perspektywy jest mniej znaczące, zważywszy na stosowanie pogłosu syntetycznego (sztucznego), nieodbiegającego brzmieniowo od naturalnego już od wielu lat [25, 26], stąd też nagrania pop oraz rockowe charakteryzują się mniejszym pogłosem, na dodatek bar-

dziej spójnym z sygnałami bezpośrednimi pochodzącymi od źródeł naturalnych [14, 21, 27].

Zgodnie z wynikami z przeprowadzonych badań można wywnioskować, że zmiany słyszalne po kompresji nie zachodzą wyłącznie w obrębie jakiegoś jednego atrybutu sceny dźwiękowej. Jeśli zostało wykryte pogorszenie barwy, to według słuchaczy również zmieniły się atrybuty przestrzenne – zgodnie z regułą percepcji i interpretacji bodźców złożonych [16]. Pogorszeniu barwy w ok. 90% towarzyszyło zawężenie panoramy, a także zmniejszenie wrażenia głębi sygnału [21].

Okazuje się jednak, że opisane artefakty związane ze zmniejszeniem szybkości bitowej, a tym samym z redukcją przesyłanych informacji nie są czynnikiem istotnym powodującym zmniejszenie słuchalności radia cyfrowego. Z literatury przedmiotu (m.in. publikacje: [8, 10, 17, 30]) wynika, że ponad 60% słuchaczy ocenia jakość programów nadawanych w systemie DAB+ przy szybkości bitowej 64 kb/s jako dobrą lub bardzo dobrą, a 23% jako akceptowalną. Może to oznaczać, że słuchacze większą uwagę przywiązują do samej zawartości merytorycznej programów radiowych (informacje, komunikaty drogowe i pogodowe) przy określonym profilu stacji niż do oceny estetycznej nadawanej muzyki. Pozwala to nadawcom na spore oszczędności w zarządzaniu widmem i zasobami cyfrowymi.

W ocenie wyników otrzymanych w przypadku sygnałów mowy można stwierdzić, że przy przepływności 48 kb/s otrzymano zadowalającą słuchaczy jakość badanych przepływności, różnice w wartościach oceny MOS nie są duże – wartość MOS dla 128 kb/s to 4,42, a dla 48 kb/s – 4,03. W odniesieniu do jakości CD, dla której wartość MOS wynosi 4,71 dla głosu żeńskiego i 4,64 dla głosu męskiego, począwszy od szybkości bitowej 64 kb/s różnica jest bardzo mała – odpowiednio 0,4 i 0,33. Zgodnie z zaleceniem zawartym w normie ITU-T P.800 [13] dobrej jakości transmisji sygnału mowy odpowiada wartość MOS = 4,0. Oznacza to, że jakość słownych audycji radiowych nadawanych w systemie DAB+ będzie oceniana przez słuchaczy jako dobra dla szybkości bitowych od 48 kb/s.

Jeśli porówna się wyniki badań ogólnej jakości sygnałów mowy oraz sygnałów muzycznych, dla których sygnały o mniejszej złożoności zostały ocenione wyżej niż próbki muzyczne, można postawić hipotezę, że istnieje związek między rodzajem sygnału a oceną zmian wprowadzonych przez kompresję stratną stosowaną w radiofonii DAB+. Dzięki temu jest możliwe przyporządkowanie wartości szybkości bitowych do konkretnych stacji o określonym profilu czy wręcz wybranych programów nadawanych przez poszczególne rozgłośnie [10], dlatego gospodarowanie przepustowością kanałów transmisji radiowej staje się bardziej efektywne [11].

## 8.5. Podsumowanie

Po analizie wyników 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 48 kb/s. W przypadku programów muzycznych minimalna wartość przepływności zapewniająca akceptowalną jakość to 64 kb/s. Wówczas, gdy szczególnie istotna jest warstwa estetyczna nagrań (np. dla muzyki klasycznej), przepływność powinna wynosić 96 kb/s lub nawet 128 kb/s, co jest zgodne z wynikami wcześniejszych badań nad jakością dźwięku kodowanego za pomocą różnych kodeków [20, 22].

Przedstawione wyniki oceny jakości dotyczą sesji odsłuchowych przeprowadzonych w warunkach studyjnych z dużym odstępem od dźwięków zakłócających. Z tego powodu w Internecie najczęściej spotykanymi zasobami muzycznymi są pliki w formacie MP3 o przepływnościach 64 kb/s, 96 kb/s i 128 kb/s [24]. Niezwykle istotny jest wpływ warunków odsłuchowych audycji radiowych, w tym sprzętu odsłuchowego leżących u podstaw oceny jakości programów radiowych przez szerokie grono słuchaczy – nie wszystko to, co wyraźnie da się usłyszeć na profesjonalnych monitorach, będzie w równym stopniu percypowane w warunkach odsłuchu domowego [21].

Na podstawie analizy otrzymanych wyników można potwierdzić wnioski uzyskane w badaniach symulacyjnych opartych na programowym kodowaniu AAC [2], a także na ocenie audycji słownych nadawanych na żywo [16]. 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 nieznacząco się różnią od nadawania sygnału DAB+ za pomocą jednego nadajnika [6, 14, 15]. Można zatem stwierdzić, że sposób emisji (sieć jednoczęstotliwościowa vs. jeden nadajnik) nie wpływa istotnie na ocenę jakości dźwięku transmitowanego w systemie DAB+.

**Słowa kluczowe:** *radiofonia cyfrowa, jakość dźwięku, badania subiektywne.*

## Bibliografia

[1] Brachmański S., *Wybrane zagadnienia oceny jakości transmisji sygnału mowy*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 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, 2019.

[3] Brachmański S., Dobrucki A., Rurzyńska N., Zemankiewicz P., *Jakość sygnału mowy emitowanego w lokalnej radiofonii cyfrowej w wybranych punktach Wrocławia*, Krajowa Konferencja Radiokomunikacji, Radiofonii i Telewizji KKRRiT'2020, Łódź 2020.

[4] Brachmański S., Kin M., *Quality evaluation of sound broadcasted via DAB+ system based on a single frequency network*, 144th Convention of Audio Eng. Society, Milan, Italy, 2018, Convention paper 10004.

[5] 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].

[6] 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), s. 488–491.

[7] EBU, *Listening Conditions for the Assessment of Sound Programme Material*, Technical Recommendation R22-1999, EBU, Geneva, Switzerland, 1999.

[8] Falkowski-Gilski P., Brachmański S., *Subiektywna ocena jakości sygnałów mowy i muzyki emitowanych w lokalnych multipleksach radiofonii DAB+ w Gdańsku i Wrocławiu*, Krajowa Konferencja Radiokomunikacji, Radiofonii i Telewizji KKRRiT'2020, Łódź 2020.

[9] Gilski P., *DAB vs DAB+ radio broadcasting: a subjective comparative study*, „Archives of Acoustics" 2017, 42(4), s. 715–723.

[10] Gilski P., Stefański J., *Subjective and objective comparative study of DAB+ broadcast system* „Archives of Acoustics" 2017, 42(1), s. 3–11.

[11] Hoegh W., Lauterbach T., *Digital Audio Broadcasting*, Wiley, England, 2003.

[12] ITU-R Recommendation BS-1534, *Method for the subjective assessment of intermediate quality level of coding systems*, 2001–2003.

[13] ITU-T Recommendation P.800, *Method for subjective determination of transmission quality*, 1996.

[14] Kin M., *Subiektywna ocena jakości nagrań muzycznych nadawanych w systemie DAB+*, „Przegląd Telekomunikacyjny. Wiadomości Telekomunikacyjne" 2013, R. 86(6), s. 494–497.

[15] 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.

[16] Kin M., Brachmański S., *Quality assessment of musical and speech signals broadcasted via single frequency network DAB*, „International Journal of Electronics and Telecommunications" 2020, Vol. 66, No. 1, s. 139–144.

[17] McGregor I., Cunningham S., *Comparative Evaluation Of Radio And Audio Logo Sound Design*, „J. Audio Eng. Soc." 2015, 63(11), s. 876–887.

[18] Niewiadomski D., Michniewicz R., Sobolewski J., Więcek D., *Planowanie lokalnej sieci jednoczęstotliwościowej radiofonii cyfrowej DAB+*, „Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne" 2017, 6, s. 513–516.

[19] Pawłowski T., *Subiektywizm*, [w:] *Wybór pism estetycznych*, G. Sztabiński (red.), Universitas, Kraków 1987, s. 103–140.

[20] Prygoń S., Kin M., *Badania zauważalności zmian sceny dźwiękowej sygnałów poddanych różnym rodzajom kompresji*, Mat. XVII Symp. Inżynierii i Reżyserii Dźwięku ISSET 2017, Warszawa.

[21] Prygoń S., Kin M., *Ocena wybranych atrybutów sceny dźwiękowej sygnałów poddanych różnym rodzajom kompresji*, [w:] *Aspekty komputerowej inżynierii dźwięku. Od metafory do standaryzacji*, S. Brachmański, A. Miśkiewicz, P. Plaskota (red.), Wyd. JAKOPOL, Wrocław 2017, s. 39–47.

[22] Rogowska A., *Czy słyszymy kompresję?*, Mat. XVI Symp. Inżynierii i Reżyserii Dźwięku ISSET 2015,Warszawa, s. 140–145.
[23] Rumsey F., *Telling the difference – preference and prediction*, „J. Audio Eng. Soc." 2020, 68(10), s. 774–778.
[24] Sayood K., *Introduction to Data Compression*, Elsevier, Waltham, MA, 2012, s. 6386.
[25] Schroeder M.R., Logan B.F., *Colorless artificial reverberation*, IRE Transactions on Audio, 1961, Vol. AU-9, No. 6, s. 209–214.
[26] Schroeder M.R., *Natural sounding artificial reverberation*, „J. Audio Eng. Soc." 1962, 10(3), s. 219–223.
[27] 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), s. 712–723.
[28] Zieliński R.J., *Warunki uzyskania poprawnego odbioru sygnału DAB+ w sieci jednoczęstotliwościowej*, „Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne" 2017, 6, s. 509–512.
[29] Żórawski J., *O budowie formy architektonicznej*, [w:] *Wybór pism estetycznych*, D. Juruś (red.), Universitas, Kraków 2008, s. 86–95.
[30] 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].
[31] www.polskieradio.pl/240,Cyfrowe-radio-DAB/4698,zasieg [dostęp: 7.12.2020].

# 9. Subiektywny pomiar jakości sygnałów mowy i muzyki w lokalnych multipleksach radiofonii DAB+ w Gdańsku i Wrocławiu

PRZEMYSŁAW FALKOWSKI-GILSKI[1], STEFAN BRACHMAŃSKI[2]

[1] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk

[2] Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

Radiofonia cyfrowa DAB+ (*Digital Audio Broadcasting plus*) dostępna jest dla słuchaczy w Polsce od 2013 r. Standard ten oferuje szerokie możliwości konfiguracji multipleksów lokalnych nie tylko pod względem liczby, lecz także 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 – *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. Opisano rezultaty 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).

## 9.1. Wstęp

Radiofonia cyfrowa DAB+ (*Digital Audio Broadcasting plus*) jest jednym z najpopularniejszych standardów radiodyfuzji cyfrowej na świecie [8, 9, 15, 23]. Na kontynencie europejskim, a szczególnie w krajach członkowskich Unii Europejskiej, system ten uważany jest za wiodący, a określa się go 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 publikacjach [4, 10, 25].

W Polsce pierwsze regularne emisje DAB+ rozpoczęto w październiku 2013 r. w aglomeracji warszawskiej oraz śląskiej. Obecnie w zasięgu cyfrowego multipleksu znajduje się ponad 60% ludności Polski [27]. Do końca 2021 r. zaplanowane jest pokrycie ok. 64% obszaru Polski, a tym samym rozszerzenie zasięgu dla ok. 82% ludności [5]. Wraz z rozwojem ogólnopolskiej sieci nadajników DAB+ prowadzone są prace z wykorzystaniem lokalnych wariantów multipleksu.

Obecnie tylko Norwegia zrealizowała przejście (tzw. *switchover*) z nadawania w technice analogowej FM na cyfrową DAB+. Inne państwa (w szczególności w Europie) wciąż dokonują modyfikacji. Wielu nadawców publicznych i prywatnych emituje ten sam materiał równolegle w technice analogowej i cyfrowej (tzw. *simulcast*).

Warto zaznaczyć, że skoro słuchacze odbierają te same programy radiowe za pomocą odbiorników kompatybilnych ze standardem FM i DAB+ – punktem wyjściowym w ich subiektywnej ocenie jakości nadawanego materiału staje się radiofonia analogowa FM znana od wielu lat. W pracy opisano wyniki badań subiektywnych obejmujących programy radiowe nadawane jednocześnie w obu standardach oraz przybliżono proces rekonfiguracji naziemnego multipleksu. Opisane wyniki mogą stanowić wskazówkę dla wielu podmiotów zarówno tych obecnych, jak i nowych, zainteresowanych procesem cyfryzacji.

## 9.2. Przegląd literatury przedmiotu

W pracy [30] przedstawiono przebieg implementacji radiofonii cyfrowej DAB+ na terenie Czech. Opisano proces planowania sieci, w tym analizy zasięgowe multipleksu,

aż po ocenę końcową po stronie słuchaczy. Przedstawiono wyniki badań dotyczących różnych wariantów kodeka AAC (Advanced Audio Coding), a także szeregu szybkości bitowych.

Z kolei w publikacji [26] naukowcy opisali przebadaną jakość sygnałów mowy oraz muzyki z wykorzystaniem obiektywnej metryki PEAQ (Perceptual Evaluation of Audio Quality) [17]. Testy obejmowały próbki kodowane z szybkością 24–256 kb/s (dla standardu DAB+) oraz 56–320 kb/s (dla standardu DAB). Autorzy porównali otrzymane wyniki dla tych samych próbek przetworzonych z użyciem kodeka AAC (DAB+) oraz MP2 (DAB).

W Szwecji [1] natomiast porównano jakość programów radiowych emitowanych w standardzie FM oraz DAB+. Zgodnie z pierwszym scenariuszem (bez sygnałów odniesienia) przesyłano próbki fonii kodowane z szybkością bitową 96–192 kb/s, a z drugim – wykorzystano materiał kodowany z szybkością 48–192 kb/s i porównano go do analogicznego sygnału dostępnego w standardzie FM. Badania subiektywne przeprowadzono w 100-stopniowej skali MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) [18].

Międzynarodowa grupa autorów [24] skupiła się na badaniu wpływu różnych kodeków wykorzystywanych m.in. w naziemnej radiofonii oraz usługach strumieniowych w sieci na końcową jakość materiału fonicznego. Próbki sygnałów kodowane były z szybkością bitową 24–320 kb/s. Badania obejmowały testy subiektywne i obiektywne z użyciem metryki PEAQ oraz POLQA (Perceptual Objective Listening Quality Assessment) [22].

Prace własne autora z Politechniki Gdańskiej obejmujące m.in. badania subiektywne w skali MOS (Mean Opinion Score) oraz obiektywne z użyciem metryki ViSQOLAudio (Virtual Speech Quality Objective Listener Audio) [14] zostały opisane w publikacjach [12, 13]. Dotyczyły one i próbek sygnałów dźwiękowych przetworzonych z użyciem kodeka AAC, kodowanych z szybkością bitową 64–160 kb/s, i transmitowanych na żywo programów radiowych nadawanych w standardzie FM, DAB oraz DAB+ na terenie aglomeracji gdańskiej. Badania obejmowały regionalny multipleks DAB+ w Gdańsku oraz własny emitujący programy radiowe w standardzie DAB oraz DAB+ skonstruowany na potrzeby testów.

W niedawno opublikowanym artykule [6] grupa autorów z Wrocławia opisała badania obiektywne z użyciem metryki PESQ (Perceptual Evaluation of Speech Quality) [20, 21] i subiektywne z wykorzystaniem metryki MOS obejmujące próbki mowy oraz muzyki transmitowane na żywo. Materiał dźwiękowy kodowano z szybkością bitową 24–128 kb/s. Testy przeprowadzono w regionalnym multipleksie DAB+ we Wrocławiu.

## 9.3. Konfiguracja multipleksów

Lokalny multipleks w Gdańsku uruchomiono w 2015 r. z przeznaczeniem dla aglomeracji trójmiejskiej – to regionalny wariant ogólnopolskiego multipleksu. Sygnał emitowany jest z jednego nadajnika zlokalizowanego w Chwaszczynie. Zawartość programową oraz konfigurację multipleksu przedstawiono w tab. 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 dziesięć to programy radiowe emitujące sygnały audio (mowa i/lub muzyka), a dwa to programy emitujące wyłącznie dane. Programy emitujące dane (Dane i Journaline) dostępne są z szybkością bitową 16 kb/s. W przypadku programów dźwiękowych, tzw. klasycznych programów radiowych, szybkość bitowa oscyluje 64–128 kb/s.

Jednoczęstotliwościowy multipleks we Wrocławiu będący pionierskim multipleksem w Polsce uruchomiono w 2018 r. Sygnał emitowany jest z sieci trzech nadajników rozmieszczonych na terenie Polskiego Radia Wrocław S.A., Instytutu Łączności PIB we Wrocławiu oraz Politechniki Wrocławskiej. Na potrzeby badań uruchomiono multipleks emitujący pięć programów radiowych, których szybkość bitowa wynosiła 32–112 kb/s, transmitujących sygnały mowy wypowiadane przez kobietę i mężczyznę. Zawartość programową oraz konfigurację multipleksu przedstawiono w tab. 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 i muzycznych. W pracy specjalny nacisk położono zarówno na ocenę jakości programów w standardzie DAB+ w porównaniu do analogicznej oferty emitowanej w standardzie FM, jak i na ocenę jakości treści emitowanych w pionierskiej sieci jednoczęstotliwościowej [28, 29].

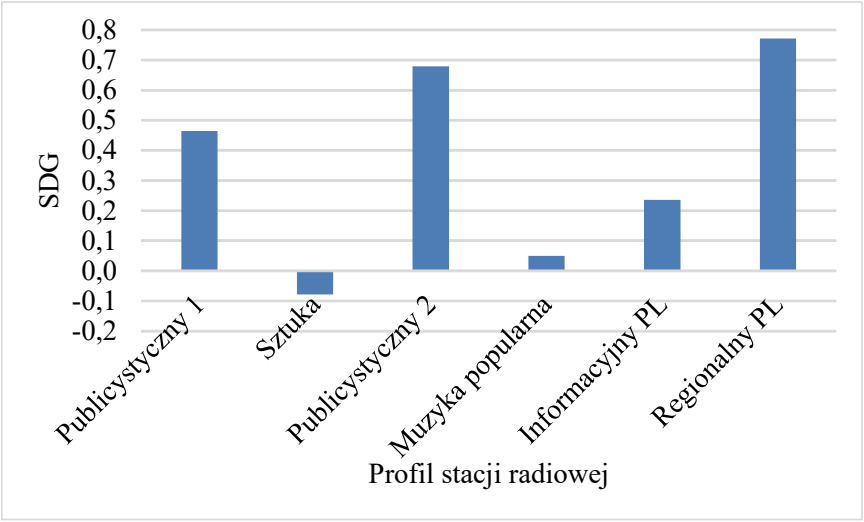
## 9.4. Badania odsłuchowe

Badania odsłuchowe mogą być oparte na kryterium jakościowym i/lub zrozumiałościowym. Testy subiektywne przeprowadzono zgodnie z zaleceniami międzynarodowymi ITU (International Telecommunication Union) oraz EBU (European Broadcast Union) i krajowymi [3, 7, 16, 19] z wykorzystaniem 5-stopniowej skali MOS w wariancie ACR (Absolute Category Rating). W tym podejściu słuchacz dokonuje oceny w odniesieniu do własnego wzorca utworzonego na podstawie wcześniejszych doświadczeń i preferencji – podaje wartość od 1 (jakość zła) do 5 (jakość doskonała).

### 9.4.1. Subiektywne testy multipleksu DAB+ w Gdańsku

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 z wykorzystaniem programów nadawanych jednocześnie w obu technikach oraz wszystkich programów dostępnych w naziemnym multipleksie.

Obecnie w tzw. *simulcascie* spośród dziesięciu regularnych emisji dostępnych jest sześć programów radiowych. Testy wykonano w środowisku wewnątrzbudynkowym z użyciem wysokiej klasy odbiornika radiowego i słuchawek zamkniętych. Zadaniem uczestników była ocena ogólnej jakości programów nadawanych na żywo. Słuchacze najpierw oceniali programy DAB+ przez ok. 10–20 s odseparowane kilkusekundową przerwą, a następnie programy 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

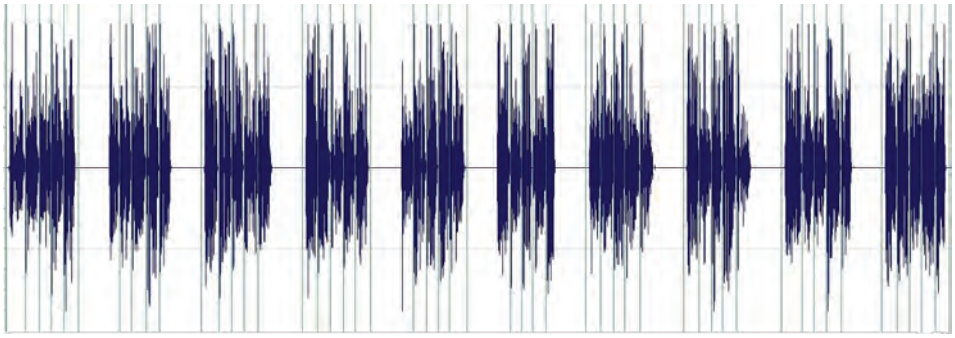
Uzyskane wyniki zostały poddane analizie statystycznej z wykorzystaniem metody ANOVA (Analysis of Variance). Przedział ufności ustawiono na 95% ($\alpha = 0{,}05$). We wszystkich przypadkach dyspersja była mniejsza niż 10% wartości średniej. Z uwagi na przejrzystość nie zaznaczono jej na wykresach.

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 języku 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.

## 9.4.2. Subiektywne testy Multipleksu DAB+ we Wrocławiu

Badania subiektywne dotyczące oceny jakości programów radiowych emitujących sygnały mowy w technologii cyfrowej DAB+ na terenie Wrocławia wykonano na podstawie materiału testowego składającego się z 10 list zdaniowych opracowanych w Pracowni Analizy i Przetwarzania Sygnałów Akustycznych Katedry Akustyki, Multimediów i Przetwarzania Sygnałów Politechniki Wrocławskiej. Każda lista zawierała po dziesięć grup składających się z pięciu zdań. Grupy oddzielone były 5-sekundowymi okresami ciszy (rys. 2). Czas trwania zestawu testowego wynosił ok. 2 min, w przypadku pojedynczej grupy natomiast 10–11 s dla głosów żeńskich i 8–9 s dla głosów męskich.



Rys. 2. Przebieg czasowy zestawu testowego emitowanego przez LokalDAB we Wrocławiu

Listy testowe zbadano pod względem zrównoważenia fonetycznego. Dla każdej listy zdaniowej obliczono częstości występowania poszczególnych fonemów wg transkrypcji IPA (International Phonetic Alphabet) [2]. Częstości występowania fonemów wyliczano jako stosunek liczby wystąpień danego fonemu w analizowanej liście do wszystkich fonemów z tej listy. Tak obliczone częstości występowania fonemów w liście zdaniowej porównano z częstościami występowania fonemów w języku polskim.

W celu postawienia hipotezy o zgodności częstości występowania fonemów na listach testowych z językiem polskim zastosowano test *t*-Studenta. Stwierdzono, że na założonym poziomie trafności $\alpha = 0{,}3$ są podstawy, by przyjąć następującą hipotezę: Częstości występowania fonemów podane dla języka polskiego i obliczone dla poszczególnych list pochodzą z tej samej populacji ogólnej.

Wypowiadane przez kobietę i mężczyznę listy zdaniowe [2] emitowane były w różnych kanałach z pięcioma szybkościami bitowymi: 32 kb/s, 48 kb/s, 64 kb/s, 96 kb/s, 112 kb/s. Przykładową listę zdaniową przedstawiono w tab. 3. Sygnały testowe zostały zareje-

strowane w dziewięciu punktach pomiarowych, równomiernie rozlokowanych na terenie Wrocławia, za pomocą odbiornika wysokiej klasy (do celów późniejszej analizy).

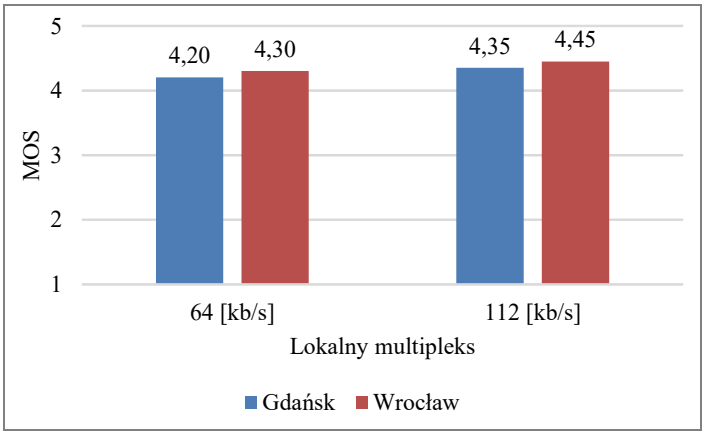
Tabela 3. Przykładowa lista zdaniowa emitowana przez LokalDAB we Wrocławiu [2]

| Grupa | Zdania | Grupa | Zdania |
|---|---|---|---|
| 1. | 1. Schroniliśmy się w bunkrze.<br>2. Zobacz jaki obrzydliwy pająk.<br>3. Dostałem zaproszenie od wujka.<br>4. Oglądałem rytuał tubylców.<br>5. On się czasem nie kontroluje. | 6. | 1. Wzięliśmy udział w paradzie.<br>2. Napadło ich trzech zbirów.<br>3. Wkrótce ujrzeliśmy nasze miasto.<br>4. Kaloryfery już nie grzeją.<br>5. Wszyscy czekają na końcowy gong. |
| 2. | 1. Pierwszy wykład ma we wtorek.<br>2. Wszystkie dzieci są ładne.<br>3. Czuję się jak nastolatka.<br>4. Samoloty robią dużo hałasu.<br>5. Kręci mi się w głowie. | 7. | 1. Chcę się stąd wydostać.<br>2. Na półkach jest dużo kurzu.<br>3. Jesteś okrutną matką.<br>4. Jechaliśmy po długim stalowym moście.<br>5. Na kongresie poznałem wiele osób. |
| 3. | 1. Chciałem jej zrobić niespodziankę.<br>2. Zostanę tu do jutra.<br>3. Podaj mi śrubokręt.<br>4. Teraz wystąpi gwiazda programu.<br>5. Zobaczyłem go za oknem. | 8. | 1. Gram w szachy co niedzielę.<br>2. Woda w rzece lśniła w słońcu.<br>3. Pojechał sprowadzić pomoc.<br>4. Byliśmy na premierze sztuki.<br>5. Mam kłopoty ze zdrowiem. |
| 4. | 1. Ptaki przestraszyły się huku.<br>2. Ktoś czaił się w krzakach.<br>3. Zrób to dla swojej rodziny.<br>4. Jadę po uszczelkę do pralki.<br>5. Łódź nabierała wody. | 9. | 1. Lodówka jest zupełnie pusta.<br>2. Jestem znanym fachowcem.<br>3. Bomba zabiła setki osób.<br>4. Kiedy zrobisz to ponownie?<br>5. Straciliśmy łączność ze statkiem. |
| 5. | 1. Zjeżdżamy na boczną drogę.<br>2. W hotelu brakło miejsc.<br>3. Spotkamy się obok pomnika.<br>4. Na szyi ma amulet z kłów.<br>5. Nie wpuścili ich do baru. | 10. | 1. Prosimy wszystkich na pokład.<br>2. Los rzucił nam wyzwanie.<br>3. Staram się ze wszystkich sił.<br>4. To może być dziwne uczucie.<br>5. Nie wejdziemy tam bez biletów. |

Materiał dźwiękowy został poddany ocenie subiektywnej przez słuchaczy, podobnie jak w badaniach w Gdańsku, w 5-stopniowej skali MOS w wariancie ACR. Ekipę odsłuchową liczącą 30 słuchaczy (10 kobiet, 20 mężczyzn) w przedziale wiekowym 18–25 lat dobrano stosownie do zalecenia zawartego w normie ITU-T P.800 [19]. Odsłuch wykonano w pomieszczeniu spełniającym określone w niej wymogi odnośnie do poziomu szumów pomieszczenia i czasu pogłosu. Sygnały testowe (listy zdaniowe nagrane w poszczególnych punktach Wrocławia z analizowanymi szybkościami bitowymi) prezentowano słuchaczom za pomocą zestawu głośnikowego o pasmie przenoszenia 50 Hz–20 kHz ±1 dB. Po wykonaniu pomiarów dla każdego warunku kodowania (szyb-

kość bitowa, punkt pomiarowy) wyznaczono średnią ocenę jakości MOS i odchylenie standardowe. Pierwszym krokiem analizy statystycznej było sprawdzenie warunku $3\sigma$, czyli rozrzutu ocen jakości mowy w grupie słuchaczy. Oceny odbiegające o wartość $3\sigma$ zostały odrzucone, a obliczenia wykonano ponownie bez odrzuconych ocen.

Z wykorzystaniem testu *t*-Studenta sprawdzono, czy różnice w wartościach średnich ocen słuchaczy, a także w wartościach średnich poszczególnych punktów pomiarowych 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 oraz średnimi między punktami pomiarowymi. Wyniki uśredniono dla każdego punktu pomiarowego i warunku transmisji – podano je jako ocenę końcową w skali MOS. Porównanie wyników lokalnych multipleksów w Gdańsku i Wrocławiu w przypadku dwóch szybkości bitowych: 64 kb/s oraz 112 kb/s, związanych z szybkością bitową regularnych audycji radiowych przedstawiono na rys. 3.




Rys. 3. Porównanie wyników oceny jakości programów radiowych emitujących sygnały mowy w standardzie DAB+ dla multipleksów w Gdańsku i Wrocławiu

Wyniki testów odsłuchowych mimo przeprowadzenia ich w różnych warunkach, przy odmiennej liczebności grupy uczestników – 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 dla sygnałów mowy o szybkości bitowej zarówno 64 kb/s, jak i 112 kb/s. Na podstawie otrzymanych wyników można przyjąć, że potwierdziła się prawdziwość przyjętych podczas prac założeń oraz prawidłowość metodologii zrealizowanych badań. To może

stanowić zachętę do prowadzenia dalszych wspólnych prac, których wyniki – z pewnością – mogą być pomocne dla instytucji rządowych i pozarządowych zainteresowanych procesem cyfryzacji radiofonii naziemnej.

## 9.5. Rekonfiguracja multipleksów

W tym podrozdziale przedstawiono proces rekonfiguracji regionalnego multipleksu DAB+ w Gdańsku oraz multipleksu LokalDAB we Wrocławiu dokonany na przestrzeni lat, w tym zawartość programową, przydzielone zasoby (szybkość bitową) oraz częstotliwości nadawania.

### 9.5.1. Rekonfiguracja multipleksu DAB+ w Gdańsku

Regionalny multipleks radiofonii cyfrowej DAB+ w Gdańsku od momentu uruchomienia w 2015 r. jako lokalna oferta publicznego nadawcy dostępny był początkowo na kanale 10D (215,072 MHz), a od 2019 r. na – 5B (176,64 MHz). Dane dotyczące oferty programowej przedstawiono w tab. 4.

Tabela 4. Rekonfiguracja programowa regionalnego multipleksu DAB+ w Gdańsku

| Lp. | Profil stacji radiowej | Szybkość bitowa [kb/s] | |
|---|---|---|---|
| | | do 2019 r. | od 2019 r. |
| 1. | Publicystyczny 1. | 112 | 112 |
| 2. | Sztuka | 128 | 128 |
| 3. | Publicystyczny 2. | 112 | 112 |
| 4. | Muzyka popularna | 112/128 | 128 |
| 5. | Informacyjny w jęz. angielskim | 64 | 64 |
| 6. | Informacyjny w jęz. polskim | 64 | 64 |
| 7. | Muzyka klasyczna/świąteczna | 128 | 128 |
| 8. | Muzyka elektroniczna/popularna | 96/– | – |
| 9. | Dziecięcy | 72 | 72 |
| 10. | Regionalny w jęz. polskim | 104/112 | 112 |
| 11. | Regionalny w jęz. angielskim | – | 72 |
| 12. | Dane | 16 | 16 |
| 13. | Journaline | 16 | 16 |

Warto zaznaczyć, że niezależnie od analizowanego okresu w multipleksie pozostawała pewna pula nieobsadzonych zasobów. W przeciągu lat wnosiła ona 192–96 jednostek CU (Capacity Unit), co odpowiadało 256–128 kb/s. Zgodnie z informacją pochodzącą od nadawcy publicznego – docelowa konfiguracja multipleksu ma zawierać 12 programów radiowych, w tym dziesięć ogólnokrajowych oraz dwa regionalne [25].

W celu sprawdzenia poprawności koncepcji rekonfiguracji multipleksu w Gdańsku wraz ze zmianą zasobów (szybkości bitowej) przydzielonych poszczególnym programom radiowym dokonano porównania otrzymanych wyników z badaniami opublikowanymi w 2017 r. [13] obejmującymi te same programy radiowe, a także ten sam sprzęt do odbioru i odsłuchu. Wyniki tego zestawienia przedstawiono w tab. 5.

Tabela 5. Porównanie wyników oceny jakości regionalnego multipleksu DAB+ w Gdańsku

| Lp. | Profil stacji radiowej | Szybkość bitowa [kb/s] | | Różnica MOS |
|---|---|---|---|---|
| | | 2017 r. | 2019 r. | |
| 1. | Publicystyczny 1. | 112 | 112 | 0,14 |
| 2. | Sztuka | 128 | 128 | –0,07 |
| 3. | Publicystyczny 2. | 112 | 112 | 0,52 |
| 4. | Muzyka popularna | 112 | 128 | –0,04 |
| 5. | Regionalny w jęz. polskim | 104 | 112 | 0,25 |

W przypadku dwóch programów radiowych, w których nie wystąpiła zmiana szybkości bitowej, uzyskane oceny są na podobnym poziomie. Różnice oceny w skali MOS oscylują od 0,07 (sztuka – 128 kb/s) do 0,14 (publicystyczny 1. – 112 kb/s). Odmienne wyniki można zaobserwować w przypadku programu publicystycznego 2. (112 kb/s) – tu różnica MOS wyniosła 0,52. Powodem może być na przykład zmiana charakteru prowadzonej w tym czasie audycji radiowej lub zmiana typu emitowanego materiału

Tabela 6. Zawartość sygnałów mowy dla wybranych programów DAB+ w godzinie największej słuchalności

| Lp. | Profil stacji radiowej | Zawartość sygnałów mowy [%] |
|---|---|---|
| 1. | Publicystyczny 1. | 45 |
| 2. | Publicystyczny 2. | 33 |
| 3. | Informacyjny w jęz. angielskim | 90 |
| 4. | Informacyjny w jęz. polskim | 100 |
| 5. | Regionalny w jęz. polskim | 65 |
| 6. | Dziecięcy | 43 |

dźwiękowego. W tabeli 6 przedstawiono procentowy udział sygnałów mowy podczas audycji w tzw. godzinie największej słuchalności dla wybranych regularnych programów radiowych emitowanych w naziemnym cyfrowym standardzie DAB+.

Jeśli wziąć pod uwagę programy radiowe, dla których wystąpiła zmiana przypisanej szybkości bitowej, różnice wyniosły od –0,04 (muzyka popularna – zmiana ze 112 kb/s na 128 kb/s) do 0,25 (regionalny w języku polskim – zmiana ze 104 kb/s na 112 kb/s). W tym wypadku zmianę przypisanych zasobów należy ocenić pozytywnie.

## 9.5.2. Rekonfiguracja multipleksu DAB+ we Wrocławiu

W pionierskiej radiofonii LokalDAB we Wrocławiu uruchomionej 19 stycznia 2018 r., w pierwszym okresie stosowano szybkość bitową 80 kb/s – w późniejszym zwiększono ją do 96 kb/s. Na kolejnych etapach prac nad poprawą jakości emisji sygnałów muzyki i mowy wprowadzono zróżnicowanie szybkości bitowych audycji słownych i muzycznych. Programy typowo muzyczne były nadawane z szybkością bitową 128 kb/s, muzyczno-słowne z szybkością 96 kb/s, a słowne z szybkością 64 kb/s.

Tabela 7. Rekonfiguracja programowa regionalnego multipleksu LokalDAB we Wrocławiu

| Lp. | Nazwa stacji radiowej | Szybkość bitowa [kb/s] | |
|---|---|---|---|
| | | do 2020 r. | od 2020 r. |
| 1. | Radio Wrocław | 112 | 104 |
| 2. | Radio Wrocław Kultura | 112 | 104 |
| 3. | Radio RAM | 112 | 104 |
| 4. | Akademickie Radio LUZ | 112 | 104 |
| 5. | Radio Rodzina | 112 | 104 |
| 6. | Radio Profeto | 112 | 104 |
| 7. | Radio LEM.fm | 112 | 104 |
| 8. | Radio Mega | 112 | 104 |
| 9. | Radio Nuta | 112 | 104 |
| 10. | Disco Radio | 112 | 104 |
| 11. | Radio Piekary | – | 104 |

W 2019 roku, gdy w multipleksie LokalDAB nadawanych było dziesięć programów, wprowadzono szybkość bitową 112 kb/s dla wszystkich programów. W tym czasie

emitowane były programy: Radio Wrocław, Radio Wrocław Kultura, Radio RAM, Akademickie Radio LUZ, Radio Rodzina, Radio Profeto, Radio LEM.fm, Radio Mega, Radio Nuta oraz Disco Radio.

W 2020 roku lista programów została wzbogacona o Radio Piekary, co skutkowało zmniejszeniem szybkości bitowej do 104 kb/s. Z tą szybkością aktualnie nadawane są wszystkie programy (stacje) LokalDAB we Wrocławiu (tab. 7).

Emisja programów radiofonii cyfrowej LokalDAB realizowana jest na kanale 11A (216,928 MHz) w sieci jednoczęstotliwościowej opartej na trzech nadajnikach rozmieszczonych na masztach budynków: Polskiego Radia Wrocław S.A., Instytutu Łączności PiB we Wrocławiu oraz przy ul. Długiej, w bezpośrednim sąsiedztwie terenów Politechniki Wrocławskiej.

## 9.6. 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 do liczby oraz jakości dostępnych programów radiowych. Warto wspomnieć, że wielu nadawców, w szczególności publicznych, emituje te same treści w standardzie FM i DAB+. Wprowadza to 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 64 kb/s uznać można za górną granicę dla kodowania sygnałów mowy w przypadku audycji radiowych. Na podstawie wcześniejszych wspólnych prac dotyczących transmisji głosowych komunikatów drogowych w radiofonii DAB+ [11] wykazano, że w przypadku prostych komend głosowych szybkość bitowa może zostać obniżona nawet do poziomu 24 kb/s.

Niepodważalnie sprawa 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 i przyspieszyła sam proces. Otrzymane rezultaty stanowią potwierdzenie poprawności zarówno przyjętych założeń, jak i metodologii badań. Są też doskonałym materiałem do dyskusji – szczególnie wśród nadawców lokalnych zainteresowanych wejściem na cyfrowy multipleks. Uzyskane wyniki zachęcają także do prowa-

dzenia dalszych prac dotyczących cyfryzacji radiofonii, rozwoju cyfrowego rynku radiowego, a docelowo – przejścia z nadawania w technice analogowej FM na cyfrową DAB+.

**Słowa kluczowe:** DAB+ (*Digital Audio Broadcasting plus*), kodowanie, kompresja, przetwarzanie sygnałów, radiofonia.

## Bibliografia

[1] Berg J., Bustad C., Jonsson L., Mossberg L., Nyberg D., *Perceived audio quality of realistic FM and DAB+ radio broadcasting systems*, „Journal of the Audio Engineering Society" 2013, 61(10), s. 755–777.

[2] Brachmański S., *Materiał testowy w pomiarach jakości mowy w radiofonii DAB+*, [w:] *Postępy akustyki*, D. Bismor (red.), Polskie Towarzystwo Akustyczne, Gliwice 2017, s. 251–262.

[3] Brachmański S., *Wybrane zagadnienia oceny jakości transmisji sygnału mowy*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2015.

[4] Brachmański S., Kin M.J., *Quality evaluation of sound broadcasted via DAB+ system based on a single frequency network*, Proceedings of the 144th AES Convention, Milan, Italy, 23–26 maja 2018, 1, s. 1–6.

[5] 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: 14.06.2021].

[6] 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, s. 1–8.

[7] EBU Technical Recommendation R22, *Listening conditions for the assessment of sound programme material*, EBU, 1999.

[8] EBU Technical Report Tech 3391, *Guidelines for DAB network planning*, EBU, 2018.

[9] ETSI Technical Specification TS 102 563, *Digital Audio Broadcasting (DAB); Transport of Advanced Audio Coding (AAC) audio*, ETSI, 2010.

[10] 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), s. 321–328.

[11] Falkowski-Gilski P., Brachmański S., Dobrucki A.B., *Transmisja głosowych komunikatów drogowych w radiofonii cyfrowej DAB+*, „Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne" 2019, 6, s. 334–337.

[12] Gilski P., *DAB vs DAB+ radio broadcasting: A subjective comparative study*, „Archives of Acoustics" 2017, 42(4), s. 715–723.

[13] Gilski P., Stefański J., *Subjective and objective comparative study of DAB+ broadcast system*, „Archives of Acoustics" 2017, 42(1), s. 3–11.

[14] Hines A., Gillen E., Kelly D., Skoglund J., Kokaram A., Harte N., *ViSQOLAudio: An objective audio quality metric for low bitrate codecs*, „Journal of the Acoustical Society of America" 2015, 137, s. 449–455.

[15] Hoeg W., Lauterbach T., *Digital Audio Broadcasting: Principles and applications of DAB, DAB+ and DMB*, John Wiley & Sons, Chichester 2009.
[16] ITU Recommendation BS.1284, *General methods for the subjective assessment of sound quality*, ITU, 2003.
[17] ITU Recommendation BS.1387, *Method for objective measurements of perceived audio quality*, ITU, 2001.
[18] ITU Recommendation BS.1534-1, *Method for the subjective assessment of intermediate quality level of coding systems*, ITU, 2003.
[19] ITU Recommendation P.800, *Methods for subjective determination of transmission quality*, ITU, 1996.
[20] ITU Recommendation P.862, *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, ITU, 2001.
[21] ITU Recommendation P.862.2, *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, ITU, 2007.
[22] ITU Recommendation P.863, *Perceptual objective listening quality assessment*, ITU, 2014.
[23] Oziewicz M., *Cyfrowa radiofonia DAB/DAB+ – multimedialny system rozsiewczy*, Dolnośląska Biblioteka Cyfrowa, Wrocław 2014.
[24] Počta P., Beerends J.G., *Subjective and objective assessment of perceived audio quality of current digital audio broadcasting systems and web-casting applications*, „IEEE Transactions on Broadcasting" 2015, 61(3), s. 407–415.
[25] Rosłan-Kuhn K., *Analogowa wyspa – czy cyfrowe radio DAB+*, „Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne" 2014, 4, s. 67–70.
[26] Ulovec K., Smutny M., *Perceived audio quality analysis in digital audio broadcasting plus system based on PEAQ*, „Radioengineering" 2018, 27(1), s. 342–352.
[27] WORLD DAB, *DAB digital radio – Europe and Asia Pacific*; https://www.worlddab.org/public_document/file/1404/WorldDAB_infographic_H1_2020_6_pager_FINAL_r1.pdf?1616438263 [dostęp: 14.06.2021].
[28] 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), s. 85–92.
[29] Zieliński R., *Fade analysis in DAB+ SFN network in Wrocław*, Proceedings of the International Symposium on Electromagnetic Compatibility, Barcelona, Hiszpania, 2–6 września 2019, 1, s. 106–113.
[30] Zyka K., *The Digital Audio Broadcasting journey from the lab to listeners – the Czech Republic case study*, „Radioengineering" 2019, 28(2), s. 483–490.

# 10. Poziom głośności nagrań dźwiękowych w zależności od rodzaju nośnika

PRZEMYSŁAW PLASKOTA, MAŁGORZATA GAWLIŃSKA

Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

W niniejszym rozdziale przedstawiono porównanie wartości parametrów akustycznych związanych z poziomem głośności nagrań dźwiękowych zapisanych na różnych nośnikach i sprawdzono, czy powszechnie panująca opinia odnośnie do zwiększającego się poziomu głośności utworów jest prawdziwa. Założonym celem był 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), poziomu głośności wyrażonego w jednostce LUFS, wartości skutecznej, zakresu dynamiki DR (Dynamic Range). Wykorzystano utwory muzyczne zapisane na płycie analogowej, płycie CD, a także pochodzące z trzech różnych serwisów internetowych (streaming).

## 10.1. Wprowadzenie

W dokumentach ITU-R [3–5] i EBU [6–10] zostały opisane metody pomiaru poziomu głośności w mediach takich jak radio, telewizja czy film z podaniem konkretnych wartości, do których należy dążyć. Wytyczne te mają pomóc profesjonalistom w tworzeniu produkcji, tak by podczas ich odtwarzania nie była konieczna regulacja poziomu głośności zarówno w trakcie transmisji, jak i przełączania się między kanałami. Poziom głośności może nadal różnić się ze względu na potrzebę artystyczną lub tech-

niczną programu. W celu uwzględnienia różnorodności głośności konkretnego programu korzysta się z metody normalizacji głośności przy wykorzystaniu średniej wartości głośności. Wyrównywanie poziomu głośności dotyczy wszystkich etapów transmisji audio od pozyskania do dystrybucji i przesyłania sygnału.

W publikacjach dotyczących głośności ich autorzy skupiali się przede wszystkim na zjawisku nazwanym „wojną głośności" (ang. Loudness War), a w szczególności na sposobach ograniczenia tego zjawiska [1], lub na zagadnieniu, czy zwiększanie głośności wpływa na poprawę odbioru utworów muzycznych przez konsumentów [2].

W tym rozdziale m.in. podjęto próbę odpowiedzi na pytanie, czy zmiana sposobu kształtowania dynamiki nagrań dźwiękowych może leżeć u podstaw wyższej oceny nagrań wydanych na płytach analogowych – zwłaszcza w konfrontacji z nagraniami na płycie kompaktowej.

Jeszcze kilka lat temu do określania poziomu głośności programów korzystano z mierników mierzących wartości szczytowe sygnałów PPM (Peak Programme Meter). Ponieważ mierniki te nie uwzględniają różnic związanych z próbkowaniem sygnału, następuje rozbieżność między wartością szczytową PEAK a rzeczywistą wartością szczytową TPL (True Peak Level) występującą w nadawanym programie. Różnica ta wynika z tego, że na wyjściu przetwornika CA możliwe jest uzyskanie wyższego poziomu, niż wynikałoby to wprost z wartości zapisanej cyfrowo [4]. Aby uniknąć wartości, które zostały pominięte podczas pomiaru, a wynikały z niedokładności miernika, został wprowadzony maksymalny poziom PML (Permitted Maximum Level). Niezależnie od 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 mimo wyższej rzeczywistej wartości PEAK programu.

Problem ten rozwiązała Międzynarodowa Unia Telekomunikacyjna (ITU), która opracowała dokument ITU-R BS. 1770 [3] opisujący algorytm pomiaru poziomu głośności i rzeczywistych poziomów szczytowych programów, a także metody pomiaru. W dokumencie EBU R128 natomiast zostały zdefiniowane konkretne wartości normalizacji głośności i sposoby bramkowania sygnału, tak by lepiej dostosować głośność programów zawierających dłuższy okres ciszy lub odizolowaną mowę i nie zaniżać wartości poziomu głośności. Do tego w celu ułatwienia twórcom programów określenia poziomu głośności wprowadzono trzy parametry: poziom głośności programu, zakres poziomu głośności, rzeczywisty poziom szczytowy.

Poziom głośności programu to długoczasowa i uśredniona głośność liczona za czas trwania programu. Parametrem jest jedna liczba wyrażona w jednostce LUFS z do-

kładnością do jednego miejsca po przecinku. Określa ona uśredniony poziom głośności programu. Pomiar poziomu głośności należy wykonywać miernikiem opisanym w dokumencie ITU-R BS. 1770 [3] uzupełnionym o funkcję bramkowania. Wartość progowa bramki została ustalona na poziomie –8,0 LU, ustawiona względem wartości docelowej –23,0 LUFS, przy długości bloku co najmniej 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 z uwzględnieniem jej statystycznego rozkładu, ale bez ekstremalnych wartości. Przedstawia się go w jednostce LU. W wymienionych rekomendacjach ITU oraz EBU nie zaproponowano określonej wartości parametru, ponieważ jest ona uzależniona od progu tolerancji słuchacza czy rozkładu programów w słuchanej stacji. Parametr ma być pomocą w 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, żeby móc uwzględnić potencjalne zaniżenie 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.

## 10.2. Pomiary

Wykonano pomiar czterech parametrów: rzeczywistego poziomu szczytowego TPL (True Peak Level), poziomu 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 liczoną na podstawie 20% czasu trwania najgłośniejszej części utworu. Szczegółowy opis matematyczny mierzonych parametrów znajduje się w rekomendacji ITU [3].

Podczas badań wykorzystano pięć różnych nośników: płytę CD, płytę gramofonową, trzy serwisy zajmujące się cyfrową usługą strumieniowania muzyki: Spotify [11], Tidal [12], iTunes [13]. Do pomiarów wybrano dwa albumy. Pierwszy z nich to *Sing-Sing* M. Rodowicz wydany w 1976 r. przez wytwórnię Pronit na płycie gramofonowej, a następnie na płycie CD w Serii: Antologia Marylii Rodowicz (2012–2013) przez

Universal Music Polska, oraz *The Dark Side of the Moon* grupy Pink Floyd, album wydany w 1973 r. przez Harvest Records / Capitol Records, który jest drugim najlepiej sprzedającym się albumem w historii muzyki.

Utwory z wszystkich nośników zostały przekonwertowane do formatu bezstratnego WAVE przy częstotliwości próbkowania 44,1 kHz i zapisie na 16 b. Utwory z płyt CD przekonwertowano z formatu CD-Audio do formatu WAVE przy częstotliwości próbkowania 44,1 kHz z rozdzielczością 16 b. Do tego celu użyto programu Extract Audio Copy – dzięki niemu pliki zostały przekonwertowane z jednoczesnym sprawdzeniem zachowania poprawności podczas wykonanej konwersji: czy nie pojawiły się różnice między materiałem z płyty a kopiami w formacie WAVE przez kilkukrotny odczyt zapisu z płyty.

Utwory z płyt gramofonowych przekonwertowano do wersji cyfrowej. Pliki zostały przegrane do formatu WAVE przy częstotliwości próbkowania 44,1 kHz z rozdzielczością 16 b. Pewną trudność sprawił układ wykorzystany do konwersji płyty z wersji analogowej do wersji cyfrowej, udało się jednak wykonać kalibrację z wykorzystaniem analogowej płyty testowej (zapisana jest ona sygnałami o znanym poziomie, stąd wprost wynika możliwość kalibracji układu pomiarowego).

Utwory z serwisów streamingowych nie są dostępne na wyłączność użytkownika. Aby pozyskać materiał, który jest strumieniowany przy użyciu aplikacji komputerowych, zarejestrowano albumy przy użyciu programu Audacity [14]. Do rejestracji wykorzystano interfejs Windows WASAPI gwarantujący przechwytywanie dźwięku bezpośrednio z urządzenia bez ponownego próbkowania go. Podczas rejestracji w aplikacji klienta serwisu poziom odtwarzania ustawiony był na 100%. Wykorzystano również najwyższą jakość dźwięku oferowaną przez dany serwis, a także zadbano o wyłączenie opcji normalizacji oferowanej przez niektóre serwisy.

## Narzędzia wykorzystane do pomiarów

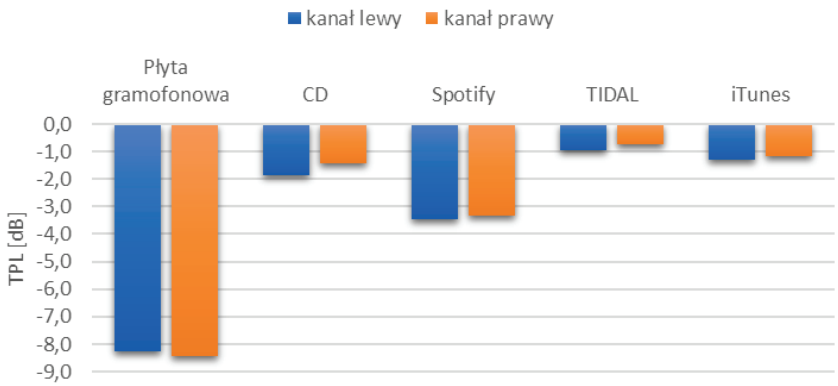
Do pomiarów czterech parametrów związanych z głośnością utworów z albumów wykorzystano programy i wtyczki. Do pomiaru wartości RMS, TPL oraz DR użyto programu TT DR Offline Meter 1.4 [15] dostępnego na licencji OpenSource.

TT DR Offline Meter został stworzony przy współpracy dwóch firm Tischmeyer Technology i Algorithmix. Służy do wykrycia kompresji utworu oraz zapobieganiu nadmiernemu wykorzystaniu jej w utworach. Ma być narzędziem, które wykona pomiar rzeczywistej wartości szczytowej, wartości skutecznej RMS i parametru DR. Na podstawie otrzymanych wyników użytkownik będzie umiał określić zakres dynamiki analizowanego nagrania.

Pomiar poziomu głośności w jednostce LUFS odbył się przy użyciu wtyczki w wersji demo iZotope Insight [16]. Insight to pakiet pomiarowy do zastosowań w postprodukcji i dystrybucji nagrań. Zapewnia on zestaw narzędzi do analizy i pomiaru dźwięku zgodnych ze standardami głośności transmisji. Podczas realizacji pomiarów i ich opisu wykorzystano dokumentację [14]. Oprogramowanie zostało przetestowane przy użyciu dostępnych testowych nagrań [5] w celu sprawdzenia miarodajności otrzymanych wyników pomiarów.

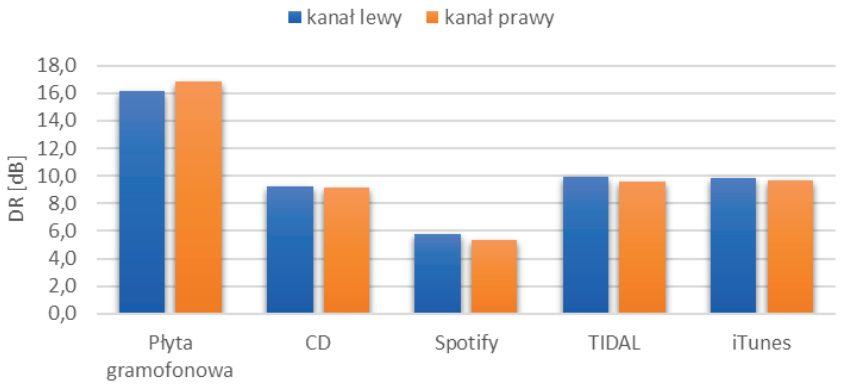
## 10.3. Wyniki pomiarów

Pomiary czterech parametrów wykonano dla pięciu różnych nośników na przykładzie dwóch wymienionych wcześniej albumów muzycznych. Na rysunku 1 zaprezentowano wyniki pomiaru parametru True Peak Level. Jak można zaobserwować, najniższe wartości szczytowe zostały zmierzone w przypadku płyty gramofonowej, kolejny najniższy wynik odnosił się do platformy Spotify, następnie płyty CD, platformy iTunes, a na końcu platformy Tidal. Mimo że Spotify, Tidal i iTunes są platformami udostępniającymi muzykę przez strumieniowanie, można zauważyć spore rozbieżności między wynikami, które otrzymano. Dla serwisu Spotify średnia wartość TPL jest o ok. 2–3 dB niższa w porównaniu do serwisów Tidal oraz iTunes. Podczas pomiarów płyty CD w programie TT DR zamiast wartości TPL pojawił się opis OVER oznaczający, że dźwięk był na granicy przesterowania. Wszystkie pomiary wykonano z uwzględnieniem całkowitego czasu trwania utworów.



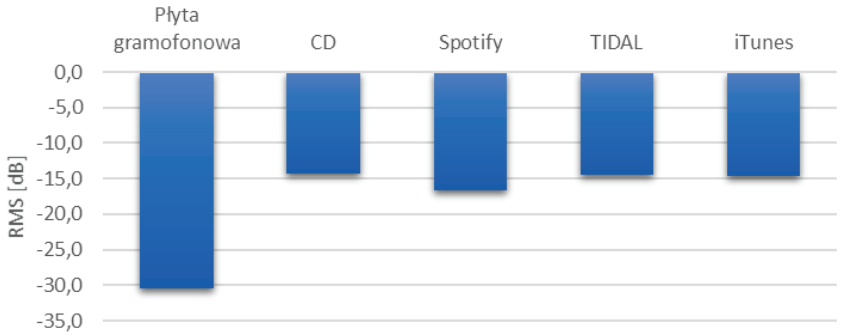

Rys. 1. Wartości średnie wyników pomiarów parametru TPL dla różnych nośników

Następnie wykonano pomiar parametru DR. Na podstawie otrzymanych wyników określono zakres dynamiki nagrania. Wyniki pomiarów przedstawiono na rys. 2 – z ich analizy wynika, że nośnik, na którym nagrania mają największy zakres dynamiki, to płyta gramofonowa. Kolejnym nośnikiem o średniej wartości parametru DR o wartości ok. 6 dB niższej od płyty gramofonowej była platforma Tidal, a następnie iTunes. Płyta CD różniła się natomiast średnimi wartościami parametru DR od płyty gramofonowej o ok. 6,5–7 dB. A najmniejszy zakres dynamiki miał materiał udostępniany w serwisie Spotify. Różnica średnich wartości parametru DR między serwisem Spotify a płytą gramofonową znalazła się w przedziale 10,0–10,5 dB.




Rys. 2. Wartości średnie wyników pomiarów parametru DR dla różnych nośników

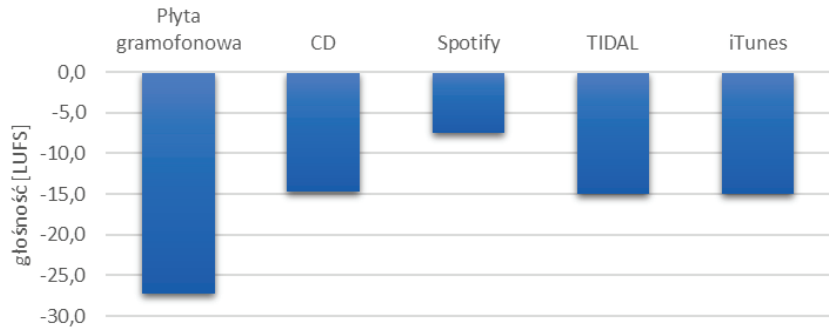
Po oszacowaniu zakresu dynamiki nagrań przystąpiono do pomiarów wartości skutecznej nagrań (RMS) – wyniki pomiarów przedstawiono na rys. 3.




Rys. 3. Wartości średnie wyników pomiarów wartości skutecznej RMS dla różnych nośników

Pomiar wartości skutecznej zrobiono w celu zestawienia otrzymanych wyników z wartościami wyników pomiarów poziomu głośności mierzonej w jednostce LUFS. W ten sposób można przeanalizować, jaki wpływ na określenie głośności ma algorytm zastosowany przez EBU do oceny średniej głośności materiału. Najmniejsza wartość średnia RMS dla całego albumu pojawia się w przypadku płyty gramofonowej. Kolejnym nośnikiem, którego wartość średnia RMS dla całego albumu jest najniższa, jest platforma Spotify, dalej – iTunes, Tidal, a na końcu płyta CD. Między platformą Spotify a iTunes różnica średnich wartości RMS wynosi ok. 2,0 dB, między platformami Tidal oraz iTunes – 0,1 dB, a między platformą Tidal oraz płytą CD – 0,1 dB. Na podstawie uzyskanych wyników można wywnioskować, że materiał udostępniany przez platformę Spotify jest najcichszy, a materiał nagrany na płytę CD – najgłośniejszy.

Następnie przystąpiono do pomiaru średniego poziomu głośności wyrażonej w jednostce LUFS. Wyniki pomiarów przedstawiono na rys. 4.




Rys. 4. Wykres wartości średnich wyników pomiarów głośności wyrażonej w jednostce LUFS dla różnych nośników

W tym przypadku również płyta gramofonowa charakteryzuje się najniższym poziomem głośności materiału dźwiękowego. Po niej kolejnym nośnikiem z najniższą głośnością jest platforma Tidal oraz iTunes, ponieważ ich średnie wartości poziomu głośności wyniosły –15,1 LUFS. Poziom głośności o 0,4 LUFS wyższy okazał się dla materiału nagranego na płycie CD. A najwyższy poziom głośności dotyczył materiału udostępnianego przez platformę Spotify. Różnica między średnim poziomem głośności odnośnie do Spotify a płytą CD wyniosła 7,1 LUFS.

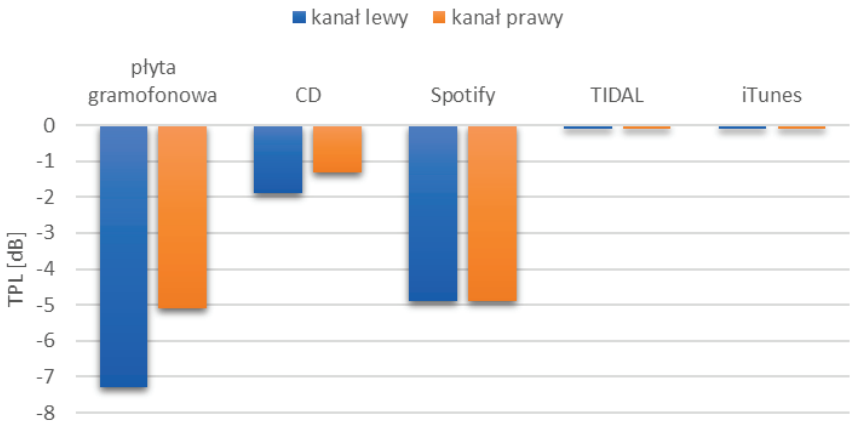
Można zauważyć, że algorytm zaproponowany przez EBU wnosi spore różnice w porównaniu do wyników uzyskanych na podstawie pomiaru wartości RMS. W przypadku pomiarów wartości skutecznej RMS platforma Spotify była najcichszym nośni-

kiem, nie biorąc pod uwagę pomiarów wykonanych dla płyty gramofonowej. Z ustalenia poziomu głośności z wykorzystaniem algorytmów EBU wynika, że platforma Spotify osiąga najwyższą głośność materiału udostępnianego użytkownikom.

Materiał umieszczony na pozostałych nośnikach przekroczył poziom docelowy, czyli wartość –23,0 LUFS. Wartości otrzymane dla płyty CD, serwisu Tidal i iTunes są bardzo podobne. Z analizy wykresu poziomu głośności w czasie otrzymanym dla serwisu Spotify można wnioskować, że materiał na tej platformie jest poddany dużej kompresji dynamiki, a głośność materiału jest na granicy przesterowania.

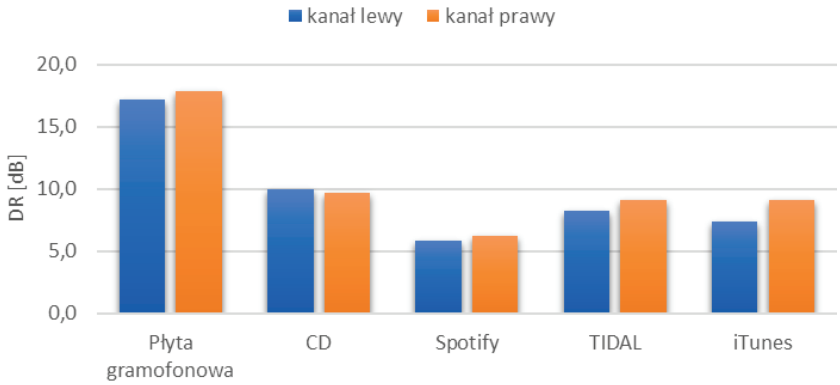
Kolejnym albumem, który został poddany pomiarom, był materiał z płyty *Sing-sing* M. Rodowicz. Zmierzono cztery parametry, tak samo jak w przypadku albumu Pink Floyd. Wyniki pomiarów zostały zamieszczone na rys. 5–8.

Na rysunku 5 przedstawiono wyniki pomiarów parametru True Peak Level (TPL) dla pięciu różnych nośników i dziesięciu utworów pochodzących z albumu.



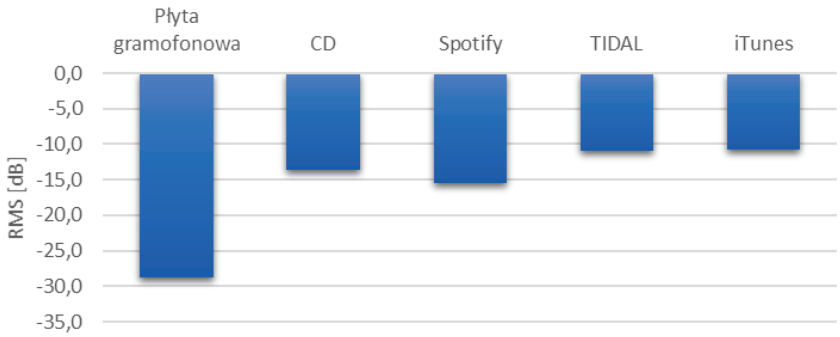

Rys. 5. Wykres wartości średnich wyników pomiarów parametru TPL dla różnych nośników

Analogicznie do albumu Pink Floyd materiał udostępniany w serwisach Tidal i iTunes ma bardzo zbliżone wartości średnie parametru TPL, a w tym przypadku dokładnie takie same – wynoszące –0,1 dB. Dla tych nośników parametr TPL osiąga najwyższe wartości. W przypadku płyty gramofonowej i materiału z albumu zanotowano najniższe wartości parametru TPL, a po niej – platforma Spotify z wartościami średnimi parametru TPL: –4,9 dB. Materiał na płycie CD jest głośniejszy od platformy Spotify. Różnica w wartości parametru TPL dla tych nośników mieści się w przedziale 3,0–3,5 dB.




Rys. 6. Wykres wartości średnich wyników pomiarów parametru DR dla różnych nośników

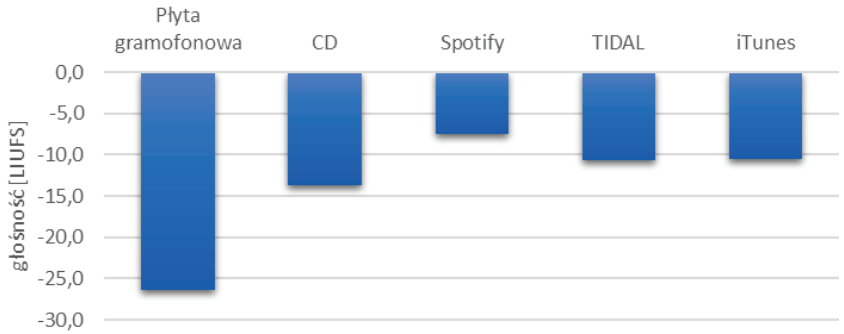
Następnie wykonano pomiar parametru DR. Wyniki przedstawiono na rys. 6. Na podstawie wyniku pomiaru DR można stwierdzić, że materiał z płyty gramofonowej ma największy zakres dynamiki. Dalej – najwyższe wartości parametru DR odnoszą się do płyty CD (ok. 9–10 dB), serwisów Tidal (ok. 8–9 dB), iTunes (ok. 7–9 dB), a na końcu serwisu Spotify (ok. 5–6 dB). Serwis Spotify udostępnia materiał o najmniejszym zakresie dynamiki, a różnica w średnich wartościach parametru DR między płytą gramofonową a serwisem mieści się w przedziale 11–12 dB.




Rys. 7. Wykres wartości średnich wyników pomiarów wartości skutecznej RMS dla różnych nośników

Na rysunku 7 przedstawiono wyniki pomiarów wartości skutecznej (RMS). Na ich podstawie można wnioskować, że, najniższe wartości RMS dotyczyły materiału nagranego na płycie gramofonowej. Jak w przypadku albumu zespołu Pink Floyd najmniejsza wartość RMS odniosła się do materiału udostępnionego na platformie Spotify. Średnia wartość RMS dla całego albumy wyniosła –15,5 dB. Kolejnym nośnikiem była

płyta CD – wartość średnia parametru RMS dla całego albumu w tym przypadku to –13,6 dB. Najwyższą wartość parametru RMS osiągnięto dla materiału udostępnianego w serwisie Tidal oraz iTunes, czyli odpowiednio –10,0 dB i –10,8 dB.




Rys. 8. Wartości średnie wyników pomiarów poziomu głośności LUFS dla różnych nośników

Na koniec zmierzono poziom głośności materiału wyrażony w jednostce LUFS. Wyniki pomiarów przestawiono na rys. 8. Jak wynika z przeprowadzonego pomiaru, najgłośniejszy materiał znacząco odbiegający od pozostałych nośników udostępniał serwis Spotify. Różnica w głośności utworów w zestawieniu z innymi serwisami mieściła się 3–6 LUFS. Po Spotify najgłośniejszy materiał pochodził z platformy Tidal oraz iTunes, których średni poziom głośności albumu to –10,6 LUFS. Materiał o najniższym poziomie głośności był na płycie CD – średnia wartość odnośnie do albumu wyniosła –13,7 LUFS, a w przypadku płyty gramofonowej –26,3 LUFS.

## 10.4. Podsumowanie

Obecnie na rynku istnieje wiele różnych nośników, na których można przechowywać muzykę. Każdy z nich oferuje dźwięk charakteryzujący się parametrami o odmiennych wartościach. W przeprowadzonych badaniach można było doskonale dostrzec różnicę między nośnikami cyfrowymi a analogowymi. Zakres dynamiki znajdującego się na nich materiału znacząco odbiegał od siebie. Również między nośnikami cyfrowymi, takimi jak: Spotify, Tidal, iTunes, można zauważyć spore rozbieżności nie tylko w poziomie głośności materiału, lecz także w zakresie dynamiki.

W rozdziale opisano i omówiono pomiar czterech parametrów akustycznych: wartości TPL (True Peak Level), DR, na podstawie którego określano zakres dynamiki mate-

riału znajdującego się na nośnikach, wartości skutecznej (RMS) i poziomu głośności materiału mierzonego w nowej jednostce LUFS wykorzystującej algorytm umożliwiający dostosowanie subiektywnego sposobu percepcji dźwięku przez człowieka do wyników pomiarów.

Materiał do pomiarów był dostępny w postaci pięciu różnych nośników: płyt gramofonowych, płyt CD, trzech serwisów oferujących strumieniowanie plików, czyli Spotify, Tidal, iTunes. Wykorzystane albumy to *Sing-Sing* M. Rodowicz i *The Dark Side of the Moon* grupy Pink Floyd stanowiące przykład muzyki ponadczasowej i kultowej, dlatego wydawanej na rożnych nośnikach w różnych okresach czasu.

Na podstawie otrzymanych wyników pomiarów można stwierdzić, że serwisy zajmujące się strumieniowaniem plików, chociaż oferują swoim użytkownikom bardzo podobną usługę, różnią się między sobą udostępnianym materiałem. Serwis Spotify dysponuje materiałem poddanym dużej kompresji dynamiki, a poziom głośności utworów jest na granicy przesterowania (por. np. [17]). Atutem pozostałych serwisów streamingowych jest odmienny format plików udostępnianych przez platformy. Sprawia to, że pliki udostępniane przez serwisy Tidal oraz iTunes dążą do uzyskania wartości parametrów związanych z głośnością zbliżonych do materiału nagranego na płyty CD.

Pomiary wartości skutecznej RMS oraz poziomu głośności wyrażonego w jednostce LUFS umożliwiły zauważenie rozbieżności wynikających z zastosowania odmiennych parametrów do określenia głośności materiału na nośnikach. Poziom głośności wyrażony w jednostce LUFS jest dokładniejszy i lepiej różnicuje materiał dźwiękowy. Przyczyną tego jest wykorzystany algorytm zaprojektowany przez EBU w celu przybliżenia wyników pomiarów do rzeczywistej wartości głośności percypowanej przez człowieka.

**Słowa kluczowe:** głośność programu, EBU R128, pomiary głośności.

## Bibliografia

[1] Katz B., *Sound Board: Can We Stop the Loudness War in Streaming?*, „J. Audio Eng. Soc." 2015, Vol. 63, No. 11, s. 939, 940.

[2] Vickers E., *The Loudness War: Do Louder, Hypercompressed Recordings Sell Better?*, „J. Audio Eng. Soc." 2011, Vol. 59, No. 5, s. 346–351.

[3] ITU-R BS.1770-4. Algorithms to measure audio programme loudness and true-peak audio level, 2015.

[4] ITU-R BS.1771-1. Requirements for loudness and true-peak indicating meters, 2012.

[5] ITU-R BT.2217. Compliance material for Recommendation ITU-R BS.1770, 2011.

[6] EBU Technical Recommendation R 128 Loudness normalisation and permitted maximum level of audio signals, 2010.
[7] EBU Tech Doc 3341 Loudness Metering: 'EBU Mode' metering to supplement loudness normalisation in accordance with EBU R 128, 2010.
[8] EBU Tech Doc 3342 Loudness Range: A descriptor to supplement loudness normalisation in accordance with EBU R 128, 2010.
[9] EBU Tech Doc 3343 Practical Guidelines for Production and Implementation in accordance with EBU R 128, 2010.
[10] EBU Tech Doc 3344 Practical Guidelines for Distribution of Programmes in accordance with EBU R 128, 2010.
[11] Spotify; https://press.spotify.com/pl/about/ [dostęp: 5.10.2020].
[12] Tidal; https://tidal.com/about [dostęp: 5.10.2020].
[13] iTunes; https://www.apple.com/pl/itunes/ [dostęp: 5.10.2020].
[14] Audacity 2.2.1 Manual; https://manual.audacityteam.org/ [dostęp: 5.10.2020].
[15] TT DR Offline Meter Software Manual Version 1.1; http://www.dynamicrange.de/sites/default/files/DR-Manual-V1_1-English.pdf [dostęp: 5.10.2020].
[16] Insight help documentation; https://www.izotope.com/content/dam/izotope/support/help-guides/izotope-insight-help-documentation.pdf [dostęp: 5.10.2020].
[17] Grimm E., *Analyzing Loudness Aspects of 4.2 Million Musical Albums in Search of an Optimal Loudness Target for Music Streaming*, Paper 10268, 2019.

# 11. Wpływ parametrów kompresji dynamicznej na subiektywną głośność nagrania instrumentu muzycznego

KAROL CZESAK, PIOTR KLECZKOWSKI

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie,
al. Mickiewicza 30, 30-059 Kraków

Na przestrzeni ostatnich dziesięcioleci zauważyć można tendencję do zawężania zakresu dynamicznego nagrań muzycznych. Przyświeca temu cel, jakim jest całościowe zwiększenie głośności nagrania. Pojęcie głośności bardzo często wiązane jest z wartością sygnału w kanale transmisji. Stąd bardzo często materiał poddawany jest agresywnemu przetwarzaniu z użyciem procesorów dynamicznych. Doświadczenie reżyserów dźwięku pokazuje jednak, że subiektywne wrażenie głośności warunkowane jest nie tylko przez stopień kompresji, lecz także przez inne jej parametry. W niniejszym rozdziale podjęto próbę wskazania parametrów kompresji dynamicznej wywierających istotny wpływ na subiektywnie postrzeganą głośność zróżnicowanego materiału muzycznego. Stwierdzono, że uzyskanie większej głośności nagrania bez stosowania wysokich wartości stopnia kompresji jest przy określonych warunkach możliwe przez: obniżenie progu kompresji, unikanie niewielkich stopni kompresji, wydłużenie czasu ataku, unikanie zbyt długiego czasu zwolnienia.

## 11.1. Wstęp

Kompresja dynamiki jest operacją dążącą do redukcji zakresu dynamicznego materiału dźwiękowego przez zmniejszenie wzmocnienia sygnału po przekroczeniu przez

niego ustalonego progu. Celem jej stosowania jest zarówno uniknięcie przesterowania kanału transmisji, jak i zwiększenie komfortu odbioru materiału. Od lat 50. XX w. zaobserwować można odbiór utworów muzycznych nagranych z wyższym poziomem skutecznym za atrakcyjniejszy. Doprowadziło to do zjawiska „wojny głośności", której apogeum przypadło na przełom XX i XXI w. [1]. Odpowiedź na to zjawisko nadeszła ze strony Międzynarodowej Organizacji Telekomunikacyjnej (ITU) oraz Europejskiej Unii Nadawców (EBU). Opublikowały one liczne dokumenty, w których spróbowano uzgodnić poziom głośności w mediach elektronicznych w ujęciu zakresu dynamicznego i wartości skutecznej sygnału. W przedstawionym kontekście wypada zadać sobie pytanie: Czy jest możliwe wyprodukowanie materiału dźwiękowego spełniającego przyjęte zalecenia dotyczące głośności, odczuwalnie jednak głośniejszego od innych sobie podobnych? W ramach poszukiwania odpowiedzi na to pytanie, w tym rozdziale badano zależność między parametrami kompresji dynamicznej a subiektywnie rozumianym poczuciem głośności i atrakcyjności materiału, wynikającymi z indywidualnego poczucia estetyki i komfortu słuchaczy.

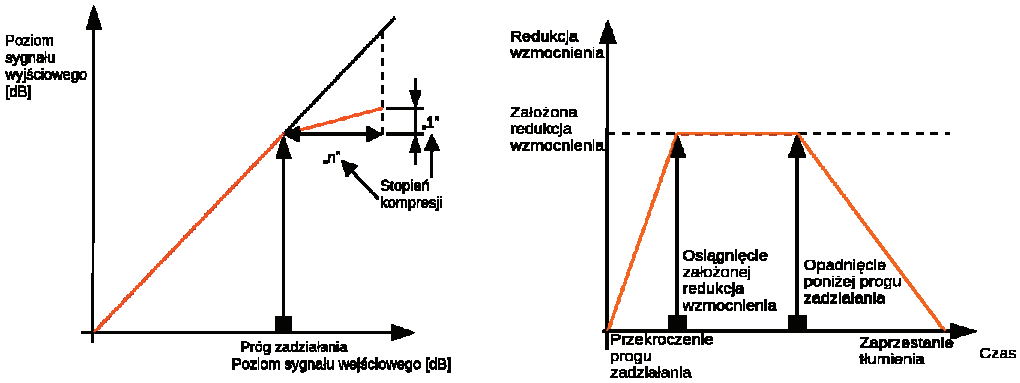
## 11.2. Rozpatrywane parametry kompresji dynamicznej oraz implementacja kompresora

Kompresor dynamiki przekształca sygnał w sposób nieliniowy i działa w dziedzinie czasu. Te nieliniowości występują w stanach przejściowych, dzięki czemu możemy traktować kompresor jako narzędzie liniowe, zmienne w czasie [2]. Wprowadzane przez kompresor zmiany są zależne od sygnału podawanego na wejście, za którym procesor niejako „podąża". Oznacza to, że zastosowanie kompresora pozostaje nie bez wpływu na barwę dźwięku – może też wprowadzać (i wprowadza) mniej lub bardziej słyszalne artefakty [3]. Mając to na uwadze, kompresor dynamiczny można wykorzystać w funkcji już nie tyle zabezpieczającej, ile w charakterze narzędzia kreacji brzmienia.

Charakterystyka statyczna kompresora jest określona parametrami takimi jak próg zadziałania i stopień kompresji. Próg zadziałania wyrażany w decybelach w odniesieniu do nominalnej wartości sygnału w kanale transmisji jest parametrem określającym poziom graniczny sygnału, po którego przekroczeniu zaczyna zachodzić tłumienie sygnału. Stopień kompresji określa się ilorazem $n:1$ definiującym stosunek „nadwyżki" sygnału wejściowego powyżej progu zadziałania kompreso-

ra do pozostającej „nadwyżki" sygnału wyjściowego. Zmienna *n* przybiera wartości od 1 wzwyż.

W celu eliminacji artefaktów związanych z kompresją dynamiki i wynikłych z wprowadzonej nieliniowości przebieg redukcji wzmocnienia jest określony pewną obwiednią czasową. W kompresorach analogowych wynika ona z pewnej bezwładności układów użytych do implementacji procesu kompresji dynamiki opisanej czasem ataku i czasem opadania. Wprowadzenie regulacji tych dwóch parametrów umożliwia zarówno lepsze dostosowanie charakteru pracy kompresora do rodzaju opracowywanego materiału, ustrzeżenie się przed artefaktami [4], jak i jego zastosowanie w charakterze narzędzia do kształtowania obwiedni amplitudowej sygnału [5] – czego przykładem jest kontrola transjentów. Opisane parametry zobrazowano na rys. 1.



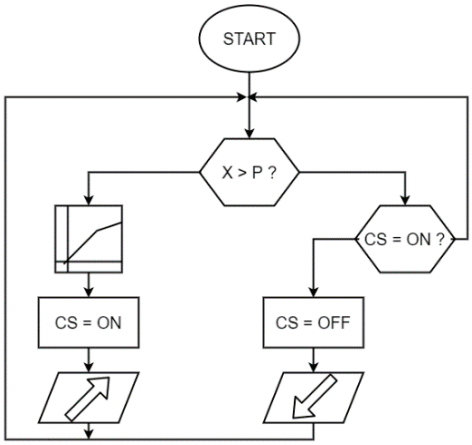

Rys. 1. Graficzna interpretacja czterech podstawowych parametrów kompresji dynamiki

Jeśli zatem oznaczy się stan pracy kompresora zmienną *CS* przyjmującą wartości: „prawda" lub „fałsz", sygnał wejściowy – zmienną *X*, a próg zdziałania – zmienną *P*, można przedstawić działanie kompresora przy użyciu schematu blokowego (rys. 2).

W celu pełnej kontroli nad wartościami parametrów kompresji dynamicznej zdecydowano o implementacji kompresora dynamiki w postaci programu komputerowego działającego w czasie rzeczywistym – najlepiej w formie wtyczki do oprogramowania DAW (Digital Audio Workstation – cyfrowa dźwiękowa stacja robocza). Ze względu na łatwość implementacji kompresor napisany został w języku C++ przy wykorzystaniu pakietów narzędzi Steinberg VST SDK i JUCE. Kompilacja projektu przebiegała w środowisku Microsoft Visual Studio 2017.

Implementacja opisanych stanów przejściowych kompresora wprowadziła trudności – ich wyeliminowanie zostało zapewnione dzięki utworzeniu licznika czasu wyzwala-

nego przejściem sygnału przez próg zadziałania kompresora. Porównywanie jego odczytów z założonymi czasami ataku i opadania kompresora umożliwiło płynne przechodzenie na wyjściu między sygnałem nieskompresowanym a sygnałem o wyliczonej uprzednio wartości docelowej.




Rys. 2. Schemat blokowy działania kompresora dynamiki

Na podstawie przedstawionych założeń stało się możliwe opracowanie kompresora najpełniej realizującego przyjęte założenia teoretyczne dotyczące kompresji dynamicznej, a jednocześnie wydajnego obliczeniowo i w postaci łatwej w uruchomieniu wtyczki VST.

Bardzo użyteczne okazało się umożliwienie zmiany parametrów kompresji w czasie pracy procesora. Dlatego stworzono graficzny interfejs użytkownika wyposażony w kontrolki do regulacji tych parametrów (rys. 3).


Rys. 3. Graficzny interfejs użytkownika kompresora dynamiki

W graficznym interfejsie użytkownika tej implementacji kompresora dynamiki widoczny jest brak opcji regulacji poziomu wyjściowego. Jego pominięcie jest jednak zabiegiem celowym. Zgodnie bowiem z przyjętym w projekcie założeniu o ewaluacji odczuwanej głośności materiału muzycznego przetworzone przez ten kompresor sygnały powinny zostać znormalizowane do jednakowego poziomu skutecznego.

## 11.3. Pozyskanie i przetwarzanie materiału muzycznego

Kolejnym założeniem było operowanie na próbkach zarówno instrumentalnych, jak i wokalnych przy maksymalnej eliminacji wpływu czynników zależnych od pomieszczenia czy samej techniki rejestracji, warunkujących charakterystykę, a zatem i brzmienie nagrywanego materiału. Pozyskanie nagrań spełniających założone kryterium nie jest łatwe. Możliwość zdobycia tego typu materiału pojawiła się latem 2018 r. w czasie sesji nagraniowej big-bandu wywodzącego się z młodzieżowej orkiestry dętej. Kapelmistrzowi zespołu chodziło o taką rejestrację materiału, by ewentualna podmiana niepoprawnie wykonanych partii była maksymalnie ułatwiona.

W studiu nagraniowym dostępne były trzy pomieszczenia nagraniowe z czasem pogłosu wyrównanym w oktawach do wartości nieco poniżej 0,5 s, a także reżyserka. W ramach jednej sekcji jednocześnie wykonywało swoją partię do trzech instrumentalistów – każdy przebywający w innym pomieszczeniu nagraniowym. Dzięki temu udało się wyeliminować całkowicie przesłuchy między instrumentami dętymi. W nagraniach instrumentów dętych wykorzystano mikrofony: AKG C2000, AKG C414XLS, Neumann TLM 103. Do przetworzenia sygnału posłużyła natomiast karta dźwiękowa wyposażona w przedwzmacniacze mikrofonowe – Universal Audio Apollo 8p.

Gitara basowa zarejestrowana została z użyciem urządzenia zapewniającego dopasowanie impedancyjne instrumentu oraz przedwzmacniacza – popularnie nazywanego Di-Boxem, dzięki któremu zapewniona została również ich separacja galwaniczna.

Jedynie nagrania sekcji smyczkowej przeprowadzono na tzw. setkę (czyli wszystkie instrumenty grały jednocześnie) w czasie sesji nagraniowej, która odbyła się w obiekcie z lat 30. XX w. pełniącym pierwotnie funkcje kina. Oprócz pary głównej pracującej w systemie M/S (mikrofony Neumann TLM 103 i AKG C414XLS) oraz mikrofonów nad sekcjami – każdy z instrumentów wyposażony był w mały mikrofon marki DPA zamocowany w okolicach podstawka. Z takiego właśnie mikrofonu w przypadku skrzy-

piec umiejscowionych skrajnie po lewej stronie sceny pochodzi wykorzystana w projekcie próbka skrzypiec.

W celu uzyskania pełni brzmienia werbla użyto dwu ścieżek zarejestrowanych mikrofonami Shure SM57 umieszczonymi w pobliżu jego naciągów, przy czym sygnał z mikrofonu umiejscowionego w pobliżu naciągu rezonansowego werbla wzięto do sumowania z odwróconą polaryzacją. Do rejestracji ścieżek wokali posłużył wymieniony już wielokrotnie mikrofon AKG C414XLS.

Wybór próbek instrumentów muzycznych do badań wiązał się z daleko idącym kompromisem. Z pozycji badacza najlepszym rozwiązaniem byłoby zebranie wyników dla możliwie największej liczby instrumentów muzycznych oraz wartości parametrów kompresji. Wynikałaby z tego jednak konieczność prowadzenia wielogodzinnych testów odsłuchowych, co zarówno z przyczyn logistycznych, jak i zmęczenia słuchaczy jest niemożliwe w realizacji. Dlatego zostało nałożone kryterium długości pojedynczego testu odsłuchowego, który nie mógł trwać więcej niż 30 min.

Stąd pojawiła się potrzeba ograniczenia do minimum liczby badanych instrumentów. Bardzo ważne przy tym było, żeby były to instrumenty relatywnie często poddawane kompresji dynamicznej, a także możliwie jak najbardziej zróżnicowane pod względem budowy czy sposobu gry, a przede wszystkim charakteru wydobywanego dźwięku. W celu pełniejszego skupienia uwagi słuchacza zapadła decyzja, by wszystkie próbki były fragmentami jednego i tego samego utworu muzycznego.

Zgodnie z przyjętymi kryteriami w badaniu użyte zostały próbki pochodzące z następujących instrumentów: perkusyjna – werbel, dęte drewniane – saksofon, dęte blaszane – puzon, strunowe szarpane – gitara basowa, smyczkowe – skrzypce, oraz wokalu. Wszystkie próbki pochodzą z utworu A. Zielińskiego *Nie całuj mnie pierwsza* do słów A. Osieckiej – w big-bandowej aranżacji W. Zwierniaka.

Nie tylko dla każdego z wybranych instrumentów, lecz także wokalu wyizolowano po jednym reprezentatywnym przykładzie obejmującym jedną dłuższą lub dwie krótsze frazy muzyczne. Długości próbek wahały się w granicach 12–15 s. Przy tak dobranej długości słuchacze mogli już przeprowadzić analizę tego, co słyszą – jednocześnie udało się zachować rozsądny czas trwania pojedynczego testu odsłuchowego.

Każdy z przykładów poddany został kompresji dynamicznej 12-krotnie – dla trzech różnych wartości każdego z czterech badanych parametrów. Szczególną uwagę poświęcono zachowaniu stałych wartości parametrów kompresji innych niż analizowany w obrębie danego pakietu próbek. Same wartości tych parametrów zostały dobrane eksperymentalnie, tak by w czasie ich dobierania była słyszalna istotna redukcja wzmocnienia, a (poza werblem i gitarą basową szeroko stosowanymi w muzyce roz-

rywkowej) nie były słyszalne wynikające z kompresji artefakty, które często są traktowane właśnie w przypadku werbla i gitary basowej jako integralny element brzmienia danego instrumentu.

Po przeprowadzeniu kompresji dynamicznej wszystkie przykłady zostały znormalizowane do jednego poziomu skutecznego. Żeby tego dokonać, we wszystkich przykładach zostały obliczone poziomy głośności zgodnie z zaleceniem zawartym w normie ITU BS.1770 [6], której europejskim odpowiednikiem jest EBU R128 [7] i suplementy: EBU Tech 3341 [8], EBU Tech 3343 [9] określające zalecany poziom głośności materiału dźwiękowego w mediach elektronicznych oraz sposób obliczania ich wartości – obliczone na ich podstawie poziomy głośności i zakresy dynamiki są równoważne.

W rozpatrywanym przypadku do pomiaru poziomu głośności materiału zastosowany został miernik Orban Loudness Meter (rys. 4) w wersji 2.9.6 (dostępny do pobrania pod adresem: www.orban.com/meter).


Rys. 4. Pomiar poziomu głośności przygotowywanego materiału

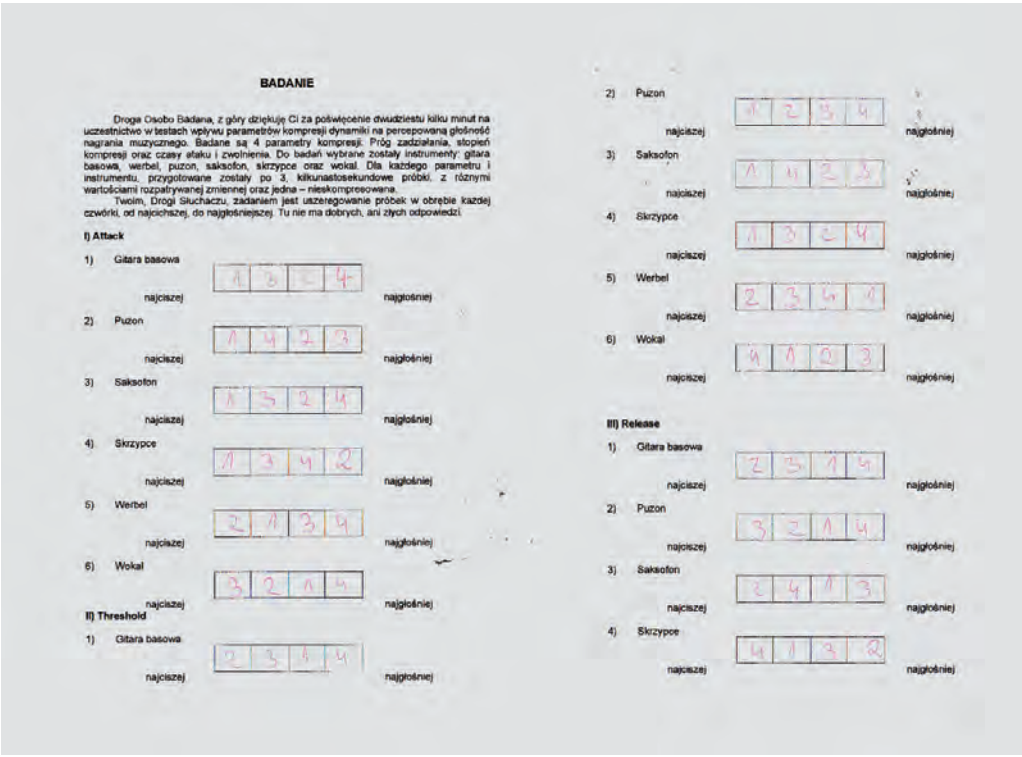
Ponieważ znane były poziomy głośności całej palety próbek biorących udział w badaniu, łatwa stała się normalizacja niemal wszystkich próbek do wartości głośności –23 LKFS (który jest równoważny wartości –23 LUFS) stanowiącego umowny standard w emisji telewizyjnej w Europie. Wyjątek stanowiły jednak próbki brzmienia werbla. Przy próbie znormalizowania jego nieskompresowanej próbki do wartości –23 LKFS transjenty powodowały przesterowanie kanału transmisji, wyjątkowo zatem dla tego instrumentu próbki znormalizowane zostały do wartości –30 LKFS. Po znormalizowaniu próbek pomiary powtórzono w celu uzyskania aktualnych wartości zakresu głośności (*loudness range*) oraz najwyższego, zrekonstruowanego poziomu szczytowego (*highest reconstructed peak level*).

Na koniec omawianego etapu próbki zostały pogrupowane w zależności od instrumentu oraz zmieniającego się parametru kompresji, a następnie ułożone w kolejności pseudolosowej – wraz z przykładem nieskompresowanym, oraz rozdzielone dwusekundowymi pauzami. Uzyskane w ten sposób pakiety, na które składały się po trzy próbki skompresowane i po jednej nieskompresowanej, zostały wyeksportowane jako pojedyncze pliki wav, tak by słuchacze wysłuchali (w obrębie danej grupy) próbek dokładnie w tej samej kolejności.

## 11.4. Metodyka prowadzonych testów odsłuchowych

Badania odsłuchowe przeprowadzono w możliwie cichych pomieszczeniach. Słuchacze otrzymywali komputer przenośny z otwartym programem VLC media player przygotowanym do odtwarzania sporządzonych wcześniej pakietów próbek oraz słuchawki Audio Technica ATH-M30x – zostali poproszeni o wyregulowanie możliwie jak najbardziej komfortowego dla siebie poziomu głośności. Następnie zostali poinstruowani o sposobie zatrzymywania i wznawiania odsłuchu (klawisz: spacja) oraz przechodzenia do kolejnej grupy próbek (klawisz: enter). Manualne wyzwalanie kolejnych zestawów próbek uwalniało słuchaczy od nadmiernego poczucia reżimu czasowego, stwarzało także możliwość spokojnego zastanowienia się nad udzieleniem odpowiedzi. W czasie testu wysłuchiwali oni czterech serii (dla każdego ze zmieniających się parametrów kompresji) po sześć pakietów (dla każdego instrumentu) po cztery próbki. Zadaniem respondentów było uszeregowanie próbek w obrębie każdego pakietu od uważanej przez nich za najcichszą do uważanej przez nich za najgłośniejszą – w okienka należało wpisać numery konkretnych próbek z pakie-

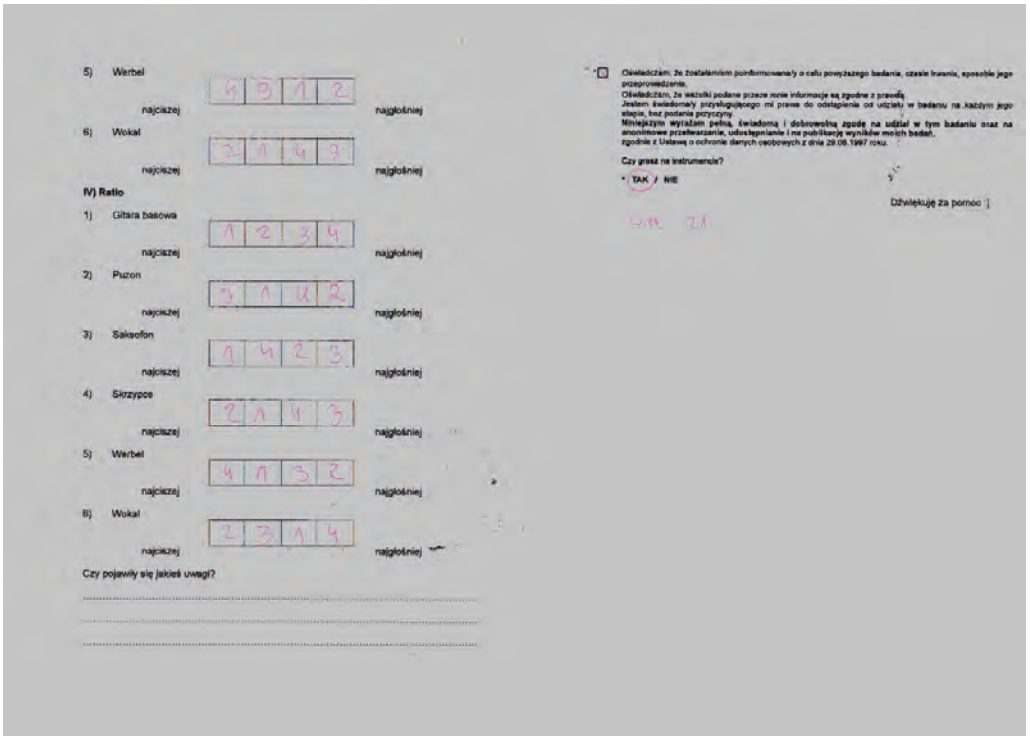
tu. Uzyskane odpowiedzi przeliczane były na wartości punktowe – próbce, której numer został umieszczony w skrajnym lewym okienku, przypisywano 1 punkt, a tej w skrajnym prawym – 4 punkty. Na rysunkach 5 i 6 przedstawiono przykładowo wypełnioną kartę odpowiedzi.




Rys. 5. Przykładowo wypełniona karta odpowiedzi – awers

Jak już wcześniej wspomniano, próbki pogrupowane były w czteroprzykładowe pakiety. Każda z serii zawierających po sześć pakietów odnosiła się do innego z badanych parametrów kompresji dynamicznej. I tak kolejno pojawiały się serie dla: czasu ataku kompresora, progu zadziałania kompresora, czasu zwolnienia kompresora i stopnia kompresji. A w obrębie każdej serii instrumenty: gitara basowa, puzon, saksofon, skrzypce, werbel i wokal.

Grupę badawczą stanowili studenci kierunku inżynieria akustyczna na Wydziale Inżynierii Mechanicznej i Robotki Akademii Górniczo-Hutniczej w Krakowie i młodzież spędzająca swój wolny czas w wiejskiej świetlicy w Gierczycach (woj. małopolskie).




Rys. 6. Przykładowo wypełniona karta odpowiedzi – rewers

W badaniu wzięło udział 18 osób, przy czym trzy z nich zadeklarowały trudności w wychwytywaniu różnic głośności próbek, dlatego uzyskane od nich wyniki zostały odrzucone. Do analizy statystycznej uwzględniono zatem wyniki pochodzące od pozostałych 15 osób – każda poproszona została o podanie swojego wieku i określenie aktywności w grze na instrumencie muzycznym lub w śpiewie. Osób aktywnie muzycznych w badanej grupie było więcej, a średnia wieku zauważalnie niższa od osób „niegrających". Żaden z respondentów nie deklarował problemów ze słuchem. Osoby wykonujące muzykę dzieliły się też nieco innymi uwagami i spostrzeżeniami niż pozostałe. To właśnie wśród osób aktywnych muzycznie pojawiały się głosy na temat sposobu detekcji i oceny głośności (słyszalność artykulacji, zawartość alikwotów).

Ciekawym przypadkiem wśród respondentów był niewykazujący aktywności muzycznej ani też nieposiadający doświadczenia w wykonawstwie muzyki – absolwent kulturoznawstwa. Wykazywał rozległą wiedzę w zakresie muzyki poważnej, której słucha, jak sam przyznał, w sposób wyuczony. Zademonstrował niebywałą wręcz łatwość w rozpoznawaniu próbek instrumentów akustycznych (np. saksofonu i skrzypiec) mimo zgod-

nego ze sztuką zastosowania na nich kompresji dynamicznej. Problematyczne w jego odbiorze okazały się natomiast próbki instrumentów poddanych agresywnej kompresji (jak werbel czy gitara basowa), które w opinii większości respondentów uchodziły za najłatwiejsze do rozpoznawania różnic w głośności.

Mając na uwadze ten przypadek, można sformułować tezę, że oprócz zarówno samych parametrów kompresji dynamicznej, jak i jakości nagrania muzycznego – wpływ na subiektywne postrzeganie głośności wywierają osobiste preferencje, doświadczenia, przyzwyczajenia oraz to, co kolokwialnie zwykliśmy nazywać „osłuchaniem z muzyką".

## 11.5. Prezentacja i omówienie uzyskanych wyników

Wyniki z przeprowadzonych testów odsłuchowych zostały najpierw pogrupowane w zależności od instrumentu muzycznego i badanego parametru kompresji dynamicznej, a następnie poddane analizie statystycznej.

Analiza przeprowadzono w następujących krokach:

- sprawdzenie normalności rozkładu z wykorzystaniem testów – Kołomogorowa–Smyrnowa, Lillieforsa, Shapiro–Wilka;
- weryfikacja statystycznej istotności różnic między odpowiedziami respondentów dla próbek z danej grupy na podstawie testów – Kruskala–Wallisa, Mediany;
- w przypadku wystąpienia statystycznie istotnych różnic uściślenie, między którymi próbkami w obrębie grupy te różnice wystąpiły – wielokrotne porównanie średnich.

W dalszej części rozdziału (w formie tabelarycznej oraz opisowej) zostały przedstawione i omówione uzyskane wyniki w przypadku próbek, między którymi zaobserwowano statystycznie istotne różnice w odpowiedziach respondentów.

### 11.5.1. Gitara basowa

Gitara basowa powszechnie traktowana jest jako instrument elektryczny. Generowanie dźwięku odbywa się tu co prawda na drodze mechanicznej, ale o ostatecznym jej brzmieniu przesądza jednak sposób konstrukcji toru wzmocnienia. Ze względu na pełnioną przez ten instrument (w muzyce rozrywkowej) funkcję kompresja dynamiczna stanowi wręcz pożądany element wpływający na jego brzmienie, dzięki czemu realizatorzy otrzymują dużą dowolność w jej kształtowaniu, a słyszalne artefakty czy zniekształcenia traktowane są jako integralne cechy brzmienia gitary basowej [10].

W tym przypadku istotne różnice wynikły wraz ze zmianą progu zadziałania kompresora. Próbka skompresowana z progiem zadziałania –13,1 dBFS jest wyraźnie głośniejsza od próbki skompresowanej z progiem zadziałania –9,1 dBFS. Między odczuwaną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat ich parametrów przedstawiono w tab. 1.

Tabela 1. Zestawienie próbek odnośnie do gitary basowej, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą progu zadziałania kompresora

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Gitara basowa | Threshold | 4. | –13,1 | 4,2,1 | 2 | 81 | 5 | –11,8 |
| 2. | Gitara basowa | Threshold | 1. | –9,1 | 4,2,1 | 2 | 81 | 4,9 | –8,1 |

Próbka uznana przez respondentów za głośniejszą charakteryzuje się niżej ustawionym progiem zadziałania kompresora, w wyniku czego mamy do czynienia z mniejszym współczynnikiem szczytu, ale też nieco większą zawartością zniekształceń w stanach przejściowych kompresora, co może prowadzić do odczucia większej głośności.

W przypadku gitary basowej zaobserwowano również dla stopnia kompresji statystycznie istotne różnice w odpowiedziach respondentów. Próbka skompresowana ze stopniem kompresji 4,2 : 1 sprawia na respondentach wrażenie głośniejszej niż skompresowana ze stopniem 1,6 : 1. Między odczuwaną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 2.

Tabela 2. Zestawienie próbek odnośnie do gitary basowej, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą stopnia kompresji

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Gitara basowa | ratio | 3. | –9,1 | 4,2 : 1 | 2 | 81 | 6,9 | –8,1 |
| 2. | Gitara basowa | ratio | 1. | –9,1 | 4,2 : 1 | 2 | 81 | 4,9 | –7,6 |

Próbka wybrana przez respondentów jako głośniejsza charakteryzuje się większym stopniem kompresji, w wyniku czego analogicznie do zmian progu zadziałania kompresora w tym przypadku wystąpił mniejszy współczynnik szczytu, a większa zawar-

tość zniekształceń w stanach przejściowych kompresora – to może prowadzić do odczucia większej głośności.

## 11.5.2. Puzon

Puzon jest instrumentem stosowanym w bardzo różnorodnych składach instrumentalnych – zarówno w orkiestrach symfonicznych, jazzowych big-bandach, jak i w orkiestrach dętych wykonujących swój tradycyjny repertuar, ale też standardy muzyki filmowej i rozrywkowej. Puzon pojawia się również w sekcjach dętych zespołów bluesowych i rockowych. Mnogość zastosowań tego instrumentu przekłada się na mnogość podejść realizatorów dźwięku do pracy z nim. I tak, o ile przy składach o charakterze tradycyjnym przesadna ingerencja w brzmienie instrumentu jest źle widziana, o tyle w zespołach wykonujących muzykę rozrywkową agresywne brzmienie puzonu jest wręcz pożądane, dzięki czemu jest on mocno osadzony w bogatych aranżacjach utworów zakładających dużą ekspresję wykonawczą.

Istotne różnice dotyczące puzonu można zaobserwować już w przypadku czasu ataku. Otóż próbka skompresowana z czasem ataku 142 ms okazuje się być wyraźnie głośniejsza od próbki nieskompresowanej. Między odczuwaną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 3.

Tabela 3. Zestawienie próbek odnośnie do puzonu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą czasu ataku

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Puzon | attack | 3. | –11,35 | 1,4 : 1 | 142 | 134 | 8,5 | –10,3 |
| 2. | Puzon | attack | 1. | bez kompresji | | | | 8,4 | –7,9 |

W opisywanym przypadku próbka skompresowana okazała się być odczuwalnie głośniejsza od nieskompresowanej. Ciekawi jednak, że statystycznie istotne różnice zachodzą dla najdłuższego z proponowanych czasów ataku, można to interpretować jako potencjalny wpływ początkowego transjentu na percepcję głośności puzonu.

W analizie dotyczącej czasu zwolnienia zostały zaobserwowane istotne statystycznie różnice. Próbka skompresowana z czasem zwolnienia 82 ms została oceniona jako wyraźnie głośniejsza niż próbka skompresowana z czasem zwolnienia 200 ms czy próbka nieskompresowana. Jednocześnie nie udało się stwierdzić wyraźnych różnic

między wynikami uzyskanymi przez te próbki a wynikami uzyskanymi przez próbkę skompresowaną z czasem zwolnienia 134 ms. Szczegółowe informacje na temat parametrów poszczególnych próbek przedstawiono w tab. 4.

Tabela 4. Zestawienie próbek odnośnie do puzonu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą czasu zwolnienia

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Puzon | release | 3. | bez kompresji | | | | 8,4 | –7,9 |
| 2. | Puzon | release | 4. | –11,35 | | 109 | 82 | 8,5 | –10,3 |
| 3. | Puzon | release | 1. | –11,35 | 3,4 : 1 | 109 | 200 | 8,4 | –10,3 |

W opisywanym przypadku próbka nieskompresowana oraz skompresowana z krótkim czasem zwolnienia okazują się być odczuwalnie głośniejsze od próbki z najdłuższym czasem powrotu powodującym kompresowanie fazy wybrzmienia instrumentu – stanowi to przykład nietrafionego doboru wartości parametru kompresji.

Jeśli wziąć pod uwagę próg zadziałania, to można także zaobserwować statystycznie istotne różnice dla puzonu w odpowiedziach respondentów. Próbka skompresowana z progiem zadziałania –15,6 dBFS oceniona została jako wyraźnie głośniejsza niż próbka nieskompresowana. Między odczuwaną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 5.

Tabela 5. Zestawienie próbek odnośnie do puzonu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą progu zadziałania kompresora

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Puzon | threshold | 4. | –15,55 | 5,5 : 1 | 92 | 134 | 8,3 | –14,1 |
| 2. | Puzon | threshold | 2. | bez kompresji | | | | 8,4 | –7,9 |

Próbka uznana przez respondentów za subiektywnie głośniejszą charakteryzuje się bardzo nisko ustawionym progiem zadziałania kompresora, w wyniku czego współczynnik szczytu jest mniejszy, ale większa zawartość zniekształceń w stanach przejściowych kompresora niż w próbce nieskompresowanej, w której te stany przejściowe nie występują – może to prowadzić do odczucia większej głośności.

## 11.5.3. Saksofon

Saksofon podobnie jak puzon wykorzystywany jest w różnorakich składach instrumentalnych – często oba instrumenty występują, jeżeli nie obok siebie, to w składach wykonujących bardzo podobną muzykę: od poważnej i filmowej, przez jazz po folk i muzykę rozrywkową. Podejście realizatorów dźwięku jest tu analogiczne do podejścia w przypadku puzonu. W zależności od rodzaju wykonywanej muzyki kompresja dynamiczna jest tu efektem mniej lub bardziej pożądanym, przy czym saksofon obdarzony łagodniejszym brzmieniem niż puzon jest bardziej wrażliwy na artefakty.

Statystycznie istotne różnice dotyczące saksofonu w odpowiedziach respondentów zostały zaobserwowane między próbką skompresowaną z progiem zadziałania kompresora –10,8 dBFS i nieskompresowaną a próbką skompresowaną z progiem zadziałania –12,7 dBFS, od której okazały się wyraźnie głośniejsze. Jednocześnie między wymienionymi próbkami a próbką skompresowaną z progiem zadziałania –7 dBFS nie zostały stwierdzone statystycznie istotne różnice. Szczegółowe informacje na temat parametrów wszystkich próbek przedstawiono w tab. 6.

Tabela 6. Zestawienie próbek odnośnie do saksofonu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą progu zadziałania kompresora

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Saksofon | threshold | 3. | –10,8 | 3,0 : 1 | 20 | 107 | 6,4 | –9 |
| 2. | Saksofon | threshold | 4. | bez kompresji | | | | 4,5 | –10,2 |
| 3. | Saksofon | threshold | 1. | –12,7 | 3,0 : 1 | 20 | 107 | 6,5 | –10,2 |

W opisywanym przypadku próbka nieskompresowana oraz skompresowana z ustawionym progiem zadziałania kompresora –7dBFS okazują się być odczuwalnie głośniejsze od próbki z najniżej ustawionym progiem zadziałania, który powodować może wrażenie stłumionego i pozbawionego ekspresji brzmienia instrumentu.

W przypadku saksofonu istotne różnice zostały zaobserwowane także dla stopnia kompresji. I tu próbka skompresowana ze stopniem kompresji równym 3,0 : 1 oceniona została jako istotnie głośniejsza niż próbka skompresowana ze stopniem kompresji 1,5 : 1. Między odczuwaną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów wspomnianych próbek przedstawiono w tab. 7.

Tabela 7. Zestawienie próbek odnośnie do saksofonu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą stopnia kompresji

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Saksofon | ratio | 3. | –12,7 | 3,0 : 1 | 20 | 107 | 6,5 | –6,4 |
| 2. | Saksofon | ratio | 1. | –12,7 | 1,5 : 1 | 20 | 107 | 6,4 | –9,2 |

Próbka uznana przez respondentów za subiektywnie głośniejszą charakteryzuje się większym stopniem kompresji – w jej przypadku występuje mniejszy współczynnik szczytu, ale też nieco większa zawartość zniekształceń w stanach przejściowych kompresora, co może prowadzić do odczucia większej głośności.

## 11.5.4. Skrzypce

Skrzypce to instrument obecny w różnorodnych składach instrumentalnych, począwszy od kwartetów smyczkowych i orkiestr symfonicznych, przez comba jazzowe po sekcje smyczkowe towarzyszące zespołom wykonującym muzykę rozrywkową czy kapele ludowe. Skrzypce bardzo chętnie wykorzystuje się w charakterze instrumentu prowadzącego oraz solowego. Niezależnie jednak od charakteru składu instrumentalnego pożądana jest czysta i niezniekształcona barwa skrzypiec. Sprawia to, że poprawne poddanie skrzypiec kompresji dynamicznej jest zadaniem relatywnie trudnym, dlatego często się z niej rezygnuje [11]. Niekiedy jednak specyfika kanału transmisji wymusza zastosowanie kompresji dynamicznej.

Statystycznie istotne różnice w odpowiedziach słuchaczy objawiają się w przypadku czasu zwolnienia. Próbka skompresowana z czasem zwolnienia 32 ms oraz próbka nieskompresowana odbierane są jako wyraźnie głośniejsze od próbki skompresowanej z czasem zwolnienia 21 ms. Jednocześnie nie zostały stwierdzone wyraźne różnice w postrzeganej głośności między wskazanymi próbkami a próbką skompresowaną z czasem zwolnienia 50 ms. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 8.

Próbka nieskompresowana i ze średnim czasem ataku okazują się być tu statystycznie odczuwalnie głośniejsze niż próbka skompresowana z krótkim czasem zwolnienia. Należy jednak zwrócić uwagę, że obie skompresowane próbki charakteryzują się podobnymi wartościami zakresu głośności i najwyższego zrekonstruowanego poziomu szczytowego. Istnieje duże prawdopodobieństwo, że pewna liczba słuchaczy w obliczu nie-

znacznych różnic między próbkami mogła uznać ostatnią prezentowaną próbkę w obrębie pakietu za najgłośniejszą.

Tabela 8. Zestawienie próbek odnośnie do skrzypiec, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą czasu zwolnienia

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Skrzypce | release | 3. | bez kompresji | | | | 4,4 | –7,1 |
| 2. | Skrzypce | release | 4. | –20,5 | 2,1 : 1 | 9 | 32 | 4,2 | –16,7 |
| 3. | Skrzypce | release | 1. | –20,5 | 2,1 : 1 | 9 | 21 | 4,2 | –16,7 |

Statystycznie istotnie różniące się odpowiedzi słuchaczy dotyczące skrzypiec odnosiły się do progu zadziałania kompresora. Próbka skompresowana z progiem zadziałania –18 dBFS okazała się wyraźnie głośniejsza od próbki skompresowanej z progiem zadziałania –20,5 dBFS. Między odczuwaną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat ich parametrów przedstawiono w tab. 9.

Tabela 9. Zestawienie próbek odnośnie do skrzypiec, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą progu zadziałania kompresora

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Skrzypce | threshold | 4. | –18,0 | 3,2 : 1 | 16 | 21 | 4,3 | –16,2 |
| 2. | Skrzypce | threshold | 1. | –20,5 | 3,2 : 1 | 16 | 21 | 4,2 | –18 |

Próbka uznana przez respondentów za subiektywnie głośniejszą charakteryzowała się wyżej ustawionym progiem zadziałania kompresora, skutkiem tego materiał wciąż był dość mocno skompresowany, ale już niekoniecznie sprawiał wrażenie stłumionego.

Liczne statystycznie istotne różnice w odpowiedziach respondentów dotyczące skrzypiec dały się zaobserwować przy zmieniającym się stopniu kompresji. Próbka skompresowana ze stopniem kompresji 2,1 : 1 oraz nieskompresowana okazały się wyraźnie głośniejsze od próbek skompresowanych ze stopniem kompresji równym 1,4 : 1 oraz 3,0 : 1. Nie zostały stwierdzone wyraźne różnice między słyszaną głośnością próbek w obrębie wyszczególnionych grup. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 10.

Tabela 10. Zestawienie próbek odnośnie do skrzypiec, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą stopnia kompresji

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Skrzypce | ratio | 3. | bez kompresji | | | | 4,4 | –7,1 |
| 2. | Skrzypce | ratio | 4. | –20,5 | 3,2 : 1 | 16 | 21 | 4,2 | –16,7 |
| 3. | Skrzypce | ratio | 1. | –20,5 | 1,4 : 1 | 16 | 21 | 4,3 | –16,7 |
| 4. | Skrzypce | ratio | 2. | –20,5 | 2,1 : 1 | 16 | 21 | 4,2 | –14,7 |

W przypadku skrzypiec wystąpiło ciekawe zjawisko: w dwóch parach próbek – pierwsza obejmowała dwa pierwsze przykłady i była istotnie głośniejsza od drugiej obejmującej dwa ostatnie przykłady. Statystycznie głośniejsza okazała się próbka nieskompresowana i skompresowana z dużym stopniem kompresji. Stało się tak prawdopodobnie dlatego, że w przypadku łagodnej kompresji efekty jej działania nie są słyszalne i w konsekwencji uchodzą za cichsze.

## 11.5.5. Werbel

Werbel to instrument o charakterze typowo transjentowym występujący w większych składach wykonujących muzykę poważną czy filmową, w orkiestrach marszowych, ale przede wszystkim jest elementem zestawu perkusyjnego w jazzie i muzyce rozrywkowej, w których wyznacza pulsację rytmiczną utworu. Bardzo głośny werbel z łatwością przebija się nawet przez kilkudziesięcioosobowe składy instrumentalne – i tego oczekuje się w produkcji muzycznej: ma być głośny, mocno zarysowany, z wyraźnie jednak zaznaczonym transjentem. Dlatego właśnie werbel poddawany jest kompresji dynamicznej często i chętnie, a wynikłe stąd zniekształcenia w tym przypadku stanowią zjawisko pożądane.

Tabela 11. Zestawienie próbek odnośnie do werbla, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą czasu ataku

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Werbel | attack | 4. | –7,6 | 3,1 : 1 | 84 | 99 | 3,6 | –6,4 |
| 2. | Werbel | attack | 1. | –7,6 | 3,1 : 1 | 32 | 99 | 3,7 | –6,4 |

W trakcie badań okazało się, że próbka skompresowana z czasem ataku 84 ms okazała się wyraźnie głośniejsza od próbki skompresowanej z czasem ataku 32 ms. Między

odczuwalną głośnością tych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów próbek przedstawiono w tab. 11.

Poza tym – próbka o dłuższym czasie ataku okazała się odczuwalnie głośniejsza od próbki o krótkim czasie ataku, co można łatwo wytłumaczyć, ponieważ kompresor ustawiony z długim czasem ataku kompresuje fazę wybrzmienia werbla i nie ingeruje w transjent aż tak bardzo jak kompresor ustawiony na krótki czas ataku [12]. W przypadku percepcji brzmienia werbla, to właśnie transjent odgrywa jedną z najistotniejszych ról.

## 11.5.6. Wokal

Ludzki głos jest uznawany za najstarszy i najbardziej rozpowszechniony instrument muzyczny. Obecny we wszystkich grupach etnicznych, w kulturze wysokiej i masowej, a także ludowej. Zależnie od okoliczności wykonania utworu oraz kwalifikacji wykonującego może być formą sztuki lub stanowić najbardziej egalitarny sposób uprawiania muzyki (śpiew zarówno w czasie obrzędów o charakterze tradycyjnym czy religijnym, jak i spotkań towarzyskich). Nie należy traktować jednak ludzkiego głosu wyłącznie w kategoriach instrumentu – jest również nośnikiem warstwy lirycznej (czy niekiedy dramatycznej) utworu. W przeszłości stanowił również metodę przekazu informacji (m.in. tzw. pieśni dziadowskie [13]). Biorąc pod uwagę wymienione czynniki oraz to, że przeciętny, niewyedukowany słuchacz skupia się zazwyczaj na linii wokalnej utworu, jest oczywista konieczność zapewnienia jego poprawnej, niezniekształconej (poza sytuacją, w której zniekształcenie jest zabiegiem celowym) słyszalności na tle zespołu instrumentalnego. W przypadku wokalu właśnie kompresja dynamiczna sprowadzona jest do swojego najbardziej typowego zadania – czyli utrzymania zakresu dynamicznego sygnału w zakresie odpowiednim nie tylko dla danego kanału transmisji, lecz także komfortu słuchacza.

Jak wynika z przeprowadzonych testów odsłuchowych, próbka wokalu skompresowana z progiem zadziałania –20,8 dBFS jest postrzegana jako istotnie głośniejsza od próbki skompresowanej z progiem zadziałania –16,9 dBFS. Między odczuwalną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 12.

Próbka uznana przez respondentów za subiektywnie głośniejszą charakteryzowała się bardzo nisko ustawionym progiem zadziałania kompresora, dlatego wystąpił mniejszy współczynnik szczytu, a większa zawartość zniekształceń w stanach przejściowych

kompresora. Nie pojawiły się natomiast statystycznie istotne różnice między próbką o progu zadziałania –20,8 dBFS a próbką o progu zadziałania –14,7 dBFS – prawdopodobnie tu tylko niewielki ułamek materiału jest poddawany kompresji dynamicznej.

Tabela 12. Zestawienie próbek odnośnie do wokalu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą progu zadziałania kompresora

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Wokal | threshold | 1. | –20,8 | 1,9 : 1 | 9 | 29 | 4,1 | –15,3 |
| 2. | Wokal | threshold | 1. | –16,9 | 1,9 : 1 | 9 | 29 | 4,4 | –13,5 |

W przypadku wokalu zaobserwowane zostały istotne różnice w postrzeganej głośności towarzyszące zmieniającemu się stopniowi kompresji. Okazało się, że próbka skompresowana ze stopniem kompresji 1,9 : 1 była istotnie głośniejsza od próbki skompresowanej ze stopniem kompresji 1,6 : 1. Między odczuwalną głośnością wskazanych próbek a pozostałych z danej grupy nie stwierdzono istotnych statystycznie różnic. Szczegółowe informacje na temat parametrów tych próbek przedstawiono w tab. 13.

Tabela 13. Zestawienie próbek odnośnie do wokalu, dla których wystąpiły istotne różnice odczuwanej głośności – ze zmianą stopnia kompresji

| Lp. | Instrument | Kryterium | Pozycja | Threshold [dB] | Ratio | Attack [ms] | Release [ms] | LRA [LU] | HRPL [dBFS] |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Wokal | ratio | 2. | –20,8 | 1,9 : 1 | 9 | 29 | 4,1 | –15,3 |
| 2. | Wokal | ratio | 3. | –20,8 | 1,6 : 1 | 9 | 29 | 4,1 | –14,3 |

Mocniej skompresowana próbka jest percepcyjnie głośniejsza od słabiej skompresowanej, a brak statystycznie istotnych różnic między każdą z omawianych próbek a próbką o stopniu kompresji 1,3 : 1 może wywodzić się z bardzo nieznacznego działania kompresora dynamicznego przy takim ustawieniu.

## 11.6. Podsumowanie

W odniesieniu do uzyskanych wyników oraz doświadczeń zdobytych przy przeprowadzaniu eksperymentu można sformułować zestaw prawidłowości o dość uniwersalnym charakterze:

1. Próg zadziałania kompresora warunkuje odczuwaną przez słuchacza głośność nagrania instrumentu muzycznego we wszystkich – oprócz werbla, analizowanych źródłach dźwięku. Czym niżej jest on ustawiony, tym mniejszym współczynnikiem szczytu charakteryzuje się nagranie, ale i większe jest prawdopodobieństwo wystąpienia odczuwalnych zniekształceń w stanach przejściowych działania kompresora. Przy mało agresywnej kompresji nadmiernie nisko ustawiony próg zadziałania powoduje wrażenie stłumionego brzmienia, co negatywnie wpływa na odbieraną głośność. Wszystko wskazuje na to, że dla wysokich stopni kompresji (powyżej 3 : 1) próg zadziałania należy ustawiać wysoko, a dla małych – nisko. Dokładne wartości progu zadziałania nie zostały zaproponowane, ponieważ jego ustawienie w bardzo dużym stopniu zależy od charakteru sygnału na wejściu kompresora. Przy ustawianiu progu zadziałania, raczej należy kierować się wskazaniami wskaźnika redukcji wzmocnienia.
2. Można wnioskować, że w większości omówionych przypadków im większy jest stopień kompresji, tym nagranie sprawia wrażenie głośniejszego. Wynika to z przyczyn analogicznych do opisanych w p. 1. Dla niewielkich jednak stopni kompresji prawidłowość ta nie zachodzi i nagranie skompresowane w nieznacznym stopniu sprawia wrażenie cichszego od nagrania nieskompresowanego, dlatego należy unikać przesadnie niskich stopni kompresji, czyli poniżej 1,5 : 1.
3. Postrzegana głośność nagrania instrumentu muzycznego wykazuje związek z czasem ataku dla instrumentów, w których naturalna faza ataku jest dość krótka lub dany instrument charakteryzuje się transjentowym czy też agresywnym charakterem brzmienia. Wówczas kompresor z ustawionym relatywnie długim czasem ataku staje się narzędziem modyfikującym obwiednię sygnału instrumentu. Jeśli czas ataku jest dłuższy od fazy ataku w obwiedni sygnału instrumentu muzycznego poddawanego kompresji dynamicznej, zachowany zostaje możliwie niezniekształcony charakter transjentu, przy stłumieniu wybrzmienia, co dodatnio wpływa na postrzeganą głośność instrumentu. Dla werbla czas ataku zapewniający ten efekt mieści się w wartościach 75–85 ms.
4. Zbyt długi czas zwolnienia kompresora przekłada się na zbyt późne zaprzestanie kompresji i redukcję wzmocnienia fragmentów nagrania o poziomie sygnału leżącym dalece poniżej progu zadziałania kompresora – w konsekwencji nierozważny dobór tego parametru negatywnie wpływa na odczuwaną głośność materiału wynikowego. Należy unikać czasów zwolnienia powyżej 120 ms. Czas zwolnienia

kompresora trzeba dobrać do charakteru pulsacji rytmicznej wykonywanego utworu muzycznego.

Praca została sfinansowana z subwencji badawczej Wydziału Inżynierii Mechanicznej i Robotyki AGH w Krakowie, nr 16.16.130.942.

**Słowa kluczowe:** kompresja, głośność, dynamika, parametry, normalizacja, percepcja, instrument, nagranie.

## Bibliografia

[1] Deruty E., *Dynamic Range' & The Loudness War*; https://www.soundonsound.com/sound-advice/dynamic-range-loudness-war [dostęp: 12.04.2019].
[2] Zieliński T.P., Korohoda P., Rumian R., *Cyfrowe przetwarzanie sygnałów w telekomunikacji*, Wydawnictwo Naukowe PWN, Warszawa 2014.
[3] Huber D.M., Runstein R.E., *Modern recording techniques*, Focal Press, 2014.
[4] Kleczkowski P., *The reduction of distortion in the dynamic compressor*, preprint No. 5445, 111 AES Convention, New York, USA, 2001.
[5] Wojtoń M., *Kompresja dźwięku*; http://www.studiomastering.net/mastering05.html [dostęp: 28.03.2019].
[6] ITU-R BS1770-4: Algorithms to measure audio programme loudness and true-peak audio level, 2015.
[7] EBU R-128: Loudness normalisation and permitted maximum levels of audio signal, Genewa 2012.
[8] EBU TECH 3341: Loudness metering: 'EBU Mode' metering to supplement EBU R-128 loudness normalization, Genewa 2016.
[9] EBU TECH 3343: Guidelines for production of programmes in accordance with EBU R-128, Genewa 2016.
[10] Hjortkjaer J., Walther-Hansen M., *Perceptual Effects of Dynamic Range Compression in Popular Music Recordings*, „Journal of Audio Engineering Society" 2014, 62(1), s. 37–41.
[11] Ronan M., Sazdov R., Ward N., *Factors influencing listener preference for dynamic range compression*, preprint No. 9176, 137 AES Convention, Los Angeles, USA, 2014.
[12] Wendl M., Lee H., *The Effect of Dynamic Range Compression on Loudness and Quality Perception in Relation to Crest Factor*, preprint No. 9021, 136 AES Convention, Berlin, Germany, 2014.
[13] Grochowski P., *Dziady. Rzecz o wędrownych żebrakach i ich pieśniach*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2009, s. 264–291.

# 12. Skuteczność klasyfikacji gatunków muzycznych za pomocą sieci neuronowej w zależności od typu danych wejściowych

MACIEJ BLASZKE[1], DAMIAN KOSZEWSKI[2], BOŻENA KOSTEK[3]

[1] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk,
[2] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Katedra Systemów Multimedialnych, ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk,
[3] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Laboratiorium Akustyki Fonicznej, ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk

Rozpoznawanie gatunku muzycznego jest jednym z podstawowych elementów inteligentnych systemów tworzenia automatycznych list muzyki. Platformy strumieniowe oferujące taką usługę wymagają rozwiązań, które umożliwią jak najdokładniej określić przynależność utworu do gatunku muzycznego. Zgodnie z aktualnym stanem wiedzy – najskuteczniejszym klasyfikatorem są sztuczne sieci neuronowe (w tym w wersji uczenia głębokiego), dla których wejście może stanowić spektrogram (postać 2D wektora wejściowego), współczynniki MFCC czy wektor parametrów. We wcześniejszych pracach autorzy opisali opracowaną przez siebie sztuczną sieć neuronową, która z 5-procentowym błędem pozwoliła wyznaczyć zestaw deskryptorów standardu MPEG-7. Mogą one zostać wykorzystane między innymi jako dane wejściowe do klasyfikatora gatunku muzycznego.

W rozdziale zaprezentowano porównanie skuteczności klasyfikatora wykorzystującego architekturę głęboką w zależności od typu danych wejściowych, takich jak: sygnał w postaci czasowej, spektrogram, MFCC, wektor parametrów oraz deskryptory niskopoziomowe standardu MPEG-7 zarówno występujące w bazie danych, jak i te obliczone z wykorzystaniem sieci neuronowej.

# 12.1. Wprowadzenie

W 2002 roku naukowcy z Uniwersytetu w Berkeley przeprowadzili badanie dotyczące dostępnych zasobów związanych z publikacją muzyki w filmie, w Internecie oraz na różnego typu nośnikach [1]. Zasoby te zostały ocenione na 5 mln TB, przy czym już wtedy – tylko w Internecie – sięgały one 170 TB. Obecnie szacuje się, że globalny wzrost dochodów (dane z 2018) w tej dziedzinie wynosi +9,7%, przychody z tytułu transmisji strumieniowej stanowią 46,8% całkowitych przychodów na świecie, wzrost przychodów z płatnych przekazów strumieniowych – +32,9%, a przychody z tytułu pobierania plików – 21,2% [2]. Oznacza to, że w miarę jak rynki muzyczne rozwijają się i ewoluują, zachodzi potrzeba efektywnego zarządzenia muzyką w światowych repozytoriach obejmujących również społeczne systemy muzyczne.

Gatunek muzyczny (definicyjnie) identyfikuje utwory muzyczne jako należące do wspólnej tradycji lub zbioru konwencji. Nie jest tożsamy z formą muzyczną i stylem muzycznym, chociaż w praktyce terminy te są niekiedy używane zamiennie. W specyfikacji kontenera danych ID3v1 można znaleźć 80 typów gatunków muzycznych, które z rozszerzeniem programu Winamp obejmują numerację typów gatunków aż do 191 [3, 4]. Mimo że podział muzyki na gatunki jest w dużym stopniu subiektywny, istnieją kryteria percepcyjne związane z określoną instrumentacją, tempem lub strukturą rytmiczną muzyki, jakie mogą posłużyć do scharakteryzowania określonego gatunku [5, 6]. Na podstawie badań subiektywnych wiadomo, że człowiek potrafi przewidzieć gatunek muzyczny jedynie na podstawie 250-milisekundowego fragmentu, jeśli dany gatunek jest mu znany [7]. Może to sugerować, że ludzie potrafią oceniać gatunki muzyczne jedynie na płaszczyźnie muzycznej, bez konstruowania opisów wyższego poziomu.

Rozpoznawanie gatunku muzycznego jest jednym z podstawowych elementów inteligentnych systemów tworzenia automatycznych list muzyki. Platformy strumieniowe (*streamingowe*) oferujące taką usługę wymagają rozwiązań umożliwiających jak najdokładniejsze określenie przynależności utworu do gatunku muzycznego. Zgodnie z aktualnym stanem wiedzy 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 [8–10].

Dzięki ogólnodostępnym narzędziom (np. MIRToolbox czy Python/libROSA) można wyznaczyć wektory parametrów, które często charakteryzują się dużą redun-

dancją. Wynikowy wektor jest zawsze nadmiarowy ze względu na obecność skorelowanych parametrów. Przykład dokładnie wyspecyfikowanego wektora parametrów to standard MPEG-7 opisany w normie ISO/IEC 15938 [11]. Z jego zastosowaniem uzyskuje się nawet 100-procentową skuteczność w kontekście klasyfikacji dźwięków instrumentów muzycznych.

Główną motywacją do przeprowadzenia badań w tym zakresie 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 własne propozycje deskryptorów konstruowanych w kontekście rozwiązywanego problemu. Kolejnym problemem jest nadmiarowość wektora cech. Aby uzyskać wysoką skuteczność klasyfikacji niezależnie od stosowanego algorytmu, należy dodatkowo stosować metody umożliwiające redukcję proponowanych cech (np. analizę składowych głównych, Principal Component Analysis (PCA) czy analizę korelacyjną) [12, 13]. Autorzy w poprzednich pracach opracowali sztuczną sieć neuronową, która z 5-procentowym błędem wyznacza taki zestaw parametrów [14]. Zaproponowaną sieć można wykorzystać zarówno jako niezależną metodę wyznaczania parametrów, jak i część bardziej złożonej architektury, np. jako jej wejście. Przykładem takiej sieci może być między innymi klasyfikator gatunku muzycznego.

W niniejszym rozdziale przedstawiono przyjęte w przeprowadzonych eksperymentach założenia oraz opis bazy danych muzycznych wykorzystanych w badaniach. W podrozdziale 12.2 opisano bazę SYNAT [12], w której są zawarte wektory cech dla ok. 32 tys. 30-sekundowych utworów muzycznych przydzielonych do 22 gatunków muzycznych. Są to: Alternative Rock, Blues, Broadway & Vocalists, Children's Music, Christian & Gospel, Classic Rock, Classical, Country, Dance & DJ, Folk, Hard Rock & Metal, International, Jazz, Latin Music, Miscellaneous, New Age, Opera & Vocal, Pop, Rap & Hip-Hop, Rock, R & B, Soundtracks. W bazie zawarte są dodatkowe informacje, czyli: nazwisko artysty, nazwa albumu, gatunek i tytuł utworu, numer ścieżki, rok nagrania, inne parametry używane do adnotacji nagrań. Jak wspomniano wcześniej, baza SYNAT zawiera również wektory cech, które zostały wykorzystane w badaniach. Na koniec omówiono podstawy parametryzacji. Następnie opisano zastosowane modele sieci neuronowej w wersji głębokiej architektury (podrozdz. 12.3). Ważnym elementem analizy skuteczności parametryzacji jest redukcja nadmiarowości wektora cech, dlatego uzyskane reprezentacje sprawdzono również w kontekście separowalności danych podawanych na wejście sieci. Kolejny etap obejmował trening i walidację stworzonych modeli. W podrozdziale 12.4 nie tylko przedstawiono wyniki

ewaluacji w kontekście macierzy pomyłek, wartości precyzji i czułości, lecz także wzięto pod uwagę wpływ eliminacji nadmiarowości wektora cech na uzyskaną skuteczność modelu. W tym celu zastosowano liniową analizę dyskryminacyjną (Linear Discriminant Analysis; LDA).

W podrozdziale 12.5 zawarto wnioski i zarysowano plan rozwoju badań.

## 12.2. Dane treningowe

W pracy została wykorzystana baza danych SYNAT zawierająca przykłady utworów muzycznych z wyznaczonymi dla nich zgodnie z definicją deskryptorami MPEG-7 oraz dokładną informację o gatunku, do którego należy dane nagranie [12]. 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. Każdy z nich zawierał odpowiednio następujące liczby przykładów: 9890, 206 i 189 z czterech gatunków: klasyczny, dance, rap i rock [12, 13].

Do porównania wykorzystano pięć typów danych reprezentujących sygnał [14–19]:

1) przebieg czasowy,
2) spektrogram,
3) współczynniki MFCC,
4) parametry wyznaczone za pomocą pyAudioAnalysis,
5) deskryptory MPEG-7 z bazy SYNAT,
6) deskryptory MPEG-7 wyznaczone za pomocą sieci neuronowej.

Dla całego wybranego zbioru wyznaczono wszystkie wymieniowe typy danych z wykorzystaniem 8192 próbek ze środka nagrania.

### 12.2.1. Deskryptory MPEG-7

Przykładem dokładnie wyspecyfikowanego wektora parametrów może być standard MPEG-7 opisany w normie ISO/IEC 15938 [11]. Jego zastosowanie umożliwia uzyskanie nawet 100-procentowej skuteczności w klasyfikacji instrumentu muzycznego. Zawiera on w sobie deskryptory opisujące zarówno przebieg czasowy sygnału, jak i jego strukturę. Można je podzielić na sześć głównych grup:

1) Basic – bazujące na wartości próbek sygnału fonicznego,
2) BasicSpectral – prosta czasowo-częstotliwościowa analiza sygnału,

3) SpectralBasis – jednowymiarowa widmowa projekcja sygnału przygotowana przede wszystkim w celu łatwiejszej klasyfikacji sygnału,
4) SignalParameters – informacja o okresowości sygnału,
5) TimbralTemporal – charakterystyka czasowa oraz barwa muzyczna,
6) TimbralSpectral – deskryptory opisujące zależności liniowo-częstotliwościowe w sygnale.

Dokładne nazwy deskryptorów dla poszczególnych grup (z ich akronimami) przedstawiono w tab. 1.

Tabela 1. Deskryptory standardu MPEG-7

| Grupa | Deskryptory niskiego poziomu | Oznaczenia |
|---|---|---|
| Basic | AudioWaveform, AudioPower | AW, AP |
| BasicSpectral | AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness | ASE, ASC, ASS, ASF |
| SpectralBasis | AudioSpectrumBasis, AudioSpectrumProjection | ASB, ASP |
| SignalParameters | AudioHarmonicity, AudioFundamentalFrequency | AH, AFF |
| TimbralTemporal | LogAttackTime, TemporalCentroid | LAT, TC |
| TimbralSpectral | SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectral-Spread, HamonicSpectralVariation | SC, HSC, HSD, HSS, HSV |

## 12.2.2. Parametry dostępne w pakiecie pyAudioAnalysis

Pakiet pyAudioAnalysis został przygotowany dla języka Python. Dzięki niemu możliwe jest wyznaczenie parametrów sygnału na podstawie jego przebiegu czasowego. Mogą one być wyznaczone i dla całego badanego sygnału, i dla ramek przetworzonych z zadanym oknem. Do eksperymentu wyznaczono takie parametry, jak:

- częstość przejść przez zero,
- energia sygnału,
- entropia energii,
- centroid widma,
- rozpiętość widma,
- tempo zmienności widma,
- częstotliwość, poniżej której znajduje się 95% energii widma,
- współczynniki MFCC.

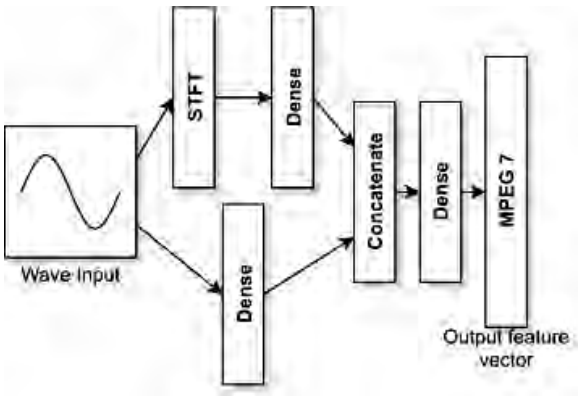
W eksperymencie wyznaczono po jednej wartości dla każdego z parametrów z uwzględnieniem całego badanego fragmentu utworu.

## 12.3. Modele sieci wykorzystanych w eksperymencie

Do badań wykorzystano dwie odrębne architektury sieci neuronowych. Pierwsza z nich stanowi model do wyznaczania deskryptorów MPEG-7 opracowany przez autorów we wcześniejszych pracach [14], druga jest modelem przeznaczonym do klasyfikacji gatunków. Modele te zostały opisane w kolejnych podrozdziałach.

### 12.3.1. Architektura modelu do wyznaczania deskryptorów MPEG-7

W poprzednich pracach autorzy opracowali model sieci neuronowej, której zadaniem jest wyznaczanie deskryptorów standardu MPEG7. Zaproponowana architektura opiera się na warstwach w pełni połączonych, bez pętli sprzężenia zwrotnego. Jako wejście przyjmowany jest wektor 8192 próbek sygnału fonicznego wybierany ze środka pliku audio. Sieć ma dwie rozdzielne ścieżki (rys. 1): zależną od czasu (*Wave Input*) i częstotliwości (STFT), aby pracować w dziedzinie widma, wykorzystano warstwę wykonującą szybką transformację Fouriera. Na końcu ścieżek sygnały były łączone (*Concatenate*) i przekazywane na wejście ostatniej warstwy (*Dense*) w pełni połączonej,




Rys. 1. Architektura modelu do wyznaczania deskryptorów MPEG7 [14];
(tłum. przyjętych na rysunku oznaczeń podano w nawiasach w 1. akapicie podrozdz. 12.3.1)

która wyznacza znormalizowane deskryptory MPEG-7. W celu przywrócenia ich faktycznych wartości należy zaaplikować wcześniej wyznaczony wektor wag normalizacji. Architektura modelu została przedstawiona na rys. 1 [14].

## 12.3.2. Architektura modelu

Do porównania skuteczności klasyfikacji gatunku i jakości klasyfikacji gatunku w zależności do badanego typu danych wykorzystano jedną architekturę modelu różniącą się jedynie wymiarami warstwy wejściowej oraz kierunkiem splotów w warstwach splotowych. Dzięki takiemu podejściu w eksperymencie badana jest tylko jedna zmienna. Model sieci neuronowej wykorzystanej w eksperymencie składa się z w pełni połączonych warstw normalizacji, splotu. Model można podzielić na dwie sekcje: realizującą sploty i redukującą wymiar wektora.


Rys. 2. Architektura modelu do klasyfikacji gatunków

Pierwsza część modelu to następujące po sobie kolejno warstwy normalizacji i splotu dwuwymiarowego o liczbie filtrów 32. Rozmiar jądra przekształcenia jest zależny od wymiarów danych wejściowych – dla danych jednowymiarowych wynosi on (2, 1), a dla dwuwymiarowych – (2, 2). Funkcją aktywacji wykorzystaną przy realizacji operacji splotu jest ReLU [20, 21].

W drugiej części zastosowano warstwy w pełni połączone mające 128 neuronów i 64 neurony oraz ReLU jako funkcję aktywacji. Wyjściem z sieci jest również warstwa w pełni połączona składająca się z liczby neuronów równej liczbie klas i bazująca na funkcji softmax jako funkcji aktywacji [21–25].

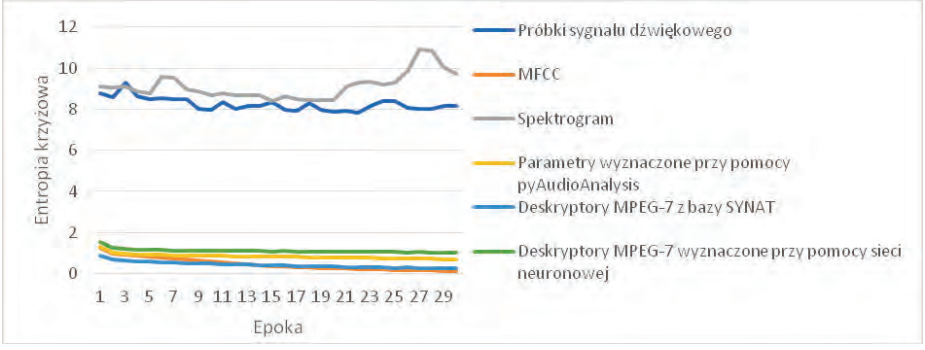
Sposób ułożenia po sobie kolejnych warstw modelu zaprezentowano na rys. 2.

### 12.3.3. Trening modelu

Do treningu modeli wykorzystano pakiet Keras dostępny dla języka Python. Dla każdego z typów danych zastosowano identyczne parametry:

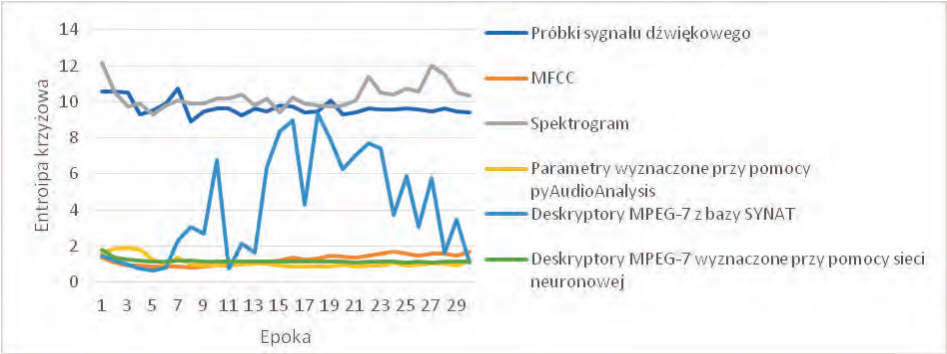
1) rozmiar pojedynczej partii przetwarzania – 64 przykłady,
2) liczba epok – 30,
3) funkcja kosztu – entropia krzyżowa [26],
4) wybór ostatecznego modelu – na podstawie najniższej wartości funkcji kosztu na zbiorze walidacyjnym,
5) optymalizator treningu – Adam [27].

Krzywe przebiegu funkcji kosztu dla zbioru treningowego i walidacyjnego przedstawiono na rys. 3 i rys. 4 – w obu zaprezentowanych przypadkach próbki sygnału dźwiękowego i spektrogram znacznie różnią się od pozostałych typów danych. Ponadto wykres dla deskryptorów MPEG-7 z bazy SYNAT mimo prostego przebiegu na zbiorze treningowym



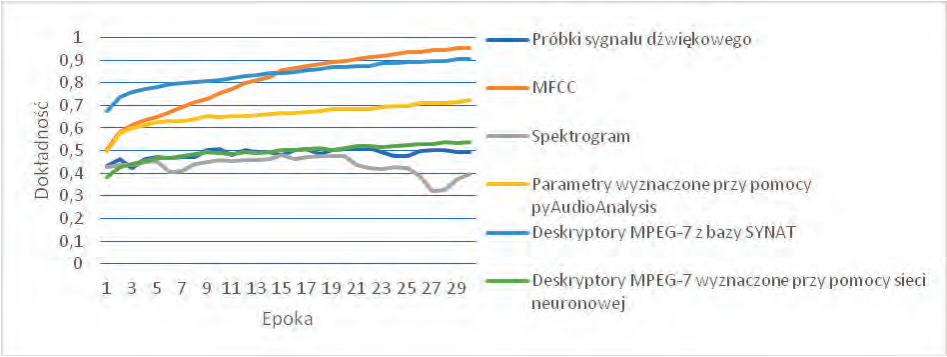

Rys. 3. Funkcja kosztu uzyskana dla zbioru treningowego

w sposób znaczący zmienia swoją wartość w kolejnych epokach na zbiorze walidacyjnym. Wskazuje to na znaczne zmiany w wagach modelu podczas treningu.

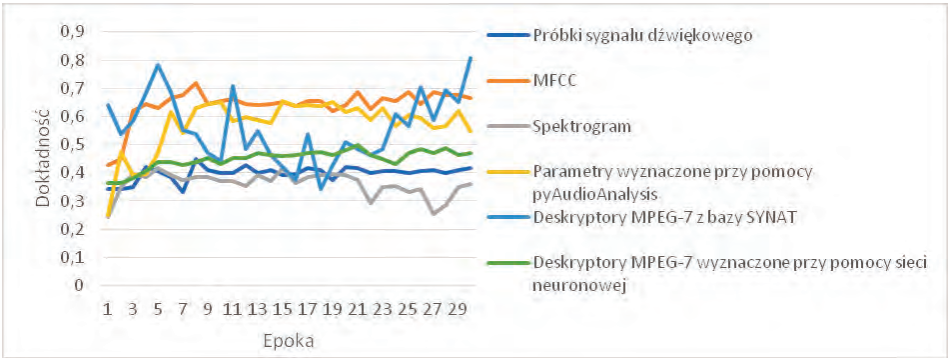



Rys. 4. Funkcja kosztu uzyskana dla zbioru walidacyjnego

Drugą metryką wyznaczaną podczas treningu jest dokładność klasyfikacji. Jej przebiegi przedstawiono na rys. 5 i rys. 6 – są widoczne większe różnice w treningu niż przy obserwacji jedynie funkcji kosztu. Dla zbioru treningowego deskryptory MPEG-7 z bazy SYNAT oraz MFCC jako jedyne osiągają dokładność wyższą niż 90%, podczas gdy kolejny wynik jest dla parametrów wyznaczonych za pomocą pyAudioAnalysis. Zbiór walidacyjny w każdym przypadku daje wartości niższe o przynajmniej 10 punktów procentowych, ale tu (podobnie jak dla zbioru treningowego) najlepiej wypadają deskryptory MPEG-7 z bazy SYNAT, MFCC oraz parametry wyznaczone za pomocą pyAudioAnalysis – osiągają dokładność odpowiednio na poziomie 81%, 72% i 66%. Pozostałe typy danych wejściowych uzyskały dokładność niższą niż 50%.




Rys. 5. Dokładność klasyfikacji uzyskana dla zbioru treningowego




Rys. 6. Dokładność klasyfikacji uzyskana dla zbioru walidacyjnego

## 12.4. Wyniki eksperymentów

Dla każdego z typów danych wybrano najlepszy model, a następnie przeprowadzono na nim testy – otrzymane wyniki w formie macierzy pomyłek przedstawiono w tab. 2–7.

Tabela 2. Macierz pomyłek dla próbek sygnału fonicznego podanych na wejście sieci neuronowej

| Próbki sygnału dźwiękowego | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **0** | 0 | 0 | 100 |
| Dance | 0 | **0** | 6,5 | 93,5 |
| Rap | 0 | 0 | **23,5** | 76,5 |
| Rock | 0 | 0 | 2,6 | **97,4** |

Tabela 3. Macierz pomyłek dla współczynników MFCC podanych na wejście sieci neuronowej

| Współczynniki MFCC | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **41,4** | 41,4 | 0 | 17,2 |
| Dance | 9,7 | **51,6** | 16,2 | 22,5 |
| Rap | 2 | 7,8 | **64,7** | 25,5 |
| Rock | 5,1 | 6,4 | 7,7 | **80,8** |

Tabela 4. Macierz pomyłek dla spektrogramów podanych na wejście sieci neuronowej

| Spektrogram | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **0** | 0 | 10,3 | 89,7 |
| Dance | 0 | **0** | 22,6 | 77,4 |
| Rap | 0 | 0 | **80,4** | 19,6 |
| Rock | 0 | 0 | 52,6 | **47,4** |

Tabela 5. Macierz pomyłek dla parametrów wyznaczonych za pomocą pyAudioAnalysis, podanych na wejście sieci neuronowej

| Parametry wyznaczone przy pomocy pyAudioAnalysis | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **34.5%** | 37.9% | 10.3% | 17.2% |
| Dance | 12.9% | **64.5%** | 3.2% | 19.4% |
| Rap | 7.8% | 2.0% | **66.7%** | 23.5% |
| Rock | 7.7% | 9.0% | 11.5% | **71.8%** |

Tabela 6. Macierz pomyłek dla deskryptorów MPEG-7 z bazy SYNAT podanych na wejście sieci neuronowej

| Deskryptory MPEG-7 z bazy SYNAT | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **62,1** | 27,6 | 0 | 10,3 |
| Dance | 19,4 | **64,5** | 0 | 16,1 |
| Rap | 0 | 2 | **80,4** | 17,6 |
| Rock | 1,3 | 2,6 | 0 | **96,1** |

Tabela 7. Macierz pomyłek dla deskryptorów MPEG-7 wyznaczonych za pomocą sieci neuronowej, podanych na wejście sieci

| Deskryptory MPEG-7 wyznaczone za pomocą sieci neuronowej | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **10,3** | 24,1 | 0 | 65,6 |
| Dance | 3,2 | **19,4** | 3,2 | 74,2 |
| Rap | 2 | 2 | **31,4** | 64,6 |
| Rock | 0 | 3,8 | 6,4 | **89,7** |

Na podstawie tab. 2–7 wyznaczono kolejne dwie miary: precyzję i czułość zdefiniowane następująco:

$$\textbf{precyzja} = \frac{prawdziwie\ dodatni}{prawdziwie\ dodatni + fałszywie\ dodatni} \tag{1}$$

$$\textbf{czułość} = \frac{prawdziwie\ dodatni}{prawdziwie\ dodatni + fałszywie\ ujemny} \tag{2}$$

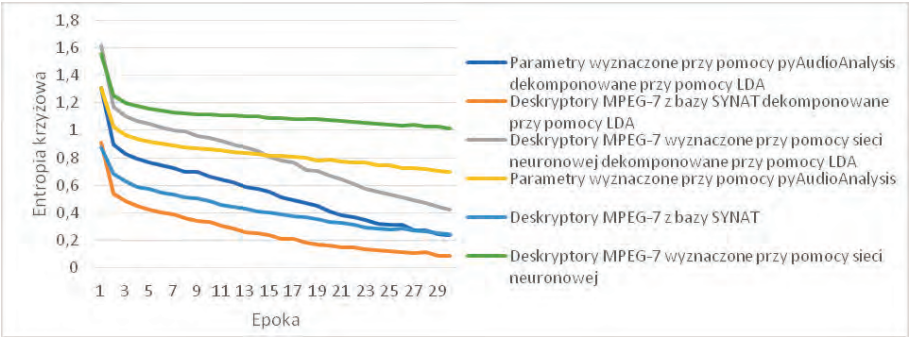
Obliczone wartości miar podano w tab. 8. Można zauważyć, że wartości pokrywają się z wynikami dokładności na zbiorze walidacyjnym – najlepszą jakość klasyfikacji zapewnił model pracujący na deskryptorach MPEG-7 z bazy SYNAT, a następnie współczynnikach MFCC i parametrów wyznaczonych za pomocą pakietu pyAudioAnalysis.

Tabela 8. Wartości miar precyzji i czułości dla poszczególnych typów danych

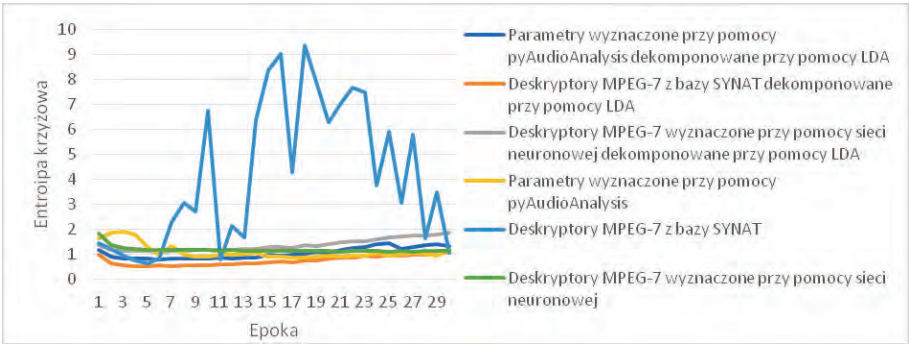
| Typ danych wejściowych | Ważona średnia precyzja na zbiorze testowym | Ważona średnia czułość na zbiorze testowym |
|---|---|---|
| Próbki sygnału dźwiękowego | 0,47 | 0,47 |
| Spektrogram | 0,41 | 0,41 |
| Współczynniki MFCC | 0,66 | 0,66 |
| Parametry wyznaczone za pomocą pyAudioAnalysis | 0,63 | 0,63 |
| Deskryptory MPEG-7 z bazy SYNAT | **0,81** | **0,81** |
| Deskryptory MPEG-7 wyznaczone za pomocą sieci neuronowej | 0,5 | 0,5 |

## Liniowa analiza dyskryminacyjna

W celu poprawienia jakości klasyfikacji przeprowadzona została liniowa analiza dyskryminacyjna Linear Discriminant Analysis; LDA) dla parametrów wyznaczonych za pomocą pyAudioAnalysis, deskryptorów MPEG-7 z bazy SYNAT oraz – sieci neuronowej. Dzięki tej operacji liczba parametrów została zredukowana do 22, co umożliwiło najlepszą separację klas w obrębie danego eksperymentu. Wyniki wykresów funkcji kosztu dla danych przed dekompozycją i po dekompozycji przedstawiono na rys. 7. i rys. 8. Na ich podstawie można wnioskować, że zastosowanie analizy LDA prowadzi do obniżenia wartości funkcji kosztu dla każdego z typów danych.




Rys. 7. Porównanie funkcji kosztu na zbiorze treningowym
dla danych przed zastosowaniem analizy LDA i po jej wykorzystaniu

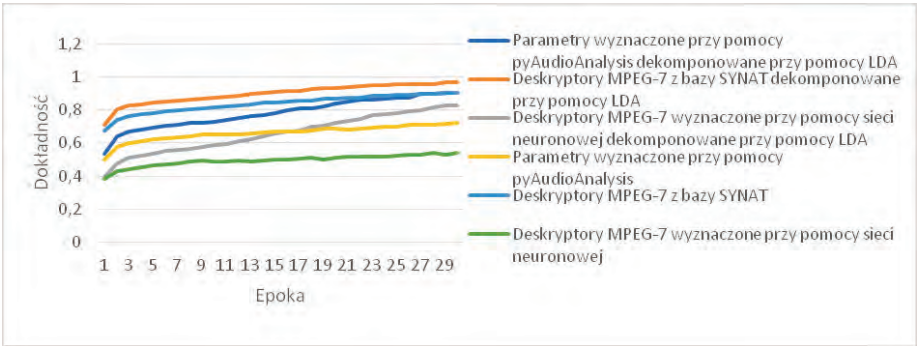



Rys. 8. Porównanie funkcji kosztu na zbiorze walidacyjnym
dla danych przed zastosowaniem analizy LDA i po jej wykorzystaniu

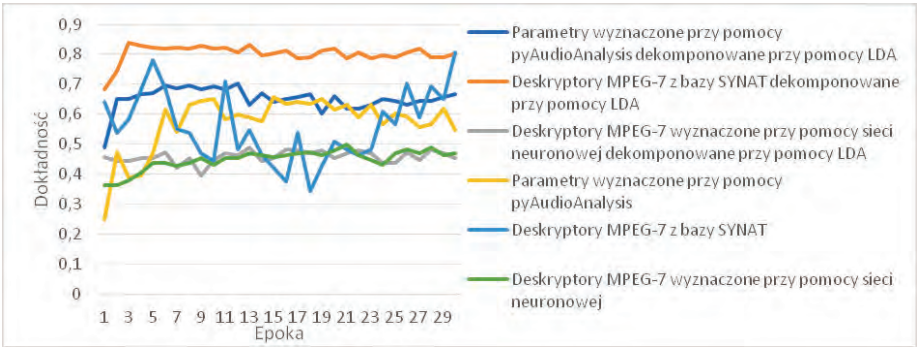
Dekompozycja ma również wpływ na dokładność uzyskaną podczas treningu. Na rysunkach 9 i 10 przedstawiono przebiegi tej wielkości. Podobnie jak w przypadku funkcji kosztu każdy z typów danych – zarówno dla zbioru treningowego, jak i walidacyjnego osiąga wyższe wyniki. Ponownie najwyższą wartość uzyskano dla deskryptorów MPEG-7 z bazy SYNAT.

Na podstawie nowych modeli przygotowano również macierze pomyłek przedstawione w tab. 9–11.

Jeśli porównać wartości miar precyzji i czułości dla modeli wytrenowanych za pomocą danych poddanych liniowej analizie dyskryminacyjnej testowanych z identycznym zbiorem ewaluacyjnym dla każdego typu danych, uzyskuje się wyższe wyniki po dekompozycji. Na podstawie analizy uzyskanych wyników, można stwierdzić, że najlepszą jakość klasyfikacji osiągnięto przy wykorzystaniu deskryptorów MPEG-7 z bazy SYNAT (patrz tab. 12).




Rys. 9. Porównanie dokładności klasyfikacji na zbiorze treningowym dla danych przed zastosowaniem analizy LDA i po jej wykorzystaniu




Rys. 10. Porównanie dokładności klasyfikacji na zbiorze walidacyjnym dla danych przed zastosowaniem analizy LDA i po jej wykorzystaniu

Tabela 9. Macierz pomyłek dla parametrów wyznaczonych za pomocą pyAudioAnalysis po przekształceniu z zastosowaniem liniowej analizy dyskryminacyjnej, podanych na wejście sieci neuronowej

| Parametry wyznaczone przez pyAudioAnalysis przekształcone za pomocą LDA | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **31,0** | 44,8 | 6,9 | 17,2 |
| Dance | 12,9 | **54,8** | 6,5 | 25,8 |
| Rap | 3,9 | 0,0 | **70,6** | 25,5 |
| Rock | 1,3 | 9,0 | 11,5 | **78,2** |

Tabela 10. Macierz pomyłek dla deskryptorów MPEG-7 z bazy SYNAT po przekształceniu za pomocą liniowej analizy dyskryminacyjnej, podanych na wejście sieci neuronowej

| Deskryptory MPEG-7 z bazy SYNAT przekształcone za pomocą LDA | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **69,0** | 27,6 | 0,0 | 3,4 |
| Dance | 12,9 | **71,0** | 3,2 | 12,9 |
| Rap | 0,0 | 0,0 | **92,2** | 7,8 |
| Rock | 1,3 | 3,8 | 1,3 | **93,6** |

Tabela 11. Macierz pomyłek dla deskryptorów MPEG-7 wyznaczonych za pomocą sieci neuronowej po przekształceniu z zastosowaniem liniowej analizy dyskryminacyjnej, podanych na wejście sieci neuronowej

| Deskryptory MPEG-7 wyznaczone za pomocą sieci neuronowej, a przekształcone dzięki LDA | | | | |
|---|---|---|---|---|
| Gatunek muzyczny | Predykcja [%] | | | |
| | Klasyczny | Dance | Rap | Rock |
| Klasyczny | **31,0** | 17,2 | 13,8 | 37,9 |
| Dance | 16,1 | **16,1** | 16,1 | 51,6 |
| Rap | 3,9 | 3,9 | **45,1** | 47,1 |
| Rock | 1,3 | 6,4 | 19,2 | **73,1** |

Tabela 12. Wartości miar precyzji i czułości dla poszczególnych typów danych poddanych liniowej analizie dyskryminacyjnej

| Typ danych wejściowych | Ważona średnia precyzja na zbiorze testowym | Ważona średnia czułość na zbiorze testowym |
|---|---|---|
| Parametry wyznaczone za pomocą pyAudioAnalysis | 0,6 | 0,6 |
| Parametry wyznaczone za pomocą pyAudioAnalysis, przekształcone z zastosowaniem liniowej analizy | 0,65 | 0,65 |
| Deskryptory MPEG-7 z bazy SYNAT | 0,81 | 0,81 |
| Deskryptory MPEG-7 z bazy SYNAT przekształcone za pomocą liniowej analizy | **0,86** | **0,86** |
| Deskryptory MPEG-7 wyznaczone za pomocą sieci neuronowej | 0,5 | 0,5 |
| Deskryptory MPEG-7 wyznaczone za pomocą sieci neuronowej, przekształcone z zastosowaniem liniowej analizy dyskryminacyjnej | 0,5 | 0,5 |

## 12.5. Podsumowanie

Z przeprowadzonych eksperymentów wynika, że najwyższą precyzję klasyfikacji gatunku muzycznego, tj. na poziomie 86%, umożliwia model wytrenowany za pomocą deskryptorów MPEG-7 wyznaczonych na podstawie definicji zawartych w standardzie. W tym przypadku wykorzystano wektory cech obliczone wcześniej w ramach projektu SYNAT, ale z uwzględnieniem analizy nadmiarowości. Najsłabszą jakość klasyfikacji uzyskano w przypadku próbek sygnału fonicznego, które były podawane na wejściu modelu sieci. Również proste przekształcenie z wykorzystaniem transformacji Fouriera nie skutkowało dużą poprawą jakości. Dopiero wyznaczenie MFCC ze spektrogramu doprowadziło do znacznego zwiększenia skuteczności klasyfikacji.

Dzięki eksperymentowi wiadomo również, jak ważny jest dobór odpowiedniego typu danych do zadania, które ma do zrealizowania sieć neuronowa. Między najlepszym i najsłabszym modelem występuje różnica 45 punktów procentowych. Ponadto trzeba zauważyć, że utwory muzyczne zostały zebrane w sposób automatyczny za pomocą robota muzycznego [12]. Oznacza to, że istnieje prawdopodobieństwo, że pewne utwory zostały niepoprawnie przypisane do danego gatunku muzycznego, ze względu jednak na wielkość zbioru błędy nie wpłynęły znacząco na poprawność działania klasyfikatora. Aby uodpornić zatem algorytm na nieprawidłowe oznaczenie danych źródłowych, przyszłe badania należy ukierunkować m.in. na hierarchiczne podejście do gatunków muzycznych, w których uwzględni się połączenia międzygatunkowe [28]. Kolejnym interesującym kierunkiem jest wykorzystanie innych dwuwymiarowych reprezentacji sygnału audio ściślej związanych z charakterystyką muzyki niż z reprezentacją widmową w postaci spektrogramu [29].

**Słowa kluczowe:** uczenie maszynowe, MPEG7, rozpoznawanie gatunków muzycznych.

## Bibliografia

[1] Lyman P., Varian H.R., *How much information*; https://groups.ischool.berkeley.edu/archive/how-much-info-2003/ [dostęp: 13.06.2021].

[2] *IFPI Global Music Report 2019*; https://groups.ischool.berkeley.edu/archive/how-much-info-2003/ [dostęp: 13.06.2021].

[3] Kontener metadanych IDv3.1; https://id3.org/ID3v1 [dostęp: 13.06.2021].

[4] *Gatunki muzyczne w kontenerze IDv3.1*; https://en.wikipedia.org/wiki/List_of_ID3v1_Genres [dostęp: 13.06.2021].

[5] Tzanetakis G., Essl G., Cook P., *Automatic Musical Genre Classification Of Audio Signals*, International Society for Music Retrieval Conference, ISMIR 2001; https://ismir2001.ismir.net/pdf/tzanetakis.pdf [dostęp: 13.06.2021].

[6] Tzanetakis G., Cook P., *Musical genre classification of audio signals*, IEEE *Transactions on Speech and Audio Processing*, July 2002, Vol. 10, No. 5, DOI: 10.1109/TSA.2002.800560, s. 293–302.

[7] Perrot D., Gjerdigen R.O., *Scanning the dial: An exploration of factors in the identification of musical style*, Proceedings of the 1999 Society for Music Perception and Cognition, 1999, s. 88.

[8] Silla Jr. C.N., Kaestner C.A.A., Koerich A.L., *Automatic music genre classification using ensemble of classifiers*, IEEE International Conference on Systems, Man and Cybernetics, Montreal, Que., 2007, DOI: 10.1109/ICSMC.2007.4414136, s. 1687–1692.

[9] Hoffmann P., Kostek B., *Bass enhancement settings in portable devices based on music genre recognition*, „J. Audio Eng. Soc." 2015, 63(12), DOI: 10.17743/jaes.2015.0087, s. 980–989.

[10] DOROCHOWICZ A., KOSTEK B., *A quantitative analysis of music-related features extracted from audio recordings samples*, „Archives of Acoustics" 2018, 43(3), DOI: 10.24425/123922, s. 505–516.

[11] LINDSAY A., HERRE J., *MPEG-7 and MPEG-7 Audio – An Overview*, „Journal Audio Eng. Soc." 2001, 49, 7/8, s. 589–594.

[12] Kostek B., Hoffmann P., Spaleniak P., Kaczmarek A., *Wyszukiwarka nagrań muzycznych – Serwis muzyczny Synat*, „Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne" 2013, 8–9.

[13] Rosner A., Kostek B., *Automatic music genre classification based on musical instrument track separation*, „J. Intell. Inf. Syst." 2018, April, Vol. 50, No. 2, DOI: 10.1007/s10844-017-0464-5, s. 363–384.

[14] Blaszke M., Koszewski D., *Determination of Low-Level Audio Descriptors of a Musical Instrument Sound Using Neural Network*, Signal Processing – Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA 2020, DOI: 10.23919/spa50552.2020.9241264, s. 238–141.

[15] Giannakopoulos T., *PyAudioAnalysis: An open-source python library for audio signal analysis*, PLoS One, 2015, t. 10, 12, DOI: 10.1371/journal.pone.0144610.

[16] MPEG 7 standard; https://mpeg.chiariglione.org/standards/mpeg-7 [dostęp: 13.06.2021].

[17] *TU-Berlin: MPEG-7 Audio Analyzer – Low Level Descriptors (LLD) Extractor*; http://mpeg7lld.nue.tu-berlin.de/ [dostęp: 13.06.2021].

[18] sklearn.discriminant_analysis.LinearDiscriminantAnalysis – scikit-learn 0.23.2 documentation; https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis. LinearDiscriminant Analysis.html#sklearn.discriminantanalysis.LinearDiscriminantAnalysis [dostęp: 13.06.2021].

[19] Feature extraction – librosa 0.7.2 documentation; https://librosa.github.io/librosa/feature.html [dostęp: 13.06.2021].

[20] tf.keras.layers.Conv2D | TensorFlow Core v2.3.1; https://www.tensorflow.org/api_docs/python/tf/keras/layers/Conv2D [dostęp: 13.06.2021].

[21] *A Gentle Introduction to the Rectified Linear Unit (ReLU)*; https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/ [dostęp: 13.06.2021].

[22] tf.keras.activations.relu | TensorFlow Core v2.3.1; https://www.tensorflow.org/api_docs/python/tf/keras/activations/relu [dostęp: 13.06.2021].

[23] *Softmax Activation Function with Python*; https://machinelearningmastery.com/softmax-activation-function-with-python/ [dostęp: 13.06.2021].

[24] tf.keras.activations.softmax | TensorFlow Core v2.3; https://www.tensorflow.org/api_docs/python/tf/keras/activations/softmax [dostęp: 13.06.2021].

[25] tf.keras.layers. Dense | TensorFlow Core v2.3.1; https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dense [dostęp: 13.06.2021].

[26] tf.keras.losses.CategoricalCrossentropy | TensorFlow Core v2.3.1; https://www.tensorflow.org/api_docs/python/tf/keras/losses/CategoricalCrossentropy [dostęp: 13.06.2021).

[27] tf.keras.optimizers.Adam | TensorFlow Core v2.3.1; https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam [dostęp: 13.06.2021].

[28] Dorochowicz A., Kurowski A., Kostek B., *Employing Subjective Tests and Deep Learning for Discovering the Relationship between Personality Types and Preferred Music Genres*, „Electronics" 2020, 9(12), 2016, DOI: 10.3390/electronics9122016.

[29] Tamulevičius G., Korvel G., Yayak A.B., Treigys P., Bernatavičienė J., Kostek B., *A Study of Cross-Linguistic Speech Emotion Recognition Based on 2D Feature Spaces*, „Electronics" 2020, 9(10), 1725, DOI: 10.3390/electronics9101725.

# 13. Automatyczne generowanie kolejności list utworów muzycznych

KAMILA PIETRUSIŃSKA[1], ADAM KUROWSKI[1, 2], BOŻENA KOSTEK[2]

[1] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Katedra Systemów Multimedialnych, ul. Gabriela Narutowicza, 80-233 Gdańsk
[2] Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Laboratorium Akustyki Fonicznej, ul. Gabriela Narutowicza, 80-233 Gdańsk

W niniejszym rozdziale przedstawiono przygotowanie algorytmu do automatycznego układania kolejności utworów muzycznych, zgrywającego je do postaci jednego, długiego miksu. Dzięki algorytmowi dobierane są utwory na podstawie analizy podobieństwa fragmentów końcowych i początkowych utworów. Podobieństwo to jest obliczane za pomocą odległości euklidesowej między wektorami parametrów wyznaczonymi przez autoenkoder oraz na podstawie analizy skupień (*data clustering*). Taki sposób ułożenia utworów umożliwia zapewnienie ciągłości listy. Jakość wyników jest weryfikowana z zastosowaniem testów odsłuchowych przez porównanie automatycznie generowanych list z listami ułożonymi w sposób losowy.

## 13.1. 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 analizie statystycznej, u której podstaw leżą metody eksploracji

dużych zbiorów danych opisujących muzyczne preferencje użytkowników [3, 17]. Szeroko stosowaną metodą łączenia utworów jest „filtracja zespołowa/społecznościowa" (*collaborative filtering*) [3]. Są to metody należące do technik tzw. filtracji zbiorowej oraz metod eksploracji danych. Mogą one niestety powodować, że utwory artystów mało znanych nie pojawią się nigdy na wysokim miejscu listy odtwarzania [17, 31]. Alternatywą dla takiego podejścia może być wykorzystanie sztucznych sieci neuronowych. Neuronowe modele uczenia maszynowego są obecnie powszechnym sposobem rozwiązywania problemów często niedających się zdefiniować w sposób ścisły. Jest to klasa zagadnień, w których rozwiązaniu nie istnieje deterministyczny algorytm, jak ma to na przykład miejsce z operacjami sortowania danych czy wyboru optymalnych rozwiązań [8]. Problemy te charakteryzują się znaczną złożonością i brakiem wyraźnych, matematycznych kryteriów sukcesu. Ocena tego, czy rozwiązanie problemu tego typu jest poprawne, bardzo często ma często charakter subiektywny, uznaniowy – w ogólności – heurystyczny [9]. Jednym z modeli uczenia maszynowego, jaki może być szczególnie przydatny w tym celu, jest uczenie głębokie. Jest to aktualnie widoczne w takich obszarach, jak: rozpoznawanie obrazu, mowy, głosu czy też przetwarzanie języka naturalnego. Wykorzystywane są nie tylko nowe techniki uczenia sieci, lecz także nowe architektury połączeń sztucznych neuronów potrafiących na przykład w automatyczny sposób dokonywać ekstrakcji parametrów ze zbioru uczącego podanego na wejście algorytmu. Dlatego uczenie głębokie nadaje się do takich zadań, jakim jest m.in. automatyczne układanie list odtwarzania. Wykorzystanie zaawansowanych struktur sieci neuronowych sprawdza się chociażby w modelach z mechanizmem skupiania uwagi (*attention-based networks*) czy w tzw. transformatorach (*transformers*) [30]. Stanowią one rozszerzenie sieci neuronowych z mechanizmem skupiania uwagi, które swoje zastosowanie znalazło także w obszarze muzyki – dzięki temu jest możliwe m.in. automatyczne tworzenie utworów muzycznych, Music Transformer [13].

Warto również podjąć temat, czy w przypadku algorytmów sieci neuronowych zasadna jest ekstrakcja parametrów sygnałów muzycznych i tworzenie wektorów parametrów, czy też ze względu na architekturę głęboką możliwe jest wykorzystanie innej reprezentacji, np. 2D, tj. spektrogramów, spektrogramów w skali melowej, chromagramów itd. [18]. Można również zastosować sieć neuronową do generowania cech badanych sygnałów tylko i wyłącznie na podstawie wiedzy wydobywanej przez sieć podczas jej treningu. Wektor parametrów tworzony klasyczną techniką przez projektowanie parametrów opisujących cechy sygnału wejściowego może zawierać zarówno parametry czasowe, częstotliwościowe (wyznaczane na podstawie estymacji widma sygnału),

deskryptory zdefiniowane w standardzie MPEG 7 [21], jak i parametry statystyczne czy proponowane w kontekście rozwiązywanego problemu, co umożliwia pogłębioną obserwację charakterystyk analizowanych sygnałów [12]. Dobór parametrów ma znaczący wpływ na efektywność opracowywanego algorytmu. Dlatego ważne jest, aby poszczególne składowe wektora nie były ze sobą skorelowane, może to bowiem potencjalnie negatywnie wpłynąć na efekty osiągane z pomocą uczenia maszynowego [4], np. przez obniżenie skuteczności klasyfikacji gatunku muzycznego czy znajdowanie mniej podobnych do siebie utworów muzycznych. Należy przy tym zaznaczyć, że wektor parametrów jest bardziej przydatny wówczas, gdy wykorzystywana baza sygnałów muzycznych nie jest duża (tysiące utworów muzycznych). Jeśli natomiast istniejące dane są charakteryzowane raczej objętością danych niż liczbą plików, niewątpliwie możliwe jest wykorzystanie struktur głębokich.

W niniejszym rozdziale przedstawiono proces automatycznego układania list utworów z wykorzystaniem uczenia głębokiego. W pierwszej kolejności przywołano pozycje z literatury przedmiotu, które dotyczą podobnej tematyki badań. Na podstawie analizy opisanych w nich badań przyjęte zostały założenia do własnych eksperymentów, w których wykorzystano sieć neuronową o strukturze autoenkodera umożliwiającą nienadzorowaną naukę parametryzacji na podstawie podanego na jej wejście zbioru wartości. Weryfikację jakości tworzonych list odtwarzania przeprowadzono za pomocą testów odsłuchowych – w ich trakcie badani porównywali listy stworzone przez algorytm z listami przygotowanymi przez generator liczb losowych. Uzyskane wyniki zweryfikowano z wykorzystaniem testów statystycznych. Na końcu rozdziału podano wnioski ogólne oraz perspektywę rozwoju prowadzonych eksperymentów.

## 13.2. Automatyczne układanie list muzycznych

Można zauważyć, że rozwiązania wykorzystujące automatyczne układanie list muzycznych znajdują zastosowanie przede wszystkim w aplikacjach związanych z funkcjonalnością społecznościowych systemów muzycznych prezentujących utwory muzyczne do odsłuchania w postaci tworzonej listy. Zagadnienie automatycznego układania list odtwarzania jest podejmowane m.in. z powodu bardzo rozwiniętych platform streamingowych. Niezwykle popularną funkcjonalnością jest zapewnienie użytkowni-

kowi specjalnie dla niego przygotowanych list utworów dopasowanych do siebie nastrojem, tempem lub zawierających pozycje mogące wpasować się w upodobania muzyczne użytkownika – z uwzględnieniem innych odtwarzanych utworów i kontekstu przypisanego do nich albumu muzycznego.

Listy takie nie są tworzone jednak tylko i wyłącznie na własne potrzeby użytkowników portali streamingowych [27] czy osób słuchających muzyki. Stanowią one podstawę pracy zarówno realizatorów emisji, redaktorów rozgłośni radiowych, DJ-ów, jak i osób przygotowujących oprawę muzyczną uroczystości lub innych wydarzeń. Ważnym elementem tworzenia oprawy muzycznej do audycji radiowych, wydarzeń czy w pracy DJ-a jest pora dnia, w czasie której będzie ona odtwarzana [2, 6]. Muzyka do audycji radiowych wybierana jest w taki sposób, aby w czasie emisji nocą utwory były stonowane – rezygnuje się na przykład z mocnych metalowych brzmień. W ciągu dnia, 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. Wyjątek stanowią autorskie audycje – często muzyczne, w czasie których redaktorzy dobierają muzykę zgodnie z wybranym przez siebie tematem. Duża liczba portali i aplikacji streamingowych [27] umożliwia układanie list odtwarzania i dzielenie się nimi ze wszystkimi użytkownikami. Dostępne są więc playlisty do odtwarzania w konkretnej sytuacji. W przypadku list przygotowanych do odtwarzania w konkretnym celu przy wyborze utworów bierze się pod uwagę ich dynamikę i nastrój. Przykładowo lista odtwarzania przygotowana do słuchania w czasie treningu zawierać będzie utwory szybkie i dynamiczne, a lista przygotowana w celu wyciszenia i odpoczynku będzie zawierała utwory jazzowe lub klasyczne, często muzykę instrumentalną [2, 6]. Listy odtwarzania są również wykorzystywane w różnych obszarach, np. w sklepach znanych sieci muzyka może też stać się wizytówką danej firmy.

Innym możliwym zastosowaniem jest automatyczne układanie list utworów muzycznych stanowiących podkład dla interaktywnej rozrywki, m.in. gier komputerowych. W takim przypadku zarówno muzyka, jak i pozostałe efekty dźwiękowe muszą być zgodne ze płaszczyzną wizualną i narracyjną, co wzmocni w ten sposób efekt artystyczny stworzony przez grę [26].

Problem automatycznego generowania list odtwarzania pojawia się w literaturze badawczej głównie dlatego [8], że nie znaleziono rozwiązania charakteryzującego się dużą skutecznością, w pełni zadowalającą użytkowników. Głównym powodem takiego stanu rzeczy jest subiektywność odbioru muzyki. Ponieważ obecnie listy muzyczne tworzy się za pomocą uczenia maszynowego (w tym uczenia głębokiego, np. z wykorzystaniem sieci neuronowych typu autoenkoder [25]), warto temat zgłębiać, mając na

uwadze tego typu algorytmy. Na przykład Van den Oord, Dieleman oraz Schrauwen w swojej pracy dotyczącej rekomendacji utworów muzycznych [29] wykorzystali sieci głębokie wytrenowane na 3 s utworów wybranych w sposób losowy. W eksperymentach została wykorzystana baza danych: The Million Song Dataset [1]. Autorzy tej pracy oparli badania na podejściu stosowanym w obszarze MIR (Music Information Retrieval), czyli poddali obserwacji lokalne cechy sygnału fonicznego. W ten sposób pozyskali wektor parametrów zawierający 13 współczynników MFCC (Mel-Frequency Cepstral Coefficients) z sygnału przy segmentacji na okna wielkości 1024 próbek, odpowiadającym 23 ms przy częstotliwości próbkowania równej 22,05 kHz. Dodatkowo obliczone zostały różnice między parametrami melcepstralnymi pierwszego i ostatniego rzędu – uzyskano 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 wykorzystującego model „*bag-of-words*". W przywołanej publikacji jednym z wniosków było stwierdzenie, że wiele aspektów dotyczących rekomendacji utworów charakteryzuje się subiektywnym podejściem, ale niektóre cechy mogą być przewidziane na podstawie sygnału fonicznego.

Inne rozwiązanie dotyczące rekomendacji utworów muzycznych opisano w publikacji z 2017 r. – jest nim algorytm automatyczny DJ-a [16], dzięki któremu w sposób w pełni automatyczny układana jest lista odtwarzania na podstawie narzuconego zbioru utworów. Przy tym zapewnione zostaje również płynne przejście między kolejnymi utworami. Parametry sygnału wykorzystane do wyboru kolejnego utworu są w tym konkretnym przypadku ekstrahowane za pomocą sieci głębokich wytrenowanych do rozpoznawania gatunku muzycznego.

Kolejnym przykładem opisanym w literaturze przedmiotu dotyczącym automatycznego rozpoznawania muzyki jest wykorzystanie kilku klasyfikatorów, w tym modelu generatywnych sieci przeciwstawnych (Generative Adversarial Networks; GAN) [7]. Jako narzędzie parametryzacji zastosowano w tym przypadku sieć neuronową. Stworzone zostały dwa systemy – jeden oparty na sztucznych sieciach neuronowych, w którym zbiorem danych była baza GTZAN [10, 28], drugi wykorzystujący sieci splotowe – tu zbiorem danych była baza Latin Music Database [19]). W celu zbadania skuteczności tych dwóch systemów użyty został klasyfikator bazujący na lasach losowych (Random Forest). Stwierdzono, że lepszą skuteczność rozpoznawania muzyki uzyskuje się w rozwiązaniu przy użyciu sieci splotowych.

Przytoczone algorytmiczne techniki układania list utworów muzycznych stanowiły podstawę do sformułowania założeń własnych eksperymentów.

## 13.3. Założenia eksperymentu

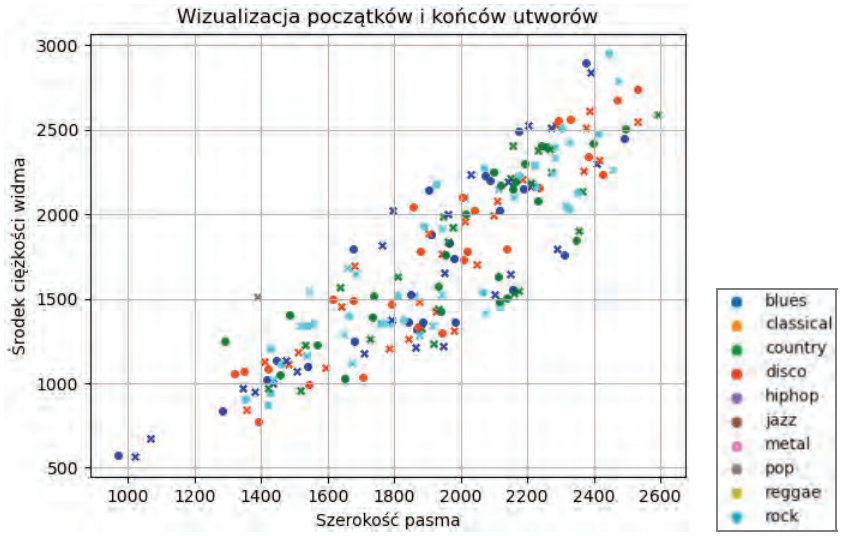
Zaprojektowanie algorytmu doboru utworów muzycznych wymaga podzielenia tego procesu na kilka etapów. Pierwszym krokiem może być wybór algorytmu. Przede wszystkim jednak konieczne jest zidentyfikowanie zbioru danych, który stanowić będzie podstawę treningu algorytmu, następnie konieczny jest wybór i opracowanie metod wstępnego przetwarzania danych. Ważne jest także określenie, według jakich reguł algorytm będzie się kierować w procesie układania list. Wybór ten będzie rzutować na dokładny wybór algorytmu najlepiej dopasowanego właśnie do preferowanego sposobu układania list utworów muzycznych [22].

W związku z tym, że projektowany algorytm wykorzystuje do działania nieprzetworzone utwory muzyczne, zrezygnowano z baz zawierających wektory parametrów, do których zalicza się m.in. bardzo popularna baza: The Milion Songs Dataset [1]. Rozważano repozytoria muzyczne zapewniające dużą liczbą danych, a jednocześnie zróżnicowane. Dlatego wybrano bazę GTZAN [10, 28] – jedną z najwcześniej opracowanych i udostępnionych baz utworów muzycznych. Podzielona jest ona na dziesięć gatunków muzycznych: blues, muzyka poważna, country, disco, hiphop, jazz, metal, pop, reggae, rock. W każdym gatunku przygotowano 100 30-sekundowych fragmentów. Pozyskany w taki sposób sygnał muzyczny musi być poddany procesowi parametryzacji, czyli wyodrębnienia cech (zwanych także parametrami czy deskryptorami).

W fazie wstępnej eksperymentu wykorzystano do parametryzacji bibliotekę librosa dostępną dla języka programowania Python [20]. Deskryptory zawarte w bibliotece librosa są podzielone na kilka grup: parametry „widmowe" (np. płaskość widma, środek ciężkości – do tej grupy przypisano jednak również częstość przejść przez zero czy wartość rms i reprezentacje 2D, tj. chromagram czy spektrogram w skali melowej), cechy związane z rytmem, parametry wynikające z przekształceń innych deskryptorów. W ekstrahowanym wektorze parametrów wykorzystane zostały deskryptory widmowe. Algorytm wczytywał pliki z danego folderu źródłowego, następnie zapisywał pierwsze i ostatnie 10 s utworów do zmiennych pomocniczych i na koniec obliczał wybrane podstawowe parametry widmowe opisane w standardzie MEPG 7, czyli np. środek ciężkości widma i płaskość widma.

Parametry obliczane dla 10-sekundowych fragmentów były następnie uśredniane. Proces dopasowania utworów polegał na obliczeniu odległości euklidesowej między wartościami odpowiadającymi końcowi aktualnie analizowanego utworu a war-

tościami odpowiadającym poszczególnym deskryptorom dla początków utworów. Przykład wyników parametryzacji zobrazowano na rys. 1. Końcowym wynikiem działania skryptu był plik tekstowy z zapisanymi ścieżkami utworów ułożonymi w „odpowiedniej" wg skryptu kolejności oraz plik w formacie m3, który umożliwiał odtworzenie ułożonych utworów zapisanych w kolejności pokrywającej się z plikiem tekstowym w wybranych odtwarzaczach obsługujących listy odtwarzania w tym formacie.




Rys. 1. Przykład wizualizacji początku („○") i końca utworu(„×");
kolorem oznaczono gatunek muzyczny

Wyniki uzyskiwane za pomocą parametryzacji z wykorzystaniem parametrów zawartych w bibliotece librosa były obiecujące, skrypt nie pracował jednak wydajnie, a parametry nie uwzględniały specyfiki subiektywnego odbioru dźwięku przez słuchacza. Z tego powodu postanowiono zastosować parametry melcepstralne oraz autoenkoder, którego zadaniem była parametryzacja pierwszych i ostatnich 10 s.

Struktura autoenkodera umożliwia nienadzorowaną naukę parametryzacji na podstawie podanego zbioru wartości na wejście sieci. W procesie nauki tego zadania algorytm jest w stanie zidentyfikować zależności między danymi – w tym zidentyfikować podobieństwa między przykładami podanymi na jego wejście. Dlatego do tego celu wykorzystano algorytm autoenkodera wariacyjnego (*variational autoencoder*), który może zostać użyty także jako algorytm generatywny. Bardziej szczegółowy opis dzia-

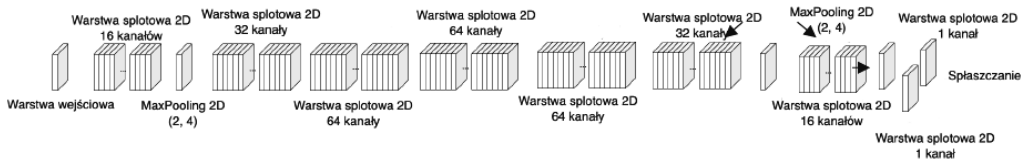
łania autoenkoderów wariacyjnych można znaleźć w licznych publikacjach, np. [5, 15]. W prezentowanym algorytmie autoenkoder wykorzystany jest do identyfikowania tych par utworów, dla których zakończenie pierwszego i początek drugiego są do siebie podobne.

Zadaniem algorytmu jest zatem ponowna analiza 10-sekundowcyh fragmentów początków i końców utworów znajdujących się w bazie. Analogicznie do poprzedniego procesu dopasowanie utworów polega na obliczeniu odległości euklidesowej między wartościami odpowiadającymi końcowi aktualnie analizowanego utworu a wartościami początku następnego.

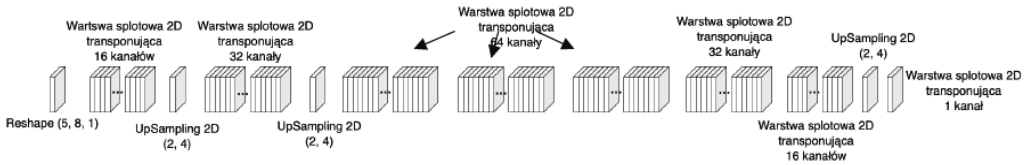
## 13.4. Architektura autoenkodera wariacyjnego

Jak już wcześniej wspomniano, do wyznaczenia kolejności utworów na listach odtwarzania wykorzystano autoenkoder wariacyjny. Jedną z cech tego algorytmu jest dążenie do umieszczania w przestrzeni reprezentacji podobnych fragmentów muzyki blisko siebie, co jest zgodne z założeniami układania list utworów muzycznych, w których zasada działania polega na umieszczaniu obok siebie podobnych 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 – część z nich została dobrana tak, aby zgodna była z bazą danych stanowiących zbiór uczący [23]. Algorytm przetwarza monofoniczne pliki zapisane w formacie wav, dla których częstotliwość próbkowania to 22 050 Hz. Dane przeznaczone do wprowadzenia na wejście autoenkodera przeliczane są na reprezentację 2D sygnału, tj. wykorzystywane są mfcc-gramy o długości 10 s (co odpowiada długości 512 próbek i rozdzielczości 40 współczynników melcepstralnych). Obliczone wartości znormalizowano do zakresu od –1 do 1. Autoenkoder składa się z dwóch części: kodera dokonującego przekształcenia mfcc-gramu na wektor reprezentacji o długości 40 współczynników i dekoder rekonstruującego mfcc-gram na podstawie wynikowych wektorów uzyskiwanych z kodera. W trakcie tworzenia architektury sieci sprawdzano jakość dekodowania uzyskaną na wyjściu autoenkodera. Na rysunkach 2 i 3 przedstawiono strukturę autoenkodera (z podziałem na enkoder i dekoder), dla której uzyskano najlepsze wyniki.

Drugą ważną cechą autoenkodera wariacyjnego jest to, że koder ma dwa wyjścia, ponieważ zamiast konkretnych wartości reprezentacji zwraca rozkład prawdopodobieństwa wartości reprezentacji powiązanych z mfcc-gramem na wejściu algorytmu. Koder ten na wyjściu generuje dwa wektory – jeden z nich reprezentuje wartości średnie, a drugi wariancje. Opisują one rozkład wartości składowych wektora losowego, którego realizacja stanowi następnie wejście dekodera. 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 między rozkładem występowania punktów kodowanych przez koder a równomiernym rozkładem punktów, co umożliwia właśnie wymuszenie takiego rozkładu punktów w przestrzeni reprezentacji. Drugim składnikiem funkcji straty jest błąd rekonstrukcji mfcc-gramu na wyjściu względem oryginału przekazanego na wejście.




Rys. 2. Struktura enkodera




Rys. 3. Struktura dekodera

Ostatnim etapem realizacji algorytmu układającego listy odtwarzania jest skrypt wykorzystujący stworzone przez autoenkoder wektory parametrów. Dla każdego utworu generowany jest wektor składający się z 40 wartości będących reprezentacją początkowego fragmentu o długości 20 s. Analogiczny wektor generowany jest dla końcowego fragmentu utworu. Następnie (o czym pisano już wcześniej) obliczana jest odległość euklidesowa między końcem a początkiem kolejnego utworu w 40-wymiarowej hiperprzestrzeni utworzonej przez parametry wyznaczone przez autoenkoder. Końcowym wynikiem działania skryptu są dwa pliki: jeden tekstowy z zapisanymi ścieżkami do utworów, drugi w formacie m3 umożliwiający odtworzenie listy w odtwarzaczu multimedialnym.

## 13.5. Testy odsłuchowe

W celu zweryfikowania wyników uzyskanych za pomocą stworzonego skryptu przeprowadzono ankietę internetową, w której respondenci udzielali odpowiedzi po wysłuchaniu przykładów list muzycznych odtwarzanych na własnym sprzęcie. Odpowiedzi z podstawowymi informacjami o samym 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 pięć zestawów list, w których jedna stworzona była przez algorytm, a druga wylosowana generatorem liczb losowych. W celu ułatwienia zadania postanowiono wybrać utwory dość popularne lub dobrze reprezentujące dany gatunek muzyczny. Ze względu na to, że celem testu było sprawdzenie, w jakim stopniu koniec jednego utworu jest dopasowany do początku kolejnego, wycięto 15-sekundowe fragmenty z początku i końca każdego utworu i połączono je we właściwej kolejności.

Test zaprojektowano zgodnie z zaleceniami zawartymi w normie ITU-R BS 1284-1 [14]. Każda lista zawierała osiem utworów, a odsłuchanie jednego zestawu (czyli listy ułożonej przez algorytm i drugiej – w losowej kolejności) trwało ok. 8,5 min. Oprócz przygotowanych zestawów zapewniono badanym możliwość odsłuchania 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 oceniono za pomocą zmodyfikowanej skali zawartej w normie ITU-R BS 1284-1 [14]. 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.

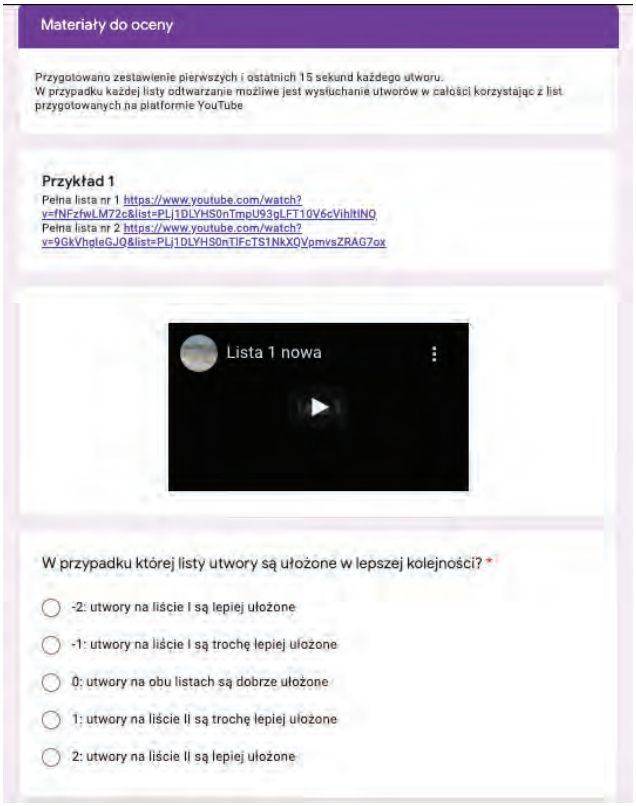
Aby ułatwić badanym proces podejmowania decyzji i możliwie skrócić czas trwania testu, którego odsłuchanie zajmowało ok. 45 min, zmniejszono skalę podaną w zaleceniu do wartości od –2 (preferowana jest lista stworzona przez algorytm) do 2 (preferowana jest lista stworzona przez generator liczb losowych). Długość testu nie była jednak ograni-

czana i ankietowani mogli odsłuchiwać przygotowaną listę utworów wielokrotnie. Pierwszą listą zawsze była ta stworzona za pomocą algorytmu. Respondenci nie posiadali informacji o tym, która lista jest stworzona przez algorytm, a która – losowa. Poinformowani zostali jednak o tym, że zawsze jedna lista jest stworzona przez wylosowanie utworów generatorem liczb losowych, a druga przez stworzony algorytm dobierający utwory przez dopasowanie ostatnich 10 s jednego utworu do pierwszych 10 s kolejnego. Przy wyborze utworów do list odtwarzania (mimo że w większości były to kompozycje uważane za popularne, a lista była krótka) zapewniono ich zróżnicowaną dynamikę, możliwość stopniowej zmiany tematu muzycznego i taki dobór, by utwory nie zawierały się tylko w obrębie jednego gatunku muzycznego lub bardzo do niego zbliżonych. Na rysunku 4 przedstawiono interfejs opracowany do realizacji testów odsłuchowych, a w tab. 1 przykład listy odtwarzania [22].

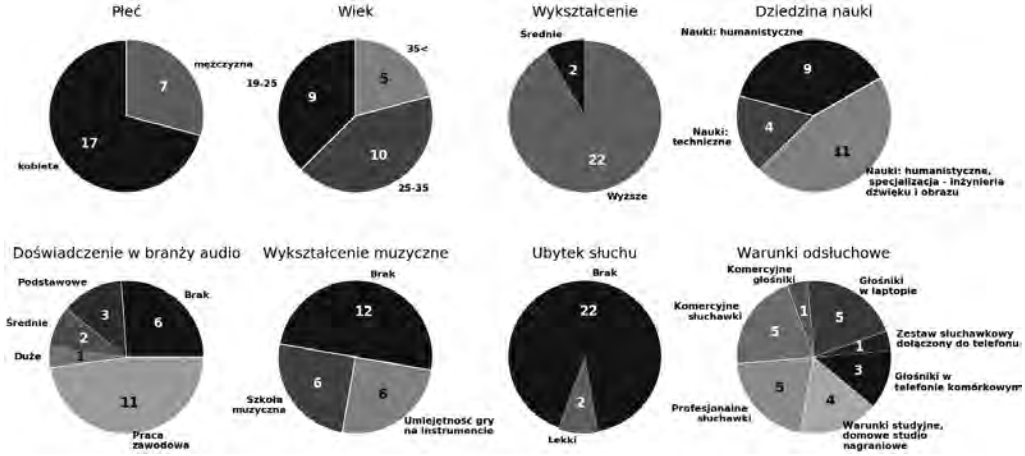
Tabela 1. Listy odtwarzania zawarte w teście odsłuchowym w pytaniu nr 1 [22]

| Numer utworu na liście | Typ listy | Utwór i wykonawca |
|---|---|---|
| 1. | stworzona przez algorytm | *Stayin Alive* – Bee Gees |
| 2. | | *You go to my head* – Przemek Dyakowski |
| 3. | | *Walking on a dream* – Empire of the Sun |
| 4. | | *Corazon espinado* – Santana ft Manna |
| 5. | | *Shine on me* – Dan Auerbach |
| 6. | | *Dirty Laundry* – Bitter Sweet |
| 7. | | *Psycho Killer* – Talking Heads |
| 8. | | *Boys don't Cry* – The Cure |
| 1. | stworzona przez losowanie kolejności | *Boys don't Cry* – The Cure |
| 2. | | *Shine on me* – Dan Auerbach |
| 3. | | *You go to my head* – Przemek Dyakowski |
| 4. | | *Corazon espinado* – Santana ft Manna |
| 5. | | *Psycho Killer* – Talking Heads |
| 6. | | *Walking on a dream* – Empire of the Sun |
| 7. | | *Stayin Alive* – Bee Gees |
| 8. | | *Dirty Laundry* – Bitter Sweet |

W badaniu łącznie udział wzięły 24 osoby: 6 kobiet i 18 mężczyzn. Szczegóły zebrane na temat badanych przedstawiono na rys. 5. Żadna z osób nie zgłosiła problemów ze słuchem.




Rys. 4. Interfejs testu odsłuchowego [22]




Rys. 5. Wykresy kołowe nt. osób biorących udział w testach – ich wykształcenia, doświadczenia w branży audio, wykształcenia muzycznego, ubytku słuchu, i warunków odsłuchowych [22]

## 13.6. Wyniki testów odsłuchowych

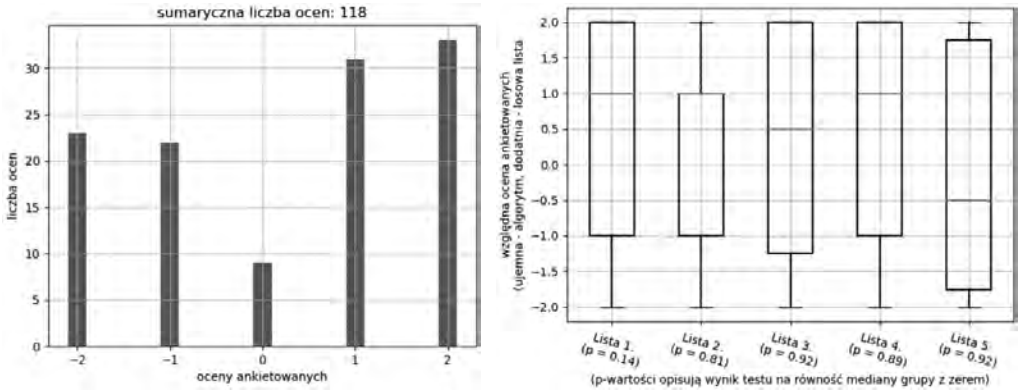
Na podstawie przeprowadzonych ankiet pozyskano zbiór 24 odpowiedzi oceniających listy w skali dyskretnej, od –2 (preferowana jest lista stworzona przez algorytm) do 2 (preferowana jest lista stworzona przez generator liczb losowych).

Do analizy statystycznej wykorzystano test Kruskala–Wallisa oraz test $\chi^2$. Wszystkie wnioski zostały oparte na przyjętym założeniu, że poziom istotności ma wartość 0,05. Test Kruskala–Wallisa został wybrany z tego względu, że dla wszystkich rozpatrywanych przypadków nie były spełnione założenia umożliwiające zastosowanie testu ANOVA [11]. Sprawdzenie dokonane zostało testem Levene'a (równości wariancji) i Shapira–Wilka (odnośnie do normalności ocen) [11]. Test Shapira-Wilka był przeprowadzany dla każdego zestawu ocen oddzielnie, w tym przypadku zastosowano zatem poprawkę na wielokrotne testowanie (metodą Simesa–Hochberga). W celu implementacji testów wykorzystano język Python i pakiet SciPy [15].

Najpierw analizę przeprowadzono na całym zbiorze badanych – bez podziału ankietowanych zgodnie z ich wykształceniem czy doświadczeniem w branży audio. Posłużono się testem Kruskala–Wallisa. Uzyskana wynikowa $p$-wartość wyniosła 0,691. Można zatem wnioskować, że badani nie mieli preferencji uzależnionych od tego, czy lista została ułożona przez zaproponowany algorytm czy w kolejności losowej. Następnie przeprowadzony został test mający na celu sprawdzenie, czy ankietowani słyszeli różnicę między wynikami algorytmów. W tym celu zastosowano test $\chi^2$ oraz test post-hoc, który polegał na wielokrotnym testowaniu poszczególnych par zmiennych za pomocą testu $\chi^2$. Obliczono w ten sposób macierz $p$-wartości, co umożliwiło stwierdzenie, czy różnice częstości wyboru danej odpowiedzi są istotne statystyczne. Także w tym przypadku zastosowano poprawkę na wielokrotne testowanie metodą Simesa–Hochberga [24]. Wyniki uzyskane dla ogółu badanych przedstawiono w tab. 2 i na rys. 6.

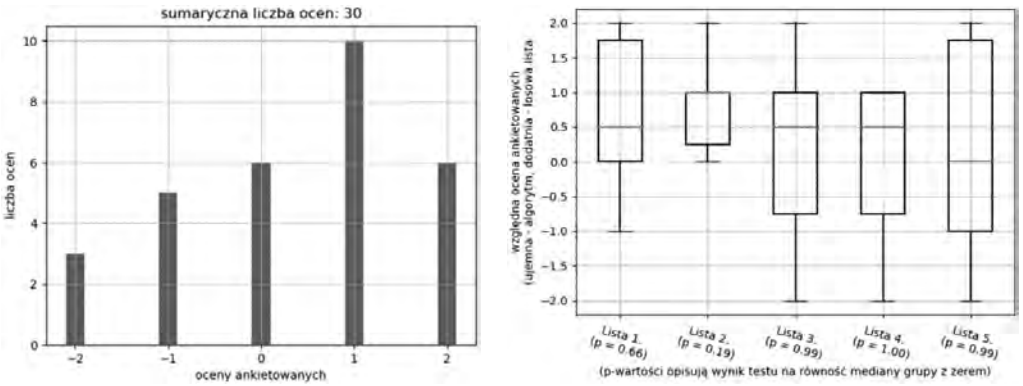
Tabela 2. Macierz testu post-hoc bazującego na teście $\chi^2$ uzyskana w analizie ogółu badanych; pogrubioną czcionką oznaczono $p$-wartości mniejsze od poziomu istotności (0,05)

| 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 | – |




Rys. 6. Histogram wyników uzyskanych dla ogółu ankietowanych za pomocą testu $\chi^2$; zobrazowanie uzyskanych odpowiedzi na wykresie pudełkowym

Na podstawie analizy wykresu pudełkowego (rys. 6) można wnioskować, że ogół ankietowanych nie miał preferencji odnośnie do tego, jaki algorytm został użyty do ustalenia kolejności utworów na odsłuchiwanych przez nich listach odtwarzania. Ale z histogramu odpowiedzi oraz istotności niektórych częstości wybierania odpowiedzi między zerem a pozostałymi ocenami wynika, że ankietowani mogli słyszeć różnicę między algorytmami. Część ankietowanych w tym przypadku jednak preferowała losowy układ list i dlatego w ogólnym ujęciu żaden algorytm nie był faworyzowany. Taka obserwacja spowodowała, że przeprowadzone zostały analizy we wszystkich możliwych podgrupach wyznaczonych za pomocą zebranych w ankiecie informacji o osobach ankietowanych.



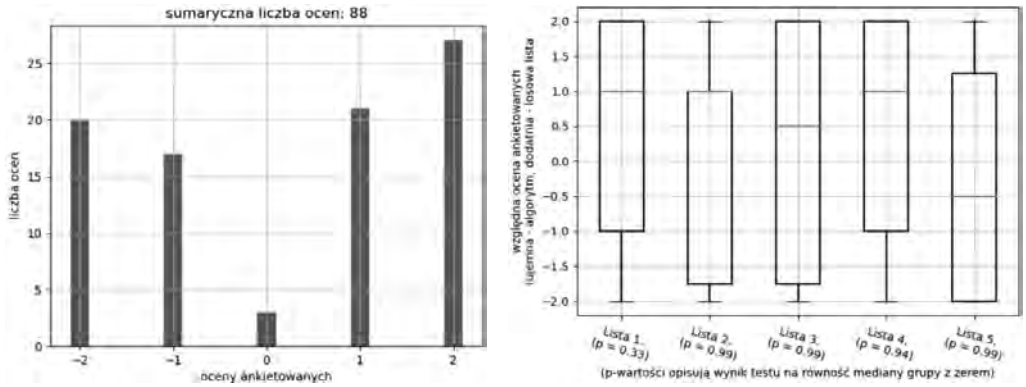

Rys. 7. Histogram uzyskany po analizie wyników otrzymanych z przeprowadzonego testu $\chi^2$ w grupie osób bez doświadczenia w branży audio – zobrazowanie odpowiedzi na wykresie pudełkowym

Jedną z przeprowadzonych w ten sposób analiz wykonano z udziałem osób, które nie posiadają żadnego doświadczenia zawodowego w branży audio – jej efekty zwizualizowano na rys. 7. Na podstawie testu post-hoc bazującego na teście $\chi^2$ wykazano, że żadna z różnic widocznych na rys. 7 nie jest istotna statystycznie. Dlatego można wnioskować, że ankietowani zaznaczali wszystkie możliwe odpowiedzi z podobną częstością. Można założyć, że brak doświadczenia zawodowego, a tym samym wytrenowania słuchu w kierunku percepcji muzyki, powoduje brak wrażliwości na postrzeganie zmian nastroju utworów, stąd ankietowani nie preferowali żadnego z przedstawionych algorytmów.

Zbadano również preferencje badanych deklarujących przynajmniej podstawowe doświadczenie w branży audio. W tym przypadku również nie można było stwierdzić, który sposób układania list odtwarzania jest lepszy z wykorzystaniem mediany uzyskanych przez algorytm ocen. Możliwe było jednak stwierdzenie, że badani słyszą

Tabela 3. Macierz testu post-hoc bazującego na teście $\chi^2$ uzyskana w analizie dotyczącej osób 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 | – |




Rys. 8. Histogram uzyskany po analizie wyników testu $\chi^2$ pochodzących od osób, które zadeklarowały doświadczenie w branży audio – zobrazowanie uzyskanych odpowiedzi na wykresie pudełkowym

różnice między prezentowanymi listami odtwarzania, ale ich preferencje są różne. Różnice w częstości zaznaczania odpowiedzi 0 względem dowolnej innej odpowiedzi są bowiem istotne statystycznie – otrzymane wyniki zobrazowano w tab. 3 oraz na rys. 8.

## 13.7. Podsumowanie

W ramach przeprowadzonych badań został opracowany algorytm mający za zadanie automatyczne układanie listy odtwarzania utworów muzycznych. Uzyskane wyniki zostały poddane testom odsłuchowym – ich analiza statystyczna umożliwiła sformułowanie następującego wniosku: badani niemający doświadczenia w branży audio najprawdopodobniej nie zauważają różnic między prezentowanymi listami odtwarzania. W grupie osób z doświadczeniem w technologii studyjnej i pracujących zawodowo z dźwiękiem natomiast część z nich preferuje sposób układania list odtwarzania realizowany przez zaprezentowany algorytm, a część – woli listy ułożone przez zwykłe losowanie kolejności. Wniosek ten oparto na analizie częstości wyboru odpowiedzi 0 wskazującej na brak słyszenia różnic między prezentowanymi listami utworów układanymi przez różne algorytmy. Praktycznym skutkiem tej obserwacji jest konkluzja, że proponowany algorytm może być elementem algorytmu układania list w inteligentnym odtwarzaczu muzyki z jednoczesnym zaleceniem, aby użytkownik takiego odtwarzacza miał możliwość wyboru, czy utwory są układane w kolejności zaproponowanej przez algorytm czy losowej.

Dodatkowe interesujące wnioski można wyciągnąć z analizy odpowiedzi osób zawodowo pracujących w branży audio (inżynier dźwięku, realizator, DJ, osoba odpowiedzialna za odprawę muzyczną wydarzeń, audycji, dziennikarz muzyczny). W tej grupie nastąpił podział – ujawniły się różne odczucia. Część z badanych deklarowała, że sztuka układania list odtwarzania jest bardzo trudna. Jeden stwierdził, że piosenki należy dobierać tak, aby przyciągnąć uwagę słuchacza od pierwszego dźwięku, np. przez charakterystyczny riff, perkusyjny beat czy nawet krzyk.

Kolejny wniosek dotyczył ułożenia utworów znanych i mniej znanych – żeby nie zestawiać utworów wyłącznie mniej znanych ze sobą, ponieważ słuchacze preferują utwory znane lub podobne.

Inny respondent z kolei – że utwory należy ułożyć w taki sposób, aby utrzymać uwagę słuchacza, co może odwoływać się do umiejętnego rozkładania znanych akcentów w kolejności ułożonych piosenek.

W perspektywie zarysowanego tematu można przyjąć za interesujący kierunek badań w przyszłości – próbę obiektywizacji układanych list. Należałoby zatem wykorzystać utwory i gatunki muzyczne mniej znane 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.

**Słowa kluczowe:** gatunek muzyczny, rozpoznawanie gatunków muzycznych, automatyczne tworzenie list muzycznych, autoenkoder.

## Bibliografia

[1] Bertin-Mahieux T., Ellis D.P.W., Whitman B., Lamere P., *The Million Song Dataset*, Proceedings of the 12th International Society for Music Information Retrieval Conference, Miami, Florida, USA, 24–28 listopada 2011; http://ismir2011.ismir.net/papers/OS6-1.pdf [dostęp: 26.06.2021].

[2] Błażejczyk W., *Postawy kompozycji muzycznej dla wykonawców – materiały pomocnicze*; https:// docplayer.pl/4378209-Podstawy-kompozycji-muzycznej-dla-wykonawcow.html [dostęp: 26.06.2021].

[3] Chen R., Hua Q., Chang Y., Wang B., Zhang L., Kong X., *A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks*, „IEEE Access" 2018, 6, DOI: 10.1109/ACCESS.2018.2877208, s. 64301–64320.

[4] 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*, 2018, „ BMC Genet" 2018, 65(19); https://doi.org/10.1186/s12863-018-0633-8 [dostęp: 26.06.2021].

[5] Diederik P.K., Welling M., *An Introduction to Variational Autoencoders*, „Foundations and Trends in Machine Learning" 2019, 12(4); http://dx.doi.org/10.1561/2200000056, s. 307–392.

[6] DJ top tips. URL; https://www.djtoptips.com/create-best-dj-set/ [dostęp: 5.12.2020].

[7] Géron A., *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media, Eddison, NJ, US, 2017.

[8] Ghias A., Logan J., Chamberlin D., Smith B.C., *Query by humming*, 3rd ACM Intern. Conference on Multimedia, San Francisco, CA, USA, 3–5 listopada 1995, s. 231–236.

[9] Ghosal D., Kolekar M.H., *Music genre recognition using deep neural networks and transfer learning*., Proceedings of the Annual Conference of the International Speech Communication Association, 19th INTERSPEECH, Hyderabad, Indie, 2–6 września 2018, s. 2087–2091.

[10] GTZAN – baza danych; http://marsyas.info/downloads/datasets.html [dostęp: 26.06.2021].

[11] Górecki T., *Podstawy statystyki z przykładami w R*, BTC, Legionowo 2011.

[12] Hoffmann P., Kostek B., *Koncepcja korekcji sygnału dźwiękowego z uwzględnieniem charakterystyk częstotliwościowych pomieszczenia oraz gatunku muzycznego*, „Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej" 2016, 51, s. 63–66.

[13] Huang C.-Z.A., Simon I., Dinculescu M., *Music Transformer: Generating Music with Long-Term Structure*; https://magenta.tensorflow.org/music-transformer [dostęp: 26.06.2021].

[14] ITU: BS.1284-1 General methods for the subjective assessment of sound quality. Intern. Telecom. Union Radiocom. Sector, 2003, s. 1–13.

[15] Jones E., Oliphant T., Peterson P., *SciPy: Open Source Scientific Tools for Python*; http://www.scipy.org/ [dostęp: 26.06.2021].

[16] Kim A., Park S., Park J., Ha J.-W., Kwon T., Nam J., *Automatic DJ Mix Generation Using Highlight Detection*, Proc. 18th International Society for Music Information Retrieval Conference, „ISMIR", Singapore, 23–28 października 2017.

[17] Kostek B., *Music Information Retrieval – Soft Computing Versus Statistics*, [w:] *Computer Information Systems and Industrial Management* K. Saeed, W. Homenda (red.), *Lecture Notes in Computer Science*, LNCS, 9339, s. 36–47, 2015, Springer, Cham; https://doi.org/10.1007/978-3-319-24369-6_3

[18] Korvel G., Treigys P., Tamulevicus G., Bernataviciene J., Kostek B., *Analysis of 2D Feature Spaces for Deep Learning-Based Speech Recognition*, „J. Audio Eng. Soc." 2018, 66(12), DOI: https://doi.org/10.17743/jaes.2018.0066, s. 1072–1081.

[19] Latin Music Database. URL; https://sites.google.com/site/carlossillajr/resources/the-latin-music-database-lmd [dostęp: 26.06.2021].

[20] Librosa; https://librosa.org/; McFee B., Raffel C., Liang D., Pw Ellis D., McVicar M., Battenberg E., Nieto O., *Librosa: Audio and music signal analysis in python*, Proceedings of the 14th Python in science conference, Austin, Texas, USA, 6–12 lipca 2015, s. 18–25.

[21] MPEG 7; http://mpeg.chiariglione.org/standards/mpeg-7 [dostęp: 26.06.2021].

[22] Pietrusińska K., *Opracowanie algorytmu do automatycznego układania list odtwarzania utworów muzycznych*, pr. mag., Katedra Systemów Multimedialnych, Wydział ETI, Politechnika Gdańska, 2020 [promotor: B. Kostek, konsultant: A. Kurowski].

[23] Prokhorov V., Shareghi E., Li Y., Pilehvar M.T., Collier N., *On the importance of the Kullback–Leibler divergence term in variational autoencoders for text generation*, Proceedings of the 3rd Workshop on Neural Generation and Translation Honk Kong, 2019, DOI: 10.18653/v1/D19-5612, s. 118–127.

[24] Simes R.J., *An improved Bonferroni procedure for multiple tests of significance*, „Biometrika" 1986, December, 73(3); https://doi.org/10.1093/biomet/73.3.751, s. 751–754.

[25] Setiono R., Baesens B., Mues C., *Recursive Neural Network Rule Extraction for Data With Mixed Attributes*, „IEEE Transactions on Neural Networks" 2008, 2, s. 299–307.

[26] Sporka A.J., Valta J., *Design and implementation of a non-linear symphonic soundtrack of a video game*, „New Review of Hypermedia and Multimedia" 2018, 23(4), DOI: 10.1080/13614568.2017.1416682, s. 229–246.

[27] Spotify. URL; https://www.spotify.com/pl/ [dostęp: 26.06.2021].

[28] Tzanetakis G., Cook P., *Musical genre classification of audio signals*, „IEEE Transactions on Speech and Audio Processing" 2002, 10(5), s. 293–302.

[29] van den Oord A., Dieleman S., Schrauwen B., *Deep content-based music recommendation*, [w:] *Advances in Neural Information Processing Systems 26*, C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weinberger (red.), Curran Associates, 2013, Red Hook, NY, USA, s. 2643–2651.

[30] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N. i in., *Attention is all you need*, „Adv Neural Inf Process Syst." 2017, December, s. 5999–6009.

[31] Zhang F., Gong T., Lee V.E., Zhao G., Rong C., Qu G., *Fast algorithms to evaluate collaborative filtering recommender systems*, „Knowledge-Based Systems" 2016, 96; https://doi.org/10.1016/j.knosys.2015.12.025, s. 96–103.

# 14. Pomiary wskaźnika odbicia dźwięku

PAWEŁ DZIECHCIŃSKI

Politechnika Wrocławska, Wydział Elektroniki, Fotoniki i Mikrosystemów, Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

W pracy omówiono metodę pomiarów wskaźnika odbicia dźwięku znaną jako QUIESST. Metoda ta jest m.in. przedmiotem normy PN-EN 1793-5. Poza wyjaśnieniem algorytmu pomiarowego przedstawiono również wymagania co do systemu pomiarowego zarówno w kontekście wymagań określonych w normie, jak i własnych doświadczeń autora. Pokazano, jaki wpływ na wyniki mogą mieć zastosowane urządzenie głośnikowe i rama, na której zamocowano mikrofony pomiarowe – jej konstrukcja wydaje się być istotnym elementem systemu pomiarowego, czego nie zaznaczono w normie PN-EN 1793-5. W pracy podano szereg wskazówek umożliwiających poprawne wykonanie tego typu ramy. Zaprezentowano i przeanalizowano wyniki pomiarów dla różnych wersji i ustawień systemu, a następnie porównano je z wynikami uzyskanymi innymi metodami.

## 14.1. Wprowadzenie

Ochrona środowiska przed hałasem komunikacyjnym sprawia, że instaluje się coraz więcej ekranów akustycznych. Do ich prawidłowego działania poza odpowiednim projektem geometrii wymagane jest zastosowanie elementów o właściwej izolacyjności akustycznej i w wielu przypadkach – pochłanianiu dźwięku od strony źródła hałasu. W warunkach laboratoryjnych właściwości w zakresie pochłaniania dźwięku ocenia się za pomocą współczynnika pochłaniania dźwięku wyznaczanego w warunkach pola po-

głosowego (PN-EN ISO 354 [18], PN-EN 1793-1 [14], PN-EN 16272-1 [16]) albo współczynnika pochłaniania dźwięku przy prostopadłym padaniu fali dźwiękowej (PN-EN ISO 10534-1 [20], PN-EN ISO 10534-2 [21]). Błędy w czasie budowy ekranów oraz wpływ czasu mogą sprawiać, że ich właściwości mogą znacząco różnić się od wartości wyznaczonych w warunkach laboratoryjnych. Między innymi dlatego stosowana jest metoda pomiaru wskaźnika odbicia dźwięku *RI* (*reflection index*) (PN-EN 1793-5 [15], CEN/TS 16272-5 [1]) umożliwiająca ocenę właściwości odbijających ekranów akustycznych w miejscu ich zainstalowania. Metodę pomiaru *RI* można też z pewnymi ograniczeniami wykorzystywać do oceny innych elementów budowlanych [4].

Najnowszą normą, której przedmiotem jest pomiar wskaźnika odbicia dźwięku, jest PN-EN 1793-5:2016-05+AC:2018-08 [15]. Metoda pomiarów *RI* jest efektem badań prowadzonych od początku lat 90. XX w. Pierwsza wersja zaproponowana przez Garai [5] oparta była na pomiarach wykonanych z użyciem jednego mikrofonu i przy prostopadłym do badanej płaszczyzny kierunku padania fali. Do pomiarów wykorzystywano sygnał MLS generowany analogowo. W celu separacji niepożądanych składników odpowiedzi impulsowej zaprojektowano okno czasowe, którego długość ograniczała zakres pomiarowy dla małych częstotliwości. Zakres metody dla dużych częstotliwości ogranicza częstotliwość $f_{max}$ odpowiadającą 1/4 długości fali równej głębokości nieregularności powierzchni badanej próbki $e$:

$$f_{\max} = \frac{c}{4e} \tag{1}$$

gdzie:

$c$ – prędkość dźwięku.

Mommertz [11] zaproponował modyfikację metody Garaiego polegającą na zastosowaniu techniki odejmowania odpowiedzi impulsowych, by rozdzielić impuls fali padającej i odbitej w sygnale oraz wykonanie pomiarów dla kilku kątów padania fali i zastosowanie w układzie pomiarowym cyfrowej korekcji zniekształceń liniowych toru pomiarowego.

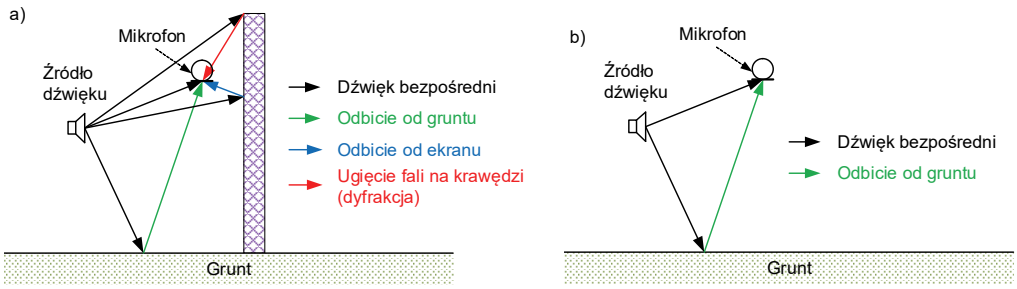
Kontynuację badań w tej tematyce zrealizowano w ramach projektu Adrienne – w jego wyniku powstała specyfikacja techniczna CEN/TS 1793-5:2003. Metoda będąca przedmiotem tego standardu znana jest jako metoda Adrienne, a polega na wykonywaniu pomiarów odpowiedzi impulsowych w dziewięciu punktach, z rozdzielczością 10°, za pomocą głośnika połączonego na sztywno z mikrofonem. Szereg prac mających na celu doskonalenie metody [6, 9] zaowocowało wprowadzeniem normy EN 1793-5:2016 (i jej polskiego odpowiednika [15]). Wprowadzono w niej zmodyfiko-

waną metodę pomiarową – QUIESST (Quietening the Environment for a Sustainable Surface Transport).

W pracy opisano system do pomiarów *RI* wykorzystujący powszechnie dostępne urządzenia. Zastosowanie takich urządzeń może świadczyć o powszechności, a tym samym atrakcyjności metody. Omówiono też aspekty pomiarowe wpływające w istotny sposób na uzyskiwane wyniki, a nie zawsze podkreślone w normie. Zwrócono uwagę na nieprecyzyjne oraz nieaktualne zapisy w normie PN-EN 1793-5:2016-05 dotyczące głównie metodyki wyznaczania odpowiedzi impulsowych stanowiących podstawę procedury pomiarowej. Badano płaskie materiały o powierzchni jednorodnej w zakresie właściwości pochłaniających dźwięk. Zarówno z literatury przedmiotu [22, 23], jak i doświadczeń autora wynika, że w innych przypadkach metoda ta może się nie sprawdzać, mimo że w normie PN-EN 1793-5 nie podano takich ograniczeń. Uzyskane wyniki porównano z wartościami uzyskanymi na podstawie pomiarów czasu pogłosu oraz z danymi z literatury.

## 14.2. Podstawy algorytmu pomiaru metodą różnicową

Pomiary wskaźnika odbicia omówione w normie PN-EN 1793-5:2016 oparte są na idei tzw. metody różnicowej, której podstawę stanowi analiza podstawowych zjawisk akustycznych zachodzących w obecności przeszkody po stronie źródła (rys. 1). Do punktu odbiorczego znajdującego się w obszarze między źródłem dźwięku a płaskim ekranem dociera dźwięk: bezpośredni oraz będący wynikiem odbić od ekranu i od gruntu, a także związany z dyfrakcją na krawędzi ekranu. W efekcie w punkcie obserwacji uzyskiwana jest odpowiedź impulsowa o schematycznym rozkładzie przedsta-

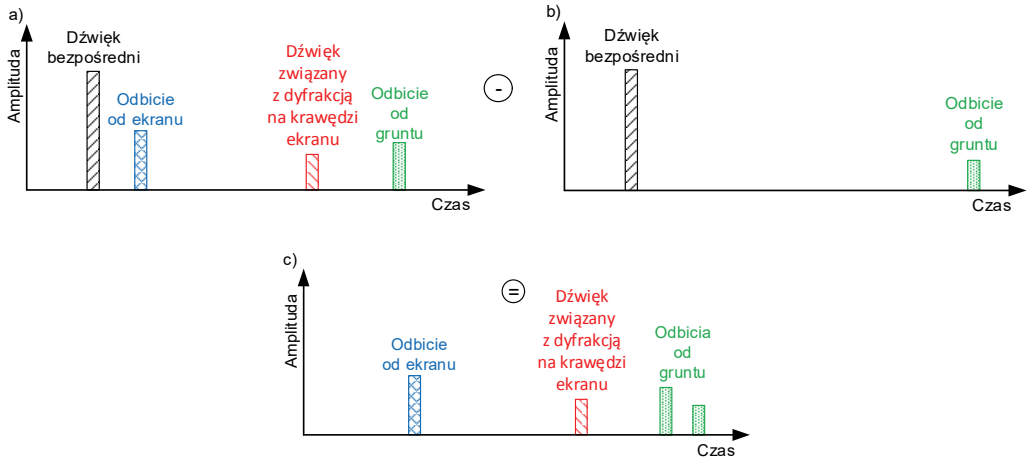



Rys. 1. Podstawowe zjawiska obrazujące propagację dźwięku:
a) w obecności przeszkody po stronie źródła, b) nad powierzchnią gruntu

wionym na rys. 2a. Ogólna idea pomiaru polega na tym, że w wyniku dwóch niezależnych pomiarów należy uzyskać dwie odpowiedzi impulsowe:

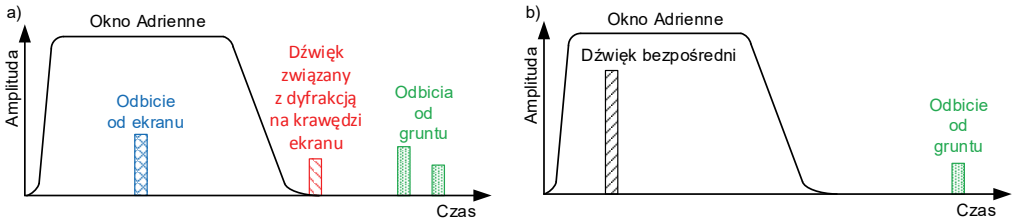
1) dźwięku bezpośredniego i odbitego od badanego ekranu,
2) dźwięku bezpośredniego.

Obie odpowiedzi impulsowe należy wyznaczyć dla tej samej odległości głośnika od mikrofonu i bez zmiany wzmocnienia systemu. W praktyce pomiar odpowiedzi impulsowej dźwięku bezpośredniego wykonuje się również nad powierzchnią gruntu (ewentualnie przy zwiększonej wysokości układu), ale w znacznej odległości od ekranu. W wyniku opisanych zjawisk uzyskuje się specyficzną odpowiedź impulsową o schematycznym rozkładzie przedstawionym na rys. 2b. Odpowiedź impulsową związaną z odbiciem od badanego ekranu uzyskuje się dzięki wyznaczeniu różnicy między pierwszą a drugą z wymienionych odpowiedzi (schemat przedstawiono na rys. 2c). Odpowiedź impulsowa zarówno z dźwiękiem bezpośrednim (rys. 2b), jak i odbitym od ekranu (rys. 2c) zawiera 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 przez specjalnie zaprojektowane okno czasowe zwane oknem Adrienne. Tę operację dla uzyskanej odpowiedzi różnicowej przedstawiono na rys. 3a. To samo




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)

okno czasowe wykorzystuje się w celu eliminacji odbicia od gruntu dla odpowiedzi impulsowej mającej na celu wyznaczenie dźwięku bezpośredniego (rys. 3b). Wskaźnik odbicia zostaje obliczony na podstawie stosunku energii dźwięku odbitego do dźwięku bezpośredniego w pasmach 1/3 oktawy z uwzględnieniem odpowiednich współczynników korekcyjnych.




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

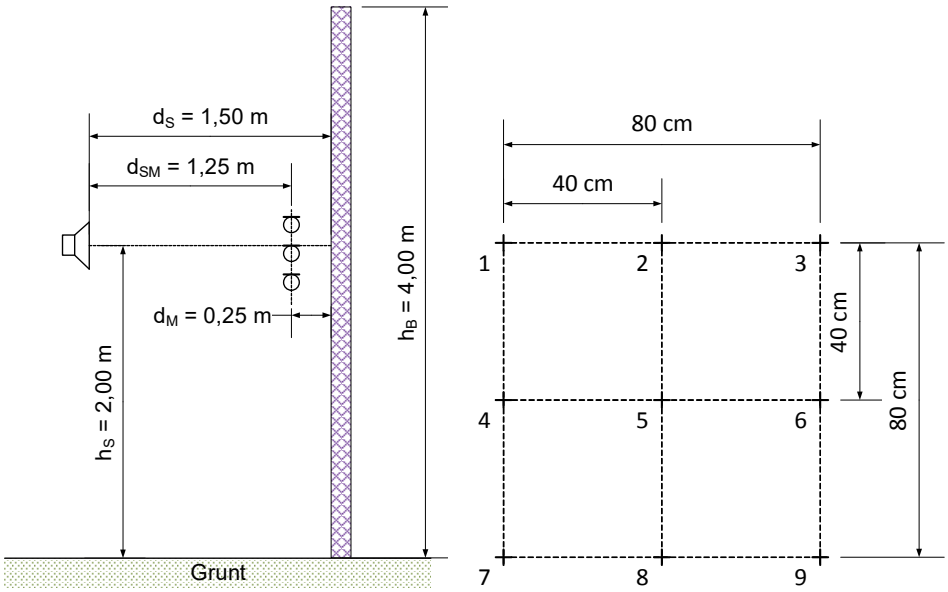
## 14.3. Metoda QUIESST

W metodzie QUIESST układ pomiarowy składa się z głośnika i 9-mikrofonowej matrycy umieszczonej równolegle do badanego ekranu. Geometrię układu pomiarowego określono w normie PN-EN 1793-5 (rys. 4). Mikrofony powinny znajdować się w odległości $d_M$ = 25 cm od badanego ekranu. Źródło dźwięku należy umieścić na osi mikrofonu nr 5 w odległości od mikrofonu $d_{SM}$ = 125 cm, czyli w odległości od ekranu $d_S$ = 150 cm. Głośnik i mikrofon pomiarowy nr 5 powinny znajdować się na wysokości $h_s$ równej połowie wysokości badanego elementu $h_B$. Uzyskanie wartości *RI* w zakresie częstotliwości 100–5000 Hz jest możliwe dla minimalnej wysokości oraz szerokości badanego elementu wynoszącej co najmniej 4 m. Zgodnie z normą dopuszczalna wysokość $h_s$ = 2 m – również w przypadku, w którym wysokość badanego elementu $h_B > 4$ m.

W przypadku ekranów o bardziej złożonej geometrii matryca powinna zostać umieszczona równolegle do powierzchni stycznej do czoła ekranu. Przykłady bardziej skomplikowanych struktur ekranów i związanych z nimi geometrii pomiarowych przedstawiono w normie PN-EN 1793-5.

Podstawą do analiz są 9-kanałowe odpowiedzi impulsowe wyznaczone w obecności badanego elementu oraz dla dźwięku bezpośredniego. Pomiar dźwięku bezpośredniego jest wykonywany dla analogicznej geometrii, ale w odległości głośnika i mikrofo-

nów od badanego obiektu na tyle dużej, aby wykorzystywane okna czasowe wyeliminowały wszelkie odbicia od tego elementu.




Rys. 4. Podstawowa geometria układu pomiarowego wraz z typowymi wymiarami i schemat rozmieszczenia mikrofonów (widok od strony głośnika)

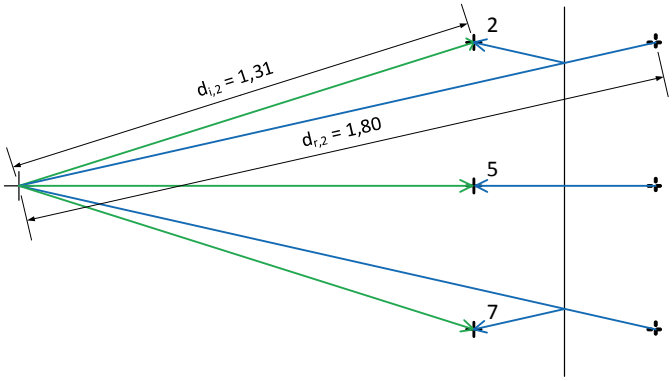
$$RI_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\left[\frac{\int\limits_{\Delta f_j}\left|F[h_{r,k}(t)\cdot w_{r,k}(t)]\right|^2 \mathrm{d}f}{\int\limits_{\Delta f_j}\left|F[h_{i,k}(t)\cdot w_{i,k}(t)]\right|^2 \mathrm{d}f}\cdot C_{geo,k}\cdot C_{dir,k}(\Delta f_j)\cdot C_{gain,k}(\Delta f_g)\right] \tag{2}$$

gdzie:

- $h_{i,k}(t)$ – składowa odpowiedzi impulsowej wyznaczonej dla pola swobodnego dotycząca dźwięku bezpośredniego w $k$-tym punkcie pomiarowym,
- $h_{r,k}(t)$ – składowa odpowiedzi impulsowej wyznaczonej w obecności badanego elementu dotycząca odbicia od badanego elementu w $k$-tym punkcie pomiarowym,
- $w_{i,k}(t)$ – okno czasowe (Adrienne) nakładane na odpowiedź impulsową zmierzoną w polu swobodnym w $k$-tym punkcie pomiarowym,
- $w_{r,k}(t)$ – okno czasowe (Adrienne) nakładane na składową odpowiedzi impulsowej wynikającą z odbicia od badanego elementu w $k$-tym punkcie pomiarowym,

$F$ – transformata Fouriera,
$j$ – indeks $j$-tego pasma 1/3 oktawy o częstotliwości środkowej z zakresu 100 Hz–5 kHz,
$\Delta f_j$ – szerokość $j$-tego pasma 1/3 oktawy,
$n$ – liczba punktów pomiarowych,
$k$ – numer punktu pomiarowego zgodnie z rys. 4,
$C_{geo,k}$ – współczynnik korekcji wynikający z różnicy dróg dla dźwięku bezpośredniego i odbitego, dla poszczególnych punktów pomiarowych (wzór (3)),
$C_{dir,k}(\Delta f_j)$ – współczynnik korekcji wynikający z charakterystyki kierunkowości źródła dla $k$-tego punktu pomiarowego (wzór (4)),
$C_{gain,k}(\Delta f_g)$ – współczynnik korekcji uwzględniający i kompensujący zmianę ustawień wzmocnienia źródła i ustawień czułości poszczególnych mikrofonów po zmianie konfiguracji między pomiarem w polu swobodnym a pomiarem w obecności badanego elementu (wzór (5)).

Współczynnik korekcji $C_{geo,k}$ wynika z różnicy dróg, jaką między głośnikiem a mikrofonem pokonuje dźwięk bezpośredni i dźwięk odbity (rys. 5).




Rys. 5. Droga dźwięku bezpośredniego i odbitego dla mikrofonu nr 2 określona na podstawie źródła pozornego

Wskaźnik odbicia dźwięku *RI* wyznacza się na podstawie 9-kanałowej odpowiedzi impulsowej ze wzoru (2). Droga dźwięku bezpośredniego $d_{i,k}$ i odbitego $d_{r,k}$ oraz współczynnik $C_{geo,k}$ dla typowej geometrii układu zostały dla poszczególnych pozycji mikrofonów podane w normie (tab. 1). Współczynnik ten wyznaczono przy założeniu, że głośnik jest źródłem punktowym i w związku z tym można go wyznaczyć ze wzoru (3):

$$c_{geo,k} = \left( \frac{d_{r,k}}{d_{i,k}} \right)^2 \tag{3}$$

gdzie:

$d_{r,k}$ – odległość od przedniego panelu głośnika do płaszczyzny pomiarowej i z powrotem do $k$-tego punktu pomiarowego wyznaczona zgodnie z zasadą – kąt padania równa się kątowi odbicia, a $d_{i,k}$ to odległość od przedniego panelu głośnika do $k$-tego punktu pomiarowego.

Tabela 1. Droga dźwięku bezpośredniego i odbitego oraz wartości współczynnika $C_{geo,k}$ dla poszczególnych pozycji mikrofonu

| $k$ | $d_{i,k}$ [m] | $d_{r,k}$ [m] | $C_{geo,k}$ |
|---|---|---|---|
| 1. | 1,37 | 1,84 | 1,80 |
| 2. | 1,31 | 1,80 | 1,87 |
| 3. | 1,37 | 1,84 | 1,80 |
| 4. | 1,31 | 1,80 | 1,87 |
| 5. | 1,25 | 1,75 | 1,96 |
| 6. | 1,31 | 1,80 | 1,87 |
| 7. | 1,37 | 1,84 | 1,80 |
| 8. | 1,31 | 1,80 | 1,87 |
| 9. | 1,37 | 1,84 | 1,80 |

Współczynnik korekcji wynikający z charakterystyki kierunkowości źródła $C_{dir,k}$ należy wyznaczyć dla wykorzystywanego do pomiarów głośnika ze wzoru (4):

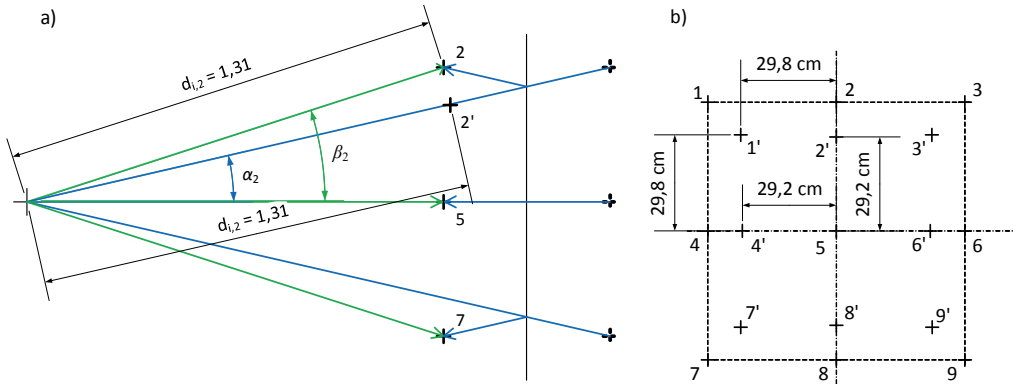
$$C_{dir,k}(\Delta f_j) = \frac{\int\limits_{\Delta f_j} \left| F[h_{i,k}(t, \alpha_k) * w_{i,k}(t)] \right|^2 \mathrm{d}f}{\int\limits_{\Delta f_j} \left| F[h_{i,k}(t, \beta_k) * w_{i,k}(t)] \right|^2 \mathrm{d}f} \tag{4}$$

gdzie:

$\alpha_k$ – kąt między linią łączącą środek głośnika z mikrofonem nr 5 a linią łączącą środek głośnika z $k$-tym mikrofonem,

$\beta_k$ – kąt między linią łączącą środek głośnika z mikrofonem 5 a linią łączącą środek głośnika z punktem odbicia od badanego ekranu.

Geometrię układu do wyznaczenia współczynnika $C_{dir,k}$ dla mikrofonu nr 2 przedstawiono na rys. 6a, a lokalizacje pozostałych punktów na rys. 6b.




Rys. 6. Geometria układu do wyznaczenia współczynnika korekcji wynikającego z charakterystyki kierunkowości źródła: a) dla $k = 2$, b) lokalizacje dla pozostałych punktów

W praktyce na niepewność pomiaru wskaźnika odbicia dźwięku znaczny wpływ może mieć dokładność ustalania geometrii pomiarowej oraz zmiana wzmocnienia układu pomiarowego. W celu oceny i ewentualnej korekcji *RI* w związku z sygnalizowanym problemem w normie PN-EN 1793-5:2016, w której początkowo zależność zapisano z błędem, a następnie skorygowano w PN-EN 1793-5:2016-05/AC:2018-08) – wprowadzono współczynnik korekcyjny $C_{gain}$:

$$C_{gain,k}(\Delta f_g) = \frac{\int\limits_{\Delta f_g} \left| F[h_{i,k,FF}(t) \cdot w_{i,k}(t)] \right|^2 \mathrm{d}f}{\int\limits_{\Delta f_g} \left| F[h_{i,k,D}(t) \cdot w_{i,k}(t)] \right|^2 \mathrm{d}f} \tag{5}$$

gdzie:

$\Delta f_g$ – pasmo częstotliwościowe o szerokości 1/3 oktawy z zakresu 500 Hz–2 kHz,

$h_{i,k,FF}(t)$ – składowa odpowiedzi impulsowej wyznaczonej dla pola swobodnego dotycząca dźwięku bezpośredniego w $k$-tym punkcie pomiarowym,

$h_{i,k,D}(t)$ – składowa odpowiedzi impulsowej wyznaczona w obecności badanego elementu dotycząca dźwięku bezpośredniego w $k$-tym punkcie pomiarowym.

Wydzielenie dźwięku bezpośredniego z odpowiedzi impulsowej wyznaczonej w obecności badanego elementu wymaga zastosowania okna Adrienne o długości 1,3 ms. W idealnym przypadku $C_{gain}$ powinno być równe 1,00, ale dopuszcza się odchyłki od

tej wartości do 5% i w tym zakresie współczynnika nie uwzględnia się we wzorze (2) ($C_{gain}$ = 1,00). Wyliczoną ze wzoru (5) wartość $C_{gain}$ uwzględnia się natomiast bezpośrednio we wzorze (2) wtedy, kiedy różni się ona od 1,00 o więcej niż 5%, a nie więcej niż 20%. Odchyłki $C_{gain}$ powyżej 20% oznaczają, że ustawienia systemu pomiarowego wpływające na poziomy sygnału uległy pokaźnej zmianie i należy sprawdzić ich poprawność i powtórzyć pomiar.

W normie PN-EN 1793-5 sporą uwagę poświęcono procedurze przesuwania w dziedzinie czasu odpowiedzi impulsowej wyznaczonej w polu swobodnym i jej odejmowania od odpowiedzi impulsowej wyznaczonej przed badanym ekranem. Realizacja tej procedury ma istotny wpływ na uzyskiwane wartość *RI*. Ocena jakości zrealizowanej procedury odejmowania jest możliwa za pomocą współczynnika redukcji $R_{sub}$ wyrażanego w decybelach, który można wyznaczyć ze wzoru (6):

$$R_{sub} = 10 \cdot \log \left[ \frac{\int_{t_{p,k}-0,5ms}^{t_{p,k+0,5ms}} \left| h_{i,k,FF}(t) \right|^2 \mathrm{d}t}{\int_{t_{p,k}-0,5ms}^{t_{p,k+0,5ms}} \left| h_{i,k,RES}(t) \right|^2 \mathrm{d}t} \right] \tag{6}$$

gdzie:

$h_{i,k,RES}(t)$ – odpowiedź impulsowa w polu swobodnym w *k*-tym punkcie pomiarowym,

$h_{i,k,RES}(t)$ – resztkowa składowa odpowiedzi impulsowej związana z odbiciem od badanego elementu w *k*-tym punkcie pomiarowym (uzyskana po odjęciu sygnału),

$t_{p,k}$ – moment czasowy, w którym znajduje się pierwszy szczyt odpowiedzi impulsowej w *k*-tym punkcie pomiarowym (przed odejmowaniem sygnałów).

Przyjmuje się, że wartości $R_{sub}$ < 10 dB wskazują na to, że procedura odejmowania nie została wykonana idealnie.

Do oceny jakości odbicia dźwięku od ekranu można wykorzystać wartość jednoliczbową $DL_{RI}$ (określoną w decybelach) wyznaczoną z zależności:

$$DL_{RI} = -10 \cdot \log \left[ \frac{\sum_{i=m}^{18} RI_i \cdot 10^{0,1L_i}}{\sum_{i=m}^{18} 10^{0,1L_i}} \right] \tag{7}$$

gdzie:

$m$ – numer najmniejszego wiarygodnego pasma 1/3 oktawy,

$L_i$ – skorygowany poziom znormalizowanego widma hałasu drogowego zgodny z normą PN-EN 1793-3 lub kolejowego – PN-EN 16272-3-2.

Kiedy stosunek sum we wzorze (7) jest większy od 1,00, wtedy należy go zastąpić wartością 0,99.

Do przeprowadzenia pomiarów w zakresie częstotliwości określonym w normie 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 sytuacji, w której 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 $r$ – można go wyznaczyć ze wzoru:

$$r = \frac{1}{d_S + d_M + cT_w}\sqrt{\left(d_S + d_M + \frac{cT_w}{2}\right)\left(d_S + \frac{cT_w}{2}\right)(2d_M + cT_w)cT_w} \tag{8}$$

gdzie:

$c$ – prędkość dźwięku,

$T_w$ – długość okna czasowego,

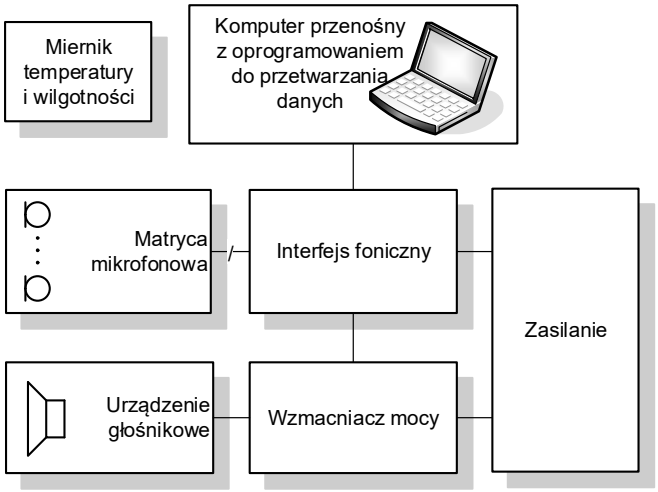
$d_S$ – odległość między głośnikiem i badaną próbką,

$d_m$ – odległość między mikrofonami i badaną próbką.

Okno Adrienne składa się z trzech części: narastającej, stałej i opadającej. Część narastająca i opadająca są odpowiednio lewą i prawą stroną okna Blackmana–Harrisa. Niezależnie od długości całego okna czasowego długość części narastającej to 0,5 ms, długość części płaskiej w stosunku do opadającej powinna być natomiast w proporcji 7/3.

## 14.4. 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, w których będzie wymagał zasilania akumulatorowego. Dziewięciokanałowe odpowiedzi impulsowe stanowiące podstawę pomiarów wyznaczano za pomocą programu EASERA, a przetwarzano programem DIRAC oraz własnymi programami.




Rys. 7. Układ do pomiaru *RI*

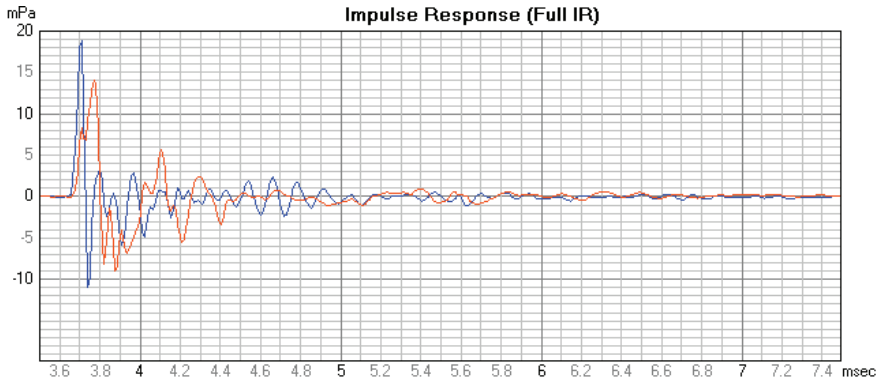
Zgodnie z normą PN-EN 1793-5:2016 cały system pomiarowy z wyjątkiem mikrofonów powinien spełniać wymagania dla mierników pierwszej klasy dokładności [17], a mikrofony – dla urządzeń klasy drugiej i powinny mieć średnicę nie większą niż 1/2″. Matryca mikrofonowa składa się z 9 mikrofonów równomiernie rozmieszczonych w rastrze 3 × 3 z rozstawem 40 cm. Do wykonania matrycy mikrofonowej zastosowano mikrofony o średnicy 1/4″ 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, a także mogą zasłaniać dźwięk bezpośredni lub odbity dochodzący do mikrofonu i tym samym – mieć wpływ na wyniki pomiarów. W celu oceny wpływu ramy na wartości *RI* sprawdzono jej kilka konstrukcji. Biorąc pod uwagę mobilność systemu, przydatnym aspektem funkcjonalnym ramy są jak najmniejsze jej wymiary. Testowane ramy i ich wpływ na wyniki pomiarów *RI* przedstawiono w dalszej części pracy.

Układ do pomiarów odpowiedzi impulsowych przedstawiony w normie PN-EN 1793-5:2016 i część związanych z nim wymagań wydają się być 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. Zgodnie z normą częstotliwości graniczne filtrów antyaliasingowych $f_{co}$

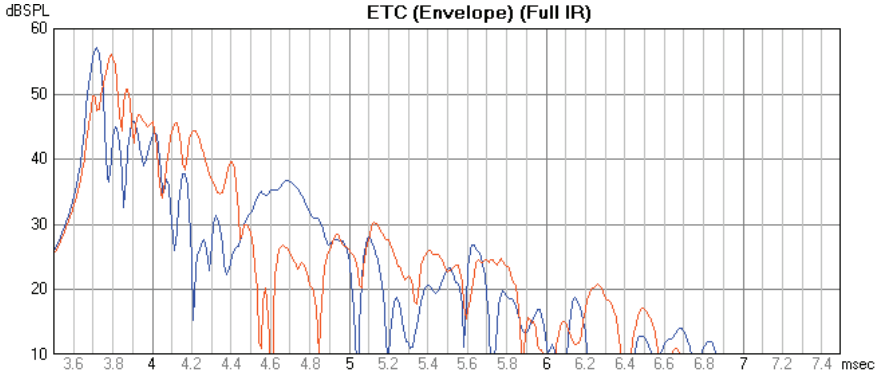
mają być relatywnie małe w stosunku do częstotliwości próbkowania $f_s$. Częstotliwości graniczne filtrów $f_{co}$ – mniejsze lub równe $1/3f_s$ dla filtrów Czebyszewa i $1/4f_s$ dla filtrów Butterwortha. Współczesne techniki [12] i systemy do pomiarów odpowiedzi impulsowych wykorzystują sygnały pomiarowe generowane cyfrowo. I przetworniki A/C, i C/A współczesnych systemów pomiarowych wykorzystywanych w technice ΔΣ pracują zatem 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 „Phantom", w związku z tym dokonano jego odpowiedniej modyfikacji. Zmierzono parametry elektryczne wykorzystywanego interfejsu potwierdzające, że spełnia on wymagania dla mierników klasy pierwszej według normy PN-EN 61672-1:2014 [17].

Urządzenie głośnikowe zastosowane do badań musi być urządzeniem jednogłośnikowym w obudowie zamkniętej, o charakterystyce częstotliwościowej bez ostrych nieregularności w zakresie pasm 1/3 oktawy z zakresu 100 Hz–5 kHz, a jego odpowiedź impulsowa powinna być nie dłuższa niż 3 ms. Do badań początkowo użyto zmodyfikowany tzw. projektor głośnikowy z systemów rozgłoszeniowych (obudowa w kształcie walca o średnicy 14 cm); w oprac. ozn.: 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 oprac. ozn.: głośnik #2. W stosunku do projektora urządzenie to zapewnia większe poziomy ciśnienia akustycznego oraz ma bardziej wyrównaną charakterystykę. Duże poziomy ciśnienia akustycznego są wymagane ze względu na konieczność zapewnienia odpowiednio dużego stosunku sygnału do szumu w warunkach terenowych. Ich uzyskanie jest trudne za pomocą urządzenia jednodrożnego w obudowie zamkniętej mającego wyrównaną charakterystykę w zakresie częstotliwości 89 Hz–5,6 kHz. W normach nie jest zdefiniowane kryterium, zgodnie z którym należy określić koniec odpowiedzi impulsowej głośnika. Z analizy odpowiedzi impulsowej przedstawionej na rys. 8 wynika, że po 3 ms jej amplitudy są dla obu głośników relatywnie małe, skala liniowa nie umożliwia jednak precyzyjnej obserwacji w dużym zakresie dynamiki. Na podstawie obwiedni ETC (rys. 9) można

stwierdzić, że zanik energii po 3 ms dla obu głośników jest większy niż 40 dB, co intuicyjnie wydaje się wartością wystarczającą.




Rys. 8. Odpowiedzi impulsowe głośników: głośnik #1 (ozn. kolorem czerwonym), głośnik #2 (ozn. kolorem niebieskim)

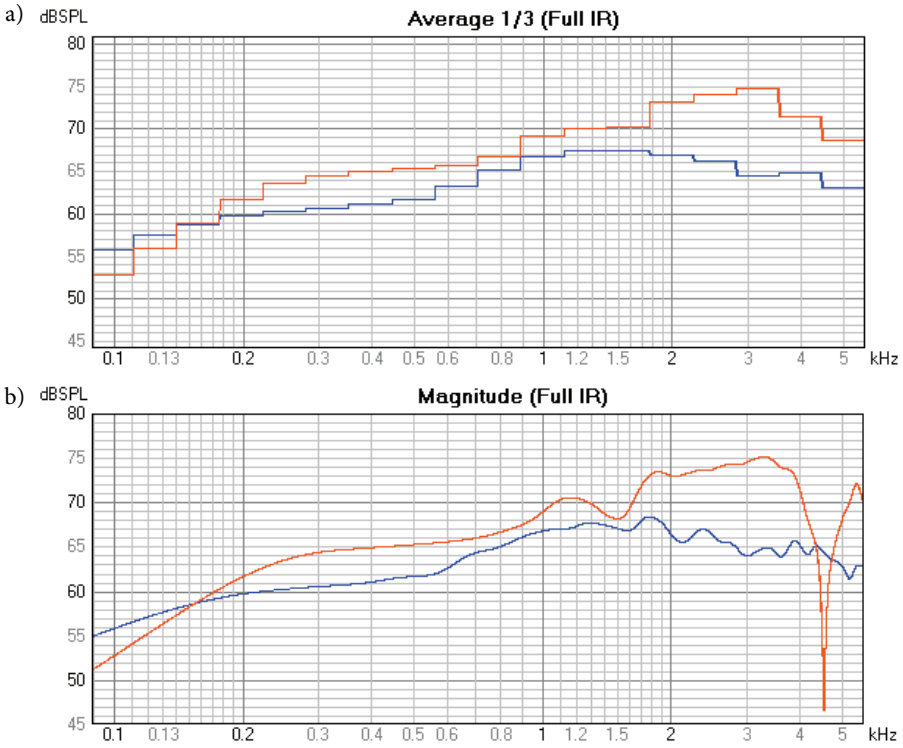



Rys. 9. Obwiednia ETC głośników: głośnik #1 (ozn. kolorem czerwonym), głośnik #2 (ozn. kolorem niebieskim)

Wprawdzie charakterystyka częstotliwościowa wykorzystywanego projektora głośnikowego w pasmach 1/3 oktawy wydawała się relatywnie równomierna (rys. 10b), ale znaczna nierównomierność charakterystyki częstotliwościowej między 4 i 5 kHz (rys. 10a) wpływa na wyniki pomiarów, co zostanie omówione w dalszej części pracy.

W normie PN-EN 1793-5:2016 nie określono żadnych dodatkowych wymagań odnośnie do wzmacniacza mocy. To oczywiste, że powinien on mieć wyrównaną charakterystykę częstotliwościową w zakresie nie mniejszym niż wymagania zawarte

w przedmiotowych normach (89 Hz–5,7 kHz). Wykorzystano wzmacniacz, który można zasilać z akumulatorów, co stanowi ułatwienie pomiarów w warunkach terenowych. Moc znamionowa jest wystarczająca do odpowiedniego wysterowania urządzeń głośnikowych. Parametry wykorzystanego wzmacniacza zweryfikowano pomiarowo: charakterystyka częstotliwościowa 20 Hz–20 kHz ±0,5 dB, THD < 0,02%, S/N (CCIR) > 80 dB, odchyłki od liniowości spełniające wymagania dla mierników klasy pierwszej.




Rys. 10. Charakterystyki częstotliwościowe głośnika #1 (ozn. kolorem czerwonym) i głośnika #2 (ozn. kolorem niebieskim): a) bezpośrednio po nałożeniu okna Adrienne, b) uśrednione w pasmach 1/3 oktawy

## 14.5. Wpływ systemu pomiarowego na wyniki badań *RI*

Walidacji wykorzystanego systemu dokonano przez porównanie uzyskanych wyników *RI* z wynikami otrzymanymi przy zastosowaniu innych metod. Ponadto oceniono powtarzalność i odtwarzalność uzyskiwanych wyników [19].

### 14.5.1. Badane elementy budowlane

Do badań wskaźnika odbicia dźwięku wytypowano dwa obiekty o bardzo zróżnicowanych właściwościach odbijających. Obiektem mocno odbijającym była betonowa podłoga komory akustycznej o wymiarach 13,2 × 9,3 × 6,5 m, a pochłaniającym – 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 jako 1 – $\alpha$, gdzie $\alpha$ oznacza pogłosowy współczynnik pochłaniania dźwięku. Wartości odbicia dźwięku dla ścian uzyskano na postawie pomiarów czasu pogłosu komory. Parametr ten w dalszej części pracy będzie oznaczany: 1 – $\alpha$. Przyjęto, że ściany i sufit badanego pomieszczenia wykonane są z takiego samego materiału. 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 pod belkami konstrukcyjnymi),
5) strop betonowy.

Nie oczekiwano dokładnej zgodności uzyskanych wyników *RI* i 1 – $\alpha$. Pogłosowy współczynnik pochłaniania dźwięku mierzony jest w warunkach pola rozproszonego, czyli dla wszystkich możliwych kątów padania fali akustycznej [18]. Wskaźnik odbicia *RI* wyznacza się w warunkach pola zbliżonego do swobodnego dla kątów padania fali akustycznej na próbkę z zakresu 0–20°. Warunki, w których wyznacza się *RI* to zatem warunki pośrednie między pogłosowymi a warunkami, dla których wyznacza się fizyczny współczynnik pochłaniania dźwięku [20, 21]. Relacje między fizycznym i pogłosowym współczynnikiem pochłaniania dźwięku podał między innymi London [10, 21]. Należy jednak zauważyć, że specyfika wykorzystywanej komory sprawia, że panujące w niej warunki (odbijająca podłoga, pochłaniające ściany i sufit) znacznie odbiegają od pola rozproszonego. Tym samym uzyskane w wyniku pomiarów czasu pogłosu współczynniki pochłaniania dźwięku ścian i sufitu będą się różnić od wartości pogłosowego współ-

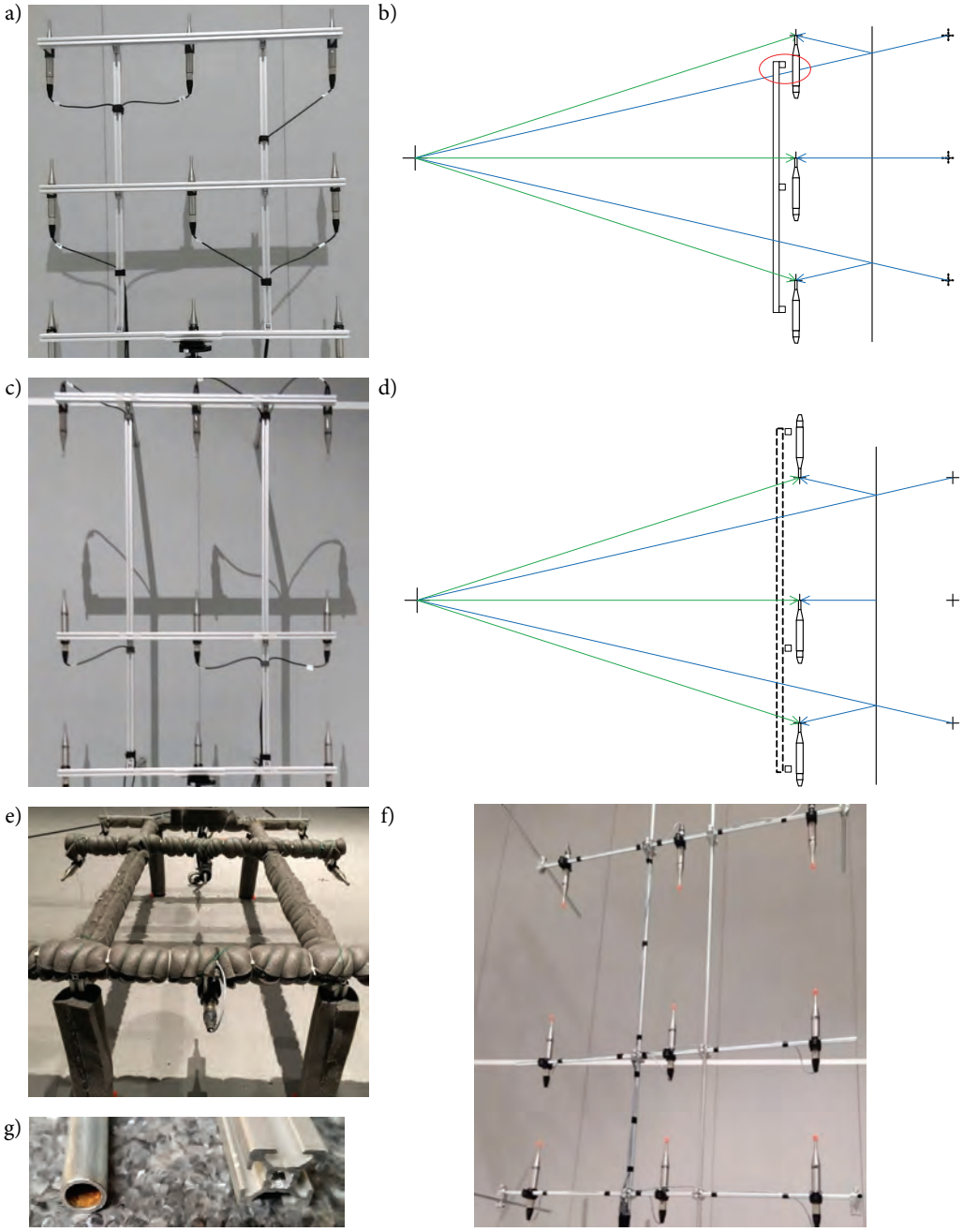
czynnika pochłaniania uzyskanych w wyniku pomiarów laboratoryjnych. Wprawdzie nie są znane wyniki pomiarów laboratoryjnych współczynnika pochłaniania całej struktury, z której zbudowana jest ściana, ale wartości dwóch pierwszych warstw podawane przez producentów są w zakresie pasm 1/1 oktawy 500 Hz–4 kHz zbliżone do 1,0, podczas gdy w wyniku pomiarów czasu pogłosu pomieszczenia uzyskano mniejsze wartości (rys. 15). Pogłosowe współczynniki pochłaniania betonowej podłogi zaczerpnięto z literatury [13]. Ich wartości rozciągają się w zakresie od 0,02 dla pasm 1/1 oktawy 125 i 250 Hz do 0,05 dla pasm 1/1 oktawy 2 i 4 kHz.

## 14.5.2. Wpływ elementów systemu na wyniki *RI*

Elementami systemu mogącymi mieć istotny wpływ na wyniki pomiarów są: rama, na której zamocowane są mikrofony pomiarowe, oraz głośnik. Kolejne rozwiązania konstrukcyjne ramy (por. rys. 11) testowano z użyciem głośnika #1 i betonowej podłogi komory. W przypadku elementów mocno odbijających spodziewano się szczególnie dużego wpływu ramy na wyniki pomiaru *RI*. Ponadto dla betonu oczekiwane wyniki *RI* powinny były znaleźć się w relatywnie wąskim zakresie między 1,00 (materiał idealnie odbijający) a wartościami 1 – $\alpha$ dla betonu, podanymi w literaturze przedmiotu.

Najmniejszą z badanych ram była rama #1 (rys. 11a) o wymiarach 80 × 82 cm wykonana z elementów o przekroju zbliżonym do kwadratu o boku 20 mm (rys. 11g). Wadą takiej konstrukcji są relatywnie małe odległości między mikrofonami a poziomymi elementami konstrukcji ramy. Ponadto w przypadku mikrofonów nr 1–3 element oraz same mikrofony znajdują się na drodze lub blisko promienia dźwiękowego padającego na badaną próbkę i skierowanego po odbiciu na dany mikrofon (rys. 11b), co zostało potwierdzone pomiarowo przez porównanie wyników *RI* jednorodnego materiału dla mikrofonów nr 1–3 i nr 7–9. Teoretycznie wyniki te powinny być symetryczne, okazało się jednak, że np. dla pasm 1/3 oktawy 1,6 kHz i 3,15 kHz różnice wyników przekraczały 30%. Wartości *RI* betonu uzyskiwane z wykorzystaniem rozpatrywanej ramy różniły się od oczekiwanych szczególnie w zakresie powyżej 2 kHz (rys. 12).

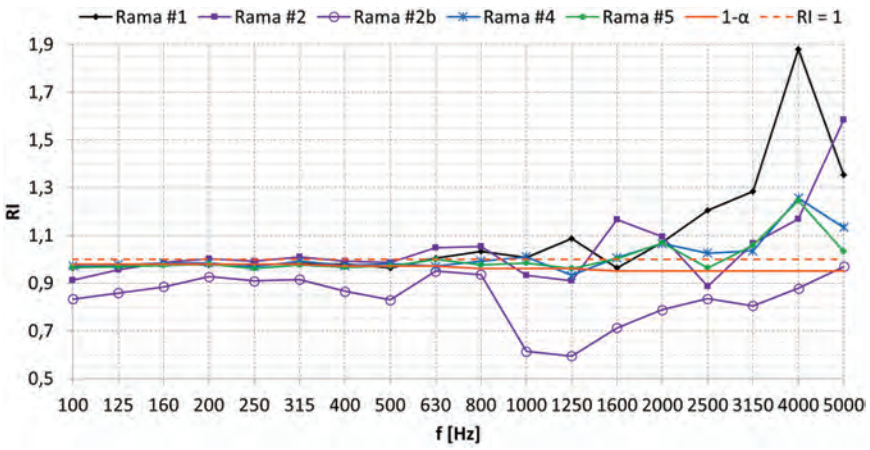
W ramie#2 (rys. 11c) zwiększono wymiar pionowy do 112 cm – to zapewniło większe odległości poziomych elementów ramy od mikrofonów oraz symetryczny układ mikrofonów (rys. 11d). Wprawdzie wprowadzona modyfikacja znacząco poprawiła wyniki w zakresie 2,5 kHz–4 kHz, ale nadal mocno różniły się one od oczekiwanych (rys. 12). Niewielka modyfikacja polegająca na zwiększeniu wymiaru poziomego ramy do 200 cm i przesunięcie jej pionowych elementów na końce (rama #3) poprawiła wyniki, ale nadal nie były one zadowalające.




Rys. 11. Wybrane ramy wykorzystywane do badań: a) rama #1,
b) promienie dźwiękowe dla ramy #1, c) rama #2, d) promienie dźwiękowe dla ramy #2,
e) rama #2 pokryta materiałem pochłaniającym dźwięk, f) rama #5, g) elementy konstrukcyjne ram

Pokrycie ramy #2 materiałem pochłaniającym dźwięk (rys. 11e) wprawdzie zmniejszyło niekorzystne efekty związane z odbiciami od ramy, ale z kolei miało wpływ na energię dźwięku bezpośredniego docierającego do badanej próbki i dla niektórych zakresów częstotliwości uzyskiwane wartości *RI* okazały się zaniżone (rys. 12 – rama #2b).

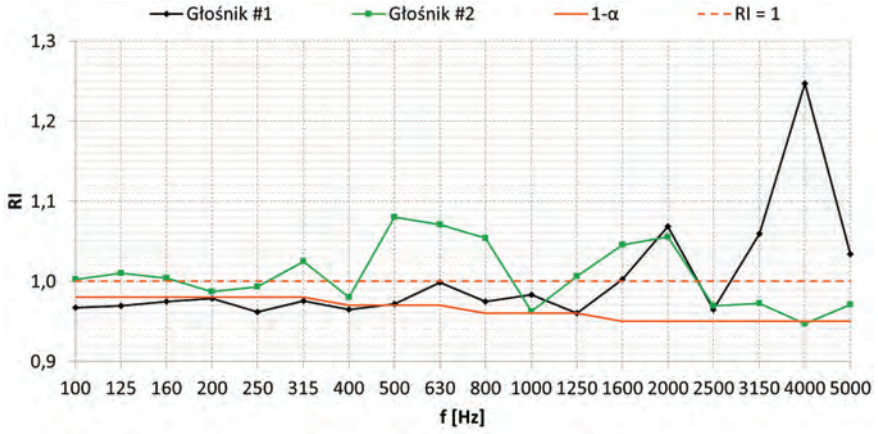
Kolejne wersje ram wykonano z rur okrągłych o średnicy 12 mm (rys. 11g). Takie rury mają gorsze właściwości mechaniczne i funkcjonalne niż te, których profil jest w przekroju zbliżony do kwadratu. Po ich zastosowaniu spodziewano się mniejszego wpływu konstrukcji ramy na wyniki *RI* w związku z ich właściwościami dyfrakcyjnymi. Mają one również relatywnie małą powierzchnię, co korzystnie wpływa na dźwięk bezpośredni docierający do badanego elementu. Rama #4 wykonana z takich rur miała geometrię zbliżoną do ramy #3. Ze względu na jej sztywność, nie można było jednak zainstalować ramy na statywie. W związku z tym skonstruowano kolejną ramę #5 (rys. 11f) o geometrii przypominającej ramę #2. W konstrukcji #5 dodatkowo zwrócono uwagę na niekorzystne odbicia, których źródłem są elementy mocujące ramę do statywu i jego głowica. Elementy te odsunięto od ramy i ich większe powierzchnie pokryto materiałem pochłaniającym dźwięk.




Rys. 12. Wyniki badań *RI* za pomocą różnych ram i głośnika #1 podłogi betonowej

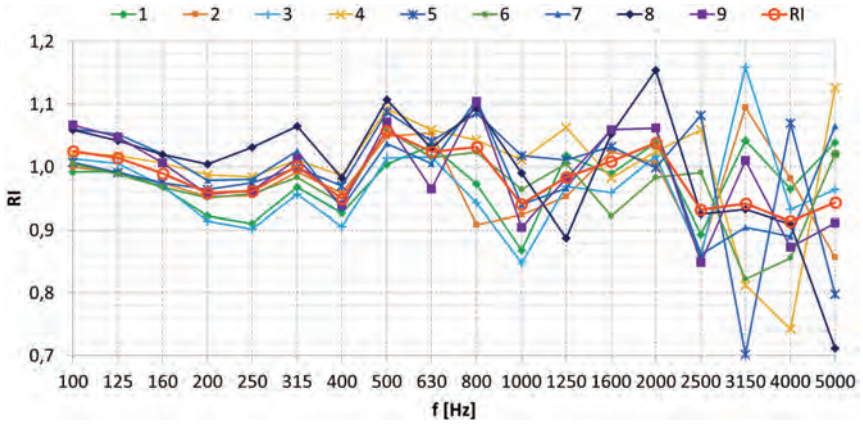
Wyniki pomiarów *RI* z użyciem ram – #4 i #5 były bardzo podobne (rys. 12), ale nieco bliższe oczekiwanym dotyczyły ramy #5. Największe odstępstwo od pożądanych wartości *RI* odnotowano dla częstotliwości 4 kHz. Zgodnie z przewidywaniami były one wynikiem znaczącej nierównomierności charakterystyki częstotliwościowej głośnika #1 (rys. 10b). Potwierdziły to pomiary *RI* z wykorzystaniem ramy #5 i głośnika #2 (rys. 13). Zastosowanie tego głośnika zapewniło bardzo dobrą zgodność z oczekiwa-

niami w zakresie 2,5 kHz–5 kHz, ale w porównaniu do głośnika #1 zgodność z oczekiwaniami odnośnie do zakresu 500–800 Hz się pogorszyła. W przypadku obu głośników występuje odstępstwo dla pasma 2 kHz, co nadal może sugerować wpływ konstrukcji ramy. Sumaryczne odchylenie *RI* zarówno od wartości 1 – $\alpha$ betonu, jak i 1,00 jest mniejsze dla głośnika #2.




Rys. 13. Wyniki badań *RI* za pomocą różnych głośników, ramy #5 i podłogi betonowej

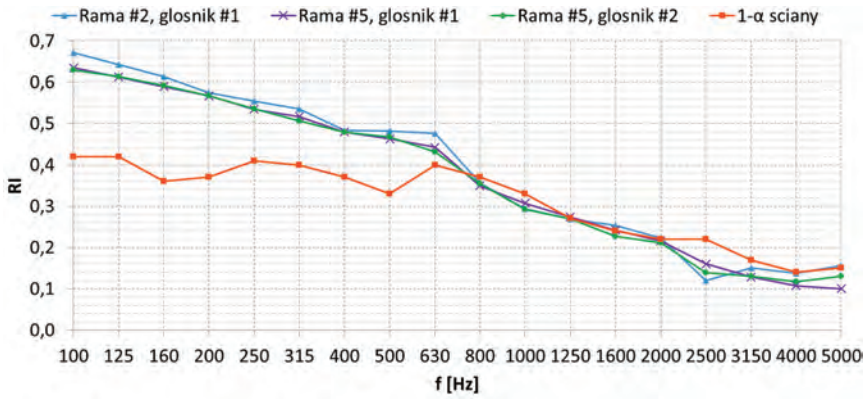
Analiza wartości *RI* wyznaczonych dla poszczególnych pozycji mikrofonów wskazuje na niedoskonałości ramy #5 – otrzymane wyniki w przypadku jednorodnej powierzchniowo struktury powinny być bardzo podobne dla tego samego kąta odbicia,




Rys. 14. Wyniki badań *RI* podłogi betonowej uzyskiwane w poszczególnych pozycjach mikrofonów (wykresy 1–9) i ich wartość średnia (wykres *RI*) za pomocą głośnika #2 i ramy #5

czyli dla mikrofonów nr 1, 3, 7 i 9 oraz nr 2, 4, 6 i 8. Różnice między wynikami dla tych mikrofonów mogą być wynikiem nieprecyzyjnej geometrii układu pomiarowego lub efektem wpływu elementów ramy na dźwięk. Największe skupienie elementów konstrukcyjnych ramy znajduje się w pobliżu mikrofonu nr 8 i wpływ tych elementów ma najprawdopodobniej odzwierciedlenie w wynikach *RI* uzyskiwanych dla tej pozycji mikrofonu, które dla niektórych pasm 1/3 oktawy znacząco odbiegają od wartości uzyskiwanych w przypadku pozostałych pozycji.

Wpływ ramy i głośnika na uzyskiwane wartości *RI* był znacznie mniejszy dla ściany pochłaniającej (rys. 15) – uzyskano tu jednak relatywnie duże różnice między wartościami *RI* a 1-α dla częstotliwości mniejszych od 630 Hz.




Rys. 15. Wyniki badań *RI* przy użyciu różnych ram i głośników dla ściany o dużym współczynniku pochłania dźwięku

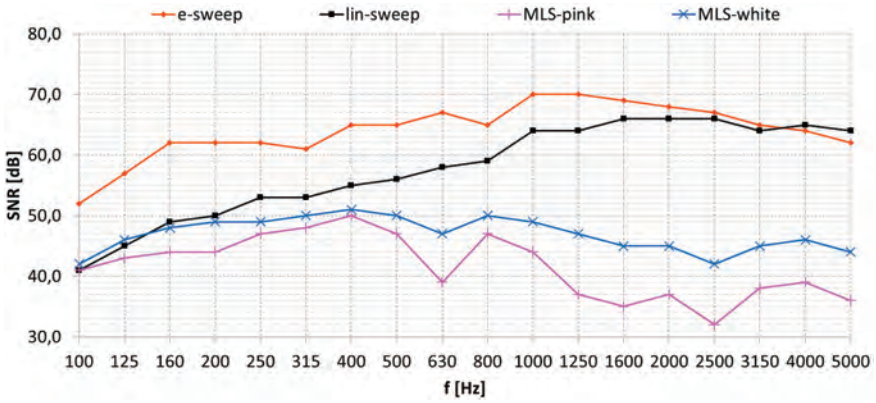
## 14.5.3. Wpływ parametrów pomiaru i analiz na wyniki *RI*

W przypadku ramy #5 i głośnika #2 sprawdzono powtarzalność pomiarów przy teoretycznie najlepszych parametrach systemu, czyli dla sygnału pomiarowego e-sweep o częstotliwości próbkowania 96 kHz, długości 5,5 s i okna czasowego 7,9 ms. Wykonano po dziesięć pomiarów dla obu badanych elementów budowlanych. Czas między pomiarami wynikał z czasu niezbędnego do generacji cyklu rozruchowego sygnału oraz wyznaczenia i zapisania odpowiedzi impulsowych. Dla betonu maksymalne różnice między wynikami dziesięciu pomiarów *RI* dla danego pasma 1/3 oktawy w większości pasm nie przekraczały 0,01. Różnice przekraczające 0,015 uzyskano tylko dla pasm 100 Hz i 3150 Hz. Dla ściany maksymalne różnice między wynikami dziesięciu pomiarów *RI* nie przekraczały wartości 0,00.

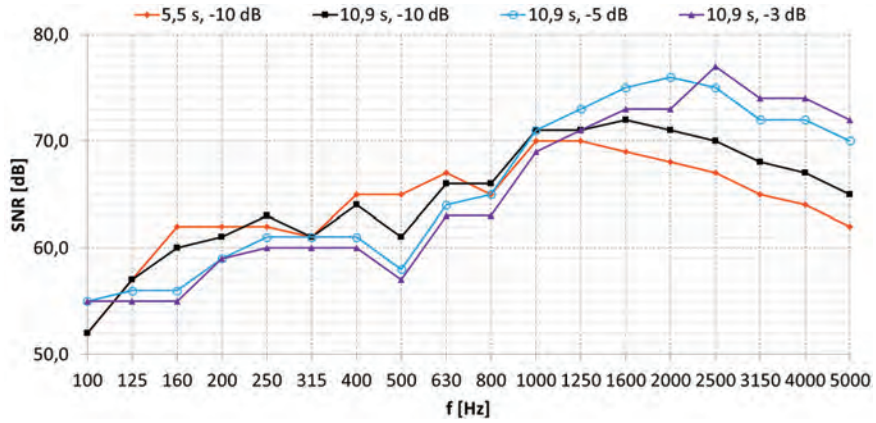
Jednym z istotnych elementów środowiska pomiarowego, który może mieć wpływ na wyniki pomiarów *RI*, jest panujący w nim poziom zakłóceń akustycznych. Uzyskanie odpowiednio dużego stosunku sygnału do szumów *SNR* może zostać zrealizowane przez system na kilka sposobów. Oczywiste jest, że należy zapewnić jak największy poziom sygnału pomiarowego. W praktyce jest to jednak utrudnione ze względu na wymagania co do wykorzystywanego urządzenia głośnikowego. Uzyskanie równomiernej charakterystyki częstotliwościowej 89 Hz–5,6 kHz za pomocą jednodrożnego urządzenia głośnikowego w obudowie zamkniętej wymaga w praktyce zastosowania przetwornika, którego efektywność nie przekracza 90–95 dB, a moc znamionowa 50–100 W. Ponadto przetwornik ten nie powinien pracować z mocami bliskimi mocy znamionowej, gdyż w takim przypadku należy się liczyć z dużymi zniekształceniami nieliniowymi, a system do wyznaczania odpowiedzi impulsowej powinien być systemem liniowym. Wpływ na stosunek sygnału do szumu mają też: rodzaj zastosowanego sygnału pomiarowego, czas trwania sygnału, liczba uśrednień.

W badaniach zastosowano cztery sygnały pomiarowe: e-sweep, sweep liniowy, MLS o widmie szumu białego i MLS o widmie szumu różowego. W pomieszczeniu, w którym wykonywano badania, zakłócenia akustyczne miały relatywnie mały poziom ($L_{Aeq}$ < 10 dB). W związku z tym wpływ rodzaju sygnału pomiarowego na wyniki *RI* był mały. Dla betonu maksymalne różnice między *RI* wyznaczonym sygnałem e-sweep a pozostałymi sygnałami dla poszczególnych pasm 1/3 oktawy wynosiły najczęściej 0,00 i 0,01, dla ściany pochłaniającej natomiast – 0,00. Znacznie większego wpływu sygnału pomiarowego na wyniki pomiarów *RI* należy spodziewać się w środowisku o większych poziomach zakłóceń – np. przy badaniach ekranów drogowych w miejscu ich zainstalowania. Rodzaj sygnału pomiarowego ma wpływ na *SNR* ze względu na swoje właściwości widmowe oraz współczynnik szczytu [3]. Właściwości te są znane z literatury przedmiotu, dokonano jednak ich sprawdzenia w wykorzystywanym systemie – m.in. dlatego, że nie ma w normie stosownych informacji. Wartości *SNR* uzyskane dla systemu pracującego z poziomem –10 dB (co odpowiada wysterowaniu głośnika sygnałem e-sweep do 1/10 mocy znamionowej głośnika) przedstawiono na rys. 15. Głównym źródłem zakłóceń w pomieszczeniu była wentylacja oraz wentylator komputera. Ponieważ mogły być one zmienne, uzyskane wartości *SNR* są orientacyjne. Zgodnie z przewidywaniami największy odstęp od zakłóceń uzyskano dla sygnału e-sweep. Dalsze zwiększanie *SNR* jest możliwe przez zwiększenie czasu trwania sygnału pomiarowego oraz zwiększenie jego poziomu. Zwiększanie poziomu może jednak równocześnie powodować wzrost zniekształceń nieliniowych, a tym samym zmniejszać wartości *SNR*. W celu zmniejszenia nieliniowości systemu przy pracy z więk-

szymi poziomami sygnał e-sweep dla poziomów –5 dB i –3 dB generowano w zakresie od 50 Hz. Wydłużenie czasu trwania sygnału ze względu na ograniczenia wykorzystywanego systemu wymagało z kolei zmniejszenia częstotliwości próbkowania z 96 kHz do 48 kHz. Wpływ czasu trwania sygnału i jego poziomu na wartości *SNR* przedstawiono na rys. 17. Czasu trwania sygnału i jego poziom zwiększały wartości *SNR* głównie dla częstotliwości większych od 1 kHz. Dla tych częstotliwości wzrost *SNR* był w większości pasm zgodny z oczekiwaniami z wyjątkiem poziomów –5 dB i –3 dB, w których nieliniowości systemu sprawiły, że spodziewany wzrost *SNR* nastąpił tylko dla zakresu częstotliwości odpowiednio 2,0 kHz–5 kHz i 2,5–5 kHz.




Rys. 16. Wpływ rodzaju sygnału pomiarowego na stosunek sygnału do zakłóceń akustycznych *SNR*




Rys. 17. Wpływ poziomu i czasu trwania sygnału e-sweep na stosunek sygnału do zakłóceń akustycznych *SNR*

W warunkach terenowych trudno jest zapewnić precyzyjne ustawienie geometrii pomiarowej. Zgodnie z normą dopuszcza się pewne odchyłki od standardowej geometrii pomiarowej, nie wiadomo jednak, jak te różnice wpływają na wyniki pomiarów. Sprawdzono, że przy odległości $d_S$ odbiegającej od znormalizowanej o ±2,5 cm wyniki *RI* w funkcji częstotliwości zmieniały się maksymalnie o 5%, a średnio do 3%, dla odchyłek ±1 cm maksymalnie o 3%, a średnio do 2%. Odchyłki o ±2,5 cm dla odległości $d_M$ powodowały zmiany wartości *RI* w funkcji częstotliwości maksymalnie do 8%, a średnio do 4%.

## 14.6. Podsumowanie

W pracy omówiono metodę pomiarów wskaźnika odbicia dźwięku *RI*. Badano płaskie elementy o jednorodnych powierzchniowo właściwościach pochłaniających. Okazuje się, że szczególne wymagania muszą zostać spełnione w przypadku głośnika i ramy, zwłaszcza jeśli badane są elementy o dużych wartościach *RI*. W normie wprawdzie zaznacza się, że charakterystyka częstotliwościowa głośnika ma być wyrównana, ale nie określono tego wymagania precyzyjnie. Jak pokazano w pracy, nawet zastosowanie głośnika o nierównomiernościach ±4 dB w zakresie 0,5–2 kHz może wprowadzać znaczne błędy pomiarowe. Oznacza to, że należałoby dokładniej określić wpływ właściwości głośnika na uzyskiwane wyniki i doprecyzować w tym zakresie wymagania. Wydaje się, że istotne znaczenie może mieć kształt obudowy głośnika i związanie z tym zniekształcenia dyfrakcyjne. Wprawdzie w normie wyklucza się stosowanie aktywnych bądź pasywnych elementów mogących wpływać na pasmo przenoszenia systemu, ale odpowiednim rozwiązaniem może być cyfrowa korekcja zniekształceń liniowych wykorzystywanego głośnika [2].

W normie nie zaznaczono, że istotnym elementem systemu jest rama, na której mocowane są mikrofony pomiarowe – w pracy wykazano, że jej konstrukcja może mieć istotne znaczenie. W publikacjach dotyczących pomiarów *RI* stosowane są bardzo różnorodne konstrukcje ram. Może mieć to wpływ na wyniki *RI* uzyskiwane w różnych laboratoriach. Tym samym zwiększa to niepewność pomiarów podaną w normie PN-EN 1793-5, określoną na podstawie porównań międzylaboratoryjnych. W pracy podano szereg wskazówek umożliwiających wykonanie ramy mającej stosunkowo mały wpływ na uzyskiwane wyniki.

Można przyjąć, że do testowania właściwości systemu pomiarowego bardzo dobrym elementem jest duża powierzchnia o współczynniku pochłaniania dźwięku jak

najbliższym wartości 0,00. Dla takiego obiektu (np. betonowej podłogi) duże odstępstwa uzyskiwanych wartości *RI* od 1,00 wskazują na niedoskonałości wykorzystywanego systemu.

Wprawdzie badano obiekt o bardzo specyficznych właściwościach akustycznych (mocno odbijająca podłoga i mocno pochłaniające pozostałe ściany), wydaje się jednak, że metodyka pomiarowa przedstawiona w normie PN-EN 1793-5 może być wykorzystywana i do oceny właściwości pochłaniających ekranów akustycznych, i 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.

Publikacja została opracowana w ramach realizacji projektu: „Samoczyszczące, wydajne panele fotowoltaiczne na podłożu elastycznym zintegrowane z ekranem akustycznym i inteligentnym systemem sterowania" otwartego w ramach konkursu nr 1/4.1.1/2017 Priorytet IV Zwiększenie potencjału naukowo-badawczego Poddziałanie 4.1.1 Strategiczne programy badawcze dla gospodarki, Wspólne Przedsięwzięcie BRIK (Badania i Rozwój w Infrastrukturze Kolejowej).

**Słowa kluczowe:** wskaźnik odbicia dźwięku, współczynnik pochłaniania dźwięku, QUIESST.

## Bibliografia

[1] CEN/TS 16272-5:2014. Railway applications. Track. Noise barriers and related devices acting on airborne sound propagation. Test method for determining the acoustic performance. Intrinsic characteristics. In situ values of sound reflection under direct sound field conditions.

[2] Dziechciński P., *Comparison of digital loudspeaker – equalization techniques*, „Archives of Acoustics" 2005, Vol. 30, No. 2, s. 193–216.

[3] Dziechciński P., *Wybrane problemy pomiarów odpowiedzi impulsowych pomieszczeń*, 56. Otwarte Seminarium z Akustyki, OSA'2009, Goniądz nad Biebrzą, 15–18 września 2009.

[4] Dziechciński P., *Wykorzystanie wskaźnika odbicia dźwięku do oceny właściwości pochłaniających elementów budowlanych*, XVIII Sympozjum Nowości w Technice Audio i Wideo, Wrocław, 15 października 2020.

[5] Garai M., *Measurement of the sound-absorption coefficient in situ: The reflection method using periodic pseudo-random sequences of maximum length*, „Applied Acoustics" 1993, 39, s. 119–139.

[6] 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." 2008, 56, s. 342–355.

[7] Garai M., Guidorzi P., *Sound reflection measurements on noise barriers in critical conditions*, „Building and Environment" 2015, December, Vol. 94, pt. 2, s. 752–763.
[8] Guidorzi P., Klepáček J., Garai M., *On the repeatability of reflection index measurements on noise barriers*, Proceeding of Euronoise 2012, Praga 2012, s. 1314–1319.
[9] Guidorzi P., Garai M., *Advancements in sound reflection and airborne sound insulation measurement on noise barriers*, „Open Journal of Acoustics" 2013, 3, s. 25–38.
[10] London A., *The determination of reverberant sound absorption coefficient from acoustic impedance measurements*, „JASA" 1950, 22(2), s. 263–269.
[11] Mommertz E., *Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique*"„Applied Acoustics" 1995, 46, s. 251–263.
[12] Müller S., Massarani P., *Transfer-Function Measurement with Sweeps*, „J. Audio Eng. Soc." 2001, June, Vol. 49, No. 6, s. 443–471.
[13] Müller G., Möser M. (red.), *Handbook of Engineering Acoustics*, Springer, Berlin 2013.
[14] PN-EN 1793-1 – Drogowe urządzenia przeciwhałasowe – Metoda oznaczania właściwości akustycznych – Część 1: Podstawowe właściwości pochłaniania dźwięku w warunkach rozproszonego pola akustycznego.
[15] 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.
[16] PN-EN 16272-1:2013-04 – Kolejnictwo – Tor – Ekrany akustyczne i obiekty oddziałujące na rozchodzenie się dźwięku w powietrzu – Metoda badawcza do określania właściwości akustycznych – Część 1: Cechy charakterystyczne – Badania laboratoryjne pochłaniania dźwięku rozchodzącego się w powietrzu.
[17] PN-EN 61672-1:2014-03 – Elektroakustyka. Mierniki poziomu dźwięku – Część 1: Wymagania.
[18] PN-EN ISO 354:2005 – Akustyka – Pomiar pochłaniania dźwięku w komorze pogłosowej.
[19] PN-ISO 5725-1:2002 – Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów.
[20] PN-EN ISO 10534-1:2004 – Akustyka – Określanie współczynnika pochłaniania dźwięku i impedancji akustycznej w rurach impedancyjnych – Część 1: Metoda wykorzystująca współczynnik fal stojących.
[21] PN-EN ISO 10534-2:2003 – Akustyka – Określanie współczynnika pochłaniania dźwięku i impedancji akustycznej w rurach impedancyjnych – Część 2: Metoda funkcji przejścia.
[22] Sipar P., Jalkanen T., Siponen D., *Sound reflection from different noise barriers*, Research reports from the Finnish Transport Agency, Helsinki 2017.
[23] 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.

# Indeks nazwisk autorów


Blaszke Maciej **207**
Brachmański Stefan **137**, **157**

Czesak Karol **185**

Dziechciński Paweł **243**

Falkowski-Gilski Przemysław **157**

Gawlińska Małgorzata **173**
Głowiak Maciej **29**

Kin Maurycy **137**
Kleczkowski Piotr **185**
Kostek Bożena **67**, **207**, **225**
Koszewski Damian **207**
Kozłowski Piotr Z. **9**
Kruk Bartłomiej **123**
Kurowski Adam **225**
Łuczyński Michał **105**
Łukasik Ewa **83**

Mróz Bartłomiej **67**

Nowak Tomasz **123**

Odya Piotr **47**, **67**
Opieliński Krzysztof **5**

Pietrusińska Kamila **225**
Pietrzak Cezary **47**
Plaskota Przemysław **173**

Skorupa Jan **29**

Walczak Marcin **83**

**W niniejszej monografii, będącej kolejnym tomem z cyklu, *Postępy badań w inżynierii dźwięku i obrazu*, przedstawiamy Czytelnikom „wycinek" akustyki związany z progresją w zakresie nowych trendów i zastosowań technologii dźwięku wielokanałowego oraz badań jakości dźwięku. W książce zawarto 14 obszernych rozdziałów opracowanych przez polskich akustyków z różnych ośrodków naukowo-badawczych: Katedry Akustyki, Multimediów i Przetwarzania Sygnałów Politechniki Wrocławskiej, Katedry Systemów Multimedialnych oraz Laboratorium Akustyki Fonicznej Politechniki Gdańskiej, Katedry Mechaniki i Wibroakustyki Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, Instytut Informatyki Politechniki Poznańskiej, Poznańskiego Centrum Superkomputerowo-Sieciowego.**

**(fragm. ze Słowa wstępnego)**

**Wydawnictwa Politechniki Wrocławskiej**
**są do nabycia w sprzedaży wysyłkowej:**
**zamawianie.ksiazek@pwr.edu.pl**

**ISBN 978-83-7493-183-0**