Postępy badań w inżynierii dźwięku i obrazu

Pomiary, przetwarzanie, klasyfikacja i ocena jakości sygnałów audio-wideo

pod redakcją

KRZYSZTOFA J. OPIELIŃSKIEGO

Oficyna Wydawnicza Politechniki Wrocławskiej
Wrocław 2023

Postępy badań w inżynierii dźwięku i obrazu

Pomiary, przetwarzanie, klasyfikacja i ocena jakości sygnałów audio-wideo

pod redakcją
Krzysztofa J. Opielińskiego

Oficyna Wydawnicza Politechniki Wrocławskiej
Wrocław 2023

Recenzenci
Krzysztof J. OPIELIŃSKI, Andrzej BRZOSKA, Andrzej CZYŻEWSKI,
Andrzej DOBRUCKI, Tadeusz KAMISIŃSKI, Piotr KLECZKOWSKI,
Bożena KOSTEK, Mirosław MEISSNER, Andrzej MIŚKIEWICZ, Janusz PIECHOWICZ,
Anna PREIS, Tomira ROGALA, Aleksander SĘK, Ewa SKRODZKA, Paweł STRUMIŁŁO,
Jerzy WICIAK, Wiesław WOSZCZYK, Jan ŻERA

Opracowanie redakcyjne
Eric HILTON (rozdz. 8),
Dorota RAWA (rozdz. 9–13),
Katarzyna SOSNOWSKA (rozdz. 1–7)

Korekta
Katarzyna SOSNOWSKA

Projekt okładki
Paweł SPALENIAK

Opracowanie typograficzne
Janusz M. SZAFRAN

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

ISBN 978-83-7493-258-5

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

Spis treści

Słowo wstępne

W kolejnej monografii z cyklu „Postępy badań w inżynierii dźwięku i obrazu" przedstawiamy czytelnikom zagadnienia z obszaru akustyki dotyczące pomiarów, przetwarzania, klasyfikacji i oceny jakości sygnałów audio-wideo w kontekście aktualnych osiągnięć naukowo-badawczych w tym zakresie.

Książka zawiera 13 obszernych rozdziałów, opracowanych przez polskich akustyków z różnych ośrodków naukowo-badawczych:

- Katedry Akustyki, Multimediów i Przetwarzania Sygnałów na Wydziale Elektroniki, Fotoniki i Mikrosystemów Politechniki Wrocławskiej,
- Katedry Informatyki Stosowanej na Wydziale Informatyki i Telekomunikacji Politechniki Wrocławskiej,
- Katedry Inżynierii Systemów, Sygnałów i Elektroniki na Wydziale Elektrycznym Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie,
- Katedry Mechaniki i Wibroakustyki na Wydziale Inżynierii Mechanicznej i Robotyki Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie,
- Katedry Systemów Multimedialnych oraz Laboratorium Akustyki Fonicznej na Wydziale Elektroniki, Telekomunikacji i Akustyki Politechniki Gdańskiej,
- Poznańskiego Centrum Superkomputerowo-Sieciowego afiliowanego przy Instytucie Chemii Bioorganicznej Polskiej Akademii Nauk,

we współpracy z firmami Advanced Bionics Polska, Sonova Audiological Care (Łódź) i KFB Acoustics (Wrocław).

Rozdział 1 monografii zawiera wyniki badań nad oceną jakości sygnałów wideo poddanych kodowaniu H.264/AVC (ang. *Audio Video Coding*) i H.265/HEVC (ang. *High-Efficiency Video Coding*) dokonaną przez młodych widzów. W rozdziale 2 analizie poddana jest zależność między rodzajem kodowania i przepływnością sygnału audio a subiektywną oceną jakości dźwięku różnych gatunków muzycznych. Wyniki eksperymentów potwierdziły m.in. wpływ technik kodowania na ocenę jakości dźwięku utworu muzycznego każdego z badanych gatunków. Celem autorów rozdziału 3 jest analiza metod strojenia systemów nagłaśniania, w których stosowane są algorytmy automatycznej korekcji sygnału audio oraz zadane manualne algorytmy. Badania te przeprowadzono z wykorzystaniem stereofonicznych systemów nagłośniania, przeznaczonych do reprodukcji dźwięku w pomieszczeniach zamkniętych o małej kubaturze. Rozdział 4 dotyczy zagadnień związanych z pomiarami charakterystyk kierunkowości głośników modów

rozproszonych (ang. *Distributed Mode Loudspeakers – DML*), które ze względu na charakterystyki czasowo-częstotliwościowe silnie zależne od wyboru punktu pomiarowego wykazują właściwości inne niż głośniki tłokowe. Autorzy rozdziału 5 proponują modyfikację powszechnie stosowanego kwestionariusza oceny korzyści użytkowania aparatów słuchowych (ang. *Abbreviated Profile of Hearing Aid Benefit* – APHAB), polegającą na połączeniu go z testem MUSHRA (ang. *MUltiple Stimuli with Hidden Reference and Anchor*) stosowanym w ocenie jakości dźwięku oraz przekształceniu skali APHAB na 100-punktową skalę MUSHRA za pomocą logiki rozmytej. Rozdział 6 dotyczy opracowania aplikacji do automatycznego wykrywania mowy patologicznej przez sieci neuronowe wytrenowane na podstawie bazy nagrań. W rozdziale 7 autorzy przedstawiają zagadnienia związane z badaniem niektórych właściwości akustycznych pomieszczeń z wykorzystaniem przenośnego, zintegrowanego systemu pomiarowego zdolnego do wyznaczania przestrzennych natężeniowych odpowiedzi impulsowych, jak również określają możliwe obszary zastosowania tych danych do obiektywnej oceny wybranych właściwości akustyki wnętrz i uprzestrzennienia nagrań muzycznych. Rozdział 8 obejmuje opis zaawansowanego, złożonego systemu monitorowania i identyfikacji zagrożeń wibroakustycznych integrującego techniki uczenia maszynowego i algorytmy deterministyczne w celu wielopoziomowej analizy i przetwarzania sygnałów akustycznych. Niewątpliwie ważna jest przedstawiona w rozdziale 9 koncepcja kształtowania frontu fali akustycznej przez zastosowanie soczewki w formie matrycy falowodów, które dzielą czoło fali na skończoną ilość fragmentów i wprowadzają kontrolowane opóźnienie każdego z nich. Stwarza ona bowiem możliwość kształtowania rozkładu fazy na powierzchni źródła dźwięku. W rozdziale 10 przedstawione są dwa autorskie projekty techniczne instalacji wielogłośnikowych do prezentacji kompozycji przestrzennych. Stanowią one podstawę omówienia problematyki związanej z projektowaniem i konstruowaniem systemów komputerowych pozwalających na zarządzanie sygnałem w instalacjach wielogłośnikowych. Rozdział 11 dotyczy syntezy dźwięku przestrzennego w środowisku wirtualnym, w oparciu o spersonalizowane pomiary HRTF (ang. *Head Related Transfer Functions*). Kolejny rozdział zawiera propozycję wykorzystania uczenia maszynowego w kształtowaniu wiązki w polu bliskim i sygnału szerokopasmowego (mowy ludzkiej). Wyniki symulacji wykazały bowiem, że opracowany system dostosowuje się do zmian położenia mówcy. Przedmiotem rozdziału 13 jest opis algorytmu uprzestrzenniającego sygnały dźwiękowe, działającego w oparciu o metody modyfikacji fazy sygnału na składowych harmonicznych lub w pasmach częstotliwości.

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 Oddziału Wrocławskiego Polskiego Towarzystwa Akustycznego.

Krzysztof J. Opieliński

1. Wpływ kodowania w standardzie H.264 i H.265 na ocenę jakości sygnału wideo przez dwudziestoletniego widza

Stefan Brachmański, Michał Łuczyński

Politechnika Wrocławska,
Wydział Elektroniki, Fotoniki i Mikrosystemów,
wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław

1.1. Wprowadzenie

Przez wiele lat telewizja była najczęściej używaną, a tym samym najważniejszą usługą multimedialną. Współcześnie obserwuje się w tym zakresie duże zmiany. Dotyczy to zwłaszcza młodego pokolenia, które coraz chętniej wybiera oglądanie transmisji telewizyjnych, w tym filmów, za pośrednictwem internetu. Nie należy jednak lekceważyć telewizji jako popularnego medium [11].

Według danych firmy Nielsen [15] duże zmiany w preferencji oglądalności obserwuje się począwszy od 2014 r. Przykładowo, w czwartym kwartale 2014 r. ogólna miesięczna oglądalność tradycyjnej telewizji (obejmującej odtwarzanie na żywo i odtwarzanie z przesunięciem czasowym) spadła o 6 h i 18 min, czyli o 4%, i wyniosła 149 h i 38 min. Wprawdzie wszystkie grupy wiekowe oglądały telewizję krócej niż rok wcześniej, tj. w 2013 r., jednakże najszybciej od tego medium odchodzą młodzi. W grupie wiekowej 18–24 lat oglądalność spadła o 18 h i 37 min, czyli o 16%, w grupie 11–17 lat o 10%, natomiast w grupach 25–34 i 2–11 lat o 8% [14].

Przekaz sygnału wideo realizowany jest z wykorzystaniem różnych standardów kodowania, przy czym International Telecommunication Union (ITU) zaleca korzystanie ze standardu H.264 [6] i nowszego H.265 [7].

Standard H.264 [6], [10], znany również jako AVC (ang. *Advanced Video Coding*), jest 10 częścią standardu MPEG-4. W tym standardzie, podobnie jak we wcześniejszych MPEG2 i MPEG-4, wykorzystuje się kompresję różnicową, tzn. aktualny obraz jest tworzony na podstawie jednego lub kilku poprzednich obrazów. Uwzględnione są również

różnice czasowe występujące między kolejnymi obrazami. Standard H.264 jest wykorzystywany m.in. w różnych aplikacjach internetowych i satelitarnych, sieciach kablowych, filmach o niskiej i wysokiej rozdzielczości, a także w urządzeniach mobilnych i przeglądarkach [2]. Standard ten stał się jednak mniej efektywny przy kompresji sygnałów wideo o bardzo dużej rozdzielczości, np. 4K.

Standard H.265 [7], znany także pod nazwą *High Efficiency Video Coding* (HEVC) (ang.), jest obecnie najnowszym systemem kodowania sygnałów wideo, oferującym znacznie większą kompresję sygnału niż standard H.264. Ponadto obsługuje bardzo duże rozdzielczości, tj. do 8190 × 4320 pikseli (8K). Podstawową jednostką standardu H.265 jest CTB (ang. C*oding Tree Block*) o maksymalnych rozmiarach 64 × 64 pikseli, podczas gdy w standardzie H.264 rozmiary makrobloków wynoszą 16 × 16 pikseli.

Na jakość transmisji sygnału wideo, a tym samym jakość obrazu, wpływ ma m.in. szybkość transmisji i rozdzielczość obrazu.

Głównym celem autorów niniejszego rozdziału było zbadanie wpływu na ocenę jakości sygnału wideo przez młodego użytkownika końcowego takich parametrów jak:

- kodowanie według standardu H.264 i H.265,
- szybkość bitowa transmisji (300–6000 kb/s),
- rozdzielczość SD (ang. *Standard Definition*) 640 × 360, HD (ang. *High Definition*) 1280 × 720 i Full HD (ang. *Full High Definition*) 1920 × 1080).

1.2. Eksperyment

Ocenę jakości sygnału wideo wykonano w standardzie H.264 (AVC) [6] i H.265 (HEVC) [7].

Spośród różnych metod oceny jakości wideo [3], [5], [8], [9], [10], [13] w badaniach zastosowano metodę porównawczą *Double Stimulus Impairment Scale Method* (DSISM) (ang.). Ocena polegała na porównaniu wzorcowej sekwencji wideo (sygnału wzorcowego) z sekwencją ocenianą. Sygnał wzorcowy prezentowany był jako pierwszy, natomiast oceniany jako drugi. Zadaniem obserwatora (widza) była ocena stopnia pogorszenia sygnału drugiego w odniesieniu do sygnału pierwszego. Ocena była podawana w pięciostopniowej skali MOS, przy czym wartość 5 oznaczała pogorszenie jakości niedostrzegalne, a 1 – bardzo dokuczliwe [8].

Wzorcowym materiałem testowym była 20-sekundowa sekwencja wideo (bez dźwięku) o rozdzielczości 1920 × 1080 w formacie avi [1]. Młodzież ogląda przeważnie filmy o dużej ekspresji, w tym filmy akcji, dlatego też materiał testowy zawierał sceny dynamiczne, a mianowicie fragment startu wyścigów konnych. Przykładowy kadr z filmu testowego pokazano na rys. 1.1.

Wzorcowa, oryginalna sekwencja wideo została poddana kodowaniu w standardzie H.264 i H265 z różnymi szybkościami bitowymi i różną rozdzielczością. W badaniach uwzględniono trzy rozdzielczości: 640 × 360, 1280 × 720 i 1920 × 1080. W przypadku obu standardów kodowania i każdej rozdzielczości warunki transmisji symulowane były za

Rys. 1.1. Testowy sygnał wideo – kadr z filmu

pomocą 18 szybkości bitowych: 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500 i 6000 kb/s. Przygotowany w ten sposób materiał testowy prezentowano widzom w laboratorium zaadaptowanym do oceny sygnałów wideo na ekranie telewizora o przekątnej 60 cali. Pomieszczenie laboratoryjne spełniało wymogi zaleceń International Telecommunication Union [3], [8]. Sygnały testowe o różnych warunkach transmisji (różnej szybkości bitowej) były prezentowane widzom w sposób losowy.

Grupę obserwatorów tworzyli studenci Politechniki Wrocławskiej w wieku 20–21 lat, o prawidłowej ostrości widzenia i poprawnym rozróżnianiu kolorów. Zgodnie z zaleceniem International Telecommunication Union BT. 500 [3] minimalna liczba obserwatorów powinna wynosić 15 osób. W prezentowanych badaniach zostały utworzone dwie grupy. Każda grupa oceniała inny typ kodowania. Liczebność poszczególnych grup była różna i wynosiła 45 osób, gdy kodowanie odbywało się w standardzie H.264, i 35 osób, gdy kodowanie odbywało się w standardzie H.265. Przed rozpoczęciem pomiarów uczestnicy zapoznali się z metodą oceny i odbyli jedną sesję treningową, natomiast już w trakcie treningu ze sposobem prezentacji materiału testowego oraz oceny pogorszenia jakości sygnału wideo. Po odbyciu treningu przystąpiono do właściwych pomiarów. Po obejrzeniu sekwencji oryginalnej i kodowanej każdy uczestnik badań zapisywał swoją ocenę pogorszenia jakości na specjalnym formularzu. Zapisane oceny zostały wprowadzone do arkusza kalkulacyjnego i poddane analizie statystycznej zgodnie z procedurą opisaną w zaleceniu ITU-R BT.500 [3]. Pozwoliło to wyeliminować oceny wykraczające poza przyjęty 95-procentowy przedział ufności.

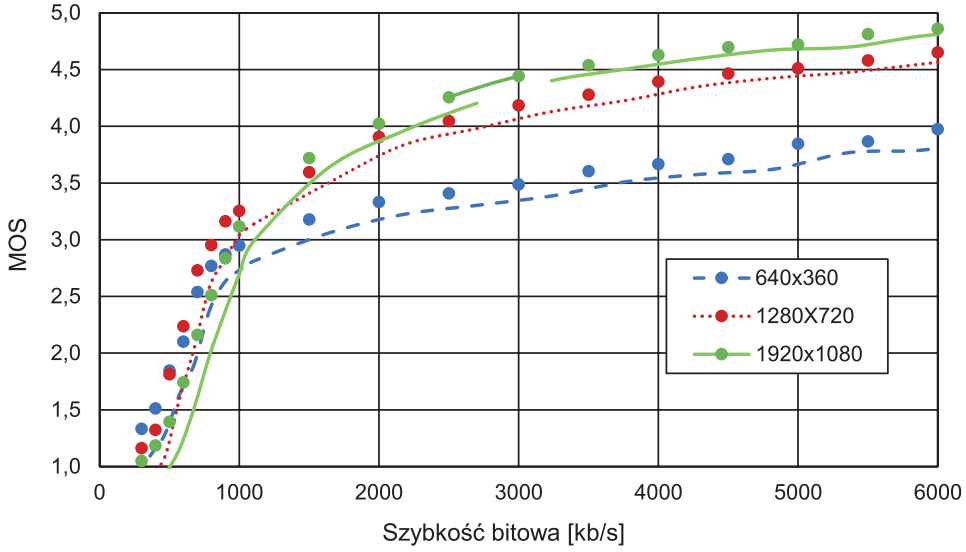
1.3. Wyniki

Ocena degradacji jakości sygnału wideo kodowanego w standardzie H.264 została wykonana w grupie 45-osobowej (złożonej z 40 mężczyzn i 5 kobiet). Otrzymane wyniki zostały przedstawione na rys. 1.2 i w tabeli 1.1.W tabeli 1.1 podano obliczone zgodnie z zaleceniem ITU-R BT.500 [3] wartości średniej oceny jakości wideo (MOS), odchylenia standardowego (S_{cri}) i współczynnika przedziału ufności (δ_{cri}).

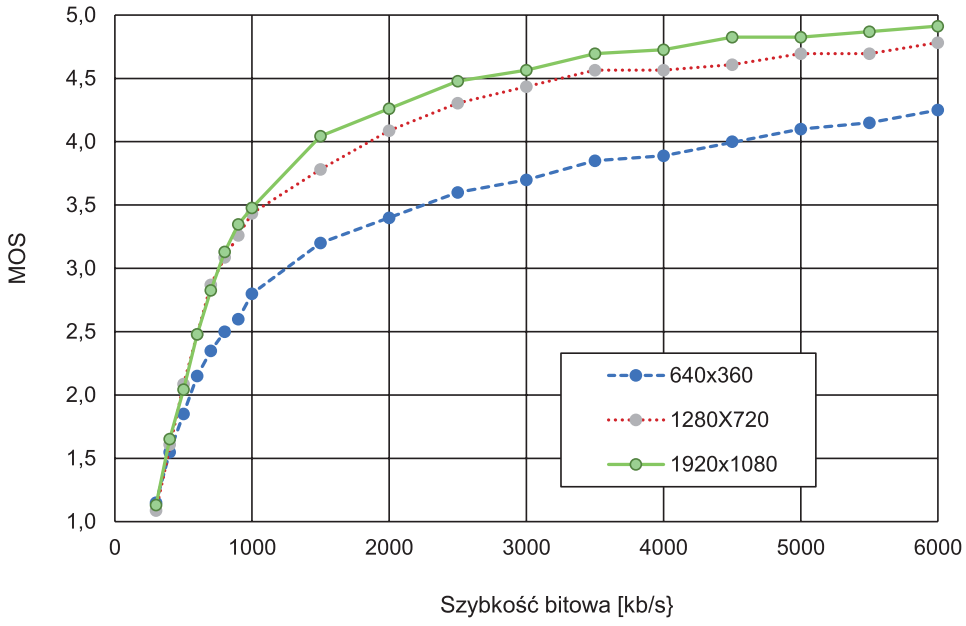
Analiza statystyczna wyników wykazała, że rozdzielczość nie wpływa na ocenę jakości obrazu sygnału wideo przy szybkości bitowej poniżej 1000 kb/s. Powyżej tej wartości jakość sygnału wideo zależy wyraźnie od rozdzielczości i zgodnie z oczekiwaniami najwyżej oceniany był sygnał wideo o rozdzielczości 1920 × 1080. W przypadku tej rozdzielczości ocenę MOS o wartości wynoszącej co najmniej 4,0 osiągnięto, gdy szybkość bitowa była nie mniejsza niż 2000 kb/s. Z kolei przy rozdzielczości 1280 × 720 wartość MOS = 4 osiągnięto, gdy szybkość bitowa nie spadała poniżej 2500 kb/s. W przypadku

Tabela 1.1. Wpływ kodowania w standardzie H.264 I H.265 na ocenę jakości sygnału wideo przez 20-letniego widza

Szybkość bitowa [kb/s]	Rozdzielczość 640 × 360			Rozdzielczość 1280 × 720			Rozdzielczość 1920 × 1080		
	MOS	S	δ	MOS	S	δ	MOS	S	δ
300	1,33	0,48	0,15	1,16	0,43	0,13	1,05	0,32	0,10
400	1,51	0,51	0,16	1,33	0,57	0,17	1,19	0,45	0,13
500	1,85	0,71	0,22	1,81	0,55	0,16	1,40	0,58	0,17
600	2,10	0,75	0,24	2,24	0,66	0,20	1,74	0,58	0,17
700	2,54	0,55	0,17	2,73	0,55	0,17	2,16	0,48	0,14
800	2,77	0,48	0,15	2,95	0,58	0,19	2,51	0,51	0,15
900	2,87	0,52	0,16	3,16	0,37	0,16	2,84	0,48	0,14
1000	2,95	0,65	0,20	3,26	0,49	0,16	3,12	0,50	0,15
1500	3,18	0,56	0,17	3,60	0,59	0,20	3,72	0,45	0,14
2000	3,33	0,66	0,21	3,91	0,43	0,21	4,02	0,46	0,14
2500	3,41	0,59	0,19	4,05	0,49	0,18	4,26	0,49	0,15
3000	3,49	0,56	0,17	4,19	0,50	0,17	4,44	0,50	0,15
3500	3,61	0,75	0,24	4,28	0,45	0,15	4,54	0,55	0,17
4000	3,67	0,66	0,21	4,40	0,49	0,15	4,63	0,49	0,15
4500	3,71	0,65	0,21	4,47	0,55	0,16	4,70	0,46	0,14
5000	3,85	0,67	0,21	4,51	0,55	0,16	4,72	0,45	0,14
5500	3,87	0,58	0,18	4,58	0,50	0,14	4,81	0,39	0,12
6000	3,97	0,58	0,18	4,65	0,48	0,13	4,86	0,35	0,10



Rys. 1.2. Zależność oceny jakości wideo (MOS) kodowanego w standardzie H.264 w funkcji szybkości bitowej przy rozdzielczości 640 × 360, 480, 1280 × 720 i 1920 × 1080



Rys. 1.3. Zależność oceny jakości wideo (MOS) kodowanego w standardzie H.265 w funkcji szybkości bitowej przy rozdzielczości 640 × 360, 480, 1280 × 720 i 1920 × 1080

Tabela 1.2. Wpływ kodowania w standardzie H.265 na ocenę jakości sygnału wideo przez dwudziestoletniego widza

Szybkość bitowa [kb/s]	Rozdzielczość 640 × 360			Rozdzielczość 1280 × 720			Rozdzielczość 1920 × 1080		
	MOS	S	δ	MOS	S	δ	MOS	S	δ
300	1,15	0,37	0,16	1,09	0,29	0,12	1,13	0,34	0,14
400	1,55	0,51	0,22	1,61	0,50	0,20	1,65	0,49	0,20
500	1,85	0,75	0,33	2,09	0,42	0,17	2,04	0,64	0,26
600	2,15	0,59	0,28	2,48	0,59	0,24	2,48	0,51	0,21
700	2,35	0,59	0,31	2,87	0,34	0,14	2,83	0,58	0,24
800	2,50	0,51	0,22	3,09	0,42	0,17	3,13	0,76	0,31
900	2,60	0,50	0,22	3,26	0,45	0,18	3,35	0,49	0,20
1000	2,80	0,62	0,27	3,43	0,51	0,21	3,48	0,59	0,24
1500	3,20	0,52	0,23	3,78	0,42	0,17	4,04	0,37	0,18
2000	3,40	0,50	0,22	4,09	0,67	0,27	4,26	0,45	0,21
2500	3,60	0,75	0,33	4,30	0,56	0,23	4,48	0,51	0,21
3000	3,70	0,47	0,21	4,43	0,59	0,24	4,57	0,51	0,21
3500	3,85	0,37	0,16	4,57	0,59	0,24	4,70	0,47	0,19
4000	3,89	0,32	0,15	4,57	0,51	0,21	4,73	0,46	0,19
4500	4,00	0,00	0,00	4,61	0,50	0,20	4,83	0,39	0,16
5000	4,10	0,45	0,20	4,70	0,47	0,19	4,83	0,39	0,16
5500	4,15	0,37	0,16	4,70	0,47	0,19	4,87	0,34	0,14
6000	4,25	0,44	0,19	4,78	0,42	0,17	4,91	0,29	0,12

najmniejszej badanej rozdzielczości, tj. 640 × 360, przy szybkości bitowej 6000 kb/s osiągnięto wartość MOS = 3,97, czyli prawie poziom 4,0.

Ocena degradacji jakości sygnału wideo kodowanego w standardzie H.265 została wykonana w grupie 35-osobowej (złożonej z 31mężczyzn i 4 kobiet). Otrzymane wyniki przedstawiono na rys. 1.3 i w tabeli 1.2. Identycznie jak w przypadku kodowania w standardzie H.264 (tabela 1.1) przedstawiono obliczone zgodnie z zaleceniem ITU-R BT.500 [3] średnią ocenę jakości wideo (MOS), odchylenie standardowe (S_{cri}) i współczynnik przedziału ufności (δ_{cri}).

Analiza statystyczna wyników wykazała, że rozdzielczość nie wpływa na MOS przy szybkości bitowej równej 500 kb/s lub mniejszej. Porównując sygnał wideo o rozdzielczości 1280 × 720 i 1920 × 1080, można zauważyć, że w ocenie jakości nie ma istotnej różnicy, gdy szybkość bitowa wynoszącej min. 1000 kb/s; powyżej tej szybkości jakość sygnału wideo zależy nieznacznie od rozdzielczości. Z kolei ocena wideo o najmniejszej badanej rozdzielczości, czyli 640 × 300, wyraźnie odbiega od oceny dwóch pozostałych rozdzielczości. W przypadku sygnału wideo o rozdzielczości 1920 × 1080 MOS prze-

kracza wartość 4,0 przy szybkości bitowej począwszy od 1500 kb/s. Z kolei przy rozdzielczości 1280 × 720 do osiągnięcia wartości MOS powyżej konieczna była szybkość większa od 2000 kb/s. Dla najmniejszej badanej rozdzielczości, tj. 640 × 360, otrzymano maksymalną wartość MOS równą 4,25 przy szybkości bitowej 6000 kb/s.

W przypadku standardu H.264 można zauważyć, że standard H.265, zgodnie z oczekiwaniami, pozwala na uzyskanie wyższych wartości MOS.

1.4. Podsumowanie

Subiektywna ocena jakości sygnału wideo dokonana przez młodego widza potwierdza teoretyczne sugestie o większych możliwościach standardu H.265 w odniesieniu do H.264 zarówno pod względem parametrów kompresji, jak i jakości kompresowanego sygnału wideo przy dużych rozdzielczościach. Standard H.265 umożliwia uzyskanie bardzo dobrej jakości (MOS = 4,5) sygnału wideo Full HD (1920 × 1080) począwszy od szybkości bitowej wynoszącej 3500 kb/s. Z kolei w standardzie H.264 MOS = 4,5 otrzymuje się wtedy, gdy szybkość jest większa niż 3500 kb/s. Uwzględniając zalecenia ITU określające jako dopuszczalną wartość MOS = 4,0 (czego skutkiem jest zauważalne, lecz niedokuczliwe pogorszenie jakości), można stwierdzić, że standard H.265 zapewnia taką ocenę już przy szybkości 1500 kb/s, natomiast H.264 przy szybkości powyżej 2000 kb/s.

Porównując sygnał wideo HD (1280 × 720) w obu standardach, można zauważyć, że w przypadku standardu H.264 MOS = 4,5 uzyskuje się przy prędkości bitowej 5000 kb/s, natomiast w przypadku H.265 już przy prędkości bitowej 3500 kb/s. Z kolei wartość MOS = 4,0 uzyskano odpowiednio przy 2500 i 2000 kb/s.

Najgorzej wypada sygnał wideo SD o rozdzielczości 640 × 360, który maksymalną ocenę osiąga przy szybkości 6000 kb/s, przyjmując w standardzie H.264 wartość MOS = 3,97, a w standardzie H.265 wartość 2,25.

Można zauważyć, że różnice między obydwoma standardami są nieduże, co wynika z ocenianych rozdzielczości. Sygnały wideo kodowane zgodnie ze standardem H.264 cechują się w miarę dobrą efektywnością przy rozdzielczości 1280 × 720 i 1920 × 1080. Opracowując standard H.265, założono, że będzie bardzo dobrze radził sobie zarówno z niskimi rozdzielczościami, jak i z tymi największymi, np. 4K UHD (3840 × 2160) czy 8K UHD (7680 × 4320).

Bibliografia

[1] Brachmański S., Klink J., *Subjective assessment of the quality of video sequences by the young viewers*; 30th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2022, Split, Croatia, FESB; University of Split, IEEE2022, s. 1–6.

[2] Buchowicz A., Galiński G., *Strumieniowanie danych wideo kodowanych w standardzie MPEG-4 AVC/H.264*, „Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne" 2010, nr 12, s. 1727–1731.
[3] ITU-R Recommendation BT. 500-14, *Methodologies for the subjective assessment of the quality of television images*, International Telecommunication Union, 2019.
[4] ITU-R Recommendation BT. 709-6, *Parameter values for the HDTV standards for production and international programme exchange*, International Telecommunication Union, 2015.
[5] ITU-R Recommendation BT. 1129-2, *Subjective assessment of standard definition digital television (SDTV) systems*, International Telecommunication Union, 2019
[6] ITU-T Recommendation H.264, *Advanced video coding for generic audiovisual services,* International Telecommunication Union, 2021.
[7] ITU-T Recommendation H.265, *High efficiency video coding,* International Telecommunication Union, 2021.
[8] ITU-T Recommendation P.910, *Subjective video quality assessment methods for multimedia applications,* International Telecommunication Union, 1998
[9] Pinson M., Wolf S., *Comparing subjective video quality testing methodologies*, „Visual Communications and Image Processing" 2003, s. 573–582.
[10] Pongsapan F.P., Hendrawan, *Evaluation of HEVC vs H.264/A VC video compression transmission on LTE network*, *11th International Conference on Telecommunication Systems Services and Applications (TSSA),* 2017, s. 1–4.
[11] Shaikh S.J., *Television Versus the Internet for Information Seeking: Lessons From Global Survey Research*, „International Journal of Communication" 2017, Vol. 11, s. 4744–4756.
[12] Schäfer R., Wiegand T., Schwarz H., *The emerging H.264/AVC standard*, „EBU Technical Review" 2003, January.
[13] Winkler S., *Video quality measurement standards – current status and trends*, 2009 7th International Conference on Information, Communications and Signal Processing (ICICS), IEEE, 2009, s. 1–5.
[14] https://nscreenmedia.com/young-millennials-watch-16-less-tv-21-internet-video/ [dostęp: 08.08.2023].
[15] www.nielsen.com/ [dostęp: 08.08.2023].

Słowa kluczowe: ocena jakości video, kodowanie video, badania subiektywne.

Wpływ kodowania w standardzie H.264 I H.265 na ocenę jakości sygnału wideo przez młodych widzów

W rozdziale przedstawiono wyniki subiektywnych testów oceny jakości sygnałów wideo poddanych kodowaniu w standardzie H.264/AVC (ang. *Audio Video Coding*) i H.265/HEVC (ang. *High-Efficiency Video Coding*). Oceny dokonano w warunkach laboratoryjnych. Sygnał wideo oceniali młodzi widzowie, którzy nie byli ekspertami w dziedzinie oceny jakości. Testy przeprowadzono z uwzględnieniem trzech rozdzielczości obrazu (640 × 360, 1280 × 720, 1920 × 1080) i różnych przepływności (3000–6000 kb/s). Uzyskane wyniki pokazały, że w przypadku standardu H.265 najlepszą jakość sygnału wideo Full HD uzyskano przy minimalnej przepływności 3000 kb/s, natomiast akceptowalną jakość obrazu zapewniała już wartość 1500 kb/s. W standardzie H.264 przepływności są wyższe i wynoszą odpowiednio 3500 i 2500 kb/s.

Effect of H.264 and H.265 coding on the assessment of the quality of the video signal by a young viewer

The chapter presents the results of subjective tests for assessing the quality of video signals subjected to H.264/AVC (Audio Video Coding) and H.265/HEVC (High-Efficiency Video Coding) encoding. Evaluation was carried out under laboratory conditions. The video signal was assessed by young viewers who were not experts in the field of quality assessment. Tests were carried out taking into account three image resolutions (640 × 360, 1280 × 720,1920 × 1080) and different bit rates (from 300 kbps to 6000 kbps). The results obtained showed that for the H.265 standard the best quality was obtained for the minimum bit rate of 3000 kbps, while the value of 1500 kbps already ensures acceptable image quality. In turn, in the H.264 standard, the bit rates are higher and amount to 3500 and 2000 kbps, respectively.

2. Badanie wpływu wybranych technik kodowania na jakość dźwięku w nagraniach różnych gatunków muzyki

Stefan Brachmański, Maurycy Kin, Piotr Nowak

Politechnika Wrocławska,
Wydział Elektroniki, Fotoniki i Mikrosystemów,
wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław

2.1. Wprowadzenie

Rozwój technologii cyfrowych dostarczył nowych możliwości przetwarzania sygnałów fonicznych, których najważniejszą zaletą jest łatwość przechowywania, przesyłania i przetwarzania. Wciąż jednak istnieją ograniczenia związane z magazynowaniem i przepustowością kanałów transmisyjnych. W wielu przypadkach wskazane jest, aby sygnały były zakodowane jak najwydajniej, tzn. aby miały jak najmniejszą objętość przy zachowaniu bardzo dobrej jakości.

Najprostszym sposobem ograniczenia ilości informacji wydaje się zastosowanie niższej częstotliwości próbkowania i mniejszej rozdzielczości bitowej. Skutkiem tego zabiegu jest jednak znaczna degradacją jakości sygnału. Konieczne stało się zatem znalezienie innych metod pozwalających na zmniejszenie ilości informacji. Efektywna w telekomunikacji kwantyzacja liniowa nie pozwala na uzyskanie wystarczającej jakości złożonych sygnałów fonicznych (np. muzyki).

Najbardziej popularną metodą, mającą zastosowanie w przypadku wszystkich rodzajów sygnałów, jest kompresja danych. Może ona odbywać się z wykorzystaniem algorytmów stratnych i bezstratnych. Zasada działania kompresji bezstratnej opiera się na skondensowaniu informacji do postaci zawierającej mniejszą liczbę bitów. Warunkiem jest gwarancja możliwości takiego odtworzenia informacji, aby była ona identyczna z oryginałem. Kompresja stratna ma również spowodować zmniejszenie liczby bitów potrzebnych do przedstawienia informacji, jednak ze względu na usuwanie z sygnału

oryginalnego informacji redundantnych nie gwarantuje zachowania identyczności przetworzonej informacji z oryginałem.

Do algorytmów kompresji stratnej zalicza się popularny format MP3. Algorytm stosowany podczas tej kompresji jest oparty na modelu psychoakustycznym. Główną ideą tego rodzaju kompresji jest eliminacja z sygnału tych informacji, których człowiek nie jest w stanie usłyszeć w danych warunkach. Uzyskane dane są poddawane dodatkowej kompresji bezstratnej w celu eliminacji nadmiarowości i skondensowaniu informacji [1], [3].

Innym popularnym standardem jest AAC (ang. *Advanced Audio Coding*). Ta metoda kompresji została zaprojektowana głównie w celu podniesienia jakości dźwięku przy podobnym do MP3 rozmiarze danych. Składa się z zaawansowanych algorytmów mających na celu poprawę jakości przy niskich przepływnościach. Kompresja AAC opiera się na zastosowaniu dwóch strategii kodowania, które mają na celu zmniejszenie ilości danych potrzebnych do reprezentowania dźwięku cyfrowego o jak najwyższej jakości. Obecnie technika AAC jest standardowym formatem audio urządzeń elektronicznych takich jak telefony, konsole do gier czy aplikacje streamingowe, a także jest wykorzystywana w cyfrowym radiu DAB.

Rozszerzeniem formatu AAC jest kodowanie HE-AAC (ang. *High-Efficiency Advanced Audio Coding*), które zostało zaprojektowane w celu otrzymywania możliwie najlepszej jakości dźwięku przy niskich przepływnościach strumieni binarnych. Wykorzystano w nim technologię *Spectral Band Replication* (ang.) [6], [10], [17]. Format HE-AAC sprawdza się głównie w usługach audio świadczonych za pośrednictwem sieci WiFi lub 3G, ale także w cyfrowej telewizji satelitarnej i kablowej czy w cyfrowym radiu DAB+.

Głównym celem autorów niniejszego rozdziału jest zbadanie, jaki wpływ na subiektywną ocenę jakości dźwięku różnych gatunków muzyki ma rodzaj techniki kodowania i zastosowana prędkość bitowa. W ramach eksperymentu słuchacze oceniali ogólną jakość dźwięku oraz dwa atrybuty wrażenia słuchowego: barwę dźwięku i przestrzenność.

2.2. Metoda badań

2.2.1. Sygnały testowe

Do przygotowywania próbek testowych zastosowano GX-Transcoder, natomiast edycji długości materiału dokonywano za pomocą edytora Samplitude v.8.

Materiał testowy składał się z następujących gatunków i reprezentujących je utworów muzycznych:

- blues: Dżem *Wehikuł czasu,*
- heavy metal: Metallica *Enter Sandman,*
- hip-hop: 2Pac *Changes*,

- muzyka klasyczna: Wolfganag Amadeus Mozart *Requiem*,
- pop: Michael Jackson *Billie Jean*,
- rock: Linkin Park *Numb*.

Źródłem powyższych utworów były płyty CD. Wybrane fragmenty trwały ok. 15–20 s, przy czym zostały one wyedytowane z zachowaniem prawideł kontekstu muzycznego [9]. Materiał testowy z każdego gatunku muzycznego składał się łącznie z 6 próbek różniących się zastosowanymi technikami kodowania oraz przepływnością danych.

2.2.2. Metody oceny

Ocenę ogólnej jakości dźwięku i badanych atrybutów dokonano z zastosowaniem metody ACR (ang. *Absolute Category Rating*) z pięciostopniową skalą ocen, zgodnie z zaleceniami ITU [7]:

1 – bardzo słaba barwa/przestrzenność dźwięku i ogólna jakość materiału,
2 – słaba barwa/przestrzenność dźwięku i ogólna jakość materiału,
3 – przeciętna barwa/przestrzenność dźwięku i ogólna jakość materiału,
4 – dobra barwa/przestrzenność dźwięku i ogólna jakość materiału,
5 – znakomita barwa/przestrzenność dźwięku i ogólna jakość materiału.

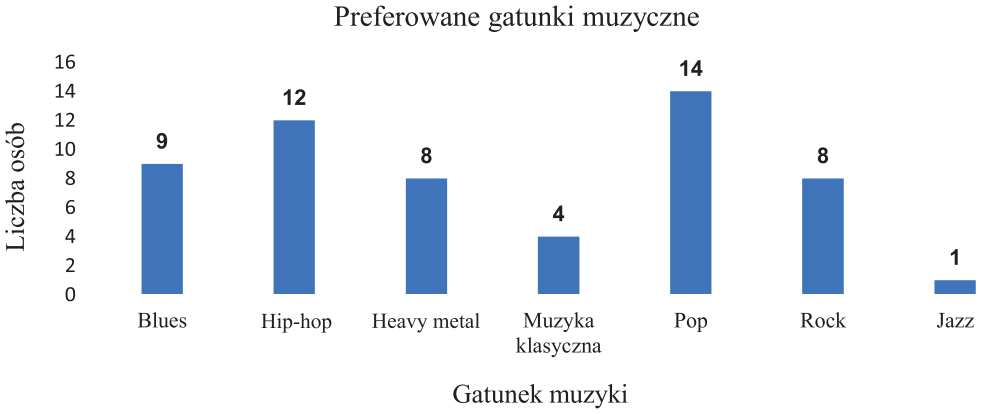
Ankieta została przygotowana z wykorzystaniem formularza Google: zawierała pytania dotyczące płci, przedziału wiekowego, częstości słuchania muzyki, umiejętności gry na instrumentach i ulubionych gatunków muzycznych. Uczestnik miał też za zadanie odpowiedzieć, czy występuje u niego ubytek słuchu oraz w jaki sposób będzie dokonywany odsłuch materiału testowego, ponieważ miejsce odsłuchiwania próbek (np. dom, park czy zatłoczony tramwaj) mogło być istotne ze względu na dokonywane oceny.

Każdy z uczestników odsłuchiwał materiał testowy indywidualnie, w najbardziej dogodnych dla siebie warunkach. Miało to na celu zapewnienie jak najbardziej naturalnych i preferowanych okoliczności odbioru muzyki, ponieważ nie badano zniekształceń wprowadzanych przez kodeki, ale wrażenia słuchaczy jako konsumentów, charakterystyczne przy ocenie *Quality of Experience* (ang.). Dodatkowym argumentem przemawiającym za takimi warunkami odsłuchu było niestwierdzenie różnic między wynikami oceny jakości sygnału mowy przeprowadzonymi w warunkach studyjnych oraz domowych [2]. Takie warunki oceny pozwoliły także na zachowanie naturalnych warunków odbioru audycji radiowych (na ogół w miejscu zamieszkania). Zaletą takiego rozwiązania była łatwość odsłuchu przygotowanych materiałów przez uczestników ankiety oraz swobodny czas realizacji testu.

Badaniom poddano metody kodowania AAC, HE-AAC i MP3 przy przepływnościach 64 i 128 kb/s. Słuchacze oceniali ogólną jakość dźwięku oraz dwa atrybuty wrażeń słuchowych: barwę i przestrzenność. Próbki dźwiękowe zawierające kilkunastosekundowe fragmenty utworów podanych w podrozdz. 2.2.1 były wybierane losowo. Każda próbka była prezentowana jednokrotnie, a czas na udzielenie odpowiedzi był regulowany przez słuchacza.

2.3. Ekipa słuchaczy

W eksperymencie wzięło udział 21 osób, przy czym 15 osób (71,4%) było z przedziału wiekowego 18–26 lat, 4 osoby z przedziału wiekowego 27–35 lat. Jedna osoba miała więcej niż 36 i jedna mniej niż 18 lat.



Rys. 2.1. Gatunki muzyczne preferowane przez uczestników ankiety badawczej

Rys. 2.2. Warunki odsłuchowe przygotowanego materiału dźwiękowego

Spośród 21 ankietowanych 18 zaznaczyło, że słucha muzyki codziennie przez 2–3 h, a pozostałe osoby, że poświęcają na odsłuch materiałów dźwiękowych 2–6 dni w tygodniu przez 2–4 h dziennie.

Na pytanie związane z grą na instrumentach muzycznych 15 osób odpowiedziało, że nie gra na żadnym instrumencie, natomiast 6 osób, że ma doświadczenie muzyczne w postaci gry na instrumentach.

Kolejnym etapem było zaznaczenie na podanej liście ulubionego gatunku muzycznego. Można było jednak nie ograniczać się do jednej odpowiedzi, ale wskazać kilka. Wyniki ilustrujące liczbę osób preferujących dany gatunek muzyki przedstawiono na rys. 2.1. Jak można zauważyć, wykorzystane w badaniach gatunki muzyczne pokrywają się z preferencjami ankietowanych.

Grupa odsłuchowa złożona była z osób, spośród których żadna nie deklarowała ubytku słuchu. Ostatnie pytanie ankiety dotyczyło warunków odsłuchowych przygotowanych materiałów dźwiękowych, które są istotne z punktu widzenia analizowanej percepcji dźwięku. Otrzymane wyniki pokazano na rys. 2.2, a cyfry podane w poszczególnych sektorach określają liczbę osób wykorzystujących do odsłuchu określone urządzenie.

Na podstawie otrzymanych wyników można wnioskować, że większość, bo 61,9%, wszystkich odsłuchiwało próbki testowe z laptopa, za pomocą wbudowanych głośników (6 osób) lub słuchawek przewodowych (7 słuchaczy), 23,8% użyło telefonu, jedna osoba wykorzystała monitory odsłuchowe, a dwie osoby tablet.

2.4. Wyniki badań

W tabeli 2.1 oceny każdego gatunku muzycznego podano uśrednione wyniki oceny każdego gatunku muzycznego. Odnoszą się one do całej grupy (odchylenie standardowe poszczególnych wyników nie przekraczało 20% prezentowanych wartości), ponieważ test oparty na analizie wariancji (ANOVA) wykazał, że na poziomie istotności $\alpha = 0{,}05$ nie odnotowano istotnego statystycznie wpływu poszczególnych słuchaczy na uzyskane wyniki ($F = 1{,}235$ przy wartości krytycznej $F_\alpha = 2{,}208$). Oznacza to, że uzyskane odpowiedzi są jednolite w całej grupie. Stwierdzono także, że wpływ warunków odsłuchowych na uzyskane wyniki także nie jest istotny statystycznie ($F = 1{,}193$, $F_\alpha = 1{,}912$).

Na podstawie otrzymanych wyników można stwierdzić, że przy przepływności 64 kb/s żaden z badanych gatunków nie został oceniony jako dobry (MOS ≥ 4) względem ogólnej oceny jakości czy badanych atrybutów, tzn. barwy dźwięku oraz wrażenia przestrzenności. Należy jednak zaznaczyć, że przy tej przepływności najwyższe oceny uzyskano w przypadku hip-hopu, co może wynikać ze sposobu produkcji i zgrania materiału: w tych nagraniach stosunkowo najmniej jest przestrzeni, a wszystkie elementy dźwiękobrazu zlokalizowane są blisko środka panoramy oraz poddane mocnej kompresji, co skutkuje wrażeniem małej głębi [12], [14], [15]. Można zatem pokusić się o stwierdzenie, że pod względem pogorszenia jakości nagrania gatunek jest mało wraż-

Tabela 2.1. Uśrednione wyniki oceny jakości dźwięku różnych gatunków muzyki kodowanej wybranymi technikami

Gatunek	Przepływność/ kodowanie [kb/s]	Ocena ogólna			Barwa dźwięku			Przestrzenność		
		MP3	AAC	HE-AAC	MP3	AAC	HE-AAC	MP3	AAC	HE-AAC
Blues	64	2,2	3,2	3,5	2,1	2,7	3,1	2,2	2,5	3,1
	128	3,6	3,8	3,9	3,5	3,4	3,9	3,7	3,8	4,0
Heavy metal	64	2,5	2,6	2,8	2,8	3,6	3,8	2,6	2,7	2,7
	128	3,7	4,0	4,1	3,6	3,7	4,0	3,6	3,9	4,1
Hip-hop	64	3,7	3,7	3,8	3,6	3,5	3,7	3,8	3,7	3,8
	128	3,9	4,0	4,0	3,9	3,9	4,1	4,1	4,0	4,2
Pop	64	3,4	3,6	3,5	3,3	3,6	3,6	3,1	3,2	2,9
	128	3,8	4,0	3,9	3,6	4,2	4,3	3,6	3,9	3,8
Rock	64	2,4	2,8	2,9	2,5	2,7	3,1	2,2	2,2	2,7
	128	2,9	3,2	3,7	3,6	4,0	3,8	2,6	3,7	3,7
Muzyka klasyczna	64	2,0	2,2	2,3	2,1	2,5	2,5	2,0	2,3	2,4
	128	3,1	3,4	3,7	3,2	3,6	3,7	3,2	3,5	3,6

liwy na usuwanie elementów redundantnych. Dwukrotne zwiększenie przepływności (do 128 kb/s) spowodowało, że także nagrania hip-hopowe uzyskały najwyższe oceny, co potwierdza postawioną wcześniej tezę o najmniejszej wrażliwości tego gatunku na działanie kompresji stratnej. Otrzymane subiektywne oceny jakości wszystkich testowanych metod kodowania mogą zatem sugerować stosowanie niższych prędkości bitowych przez stacje radiowe nadające taką właśnie muzykę.

Drugą co do odporności na zabiegi związane z kodowaniem stratnym okazała się muzyka pop: przy przepływności 128 kb/s oraz kodowaniu w standardach AAC i HE-AAC uzyskano zadowalające rezultaty, jeśli chodzi o ogólną jakość dźwięku i jego barwę, natomiast wrażenie przestrzenności nagrania zostało ocenione jako nieznacznie gorsze (odpowiednio MOS = 3,9 i MOS = 3,8). Podobnie jak w przypadku hip-hopu autorzy upatrują przyczyny takiego stanu rzeczy w sposobie produkcji muzyki z tego gatunku – współczesne tendencje opierają się także na kompresji, jednak wrażenie przestrzenności nagrań jest większe niż w przypadku hip-hopu [16].

Najniższe oceny otrzymały próbki muzyki klasycznej: przy przepływności 64 kb/s wartości MOS zawierały się w przedziale 2–2,5, natomiast przy przepływności 128 kb/s w przedziale 3,1–3,7. Jest to niewątpliwie związane z większym znaczeniem i udziałem przestrzenności (z szerszą panoramą stereofoniczną) oraz większą wrażliwością słuchaczy na zmiany barwy dźwięku instrumentów akustycznych spowodowane usuwaniem składowych wyższych rzędów, które zostały, jak się wydaje błędnie, uznane za nadmiarowe z punktu widzenia kodowania percepcyjnego.

Otrzymane wyniki były więc zgodne z oczekiwaniami, ponieważ spodziewano się uzyskania słabej (w przypadku MP3) lub przeciętnej (w przypadku AAC i HE-AAC) przestrzenności brzmienia oraz ogólnej jakości dźwięku badanego materiału dźwiękowego.

2.5. Omówienie wyników

Wyniki uzyskane w eksperymencie są zgodne z wynikami uzyskanymi we wcześniejszych badaniach. Jednoznacznie potwierdzają wpływ technik kodowania na ocenę jakości brzmienia utworu muzycznego z każdego badanego gatunku. Trudno jest wybrać najlepszą metodę, jednak na podstawie przeprowadzonych badań można wnioskować, że AAC i HE–AAC umożliwiają uzyskanie zadowalającej jakości (MOS ≥ 3,5) próbki dźwiękowej przy niższych przepływnościach niż kodowanie MP3. Zastosowanie największej możliwej przepływności w przypadku badanych gatunków muzycznych nie zawsze wpływa na uzyskanie najlepszej barwy dźwięku, przestrzenności nagrania czy ogólnej jakości dźwięku, o czym świadczą uzyskane wyniki. Odpowiednią wartością przepływności w przypadku każdego badanego gatunku jest zatem 128 kb/s, ponieważ przeważnie daje zadowalające rezultaty. Należy zrezygnować z niższych wartości przepływności, gdyż skutkują znaczącym pogorszeniem parametrów próbek dźwiękowych. Nieistotny zdaje się wpływ warunków odsłuchowych, pomieszczenie i sprzęt odsłuchowy, chociaż nie wszystko to, co wyraźnie dało się usłyszeć na profesjonalnych monitorach, było w równym stopniu percypowane w warunkach odsłuchu domowego. Domowe zestawy odsłuchowe najczęściej nie należą bowiem do tych o najwyższej jakości, jednak badani słuchacze, wskutek przyzwyczajenia do brzmienia tego typu urządzeń, akceptują taką jakość dźwięku. Autorzy sugerują, że wpływ na taki stan rzeczy miały preferencje słuchaczy i znajomość poszczególnych gatunków. Również zastosowana metoda skalowania absolutnego (bez porównania z wzorcem) mogła mieć wpływ na uzyskane wyniki. Nie dziwi więc fakt, że w Internecie najczęściej spotykanymi zasobami są MP3 o przepływnościach 96 i 128 kb/s, ponieważ przepływność 64 kb/s, poza nielicznymi wyjątkami, nie zapewnia wystarczającej jakości dźwięku, jeśli chodzi o sygnały muzyczne [13]. Przy przepływnościach większych odbiór konstrukcji dźwiękowej oparty jest na percypowaniu całości, podobnie jak to ma miejsce w przypadku przestrzeni wizualnej: jeżeli z pojedynczych form lub elementów (części) utworzona jest spoista całość, to dodanie lub usunięcie elementów tworzących tę spoistość nie jest zawsze wyraźnie spostrzegane, podobnie jak w przypadku zmysłu wzroku [18]. Dlatego celowe wydaje się zachowanie w procesie redukcji danych lub wykreowanie na etapie produkcji nagrania swoistej aury dźwiękowej, która mogłaby pomóc w kształtowaniu pewnych wrażeń zmysłowych na zasadzie tła, na którym można byłoby umieścić wszystkie zdarzenia dźwiękowe wywołujące określone wrażenia. Mogłoby to zmniejszyć stopień degradacji atrybutów przestrzennych ocenianego dźwięku, poddanego kompresji stratnej.

Dostrzegalny wydaje się także związek między rodzajem sygnału a zarejestrowanymi zmianami: sygnały o mniejszej złożoności, w których dominuje mowa lub efekty dźwiękowe (jak w przypadku hip-hopu), były zawsze oceniane wyżej niż próbki zawierające elementy typowo muzyczne (rock czy nawet pop) [8], [16]. Duży wpływ na percepcję zmian spowodowanych kodowaniem stratnym mają też indywidualne preferencje słuchaczy. Osoby, które na co dzień słuchają przede wszystkim danego gatunku (np. muzyki rockowej), najłatwiej zauważą zmiany w obrębie tego właśnie gatunku, natomiast bardziej tolerują zmianę jakości dźwięku w przypadku innych gatunków. Podobne preferencje były już notowane w literaturze przedmiotu w odniesieniu do percepcji sygnałów poddanych kompresji amplitudy [12], [15]. Kompresja danych utworów muzyki klasycznej objawia się natomiast zmianą wyobrażenia o pomieszczeniu, w którym nagranie zostało dokonane [4], [5]. Aby taką zmianę usłyszeć, konieczne jest osłuchanie z tego rodzaju sygnałami, zwłaszcza przy braku sygnałów wzorcowych, co także wiąże się z preferencjami słuchaczy biorących udział w badaniu.

Kolejnym ważnym aspektem jest brak wpływu warunków odsłuchowych materiału testowego, które było bardzo zróżnicowane: poczynając od laptopa z wbudowanym głośnikiem lub z podłączonymi słuchawkami, telefonu czy tabletu, a kończąc na wyspecjalizowanych monitorach odsłuchowych. Wyniki uzyskane w eksperymencie dają podstawę do ograniczania wartości przepływności sygnałów kodowanych stratnie przez nadawców, dla których zmniejszenie objętości przesyłanych informacji jest sprawą kluczową.

Należy także wspomnieć, że w literaturze przedmiotu odnotowane jest występowanie korelacji między ogólną oceną jakości dźwięku a oceną poszczególnych atrybutów dźwięku muzycznego: jeśli jakikolwiek atrybut wrażenia słuchowego jest percypowany jako słaby (nawet bez konieczności jego oceny) i przeszkadza w odbiorze, w konsekwencji dochodzi do obniżenia oceny ogólnej jakości dźwięku odsłuchiwanego fragmentu. Kiedy słuchacze posiadają pozytywne odczucia odnośnie do poszczególnych atrybutów dźwięku, to automatycznie ocena jakości dźwięku jest bardziej korzystna [9], [12], [16].

2.6. Wnioski

Badania wykazały, że najwyższe oceny jakości ogólnej, a także dwóch badanych atrybutów uzyskały próbki zakodowane za pomocą metody HE-AAC, natomiast najgorzej wypadło kodowanie MP3. Rezultaty zdecydowanie polepszyły się wraz ze zwiększeniem przepływności – w tym przypadku również najlepsza okazała się technika HE-AAC.

Najniższe oceny otrzymały próbki muzyki klasycznej: przy przepływności 64 kb/s wartości MOS zawierają się w przedziale 2–2,5, natomiast przy przepływności 128 kb/s MOS wynosi 3,1 w przypadku próbek MP3 i 3,7 w przypadku kodowania metodą HE-AAC. Jest to niewątpliwie związane z większym znaczeniem i udziałem przestrzenności ocenianych nagrań oraz większą wrażliwością słuchaczy na zmiany barwy dźwięku instrumentów naturalnych w porównaniu do innych gatunków muzyki. W poprzednich

badaniach minimalna przepływność w przypadku tych samych fragmentów muzyki klasycznej, przy której uzyskano MOS = 4, w znormalizowanych warunkach odsłuchowych [11] wyniosła 96 kb/s. Różnica w wynikach mogła być spowodowana wymuszonym sposobem wcześniejszych pomiarów, które wymagały od słuchaczy dokonania oceny w ściśle określonym czasie. Prezentowane badania cechowało swobodne podejście do zadania, choćby z tego względu, że słuchacze oceniali przesłany materiał o dogodnej dla siebie porze i z wykorzystaniem sprzętu, do którego są przyzwyczajeni. Można więc zaryzykować stwierdzenie, że na podstawie uzyskanych obecnie wyników jedynie kodowanie metodą HE-AAC z prędkością bitową 96 kb/s i większą może być wykorzystane do kompresji i transmisji tego gatunku muzyki.

Uzyskane wyniki potwierdzają wprawdzie wpływ technik kodowania na ocenę jakości brzmienia muzyki z każdego z badanych gatunków [12], ale zastosowanie większej przepływności nie zawsze wpływa na uzyskanie znacząco wyższej oceny barwy czy ogólnej jakości dźwięku. Przepływnością optymalną dla każdego badanego gatunku jest zatem 128 kb/s, ponieważ w zdecydowanej większości przypadków daje zadowalające rezultaty.

Bibliografia

[1] Bosi M., Goldberg R., *Introduction to Digital Audio Coding and Standards,* Springer, 2002.

[2] Brachmański S., Kin M., Zemankiewicz P., *Subjective assessment of the Speech Signal Quality Broadcasted by Local Digital Radio in Selected Locations in Wroclaw under Studio and Home Conditions,* „International Journal of Electronics and Telecommunication" 2022, Vol. 68, No. 4 (w druku).

[3] Brandenburg K., *Mp3 and AAC explained,* AES 17th International Conference on High Quality Audio Coding, Florence, Italy, 1999.

[4] Fastl H., Zwicker E., *Psychoacoustics – Facts and Models,* 3rd Edition, Springer, Berlin 2007.

[5] Haverkamp M., *The Role of The Iconicity of Sound within Multisensory Environment,* „Vibration in Physical Systems" 2022, Vol. 33, No. 1, s. 1–9.

[6] Hoeg W., *Digital Audio Broadcasting: Principles and Applications of DAB, DAB+ and DMB*, John Wiley & Sons, 2009.

[7] ITU-T P.800 *Methods for objective and subjective assessment of quality,* 1996.

[8] Jekoschu U., *Assigning Meaning To Sound,* w: Communication Acoustics, J. Blauert (ed.), Springer, Berlin 2005, s. 193–221.

[9] Łętowski T., *Słuchowa ocena sygnałów i urządzeń,* Warszawa 1984.

[10] Meltzer S., Moser G., MPEG-4 HE-AAC v2 – *audio coding for today's digital media world*, „EBU technical Review" 2006, No. 305, s. 37–38; https://tech.ebu.ch/docs/techreview/trev_305-moser.pdf [dostęp: 14.08.2023].

[11] Prygoń S., Kin M., *Ocena wybranych atrybutów sceny dźwiękowej sygnałów poddanych różnym rodzajom kompresji,* w: Materiały XVII Międzynarodowego Sympozjum Inzynierii i Reżyserii Dźwięku, Warszawa 2017.

[12] Ronan M., Ward N., Sazdov R., Lee H., *The Perception of Hyper-compression by Mastering Engineers,* „Journal of the Audio Engineering Society" 2017, Vol. 65, No. 7/8, s. 613–621.

[13] Sayood K., *Introduction to Data Compression,* 4th Edition, Elsevier, Waltham 2021.
[14] Steinmetz CH. J., Bryan N. J., Reiss J. D., *Style Transfer of Audio Effects with the Differentiable Signal Processing,* „Journal of AES" 2022, Vol. 70, No. 9, s. 708–721.
[15] Wendl M., LeeH., *The Effect of Dynamic Range Compression on Loudness and Quality Perception in Relation to Crest Factor,* 136th AES Convention, Preprint 9021, 2014.
[16] Wilmering T., Moffat D., Milo A., Sandler M.B., *A History of audio effects,* „Applied Science" 2020, Vol. 10, No. 3, s. 791.
[17] Żernicki T., *Kompresja cyfrowych sygnałów fonicznych z* łącznym *wykorzystaniem rozszerzania widma i modelowania*, rozprawa doktorska, Politechnika Poznańska, Poznań 2010.
[18] Żórawski J., *O budowie formy architektonicznej,* w: Wybór pism estetycznych, Universitas, Kraków 2008.

Słowa kluczowe: kodowanie, ocena jakości, badania subiektywne.

Badanie wpływu wybranych technik kodowania na jakość dźwięku w nagraniach z różnych gatunków muzyki

Celem autorów niniejszego rozdziału było zbadanie, jak rodzaj kodowania i przepływność wpływają na subiektywną ocenę jakości dźwięku w różnych gatunkach muzycznych. Badania przeprowadzono z wykorzystaniem metod kodowania AAC, HE-AAC i MP3, przy przepływności 64 i 128 kb/s. Zrealizowano eksperyment, w którym słuchacze oceniali ogólną jakość dźwięku oraz dwa atrybuty wrażeń słuchowych: barwę i przestrzenność. Badanie przeprowadzono metodą ACR (ang. *Absolute Category Rating*). Wzięło w nim udział 21 osób. Wyniki eksperymentu potwierdzają wpływ technik kodowania na jakość dźwięku utworu muzycznego z każdego badanego gatunku. Ponadto z badania można wnioskować, że techniki AAC i HE-AAC umożliwiają uzyskanie zadowalającej jakości próbki dźwięku przy niższych przepływnościach niż kodowanie MP3. Zastosowanie wyższych przepływności w przypadku niektórych gatunków muzycznych nie zawsze skutkuje lepszą barwą, przestrzennością czy ogólną jakością dźwięku.

Research on the influence of selected coding techniques on the quality assessment of various music genres

The main aim of this chapter is to examine how the type of coding and bit rate affect subjective quality assessment of different music genres. The study was conducted for AAC, HE-AAC and MP3 encoding methods, at bit rates of 64 and 128 kbps. A listening experiment was conducted in which listeners evaluated overall sound quality and two attributes of the listening experience: timbre and spatiality. The study was performed using the Absolute Category Rating method (ACR), and 21 subjects participated in the study. The results of the experiment confirm the influence of coding techniques on the evaluation of the sound quality of a piece of music for each genre studied. In addition, it can be deduced from the study that AAC and HE-AAC techniques make it possible to obtain satisfactory sound sample quality at lower bit rates than MP3 coding. The use of higher bit rates for some music genres does not always result in better timbre, spatiality or overall sound quality.

3. Analiza metodologii algorytmów stosowanych w strojeniu systemów nagłaśniania

Łukasz Burek, Bartłomiej Kruk

Politechnika Wrocławska,
Wydział Elektroniki, Fotoniki i Mikrosystemów,
wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław

3.1. Wprowadzenie

Systemy nagłaśniania projektowane są w zależności od ich przeznaczenia oraz odpowiednio do miejsc, w których będą wykorzystywane. Pierwszym, kluczowym krokiem jest zdefiniowanie przeznaczenia tworzonego systemu, podobnie jak zdefiniowanie oczekiwań odbiorcy, ponieważ pozwala to na zawężenie listy urządzeń, które w danym przypadku będą odpowiednie. Najbardziej ogólnie systemy dzielone są na służące do przekazywania informacji (mowy) oraz na służące do przekazywania muzyki [1].

Następnym krokiem jest poznanie pomieszczenia docelowego – zbadanie jego parametrów fizycznych takich jak kształt, wymiary, jak również rodzaj powierzchni ścian, podłogi i sufitu. Istotne jest także wyposażenie, które będzie się znajdowało w pomieszczeniu. Informacje te pozwalają na przeprowadzenie wstępnych symulacji lub wyliczeń parametrów akustycznych, m.in. RT60, STI oraz, w dalszych etapach, C80. Istotne jest również założenie poziomu hałasu i jego charakterystyki częstotliwościowej. Jeśli przestrzeń, w której ma działać projektowany system, już istnieje, konieczne jest wykonanie jej pomiaru. Zadaniem projektanta, po zapoznaniu się z podstawowymi informacjami, jest przystąpienie do doboru odpowiednich urządzeń głośnikowych, wzmacniaczy mocy, procesorów i pozostałych elementów toru elektroakustycznego. Rozmieszczenie urządzeń głośnikowych powinno zostać wykonane zgodnie z wynikami symulacji, norm, zaleceń, jeśli są one dostępne. Końcowym etapem, tj. po wykonaniu instalacji urządzeń, jest sprawdzenie zgodności instalacji pod względem akustycznym i fizycznym. Na podstawie wyników należy wprowadzić odpowiednie korekcje w celu uzyskania zakładanej jakości systemu [2].

Jedną z procedur, która wykonywana jest w opisanym powyżej etapie pracy, jest strojenie systemu nagłaśnia polegające na dopasowaniu systemu do pomieszczenia odbiorczego. Do zakresu działań może zostać zaliczona korekcja charakterystyki częstotliwościowej, jak również aplikacja opóźnienia wybranych sygnałów. Ponadto istotne jest wyrównanie poziomów ciśnienia akustycznego emitowanego przez zestawy głośnikowe względem siebie oraz wartości ciśnienia akustycznego uzyskiwanego w zadanym punkcie przy zakładanym poziomie wejściowym wzmacniacza mocy. Do głównych celów kalibracji mogą zostać zaliczone: minimalizacja nierównomierności amplitudy w spektrum częstotliwościowym, maksymalizacja koherencji dzięki jak najlepszemu dostosowaniu dźwięku bezpośredniego do dźwięku odbitego, uzyskanie dobrej zrozumiałości mowy, jak i osiągniecie pełnej kontroli nad obrazem dźwiękowym.

Wszystkie wymienione operacje wprowadzane są zazwyczaj za pomocą procesorów sygnałowych. Znajdują się one we wzmacniaczach mocy, dedykowanych procesorach sygnałowych, procesorach wbudowanych w amplitunerze (ang. *audio video reciver* – AVR), a także bezpośrednio w komputerach osobistych będących źródłem sygnału.

3.2. Metody strojenia systemów nagłaśniania

Strojenie systemów nagłaśniania odbywa się z wykorzystywaniem algorytmów, reguł i przyrządów. Na potrzeby niniejszego rozdziału dokonano podziału dostępnych metod i technologii ze względu na stopień zaangażowania inżyniera w procedurę kalibracji. Opisane strojenie to strojenie manualne, automatyczne i hybrydowe.

Podejście manualne

Inżynier, który decyduje się na wykonanie kalibracji w pełni manualnie, ma obowiązek dobrać odpowiednią aparaturę pomiarową i korygującą system elektroakustyczny. Jest odpowiedzialny za dokładne wykonanie pomiarów pomieszczenia, które w późniejszym etapie będą decydować o jakości systemu.

Podstawą każdej kalibracji jest sprawdzenie właściwości akustycznych pomieszczenia, jak również parametrów elektroakustycznych samego systemu. W tym celu konieczne jest odpowiednie dobranie wykorzystywanych narzędzi, tj. aparatury do pomiarów fizycznych, akustycznych i elektroakustycznych. Do pierwszej z trzech wymienionych kategorii zaliczają się m.in. dalmierz laserowy, kątomierz laserowy, termometr i higrometr. Dwa pierwsze urządzenia pozwalają na weryfikację kątów i odległości zestawów głośnikowych od wyznaczonego punktu odsłuchowego, natomiast pozostałe na określenie warunków atmosferycznych (temperatury i wilgotności) w celu określenia dokładnej prędkości dźwięku w danym pomieszczeniu. Wykonanie pomiarów fizycznych wpływa na dokładność dalszych działań. Kluczowym narzędziem w pomiarach akustycznych jest mikrofon, którego charakterystyka częstotliwościowa o dookólnej kierunkowości

powinna być maksymalnie płaska, a całkowite zniekształcenia harmoniczne i szumy własne jak najmniejsze. Duże znaczenie ma także stabilność pracy w funkcji temperatury i czasu [3]. Dobór mikrofonu powinien być determinowany tym, do czego miałby być przeznaczony, i warunkami, w jakich miałby pracować. Kolejnym narzędziem jest miernik poziomu dźwięku z możliwością aplikowania krzywych korekcyjnych A i C, a także znormalizowanych stałych czasowych charakterystyk fast i slow. Przydatna jest również funkcja analizy widma częstotliwościowego. W celu uzyskania jak najbardziej wiarygodnych wyników analizator powinien wyświetlać dane o rozdzielczości 24–48 punktów na oktawę [2]. Dodatkowo powinien wyznaczać funkcję transmitancji operatorowej (*ang. transfer function* –TF) na podstawie sygnału referencyjnego podawanego z generatora i sygnału badanego.

Do przydatnych funkcji należy badanie fazy lub koherencji sygnałów [2]. W celu zmniejszenia liczby urządzeń wykorzystywanych podczas przeprowadzania analiz akustycznych często używane są programy komputerowe, o ile spełniają wcześniej wymienione wymagania, np. Rational Acoustics Smaart v8 lub AMFG SysTune. Przy założeniu pracy wyłącznie z oprogramowaniem komputerowym wymagany jest interfejs audio o liczbie wejść i wyjść uzależnionej od wybranej techniki. Opcjonalny może być także zewnętrzy generator sygnałowy. Przy doborze sprzętu pomiarowego należy pamiętać, że jakość pomiaru będzie podyktowana parametrami najsłabszego z użytych urządzeń.

Następną grupą w torze są procesory sygnałowe (ang. digital signal processing – DSP), które umożliwiają wprowadzenie wszystkich wymaganych korekcji w systemie elektroakustycznym. Kluczowe jest, aby pozwalały na korekcję opóźnień na każdym z kanałów. W małych przestrzeniach są to opóźnienia rzędu od kilku do kilkudziesięciu ms, w dużych (takich jak sale koncertowe lub otwarta przestrzeń) kilku s [3]. Korekta częstotliwości realizowana jest najczęściej za pomocą filtrów parametrycznych lub tercjowych korektorów graficznych. Kolejnym przydatnym elementem procesora jest pomijanie filtrów (ang. *bypass*), dzięki czemu możliwe jest porównanie wprowadzonych zmian z poprzednim stanem systemu. Przykładem takich opisanych urządzeń mogą być dedykowany procesor sygnałowy Peavey Nion, Avid MTRX, który jest wyposażony w odpowiednią kartę rozszerzeń, lub zestawy głośnikowe z wbudowywanym cyfrowym procesorem (np. Neumann KH 80 DSP).

Podejście automatyczne i hybrydowe

Do bardzo popularnych rozwiązań w sprzęcie użytku codziennego może zostać zaliczona automatyczna kalibracja systemu nagłośnienia. Podobnie jak w podejściu manualnym wymagany jest mikrofon pomiarowy, oprogramowanie analizujące wyniki i procesor sygnałowy wykonujący odpowiednie korekcje. Ponadto często do urządzenia powszechnego użytku dodawany jest przez producenta mikrofon pomiarowy. Procesor DSP zazwyczaj występuje jako jeden z elementów urządzenia. Przykładem mogą być AVR-y wyposażone w specjalne algorytmy służące do wykonania kalibracji. W zależności od producenta używane są różne systemy:

- Denon i Marantz – system Audyssey,
- Onkyo – AcuuEQ,
- Pioneer – Multi-Channel Acoustic Calibration (MCACC),
- Sony – Digital Cinema Auto Calibration (DCAC) EX,
- Yamaha – Prometric Room Optimizer (YAPO),
- ARCAM i NAD – Dirac Live correction,
- Trinnov – Trinnov Optimizer.

Większość automatycznych systemów nie pozwala użytkownikowi na wprowadzanie zaawansowanych korekt do zrealizowanej automatycznie kalibracji. Zakres, w którym możliwe są korekty manualne, jest najczęściej niewielki. W przypadku niektórych droższych modeli urządzeń domowych lub profesjonalnych funkcjonalności te są rozszerzone, dzięki czemu bardziej doświadczony użytkownik może przeprowadzić dokładną analizę, podobną do metody manualnej. W związku z tym na potrzeby prowadzonych badań autorzy nazwali to podejście hybrydowym. Dużą zaletą tego rozwiązania jest możliwość rozpoczęcia kalibrację od przeanalizowania wyników z automatycznego strojenia. W zależności od jakości algorytmu uzyskane wyniki mogą być jedynie wskazówką w procesie dalszej manualnej kalibracji. Do profesjonalnych rozwiązań mogą zostać zaliczone m.in. AVR-y z serii Trinnov Altitude i dedykowany procesor Trinnov ST2 Pro. Innymi przykładami są procesory sygnałowe znajdujące się bezpośrednio w zestawach głośnikowych. Takie rozwiązanie wykorzystuje firma Genelec w systemie kalibrującym GLM. Istnieją ponadto rozwiązania, w których przetwarzanie dźwięku odbywa się za pomocą specjalnej aplikacji instalowanej na komputerze osobistym. Firmy sprzedające oprogramowanie tego typu zazwyczaj mają w ofercie dedykowane mikrofony pomiarowe. Reprezentantami takiego rozwiązania są m.in. Sonarworks SoundID Reference oraz IK Multimedia ARC System 3.

3.3. Wybrane algorytmy

Do badań zostały wybrane różne systemy. Jeden z nich jest strojony w sposób manualny, a dwa systemy są strojone automatycznie. Pierwszy z systemów strojenia automatycznego był przeznaczony do powszechnego użytku, natomiast drugi – do profesjonalnego.

Przy podejściu manualnym wybrano programowalny procesor sygnałowy Peavey Nion n6 z uwagi na możliwość stworzenia oprogramowania od podstaw. Ta funkcjonalność pozwala nie tylko na kreację spersonalizowanego łańcucha przetwarzania sygnału pod kątem kalibracji, ale także umożliwia opracowanie systemu routingowego (łączącego wszystkie wybrane urządzenia w jeden duży system). Dodatkową możliwością jest przygotowanie prostego i czytelnego interfejsu użytkownika, który będzie konieczny do przeprowadzenia badań subiektywnych [4]. Podczas analizy sygnału zostało wykorzystane oprogramowanie komputerowe Smaart v8 firmy Rational Acoustics [5], posiadające funkcję RTA, Transfer Function, wbudowany generator sygnałów, a także wiele innych,

które wpływają na jego przewagę. Jako główny interfejs audio wybrano RME Babyface Pro FS [6] , tj. wysokiej klasy interfejs, którego przetwornik analogowo-cyfrowy ma wartość THD+N poniżej 0,00035 % oraz zakres wzmocnienia sygnału od –11 do +65 dB. Użyteczny zakres częstotliwości wynosi od 7 Hz do 45,8 kHz przy częstotliwości próbkowania 96 kHz i odchyleniach od liniowości równych –0,5 dB. Mikrofon pomiarowy, który został użyty do pomiarów to Earthworks M30 [7]. Według noty katalogowej jego zakres użytecznych częstotliwości przy spadku –3 dB wynosi od 3 Hz do 30 kHz, a nierównomierności wynoszą ±1 dB. Szumy własne mikrofonu są na poziomie 20 dB SPL, a średnia czułość na poziomie 34 mV/Pa. Do przeprowadzenia dokładnych pomiarów ciśnienia akustycznego konieczne było również przestrzeganie procedury kalibracyjnej mikrofonu, aby możliwe było uzyskanie dokładnej wartości czułości. W tym celu użyto pistofonu klasy 1 [8] Larson Davis CAL200 [9], generującego sygnał o częstotliwości 1 kHz i poziomie ciśnienia 94 lub 114 dB.

Podstawą drugiego z wybranych algorytmów jest podejście automatyczne. Kalibracja systemu nagłośnienia AVR Marantz SR6008 [10] została przeprowadzona z wykorzystaniem urządzenia bazującego na Audyssey MultiEQ XT. Jej realizacja odbywa się za pomocą dołączonego do zestawu mikrofonu pomiarowego. Dokładna specyfikacja urządzenia nie jest podana przez producenta. Spośród potrzebnych informacji w nocie katalogowej AVR-a podano jedynie użyteczny zakres częstotliwości wejścia liniowego przetwornika A/C od 10 Hz do 100 kHz przy nierównomiernościach równych +1 dB, –3 dB. Producent zastrzega jednak, że pomiary zostały wykonane w trybie DIRECT, który pozwala na odtwarzanie dźwięku w wysokiej jakości. Dodatkowo tryb blokuje rożnego rodzaju ustawienia korekcji częstotliwościowej i dynamicznej głośności w urządzeniu [11].

Ostatnim z wybranych algorytmów jest system strojenia wbudowany w zestawy głośnikowe GLM firmy Genelec, który pozwala na przeprowadzenie strojenia automatycznego w sposób hybrydowy. Urządzeniami wybranymi do systemu są aktywne dwudrożne monitory pola bliskiego Genelec 8330A [12]. Charakteryzują się użytecznym zakresem częstotliwości od 45 Hz do 23 kHz (przy spadku 6 dB) oraz liniowością charakterystyki w zakresie ±1,5 dB od 58 Hz do 20 kHz. Dodatkowo do każdego z poszczególnych głośników, do których dostarczone są sygnały podzielone za pomocą zwrotnicy aktywnej, posiadają dedykowany wzmacniacz. Zestaw głośnikowy może emitować w sposób ciągły poziom ciśnienia 96 dB SPL w odległości 1 m przy generowanym szumie zgodnym z ważeniem IEC. Do przeprowadzania kalibracji i wglądu w ustawienia DSP urządzeń konieczny jest adapter GLM, który jednocześnie spełnia rolę interfejsu audio. Jest on podłączony do komputera osobistego przewodem USB2.0. Użytkownik ma możliwość kontroli wszystkich dostępnych ustawień z poziomu darmowej aplikacji GLMv4. Wszystkie urządzenia głośnikowe składające się na dany system nagłaśniania (w tym przypadku stereo) muszą zostać wcześniej podłączone szeregowo do adapteru GLM za pomocą złączyRJ45. Mikrofon pomiarowy dołączony do zestawu wykorzystuje złącze TRS 3,5 mm. Producent nie udostępnia jednak jego dokładnej noty katalogowej. Analizując jego konstrukcję, można zauważyć jeden przetwornik. Możliwe jest, że

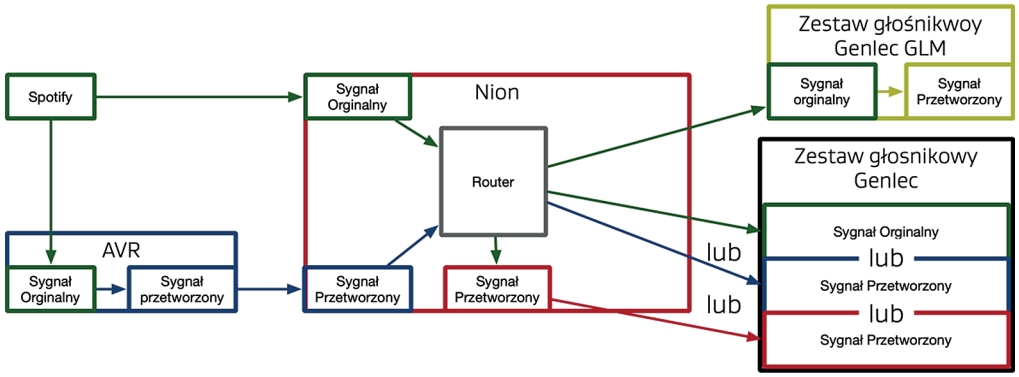
sygnał mikrofonowy jest przesyłany połączeniem symetrycznym. Dodatkowo oprogramowanie GLMv4 po wprowadzeniu numeru seryjnego mikrofonu jest w stanie wczytać charakterystykę korekcyjną mikrofonu i dokładną czułość. Rozwiązanie to umożliwia przeprowadzenie bardzo dokładnych pomiarów.

W badaniu użyto dwóch par wyżej opisanych urządzeń stereo Genelec 8330A: jedna para przyjmuje sygnał oryginalny, który następnie jest korygowany za pośrednictwem wbudowanego systemu GLM, druga – sygnał przetworzony przez algorytm manualny Nion lub algorytm automatyczny Audyssey z AVR.

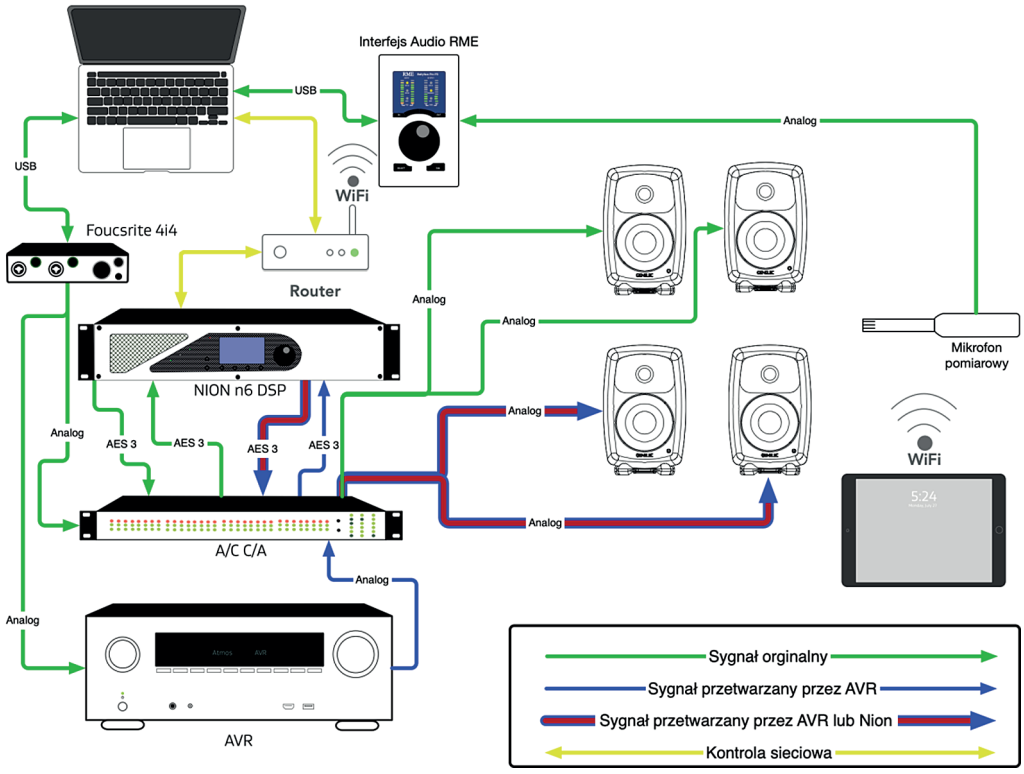
3.4. Stanowisko badawcze

Badanie trzech rożnych strojeń wymaga przygotowania odpowiednio zaawansowanego stanowiska badawczo-pomiarowego. Pomieszczenie badawcze jest wybierane z uwzględnieniem dwóch aspektów: poziomu tła i czasu pogłosu, a wynika to z zamiaru stworzenia jak najkorzystniejszych warunków dla systemów automatycznych. Pomieszczenie ma 6,8 m długości, 4,9 m szerokości i 2,8 m wysokości. Jego budowa polega na zasadzie „pudełka w pudełku". Czas pogłosu w pomieszczeniu wynosi około 0,5 s, a poziom tła jest zgodny z NC35. Równocześnie spełniane są założenia przedstawione m.in. w normie ITU-T BS 1116-3. Ważne podczas projektowania stanowiska jest logiczne połączenie wszystkich urządzeń w jeden system elektroakustyczny, zachowanie wszystkich standardów związanych z transmisją sygnałów elektroakustycznych, a także zapewnienie jak najlepszych warunków odsłuchowych w pomieszczeniu odbiorczym. Dodatkowo istotne jest opracowanie możliwie wygodnego i przejrzystego sposobu symultanicznego przełączania się między strojeniami, a powodem jest przede wszystkim konieczność przeprowadzenia subiektywnych testów odsłuchowych, które wymagają płynnej zmiany strojenia. Zapewnienie takiej funkcjonalności pozwala także na łatwiejsze przełączanie strojeń podczas badań obiektywnych.

Podstawowym elementem stanowiska jest źródło sygnału – komputer osobisty podłączony do interfejsu dźwiękowego Focusrite 4i4. Ma on spowodować, że (kolejno) na przetwornik A/C, C/A Lynx Aurrora 8 i na wejście AVR Marantz SR6008 dotrze taka sama para sygnałów. Sygnał po przetworzeniu przez procesor Audyssey przepływa na przetwornik A/C za pośrednictwem wyjść liniowych pre-out. Następnie, po przetworzeniu do standardu AES3, dociera do procesora Nion, a wewnątrz samego procesora – do router-a audio. W zależności od ustawień router-a sygnał oryginalny może trafiać na zestaw głośnikowy z aktywną funkcją GLM lub na blok odpowiedzialny za strojenie manualne. Po przejściu przez blok sygnał docierał do urządzeń głośnikowych z wyłączoną funkcją GLM. Przy innych ustawieniach router-a mógł zostać wysłany jedynie sygnał oryginalny lub przetworzony przez algorytm Audyssey. Uproszczony schemat opisanych powyżej połączeń przedstawiono na rys. 3.1, a schemat szczegółowy (ze wszystkimi fizycznymi połączeniami) na rys. 3.2.



Rys. 3.1. Uproszczony schemat połączeń stanowiska badawczego



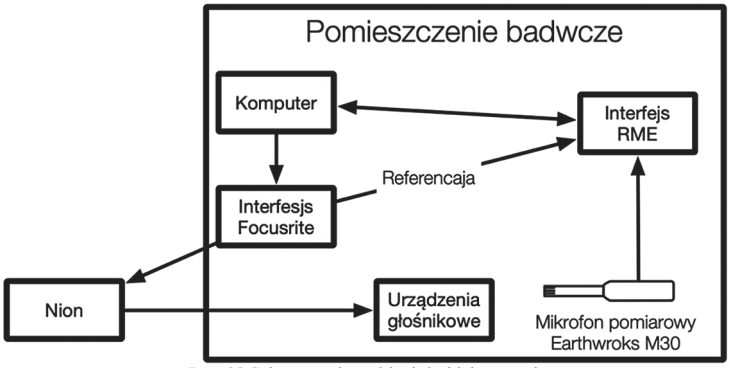
Rys. 3.2. Szczegółowy schemat stanowiska badawczego z uwzględnieniem fizycznych połączeń

Procesor sygnałowy Peavey Nion wraz z przetwornikiem Aurrora 8 został umieszczony w pomieszczeniu odizolowanym akustycznie od pomieszczenia badawczego w celu eliminacji hasła generowanego przez nie hałasu w trakcie pracy. Wspomniane miejsca są połączone ze sobą przyłączami naściennymi, co umożliwia zbudowanie powyższej konfiguracji.

Urządzenia głośnikowe są rozmieszczone w pomieszczeniu zgodnie z założeniami ITU [15], według których linie poprowadzone wzdłuż centrum akustycznego urządzeń głośnikowych powinny przecinać się w punkcie odsłuchowym pod kątem 60°.

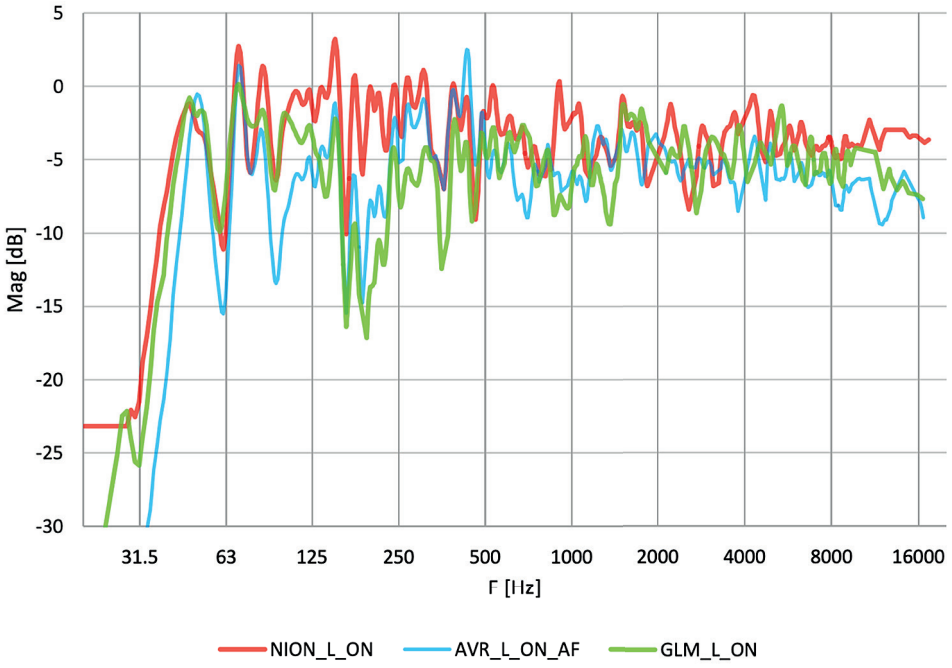
Wyniki i ich dyskusja

Uzyskane rezultaty wszytkich kalibracji zostały poddane dokładnej analizie. Podczas przeprowadzania badań użyto zestawu narzędzi zbliżonego do zestawu użytego podczas strojenia manulanego. Wszystkie analizy przeprowadza się z wykorzystaniem programu Smaart v8 dla wcześniej skorygowanego mikrofonu pomiarowego Earthworks i sygnału odniesienia. Źródłem sygnału jest generator wbudowany w oprogramowanie komputerowe. Zmiana punktu emisji sygnału jest konieczna w celu przetwarzania go przez procesor Audyssey, na który wysyła się oryginał bezpośrednio z interfejsu Focusrite. Uproszczony schemat blokowy przedstawiono na rys. 3.3.



Rys. 3.3. Schemat blokowy stanowiska badawczego do badań obiektywnych

Pierwszym krokiem, podobnie jak w przypadku strojenia manualnego, jest analiza charakterystyk częstotliwościowych. Z rysunku 3.4 wynika, że algorytm GLM zastosował najmniejsze korekcje spośród wszystkich pozostałych. Wszystkie strojenia zredukowały podbicia przy częstotliwości 50, 70 i 80 Hz w różnym stopniu. Kalibracja AVR wykonała dostrojenia za pomocą tylko jednego filtru o małej wartości dobroci. Skutkowało to jednocześnie pogłębieniem minimum w przypadku częstotliwości w granicach



Rys. 3.4. Porównanie wszystkich aktywnych strojeń w przypadku kanału lewego

60 Hz. W pozostałych przypadkach użyta została większa liczba filtrów o dużej dobroci w celu uniknięcia zwiększenia tłumienia dla tego pasma. Strojenie wykonane przez system Audyssey różniło się od pozostałych dużą ilością zastosowanych filtrów w zakresie średnich i dużych częstotliwości. W przypadku strojenia GLM może nastąpić delikatne wzmocnienie częstotliwości o charakterze filtru półkowego, jednakże podczas analizy danych po zakończeniu kalibracji nie zauważono zastosowania żadnej implementacji filtrów. Może to świadczyć o zmianie wzmocnienia dla wzmacniacza głośnika wysokotonowego. W zakresie od 4 kHz widoczne są różnice nachyleń liniowości charakterystyk w przypadku każdego ze strojeń.

Po analizie badań obiektywnych nasuwają się następujące wnioski:

- Różnica poziomów między kanałami w przypadku każdego z algorytmów była skorygowana poprawnie. Jej maksymalna wartość 0,2 dB została odnotowana, gdy strojenie zostało wykonane za pośrednictwem AVR, a pomiar za pomocą oprogramowania Smaart z wczytaną korekcją czułości mikrofonu.
- Opóźnienia między urządzeniami głośnikowymi również zostały poprawnie skorygowane przez każdy z algorytmów.

Na podstawie powyższych wyników z analiz zaobserwowano, że najbardziej płaska charakterystyka częstotliwościowa została uzyskana podczas strojenia manualnego.

3.5. Badania subiektywne

Celem badań było sprawdzenie percepcji wrażeń wynikających ze strojenia przez grupę badawczą. Z uwagi na chęć przeprowadzenia szczegółowych badań zadanych parametrów do badań została wybrana jedynie doświadczona grupa ekspercka składająca się z siedmiu osób w wieku 25–35 lat. Wszyscy słuchacze ukończyli studia bezpośrednio związane z akustyką, są związani z branżą dźwiękową (są aktywni zawodowo), gdzie często biorą udział w krytycznych testach odsłuchowych.

Metodyka badań

Z uwagi na brak istniejącej metody badań wprowadza się nową metodę, której podstawą są metody opisane w normach ITU-R-BS.1116-3_2015_02 [12] i EBU-tech.3286 [13]. Główną przeszkodą w istniejących zaleceniach była obecność materiału referen-

Tabela 3.1. Definicja parametrów

Parametr	Definicja	Odniesienie do skali ocen
Niskie	wrażenie balansu niskich częstotliwości	defektem jest zbyt duże podbicie lub tłumienie w danym pasmie częstotliwości
Średnie	wrażenie balansu średnich częstotliwości	
Wysokie	wrażenie balansu wysokich częstotliwości	
Wrażenie stereo	percepcja stereo: czy balans między kanałami jest odpowiedni, a obraz dźwiękowy prawidłwo reprodukowany?	
Przejrzystość	łatwość rozróżniania instrumentów w utworze	defektem jest łatwość rozróżniania instrumentów
Pierwsze wrażenie	pierwsze wrażenie odnośnie do danego strojenia	

Tabela 3.2. Skala ocen parametrów subiektywnych

Ocena	Jakość	Wrażenie
1	zła	bardzo denerwujące defekty
2	słaba	dużo denerwujących defektów
3	dostatecznie dobra	kilka denerwujących defektów
4	dobra	kilka delikatnie denerwujących defektów
5	bardzo dobra	kilka słyszalnych, ale niedenerwujących defektów
6	znakomita	brak denerwujących defektów

cyjnego. Ze względu na charakter badań nie było to możliwe. Porównanymi obiektami nie były próbki dźwiękowe, a różnego rodzaju wrażenia związane z poprawnym zestrojeniem systemu odsłuchowego (tabela 3.1).

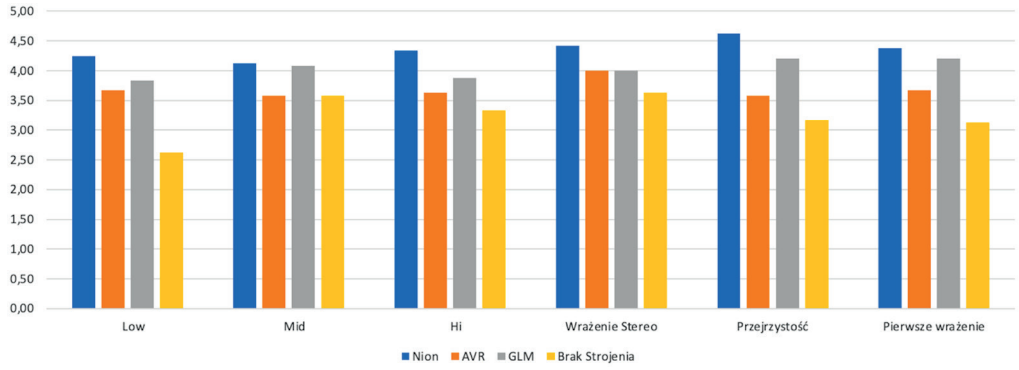
Każdy ze słuchaczy miał za zadanie odsłuchać dwa utwory narzucone przez badaczy i dwa wybrane przez siebie. W trakcie odsłuchu każda z osób mogła dowolnie odtwarzać fragmenty utworów, a także zmieniać je odpowiednio w stosunku do preferencji. Do oceny zostały podane cztery możliwości strojenia, które nie były znane słuchaczowi. Zostały one oznaczone literami od A do D. Słuchacz w sposób symultaniczny mógł przełączać się między badanymi strojeniami. Przypisane litery oznaczały kolejno:

- wykonanie manualne za pomocą DSP Nion (A);
- strojenia wykonane automatycznie za pomocą systemu Audyssey w AVR (B);
- strojenia wykonane automatycznie za pomocą systemu GLM w zestawach głośnikowych Genelec (C);
- brak strojenia częstotliwościowego (D).

Sygnał ze wszystkich strojeń został wyrównany z użyciem szumu różowego, tak aby przy zadanym średnim poziomie źródła (portal Spotify) każde strojenie wytwarzało równy poziom ciśnienia akustycznego na zestaw głośnikowy. Średni poziom generowany przez źródło równy był poziomowi ciśnienia L_{CS} = 79 dB SPL. Wszystkie testy zostały przeprowadzone dla jednakowych warunków fizycznych i akustycznych. Do przeprowadzenia badań konieczne było przegotowanie zestawu parametrów, skali ocen, zestawu odpowiednich definicji i opisanie sposobu przeprowadzania badań (tabela 3.1).

Wyniki i ich dyskusja

Na podstawie wykresu na rys. 3.5 można stwierdzić, że najlepiej ocenianym algorytmem strojenia systemów na podstawie zadanych parametrów oceny jest kalibracja wykonana manualnie za pomocą systemu DSP Nion. Średnia ocen w przypadku tych



Rys. 3.5. Uśrednione wyniki ocen badanych utworów, wystawionych przez uczestników grupy eksperckiej

samych prób nieznacznie różni się od wyników uśrednionych wszystkich prób. Istotna jest także obserwacja wpływu średniej ocen utworów wybieranych przez osoby badane na ogólny wynik. W przypadku każdego z parametrów widoczny jest dodatni wpływ ocen na strojenie wykonane za pomocą systemu GLM. Pozostałe strojenia cechuje spadek średniej ocen wszystkich parametrów, poza pierwszym wrażeniem oraz odbiorem małych i średnich częstotliwości w przypadku braku strojenia częstotliwościowego.

Na poziomie istotności $\alpha = 0{,}05$ zweryfikowano hipotezę, że oceny wszystkich członków grupy eksperckiej są takie same (tabela 3.3). W tym celu autorzy niniejszego rozdziału przeprowadzili test analizy wariancji dla wielu średnich [14].

Tabela 3.3. Skala ocen parametrów subiektywnych

	Nion	AVR	GLM	Brak strojenia
Paramater	ocena			
Niska	4,25	3,67	3,83	2,63
Średnia	4,13	3,58	4,08	3,58
Wysoka	4,33	3,63	3,88	3,33
Wrażenie stereo	4,42	4,00	4,00	3,63
Przejrzystość	4,63	3,58	4,21	3,17
Pierwsze wrażenie	4,38	3,67	4,21	3,13

Wyniki testu istotniści przeprowadzone za pomocą F Snedecora wynoszą odpowiednio 0,294, 0,239, 0,367 i 1,174. Wszytkie wymienione wartości są mniejsze od wartości krytycznej $F_\alpha = 2{,}492$. Ponieważ wartości F leżą poza obszartek krytycznym, nie ma podstaw do odrzucenia sprwdzanej hipotezy. Można zatem stwierdzić, że oceny słuchaczy były jednorodne. Różnice otrzymane w przypadku poszczególnych wartości parametrów są w różnych systemów istotne statystycznie: we wszystkich przypadkach uzyskano wartości F Snedecora większe od $F_\alpha = 2{,}682$.

3.6. Podsumowanie

Na podstawie otrzymanych wyników można stwierdzić, że średnio najlepiej ocenianym algorytmem strojenia systemów na podstawie zadanych parametrów oceny jest kalibracja wykonana manualnie za pomocą systemu DSP Nion. Automatyczna kalibracja, nawet w przypadku starszych algorytmów, przyniosła lepsze wrażenia słuchowe niż brak strojenia systemu. Dlatego może być ona dobrym punktem wyjścia dla inżyniera wykonującego kalibrację w sposób zaawansowany. Profesjonalne rozwiązania takie jak algorytm GLM przynoszą rezultaty niewiele gorsze niż strojenie manualne przy wyma-

ganej mniejszej ilości sprzętu. Dużą zaletą takiego rozwiązania jest fakt, że urządzenia te są w stanie jeszcze lepiej zarządzać ustawieniami wzmacniaczy mocy oraz częstotliwości podziału, co w przypadku strojenia aktywnych zestawów głośnikowych w innych rozwiązaniach nie jest możliwe. Za bardzo subiektywne należy uznać także percypowanie przestrzeni stereo, która w przypadku strojenia manualnego okazała się bardzo szeroka w porównaniu do strojenia automatycznego wykonanego z wykorzystaniem urządzenia Genelec. Było to spowodowane użyciem zbyt różnych korekcji na lewym i prawym kanale.

W przyszłości, w podobnych badaniach subiektywnych wartościowe mogłoby okazać się rozszerzenie ich o dodatkowe pytania dotyczące zakresu percypowanych atrybutów wrażenia słuchowego. Słuchacze byliby proszeni o uszeregowanie ocenianych strojeń od najlepszego do najgorszego. Dodatkowo wszelkie komentarze, odczucia mogłyby być rejestrowane, a następnie spisywane przez badaczy. Taki sposób zapewniłby podobną swobodę wypowiedzi, jaka miała miejsce po zakończeniu prezentowanych badań, z tą różnicą, że możliwe byłoby lepsze udokumentowanie dodatkowych spostrzeżeń.

Bibliografia

[1] Ballou G., *Handbook for sound engineers*, Focal Press, New York 2015.
[2] McCarthy B., *Sound Systems: Design and Optimization*, Focal Press, New York 2016.
[3] Dobrucki A., *Pomiary w akustyce.*
[4] PEC, „MediaMatrix NION n6 Spec Sheet" [online]; https://peaveycommercialaudio.com/wp-content/uploads/2019/03/MediaMatrix-NION-n6-Spec-Sheet.pdf [dostęp: 2021].
[5] Rational Acoustics, „Smaart v8 User Guide", Woodstock, CT 06281 USA, 2018.
[6] RME, „Specyfikacja Techniczna Babyface Pro FS" [online]; https://www.rme-audio.de/babyface-pro-fs.html [dostęp :2021].
[7] Earthworks Audio, „Specyfikacja technicza Earthworks M30" [online]. Available: https://earthworksaudio.com/wp-content/uploads/2020/11/Earthworks-Audio-M30-Data-Sheet-V1.pdf [dostęp: 2021].
[8] IEC, „Electroacoustics – Sound calibrators," 2017.
[9] MTS, „Specyfikacja kaliblatora akustycznego CAL200" [online]; http://www.larsondavis.com/products/calibrators/modelcal200 [dostęp: 2021].
[10] Marantz, „Specyfikacja techniczna AVR Marantz SR6008" [online] https://www.marantz.com/-/media/files/documentmaster/marantzna/us/sr6008_specification_sheet.pdf [dostęp: 2021].
[11] Marantz, „Instukcja obsługi Marntz SR6008" [online]; https://www.marantz.com/-/media/files/documentmaster/marantzna/us/sr6008u_eng_cd-rom_ug_v00.pdf [dostęp: 2021].
[12] ITU, „ITU-T BS 1116-3".
[13] EBU, „EBU – tech.3286", Geneva 1997.
[14] Greń J., *Statystyka matematyczna modele i zadania*, PWN, Warszawa 1974

Słowa kluczowe: strojenie systemów nagłaśniania, systemy elektroakustyczne, DSP.

Analiza metodologii algorytmów stosowanych w strojeniu systemów nagłaśniania

Celem autorów niniejszego rozdziału jest analiza metod strojenia, w których stosowane są algorytmy automatycznej korekcji sygnału i zadane manualne algorytmy. Do realizacji założenia konieczne jest wykonanie analizy dostępnych metod strojenia systemów nagłaśniania, zaprojektowanie stanowiska pomiarowego, wykonanie pomiarów z wykorzystaniem metod subiektywnych i obiektywnych. W badaniu wykorzystywane są systemy nagłośnienia dwukanałowego (stereo), przeznaczone do reprodukcji dźwięku w pomieszczeniach zamkniętych o małej kubaturze.

Analysis of methodologies used in sound system tuning algorithms

The key aspect of research was to compare various methods of sound system tuning with use of subjective and objective approach. Three methods have been chosen: manual tuning with use of external DSP system, automatic with use of professional GLM and consumer Audyssey algorithm. In order to achieve the goal, it was required to design advance electroacoustic system with ability of simultaneous switching between different tunings. Listening tests have been carried out with custom designed method, which have been based on ITU-T BS 1116-3 and EBU – tech.3286. The main obstacle in use of existing recommendations was presence of reference sample in it. Ultimately the best algorithm of sound system tuning is hard to be define. Everything depends of listener preferences.

4. Wybrane aspekty charakterystyk kierunkowości głośników modów rozproszonych

Karol Czesak, Piotr Kleczkowski

Akademia Górniczo-Hutnicza im. Stanisława Staszica,
Wydział Inżynierii Mechanicznej i Robotyki,
al. Adama Mickiewicza 30, 30-059 Kraków

4.1. Wprowadzenie

Głośniki modów rozproszonych (ang. *Distributed Mode Loudspeaker – DML*) przeważnie mają postać utwierdzonej na brzegach, prostokątnej płyty o bokach różniących się długością, wykonanej z materiału podatnego na drgania giętne (rys. 4.1). Do drgań giętnych [1] płyta pobudzana jest za pomocą jednego lub więcej wzbudników elektrodynamicznych zamocowanych w jej tylnej części. Publikowane były modele działania głośników tego typu [2], [3], [4]. Charakterystyka uzyskiwanych drgań jest silnie związana z modami drgań własnych płyty, skąd biorą się obserwowane w pomiarach charakterystyk amplitudowo-częstotliwościowych liczne minima lokalne. Drgania te mają charakter niekoherentny, stąd charakterystyki amplitudowo-częstotliwościowe DML zdejmowane w punktach oddalonych od siebie na półsferze pomiarowej o niewielkie wartości kąta mogą się znacznie od siebie różnić. Ta właściwość utrudnia, a wręcz wyklucza możliwość dokonywania aproksymacji charakterystyk kierunkowości DML na podstawie pomiaru w zaledwie kilku punktach [5], przy zachowaniu podejścia właściwego dla opisu głośników tłokowych. Źródłem rzetelnej informacji na temat promieniowania DML może być wyłącznie pomiar na sferze lub półsferze w gęstej siatce pomiarowej.

W związku z niekoherentnym promieniowaniem wyznaczanie charakterystyki fazowej DML jest bezcelowe, jeżeli prowadzona analiza dotyczy więcej niż jednego kierunku promieniowania. Za charakterystykę częstotliwościową tych przetworników należy uznawać jedynie charakterystykę amplitudową, uzyskaną w wyniku uśredniania charakterystyk amplitudowych mierzonych w wielu punktach na półsferze pomiarowej.

Rys. 4.1. Widok głośnika modów rozproszonych (DML), zainstalowanego w komorze bezechowej

Przywołane właściwości DML przekładają się na ich zachowanie w rzeczywistym pomieszczeniu, gdzie mają miejsce odbicia dźwięku od podłogi, ścian, sufitu i elementów wyposażenia. Pomimo że sprawność DML jest niższa od sprawności przetworników tłokowych, rozkład postrzeganej głośności uzyskiwanej w pomieszczeniu jest bardziej równomierny niż w przypadku zastosowania konwencjonalnych przetworników elektroakustycznych [6]. Uzyskanie zadowalającego poziomu ciśnienia akustycznego wymaga również amplitudy drgań DML mniejszej niż amplituda drgań głośnika tłokowego uzyskiwana w wyniku zastosowania większej powierzchni drgającej. Zjawiska te pozwalają także wnioskować, że występujące w rzeczywistym pomieszczeniu odbicia dźwięku są z perspektywy słuchacza mniej destrukcyjne, jeżeli zostaną w nim zastosowane DML, co ma związek z niekoherentną emisją dźwięku.

4.2. Procedura pomiarowa

Procedura pomiarowa zakładała zdjęcie charakterystyk amplitudowo-częstotliwościowych DML w 325 punktach półsfery, których rozmieszczenie warunkowane było zachowaniem stałej rozdzielczości kątowej pomiaru wynoszącej 10°. Jeden z punktów znajdował się na osi prostopadłej do płaszczyzny głośnika, natomiast pozostałe punkty zostały rozlokowane w grupach liczących po 36 punktów na równoleżnikowych okrę-

gach o wartościach kąta elewacji odpowiadających pełnym dziesiątkom stopni kątowych (10, 20, 30, …). W celu wyeliminowania efektów pola bliskiego zdecydowano się na promień 2,5-metrowy – najdłuższy spośród dostępnych w komorze bezechowej Laboratorium Akustyki Technicznej (LAT) w Katedrze Mechaniki i Wibroakustyki AGH. Precyzyjne zadawanie kąta azymutalnego zapewnił napędzany silnikiem krokowym stolik obrotowy, zaś kąt elewacji zadawany był z wykorzystaniem zrobotyzowanego ramienia, na którym został umieszczony mikrofon pomiarowy. Praca urządzeń kontrolujących geometrię pomiaru nadzorowana była w środowisku LABVIEW za pomocą aplikacji specjalnie napisanej przez personel LAT.

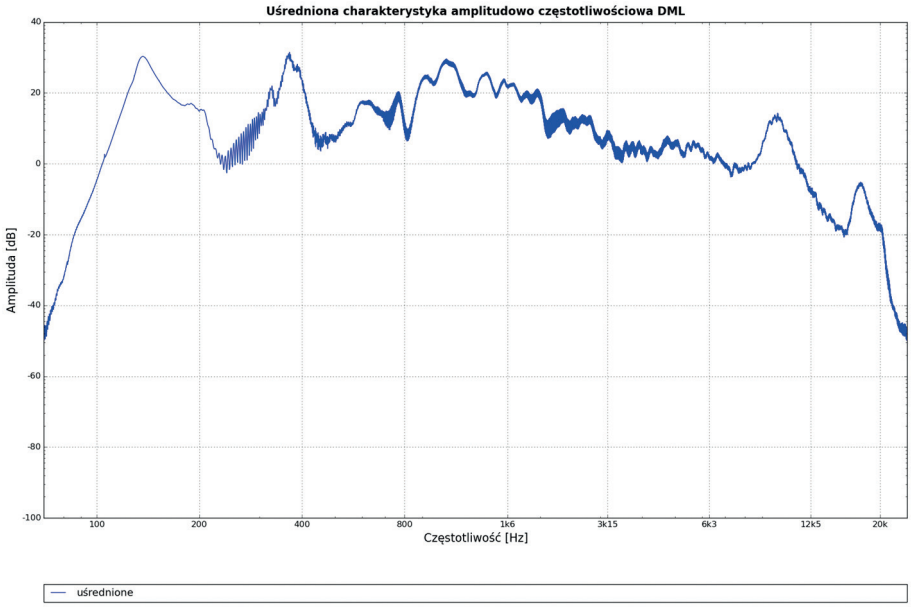
W badaniach wykorzystany został mikrofon pola swobodnego G.R.A.S 46AE, z którego sygnał trafiał do kondycjonera G.R.A.S 12AK, a dalej – przez urządzenie symetryzujące i zapewniające separacje galwaniczną (Di-box) – do przetwornika A/C połączonego z komputerem, gdzie rejestrowany był z częstotliwością próbkowania 96 kS/s i rozdzielczością 24 bitów. Sygnał wymuszający dostarczany był do badanego DML przez wzmacniacz mocy Anthem PVA-7 i miał postać sinusa przestrajanego liniowo w zakresie częstotliwości 70–20 000 Hz, o długim czasie trwania. Sygnał ten zapisany był w pliku wav o częstotliwości próbkowania 96 kS/s i rozdzielczości 24 bity. Długość wektora próbek wynosiła 2^{22} = 4 194 304 próbki. Przy takim doborze parametrów aktywny czas trwania pojedynczego sygnału wymuszającego wynosił 36 s. Pobudzanie DML w częstotliwościach niższych niż 70 Hz jest – zgodnie z zaleceniami producenta – niewskazane i może skutkować uszkodzeniem przetwornika. W czasie trwania pomiaru wskazania woltomierza podłączonego do wzmacniacza mocy równolegle z DML oscylowały wokół 2,83 V, co przy rezystancyjnym charakterze impedancji przetwornika równej 4 Ω daje moc 0,7 W. Dodatkowo zarówno część nadawcza, jak i odbiorcza toru pomiarowego (poza mikrofonem i kondycjonerem) zostały przebadane pod kątem wprowadzania zniekształceń za pomocą urządzenia Prism dScope. W przypadku części odbiorczej nie stwierdzono występowania zniekształceń, natomiast w przypadku części nadawczej wyznaczono charakterystykę amplitudowo-częstotliwościową układu, która posłużyła do wyznaczenia krzywej korekcyjnej. Charakterystykę tę i dalsze szczegóły dotyczące techniki pomiaru przedstawiono w [7]. Obiektem badań był głośnik DML Amina Edge 5i zawierający dwa wzbudniki drgań.

4.3. Charakterystyki amplitudowo-częstotliwościowe DML

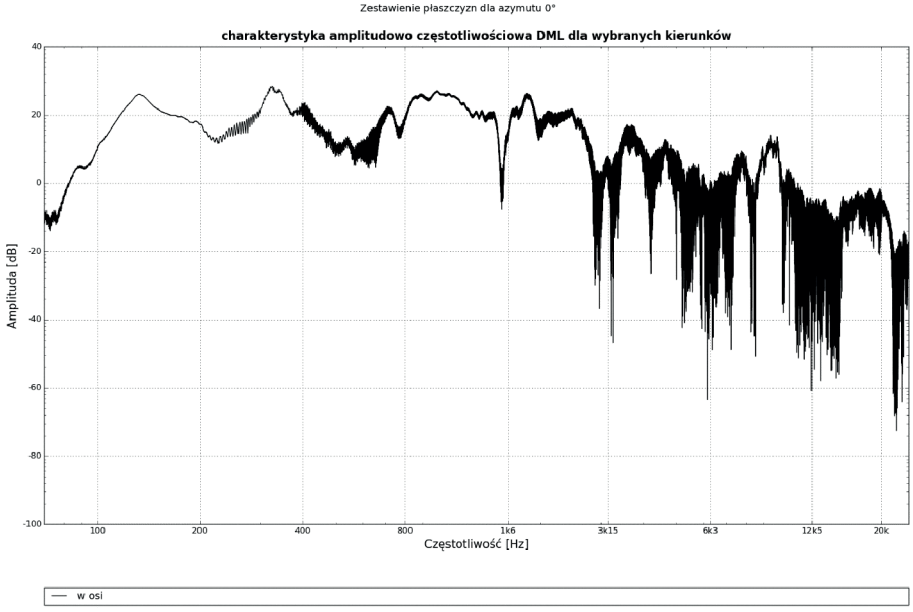
Uzyskane charakterystyki amplitudowo-częstotliwościowe dalekie są od płaskich. Po uśrednieniu charakterystyk z wszystkich 325 punktów pomiarowych uwydatnieniu ulegają maksima lokalnew okolicy 100, 300 i 10 kHz, jednak zupełnie zanikają widoczne w poszczególnych charakterystykach znaczące minima lokalne. Zjawisko to dowodzi dużej zmienności lokalizacji tych minimów w dziedzinie częstotliwości, zależnie od kąta

pomiaru. Uśredniona charakterystyka amplitudowo-częstotliwościowa DML w postaci widma mocy została przedstawiona na rys. 4.2. Na wykresie daje się zauważyć stopniowe opadanie charakterystyki przy częstotliwości powyżej 2 kHz, z wyjątkiem wyraźnego maksimum lokalnego w okolicy 10 kHz. To zjawisko jest tym bardziej widoczne na wykresach charakterystyk amplitudowo-częstotliwościowych DML z pojedynczych punktów pomiarowych. Zanika ono dopiero przy osiągnięciu przez kąt elewacji wartości 60° (rys. 4.3 i 4.5). Na rysunku 4.6 zaprezentowano zmienność uzyskiwanych charakterystyk amplitudowo-częstotliwościowych DML wraz ze wzrostem kąta elewacji punktów pomiarowych. Tłumienie charakterystyki powyżej 20 kHz wynika z tłumienia celowo wprowadzonego w sygnale wymuszającym.

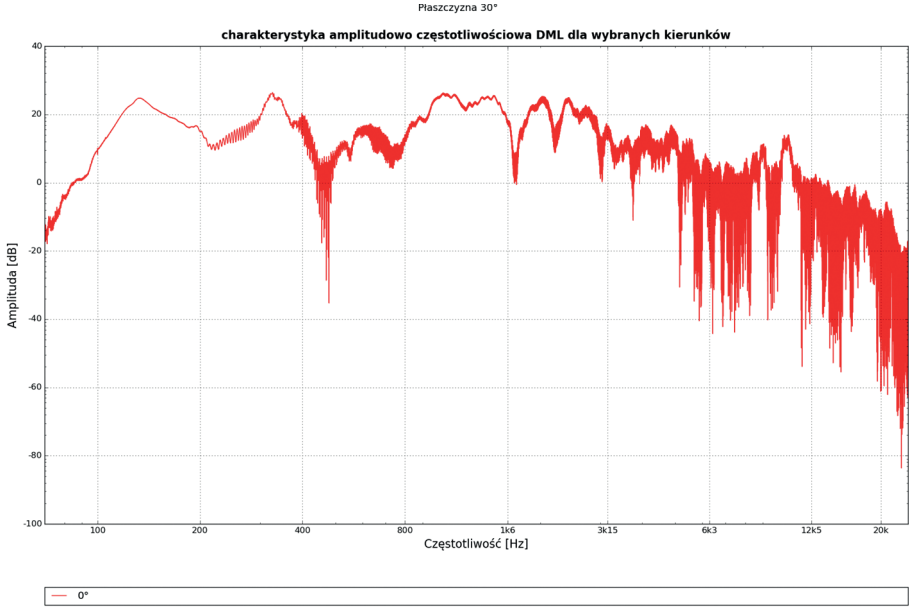
W punktach o dużych wartościach elewacji obserwowana dystrybucja minimów lokalnych widma jest bardziej równomierna, toteż różnice w charakterystykach amplitudowo-częstotliwościowych uzyskiwanych w poszczególnych punktach pomiarowych są łatwiej zauważalne (rys. 4.7 i 4.8). Uśrednianie charakterystyk amplitudowo-częstotliwościowych DML z punktów leżących na jednym okręgu równoleżnikowym niesie za sobą ryzyko utraty istotnych informacji na temat charakteru promieniowania przetwornika. Ten sposób analizy danych został wykorzystany przy wykreśleniu zgrubnych charakterystyk kierunkowości promieniowania DML w funkcji kąta elewacji przy częstotliwości 630, 1000, 1600 i 20 000 Hz. Widoczne na rys. 4.9 wykresy są symetryczne z uwagi na uśrednianie dla kątów azymutalnych w zakresie 10–90°.



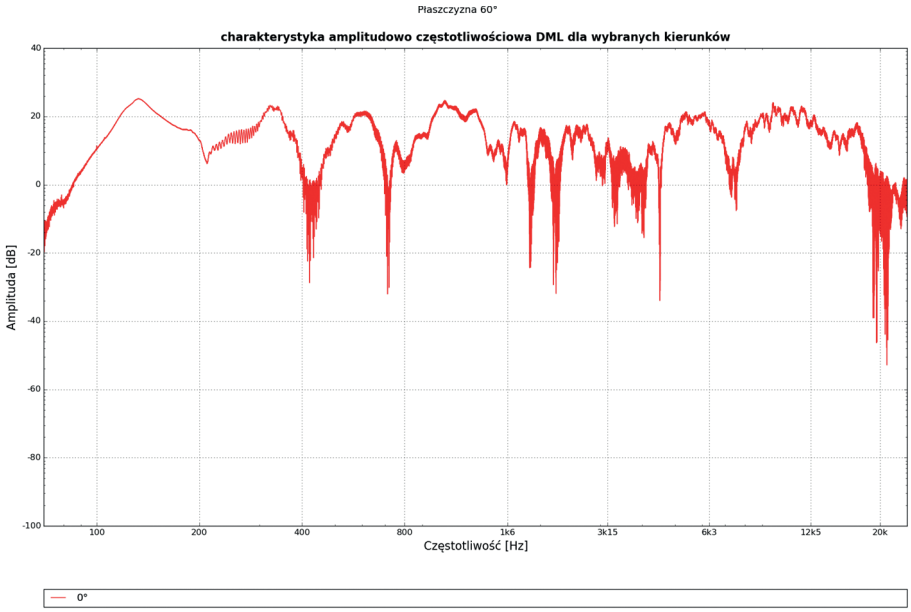
Rys. 4.2. Uśredniona charakterystyka amplitudowo-częstotliwościowa DML Amina Edge 5i



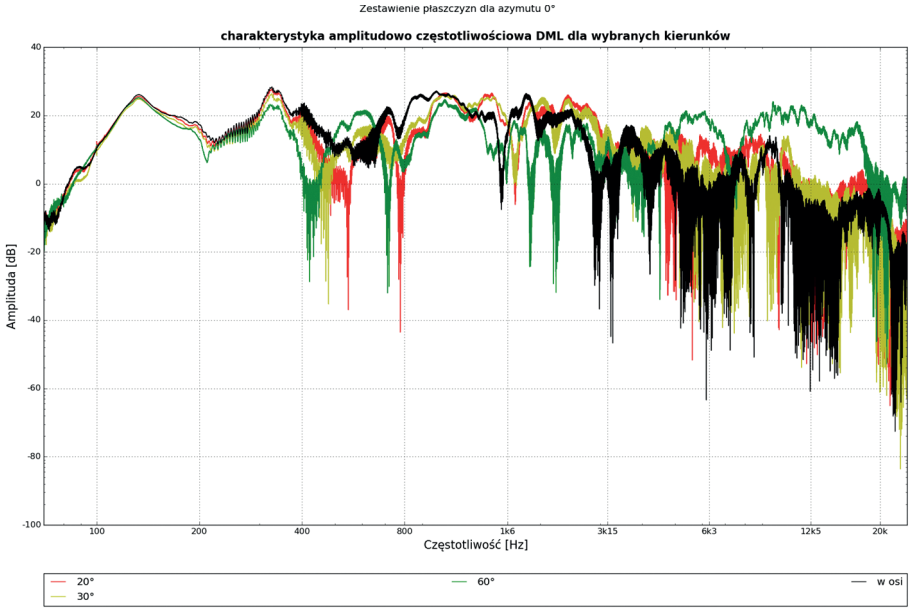
Rys. 4.3. Charakterystyka amplitudowo-częstotliwościowa DML w osi przetwornika



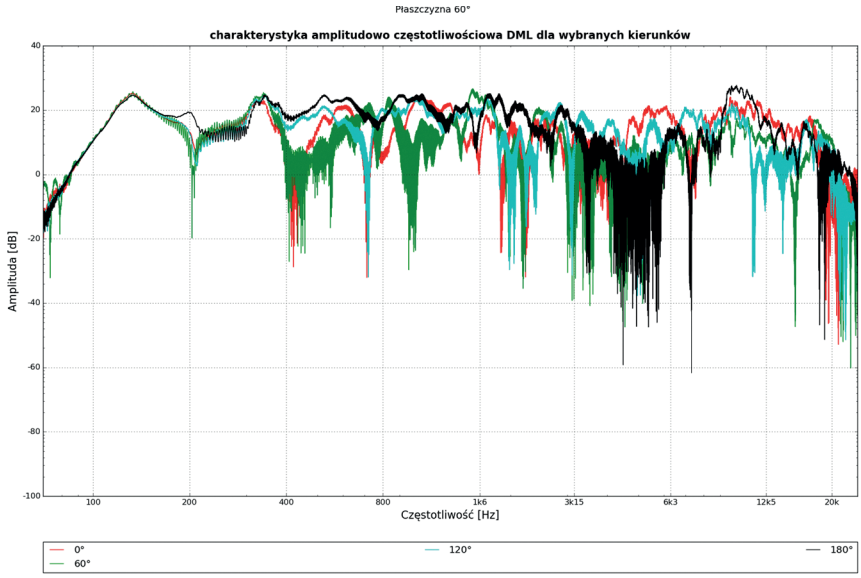
Rys. 4.4. Charakterystyka amplitudowo-częstotliwościowa DML 30° od osi przetwornika



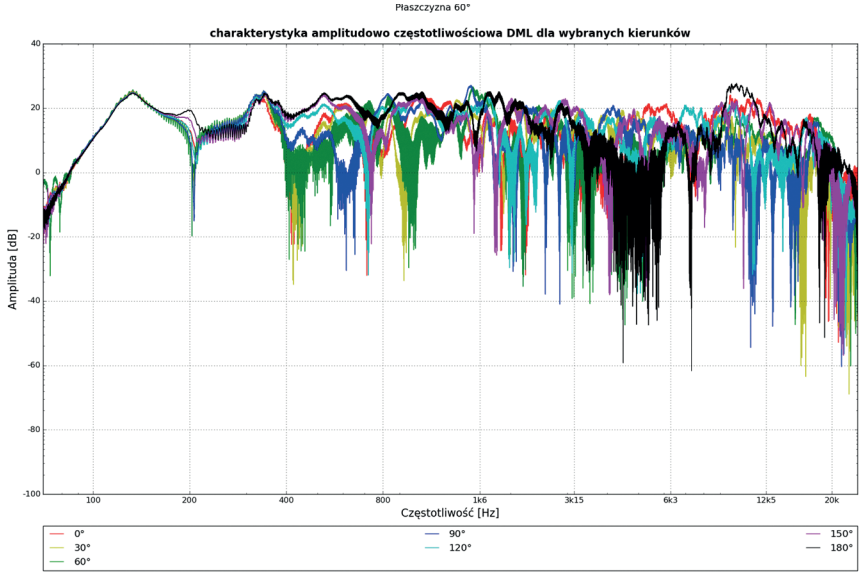
Rys. 4.5. Charakterystyka amplitudowo-częstotliwościowa DML 60° od osi przetwornika



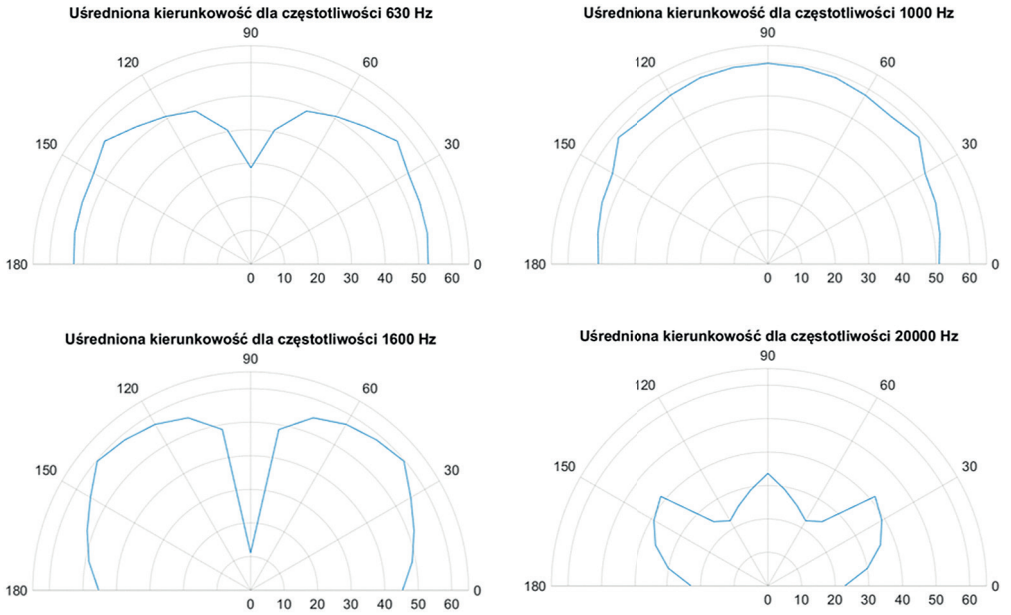
Rys. 4.6. Charakterystyki amplitudowo-częstotliwościowe DML w funkcji kąta elewacji



Rys. 4.7. Charakterystyki amplitudowo-częstotliwościowe DML w punktach leżących na okręgu równoleżnikowym 60°, z krokiem między punktami 60°



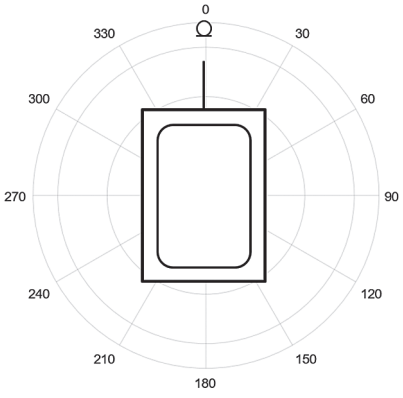
Rys. 4.8. Charakterystyki amplitudowo-częstotliwościowe DML w punktach leżących na okręgu równoleżnikowym 60°, z krokiem między punktami 30°



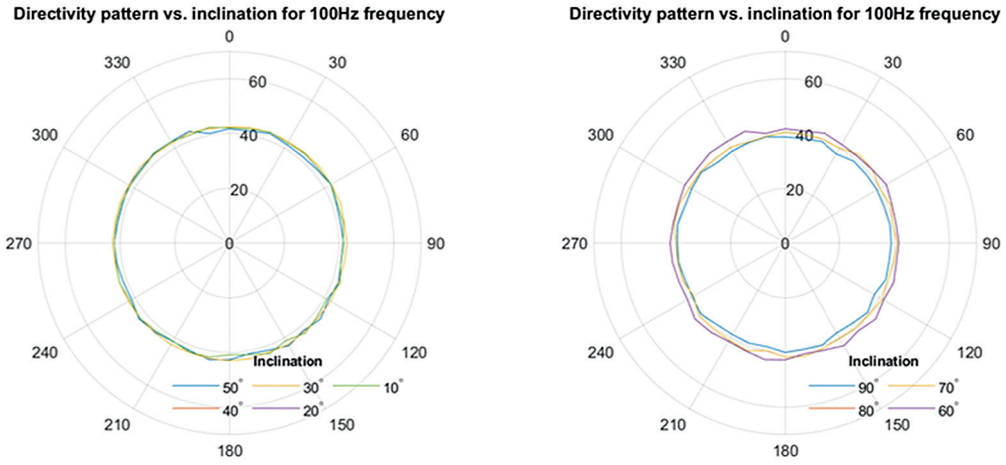
Rys. 4.9. Uśrednione wzdłuż okręgów równoleżnikowych charakterystyki kierunkowości promieniowania DML w funkcji kąta elewacji przy częstotliwości 630, 1000, 1600 i 20000 Hz

4.4. Wybrane charakterystyki kierunkowości promieniowania DML

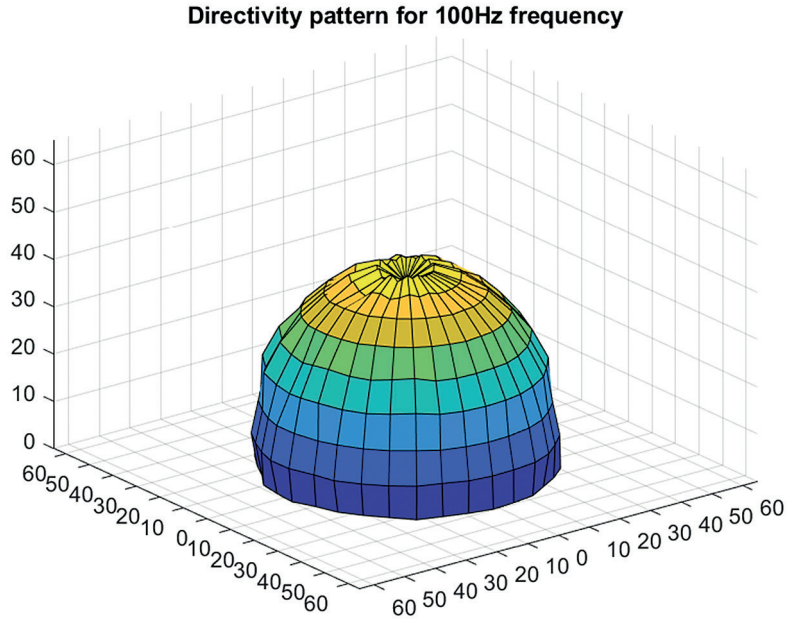
Przedstawione na rys. 4.9 zgrubne wykresy kierunkowości promieniowania DML zestawiono z wykresami kołowymi skuteczności DML w funkcji kąta azymutalnego, gdzie każdemu okręgowi równoleżnikowemu odpowiada inna seria danych. Serie danych zostały pogrupowane w pary wykresów kołowych. Jeden wykres z pary przedstawia serię danych dla elewacji z zakresu 10–50°, drugi – z zakresu 60–90°. Wykresy te zostały uzupełnione wykresami trójwymiarowymi (tzw. balonami kierunkowości), obejmującymi również skuteczność przetwornika dla danej częstotliwości w jego osi. Wykresy trójwymiarowe zostały przedstawione w rzutach izometrycznych. W ten sposób zostały przedstawione cha-



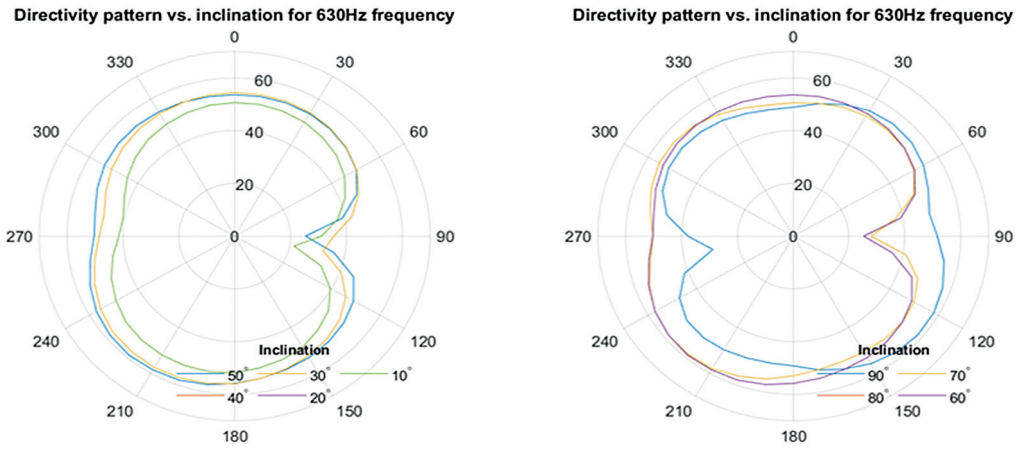
Rys. 4.10. Umiejscowienie DML w układzie pomiarowym



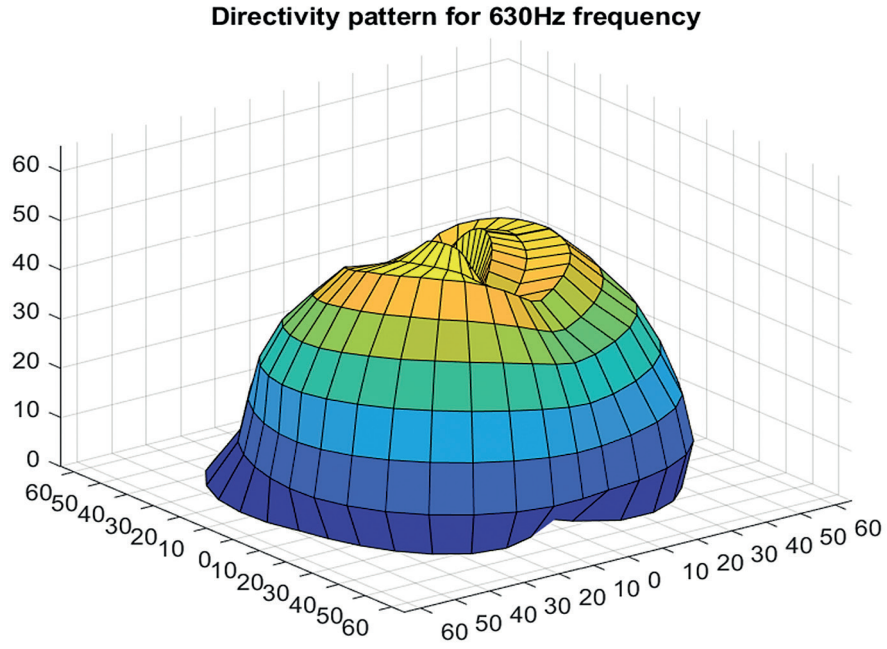
Rys. 4.11. Charakterystyki kierunkowości DML w funkcji kąta azymutalnego przy częstotliwości 100 Hz



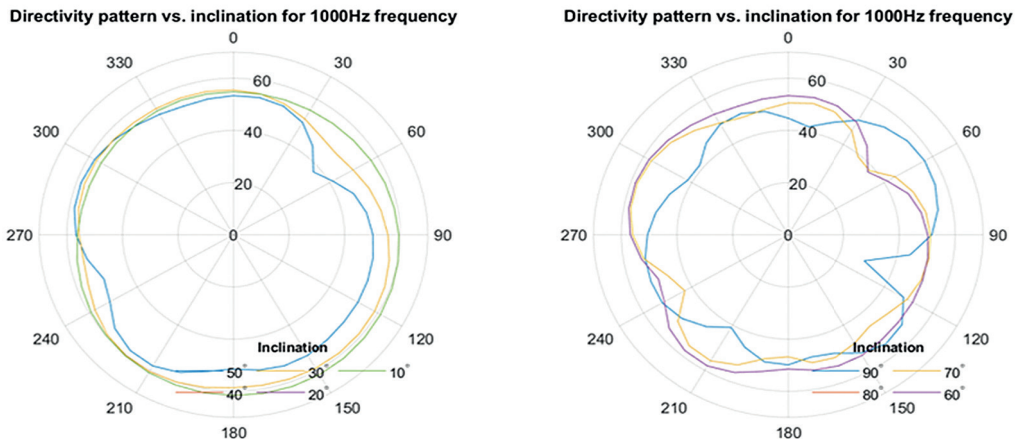
Rys. 4.12.Trójwymiarowy wykres kierunkowości promieniowania DML przy częstotliwości 100 Hz



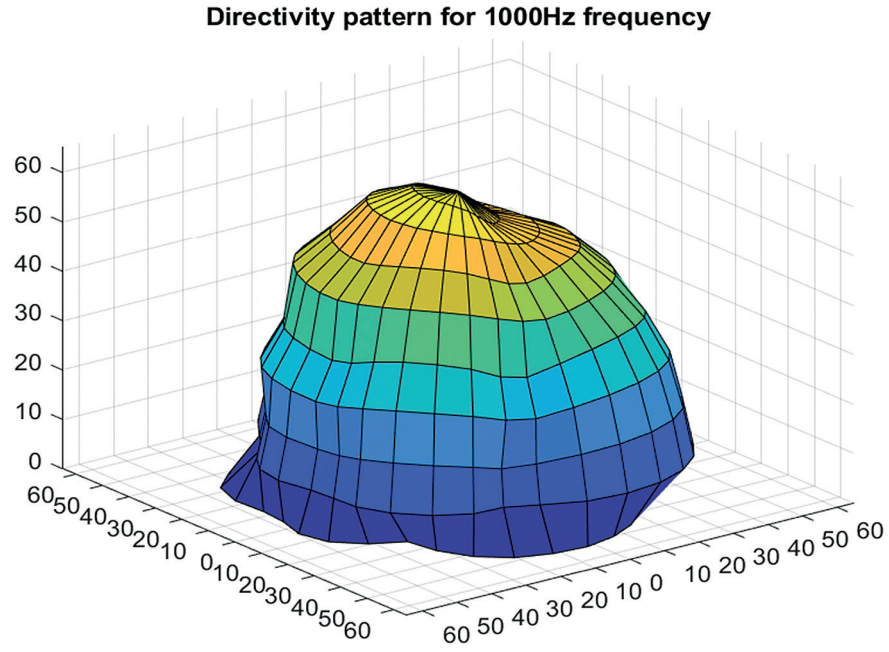
Rys. 4.13. Charakterystyki kierunkowości DML w funkcji kąta azymutalnego przy częstotliwości 630 Hz



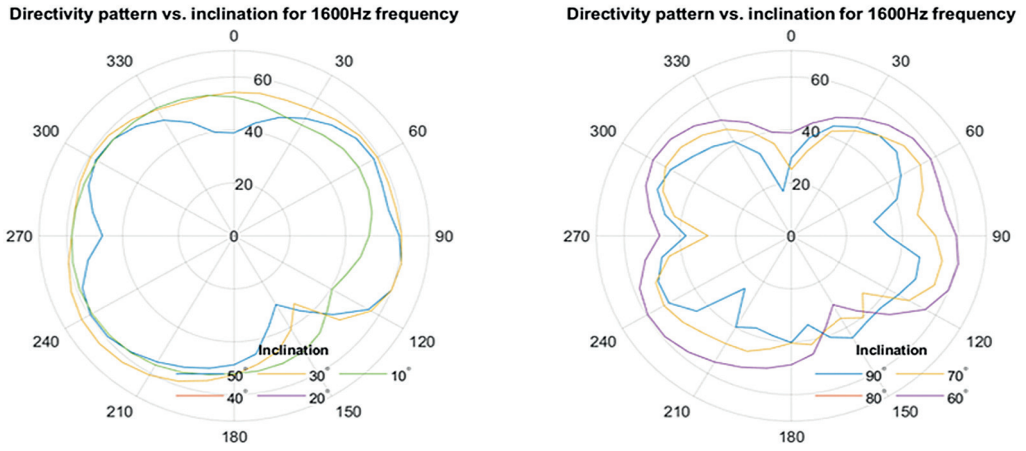
Rys. 4.14.Trójwymiarowy wykres kierunkowości promieniowania DML przy częstotliwości 630 Hz



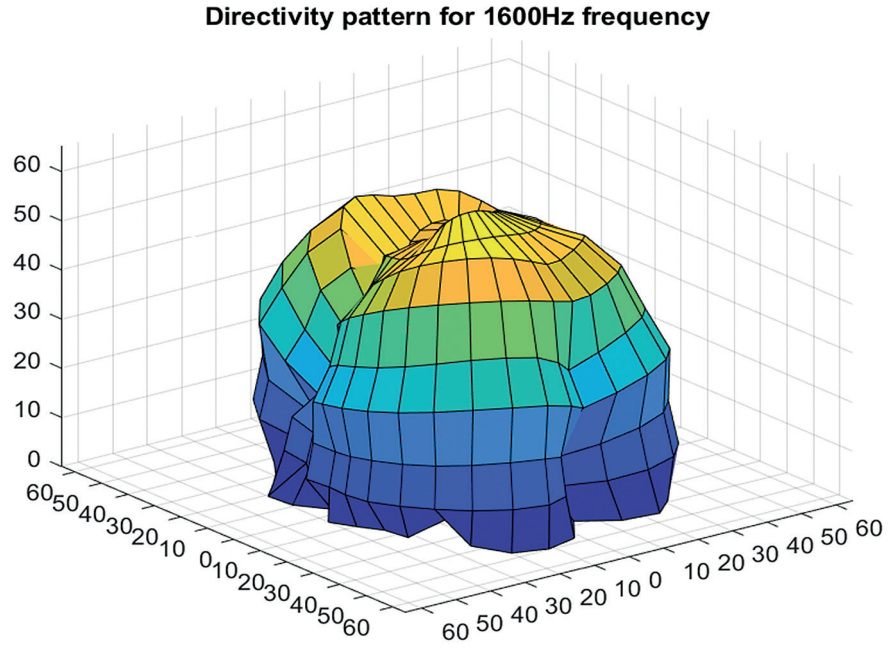
Rys. 4.15. Charakterystyki kierunkowości DML w funkcji kąta azymutalnego przy częstotliwości 1000 Hz



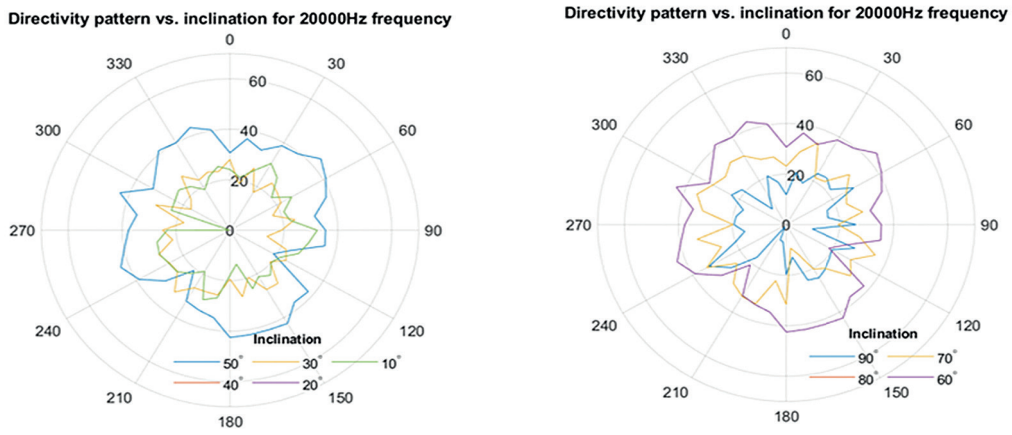
Rys. 4.16.Trójwymiarowy wykres kierunkowości promieniowania DML przy częstotliwości 630 Hz



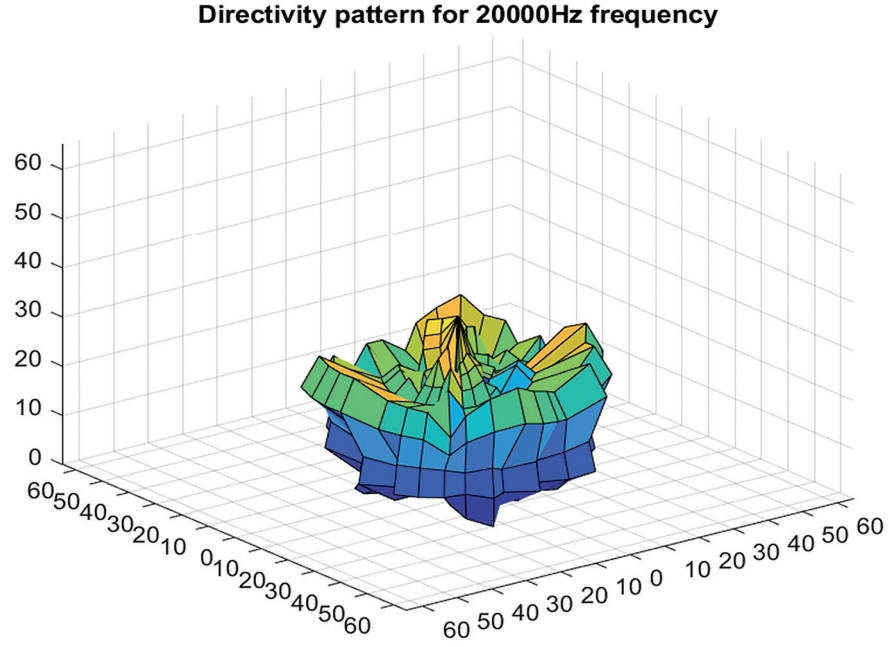
Rys. 4.17. Charakterystyki kierunkowości DML w funkcji kąta azymutalnego przy częstotliwości 1600 Hz



Rys. 4.18.Trójwymiarowy wykres kierunkowości promieniowania DML przy częstotliwości 1600 Hz



Rys. 4.19. Charakterystyki kierunkowości DML w funkcji kąta azymutalnego przy częstotliwości 20 000 Hz



Rys. 4.20.Trójwymiarowy wykres kierunkowości promieniowania DML przy częstotliwości 20000 Hz

rakterystyki kierunkowości promieniowania DML przy częstotliwości wynoszącej 630 (rys. 4.13 i 4.14), 1000 (rys. 4.15 i 4.16), 1600 (rys. 4.17 i 4.18), 20 000 Hz (rys. 4.19 i 4.20) oraz – dodatkowo – przy 100 Hz (rys. 4.11 i 4.12).

Wykresy potwierdzają wskazane we *Wprowadzeniu* zjawiska, takie jak dookólność promieniowania w niskich częstotliwościach, pojawianie się listków bocznych czy występowanie w obrębie półsfery punktów, w których notowana skuteczność przetwornika jest wyższa niż w osi prostopadłej do jego płaszczyzny. Co więcej – tak szczegółowo sporządzone wykresy stanowią informację o asymetrycznościach wiązek promieniowania DML w konkretnych częstotliwościach. W celu jednoznacznej interpretacji wyników na rys. 4.10 przedstawiona została orientacja badanego DML w układzie pomiarowym, w którym wartość azymutu wynosi 0^o. Zaznaczono też położenie mikrofonu, które pozostawało niezmienne. Obracany był wyłącznie DML. Co drugi obrót odbywał się przeciwnie do ruchu wskazówek zegara. Umiejscowienie DML w układzie ułatwiało osiowosymetryczne położenie gniazda sygnału zasilającego (przewód wrysowano czarną linią).

4.5. Wnioski

Głośniki modów rozproszonych wykazują całkiem inne właściwości promieniowania niż konwencjonalne głośniki tłokowe. Dotyczy to zarówno samych charakterystyk amplitudowo-częstotliwościowych, jak i charakterystyk kierunkowości. Prawidłowości, obserwowane w przypadku głośników tłokowych, np. zawężanie wiązki promieniowania wraz ze wzrostem częstotliwości pobudzenia, w przypadku DML okazują się nie zachodzić. Promieniowanie DML ma charakter równomiernie wszechkierunkowy tylko w niskich częstotliwościach – spośród tu przedstawionych tylko przy 100 Hz i w tym zakresie częstotliwości jest bardzo zbliżone do głośników tłokowych. Począwszy już od częstotliwości średnich zaczyna występować nierównomierność promieniowania w sposób istotnie nieregularny. Przy częstotliwości 630 Hz promieniowanie jest szersze wzdłuż dłuższej osi DML (odwrotnie, niż miałoby miejsce w przypadku źródła koherentnego), a przy częstotliwości 1600 Hz wzdłuż osi krótszej. Najwyższa nierównomierność występuje w najwyższych częstotliwościach, przy czym we wszystkich częstotliwościach promieniowanie ma charakter nierównomiernie wszechkierunkowy. Kształty wiązek promieniowania generowanych przez DML są na tyle nieregularne i asymetryczne, że estymacja „balonów kierunkowości" na podstawie serii pomiarów charakterystyk amplitudowo-częstotliwościowych głośnika w dwóch prostopadłych płaszczyznach daje wyniki mylące, niemające pokrycia w stanie faktycznym.

Dokładna znajomość charakterystyk promieniowania DML umożliwia wyliczenie wartości ich estymatorów takich jak indeks kierunkowości oraz pozwala na modelowanie tego typu przetworników w oprogramowaniu predykcyjnym czy też zaprojektowanie krzywych korekcyjnych do zaimplementowania w procesorach sygnałowych w torze elektroakustycznym zakończonym DML.

W obliczu zaobserwowanych zjawisk nasuwa się wniosek, że wszystkie zależne od kierunku pomiaru sposoby opisu przetworników elektroakustycznych, czyli charakterystyka amplitudowo-częstotliwościowa, fazowo-częstotliwościowa, pasmo przenoszenia, skuteczność czy odpowiedź impulsowa, jeżeli dokonywane są tylko w osi prostopadłej do powierzchni przetwornika, nie są właściwe dla głośników DML.

Bibliografia

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

[2] Harris N.J., HawksfordM., *Introduction to distributed mode loudspeakers (DML) with first-order behavioural modelling*, „IEE Proceedings. Circuits, Devices and Systems" 2000, Vol. 147, No. 3, s. 153– 57.

[3] Zhang S.Z., Shen Y., Shen X.X., Zhou J.L., *Model optimization of distributed-mode loudspeaker using attached masses*, „Journal of the Audio Engineering Society" 2006, Vol. 54, Iss. 4, s. 295–305.

[4] Anderson D.A., Bocko M.F., *A Model for the Impulse Response of Distributed-Mode Loudspeakers and Multi-Actuator Panels*, w: 139th AES Convention, October 29–November 1 New York, 2015.

[5] Gontcharov V.P., Hill N.P.R., Taylor V.J., *Measurement Aspects of Distributed Mode Loudspeakers*, w: 106th AES Convention, May 8–11 Munich, 1999.

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

[7] Czesak K., Kleczkowski P., Król-Nowak A., *Metodyka wyznaczania charakterystyk kierunkowości głośników modów rozproszonych*, w: Postępy w inżynierii dźwięku i psychoakustyce, Wydawnictwa AGH, Kraków 2022 (otwarty dostęp), s. 111–120.

Słowa kluczowe: głośniki modów rozproszonych, DML, charakterystyki, kierunkowość, wiązka.

Wybrane aspekty charakterystyk kierunkowości głośników modów rozproszonych

Głośniki modów rozproszonych (ang. *Distributed Mode Loudspeakers* – DML) charakteryzują się odmiennymi właściwościami niż głośniki tłokowe. Charakterystyki czasowo-częstotliwościowe silnie zależą od wyboru punktu pomiarowego. W ramach niniejszego rozdziału przeprowadzono szczegółowe pomiary charakterystyki kierunkowej jednego typu głośnika DML: Amina Edge 5i. Z uwagi na zalecany i łatwy do wykonania sposób montażu głośnika DML (na powierzchni ściany) zbadano charakterystykę kierunkową na półsferze. Stwierdzono, że wiązka promieniowania głośnika DML nie ulega zwężeniu w kierunku osi głośnika wraz ze wzrostem częstotliwości pobudzenia. Zauważalne jest zjawisko pojawiania się listków bocznych, których liczba wzrasta razem z częstotliwością pobudzenia, co jest widoczne już od 500 Hz. Poniżej tej częstotliwości głośniki DML promieniują dookólnie. Z pomiarów wynika, że na osi głośnika średnie ciśnienie akustyczne jest niższe niż w wielu innych kierunkach, a charakterystyki amplitudowe w większości pasma znacząco różnią się zależnie od kierunku. Stąd charakterystyka amplitudowa w osi przetwornika nie

jest miarodajną oceną jego zachowania w całym przetwarzanym paśmie. Przedstawiono wybrane uśrednione charakterystyki amplitudowo-częstotliwościowe, a także wybrane wykresy charakterystyk kierunkowych w dwóch oraz w trzech (tzw. balony kierunkowości) wymiarach.

Directivity characteristics of Distributed Mode Loudspeakers. Selected Aspects

Distributed Mode Loudspeakers (DML) are characterized by properties, significantly differing from those related to pistonic loudspeakers. The differences occur due to design assumptions of the DML, that are totally differing from design principles of pistonic loudspeakers. The oscillations of the DML consist of bending waves, propagating across the loudspeaker surface, which is rectangular. That is the cause, why transducers of this type present frequency characteristics with sharp local minima occurring at various frequencies, depending on the angle between the loudspeaker's surface and a measurement microphone. Also, the directivity characteristics of the DML are differing from those related to pistonic loudspeakers. In this work detailed measurements of directional characteristics of one type of the DML: Amina Edge 5i have been carried out. As a consequence of the preferred and easy to implement method of mounting the DML (flush mounting) the half-sphere directional characteristics have been evaluated. It has been found out, that the beam was not tightening with increasing frequency, but started splitting into several side beams, what occurred for frequencies above 500 Hz. The measurements have demonstrated that the average sound pressure in the axis of the loudspeaker was lower than in many other directions, and amplitude characteristics in most of the band considerably differ depending on the direction. Hence, the on-axis amplitude characteristic is not a reliable assessment of its performance in the entire frequency band. Selected averaged amplitude-frequency characteristics are presented, as well as selected plots of directional characteristics in two and three (so called directivity balloons) dimensions.

5. Wykorzystanie testu MUSHRA w badaniu korzyści użytkowania protez słuchowych

Piotr Szymański[1], Tomasz Poremski[2], Bożena Kostek[3]

[1] Sonova Audiological Care Sp. z o.o,
ul. Gabriela Narutowicza 130, 90-146 Łódź
[2] Advanced Bionics Polska,
Sonova Audiological Care Polska Sp. z o.o.,
ul. Gabriela Narutowicza 130, 90-146 Łódź
[3] Politechnika Gdańska,
Wydział Elektroniki i Telekomunikacji,
ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk

5.1. Wprowadzenie

Ocena jakości dopasowania aparatów słuchowych w kontekście korzyści, jakie może przynieść ta proteza, jest złożonym zagadnieniem. W łatwy sposób można jednak wyznaczyć obiektywne parametry aparatów, m.in. wzmocnienie, zniekształcenia harmoniczne, pasmo przenoszenia. Nie zawsze mają one jednak bezpośredni i decydujący wpływ na subiektywną ocenę przez pacjenta jakości dopasowania protezy słuchowej. Współczesne aparaty słuchowe posiadają szereg zaawansowanych rozwiązań, które ułatwiają i poprawiają (zwłaszcza) rozumienie mowy w różnych trudnych sytuacjach akustycznych, ale ich porównanie lub pomiar nie jest w pełni możliwy.

W większości wymienionych rozwiązań, np. w układach redukcji hałasów, modułach poprawy jakości sygnału mowy, mikrofonach kierunkowych (w tym w układach odpowiadających właściwościom małżowiny usznej), dąży się do poprawy stosunku sygnału do szumu (SNR). Nowoczesny aparat słuchowy ma zapewnić jak najlepsze rozumienie mowy, ale jednocześnie naturalne wrażenia słuchowe w celu zapewnienia komfortu przebywania w różnych sytuacjach akustycznych. Realizację tego założenia umożliwia funkcja automatycznego rozpoznawania warunków akustycznych i adaptacyjny dobór poszczególnych układów oraz regulacja ich nastaw. Teoretycznie zastosowanie wszystkich tych rozwiązań powinno zapewnić poprawę jakości słyszenia i satysfakcję użyt-

kowników aparatów słuchowych. W codziennej praktyce zarówno pacjent, jak i protetyk słuchu muszą jednak dokonywać pewnych wyborów czy iść na kompromis.

Różnorodność rozwiązań technicznych w oferowanych przez producentów aparatach słuchowych sprawia, że są one trudne do porównania i obiektywnej oceny. Wynika to z faktu, że rozwiązania oferowane w aparatach słuchowych, mimo iż są podobne do siebie, to ich jakość i wpływ na poprawę percepcji słuchowej zależą m.in. od indywidulanej konfiguracji oraz pracy algorytmów, które nimi zarządzają. Protetyk słuchu zazwyczaj nie ma pełnego wglądu i dostępu do mechanizmów ich działania. W codziennej praktyce powinien opierać się na wskazaniach producenta, swoim doświadczeniu i informacji zwrotnej od pacjenta. Z kolei pacjent, który podjął decyzję o zakupie aparatu słuchowego, chciałby otrzymać rozwiązanie, które zapewni mu odzyskanie pełnej zdolności słyszenia oraz rozumienia mowy we wszystkich sytuacjach, w których przebywa.

Pomiary efektywności aparatu słuchowego mogą dotyczyć wielu aspektów, m.in. kompensacji niedosłuchu, akceptacji, zysku czy też satysfakcji z protezowania. Ze względu na specyficzny zakres wiedzy najnowsze narzędzia do pomiaru efektywności protezowania dostępne są jedynie dla specjalistów. Stworzenie łatwej w obsłudze i intuicyjnej aplikacji internetowej dałoby taką możliwość zarówno protetykom słuchu, jak i pacjentom. W ten sposób zobiektywizowana ocena efektywności protezowania byłaby pomocna przy wyborze optymalnego rozwiązania poprawiającego słuch oraz w jego precyzyjnym dopasowaniu i regulacji. W późniejszym okresie użytkowania protezy słuchowej aplikacja ta służyłaby do monitorowania postępów w rehabilitacji słuchu. Uzyskiwane wskaźniki mogłyby być wykorzystywane do przewidywania długoterminowych efektów protezowaniu już po krótkim, próbnym okresie użytkowania protez słuchowych.

Rozwijana przez autorów niniejszego rozdziału aplikacja internetowa, jak również prowadzone badania mogą przyczynić się do powstania innowacyjnego narzędzia oceny efektywności dopasowania aparatów słuchowych. Może ono zostać zaimplementowanie w dużej liczbie punktów protetycznych oraz udostępnione w odpowiednio przygotowanej formie pacjentom. Dzięki temu pacjenci mogliby dokonywać oceny protezowania nie tylko w punkcie protetycznym, ale również np. w domu lub innym otoczeniu akustycznym, które jest dla nich szczególnie ważne. Wyniki oceny mogłyby zatem służyć jako narzędzie do bardziej zobiektywizowanej oceny słyszenia w aparatach i ułatwić pacjentowi dokonanie wyboru między różnymi dostępnymi rozwiązaniami już po krótkim okresie ich użytkowania (testowania).

5.2. Cel badania

Opracowana metoda oceny korzyści użytkowania protez słuchowych powinna:

- poddawać ocenie najbardziej typowe sytuacje akustyczne, z którymi boryka się osoba niedosłysząca w podeszłym wieku;

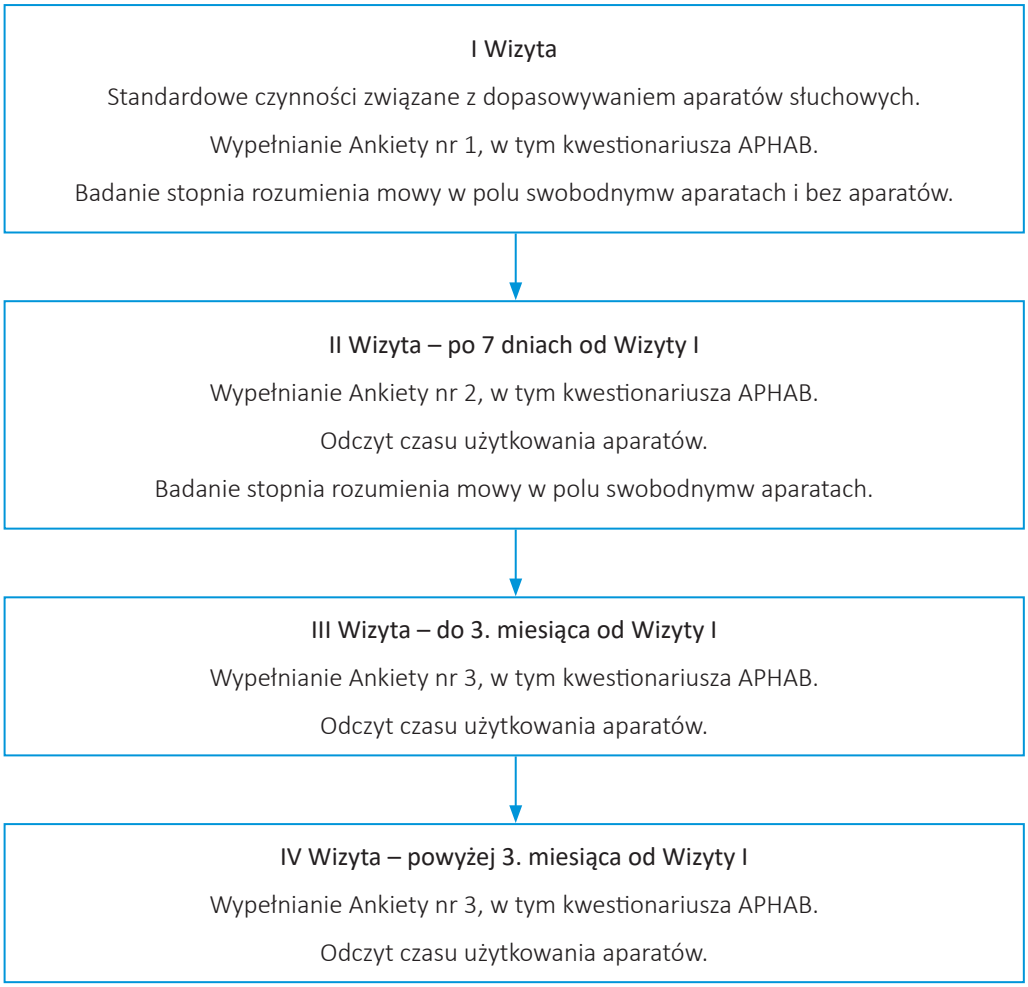
- w ocenie korzyści z użytkowania aparatów uwzględniać: stopień niedosłuchu, doświadczenie pacjenta, rodzaj zastosowanych aparatów;
- poddawać ocenie pozaakustyczne wskaźniki i aspekty użytkowania aparatów słuchowych;
- być łatwa do wdrożenia w punktach protetycznych, wymagać zaangażowania istniejących zasobów personalnych i wykorzystania typowego wyposażenia audiologicznego;
- zostać wdrożona w postaci łatwej do obsługi aplikacji komputerowej.

5.3. Opis metody

W celu zebrania danych służących do oceny korzyści z zastosowania aparatów słuchowych przygotowano aplikację internetową, pozwalającą na uporządkowane podejście do zbierania otrzymanych wyników. Aplikacja ta została przygotowana z wykorzystaniem platformy LMS (ang. *Learning Management System*) Moodle [10]. Ponieważ ta platforma e-learningowa powstała na bazie języka skryptowego PHP, jest m.in. wysoce elastyczna i w pełni konfigurowalna. Za jej wykorzystaniem przemawiała również dostępność w punktach protetycznych, znajomość jej obsługi przez potencjalnych użytkowników oraz możliwość wykorzystania zaimplementowanego modułu bazy danych. Zaprojektowany interfejs użytkownika bazy danych ma postać formularza, którego strukturę i formę można modyfikować. Moduł bazy danych pozwala również na konfigurację zakresu eksportowanych danych. Dane można pobrać w dowolnie zdefiniowanym zbiorze pól lub w całości. Eksportu można dokonać w dwóch formatach: CSV lub ODS. Aplikacja zawiera ankiety ściśle związane z kolejnymi wizytami pacjenta. Służą one do porządkowania danych i wskazują, co należy wykonać na kolejnych etapach obsługi użytkownika protezy słuchu, dlatego należy je wypełniać w odpowiedniej kolejności. Na rysunku 5.1 przedstawiono schemat zbierania danych.

Jednym z najważniejszych elementów ankiet jest powszechnie stosowany zamknięty, wypełniany samodzielnie przez pacjenta kwestionariusz oceny korzyści użytkowania aparatów słuchowych APHAB (ang. *Abbreviated Profile of Hearing Aid Benefit*) [4], [7], [16]. Kwestionariusz ten składa się z 24 elementów (stwierdzeń) w czterech podkategoriach (po 6 stwierdzeń na kategorię) dotyczących:

- EC (ang. *Ease of Communication*) – zdolności komunikacji w ciszy, wysiłku związanego z komunikacją w relatywnie łatwych warunkach odsłuchu;
- RV (ang. *Reverberation*) – zdolności komunikacji i rozumienia mowy w warunkach umiarkowanego pogłosu;
- BN (ang. *Background Noise*) – komunikowania się w obecności szumu otoczenia, rozumienia mowy w obecności wielu rozmówców lub innych konkurencyjnych warunkach akustycznych (hałasu środowiskowego);
- AV (ang. *Aversiveness of Sounds*) – stopnia akceptacji nieprzyjemnych dźwięków, negatywnych reakcji na dźwięki środowiskowe [2], [3], [5].



Rys. 5.1. Schemat zbierania danych

Poszczególne punkty oceniane są w skali 7-stopniowej. Każdy stopień skali, od A do G, zawiera opis i związaną z nim wartość procentową (tabela 5.1).

Celem stosowania kwestionariusza APHAB może być:

- przewidywanie prawdopodobnego powodzenia z zastosowania aparatu słuchowego [3] lub alternatywnych urządzeń wspomagających słyszenie [8];
- porównanie funkcjonowania osoby stosującej aparat słuchowy (aparaty słuchowe) z wynikami grupy referencyjnej, używającej z sukcesem aparatów słuchowych [3];
- dokumentacja korzyści z zastosowania aparatów słuchowych w różnych środowiskach w celu poprawy (wyeliminowania) nieskutecznego dopasowania, jak i po-

równania zysku przy zastosowaniu różnych aparatów słuchowych lub różnych programów w danym aparacie słuchowym [3];
- potwierdzenie skuteczności nowych procedur doboru i strojenia aparatów słuchowych czy też innych urządzeń wspomagających słyszenie [12].

Tabela 5.1. Skala kwestionariusza APHAB

A	Zawsze	Always	99%
B	Prawie zawsze	Almost Always	87%
C	Na ogół	Generally	75%
D	Pół-na-pół	Half-the-time	50%
E	Czasami	Occasionally	25%
F	Rzadko	Seldom	12%
G	Nigdy	Never	1%

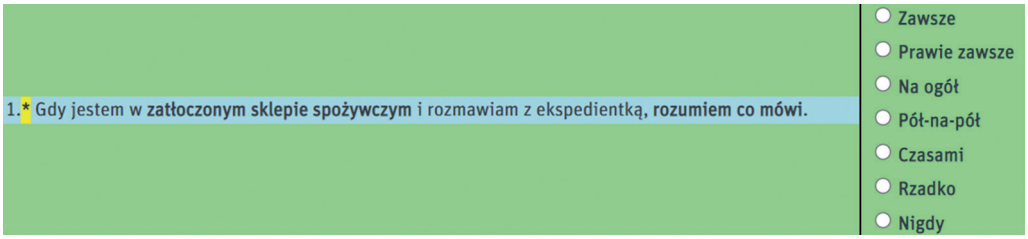
Korzyść wynikającą z zastosowania aparatu słuchowego można ocenić, analizując średnie wartości procentowe w poszczególnych kategoriach (EC, RV, BN, AV) [3] [5], jak również wartość średnią w kilku kategoriach (ang. *Global Score*). Według Hojana i in. [5] jest to wartość średnia z 4 kategorii, według Jani i in. [6] jest to wartość średnia z kategorii EC, RV, BN. Kwestionariusz ten używany jest w wielu krajach i w różnych wersjach językowych.

Porównując wyniki uzyskane przez autorów w dwóch grupach badanych [11], [15], stwierdzono, że członkowie jednej z grup uzyskali w ogólności lepsze wyniki. Może to wynikać ze skali oceny słyszenia stosowanej w formularzu APHAB, jak również przekroju wiekowego użytkowników protez słuchowych. Połączenie skali literowej, procentowej i opisowej może stanowić trudność w interpretacji dla osób badanych, którymi w przeważającej większości są osoby w podeszłym wieku.

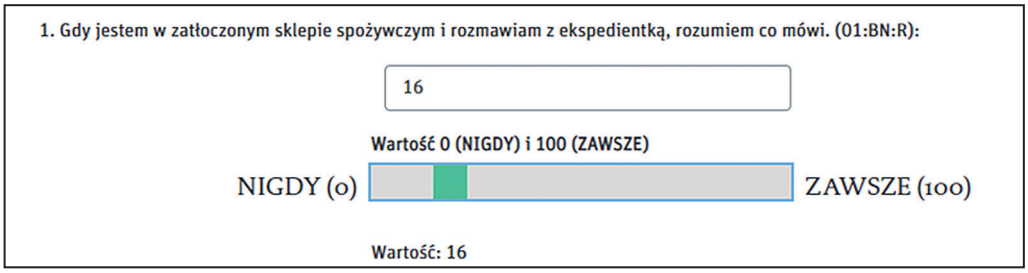
5.4. Modyfikacja metody

W związku z powyższym została opracowana koncepcja modyfikacji kwestionariusza, tj. przekształceniua skali APHAB na skalę zgodną ze skalą testu MUSHRA (ang. *MUltiple Stimuli with Hidden Reference and Anchor*) [9] przy jednoczesnym wykorzystaniu logiki rozmytej. Test MUSHRA znajduje zastosowanie w ocenie jakości dźwięku aparatów słuchowych zarówno przez osoby z ubytkiem słuchu, jaki i przez osoby z prawidłowym słuchem [1], [17], [13]. Z kolei logika rozmyta znajduje zastosowanie w procedurach regulacji aparatów słuchowych z wykorzystaniem skalowania głośności [14].

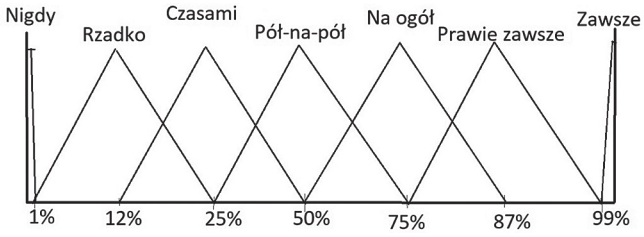
APHAB



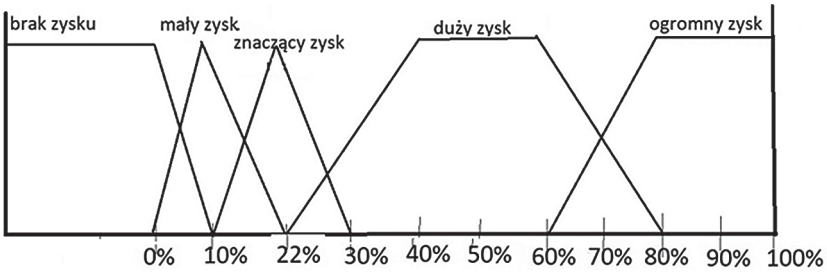
MUSHRA



Rys. 5.2. Przekształcenie skali APHAB na skalę MUSHRA



Rys. 5.3. Przykładowa funkcja przynależności w kategorii EC



Rys. 5.4. Przykładowa skala zysku w kategorii EC

Zadaniem użytkownika aparatów słuchowych jest ocena słyszenia (jakości dźwięku) bez aparatów i w aparatach według skali 100-punktowej (patrz rys. 5.2) w przykładowej sytuacji w kategorii EC (łatwość komunikacji). W pierwszym kroku przekształcenia zamienia się 7-stopniową skalę zastosowaną w ankiecie APHAB, w której każdemu stopniowi przypisana jest wartość procentowa, np. rzadko – 12% (tj. wartość 0,12), na skalę ciągłą z testu MUSHRA (0–100). W związku z tym na bazie ankiety APHAB została zbudowana funkcja przynależności w poszczególnych kategoriach przez analogię do procentowej skali MUSHRA. W tabeli 5.2 – w przypadku badanego nr 1 i pytania nr 4 – uzyskano oceny według skali APHAB 0,12 (12%), a według skali MUSHRA 0,04 (4%). Następnym krokiem jest rozmycie wartości uzyskanej według skali MUSHRA zgodnie z funkcją przynależności z rys. 5.3.

Kategoria EC, pytanie nr 1: Gdy jestem w zatłoczonym sklepie spożywczym i rozmawiam z ekspedientką, rozumiem co mówi.

Odpowiedzi użytkownika zostaną przypisane do odpowiedniej funkcji przynależności w danej kategorii. Na rysunku 5.3 pokazana jest przykładowa funkcja przynależności w kategorii EC.

Na podstawie oceny użytkownika i z zastosowaniem reguł logiki rozmytej zostanie wyznaczony zysk w określonej kategorii i/lub w kilku kategorii. Na rysunku 5.4 przedstawiono przykładową skalę zysku w kategorii EC.

5.5. Podsumowanie

Należy zauważyć, że uzyskane wyniki i wnioski, które nasuwają się po przeprowadzeniu analiz z wykorzystaniem tego typu metodologii są na wczesnym etapie implementacji ankiety. Obecnie dostępne są jedynie wstępne dane z zastosowania modyfikacji przedstawionej metody. W tabeli 5.2 przedstawiono wyniki oceny rozumienia mowy w kategorii EC: 1 oznacza zawsze, 0 – nigdy.

Pytania/stwierdzenia w kategorii EC (tabela 5.2):

- Nr 4: W domu, gdy jest cicho z trudnością słyszę słowa, które ktoś do mnie mówi.
- Nr 9: Gdy cicho rozmawiam ze swoim lekarzem w jego gabinecie, trudno mi zrozumieć, co do mnie mówi.
- Nr 12: W cichym pomieszczeniu, podczas rozmowy z jedną osobą, muszę prosić rozmówcę o powtarzanie słów.
- Nr 15: Gdy jestem w małym biurze, trudno mi zrozumieć kierowane do mnie pytanie.
- Nr 17: Kiedy rozmawiam z kimś w spokojnym miejscu, trudno mi go zrozumieć.
- Nr 18: Gdy mówca przemawia do niewielkiej grupy osób i wszyscy słuchają w ciszy, ja muszę się wysilić (bardzo uważnie słuchać), by zrozumieć tekst.

Im niższa jest wartość, tym mniejsze są trudności odczuwane w tej kategorii. Zaznaczono największe różnice w wartości średniej w danej kategorii. Różnice te mogą

Tabela 5.2. Wyniki oceny sytuacji w kategorii EC bez aparatów słuchowych (dla czterech badanych)

MUSHRA	Kategoria EC						
Pytanie nr	4	9	12	15	17	18	Wartość średnia w danej kategorii
Badany nr 1	0,04	0,09	0,06	0,02	0	0,17	0,06
Badany nr 2	0,05	0,07	0,04	0,05	0,02	0,03	0,04
Badany nr 3	0,16	0,1	0,07	0,06	0,17	0,08	0,11
Badany nr 4	0,4	0,5	0,7	0,64	0,3	0,7	0,54
APHAB	Kategoria EC						
Pytanie nr	4	9	12	15	17	18	Wartość średnia w danej kategorii
Badany nr 1	0,12	0,12	0,12	0,12	0,12	0,12	0,12
Badany nr 2	0,01	0,01	0,01	0,12	0,01	0,01	0,03
Badany nr 3	0,12	0,12	0,12	0,12	0,12	0,12	0,12
Badany nr 4	0,5	0,5	0,75	0,87	0,75	0,87	0,71

przełożyć się na precyzyjniejsze szacowanie zysku z używania aparatów słuchowych, szczególnie przy zastosowaniu reguł logiki rozmytej.

Średni czas wypełnienia ankiety z zastosowanie skali MUSHRA był krótszy o 20%, jednak za wcześnie jest wnioskować, czy będzie tak w przypadku wszystkich wypełniających ankietę. Kolejny wniosek dotyczy łatwości udzielania odpowiedzi na pytania – w tym przypadku zdania badanych były podzielone. Dwie osoby oceniły skalę MUSHRA jako łatwiejszą w użyciu, a dwie pozostałe jako trudniejsza. Na tym etapie trudno jest jeszcze wnioskować, czy taki trend utrzyma się przy większej liczbie osób testowanych.

Bibliografia

[1] Beck D., Tryanski D., Kai Loong Man B., *Sound Quality and Hearing Aids,* „Hearing Review" 2021, Vol. 28, No. 8, s. 30–31; https://hearingreview.com/hearing-products/hearing-aids/psap/sound-6 [dostęp: 27.10.2022].

[2] Briant T.A., *Self-Report Assessment of Hearing Aid Outcome – An Overview*, 2007; http://www.audiologyonline.com/articles/self-report-assessment-hearing-aid-931 [dostęp: 27.10.2022].

[3] Cox R.M., Alexander G.C., *The abbreviated profile of hearing aid benefit*, „Ear and Hearing" 1995, Vol. 16, No. 2, s. 176–186.

[4] Giordano P., Argentero P., Canale A., Lacilla, M. Albera, R., *Evaluation of hearing aid benefit through a new questionnaire: CISQ (Complete Intelligibility Spatiality Quality)*, „Acta otorhinolaryngologica Italica" 2013, Vol. 33, No. 5, s. 329–336.

[5] Hojan E., *Protetyka słuchu.* Wydawnictwo Naukowe UAM, Poznań 2014, s. 715–721.
[6] Jani A.J., Robyn M.C., Genevieve C.A., *Development of APHAB Norms for WDRC Hearing Aids and Comparisons with Original Norms*, „Ear and Hearing" 2010, Vol. 31, No. 1, s. 47–55.
[7] Löhler J., Gräbner F., Wollenberg B., Schalttmann P., Schönweiler R., *Sensitivity and specificity of the abbreviated profile of hearing aid benefit (APHAB)*, „European Archives of Oto-Rhino-Laryngology" 2017, Vol. 274, No. 10, s. 3593–3598.
[8] Maidment D.W., Barker A.B., Xia J., Ferguson M.A., *A systematic review and meta-analysis assessing the effectiveness of alternative listening devices to conventional hearing aids in adults with hearing loss*, „International Journal of Audiology" 2018, Vol. 57, No. 10, s. 721–729.
[9] ITU-R, *Method for the subjective assessment of intermediate quality level of audio systems*, Recommendation ITU-R BS.1534-3 (10/2015); https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en [dostęp: 27.10.2022].
[10] *Moodle Open source learning platform*; https://moodle.org/ [dostęp: 27.10.2022].
[11] Poremski T., Szymański P., Kostek B., *Assessment of the Effectiveness of a Short-term Hearing Aid Use in Patients with Different Degrees of Hearing* Loss, „Archives of Acoustics" 2019, Vol. 44, No. 4, s. 719–729.
[12] Sabin A.T., Van Tasell D.J., Rabinowitz B., Dhar S., *Validation of a Self-Fitting Method for Over-the-Counter Hearing* Aids, „Trends in Hearing" 2020, Vol. 24, No. 3, s. 1–19.
[13] Simonsen C.S., Legarth S.V., *A Procedure for Sound Quality Evaluation of Hearing Aids*, „Hearing Review" 2010, Vol. 25, No. 3, s. 32–37; https://hearingreview.com/practice-building/practice-management/a-procedure-for-sound-quality-evaluation-of-hearing-aids [dostęp: 27.10.2022].
[14] Suchomski P., Kostek B., Czyżewski A., *Hearing adi fitting method based on fuzzy logic processing*, „Archives of Acoustics" 2008, Vol. 33, No. 4, s. 153–158; https://acoustics.ippt.gov.pl/index.php/aa/article/view/851/730 [dostęp: 27.10.2022].
[15] Szymański P., Poremski T., Kostek B., *Pursuing Analytically the Influence of Hearing Aid Use on Auditory Perception in Various Acoustic Situations*, „Vibrations in Physical Systems" 2022, Vol. 33, No. 1.
[16] Turan S., Unsal S., Kurtaran H., *Satisfaction assessment with Abbreviated Profile of Hearing Aid Benefit (APHAB) questionnaire on people using hearing aid having Real Ear Measurement (REM)* eligibility, „The International Tinnitus Journal" 2019, Vol. 23, No. 2, s. 97–102.
[17] Völker C., Bisitz T., Huber R., Kollmeier B., Ernst S., *Modifications of the MUlti stimulus test with Hidden Reference and Anchor (MUSHRA) for use in* audiology, „International Journal of Audiology" 2018, Vol. 57 (sup3), s. 92–104.

Słowa kluczowe: aparaty słuchowe, ubytek słuchu, korzyść z aparatu słuchowego, MUSHRA, APHAB.

Wykorzystanie testu MUSHRA w badaniu korzyści użytkowania protez słuchowych

Ocena jakości dopasowania aparatów słuchowych w kontekście korzyści, jakie może przynieść tego rodzaju proteza, jest złożonym zagadnieniem. Obiektywne parametry aparatów, które można wyznaczyć (np.

wzmocnienie czy pasmo przenoszenia), nie zawsze mają bezpośredni i decydujący wpływ na subiektywną ocenę jakości dopasowania protezy słuchowej przez pacjenta. Pomiary efektywności aparatu słuchowego mogą dotyczyć wielu aspektów, m.in. kompensacji niedosłuchu, akceptacji, zysku czy też satysfakcji z protezowania. Autorzy przedstawiają modyfikację powszechnie stosowanego kwestionariusza oceny korzyści użytkowania aparatów słuchowych APHAB (ang. *Abbreviated Profile of Hearing Aid Benefit*), polegającą na połączeniu go z testem MUSHRA ang. (*MUltiple Stimuli with Hidden Reference and Anchor*), który stosowany jest w ocenie jakości dźwięku oraz przekształceniu skali APHAB na 100-punktową skalę MUSHRA za pomocą logiki rozmytej.

Employing the MUSHRA test in the study of the benefits of using hearing prostheses

Evaluating the quality of hearing aid fitting in terms of the benefits a prosthesis can provide is a complex issue. Objective parameters of hearing aids, such as gain, harmonic distortion, frequency response, etc., can be easily determined; however, they do not always have a direct and decisive influence on the patient's subjective assessment of the quality of the hearing aid fitting. Measurements of hearing aid effectiveness can address many aspects, including hearing loss compensation, acceptance, gain, or satisfaction with the prosthesis.

The authors present a modification of the commonly used hearing aid benefit assessment questionnaire APHAB (Abbreviated Profile of Hearing Aid Benefit) by combining it with the MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) test, which is used to assess the sound quality. In the paper, a concept of modifying the questionnaire was developed, i.e., mapping the 7-point APHAB scale to a 100-point scale consistent with the MUSHRA test scale using fuzzy logic.

6. Automatyczna klasyfikacja mowy patologicznej

Martyna Włoszczyńska, Bożena Kostek

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

6.1. Wprowadzenie

Celem niniejszego rozdziału jest przedstawienie wyników badań oraz projektu aplikacji służących do automatycznego wykrywania mowy patologicznej na podstawie bazy nagrań. Zaburzenia mowy mają istotny wpływ na komunikację osób nimi dotkniętych z otoczeniem. Trudności w wypowiadaniu pojedynczych głosek, pełnych słów i zdań zaburzają proces efektywnego porozumiewania się, co znacznie utrudnia adaptację w środowisku społecznym oraz może mieć wpływ na funkcje poznawcze zarówno w przypadku dzieci, jak i osób dorosłych. Wśród przyczyn powstawania patologii głosu/wymowy wymienia się m.in. niewłaściwą budowę anatomiczną aparatu głosowego, problemy ze słuchem, a także zaburzenia rozwoju czy schorzenia neurologiczne. Niepożądane zachowania narządu mowy mogą być uwarunkowane m.in..budową fałdów głosowych, których nieregularne drgania – w zależności od rodzaju zaburzenia i lokalizacji choroby – powodują wytwarzanie przez aparat głosowy różnych tonów podstawowych [28].

Wśród patologii mowy można wyróżnić wiele typów zaburzeń. Prozodia jest istotnym elementem głosowej komunikacji. Jest to cecha mowy odpowiadająca za jej brzmienie, czyli m.in. akcent, intonację czy iloczas. Zaburzenia funkcji prozodycznych wiążą się z trudnościami w odbiorze, ekspresji oraz interpretacji wypowiedzi i występują w wielu powszechnych patologiach mowy, takich jak: pragnozja, afazja, dyzartria, apraksja, specyficzne zaburzenia językowe (ang. *specific language impairment* – SLI) czy oligofazja [46]. Należy zauważyć, że niektóre zaburzenia są wprawdzie trudne do zdiagnozowania, ale ich wczesne wykrycie może pomóc w spowolnieniu, a nawet w całkowitym wyeliminowaniu skutków wspomnianych niesprawności. Pomimo że badaniami zaburzeń

mowy zajmują się lekarze różnych specjalności – głównie foniatrzy, ale też m.in. neurolodzy, wykrywanie problemów tego typu może wspomóc diagnostykę.

Metody automatycznej detekcji lub identyfikacji patologii w mowie są obecne w literaturze przedmiotu od kilku dekad. Jednak dopiero pojawienie się algorytmów uczenia głębokiego spowodowało realną możliwość zastosowania tego typu metod w diagnostyce medycznej w celu wsparcia foniatry w praktyce. Istotny jest fakt, iż analiza obiektywna sygnału mowy, jako że jest w dużym stopniu pozbawiona czynnika subiektywnego, może wspomóc lekarzy w podejmowaniu prawidłowej decyzji, w odróżnieniu od pomiaru jakości głosu opartego jedynie na własnym doświadczeniu badającego [28]. Wskazują na to wyniki eksperymentów przedstawianych w literaturze przedmiotu, np. wykorzystanie algorytmów uczenia maszynowego w celu detekcji mowy zaburzonej [7], jak i ogólna dogłębna analiza sygnału mowy [18], [20], [31].

Warto zauważyć, że większość analiz tego typu wywodzi się z obszaru automatycznego rozpoznawania mowy (ang. *automatic speech recognition* – ARS) [5],[10], [18], [19] i pomimo różnic w sygnale mowy zaburzonej i niezaburzonej znajduje szerokie zastosowanie w systemach automatycznego wykrywania wad wymowy. System automatycznego rozpoznawania mowy składa się z dwóch zasadniczych segmentów: ekstrakcji cech i klasyfikacji. Technika ekstrakcji cech odgrywa istotną rolę w rozpoznawaniu mowy, dlatego w kolejnych rozdziałach zostaną przywołane metody analizy sygnału mowy służące do ekstrakcji cech i przygotowania danych do automatycznego rozpoznawania mowy zaburzonej, jak również przykłady baz mowy patologicznej. W podrozdziale 6.4. przedstawione zostaną.wykorzystane algorytmy, wyniki przeprowadzanych analiz, jak również opis opracowanej aplikacji, natomiast w podrozdz. 6.6 wnioski i kierunki rozwoju opisanych badań mowy zaburzonej.

Niniejszy rozdział powstał na podstawie wyników pracy dyplomowej Martyny Włoszczyńskiej [45].

6.2. Analiza sygnału mowy

Analiza mowy – w kontekście ASR – odnosi się przede wszystkim do ekstrahowania z sygnału mowy istotnych informacji, które są wynikiem procesu wytwarzania mowy. Wydobywane cechy mowy czy parametry sygnału, ich liczba oraz jakość jest w pełni uzależniona od celu badania, np. czy dotyczy ono analizy barwy głosu, intonacji, płynności wypowiedzianych zdań. Pełna analiza sygnału mowy obejmuje mechanizm wytwarzania mowy, który jest procesem złożonym [42], [43].

W niniejszym rozdziale przedstawione zostaną pokrótce metody analizy sygnału mowy stosowane w celu przygotowania danych do automatycznego rozpoznawania mowy. U podstaw metod analizy sygnału leży analiza czasowa, widmowa, czasowo-częstotliwościowa, cepstralna, jak również analiza LPC (ang. *linear predictive coding*), cepstralne współczynniki liniowego kodowania predyktywnego (ang. *linear predictive*

cepstral coefficients – LPCC), współczynnik perceptualnej liniowej predykcji (ang. *perceptual linear predictive coefficients* – PLP). Analizy te służą wydobyciu cech sygnału, które składają się na jego charakterystykę.

Zebranie wielu informacji o badanym sygnale może pozwolić np. na rozpoznawanie mowy, identyfikację mówcy, rozpoznanie emocji, transkrypcję mowy na tekst (ang. *speech-to-text* – STT), a także detekcję zmian patologicznych w wypowiedzi mówcy.

Oprócz głośności/intensywności, dźwięczności, charakterystyki widma, czasu trwania fonemów, gęstości przejść przez zero i parametrów standardu MPEG-7 [17], np. środka ciężkości, wyznacza się współczynniki mel-cepstralne [25]. Podstawą wielu prac dotyczących analizy mowy są systemy automatycznego rozpoznawania mowy. Z dostępnych publikacji jednoznacznie wynika, że najbardziej skuteczne rozwiązania opierają się obecnie na reprezentacjach sygnału w dziedzinie czasowo-częstotliwościowej, a najpopularniejsze z nich to te, w których wykorzystywane są spektrogramy oraz spektrogramy w skali melowej [1], [18], [31]. Przykładem zastosowania współczynników MFCC w akustycznej analizie mowy są badania sygnałów mowy u pacjentów z dysfonią spowodowaną obecnością guzków śpiewaczych opisane w pracy [29]. Istnienie większej korelacji między akustycznymi cechami mowy a wektorami MFCC potwierdziła również analiza współczynników MFCC przeprowadzona w pracach [5], [25]. W badaniach tych udowodniono, że istnieje większa korelacja, gdy mierzy się ją w obrębie określonych fonemów, a nie globalnie, we wszystkich dźwiękach mowy.

Obecnie, w celu modelowania cech akustycznych sygnału, w przeważającej liczbie systemów ASR wykorzystywane są głębokie sieci neuronowe (ang. *deep neuronal network* – DNN) [18], [19], [20], [34], [41], co jest spowodowane ich wysoką wydajnością w przetwarzaniu danych wejściowych o wysokiej wymiarowości. Dodatkowo nowoczesne komputery są wyposażone w karty graficzne, co powoduje, że problem obciążenia zasobów w procesie przetwarzania mowy przestaje być istotny [41]. Modele te cechuje możliwość posiadania różnych architektur, nieograniczona liczba warstw ukrytych i zdolność do przetworzenia bardzo dużych zbiorów danych [34], [41]. Obecnie modele głębokie znajdują zastosowanie w rozpoznawaniu emocji w mowie, detekcji patologii aparatu mowy czy identyfikacji mówcy [1], [2].

Pomimo zdolności sieci neuronowych do przetwarzania danych jednowymiarowych (np. sygnału mowy podanego w postaci wektora parametrów na wejście sieci) wykorzystanie dwuwymiarowej reprezentacji danych może znacznie zwiększyć skuteczność algorytmu. Przykładowo, w przypadku sieci splotowych splot dwuwymiarowy jest w stanie wyekstrahować więcej istotnych szczegółów [18]. Taka dwuwymiarowa reprezentacja jednowymiarowej próbki sygnału mowy może zostać uzyskana w wyniku wyodrębnienia konkretnych cech sygnału i przedstawienia ich w dziedzinie czasu. Najprostszym przykładem takiego zabiegu jest spektrogram, który prezentuje widmo danego sygnału w czasie. Ponadto niesie on pełną informację o cechach akustycznych danego sygnału (zarówno statycznych, jak i dynamicznych) [18].

Interesującym przykładem zastosowania spektrogramów w skali melowej jest praca zawierająca omówienie badania kaszlu w celu detekcji choroby COVID-19 [48]. Jest

dowód, że analiza mel-spektrogramów jest bardzo obiecującym narzędziem, mającym wiele wartościowych zastosowań, również w kontekście biomedycznym. Przykładem praktycznego zastosowania współczynników mel-cepstralnych pochodzących z spektrogramów w skali melowej (zamiennie stosowana nazwa mel-spektrogramy) jest stan wiedzy ASR – HMM/GMM (ang. *Hidden Markov Model*/*Gaussian Mixture Model*), czyli technika łącząca ukryty model Markowa oraz model mieszaniny rozkładów Gaussa. Algorytm ten generował cechy o wysokim stopniu nieskorelowania przy jednoczesnym zachowaniu stosunkowo małej wymiarowości [41].

6.3. Ogólnodostępne bazy sygnałów mowy

Bazy mowy zawierające nagrania mowy lub wyekstrahowane cechy z sygnałów, np. współczynniki MFCC, są obecnie ogólnie dostępne w Internecie. Wybrane bazy danych złożone z nagrań wypowiedzi przedstawiono w tabeli 6.1 [2]. Należy zauważyć, że przeważająca liczba zestawów nagrań mowy jest w języku angielskim. Część z nich jest przygotowana do treningu systemów ASR i przeprowadzania badań związanych z automatycznym rozpoznawaniem mowy. Obecnie bazy danych z nagraniami wypowiedzi są przygotowywane w celu wykrywania emocji – wiele dostępnych zestawów jest dodatkowo zaopatrzona w stosowne adnotacje.

Ze względu na szeroki zakres zagadnienia, jakim jest badanie mowy patologicznej, dostępne są zbiory danych prezentujące różnorodne zaburzenia i mające różny charak-

Tabela 6.1. Wybrane bazy danych mowy

Nazwa bazy	Język bazy (przeważający)	Charakterystyka danych			Liczba nagranych osób
		pełne zdania	pojedyncze wyrazy	typ mowy	
IITG-MV SR [14]	angielski	tak	nie	spontaniczna	100
XM2VTS [40]	angielski	nie	nie	wymuszona	295
Brent [9]	angielski	tak	tak	wymuszona	100
SpeechDat [36]	angielski	tak	tak	wymuszona	5120
EUROM-1 [4]	duński	tak	nie	wymuszona	60
TIMIT/NTIMIT [11]	angielski	tak	nie	wymuszona	630
YOHO [47]	angielski	nie	nie	wymuszona	138
Switchboard-1 [38]	angielski	nie	nie	spontaniczna	325
KING-92 [16]	angielski	nie	nie	spontaniczna	51
LLHDB [22]	angielski	tak	nie	spontaniczna	53

Źródło: [2].

ter. Zawierają one nie tylko sygnał mowy, ale również spektrogramy bądź liczbowe parametry akustyczne sygnałów.

Z uwagi na wiele pól zastosowań najbardziej za wartościowe należy uznać bazy z surowymi sygnałami, czyli z nagraniami mowy. Na podstawie danych pochodzących z takich baz można dokonać obróbki sygnału, wygenerować spektrogramy według własnych potrzeb, a także wyznaczyć interesujące parametry akustyczne, które później można wykorzystać w analizie. Znaczącą wadą baz dźwiękowych jest ich słaba jakość.

Dotyczy to wielu wartościowych zestawów danych, w tym również tych powstałych w ubiegłym wieku.

Przykładem bazy mowy zaburzonej składającej się z nagrań wypowiedzi pacjentów z zaburzeniami mowy jest DementiaBank [27]. Baza ta powstała w 2005 r. i zawiera około 500 próbek mowy w języku angielskim. Grupę badanych tworzyło 188 osób cierpiących na demencję oraz 99 osób z grupy kontrolnej (wszyscy w wieku 45–90 lat). Nagrane wypowiedzi pacjentów składają się nie tylko z pojedynczych wyrazów, ale również z pełnych zdań [27]. Innym zestawem danych mowy patologicznej jest PC-GITA. Baza ta została utworzona w 2014 r., a zgromadzono w niej nagrania mowy osób hiszpańskojęzycznych, cierpiących na chorobę Parkinsona [30], wśród których było 50 osób z zaburzeniami oraz 50 osób zdrowych (w ramach grupy kontrolnej) w wieku 33–77 lat. W sumie zebrano 3000 próbek mowy osób z chorobą Parkinsona oraz 3000 nagrań mowy osób z grupy kontrolnej. Badani zostali poproszeni o wypowiedzenie pełnych zdań, pojedynczych słów, a także samogłosek [30]. Zbiór Saarbrücken Voice Database zawiera nagrania zebrane od 1002 mówców z 71 różnymi zaburzeniami mowy (m.in. oznaczonymi jako porażenie fałdu głosowego czy dysfonia hiperfunkcjonalna). W przy-

Tabela 6.2. Zestawienie wybranych baz danych mowy patologicznej

	PC-GITA	DementiaBank	Saarbrücken Voice Database
Rok powstania	2014	2005	2007
Język bazy	hiszpański	angielski	niemiecki
Liczba próbek	choroba Parkinsona: 3000 grupa kontrolna: 3000	demencja: 255 grupa kontrolna: 242	zaburzenia: 1002 grupa kontrolna: 3000
Rodzaj próbek	pełne zdania, pojedyncze słowa, długie samogłoski	pełne zdania, pojedyncze słowa	pełne zdanie, długie samogłoski
Liczba badanych osób	choroba Parkinsona: 50 grupa kontrolna: 50	demencja: 188 grupa kontrolna: 99	zaburzenia: 1002 grupa kontrolna: 851
Zróżnicowanie grupy badanych	wiek: 33–77 lat płeć: 25 kobiet i 25 mężczyzn	wiek: 45–90 lat płeć: 343 kobiety i 208 mężczyzn	wiek: 6–94 lat
Typ danych	audio	audio	audio

Źródło: [26], [27], [30].

padku tego zestawu danych wiek mówców mieścił się w przedziale 6–94 roku życia, a w sumie zebrano 2225 nagrań. Każda sesja nagraniowa zawierała nagrania samogłosek w różnych wysokościach dźwięku. Ponadto badani zostali poproszeni o wypowiedzenie krótkiej frazy: Guten Morgen, wie geht es Ihnen? [26].

Z innych baz można wymienić MDVR-KCL [15], Italian Parkinson's Voice and Speech database [4]) czy zestaw pochodzący z korpusu TalkBank [35], [39]. Jak można zauważyć, wśród wymienionych baz brakuje nagrań mowy patologicznej w języku polskim. Przykładowe bazy danych są wymienione w tabeli 6.2.

6.4. Analiza mowy patologicznej

Na podstawie przeglądu literatury przedmiotu w zakresie automatycznego rozpoznawania patologii w mowie można zauważyć, *że* częstym podejściem jest wykorzystanie reprezentacji sygnału mowy jako mapy wyekstrahowanych cech akustycznych (m.in. MFCC) oraz reprezentacji dwuwymiarowych (np. spektrogramów czy spektrogramów w skali melowej) [18], [41]. W aktualnych badaniach nad wykrywaniem patologii mowy wykorzystywane są obok algorytmów klasycznych także architektury modeli głębokich sieci neuronowych [10], [18]. Z kolei zaburzeniami najczęściej poddawanymi analizie są zaburzenia związane z chorob*ą* Alzheimera i z chorob*ą* Parkinsona [3], [23], [24], [30].

W ramach prowadzonych analiz zaproponowano różne konfiguracje sieci neuronowych oraz parametryzacje sygnału mowy. Wykorzystano sieci splotowe oraz wektor cech zawierający współczynniki mel-cepstralne (ang. *Mel Frequency Cepstrum Coefficients* – MFCCs), jak również reprezentacje 2D, tj. spektrogramy i spektrogramy w skali melowej. Trening modeli został oparty na trzech bazach danych: PC-GITA (nagrania mowy w języku hiszpańskim osób z chorobą Parkinsona), ADReSSo (zestaw próbek mowy pacjentów cierpiących na demencję wywołaną chorobą Alzheimera) oraz SVD (Saarbrücken Voice Database; zawierającą 71 klas zaburzeń mowy).

Tabela 6.3. Podsumowanie liczbowe danych w bazach mowy zaburzonej

Baza danych	Liczba nagrań mowy		Łączna liczba nagrań w bazie
	zaburzonej	niezaburzonej	
ADReSSo	1476	1358	2834
PC-GITA	1150	1150	2300
Saarbrücken Voice Database	750	750	1500
ADReSSo, PC-GITA, Saarbrücken Voice Database	3376	3258	6634

Źródło: [26], [27], [30].

Należy zauważyć, że dane w wybranych bazach zostały podzielone na dwie klasy: próbki mowy osób z zaburzeniami (ang. *pathological*) oraz sygnały pochodzące z nagrań w grupie kontrolnej (ang. *healthy*). Rozkład liczbowy w poszczególnych zestawach próbek przedstawiono w tabeli 6.3.

Dane te należało poddać wstępnemu przetworzeniu i wydzielić spośród nich trzy zbiory: treningowy, testowy i walidacyjny. Przetworzenie wstępne danych polegało na wyekstrahowaniu cech akustycznych i spektrogramów za pomocą odpowiednich bibliotek.

Kolejnym etapem był wybór algorytmów uczenia głębokiego. Ze względu na wysoką skuteczność uzyskiwaną przez modele głębokie również w kontekście klasyfikacji mowy zaburzonej zdecydowano się na zastosowanie sieci splotowej (CNN) w dwóch konfiguracjach:

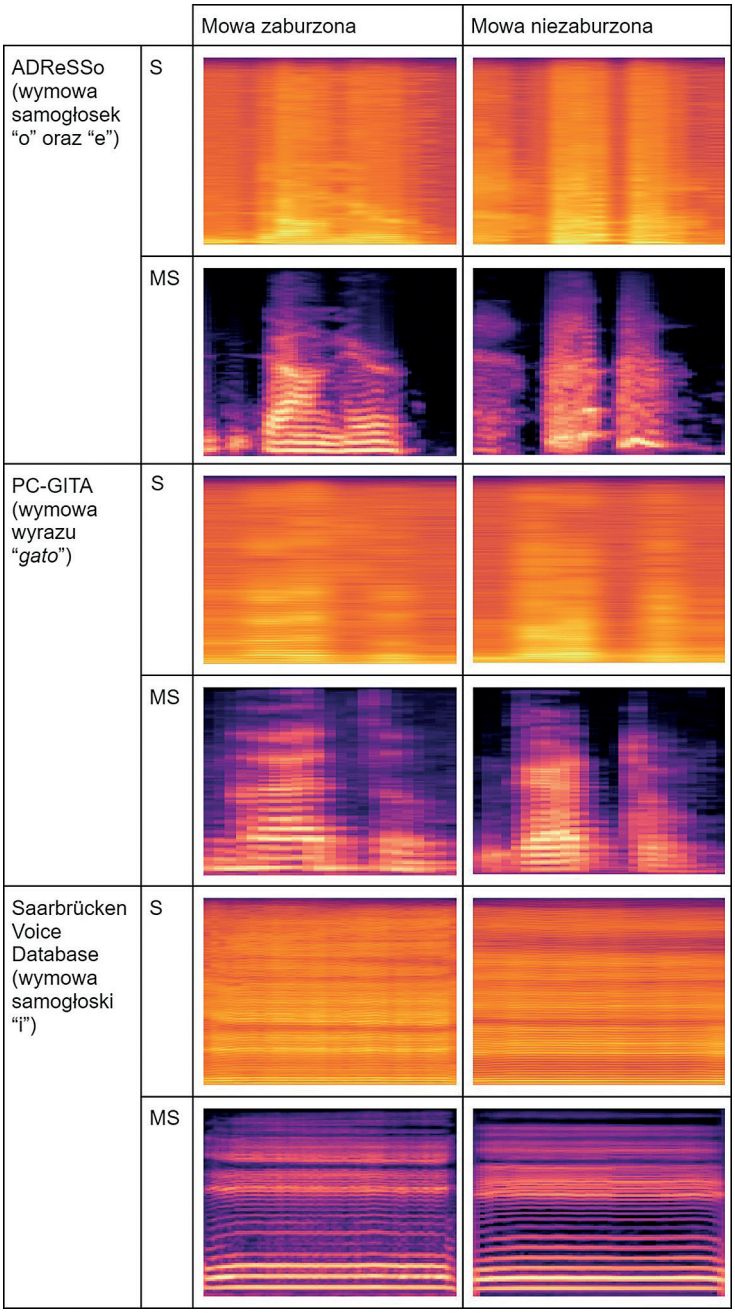
- sieci splotowej przyjmującej na wejściu zbiór wyekstrahowanych cech akustycznych z sygnału;
- sieci splotowej przyjmującej na wejściu spektrogramy oraz spektrogramy w skali melowej.

W strukturze sieci splotowej można wyróżnić takie warstwy, jak warstwa [33]:

- splotowa – splata dane wejściowe i przekazuje wynik do następnej warstwy;
- *łącząca* (ang. *pooling*) – jej zadaniem jest zmniejszenie danych wejściowych w celu zredukowania obciążenia obliczeniowego, wykorzystania pamięci i liczby parametrów;
- *maxpooling* (ang.) – przesuwa filtry NxN przez całą macierz, pobierając największą wartość z okna filtra i zapisuje ją do kolejnej mapy (gromadzenie danych za pomocą funkcji agregacyjnej – w tym przypadku maksymalizującej, czyli jedynie maksymalna wartość z każdego jądra zostaje przekazana do następnej warstwy);
- *flatten* (ang.) – spłaszcza macierze do jednego wektora;
- *dropout* (ang.) – ustawia losowo wychodzące krawędzie neuronów tworzących ukryte warstwy na 0 przy każdej aktualizacji fazy treningu;
- *dense* (ang.) – łączy wszystkie neurony z warstwy wcześniejszej z neuronami warstwy następnej;
- filtr (ang. *kernel*) – jest zbiorem wag (ma postać małej, kwadratowej macierzy o wymiarach $n \times n$);
- rozmiar jądra/rozmiar filtra (ang. *kernel size*);
- wypełnienie (ang. *padding*);
- *batch normalization* (ang.) – normalizuje wejścia warstw poprzez operacje wyśrodkowania i skalowania.

W strukturze sieci splotowych można wyróżnić trzy główne typy warstw: wejściową (zwykłą warstwę splotową), łączącą (ang. *pooling*) i w pełni połączoną. Wyjście sieci jest typową warstwą w pełni połączoną.

Zaimplementowane splotowe sieci neuronowe zostały następnie poddane procesowi trenowania na wcześniej przygotowanych zestawach danych. Po przeprowadzeniu serii



Rys. 6.1. Spektrogramy i mel-spektrogramy pochodzące z baz ADReSSo, PC-GITA i Saarbrücken Voice Database; S – reprezentacja w formie spektrogramu, MS – spektrogram w skali melowej

treningów uzyskane modele sieci neuronowej zostały przetestowane na zbiorach walidacyjnych oraz testowych. Wyniki procesu testowania zestawiono również z rezultatami podanymi w literaturze przedmiotu, aby porównać skuteczność algorytmów.

W celu znalezienia najwyższej skuteczności algorytmu przyjmującego na wejściu dane dwuwymiarowe postanowiono przeprowadzić dwie konfiguracje treningu: jedną na danych zawierających tylko spektrogramy, a drugą na danych zawierających mel-spektrogramy (patrz rys. 6.1).

W porównaniu reprezentacji 2D, tj. spektrogramów i mel-spektrogramów w nagraniach osób zdrowych i chorych, można zauważyć, że w przypadku mowy kontrolnej granice wykresu widma są wyraźniejsze.

Można też zauważyć, że na spektrogramach w skali melowej widocznych jest znacznie więcej różnic między mową zaburzoną a niezaburzoną. Taka obserwacja już w momencie przygotowania danych może być podstawą do stwierdzenia, że trening algorytmu na zestawie zawierającym mel-spektrogramy może dać lepsze rezultaty niż w przypadku spektrogramów.

Przygotowanie danych obejmowało też augmentację zbiorów podawanych na wejście sieci. W tym celu posłużono się białym szumem dodawanym do sygnału przy różnych wartościach stosunku sygnału do szumu (ang. *signal-to-noise-ratio* – SNR).

W celu oceny wyników przeprowadzonych analiz wykorzystano miary ilościowe, takie jak: dokładność, precyzja, czułość, swoistość i *F*1-score [12] – wzory (6.1)–(6.5).

$$dokładność = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.1}$$

gdzie:

TP – wartości prawdziwie pozytywne,
TN – wartości prawdziwie negatywne,
FP – wartości fałszywie pozytywne,
FN – wartości fałszywie negatywne.

$$precyzja = \frac{TP}{TP + FP} \tag{6.2}$$

$$czułość = \frac{TP}{TP + FN} \tag{6.3}$$

$$swoistość = \frac{TN}{TN + FP} \tag{6.4}$$

$$F1 = 2 \cdot \frac{precyzja \cdot czułość}{precyzja + czułość} \tag{6.5}$$

Szczegółowe wyniki przedstawiające wyniki ewaluacji na zbiorze testowym (z dokładnością do pierwszego miejsca po przecinku) zostały przedstawione w tabelach 6.4 i 6.5.

Tabela 6.4. Przedstawienie wyników eksperymentów związanych z badaniem spektrogramów i mel-spektrogramów

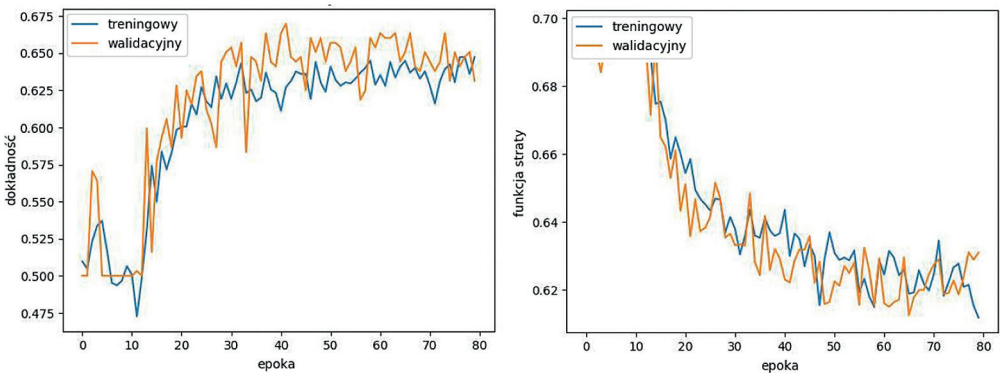
Rodzaj analizowanej reprezentacji sygnału	Baza	Dokładność (na zbiorze testowym) [%]
Spektrogram	PC-GITA	63,1
	ADReSSo	62,4
	Saarbrücken Voice Database	65,7
Mel-spektrogram	PC-GITA	70,4
	ADReSSo	67,4
	Saarbrücken Voice Database	68

Tabela 6.5. Przedstawienie wyników eksperymentów związanych z badaniem spektrogramów i mel-spektrogramów

Baza danych	Dokładność (na zbiorze testowym) [%]
PC-GITA	64,4
ADReSSo	61
SVD	62,9

Na podstawie uzyskanych wyników można zauważyć, że w przypadku:

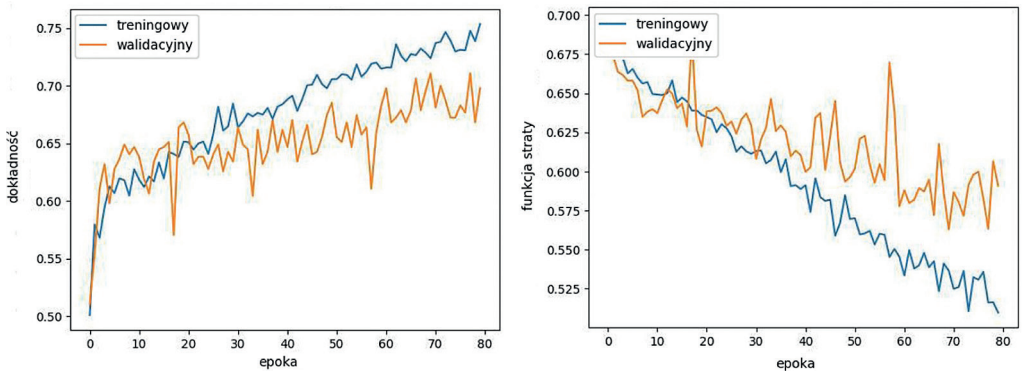
- algorytmu badającego spektrogramy i mel-spektrogramy najbardziej skuteczne okazały się konfiguracje:



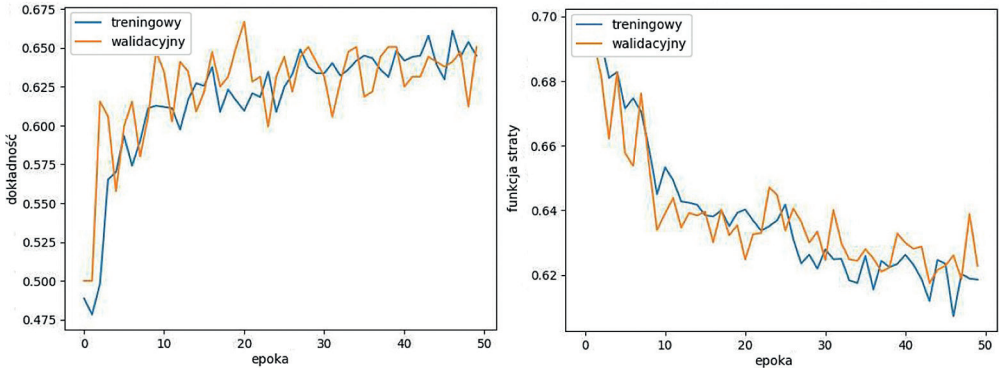
Rys. 6.2. Wykresy dokładności i funkcji straty modelu po treningu przeprowadzonym na spektrogramach wygenerowanych z bazy Saarbrücken Voice Database; linia niebieska – wyniki osiągane w zbiorze treningowym, linia pomarańczowa – wyniki osiągane w zbiorze walidacyjnym

  - badania spektrogramów – najlepszy wynik osiągnięto po treningu na bazie SVD (dokładność 65,7%),
  - badania mel-spektrogramów – najlepszy wynik osiągnięto po treningu na bazie PC-GITA (dokładność 70,4%);
- algorytmu badającego cechy akustyczne najlepsze rezultaty osiągnięto za pomocą konfiguracji treningu na bazie PC-GITA, zaś dokładność wyniosła 64,4%.

W celu zwizualizowania przebiegu treningu w przypadku tych konfiguracji, które osiągnęły najwyższe miary dokładności, na rys. 6.2–6.4 przedstawiono wykresy zmian dokładności i funkcji straty wraz z każdą kolejną epoką.



Rys. 6.3. Wykresy dokładności i funkcji straty modelu po treningu przeprowadzonym na spektrogramach wygenerowanych z bazy PC-GITA



Rys. 6.4. Wykresy dokładności i funkcji straty modelu po treningu przeprowadzonym na współczynnikach MFCC wyekstrahowanych z bazy PC-GITA

W tabelach 6.6–6.8 podano macierze pomyłek dla najlepszych konfiguracji przeprowadzanych eksperymentów, natomiast w tabeli 6.9 wyniki z raportu klasyfikacyjnego, czyli rezultaty dla miar ilościowych.

Tabela 6.6. Macierz pomyłek algorytmu badającego spektrogramy po treningu na bazie SVD

		Klasa rzeczywista	
		mowa zaburzona	mowa prawidłowa
Klasa przewidywana	mowa zaburzona	85	49
	mowa prawidłowa	54	112

Tabela 6.7. Macierz pomyłek algorytmu badającego mel-spektrogramy po treningu na bazie PC-GITA

		Klasa rzeczywista	
		mowa zaburzona	mowa prawidłowa
Klasa przewidywana	mowa zaburzona	170	81
	mowa prawidłowa	55	154

Tabela 6.8. Macierz pomyłek algorytmu badającego współczynniki MFCC po treningu na bazie PC-GITA

		Klasa rzeczywista	
		mowa zaburzona	mowa prawidłowa
Klasa przewidywana	mowa zaburzona	151	90
	mowa prawidłowa	74	145

Tabela 6.9. Zestawienie wyników miar ilościowych dla trzech algorytmów osiągających najlepsze rezultaty

	Miary ilościowe [%]				
Metoda	czułość	swoistość	precyzja	F1 Score	dokładność
Algorytm CNN klasyfikujący spektrogramy	61,15	69,57	63,43	62,27	65,67
Algorytm CNN klasyfikujący mel-spektrogramy	75,56	65,53	67,73	71,43	70,43
Algorytm DNN klasyfikujący współczynniki MFCC	67,11	61,7	62,66	64,81	64,35

W dostępnej literaturze przedmiotu można znaleźć wiele publikacji, w których autorzy postawili sobie za cel wykorzystanie parametrów akustycznych sygnału do kla-

syfikacji mowy zaburzonej. Z przykładowego zestawienia w formie tabeli 6.10 można wnioskować, że wyniki uzyskane przez tę grupę badaczy wyniki uzyskane przez autorów niniejszego rozdziału nie odbiegają znacznie od siebie. Powodem jest zastosowanie bardzo podobnej metody przygotowania i klasyfikacji danych.

Tabela 6.10. Wyniki przeprowadzonych badań związanych z klasyfikacją parametrów akustycznych sygnałów mowy z rezultatami z literatury przedmiotu

Rodzaj algorytmu	Baza danych	Dokładność [%]
DNN	PC-GITA	64,4
DNN	ADReSSo	61
DNN	SVD	62,9
SVM [32]	PC-GITA	69,2
SVM [23], [24]	ADReSSo	64,8
CNN+LSTM [13]	SVD	68,1

Na podstawie wyżej wspomnianych wyników podanych w literaturze przedmiotu można wnioskować, że wykorzystanie reprezentacji dwuwymiarowych w zadaniu detekcji mowy zaburzonej sprawdza się bardzo dobrze i spełnia swoją rolę lepiej niż analiza za pomocą parametrów akustycznych (współczynników mel-cepstralnych).

Tabela 6.11. Wyniki przeprowadzonych badań związanych z klasyfikacją spektrogramów i mel-spektrogramów sygnałów mowy oraz wyniki przedstawione w literaturze przedmiotu

Rodzaj algorytmu	Stosowane podejście	Baza danych	Dokładność [%]
CNN	analiza spektrogramu	PC-GITA	63,1
CNN	analiza spektrogramu	ADReSSo	62,4
CNN	analiza spektrogramu	SVD	65,7
CNN	analiza mel-spektrogramu	PC-GITA	70,4
CNN	analiza mel-spektrogramu	ADReSSo	67,4
CNN	analiza mel-spektrogramu	SVD	68
ResNet-101+LSTM [10]	analiza mel-spektrogramu	PC-GITA	98,61
DNN [5]	analiza spektrogramu	DementiaBank	93,3
CNN [26]	analiza spektrogramu	SVD	64

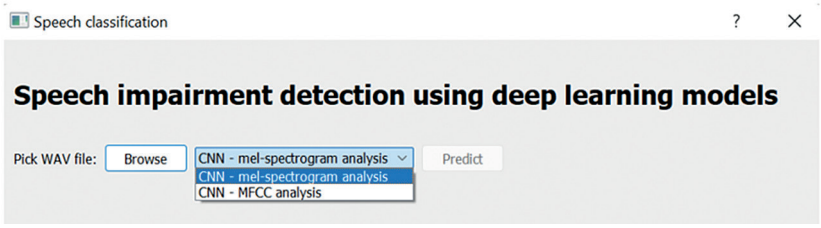
Podsumowując przeprowadzone analizy, można stwierdzić, że możliwe jest zastosowanie parametrów akustycznych wyekstrahowanych z sygnału w celu detekcji mowy zaburzonej. Jednak lepiej sprawdza się w takim zadaniu reprezentacja dwuwymiarowa,

tj. spektrogramy i spektrogramy w skali melowej niż wektor współczynników mel-cepstralnych.

6.5. Projekt aplikacji

Następnym etapem pracy był projekt aplikacji demonstracyjnej, umożliwiającej klasyfikację załadowanego przez użytkownika sygnału w formacie plików dźwiękowych oferujących bezstratną jakość dźwięku (ang. *waveform audio format* – WAV). Projekt aplikacji demonstracyjnej ma na celu umożliwienie użytkownikowi przeprowadzenie detekcji zaburzenia w sygnale fonicznym i otrzymania wyniku klasyfikacji binarnej (mowa zaburzona/niezaburzona).

Aplikacja została przygotowana w języku Python za pomocą biblioteki PyQt5 [37], która pozwala na tworzenie prostych aplikacji okienkowych zawierających określone komponenty. W tle przygotowanej aplikacji pracuje algorytm klasyfikacji sygnału mowy z możliwością wyboru reprezentacji danych przez użytkownika. W celu wyodrębnienia map cech wykorzystano ogólnodostępne narzędzie *openSMILE* [8]. Wstępne przetworzenie zestawu cech akustycznych wyekstrahowanych z sygnału mowy wymagało zastosowania standardowej techniki z pakietu *sklearn*, czyli *StandardScaler* oraz PCA [33]. W ekstrakcji cech



Rys. 6.5. Widok aplikacji z wybranym algorytmem do binarnej klasyfikacji zaburzeń mowy

Rys. 6.6. Widok aplikacji po przeprowadzeniu predykcji na załadowanym pliku fonicznym zawierającym próbkę mowy zaburzonej

z sygnału wykorzystano zestawu cech ComParE [44]. Z kolei postać dwuwymiarową danych generowano za pomocą biblioteki *librosa*.

Na podstawie uzyskanych w eksperymentach wartości dokładności oraz pozostałych miar ilościowych dokonano wyboru algorytmu pracującego w tle aplikacji. Jest to sieć splotowa z możliwością wykorzystania spektrogramów, spektrogramów w skali melowej lub wektora współczynników MFCCs, obliczanych na wejściu modelu sieci.

Na rysunku 6.5 przedstawiono widok aplikacji z wybranym modelem klasyfikacji, a na rys. 6.6 widok aplikacji po przeprowadzeniu predykcji na załadowanym pliku fonicznym zawierającym próbkę mowy zaburzonej.

6.6. Podsumowanie

Automatyczna detekcja/identyfikacja zaburzeń mowy/wymowy może być wykonana z wystarczającą skutecznością. Porównanie reprezentacji danych, na których przeprowadzono trening modeli, pozwoliło na stwierdzenie, że analizowanie spektrogramów w skali melowej jest właściwym podejściem w procesie wykrywania zmian patologicznych w głosie i pozwala na uzyskanie większej dokładności niż zastosowanie współczynników mel-cepstralnych lub wykorzystanie spektrogramów w konfiguracji sieci splotowej. Uzyskane wyniki są pod względem dokładności porównywalne z wynikami prezentowanymi w literaturze przedmiotu, tj. zawierają się (średnio) w przedziale 61–71% w przypadku badań własnych oraz w przedziale 64–98% w przypadku danych w literaturze przedmiotu w zależności od zastosowanych algorytmów i baz danych.

Głównym problemem, który się pojawia w realizacji tego typu badań, jest brak baz danych zawierających zbalansowane klasy zaburzeń, charakteryzujących się podobną ilością danych przypisanych do danej klasy, jak również poprawną adnotacją odnoszącą się do poszczególnych problemów wymowy/wad mowy. Jest nim również stosowanie różnych oznaczeń klas zaburzeń przez różnych specjalistów w procesie tworzenia tego typu zbiorów danych. Dlatego główne działania pozwalające na rozwój badań, takich jak te opisane w niniejszym rozdziale, należałoby skierować na stworzenie bazy wieloklasowej z danymi równolicznymi i z ujednoliconą adnotacją.

Bibliografia

[1] Badshah A., Ahmad J., Rahim N., Baik S., *Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network*, International Conference on Platform Technology and Service, 2017, s. 1–5.

[2] Barai B., *VQ/GMM Based Speaker Identification with Emphasis on Language Dependency*, 2018.

[3] Bertini F., Allevi D., Lutero G., Calzà L., Montesi D., *An automatic Alzheimer's disease classifier based on spontaneous spoken English*, „Computer Speech Language" 2022, Vol. 72, s. 101298; https://www.sciencedirect.com/science/article/pii/S0885230821000991

[4] Chan D., Fourcin A., Gibbon D., Grandstrom B., Huckvale M., Kokkinakis G., Kvale K., Lamel L., Linberg B., Moreno A., Mouropoulos J., Senia F., Trancoso I., Veld C., Zeiliger J., *EUROM – A spoken language resource for the EU*, 4th European Conference on Speech Communication and Technology, EUROSPEECH 1995, Madrid, Spain, September 18-21, 1995, s 1995–1998.

[5] Darch J., Milner B., Vaseghi S., *Analysis and Prediction of Acoustic Speech Features from Mel-frequency Cepstral Coefficients in Distributed Speech Recognition Architectures*, „The Journal of the Acoustical Society of America" 2009, Vol. 124, No. 6, s. 3989–4000.

[6] Dimauro G., Girardi F., *Italian Parkinson's Voice and Speech*, 2019; https://dx.doi.org/10.21227/aw6b--tg17

[7] Er M.B., Isik E., Isik I., *Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition*, „Biomedical Signal Processing and Control" 2021, Vol. 70, s. 103006; https://www.sciencedirect.com/science/article/pii/S1746809421006030

[8] Eyben F., Wöllmer M., Schuller B., *Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor*, Proceedings of the 18th ACM International Conference on Multimedia. MM '10. Firenze, Italy: Association for Computing Machinery, 2010, s. 1459–1462; https://doi.org/10.1145/1873951.1874246

[9] Fletcher P., Garman M., https:// childestalkbank. org/derived/ [dostęp: 15.11.2022].

[10] Gevaert W., Tsenov G., Mladenov V., *Neural networks used for speech recognition*, „Journal of Automatic Control" 2010, Vol. 20, No. 1.

[11] Glackin C., Wall J., Chollet G., Dugan N., Cannings N., *TIMIT and NTIMIT Phone Recognition Using Convolutional Neural Networks*, 2019.

[12] Goutte C., Gaussier E., *A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation*, w: Advances in Information Retrieval, D.E. Losada, J. Fernandez-Luna (eds.), seria: „Lecture Notes in Computer Science" 2005, Vol. 3408.

[13] Haràr P., Alonso J., Mekyska J., Galáž Z., Burget R., Smekal Z., *Voice Pathology Detection Using Deep Learning: a Preliminary Study*, 2017, s. 1–4.

[14] *IITG multivariability speaker recognition database*; https://iitg.ac.in/eee/emstlab/SRdatabase/introduction.php [dostęp: 15.11.2022].

[15] Jaeger H., Trivedi D., Stadtschnitzer M. *Mobile Device Voice Recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls*, „Zenodo" 2019; https://doi.org/ 10.5281/zenodo.2867216

[16] *King speaker verification*; https://catalog.ldc.upenn. edu/LDC95S22 [dostęp: 15.11.2022].

[17] Koenen R., Pereira F., *MPEG-7: A standardized description of audiovisual content*, „Signal Processing: Image Communication" 2000, Vol. 16, No. 1–2, s. 5–13.

[18] Korvel G., Treigys P., Tamulevicus G., Bernataviciene J., Kostek B., *Analysis of 2D Feature Spaces for Deep Learning-based Speech Recognition*, „Journal of the Audio Engineering Society" 2018, Vol. 66, No. 12, s. 1072– 1081.

[19] Korzekwa D., Lorenzo-Trueba J., Drugman T., Calamaro S., Kostek B., *Weakly-Supervised Word-Level Pronunciation Error Detection in Non-Native English Speech*, Proceedings: Interspeech 2021, s. 4408–4412.

[20] Korzekwa D., Barra-Chicote R., Zaporowski S., Beringer G., Lorenzo-Trueba J., Serafinowicz A., Droppo J., Drugmant T., Kostek B., *Detection of Lexical Stress Errors in Non-Native (L2) English with Data Augmentation and Attention*, Proceedings: Interspeech 2021, s. 3915–3919.

[21] Kukhar D.T., Lutak K., *Stochastic modeling and data analysis*, 2020.

[22] LLHDB; https://catalog.ldc.upenn.edu/LDC98S68 [dostęp: 15.11.2022].

[23] Luz S., Haider F., de la Fuente S., Fromm D., MacWhinney B., *Detecting cognitive decline using speech only: The ADReSSO Challenge*, medRxiv 2021; https://www.medrxiv.org/content/early/2021/03/27/2021.03.24. 21254263

[24] Luz S., Haider F., de la Fuente S., Fromm D., MacWhinney B., *Alzheimer's Dementia Recognition through Spontaneous Speech: The ADReSS Challenge;* https://arxiv.org/abs/2004.06833 [dostęp: 15.11.2022].

[25] Mermelstein P., *Distance measures for speech recognition, psychological and instrumental,* w: Pattern Recognition and Artificial Intelligence, C.H. Chen (ed.), Academic, New York, 1976, s. 374–388.

[26] Martinez D., Leida E., Ortega A., Miguel A., Villalba J., *Voice Pathology Detection on the Saarbrücken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit*, Proceedings: IberSPEECH 2012, Vol. 328, s. 99–109.

[27] Morheidari B., Pan Y., Walker T., Reuber M., Venneri A., *Detecting Alzheimer's Disease by estimating attention and elicitation path through the alignment of spoken picture descriptions with the picture prompt*, 2019.

[28] Al-Nasheri A., Muhammad G., Alsulaiman M., Ali Z., Mesallam T. A., Faharat M., *An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification*, „Journal of Voice" 2017, Vol. 31, No. 1, s. 113.e9– 113.e18.

[29] Niebudek-Bogusz E., Grygiel J., Strumiłł P., Śliwińska-Kowalska M., *Zastosowanie analizy kepstralnej w ocenie akustycznej głosu u pacjentów z guzkami głosowymi*, „Otorynolaryngologia" 2011, Vol. 64, No. 1, s. 176–181.

[30] Orozco J.R., Arias-Londoñ D., Vargas-Bonilla J., González-Rátiva M., Noethe E., *New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease*, 2014.

[31] Özseven T., *Investigation of the effect of spectrogram images and different texture analysis methods on speech emotion recognition*, „Applied Acoustics" 2018, Vol. 142, s. 70–77.

[32] Pah N.D., Motin M.A., Kumar D.K., *Phonemes based detection of Parkinson's disease for telehealth applications*, „Scientific Reports" 2022, Vol. 12, No. 1, s. 9687.

[33] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., *Scikit-learn: Machine Learning in Python*, „Journal of Machine Learning Research" 2011, Vol. 12, s. 2825–2830.

[34] Roger V., Farinas J., Pinquier J., *Deep neural networks for automatic speech processing: a survey from large corpora to limited data*, „EURASIP Journal on Audio Speech, and Music Processing" 2022, No.1.

[35] Sarawgi U., Zulfikar W., Soliman N., Maes P., *Multimodal Inductive Transfer Learning for Detection of Alzheimer's Dementia and its Severity*, 2020; https://arxiv.org/abs/2009.00700

[36] *SpeechDat-I*; https://www.phonetik.uni-muenchen.de/forschung/BITS/TP1/Cookbook/node187.html [dostęp: 15.11.2022].

[37] Summerfield M., *Rapid GUI Programming with Python and Qt: the Definitive Guide to PyQt Programming*, 2007; http://proquest.safaribooksonline.com/book/programming/ python/9780132354189

[38] *Switchboard-1 release 2*; https://catalog.ldc.upenn.edu/LDC97S62 [dostęp: 15.11.2022].

[39] TalkBank; https://dementia.talkbank.org/ [dostęp: 15.11.2022].

[40] *The XM2VTS database*; http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/ [dostęp: 15.11.2022].

[41] Toledano D.T., Fernández-Gallego M.P., Lozano-Diez A., *Multi-resolution speech analysis for automatic speech recognition using deep neural networks: Experiments on TIMIT*, „PLoS One" 2018, Vol. 13, No. 1, s. 1–24; https://doi. org/10.1371/journal.pone.0205355

[42] Toledano D.T., Gonzalez-Dominguez J., González-Rodrĭguez J., *Speech Analysis*, w: Encyclopedia of Biometrics, S.Z. Li, A. Jain (eds.), Springer, Boston, 2009.
[43] Toledano D.T., Ramos D., Gonzalez-Dominguez J., González-Rodrĭguez J., *Speech Analysis*, w: Encyclopedia of Biometrics, S.Z. Li, A. Jain (eds.), Springer, Boston, 2009.
[44] Weninger F., Eyben F., Schuller B.W., Mortillaro M. Scherer K.R., *On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common*, „Frontiers in Psychology" 2013, Vol. 4, s. 292.
[45] Włoszczyńska M., *Opracowanie algorytmu uczenia głębokiego do detekcji mowy patologicznej*, praca dyplomowa, Wydział ETI, Politechnika Gdańska, 2022.
[46] Wysocka M., *Narzędzia badawcze do oceny prozodii mowy*, „Nowa Audiofonologia" 2015, vol. 4, nr 4, s. 20–27.
[47] *Yoho speaker verification*; https://catalog.ldc.upenn.edu/LDC94S16 [dostęp: 15.11.2022].
[48] Zhou Q., Shan J., Ding W., Wang C., Yuan S., Sun F., Li H., Fang B., *Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network*, „Frontiers in Robotics and AI" 2021, Vol. 8.

Słowa kluczowe: mowa patologiczna, automatyczna detekcja/ zaburzeń mowy/y, parametryzacja sygnału mowy, sieć splotowa.

Automatyczna klasyfikacja mowy patologicznej

Aplikacja przedstawiona w niniejszym rozdziale służy do automatycznego wykrywania mowy patologicznej na podstawie bazy nagrań. W pierwszej kolejności przedstawiono założenia leżące u podstaw przeprowadzonych badan wraz z wyborem bazy mowy patologicznej. Zaprezentowano również zastosowane algorytmy oraz cechy sygnału mowy, które pozwalają odróżnić mowę niezaburzoną od mowy patologicznej. Wytrenowane sieci neuronowe zostały następnie wykorzystane w aplikacji, która umożliwia przeprowadzenie klasyfikacji binarnej na sygnale mowy. Uzyskane wyniki klasyfikacji mowy niezaburzonej i patologicznej zostały porównane z wynikami opisanymi w literaturze przedmiotu. W podsumowaniu zamieszczono również wnioski oraz propozycje rozwoju prowadzonych badań.

Automatic classification of pathological speech

The purpose of this paper is to present an application that is used to automatically detect pathological speech based on an annotated database containing recorded words and sentences. The background of the study belonging to human-computer communication, a critical review of the literature, the algorithms used, and the features of the speech signal that discern between undisturbed and pathological speech are presented first. The assumptions underlying the experiments conducted are also given, along with the selection of the pathological speech base. In the next step, the neural network architecture and its parameters for speech type classification are proposed, and the speech signal preprocessing is presented. The obtained results of the classification of undisturbed and pathological speech are compared with literature sources. The trained neural networks are then used in an application to classify the speech signal. The summary of the experiments also includes conclusions and suggestions for the development of this research study.

7. Uproszczona metoda pomiaru przestrzennych odpowiedzi impulsowych i jej wykorzystanie do oceny jakości akustycznej pomieszczeń i generowania pogłosu surround

Witold Mickiewicz, Grzegorz Pawełkiewicz, Kaja Kosmenda

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie,
Wydział Elektryczny,
al. Piastów 17, 70-310 Szczecin

7.1. Wprowadzenie

Zapewnienie właściwości akustycznych pomieszczeń na możliwie najwyższym poziomie to zadanie dotyczące nie tylko pomieszczeń o tzw. akustyce kwalifikowanej. Świadomość konieczności zadbania o odpowiednie parametry akustyczne w salach koncertowych czy nagraniowych, w których rozchodzenie się dźwięku wpływa na jakość percepcji słuchowej muzyków i słuchaczy czy na jakość nagrań, jest wprawdzie powszechna [14], [15], jednak modelowanie zjawisk wpływających na odbiór dźwięku w tego typu obiektach wciąż nie jest w pełni poprawne, a projektowanie m.in. sal koncertowych uważa się za sztukę. W przypadku sal kongresowych, wykładowych czy lekcyjnych, w których największą rolę odgrywa komunikacja werbalna, pojawia się ze względów ekonomicznych problem zastępowania rozwiązań architektonicznych systemami elektroakustycznymi, mającymi być lekarstwem na całe zło [2], [16]. Zapomina się często o tzw. hałasie pogłosowym, powodującym, że nie tylko ulega obniżeniu funkcjonalność tego rodzaju pomieszczeń, ale również – mimo zachowania odpowiedniej zrozumiałości mowy dla systemu nagłośnieniowego – dłuższe przebywanie w nich jest męczące. Dlatego badanie akustyki pomieszczeń już istniejących w celu eliminacji wad akustycznych oraz modelowanie propagacji dźwięku w pomieszczeniach projektowanych, w których na propagację dźwięku wpływa wiele detali architektonicznych, jest ciągle aktualny i ważny.

W polu akustycznym zachodzą zjawiska fizyczne, które chcemy obserwować, związane z propagacją fal akustycznych, ich odbiciem i załamaniem. W akustyce wnętrz propagacja odbywa się w ośrodku sprężystym, którym najczęściej jest powietrze, powodując chwilowe i lokalne zaburzenia cząstek wchodzących w jego skład. Podstawowymi parametrami tych cząstek akustycznych jest ciśnienie akustyczne i prędkość akustyczna. Ciśnienie jest łatwe w pomiarze i dlatego we współczesnej akustyce technicznej większość parametrów bazuje na pomiarze tej wielkości, która jest wielkością skalarną i nie zawiera informacji o kierunku przepływu energii akustycznej. Mimo to ze względu na wspomnianą łatwość pomiaru ciśnienia akustycznego za pomocą mikrofonu na przestrzeni lat powstała zaawansowana teoria, która zawiera definicję wielu obiektywnych parametrów akustyki pomieszczeń na bazie tzw. odpowiedzi impulsowej pomieszczenia, czyli czasowego przebiegu zmian ciśnienia akustycznego mierzonego w określonym punkcie pomieszczenia pobudzonego sygnałem impulsowym. W literaturze przedmiotu wskazane są metody wyznaczania na jej podstawie takich parametrów obiektywnych jak: czas pogłosu, czas wczesnego zaniku dźwięku, przejrzystość, wyrazistość czy siła dźwięku [14], [15]. Mimo swojej użyteczności posiadają one pewną wadę: na wyznaczane wielkości nie ma wpływu kierunek, z którego rejestrowana fala akustyczna przybywa w kolejnych chwilach czasu, gdyż odpowiedzi impulsowe mierzone są za pomocą dookólnego mikrofonu ciśnieniowego.

Jedynym unormowanym wyjątkiem, w którym rozróżnia się kierunek docierającej fali dźwiękowej, jest tzw. współczynnik energii bocznej, przy pomiarze którego odpowiedzi impulsowe mierzy się za pomocą dwóch mikrofonów: jednego o dookólnej charakterystyce kierunkowości i drugiego o charakterystyce ósemkowej. Jednak również w takim pomiarze odbicia docierające z różnych kierunków są uśredniane.

Akustyka pomieszczeń jest ściśle związana z percepcją dźwięku przez człowieka, który posiada dwoje uszu i słyszy przestrzennie. Ważnym aspektem, który należy wziąć pod uwagę przy modelowaniu i pomiarach, jest więc kierunek propagacji dźwięku, gdyż wpływa na postrzeganie dźwięku przez człowieka. Wielkością fizyczną, która dostarcza informacji o kierunku przepływu dźwięku, jest natężenie dźwięku [5], [7]. Uwzględnia ono jednocześnie zmiany ciśnienia akustycznego i prędkości akustycznej. Dzięki zarejestrowaniu odpowiedzi impulsowej pomieszczenia, która ukazuje przebieg wartości chwilowych natężenia dźwięku w czasie, można w znacznie większym stopniu analizować badane pomieszczenie.

Jak wspomniano wcześniej, pomiar natężenia dźwięku jest bardziej złożony i wymaga zastosowania specjalistycznych czujników – tzw. sond natężeniowych. W zależności od konstrukcji składają się one z kilku (co najmniej dwóch) precyzyjnie dopasowanych mikrofonów ciśnieniowych lub z jednego mikrofonu i dwóch lub więcej czujników (termoelektrycznych lub ultradźwiękowych) prędkości akustycznej [1], [6], [8]. Ze względu na detale konstrukcyjne takie rozwiązania są drogie, w związku z czym są mało rozpowszechnione i nie są jeszcze znormalizowane, tak aby mogły służyć do pomiarów akustyki wnętrz.

Niniejszy rozdział ma być przyczynkiem do rozpowszechniania metod natężeniowych jako narzędzia do badania akustyki pomieszczeń. Autorzy chcą przybliżyć uproszczoną, tanią metodę mierzenia przestrzennych natężeniowych odpowiedzi impulsowych,

zaproponować pewne sposoby wizualizacji wyników oraz wskazać, na bazie przeprowadzonych eksperymentów, obszary potencjalnego zastosowania metody w akustyce architektonicznej.

7.2. Metodologia

Proponowana uproszczona metoda pomiaru natężenia dźwięku bazuje na zasadzie działania dwumikrofonowej sondy p-p (ang. *presure-presure probe*) [1], [6]. Poniżej przedstawiono pokrótce jej zręby teoretyczne.

Zgodnie z [1] natężenie dźwięku I_r jest wielkością wektorową zdefiniowaną w następujący sposób:

$$I_r = \frac{d\,E_r}{dt \cdot dA} \tag{7.1}$$

Jest to więc parametr niosący informację o przepływie energii w kierunku *r* przez obszar *dA* zorientowany prostopadle do tego kierunku w czasie *dt*. Jednostką natężenia dźwięku jest [W/m^2]. Ponieważ energia dE_r jest równoważna pracy wykonanej w obszarze *dA* w kierunku *r*, to siłę F_r można zapisać następująco:

$$d\,E_r = F_r \cdot dr = p' \cdot dA \cdot dr \tag{7.2}$$

gdzie ciśnienie $p' = p_0 + p$ jest całkowitym ciśnieniem w środowisku.
Podstawiając równanie (7.2) do równania (7.1), otrzymujemy następującą zależność:

$$I_r = p' \cdot \frac{dr}{dt} = p_0 u_r + p\,u_r \tag{7.3}$$

gdzie u_r jest znaną wcześniej prędkością dźwięku. Przy założeniu, że wartość p_0 jest stała (stacjonarność pola), w wyniku uśredniania czasowego iloczyn $p_0 u_r$ zostanie wyzerowany, co prowadzi do ważnej zależności łączącej chwilowe natężenie dźwięku z ciśnieniem akustycznym i prędkością akustyczną:

$$I_r(t) = \overline{p(t) \cdot u_r(t)} \tag{7.4}$$

Zauważmy, że powyższe równanie uwzględnia tylko jeden kierunek *r*. W ogólnym przypadku trójwymiarowym równanie (7.4) można zapisać w następujący sposób:

$$\vec{I}(t) = \overline{p(t) \cdot u_x(t) + p(t) \cdot u_y(t) + p(t) \cdot u_z(t)} = \overline{p(t) \cdot \vec{u}(t)} \tag{7.5}$$

Natężenie dźwięku jest w istocie wielkością wektorową, która zależy od wektora prędkości akustycznej i zależności fazowej między tą prędkością a ciśnieniem.

Problem w pomiarze natężenia dźwięku pojawia się przy pomiarze chwilowej prędkości akustycznej. Prędkość tę można jednak estymować, używając do tego równania Eulera. Całkując po czasie równanie opisujące dynamikę płynu w przypadku trójwymia-

rowym i wykorzystując pomiary ciśnienia akustycznego w dwóch punktach, otrzymujemy równanie:

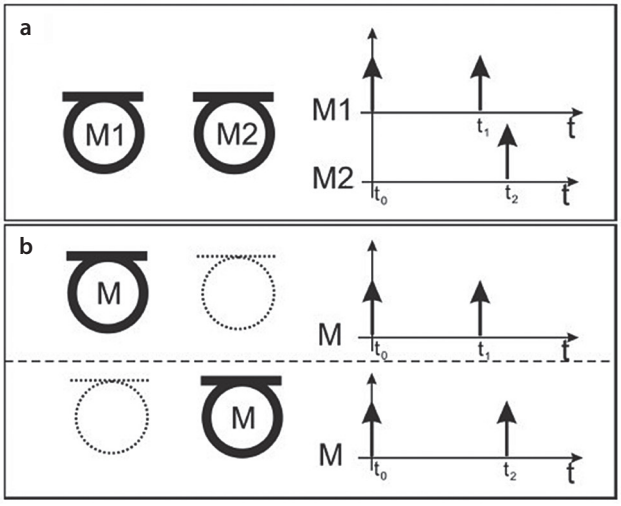
$$I_r = \frac{-1}{2\rho\,\Delta r}\overline{(p_B + p_A)\int (p_B - p_A)\,dt} \tag{7.6}$$

Wzór ten znajduje zastosowanie w pomiarze składowych natężenia dźwięku, ponieważ sprowadza się do sprawdzenia dwóch wartości ciśnienia akustycznego, a znając kierunek r i gęstość powietrza ρ wystarczy obliczyć wartość I_r. Składowe natężenia można mierzyć w dwóch lub trzech prostopadłych względem siebie kierunkach, a następnie łączyć, by otrzymać wypadkowy wektor natężenia dźwięku I w przestrzeniach dwu- lub trójwymiarowych. Taki pomiar składowej natężenia dźwięku oparty na wzorze (7.6) jest nazywany metodą bezpośrednią (w dziedzinie czasu). Podstawą zastosowanych algorytmów obliczeniowych jest właśnie wzór (7.6).

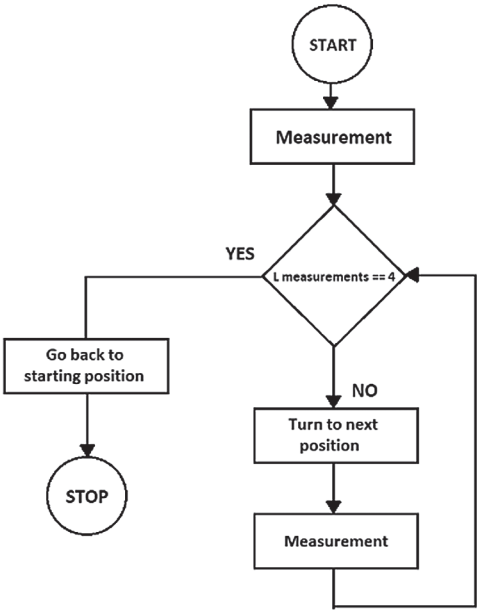
Istnieje również druga metoda, nazywana pośrednią, która oparta jest na analizie częstotliwościowej sygnałów p_A i p_B. Można wykazać, że składową natężenia dźwięku I_r opisuje następująca zależność:

$$I_r = \frac{-1}{\omega\rho\Delta r}\,\Im\left(G_{AB}\right) \tag{7.7}$$

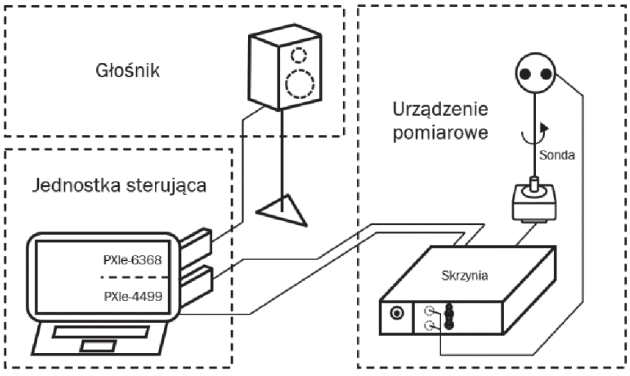
gdzie $I_m(G_{AB})$jest częścią urojoną zespolonej widmowej gęstości korelacji wzajemnej sygnałów p_A i p_B. Metoda pośrednia jest powszechnie stosowana w analizatorach dźwięku przeznaczonych do współpracy z tradycyjnymi sondami p-p.



Rys. 7.1. Schemat pomiaru natężenia dźwięku przy użyciu dwóch mikrofonów (a) oraz pomiaru synchronicznego przy użyciu jednego mikrofonu (b); oprac. własne na podstawie [10]



Rys. 7.2. Algorytm wykonywanych pomiarów



Rys. 7.3. Kompletny system pomiarowy

Wykorzystanie wzorów (7.1)–(7.7) i sondy dwumikrofonowej umożliwia rejestrację przebiegu pojedynczej składowej natężenia dźwięku wyemitowanego przez dowolne źródło dźwięku. W przypadku akustyki pomieszczeń w celu wyznaczania parametrów obiektywnych pobudza się znanym sygnałem i rejestruje odpowiedź układu na to pobudzenie. Dzięki współczesnej technologii taki eksperyment pomiarowy można powta-

rzać wielokrotnie przy zachowaniu pełnej powtarzalności wymuszenia i synchronizacji akwizycji. W tej specyficznej sytuacji przy założeniu stabilności parametrów termodynamicznych obiektu i liniowości badanego pola akustycznego (poziomach ciśnienia nieprzekraczających 120 dB SPL) pomiar natężenia dźwięku może zostać uproszczony i wykonany przy użyciu jednego mikrofonu, który w kolejnych pomiarach będzie zmieniał swoje położenie (zgodnie z geometrią sondy wielomikrofonowej), mierząc tym samym przebieg ciśnienia akustycznego w kilku pozycjach [9], [12], [13] (rys. 7.1).

W dalszej części rozdziału przedstawiono przykłady pomiarów 2D, a więc wymagających czterech pozycji pomiarowych do otrzymania zgodnie ze wzorami (7.6) lub (7.7) wektora natężenia dźwięku na płaszczyźnie.

Aby pomiar zaproponowaną metodą mógł być realizowany wygodnie, stworzono zintegrowany system pomiarowy obejmujący układ generacji i akwizycji danych bazujący na systemie NI PXI z autorskim instrumentem wirtualnym stworzonym w LabView, obrotową głowicę mikrofonową na bazie silnika krokowego oraz moduł sterownika silnika krokowego i przedwzmacniacza mikrofonowego. Na rysunku 7.2 pokazano algorytm pomiarowy, a na rys. 7.3 układ połączeń systemu pomiarowego.

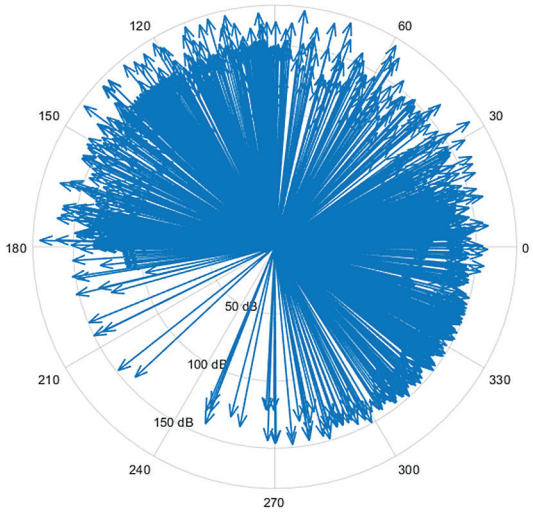
W celu uzyskania ciśnieniowych odpowiedzi impulsowych w kolejnych punktach pomiarowych obiekt pobudzany był sygnałem sinusoidalnym o logarytmicznie przestrajanej częstotliwości w funkcji czasu (sweep) [3], [4]. Generowano sygnał o czasie trwania wynoszącym 10 s i zakresie zmian częstotliwości od 20 Hz do 20 kHz. Po zakończeniu emisji system rejestrował jeszcze sygnał przez 1 s. Po wykonanym pomiarze w danym punkcie mikrofon był automatycznie obracany o 90° i pomiar powtarzał się.

7.3. Wizualizacja zmiennego w czasie wektorowego pola 2D

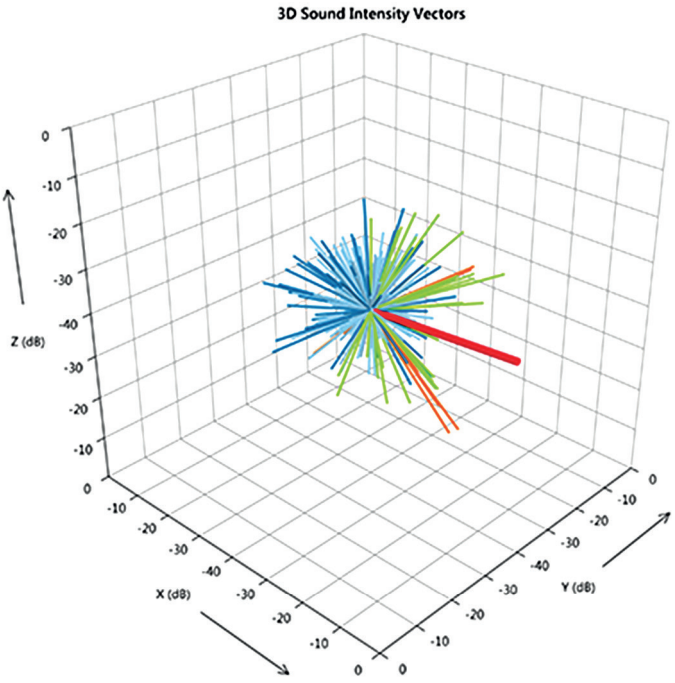
Pierwszym krokiem po wyznaczeniu ciśnieniowych odpowiedzi impulsowych jest ich przekształcenie w natężeniowe odpowiedzi impulsowe. Obliczenia realizowane są dla każdej próbki. Odbywa się to w wyniku połączenia parami ciśnieniowych odpowiedzi impulsowych (reprezentujących odpowiednie, prostopadłe kierunki), a następnie zastosowania wzoru (7.6). Dysponując przebiegami reprezentującymi kolejno składowe X i Y wektora natężenia dźwięku, można wykreślić (w postaci wektorów) szukaną, przestrzenną odpowiedź impulsową badanego pomieszczenia.

Przestrzenne odpowiedzi 2D łatwo przedstawić na wykresie biegunowym (rys. 7.4). Popularnym sposobem wizualizacji odpowiedzi impulsowych 3D jest tzw. *hedgehog pattern* (rys. 7.5).

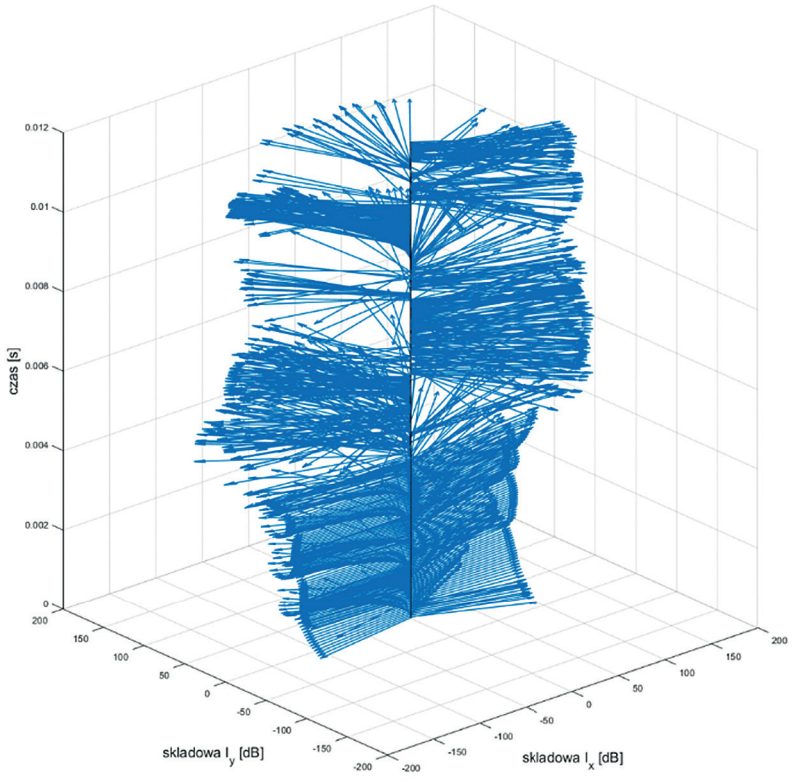
Każdy wektor lub segment natężenia dźwięku jest reprezentowany przez pojedynczą linię, wykreśloną na wykresie kartezjańskim 2D lub 3D w odniesieniu do pozycji pomiarowej (tj. początku), czyli tym samym pozycji położenia naszego mikrofonu. Długość każdej linii odpowiada względnemu poziomowi natężenia dźwięku (w dB), znor-



Rys. 7.4. Przestrzenna odpowiedź impulsowa 2D



Rys. 7.5. Wizualizacja 3D natężenia dźwięku wykonana z wykorzystaniem systemu pomiarowego IRIS [11]

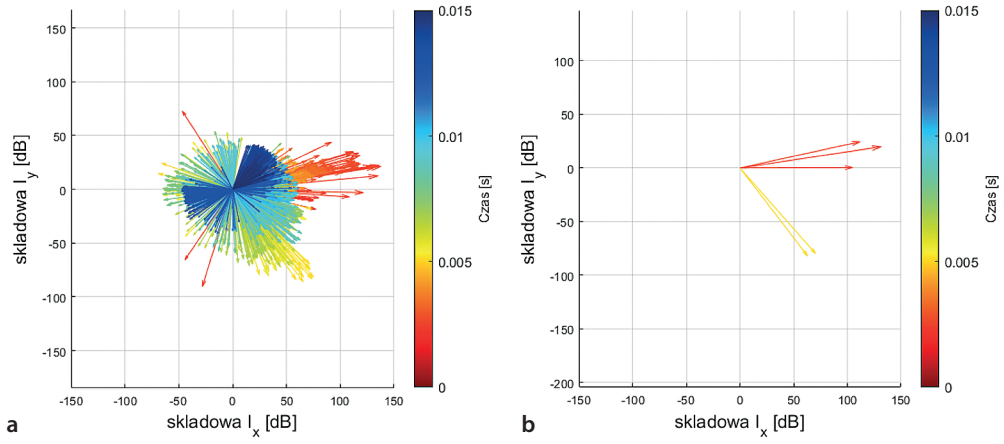


Rys. 7.6. Przestrzenna odpowiedź impulsowa z dodaną osią czasu

malizowanemu do natężenia dźwięku bezpośredniego, a jej kierunek jest szacowanym kierunkiem przepływu energii dochodzącej. Powyższe przykłady są wykresami skumulowanymi, na których wcale nie widać ewolucji czasowej (rys. 7.4) bądź jest ona zasygnalizowana kolorami poszczególnych strzałek. Uważamy, że taka prezentacja nie jest przejrzysta i trudno śledzić na niej ewolucję czasową przestrzennego przepływu energii akustycznej. Aby to poprawić, do wizualizacji 2D dodaliśmy dodatkową oś – oś czasu, dzięki której zamiast płaskiego „jeża" rysuje się coś na wzór pionowej spirali lub rozłożonego pióropusza.

Na otrzymanym wykresie 3D (rys. 7.6) zmienne w czasie długości, kierunki i zwroty wektorów są lepiej widoczne. Jednak zarówno liczba próbek, jak i szum występujący w sygnale pomiarowym są uciążliwe, jeśli konieczne jest przeprowadzenie wprost analizy takiego wykresu. W celu zwiększenia czytelności wykresu zaproponowano dwie metody uśredniania wektorów i progowe odszumianie sygnału pomiarowego.

Najprostszym sposobem uśrednienia jest wybranie konkretnego okna czasowego i wyliczenie, gdy liczba próbek odpowiada wielkości okna, jednej wartości średniej, po-



Rys. 7.7. Przestrzenna nieprzetworzona (a) i przetworzona (b) odpowiedź impulsowa komory bezechowej z umieszczoną wewnątrz ścianką, z uśrednianiem na niestałym oknie czasowym i odszumianiem przy $\epsilon = 30°$, $L = 100$ dB

sługując się standardową średnią arytmetyczną. Metoda ta umożliwia wprawdzie bardzo szybkie działanie, jednak wraz ze zwiększaniem okna (liczby N) zmniejsza się liczba dostępnych próbek w odpowiedzi natężeniowej – spada jej rozdzielczość czasowa.

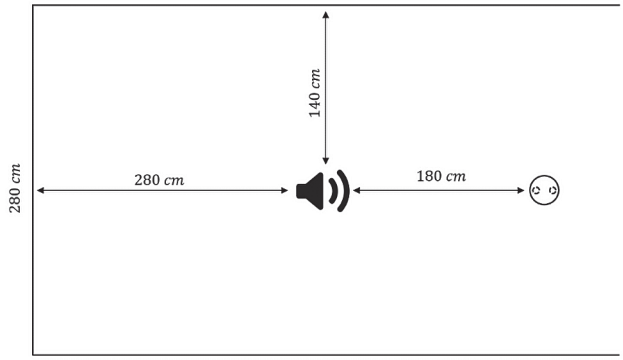
Inną metodą uśredniania może być wykorzystanie algorytmu o właściwościach odwrotnych. Wprawdzie podobnie jak metoda wyżej opisana, oparta jest również na oknie czasowym, ale zmienia swoją wartość dynamicznie w trakcie uśredniania. Uśrednianie w tej metodzie odbywa się na przestrzeni, w której kolejne próbki sygnału posiadają wartości zespolone o fazie nieodbiegającej od fazy próbki odniesienia więcej niż o dobraną tolerancję kątową ϵ. Próbką odniesienia w tym przypadku jest nie tylko pierwsza próbka w oknie, ale również każda następna, która nie spełniła warunku $(faza_{ref} - faza[k] \leq \epsilon$ (nie licząc pierwszej próbki w sygnale). Liczba próbek wewnątrz okna jest więc definiowana na podstawie orientacji pierwszego wektora i przyjętej tolerancji kątowej. Uśredniony sygnał otrzymany za pomocą tego algorytmu zachowuje informacje o odbiciach, które pojawiły się w trakcie propagacji dźwięku. Obie przedstawione metody mają swoje zalety, ale ich wspólną cechą jest to, że umożliwiają znaczne zmniejszenie liczby próbek przy jednoczesnym zachowaniu niesionych przez nie istotnych informacji, czyli przyczyniają się do zwiększenia przejrzystości prezentowanych odpowiedzi.

Poza uśrednianiem równie istotne jest odcięcie szumu tła. W naszych badaniach użyto prostej metody odcinania sygnału poniżej doświadczalnie dobranej wartości progowej. Taka prosta obróbka sygnału otrzymanego przy pomiarze pozwala na szybszą i lepszą zarówno prezentację, jak i analizę przestrzennych odpowiedzi impulsowych.

Na rysunku 7.7 przedstawiono przykład różnicy w wizualizacji 2D odpowiedzi nieprzetworzonej i przetworzonej opisanymi wyżej metodami.

7.4. Zastosowanie pomiarów natężeniowych w praktyce

W ostatniej części rozdziału zostaną podane przykłady zastosowania pomiaru natężeniowych odpowiedzi impulsowych pomieszczenia w celu obiektywnej oceny właściwości akustycznych pomieszczenia.



Rys. 7.8. Model wymiarowy korytarza

Rys. 7.9. System pomiarowy w trakcie trwania pomiaru

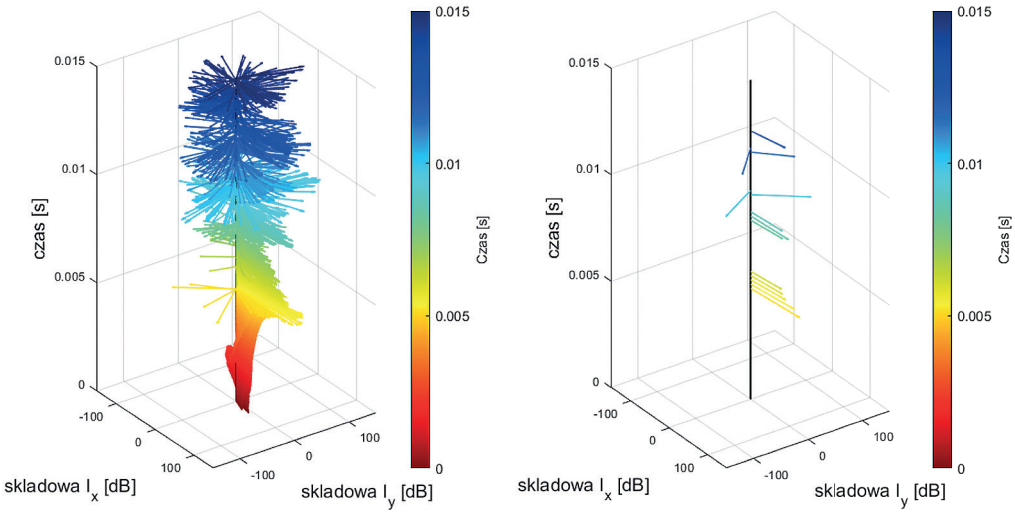
7.4.1. Analiza symetrii akustycznej pomieszczenia

Podstawą satysfakcjonującej percepcji sceny stereofonicznej przy słuchaniu zarówno orkiestry symfonicznej w sali koncertowej, jak i nagrania odtwarzanego w systemie hi-fi jest symetria akustyczna pomieszczenia odsłuchowego. Obiektywne badanie tej właściwości jest możliwe zaproponowaną metodą.

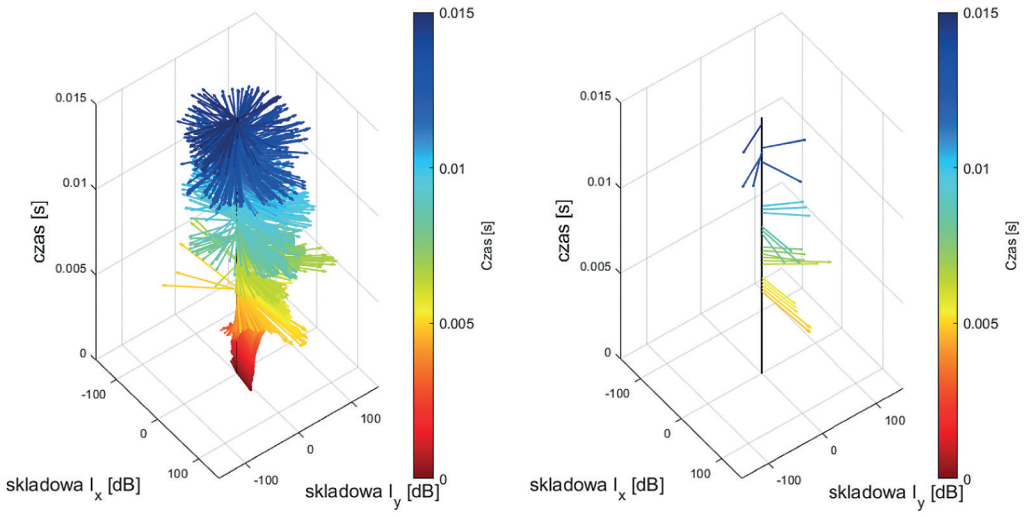
Poniżej przedstawione i opisane zostaną wyniki dwóch pomiarów, które zrealizowane zostały w tym samym pomieszczeniu o konkretnej geometrii, ale przy różnym rozmieszczeniu źródła dźwięku i sondy względem ścian tegoż pomieszczenia. Eksperyment przeprowadzono w prostopadłościennym korytarzu i w pierwszym przypadku źródło i sonda umieszczone zostały w na osi pomieszczenia (rys. 7.8 i 7.9).

Wykonany w takiej konfiguracji pomiar powinien cechować się symetrią kierunkowej odpowiedzi impulsowej. Na rysunku 7.10 przedstawiono zmierzoną przestrzenną odpowiedź impulsową bez postprocessingu oraz jej odpowiednik po zastosowaniu uśredniania i odcięcia szumów tła:

Na rysunku 7.10b widać, że niedługo po dźwięku bezpośrednim zarejestrowane zostały wektory o takim samym kierunku, lecz mniejszej amplitudzie. Są to wektory natężenia dźwięku powstałe w wyniku odbić dźwięku od podłogi. W kolejnych fragmentach odpowiedzi widoczne są wektory odpowiadające odbiciom od ścian bocznych, które cechuje wspomniana symetria. Można zauważyć, że po operacji uśredniania każdemu



Rys. 7.10. Przestrzenne nieprzetworzone (a) i przetworzone (b) odpowiedzi impulsowe badanego korytarza, z uśrednieniem w stałym oknie czasowym i redukcją szumu przy $N = 20$, $L = 67$ dB



Rys. 7.11. Nieprzetworzone (a) i przetworzone (b) odpowiedzi impulsowe korytarza przestrzennego (bez symetrii), z uśrednieniem w stałym oknie czasowym i redukcją szumów przy $N = 20$, $L = 60$ dB

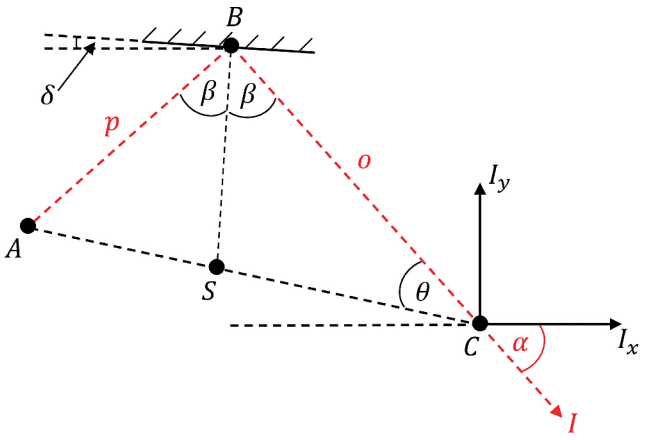
takiemu wektorowi odpowiada drugi wektor o kierunku symetrycznym względem osi I_x. Odbicia występują w bardzo bliskich chwilach czasu, czego się spodziewano.

Chcąc sprawdzić, czy opisywana symetria zostanie naruszona i odpowiednio wykryta w trakcie pomiaru, w tym samym korytarzu wykonano drugi pomiar. Tym razem jednak źródło dźwięku i sondę przesunięto równolegle względem osi korytarza w taki sposób, że odległość od jednej ze ścian bocznych wynosiła 204 cm, a od drugiej odpowiednio 76 cm. Odległość między mikrofonem a sondą zmniejszono do 167 cm. W ten sposób odbicia od jednej ze ścian powinny być widoczne na wykresie wcześniej niż odbicia zarówno od ściany drugiej, jak i podłogi. Tak faktycznie jest, co pokazano na rys. 7.11.

Zaprezentowane odpowiedzi wyraźnie pokazują, że pomiar nie jest już symetryczny względem osi I_x.

7.5. Algorytm wykrywania płaszczyzn odbijających

Pomiary zaprezentowane w podrozdz. 7.4.1 w znacznym stopniu zgadzają się z teoretycznymi przewidywaniami. Za ich pomocą można więc charakteryzować badane pomieszczenia pod kątem istnienia w nim płaszczyzn odbijających, wpływających na kierunek propagacji energii akustycznej w punkcie pomiarowym w kolejnych fazach trwania odpowiedzi impulsowej pomieszczenia. Tak więc na podstawie analizy natę-



Rys. 7. 12. Odbicie od pojedynczej przeszkody:
A – pozycja głośnika, *B* – punkt odbicia fali (środek przeszkody), *C* – pozycja środka sondy,
S – punkt przecięcia dwusiecznej kąta ∠*ABC* z odcinkiem |*AC*|,
I – zmierzony w czasie *t* wektor natężenia dźwięku o kierunku
α, *p* i *o* – promienie fali padającej oraz odbitej

żeniowej odpowiedzi impulsowej można dokonać próby detekcji obiektów, od których fala akustyczna została odbita. Przy znajomości czasu rejestracji fali odbitej, kierunku pojedynczego wektora, wzajemnego położenia sondy i głośnika oraz przy założeniu, że charakteryzujący odbicie kąt padania i kąt odbicia są takie same, problem detekcji przeszkody odpowiadającej pojedynczemu wektorowi staje się w przypadku pojedynczego odbicia problemem deterministycznym. Algorytm obliczania pozycji przeszkody zostanie opisany na podstawie poniższego rysunku:

Długość przebytej drogi *d* przez falę akustyczną poruszającą się z prędkością dźwięku v≈342 m/s i odpowiadającą wektorowi I można obliczyć na podstawie wzoru:

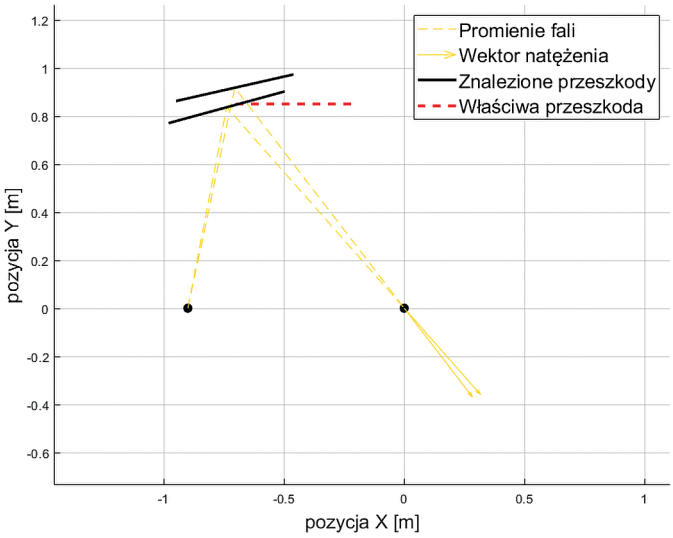
$$d = p + o = vt \tag{7.8}$$

gdzie:

d – droga jaką biegnie fala odbita,
p – odcinek od źródła do przeszkody,
o – odcinek od przeszkody do punktu pomiarowego,
v – prędkość propagacji dźwięku,
t – czas propagacji odczytany z odpowiedzi impulsowej.

Następnie, korzystając z prawa cosinusów dla kąta *θ* i przyjmując, że r = |*AC*|, otrzymujemy zależność:

$$p^2 = r^2 + o^2 - 2\,ro\cos(\theta), \tag{7.9}$$



Rys. 7.13. Wyznaczenie przeszkody dla pomiaru z rys. 7.7b

co po podstawieniu do równania (7.8) i uproszczeniu prowadzi do wyznaczenia długości promienia padającego fali:

$$p=\frac{r^2+d^2-2\,d\cos(\theta)}{2d-2\cos(\theta)}. \tag{7.10}$$

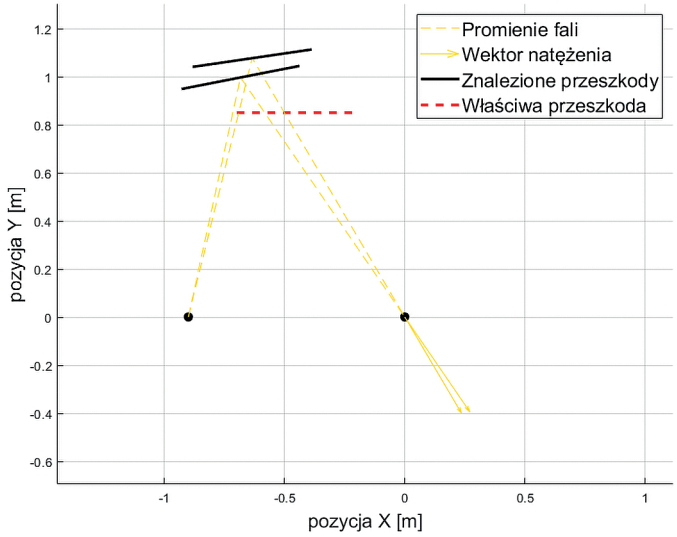
Znając długość p, można również obliczyć długość o z wykorzystaniem wzoru (7.8). Długość o wraz z kątem α jednoznacznie określa współrzędne punktu B. W sytuacji, gdy bierze się pod uwagę punkt odbicia fali B, ostatnim krokiem jest znalezienie kąta nachylenia δ względem osi I_x. Przeszkoda zorientowana zgodnie z kątem δ jest prostopadła do dwusiecznej kąta $\angle ABC$, na której leży odcinek $|BS|$. Zatem po znalezieniu współrzędnych punktu S, korzystając z twierdzenia o dwusiecznej kąta w trójkącie ΔABC, będzie można znaleźć pożądany kąt δ. Na podstawie twierdzenia otrzymamy wzór:

$$2\frac{AS}{CS}=\frac{p}{o}, \tag{7.11}$$

Podstawienie wzoru (11) do równania $|AS| = r - |CS|$ prowadzi do znalezienia długości $|CS|$.

$$CS=\frac{ro}{p+o}. \tag{7.12}$$

Biorąc pod uwagę długość $|CS|$ i kąt θ, można obliczyć współrzędne punktu S, tj. nachylenie odcinka $|BS|$ i kąt δ. Widać więc, że w przypadku pojedynczego odbicia fali za-



Rys. 7.14. Wyznaczenie przeszkody dla skompensowanego pomiaru z rys. 7.7b

rejestrowanej w czasie t i kąta α oraz znajomości względnego położenia głośnika i sondy można wyznaczyć przeszkodę skupioną w punkcie B, od której fala została odbita.

Z rysunku 7.13 wynika, że algorytmowi udaje się wprawdzie wykryć istniejącą przeszkodę, ale jej pozycja i orientacja nie pokrywają się jednak ze stanem faktycznym. Dodatkowo, z punktu widzenia algorytmu w pomieszczeniu znajdują dwie przeszkody. Te niedokładności wynikają z błędów powstałych w trakcie pomiarów.

Pierwszym aspektem wpływającym na niedokładność pomiaru jest powstałe i widoczne w każdym pomiarze rozmycie kierunków i długości wektorów, np. rozmycie wektorów związanych z dźwiękiem bezpośrednim. Rozmycie powstaje w wyniku emisji dźwięku z rzeczywistego głośnika – źródła nieidealnego, którego nie cechuje ani idealna punktowość ani dookólność.

Ponadto membrana głośnika posiada swoją bezwładność, która po emisji dźwięku właściwego wciąż może powodować zaburzenia ośrodka. Wskutek tego idealny impuls, któremu odpowiadałby pojedynczy wektor, jest tak naprawdę rozłożony w czasie, co przekłada się na większą ilość wektorów, a to z kolei na błędy w rejestrowanym czasie, kierunku i amplitudzie uśrednionego wektora reprezentującego/wektorów reprezentujących rozmycie. Kolejnym problemem jest niedokładność ułożenia początkowego sondy. Początkowe położenie sondy definiuje tak naprawdę orientację układu odniesienia całego układu. Jeżeli na samym początku sonda była lekko obrócona, to cały układ odniesienia również. Skutkiem jest obrócenie wszystkich wektorów wchodzących w skład przestrzennej odpowiedzi impulsowej i stały błąd kątowy każdego z nich. Kolejne błędy wynikać mogą z wykonanych pomiarów położenia głośnika względem sondy i pomiaru

pozycji samej przeszkody, jak również opóźnień wprowadzanych przez tor rejestracji i akwizycji danych.

Zakładając addytywność wszystkich wymienionych błędów w trakcie pomiarów, można spróbować te błędy skompensować. Zmierzona odpowiedź z rys. 7.7b została, w wyniku opisywanych błędów, obrócona o pewien dodatni kąt. Pokazują to wektory odpowiadające dźwiękowi bezpośredniemu, gdyż po uśrednieniu nie charakteryzuje ich zakładana równoległość względem osi I_x. W celu kompensacji znaleziono kąt między osią I_x a uśrednionym wektorem reprezentującym całe rozmycie dźwięku bezpośredniego, który posłużył do kompensacji pomiaru. Przepuszczając ostatecznie skompensowaną odpowiedź przez algorytm wykrywania przeszkód, otrzymuje się wyniki zaprezentowane na rys. 7.14.

Dzięki wprowadzonej kompensacji wyniki nieznacznie poprawiły się pod względem orientacji przeszkody. Nadal jednak występuje błąd niezgodności pozycji. Jedną z wciąż niewyeliminowanych przyczyn jest założenie obowiązywania zasad optyki geometrycznej w akustyce. Zaobserwowany błąd może być dowodem na występowanie dużej składowej rozproszonej w dźwięku odbitym, nawet w przypadku przeszkód pozornie płaskich i sztywnych. Podsumowując, można zatem stwierdzić, że algorytm tylko w pewnym stopniu może odwzorowywać rozmieszczenie przeszkód w pomieszczeniu i może być stosowany tylko w pobieżnej analizie przeszkód w pomieszczeniu.

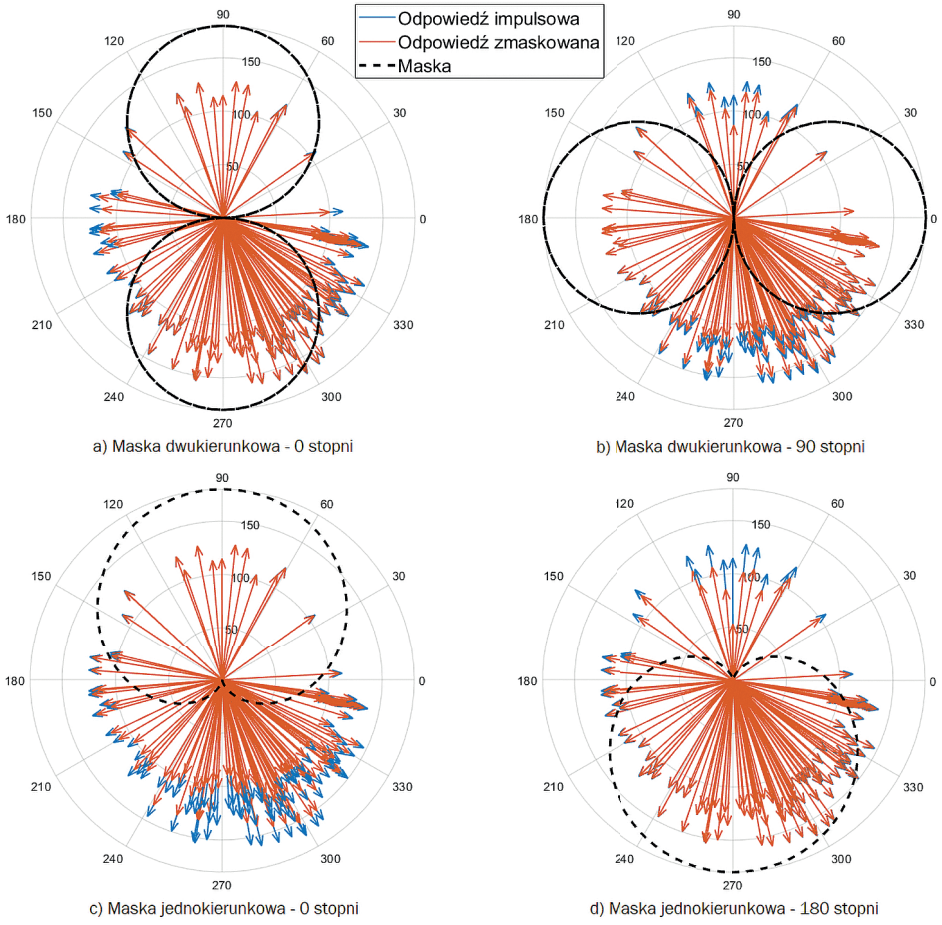
7.6. Wyznaczanie współczynnika odbić bocznych

Jednym z parametrów normatywnych, który decyduje o wrażeniu przestrzenności dźwięku odbieranego w danym pomieszczeniu jest współczynnik odbić bocznych (eng. *lateral early fraction* – LEF), gdzie odpowiedzi impulsowe będące podstawą ich wyznaczenia mierzy się z wykorzystaniem mikrofonów o charakterystyce dookólnej i ósemkowej. Współczynnik LEF definiuje się jako:

$$LEF = \frac{\int_{25ms}^{80ms} h_1^2(t)\,dt}{\int_{0ms}^{80ms} h^2(t)\,dt} \tag{7.13}$$

gdzie odpowiedź impulsowa $h(t)$ jest mierzona typowym, dookólnym mikrofonem, a odpowiedź $h1(t)$ mikrofonem o charakterystyce dwukierunkowej (ósemkowej).

Wielkość ta pozwala na określenie ilości energii bocznej w stosunku do jej całkowitej wartości we wczesnej fazie propagacji. Im większa jest jej wartość, tym większa część energii całkowitej została rozpropagowana w kierunkach bocznych. Objawem tego jest



Rys. 7.15. Wykorzystanie natężeniowej odpowiedzi impulsowej
do wyznaczania współczynników odbić bocznych dla różnych kierunków

odnoszone przez osobę znajdująca się wewnątrz takiego pomieszczenia, że dźwięk charakteryzuje się większą przestrzennością.

Dzięki pomiarowi przestrzennej natężeniowej odpowiedzi impulsowej taki współczynnik możemy wyznaczyć w bez konieczności zmiany konfiguracji sprzętu. Wystarczy, że po przeprowadzeniu pomiarów na podstawie otrzymanych danych przestrzennych przeprowadzimy proces całkowania z wykorzystaniem odpowiednich współczynników wagowych reprezentujących charakterystykę kierunkową użytego mikrofonu. Można w tym wypadku nie tylko zastosować charakterystyki dookólne i ósemkowe, ale też inne i dzięki temu wyznaczać współczynniki energii odbić z dowolnych kierunków, nie tylko z kierunku bocznego jak w przypadku LEF. Po wyliczeniu wartości natężenia dźwię-

ku zmierzonego mikrofonami o różnych charakterystykach jesteśmy w stanie obliczyć przybyłą część energii i kierunek, z którego pochodzi. Oblicza się ją, posługując się nieco zmodyfikowaną zależnością (7.13) dla sygnałów dyskretnych na całym przedziale czasowym:

$$EF_m = \frac{\sum_{0}^{\infty} I_m^2}{\sum_{0}^{\infty} I_0^2} \cdot 100\,\% \tag{7.14}$$

gdzie EF_m jest procentową częścią energii całkowitej przybyłej z kierunków, które były „najważniejsze" względem odpowiedniej operacji maskowania. Dzięki obliczaniu tak zdefiniowanej wartości można sprawdzić, z którego kierunku przybyła największa część energii.

Korzystając z przykładowych odpowiedzi impulsowych i nałożonych na nie różnych masek, można obliczyć wartości EF_m dla każdego przypadku. W przykładach pokazanych na rys. 7.15 uzyskano następujące wyniki: EF_{ma} = 77%, EF_{mb} = 23%, EF_{mc} = 0,5%, EF_{md} = 88%. Otrzymane wartości świadczą o tym, że najwięcej energii przybyło z kierunków bliskich kątowi α = 90°, zaś najmniej z kierunków bliskich kątowi α = 270°. Przyjmując za początek układu odniesienia sam mikrofon i zakładając, że na drodze między źródłem dźwięku a mikrofonem nie występowały żadne przeszkody, można dojść do wniosku, że źródło dźwięku znajdowało się właśnie na kierunku bliskim wartości $\alpha \approx 90°$.

7.7. Wykorzystanie odpowiedzi natężeniowych do generacji odpowiedzi impulsowych wykorzystywanych w splotowych reverberatorach surround

Reverberatory splotowe zapewniają bardzo realistyczne uprzestrzennianie nagrań fonicznych. Warunkiem uzyskania takich dobrych efektów jest posiadanie zestawu wysokiej jakości odpowiedzi impulsowych, odpowiadających danemu systemowi reprodukcji dźwięku. Odpowiedzi takie najczęściej rejestruje się w rzeczywistych obiektach o pożądanej akustyce, za pomocą zestawów mikrofonów stereofonicznych używanych standardowo do realizacji nagrań. Posiadając przestrzenne odpowiedzi impulsowe pomieszczenia, można postąpić inaczej. Odpowiedź impulsową pomieszczenia rejestrujemy za pomocą proponowanej, uproszczonej metody natężeniowej, a dodatkowo rejestrujemy ciśnieniową odpowiedź impulsową za pomocą wysokiej klasy mikrofonu dookólnego. Następnie, zgodnie z ewolucją przestrzenną odpowiedzi natężeniowej, segmentujemy

i rozdzielamy odpowiedź ciśnieniową na taką liczbę odpowiedzi, która odpowiada systemowi reprodukcji dźwięku. Tak otrzymany zestaw odpowiedzi impulsowych stosujemy w reverberatorze surround. Warunkiem osiągnięcia oczekiwanych rezultatów przedstawionej idei jest opracowanie metody wygładzania wynikowych odpowiedzi impulsowych w rejonach segmentacji w celu wyeliminowania możliwych artefaktów słyszalnych w uprzestrzennionym nagraniu. Nad rozwiązaniem tego problemu prowadzone są obecnie przez zespół prace badawcze.

7.8. Wnioski

Ciśnieniowe odpowiedzi impulsowe pomieszczenia są szeroko stosowane w określaniu obiektywnych wskaźników jakości akustyki pomieszczeń. Wciąż jednak wskaźniki te nie gwarantują odróżnianie pomieszczeń o akustyce poprawnej od pomieszczeń o akustyce wybitnej. Dzieje się to, zdaniem autorów niniejszego rozdziału, z powodu braku powszechnie stosowanych metod pomiarowych uwzględniających kierunkową naturę propagacji fal w pomieszczeniu oraz przestrzenne możliwości jej percepcji przez człowieka. Zaproponowana metoda pomiarowa oraz przedstawione przykłady jej wykorzystania mają na celu zachęcenie do podjęcia na nowo poszukiwania korelacji między obiektywnymi wskaźnikami bazującymi na akustyce wektorowej i subiektywnymi ocenami jakości akustycznej pomieszczeń. Zastąpienie analizy czasowej ciśnieniowych odpowiedzi impulsowych analizą czasowo-przestrzenną odpowiedzi natężeniowych wymaga opracowania nowych metod wizualizacji. Jedną z takich metod zaproponowano w tym rozdziale. Dzięki jej właściwej analizie można obiektywnie określić np. symetrię akustyczną pomieszczenia. Z kolei zaprezentowane badania nad algorytmem wykrywania przeszkód wskazują na problem rozpraszania dźwięku przy odbiciu, który powinien zostać rozwiązany. Na przykładzie współczynnika odbić bocznych pokazano, że rejestracja odpowiedzi kierunkowej daje większe możliwości analizy przestrzennych właściwości pola akustycznego w fazie opracowywania danych. Zaproponowana segmentacja ciśnieniowej odpowiedzi impulsowej z wykorzystaniem odpowiedzi natężeniowej otwiera nowe możliwości uprzestrzenniania nagrań z wykorzystaniem splotowych procesorów pogłosowych. Podsumowując, powinniśmy chętniej śledzić i brać pod uwagę w pomiarach akustycznych metody czasowo-przestrzenne. Zdaniem autorów w przyszłości mogą one, w niezbadany jeszcze sposób, umożliwić jeszcze bardziej efektywne kształtowanie akustyczne pomieszczeń i realizację nagrań.

Finansowanie: Badania były wspierane przez Szkołę Wyższą ZUT /Szkoła Orłów ZUT/ projekt koordynowany przez dr Piotra Sulikowskiego, w ramach programu Ministra Edukacji i Nauki /Grant nr MNiSW/2019/391/DIR/KH, POWR.03.01.00-00-P015/18/, współfinansowanego z Europejskiego Funduszu Społecznego, kwota finansowania 1 704 201 66 zł.

Bibliografia

[1] Gade S., *Sound Intensity (Part I Theory)*, „Technical Review" 1982, No. 3.
[2] *Handbook of Acoustics*, Rossing T.D. (ed.), Springer, 2014
[3] Farina A., *Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique*, w: Proceedings of AES 108th Convention, Paris 2000.
[4] Farina A., *Advancements in impulse response measurements by sine sweeps*, „Journal of The Audio Engineering Society" 2007.
[5] Fahy F.J, *Sound Intensity*, Spon, London 1995.
[6] Fahy F.J, *Measurement of acoustic intensity using the cross-spectral density of two microphone signals*, „Journal of the Acustical Society of America" 1997, Vol. 62, Iss. 4, s. 1057–1059.
[7] Jacobsen F., *Sound intensity and its measurement*, w: Proceedings of 5th International Congress on Sound and Vibration, 1997.
[8] Jacobsen F., *Sound intensity and its measurement and applications*, Technical University of Denmark, 2005
[9] Mickiewicz W., Jablonski M., Pyla M., *Automatized system for 3d sound intensity field measurement*, w: Proceedings of 16th International Conference Methods and Models in Automation and Robotics (MMAR), 2011
[10] Mickiewicz W., *Metrologia i przetwarzanie sygnałów w obrazowaniu wektorowego pola akustycznego*, Wydawnictwo ZUT, Szczecin 2019.
[11] Protheroe D., Guillemin B., *3D impulse response measurements of spaces using an inexpensive microphone array*, International Symposium on Room Acoustics, Toronto 2013.
[12] Mickiewicz W., Raczyński M., Parus A., *Performance Analysis of Cost-Effective Miniature Microphone Sound Intensity 2D Probe*, "Sensors" 2020, Vol. 20, No. 1, s. 271.
[13] Mickiewcz W., Raczyński M., *Modified pressure-pressure sound intensity measurement method and its application to loudspeaker set directivity assessment*, „Metrology and Measurement Systems" 2020, Vol. 27, No. 1, s. 181–194.
[14] Beranek L., *Source of practical acoustical concepts and theory, with information on microphones, loud-speakers and speaker enclosures, and room acoustics*, American Institut of Physics, 1954
[15] Sadowski J., Architectural acoustics, PWN, Warszawa, 1976
[16] Kuttruff H., *Room Acoustics*, CRC Press, 2016.
[17] Dunn M., Protheroe D., *Visualization of early reflections in control rooms*, w: Proceedings of AES 137th Convention, Los Angeles, 2014.

Słowa kluczowe: natężenie dźwięku, odpowiedź impulsowa, akustyka wnętrz, pogłos.

Uproszczona metoda pomiaru przestrzennych odpowiedzi impulsowych i jej wykorzystanie do oceny jakości akustycznej pomieszczeń i generowania pogłosu surround

W rozdziale przedstawiono zagadnienia związane z badaniem niektórych właściwości akustycznych pomieszczeń z wykorzystaniem przenośnego, zintegrowanego systemu pomiarowego zdolnego do wyznaczania przestrzennych natężeniowych odpowiedzi impulsowych. Wyjaśniono metodykę pomiaru przestrzennej natężeniowej odpowiedzi impulsowej pomieszczenia z wykorzystaniem autorskiego systemu pomiarowego. Przedstawiono metody przetwarzania i interpretacji wyników pomiarów uzyskanych za jego pomocą oraz nakreślono możliwe obszary wykorzystania tych danych do obiektywnej oceny wybranych właściwości akustyki wnętrz, jak również do uprzestrzenniania nagrań muzycznych.

Simplified method of measuring spatial impulse responses and their application in the objective assessment of acoustic quality of rooms and surround reverberation

The primary purpose of this article is to present issues related to the study of certain acoustic properties of room using a portable, integrated measurement system capable of measuring spatial intensity impulse responses. The methodology of measuring the sound intensity and spatial impulse response of a room using a broadband excitation signal will be explained. Then the methods of processing and interpreting the results are given.

8. Integration of machine learning techniques and deterministic algorithms within an advanced system for monitoring and identifying vibroacoustic threats

Bartosz Chmielewski[1], Paweł Nieradka[1,2], Arkadiusz Utko[1], Bartłomiej Golenko[2], Monika Wasilewska[2], Maciej Walczyński[2], Piotr Pruchnicki[2], Przemysław Plaskota[2]

[1] KFB Acoustics, ul. Oławska 8, 55-040 Domasław
[2] Wroclaw University of Science and Technology, Chair of Acoustics, Multimedia and Signal Processing, wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wroclaw

8.1. Introduction

Excessive noise and vibrations are serious problems in environments and workplaces like production facilities [1], [2]. Many standards and laws (both international [3] and national [4]) have been introduced to address those issues. All employers whose employees are exposed to noise are legally obliged to comply with the recommendations and requirements defined in these documents. Noise pollution introduced into the external environment (e.g., impacting homes near factories) is also regulated by law [5]. Most of those documents set limits (permissible levels and boundaries) on a number of acoustic indicators in order to monitor whether noise associated with a given control point is excessive. Those indicators often take into account human hearing characteristics by imposing psychoacoustic corrections on the measured values [6]. Examples of such acoustics indicators are A-weighted equivalent sound pressure level $L_{A,eq}$, C-weighted peak sound pressure level $L_{C,\mathrm{peak}}$, and A-weighted maximum sound pressure level $L_{A,max}$. The assessment of tonality and impulsivity is also a crucial factor, because humans are more sensitive to impulsive and tonal sounds than to stationary and broadband noise [7]. When assessing environmental noise, special correction factors are introduced to take into account the tonality and impulsivity of noise.

In order to effectively react to threats, one must answer the following questions:

- Q1: What caused the vibroacoustic threat?
- Q2: When did it happen?
- Q3: Where did it happen?

Answering the above questions becomes a challenging task, especially in large production halls with many installations and machines. This usually requires manual, advanced acoustic analysis. A competent acoustician taking the measurements in real time associates each time instant with a specific acoustic event (e.g., stating that at a given hour and work station a hammering noise was present). Based on these time stamps of sound pressure, it is possible to construct a noise source ranking and to decide which sources have to be quieted. Tonality and impulsivity can be assessed separately. Such analyses are time-consuming and costly and cannot be performed continuously. Therefore, the automation and integration of such processes into a single system that operates in real time would be very advantageous for acousticians and their clients. Currently, methods based on machine learning can partially automate the problem-solving in this context. In order to identify acoustic events, methods using convolutional neural networks [8] and various improvements of these networks are used, including hybrid convolutional-recurrent networks [9] or alternative activation functions [10]. All of these solutions are helpful in answering Question Q1. However, currently there is no solution that can answer all three questions (Q1, Q2, and Q3) within a single system.

8.2. System for monitoring and identifying vibroacoustic threats

This section presents the system that was developed in response to these problems. The system is capable of responding to Questions Q1, Q2, and Q3. Section 8.2.1 provides background information on the system, called SMiZW (system for monitoring and identifying vibroacoustic threats), and its assumptions, while Section 8.2.2 focuses on the issue of integrating the algorithm within this system.

8.2.1. System architecture and assumptions

The system consists of a grid of sensors (microphones or intensity probes) distributed across the test area and a central data analysis module (MCAD). The MCAD is a programmatic implementation of all the algorithms discussed in this chapter: a supervised DNN (deep neural network), an unsupervised clustering algorithm (DBSCAN), and a deterministic DSP (determination of tonality, impulsivity, spectral characteristics, and direction of wave arrival). To avoid overloading the MCAD with data, some basic calculations (acoustic indicators like $L_{A,eq}$, $L_{C,\mathrm{peak}}$, and sound spectra) are calculated in

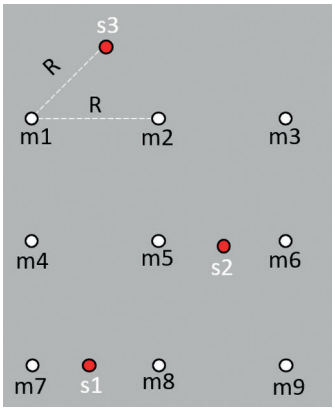


Fig. 8.1. Measurement configuration for SMiZW tests

dedicated circuit boards integrated into each sensor. The SMiZW does more than indicate when permissible noise levels are exceeded; the most important functionality of the system is answering the three questions posed in Section 8.1 in real time (What caused the vibroacoustic threat? When did it happen? Where did it happen?).

The system was tested on a dedicated test stand (Fig. 8.1) consisting of a grid of nine microphones m_i (3 × 3) and omni-directional sources (s1, s2, and s3). The grid size was set to R = 2 m. In order to reproduce near-real conditions, each source "s" was responsible for reproducing acoustic waveforms associated with a different class "c" (c1→s1, c2→s2, c3→s3), making each omni-directional source simulate the operation of a different noise source. The reconstructed waveforms were then recorded with sensors arranged in a grid and processed in the MCAD.

The architecture of the DNN used in the system is based on 2D convolution layers, dense layers, and drop-out layers. The learning process used the Mini Batch Gradient Descent method, with a mini-batch size of 64. Fig. 8.2 shows the output of the algorithm for the selected sensor. The upper part of the figure shows three signals applied to individual speakers from the test bench. In the lower part, the time that each class occurred is marked. The black color indicates the actual moment of the event, while the orange color indicates the algorithm's prediction. In order not to obscure the data, only a portion of the marked waveform is shown (the total recording time was equal to 6 h). The figure shows that when a disturbing sound occurred, the algorithm was still able to correctly label the classes on which it was trained, although the accuracy then decreased. The total accuracy of the algorithm, determined after the tests, was 82%. Further investigations are currently being conducted in order to improve the accuracy even more (by performing experiments with LSTM [11] and "squeeze and excitation" [12] layers).

The initial version of the clustering algorithm in the SMiZW was based on the K-means method, where an efficiency of 85% was achieved. The disadvantage of the

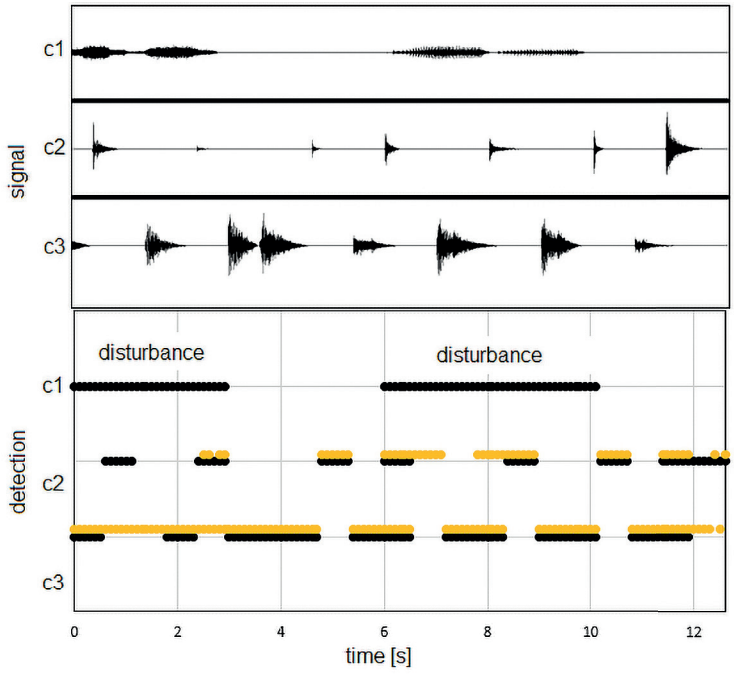


Fig. 8.2. The output of the DNN for the selected sensor

K-means method is the requirement for a priori declaration of the number of clusters used during partitioning. In the real conditions of the SMiZW system installation, this information is often impossible to obtain because the number of unidentified noise sources may be unknown. The following procedure can be carried out to select the optimal number of clusters. The distortion should be determined for successive values of the number of clusters: $n \in [2, N_{max}]$, where n is the number of clusters. Then, the inflection point of the function determined this way should be identified and selected as the optimal number of clusters. The inflection point is understood here as the point after which the distortion difference between N and $N + 1$ clusters becomes significantly smaller. Thus, the selection of the number of clusters is based on an objective parameter, but is not automated. A graph showing this procedure is presented in Fig. 8.3.

To fully automate the clustering algorithm, two alternative methods that do not require an a priori declaration of the number of clusters were tested: hierarchical clustering [13] and DBSCAN [14]. The effect of the distance threshold parameter on the accuracy of hierarchical clustering and DBSCAN was investigated, and the results are shown in Fig. 8.4. For both methods, the Euclidean metric was used to measure the distance between samples. In hierarchical clustering, the Ward technique was used, which minimizes the variance of the clusters being merged [15].

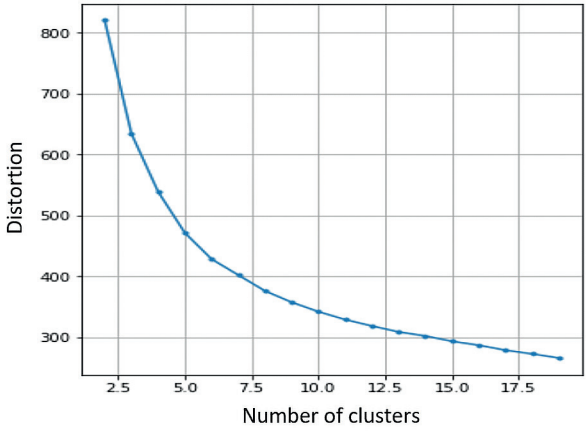


Fig. 8.3. Distortion curve versus number of clusters

For hierarchical clustering, 64% accuracy was obtained with a distance threshold value of 0.8. The best accuracy (81%) was obtained for the DBSCAN algorithm with a distance threshold value of 0.65. The accuracy obtained when using DBSCAN was slightly lower than that when using K-means (81% < 85%), but the fact that the number of clusters does not need to be declared can be considered a favorable trade-off. Fig. 8.5a shows the correctly grouped classes (reference), while Fig. 8.5b shows the effect of the DBSCAN algorithm with a distance threshold value of 0.65. The TSNE algorithm [16] was used to present the groups of classes in the 2D graph. This is a stochastic method of

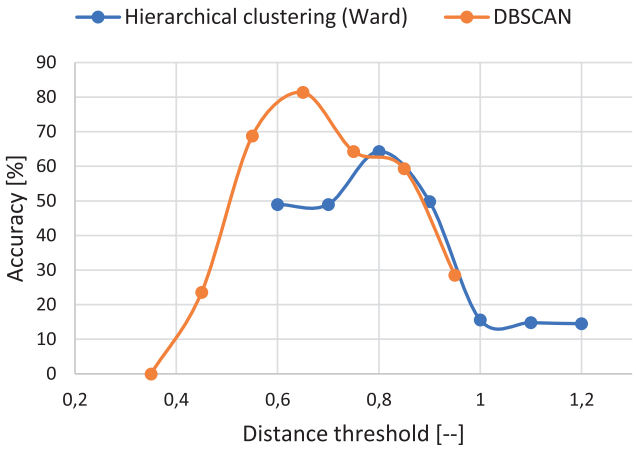


Fig. 8.4. The influence of the distance threshold on the accuracy of the tested clustering methods

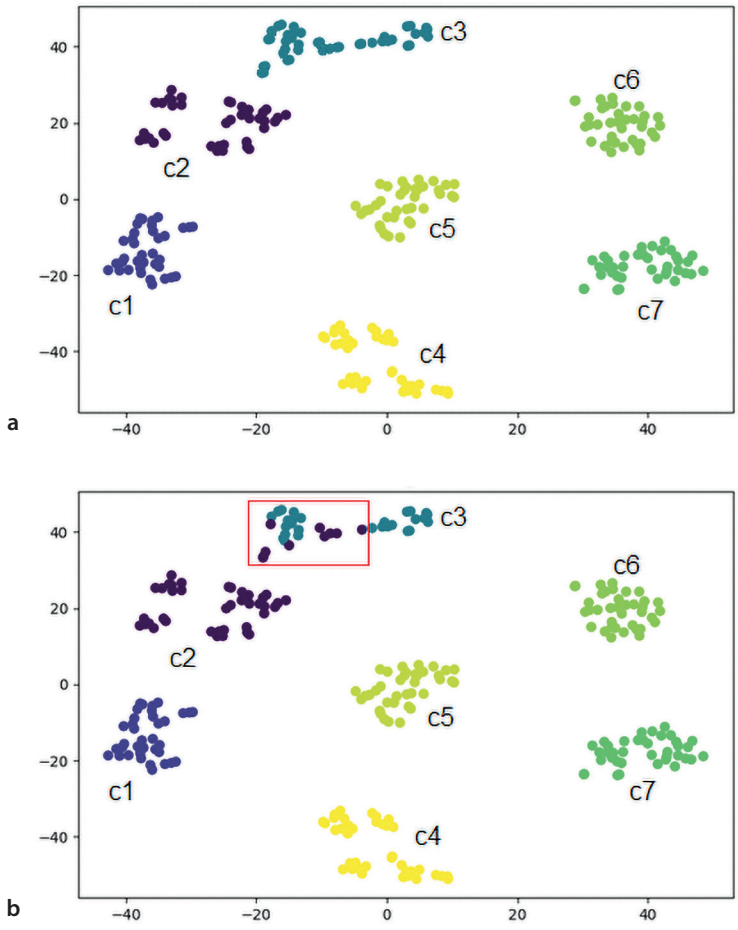


Fig. 8.5. (a) Tested signals classified according to the truth and (b) results of DBSCAN clustering

ordering neighbors based on the Student's t distribution, which allows the dimensionality of the data to be reduced while clustering similar classes. In order to allow for the correction of incorrect predictions (for example, in Fig. 8.5b there were several mistakes between Classes c2 and c3), the system provides an opportunity to listen to each sample and correct the decision made by the clustering algorithm. It should be noted that the selected optimal parameters apply to test data. In real implementations of the system, it may be necessary to tune the parameters of the algorithms under the given acoustic conditions.

The tonality of a signal is determined by the difference in levels in adjacent 1/3-octave bands. This method is described in Annex K of ISO 1996-2 [17]. It consists of comparing the time-averaged sound pressure level in the analyzed 1/3-octave band with sound

pressure levels in the neighboring two bands. For the tonal component to be detected, the level difference must exceed a fixed value. The level difference can vary with frequency. Annex K proposes to take the level difference depending on the center frequency of the 1/3-octave band: 15 dB in the low-frequency bands (25 Hz to 125 Hz), 8 dB in the mid-frequency bands (160 Hz to 400 Hz), and 5 dB in the high-frequency bands (500 Hz to 10,000 Hz). The impulsivity of a signal is determined according to NordTest NT ACOU 112 [18], where a signal is considered impulsive if its A-weighted sound level rises at a rate of at least 10 dB/sec. The direction of the wave arrival and localization algorithm is based on measuring and postprocessing the 2D sound intensity [19] on a grid of sensors and analyzing the points where they intersect.

8.3. Integration of algorithms

This section focuses on developing the ensemble of algorithms presented in the previous section. Fig. 8.6 demonstrates how the synergy between deterministic algorithms and ML algorithms can provide answers to the questions raised in Section 8.1:

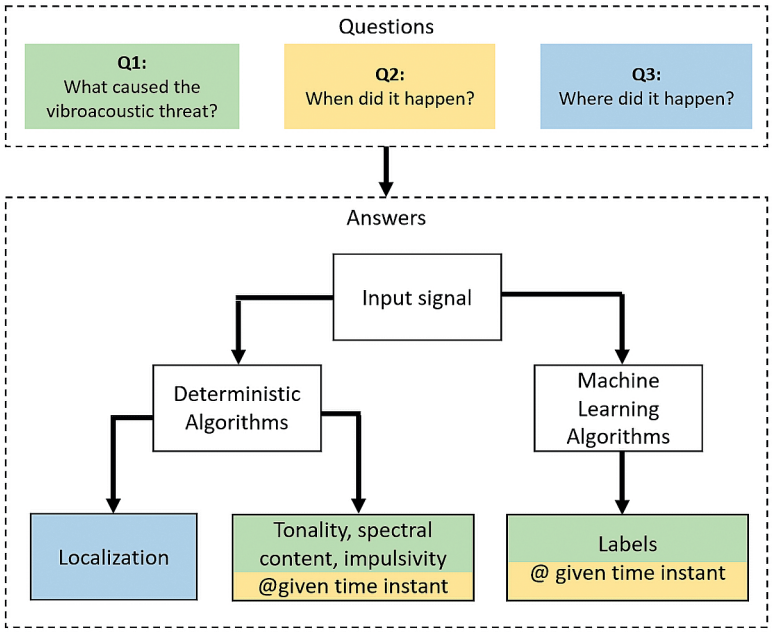


Fig. 8.6. Complementarity and synergy between different algorithms in the SMiZW to answer Questions Q1, Q2, and Q3

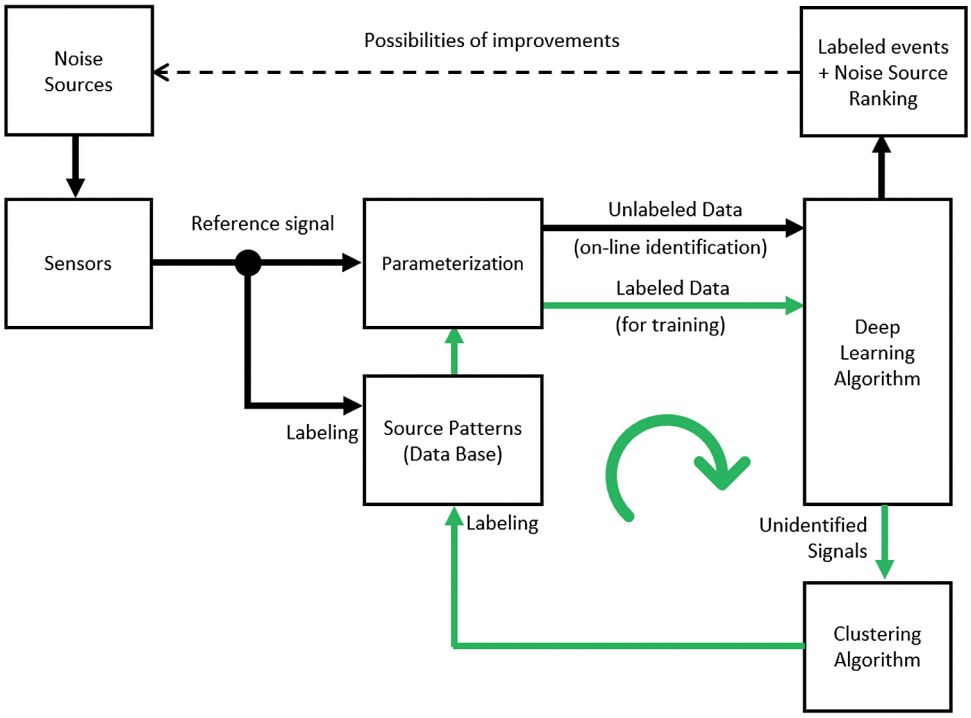


Fig. 8.7. Data flow in the SMiZW (ML part)

- Q1 can be answered based on labels obtained from the DNN, though some specific signals could only be identified with very low probability. Therefore, the inclusion of tonality information, impulsivity, direction of wave arrival, and spectral characteristics can significantly enrich the data collected and can help identify a problematic sound source based only on those deterministic characteristics. In general, having multiple independent analysis tools can infer more about the phenomena under study.
- The answer to Q2 is always obtained because the SMiZW works in real time and assigns time stamps to each identified event.
- Q3 is answered based on the localization algorithms.

Another important issue that may arise in practice is the problem of a small training dataset. The sounds recorded and recognized by the SMiZW are site-specific. Therefore, it is only possible to obtain a high-quality training dataset after the system is installed at the target site. In the SMiZW, the source patterns (training dataset) can be obtained in three ways (Fig. 8.7):

- Method 1 – using the initial database with universal and representative sounds. This approach is effective when the system is going to be installed in a typical environment with well-defined and common noise sources.

- Method 2 – data is gathered for the first few days or weeks in the target environment once the system is installed. Sensors deliver data, which is manually labeled and imported into the database.
- Method 3 – utilizing the synergy between the clustering and DNN algorithms to enrich the database with unknown signals (circular process marked in green in Fig. 8.6). The clustering algorithm gives suggestions to the user on grouping the set of unidentified signals. The user can verify these suggestions by listening to sound files associated with each signal and assigning an appropriate name for each class. Finally, the prepared and verified data can be imported into the database. The reader is referred to other articles where the topic of detecting new classes is thoroughly discussed [20], [21], [22].

Methods 2 and 3 require a long data collection process before the training dataset can be considered large enough for a DNN. To deal with this, the system currently augments the data by mixing clean sound samples with background noises. This method proved to be successful, but further improvements are planned, such as FSL techniques (meta-learning) [23].

8.4. Summary

In this chapter, an advanced system for monitoring and identifying vibroacoustic threats was presented. It was shown that integrating machine learning algorithms (DNN and clustering) with deterministic algorithms results in a more complete picture during the acoustic analysis of areas exposed to excessive noise. In particular, the interaction between the clustering and classification algorithms is discussed, and the complementarity of the results from the deterministic algorithms against the identified classes was discussed. Information on tonality, spectral character, impulsivity, and location provide valuable clues to accurately identify problematic noise sources if a given event cannot be detected with high probability by DNN algorithms. After listening to sound samples and applying the clustering algorithm, the system's user can feed the DNN algorithm's training set with new classes.

Thanks to these features, the system can improve its predictive capabilities over time. The system can perform a full, automated acoustic analysis of the noisy area in real time, including the determination of a noise source ranking. This allows for automatic analysis of a noise to identify hazards that may result in the permissible noise levels being exceeded and to detect vibroacoustic phenomena that may indicate a malfunction of machines, devices, or production lines. This solution could have functionality in both measuring noises and identifying their sources (including sources that exceed the permissible levels), and as a technological trailer (predictor) of machinery or equipment failures. This can indicate devices or areas of the factory that require urgent intervention for machinery maintenance.

Acknowledgements: (in Polish) Zadanie jest realizowane w ramach Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu społecznego i Funduszu Spójności na lata 2014-2020 w ramach projektu: nr POIR.01.01.01-00-1408/19 pn. „System monitoringu i identyfikacji zagrożeń wibroakustycznych - SMIZW", realizowanego w ramach Poddziałania 1.1.1 Programu Operacyjnego Inteligentny Rozwój 2014-2020 współfinansowanego ze środków Europejskiego Funduszu Rozwoju Regionalnego.

References

[1] Masterson E.A., Deddens J.A., Themann C.L., Bertke S., Calvert G.M., *Trends in worker hearing loss by industry sector*, 1981–2010, „American Journal of Industrial Medicine" 2015, Vol. 58, No. 4, s. 392-401.

[2] Themann C.L., Masterson E.A., *Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden*, „The Journal of the acoustical society of America" 2019, Vol. 146, No. 5, s. 3879–3905.

[3] Directive 2003/10/EC – noise: of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise) (Seventeenth individual Directive within the meaning of Article 16(1) of Directive 89/391/EEC).

[4] *Rozporządzenie Ministra Rodziny, Pracy i Polityki Społecznej z dnia 12 czerwca 2018 r. w sprawie najwyższych dopuszczalnych stężeń i natężeń czynników szkodliwych dla zdrowia w* środowisku *pracy* (Dz.U. 2018 poz. 1286 z późn. zm.).

[5] *Obwieszczenie Ministra* Środowiska *z dnia 15 października 2013 r. w sprawie ogłoszenia jednolitego tekstu rozporządzenia Ministra* Środowiska *w sprawie dopuszczalnych poziomów hałasu w* środowisku (Dz.U. 2014 poz. 112).

[6] Pierre Jr R.L.S., Maguire D.J., Automotive C.S. (2004, July). *The impact of A-weighting sound pressure level measurements during the evaluation of noise exposure*, Conference NOISE-CON, s. 12–14.

[7] Rajala V., Hongisto V., *Annoyance penalty of impulsive noise–The effect of impulse onset*, „Building and Environment" 2020, Vol. 168.

[8] Piczak K.J., *Environmental sound classification with convolutional neural networks*, *„IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)",* Boston 2015, s. 1–6.

[9] Cakir E., Parascandolo G., Heittola T., Huttunen H., Virtanen T., *Convolutional recurrent neural networks for polyphonic sound event detection*, „IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2017, Vol. 25, No. 6, s. 1291–1303.

[10] Zhang X., Zou Y., Shi W., *Dilated convolution neural network with LeakyReLU for environmental sound classification, „22nd International Conference on Digital Signal Processing (DSP)",* London 2017, s. 1–5.

[11] Hochreiter S., Schmidhuber J., *Long short-term memory*, „Neural computation" 1997, Vol. 9, No. 8, s. 1735–1780.

[12] Hu J., Shen L., Sun G., *Squeeze and excitation networks*, „Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2018, s. 7132–7141.

[13] Sibson R., *SLINK: an optimally efficient algorithm for the single-link cluster method*, „The Computer Journal" 1973, Vol. 16, No. 1, s. 30–34.

[14] Ester M., Kriegel H.P., Sander J., Xu X., *A density-based algorithm for discovering clusters in large spatial databases with noise*, „KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining", s. 226–231.

[15] Ward Jr J.H., *Hierarchical grouping to optimize an objective function*, „Journal of the American statistical association" 1963, Vol. 58, No. 301, s. 236–244.

[16] Maaten, L. van der, Hinton G., *Visualizing data using t-SNE*, „Journal of Machine Learning Research" 2008, Vol. 9, No. 11.

[17] ISO 1996-2:2017, *Acoustics — Description, measurement and assessment of environmental noise*, cz. 2, *Determination of sound pressure levels.*

[18] Nordtest method NT ACOU 112:2002 – *Acoustics – Prominence of impulsive sounds and for adjustment of LAeq*, Approved 2002-05, Taastrup, Denmark.

[19] Wiederhold C.P., Gee K.L., Blotter J.D., Sommerfeldt S.D., *Comparison of methods for processing acoustic intensity from orthogonal multimicrophone probes*, „The Journal of the Acoustical Society of America" 2012, Vol. 131, No. 4, s. 2841–2852.

[20] Chen G., Peng P., Wang X.,Tian Y., *Adversarial reciprocal points learning for open set recognition*, arXiv preprint arXiv:2103.00953, 2021.

[21] Han K., Rebuffi S.A., Ehrhardt S., Vedaldi A., Zisserman A., *Autonovel: Automatically discovering and learning novel visual categories,* „IEEE Transactions on Pattern Analysis and Machine Intelligence", 2021.

[22] Vaze S., Han K., Vedaldi A., Zisserman A., *Open-set recognition: A good closed-set classifier is all you need*, arXiv preprint arXiv:2110.06207, 2021.

[23] Parnami A., Lee M., *Learning from Few Examples: A Summary of Approaches to Few-Shot Learning*, arXiv preprint arXiv:2203.04291, 2022.

Integration of machine learning techniques and deterministic algorithms within an advanced system for monitoring and identifying vibroacoustic threats

This chapter addresses the issue of complementarity and synergy between different algorithms within a complex system. Using a concrete example of an advanced system for monitoring and identifying vibroacoustic threats, it is demonstrated how, by integrating different signal processing methods, it is possible to achieve the goals of the system at different levels of accuracy. The multi-level signal analysis included automatic, real-time labeling of acoustic waveforms (classification), grouping of unknown signals to update the training database, and additional deterministic calculations. Deterministic algorithms assure that a minimum of relevant information can always be extracted during the acoustical analysis. In that sense, deterministic digital signal processing proved to be an important complement to machine learning algorithms. The synergy between machine learning algorithms (supervised and unsupervised) and deterministic DSP (determination of tonality, impulsivity, spectral characteristics, and direction of wave arrival) made it possible to meet customers' needs better than using individual algorithms in isolation.

9. Koncepcja matrycy falowodowej do kształtowania frontu fali akustycznej

Tomasz Nowak, Andrzej Dobrucki

Politechnika Wrocławska,
Wydział Elektroniki, Fotoniki i Mikrosystemów,
Wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław

9.1. Wprowadzenie

W nauce oraz inżynierii związanej z projektowaniem przetworników elektroakustycznych i urządzeń głośnikowych pojawiają się skomplikowane ograniczenia wynikające z niedopasowania kątowego zakresu promieniowania źródeł, a także kształtów ich frontów falowych. Fizyczne wymiary i rozmieszczenie głośników, w tym urządzeń głośnikowych, w relacji do generowanych długości fali związane są z występowaniem interferencji na powierzchni wspólnego frontu falowego. W wyniku niedopasowania fazowego i amplitudowego powstaje źródło dźwięku zniekształcające sygnał akustyczny w zmienny sposób w całej objętości pola akustycznego. W celu rozwiązania wymienionych problemów w niniejszym rozdziale zaproponowano zastosowanie soczewki w formie matrycy falowodów, które dzielą czoło fali na skończoną ilość fragmentów oraz wprowadzają kontrolowane opóźnienie każdego z nich. W efekcie powstaje narzędzie wpływające na rozkład fazy na powierzchni źródła. Ta koncepcja była już przedstawiona w publikacjach naukowych [1], [2]. Dotychczasowe rozwiązania opierały się na obrocie macierzy wyjściowej względem wejściowej.

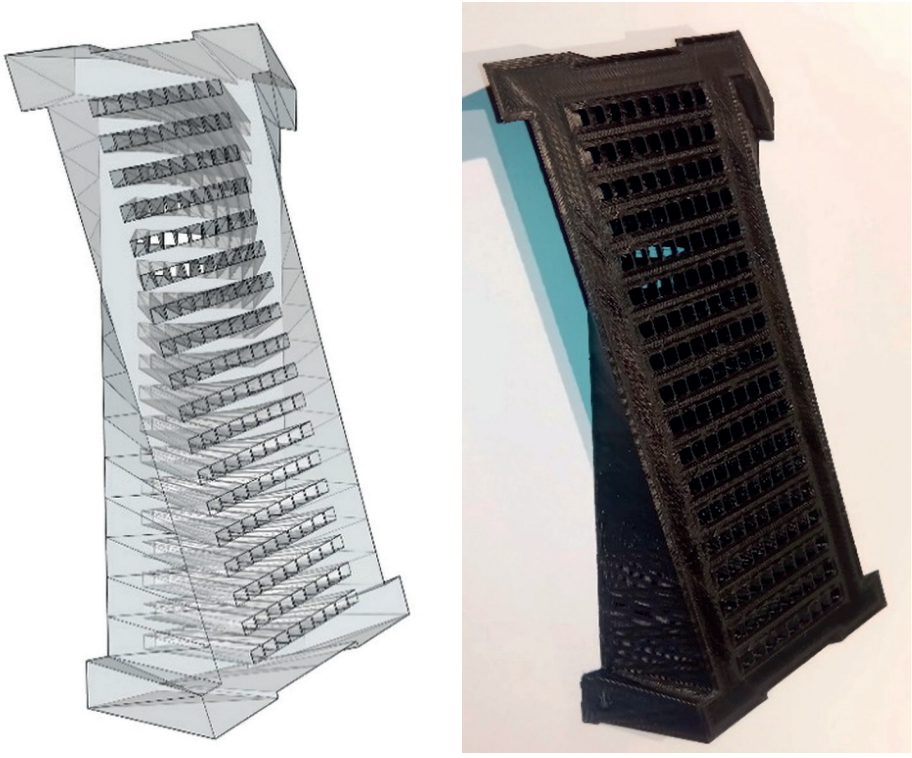
9.2. Stan obecny

Dowolny wycinek powierzchni w polu akustycznym można podzielić na fragmenty i utworzyć z nich macierz znajdującą się na wejściu soczewki. Wyjściem soczewki jest

macierz o tej samej liczbie komórek, lecz o innej orientacji przestrzennej. Komórki macierzy wejściowej i wyjściowej są połączone kanałami. Różnica w rozmieszczeniu komórek między dwiema macierzami wprowadza różnicę długości kanałów je łączących, dzięki temu można regulować opóźnienie fazowe w każdej z komórek wyjściowych.

Najprostszy przypadek, w którym rozmiary komórek w obu macierzach są jednakowe, a przetwornik i soczewka osiowosymetryczne opisano w pracy [2]. Chodziło o zwiększenie kąta promieniowania przetwornika ultradźwiękowego. Regulację opóźnień uzyskano przez obrót macierzy wyjściowej względem wejściowej wokół osi symetrii połączonych helikalnymi kanałami. Opóźnienie jest najmniejsze blisko osi symetrii, gdzie długość kanałów jest najmniejsza – stopniowo zwiększają się ono w kolejnych koncentrycznych rzędach.

Przedstawione rozwiązania zostały zastosowane w celu zakrzywienia frontu falowego przetwornika średniotonowego [1]. Kanały łączące macierze miały stałą długość, a opóźnienie wynikało z odległości komórek od punktu pomiarowego przez krzywienie wylotu soczewki.



Rys. 9.1. Projekt obrotowej soczewki akustycznej dla przetwornika AMT (z lewej) oraz wydrukowany prototyp (z prawej)

Koncepcję obracania macierzy wyjściowej można zastosować także w innych kształtach przetworników (rys. 9.1). Źródłem fali płaskiej jest głośnik wstęgowy typu AMT (Air Motion Transformer) charakteryzujący się generowaniem prostokątnego wycinka fali powierzchniowej, w którym wszystkie punkty czoła fali na powierzchni wylotu są w tej samej fazie. Problemem związanym z osiąganiem wysokiej sprawności, szerokiego pasma przenoszenia i dużej mocy znamionowej tego typu głośników jest konieczność stosowania membrany o dużym polu powierzchni czynnej i, co się z tym łączy, ze znacznymi rozmiarami wylotu przetwornika, często wielokrotnie przekraczającymi długość generowanych fal akustycznych. Zastosowanie soczewki umożliwia osiągnięcie dowolnych kątów promieniowania takiego głośnika w płaszczyźnie pionowej – niezależnie od jego rozmiarów, co zostało potwierdzone pomiarami zgodnymi z wynikami przedstawionymi w wymienionych wcześniej publikacjach naukowych [1], [2].

Dotychczasowe rozwiązania cechują się istotnym ograniczeniem: obrotowe soczewki pozwalają na regulację krzywizny czoła fali tylko w jednym kierunku. W przypadku soczewek o kształcie wielokąta foremnego lub koła za pomocą kąta obrotu oraz grubości struktury kształtowaniu podlega promień krzywizny fali kulistej – nie ma możliwości osiągnięcia kształtu fali elipsoidalnej lub cylindrycznej. Dla innych kształtów ustroju parametry wpływają na większe zakrzywienie frontu falowego w kierunku większego wymiaru. A w przypadku prostokątnej soczewki obrotowej na wyjściu powstaje fala elipsoidalna, w której stosunek promieni elipsy odpowiada stosunkowi wymiarów struktury.

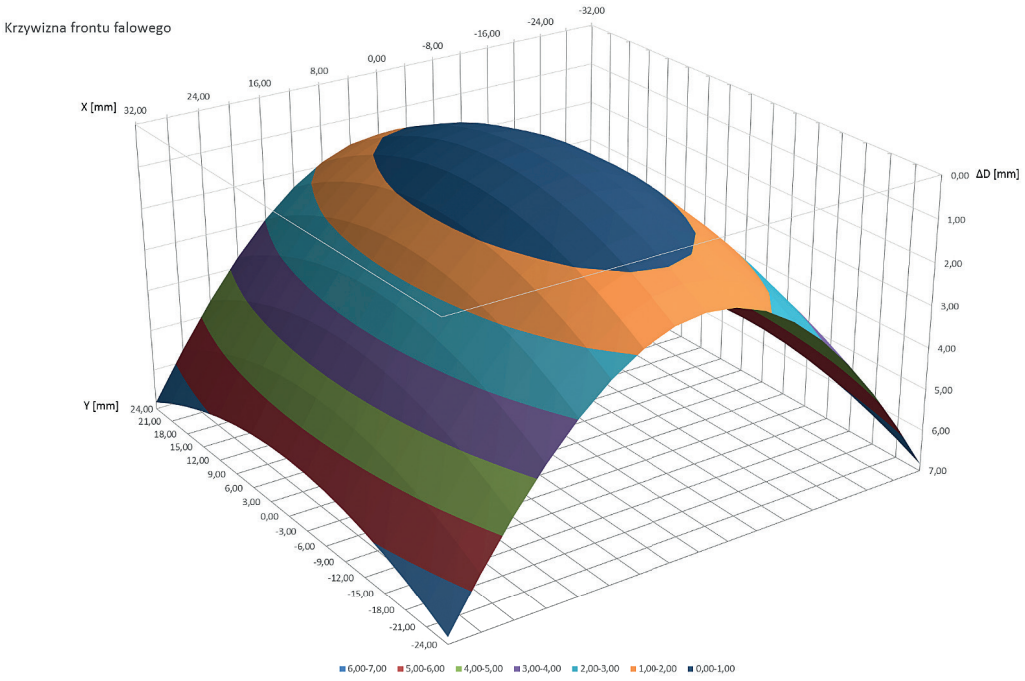
9.3. Proponowane rozwiązanie

Propozycją struktury dyskretyzującej umożliwiającej regulację krzywizny generowanej fali w dwóch płaszczyznach jest macierz wyjściowa o rozmiarach i pozycji innej niż wejściowa. Skalowanie rozmiaru oraz pozycji wyjściowej macierzy w obu kierunkach pozwala na kształtowanie czoła fali w postaci na przykład płaskich, cylindrycznych, elipsoidalnych krzywizn przestrzennych.

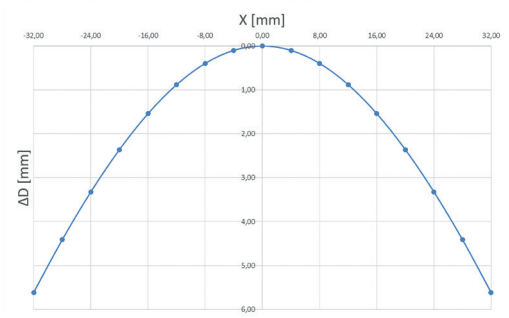
Po przyjęciu wymiarów i rozkładu komórek macierzy ze wzoru (9.1) można obliczyć dla każdej komórki macierzy drogę, czyli długość każdego tunelu łączącego komórkę wejściową z wyjściową. Gdy macierz wyjściowa znajdująca się w pozycji $t = 20$ mm (grubość soczewki) zostanie powiększona o iloczyn współrzędnych X_i ze współczynnikiem skalującym A_x oraz analogicznie współrzędnych Y_i ze współczynnikiem skalującym A_y, otrzyma się macierz wyjściową o symetrycznym rozkładzie opóźnień, których wartość rośnie dla każdej komórki wraz z odległością od środka symetrii. Dystrybucję opóźnień $\Delta t_{m,n}$ dla każdej z komórek wyjściowych obliczono ze wzoru (9.2).

$$D_{m,n} = \sqrt{\left(A_x X_m - X_m\right)^2 + \left(A_y Y_n - Y_n\right)^2 + t^2} \tag{9.1}$$

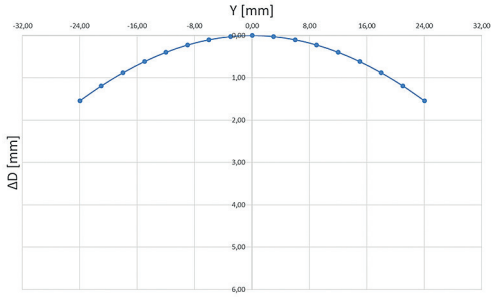
$$\Delta t_{m,n} = \frac{D_{m,n} - h}{c} \tag{9.2}$$



Rys. 9.2. Wizualna reprezentacja frontu falowego za pomocą przestrzennej dystrybucji przyrostu drogi ΔD w przypadku soczewki o grubości $t = 20$ mm i współczynnikach skalowania $A_x = 2$, $A_y = 1{,}5$



Rys. 9.3. Krzywizna frontu falowego w osi X, dystrybucja przyrostu drogi ΔD w przypadku soczewki o grubości $t = 20$ mm i współczynnikach skalowania $A_x = 2$, $A_y = 1{,}5$



Rys. 9.4. Krzywizna frontu falowego w osi Y, dystrybucja przyrostu drogi ΔD w przypadku soczewki o grubości $t = 20$ mm i współczynnikach skalowania $A_x = 2$, $A_y = 1{,}5$

gdzie $D_{m,n}$ – droga między odpowiadającymi sobie komórkami macierzy wejściowej i wyjściowej, $\Delta t_{m,n}$ – opóźnienie w poszczególnych komórkach soczewki, A_x A_y – współczynniki skalowania macierzy wyjściowej soczewki, kolejno dla osi X i Y.

Wykorzystując obliczony przyrost drogi dla każdej z komórek wyjściowych soczewki, można wizualnie przedstawić front falowy jako powierzchnię łączącą punkty tego przyrostu na wykresie przestrzennym (rys. 9.2). Na podstawie danych znajdujących się na osiach symetrii frontu falowego wyznaczono natomiast krzywiznę w kierunku X (rys. 9.3) i Y (rys. 9.4). Powyższy przykład możliwości niezależnego sterowania dwiema krzywiznami frontu falowego przedstawia narzędzie oferujące większą kontrolę niż przykłady zawarte we wcześniejszych publikacjach.

9.4. Ustroje parametryczne

Kolejnymi możliwościami, którymi cechują się proponowane ustroje, jest asymetryczne skalowanie za pomocą dowolnej funkcji oraz przesunięcie pozycji macierzy wyjściowej w dowolnym kierunku. Aby zrealizować przyrost odległości między komórkami macierzy wyjściowej w kierunku jednego z wymiarów soczewki, w tym przykładzie jest to kierunek Y, zastosowano współczynnik A_m, gdzie $i = 1$, $m = 1$ odpowiada pierwszemu, górnemu rzędowi komórek wyjściowych:

$$A_m = BA_{m-1} + C \tag{9.3}$$

$$Y_i = Y_m + A_m \tag{9.4}$$

gdzie: A_m – jednowymiarowa macierz współczynników, B – współczynnik paraboliczny, C – współczynnik liniowy, Y_i – jednowymiarowa macierz współrzędnych Y wyjściowej macierzy po skalowaniu parametrycznym.

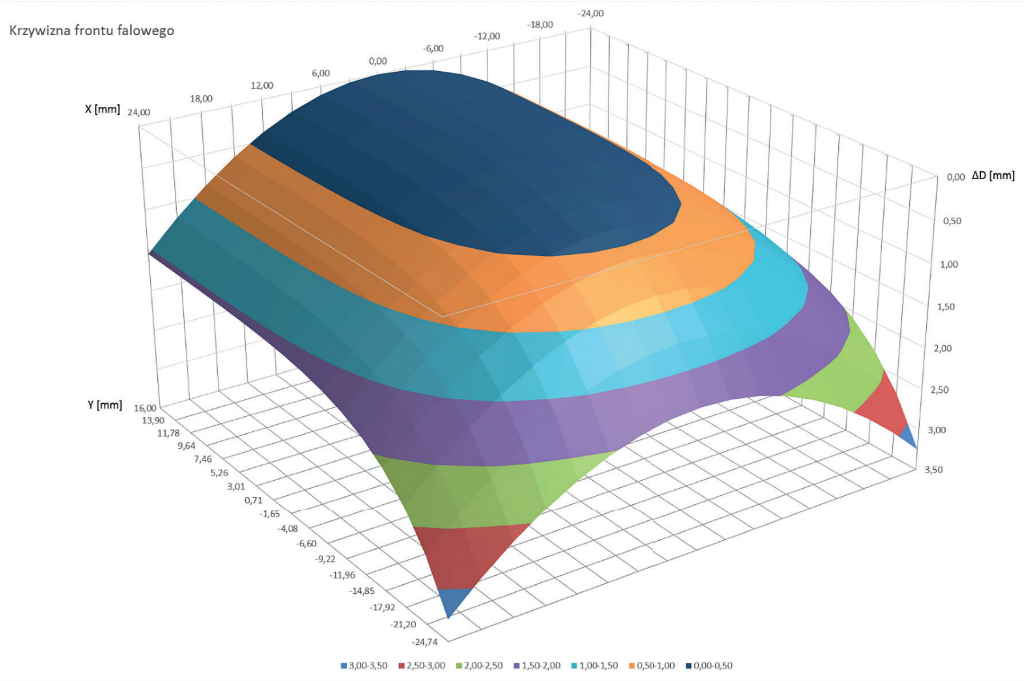
Z uwzględnieniem przyjętych parametrów B oraz C otrzymano ze wzorów (9.3) i (9.4) tabele wartości współczynników (tabela 9.1) oraz przeliczonych wartości współrzędnych Y wyjściowej macierzy (tabela 9.2). Współczynnik skalowania w tym przykładzie $A_x = 1{,}5$, współczynnik A_y pominięto, aby widoczny był tylko wpływ parametrycznego skalowania na krzywiznę frontu fali w osi Y (rys. 9.7).

Tabela 9.1. Transponowana macierz współczynników A_m, gdy $B = 1{,}2$ i $C = -0{,}1$

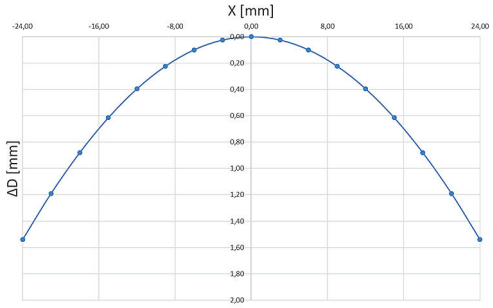
0,00	–0,10	–0,22	–0,36	–0,54	–0,74	–0,99	–1,29	–1,65	–2,08	–2,60	–3,22	–3,96	–4,85	–5,92	–7,20	–8,74

Tabela 9.2. Transponowana macierz obliczonych wartości współrzędnych Y_i [mm], gdy $B = 1{,}2$ i $C = -0{,}1$

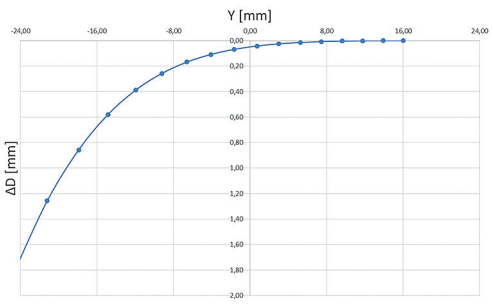
16,00	13,90	11,78	9,64	7,46	5,26	3,01	0,71	–1,65	–4,08	–6,60	–9,22	–11,96	–14,85	–17,92	–21,20	–24,74



Rys. 9.5. Wizualna reprezentacja frontu falowego za pomocą przestrzennej dystrybucji przyrostu drogi ΔD w przypadku soczewki o grubości t = 20 mm i współczynnikach skalowania A_x = 1,5, A_m = 1,2, A_{m-1} – 0,1



Rys. 9.6. Krzywizna frontu falowego w osi X, dystrybucja przyrostu drogi ΔD w przypadku soczewki o grubości t = 20 mm i współczynnikach skalowania A_x = 1,5, A_m = 1,2 A_{m-1} – 0,1



Rys. 9.7. Krzywizna frontu falowego w osi Y, dystrybucja przyrostu drogi ΔD w przypadku soczewki o grubości t = 20 mm i współczynnikach skalowania A_x = 1,5, A_m = 1,2, A_{m-1} – 0,1

Tak jak w poprzednim przykładzie dokonano wizualnej reprezentacji frontu falowego w przestrzeni (rys. 9.5) i jego zakrzywienia w osi *X* (rys. 9.6). Reprezentacja frontu falowego pokazuje płynne przejście z cylindrycznego kształtu frontu fali w kulisty.

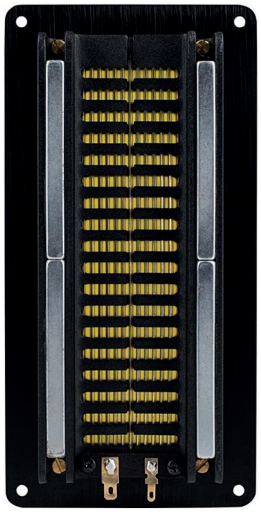
Tego typu źródło cechuje się regulowaną kierunkowością i rozkładem natężenia w funkcji wymiaru pionowego źródła. Omawiane soczewki są przykładami wielu osiągalnych typów frontów falowych. Jedynym ograniczeniem w kształtowaniu dystrybucji opóźnienia czoła fali w poszczególnych komórkach macierzy wyjściowej jest wykonalność takiej struktury, czyli warunek, aby kanały łączące obie macierze się nie przecinały.

9.5. Część eksperymentalna

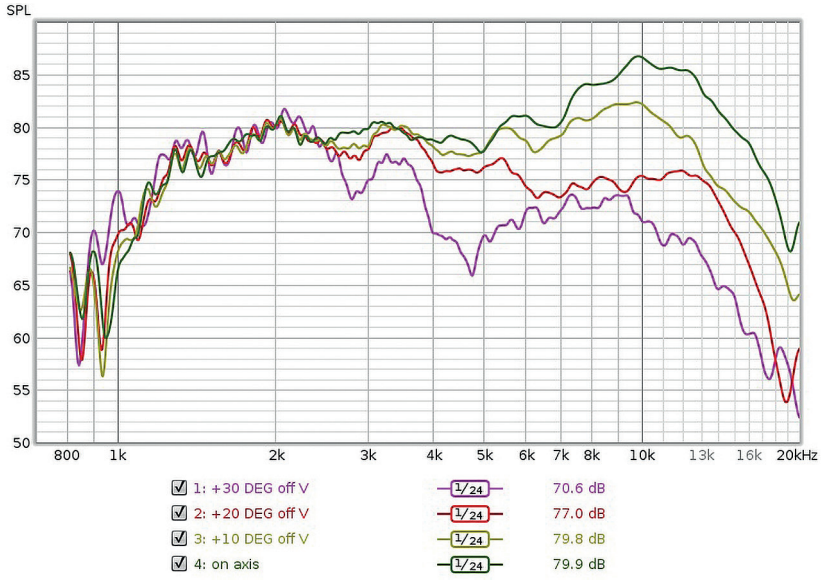
Źródłem fali płaskiej w rozpatrywanym przykładzie jest głośnik wstęgowy typu AMT (rys. 9.8). Wszystkie punkty czoła fali na powierzchni wylotu przetwornika są w tej samej fazie. Membrana wybranego modelu ma znaczące pole powierzchni czynnej oraz znaczne rozmiary wylotu przetwornika – wielokrotnie przekraczające długość generowanych fal akustycznych.

Wszystkie pomiary wykonano skalibrowanym mikrofonem w odległości 0,5 m od powierzchni membrany. Zarówno przetwornik bez soczewki, jak i przetwornik z soczewką był zamontowany w odgrodzie w warunkach bezechowych.

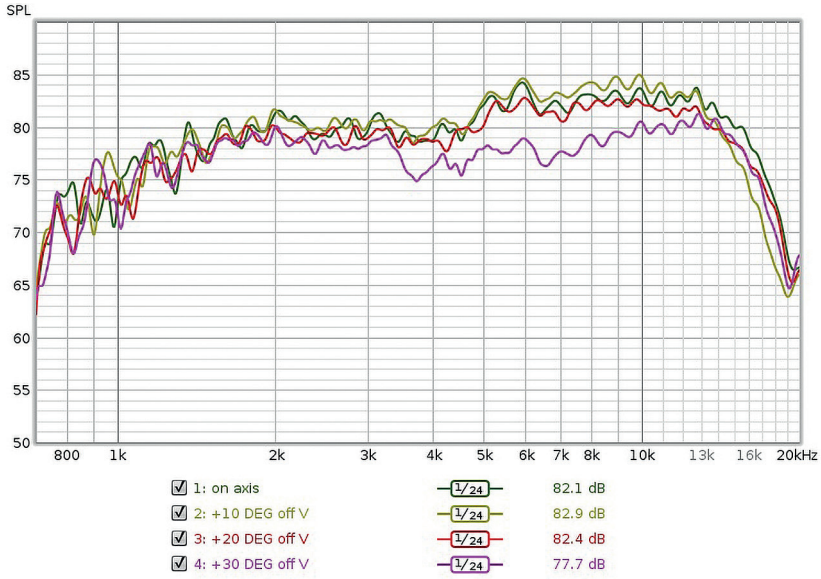
Przetwornik bez soczewki w płaszczyźnie dłuższego wymiaru charakteryzuje się znaczną kierunkowością szczególnie w zakresie dużych częstotliwości. Różnica w po-



Rys. 9.8. Przetwornik AMT, dla którego wykonano soczewkę



Rys. 9.9. Pomiar odpowiedzi amplitudowej przetwornika AMT bez soczewki na osi symetrii i pod kątem 10°, 20°, 30° poza nią w płaszczyźnie większego wymiaru



Rys. 9.10. Pomiar odpowiedzi amplitudowej przetwornika AMT z soczewką na osi symetrii i pod kątem 10°, 20°, 30° poza nią w płaszczyźnie większego wymiaru

ziomie ciśnienia akustycznego 30° od osi symetrii względem pomiaru na osi wyniosła ok. 15 dB dla częstotliwości 10 kHz (rys. 9.9). Wykonano prototyp soczewki z tego przykładu, który znacząco zwiększył jego kątowy zakres promieniowania w płaszczyźnie pionowej. Różnica poziomów ciśnienia akustycznego w przypadku tych samych kątów oraz częstotliwości wyniosła ok. 3 dB (rys. 9.10). Działanie urządzenia zostało potwierdzone pomiarami stanowiącymi dowód na zakrzywienie frontu falowego badanego przetwornika.

Różnica między zmierzonymi odpowiedziami amplitudowymi przetwornika bez soczewki (rys. 9.9) oraz z soczewką (rys. 9.10) wskazuje, że badana struktura ma wyraźny wpływ na kątowy zakres promieniowania. Z wykresów można wywnioskować, że zastosowanie soczewki znacząco zmniejsza kierunkowość przetwornika i wyrównuje poziom ciśnienia akustycznego w przestrzeni.

9.6. Podsumowanie

Zaletami proponowanego rozwiązania są:

- rozszerzenie możliwości kształtowania frontu falowego;
- nieskomplikowane obliczenia, prosta konstrukcja;
- obiecujące wyniki eksperymentalne potwierdzające działanie;
- małe wymiary (w szczególności grubość), co stanowi atrakcyjną alternatywę rozwiązań tubowych.

Rozwiązanie to nie jest jednak wolne od ograniczeń, za które należy uznać:

- brak możliwości skupiania, a jedynie rozpraszania wiązki – w przedstawianych przykładach nie ma możliwości wydłużania wewnętrznych tuneli względem zewnętrznych (ten sam problem dotyczy soczewek obrotowych);
- trudności w dopasowaniu amplitudowym matrycy wielu soczewek – zewnętrzne komórki są połączone długimi kanałami pod katem znacząco odbiegającym od prostopadłego do membrany, co może mieć wpływ na zwiększenie tłumienia.

Bibliografia

[1] Berstis V., *3D Printed Acoustic Lens for Dispersing Sound*, „Journal of the Audio Engineering Society" 2018, Vol. 66, No. 12, s. 1082–1093.

[2] Dahl T., Ealo J.L., Papakonstantinou K., Pazos J.F., *Design of Acoustic Lenses for Ultrasonic Human-Computer Interaction*, IEEE International Ultrasonics Symposium, 2011.

[3] Heil C., Unitet States Patent Number: 5, 163, 167 (US005163167A), 1992.

Słowa kluczowe: elektroakustyka, przetwornik, soczewka akustyczna.

Koncepcja matrycy falowodowej do kształtowania frontu fali akustycznej

Autorzy proponują zastosowanie soczewki w postaci matrycy falowodów, które dzielą czoło fali na skończoną liczbę fragmentów wprowadzając kontrolowane opóźnienie każdego z nich. Rezultatem jest dyskretyzująca struktura, która pozwala na regulację krzywizny generowanej fali w dwóch płaszczyznach. Osiąga się to za pomocą macierzy wyjściowej o rozmiarze i położeniu innym niż wejściowa. Skalowanie rozmiaru i położenia macierzy wyjściowej w obu kierunkach pozwala na kształtowanie czoła fali jako płaskich, cylindrycznych, elipsoidalnych i innych krzywizn przestrzennych.

W rozdziale obliczono i przedstawiono wiele przykładów frontów falowych, w tym parametrycznych, w których następuje zmiana kształtu w funkcji jednego lub obu wymiarów macierzy wyjściowej. Wykonano prototypową soczewkę dla przetwornika AMT, która znacznie zwiększyła jego kąt promieniowania w płaszczyźnie pionowej. Działanie urządzenia zostało potwierdzone pomiarami.

Waveguide matrix concept for shaping the acoustic wave front

The authors propos the use of a lens in the form of a matrix of waveguides, which divide the wave front into a finite number of fragments and introduce a controlled delay of each of them. The result is a discretizing structure that allows to adjust the curvature of the generated wave in two planes. It is achieved by an output matrix of a size and position other than the input one. Scaling the size and position of the output matrix in both directions allows the wavefront to be shaped as flat, cylindrical, ellipsoidal, and other spatial curvatures.

Multiple examples of wave fronts were calculated and presented in this paper, including parametric ones, where there is a change in shape as a function of one or both dimensions of the output matrix. A prototype lens was made for the AMT transducer, which significantly increased its angular radiation range in the vertical plane. The operation of the device has been confirmed by measurements.

10. Projektowanie i tworzenie systemów zarządzających w niekonwencjonalnych instalacjach dźwięku przestrzennego

Jan Skorupa, Maciej Głowiak

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

10.1. Wprowadzenie

Wzajemna współpraca środowisk naukowych z artystycznymi zaczęła się jeszcze w XX w. Inżynierowie w połączeniu z artystami poszukiwali nowych narzędzi do wykorzystania w muzyce elektronicznej, w której źródłem dźwięku przestaje być fizyczny instrument, a staje się głośnik odtwarzający nagranie. We wczesnych latach 50. XX w. P. Schaeffer wraz ze swoim asystentem P. Henrym we współpracy z inżynierami z Radiodiffusion-Télévision Française opracowali system wielokanałowego odsłuchu pięciu kanałów audio do czterech niezależnych głośników ustawionych wokół publiczności [3]. Praktyki tego typu były często podejmowane w kolejnych latach przez takich twórców, jak I. Xenakis czy K. Stockhausen. Xenakis – kompozytor oraz architekt – w 1958 r. wraz z Le Corbusierem zaprojektowali pawilon multimedialny Philipsa zaprezentowany podczas wystawy „Expo" w Brukseli. W pawilonie tym powstała instalacja audio-wizualna składająca się z 425 głośników oraz oświetlenia zsynchronizowanego z warstwą dźwiękową. Podczas demonstracji odtwarzana była kompozycja E. Varèse'a *Poème Électronique* i I. Xenakisa *Concert Ph* [3]. Stockhausen w 1970 r. był odpowiedzialny za stworzenie kompozycji dźwiękowych zaprezentowanych w pawilonie podczas „Expo" w Osace [8]. Na potrzeby utworów stworzono wtedy sferyczną instalację składającą się z 50 głośników. Wymienieni eksperymentatorzy w trakcie swoich poszukiwań opracowywali rozwiązania znacząco wykraczające poza ówcześnie znane systemy dźwięku przestrzennego – proponowali rozwiązania nowatorskie budzące zainteresowanie zarówno wśród kompozytorów muzyki elektronicznej, jak i inżynierów dźwięku. Praktyki tego typu obecne są do dziś, a ze względu na znacznie większe

możliwości techniczne tworzone systemy dają pełną swobodę dystrybucji kanałów oraz miksowania ich w dowolnych konstelacjach głośnikowych.

10.2. Cel przedsięwzięć

W Poznańskim Centrum Superkomputerowo-Sieciowym od 2016 r. prowadzi się badania nad wykorzystaniem dźwięku przestrzennego w projektach badawczych i rozwojowych, szczególnie w kontekście aplikacji związanych z wirtualną rzeczywistością i immersją. W 2018 roku w ramach projektu *Immersify* [5] PCSS stworzyło uniwersalne *workflow* do prowadzenia testów i eksperymentów związanych z dźwiękiem ambisonicznym i przestrzennym [6]. Zdobyte doświadczenie jest wykorzystywane w interdyscyplinarnych projektach artystycznych w kontekście adaptacji rozwiązań w eksperymentalnych wielokanałowych systemach dystrybucji dźwięku – umożliwia to testowanie i zastosowanie powszechnie niedostępnych rozwiązań związanych z dźwiękiem przestrzennym w praktycznych przedsięwzięciach artystycznych.

W niniejszym rozdziale przedstawiono dwie instalacje dźwięku przestrzennego zrealizowane we współpracy między PCSS a środowiskiem artystycznym *Akusmonium*, która powstała na potrzeby Fazma Festiwal, i system odsłuchowy zaprojektowany jako część instalacji zatytułowanej *Pokój do słuchania* autorstwa kompozytorki i piosenkarki Hani Rani oraz studia architektury Zmir. Opisano proces powstawania, projektowania i programowania systemów komputerowych zarządzających powstałymi instalacjami. Zaprezentowane opracowanie może stanowić inspirację do tworzenia kolejnych przedsięwzięć na styku inżynierii dźwięku oraz sztuki.

10.3. *Akusmonium*

10.3.1. Idea i cel

Akusmonium to projekt powstały w ramach programu pierwszej edycji festiwalu muzyki elektronicznej Fazma. Założeniem przedsięwzięcia było stworzenie instalacji wielogłośnikowej nawiązującej do historycznego *Acousmonium* zainicjowanego przez F. Byle'a we francuskiej GRM (Groupe de Recherches Musicales) [3]. Idea tej instalacji opierała się na stworzeniu orkiestry głośnikowej składającej się z kilku do kilkudziesięciu głośników, w których każdy jest inny – charakteryzuje się odmiennym pasmem przenoszenia, głośnością czy kierunkowością. Kompozytor w tym układzie przygotowywał utwór elektroniczny w wersji stereofonicznej, który podczas koncertu był panoramowany w przestrzeni wielogłośnikowej na żywo za pośrednictwem miksera audio przez sterowanie wzmocnieniem poszczególnych głośników. *Akusmonium* (rys. 10.1) to unowocześniona

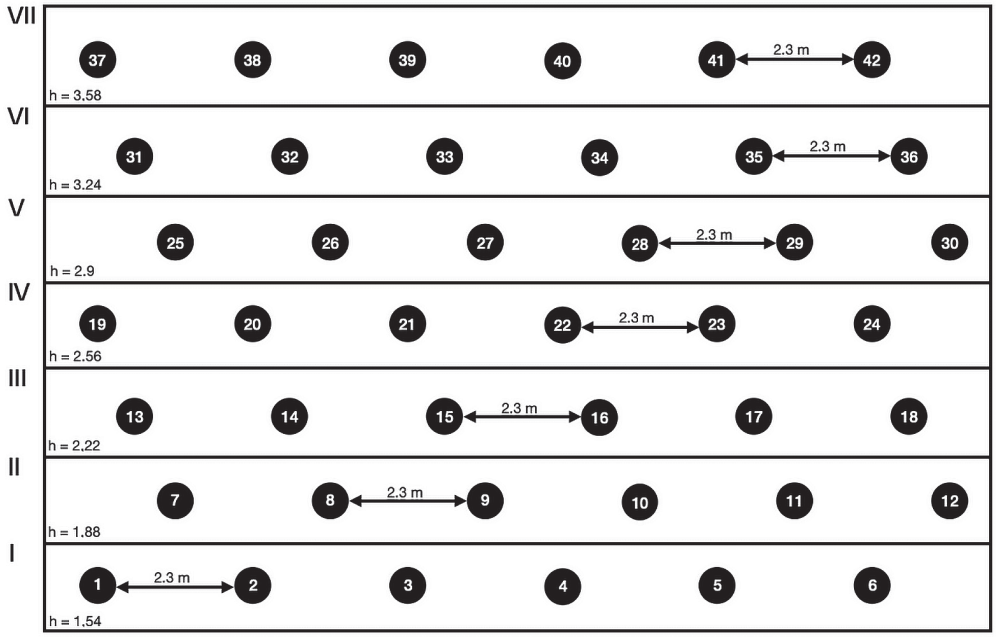
Rys. 10.1. Instalacja *Akusmonium* podczas koncertu w CK Zamek w Poznaniu

wersja historycznej instalacji Byle'a. W tym przypadku podstawowym założeniem było opracowanie systemu umożliwiającego niezależną kontrolę nad każdym kanałem oddzielnie oraz dystrybucję większej liczby sygnałów do poszczególnych głośników.

10.3.2. Opis instalacji

Historyczne *Acousmonium* to system składający się z matrycy kilkudziesięciu głośników ustawionych w sposób imitujący układ klasycznej orkiestry. Zgodnie z podstawowym założeniem głośniki adaptowały założenia orkiestry również pod względem barwy i brzmienia: część z nich o niższym paśmie przenoszenia występowała w *Acousmonium* analogicznie do obecności kontrabasów czy fagotów, a te o wyższym – analogicznie do obecności skrzypiec czy fletów. Podstawą obsługi instalacji było sterowanie na żywo wzmocnieniem poszczególnych głośników.

W przypadku instalacji *Akusmonium* projektowanej w ramach festiwalu Fazma zaproponowano układ uniwersalny i regularny. Zastosowanie głośników o tej samej charakterystyce i ich regularny układ umożliwiały panoramowanie utworu w sposób dowolny w obrębie dogłośnienia całej przestrzeni, bez ograniczeń pod względem barwy. Dodatkowo można było niezależnie kontrolować oraz dystrybuować sygnał między wszystkimi 43 kanałami audio, a dzięki temu realizować w pełni przestrzenne pozwalające na miks oraz animację dowolnej liczby ścieżek zawartych w utworze. Odległości



Rys. 10.2. Schemat ustawienia głośników w instalacji *Akusmonium*

Rys. 10.3. Schemat przedstawiający transmisję sygnału w instalacji *Akusmonium*

poszczególnych głośników zostały ustalone na podstawie subiektywnych testów i eksperymentów, tak aby ruch wirtualnego źródła między głośnikami był płynny i oferował słuchaczowi wrażenie spójności i koherentności sceny akustycznej.

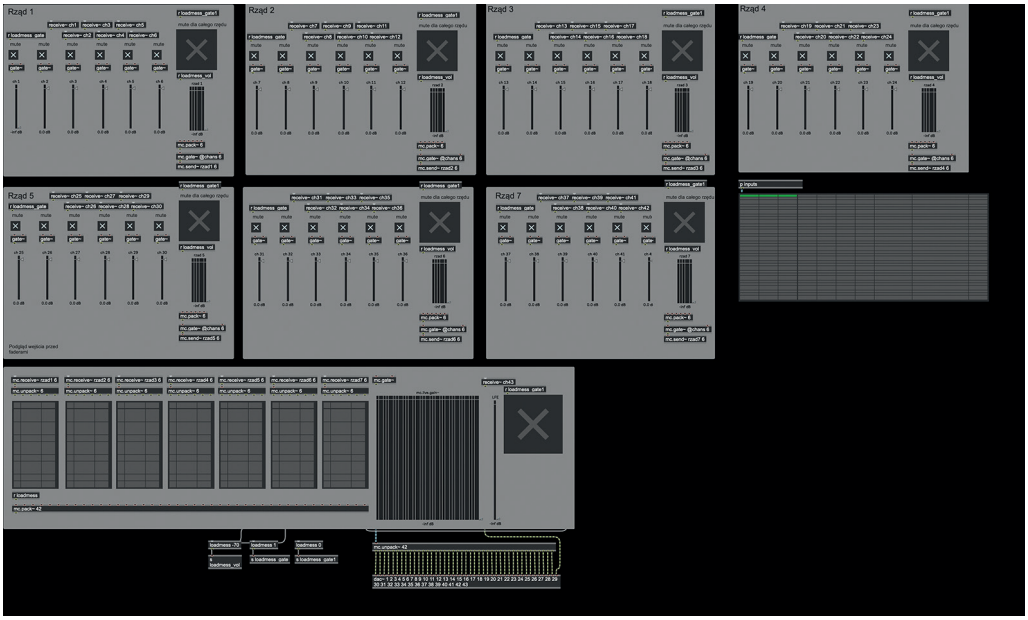
Technicznie docelowa instalacja składała się z 42 monitorów bliskiego pola Genelec 8010A. Głośniki ustawiono w siedmiu rzędach po sześć sztuk w każdym. Poszczególne głośniki w obrębie jednego rzędu były ustawione w odległości 2,30 m od siebie oraz znajdowały się na tej samej wysokości: rząd I – 1,54 m, rząd II – 1,88 m, rząd III – 2,22 m, rząd IV – 2,56 m, rząd V – 2,9 m, rząd VI – 3,24 m, rząd VII – 3,58 m. Ustawienie głośników w poszczególnych rzędach było przesunięte względem siebie, tak aby uniknąć powtarzania ich wzajemnej pozycji. Schemat ustawienia głośników przedstawiono na rys. 10.2.

W systemie sterującym instalacją były wykorzystane 43 kanały audio: 42 z nich kierowano bezpośrednio do głośników, a kanał 43 zarezerwowano dla ścieżki LFE (ang. *Low*

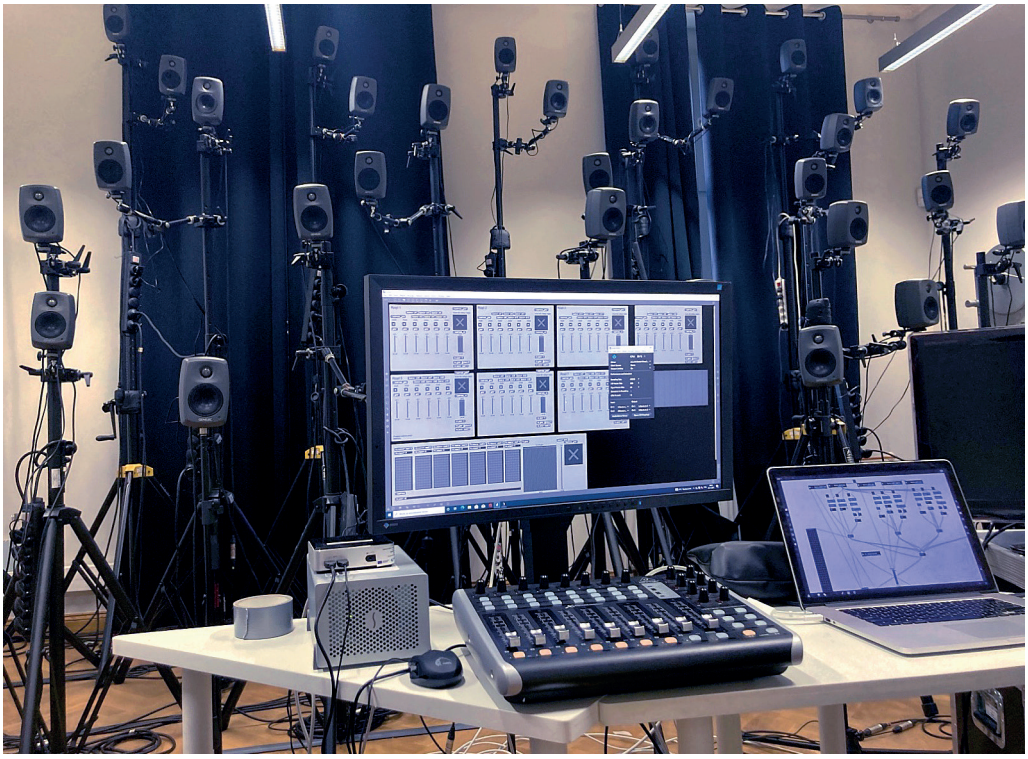
Frequency Effects). Tor audio z wykorzystaniem dwóch cyfrowych protokołów: Dante oraz ADAT. Dante służył do przesyłania sygnału audio między komputerem wykonawcy a aplikacją sterującą oraz doprowadzał sygnał do interfejsu audio. Sygnał z interfejsu był przesyłany za pośrednictwem sześciu połączeń ADAT do czterech oddzielnych przedwzmacniaczy. Każdy z przedwzmacniaczy miał po osiem wyjść analogowych, które doprowadzały sygnał bezpośrednio do głośników. Instalacja *Akusmonium* stwarzała możliwość pracy w częstotliwości próbkowania 44,1 kHz lub 48 kHz. Schemat toru transmisji sygnału audio przedstawiono na rys. 10.3.

10.3.3. Aplikacja sterująca

Za pośrednictwem środowiska programistycznego max/MSP [9] stworzono mikser pozwalający na podgląd sygnału i kontrolę głośności w torze audio niezależnie dla każdego kanału. Aplikacja obsługiwana jest za pośrednictwem kontrolera MIDI Behringer X-Touch Compact. Z pozycji kontrolera MIDI możliwe jest sterowanie głośnością poszczególnych rzędów głośnikowych, kontrola głośności kanałów natomiast jest możliwa z pozycji GUI aplikacji. Aplikacja dodatkowo wyposażona jest w ogranicznik (Limiter) zabezpieczający instalację. Na rysunku 10.4 przedstawiono GUI aplikacji w środowisku max/MSP.



Rys. 10.4. Aplikacja zarządzająca sygnałem w instalacji



Rys. 10.5. Instalacja prototypowa *Akusmonium*

10.3.4. Instalacja produkcyjna

Instalacja *Akusmonium* stanowiła temat miesięcznej rezydencji artystycznej, w której ramach zaproszeni kompozytorzy z wykorzystaniem prototypowej instalacji stworzyli utwory elektroniczne testujące oraz eksplorujące możliwości systemu odsłuchowego. Instalacja produkcyjna była pomniejszoną do 40% instalacją zaprezentowaną podczas koncertu festiwalowego. Na rysunku 10.5 przedstawiono środowisko testowe oraz wygląd instalacji prototypowej w laboratorium PSNC Future Labs.

10.3.5. Wnioski

Instalacja *Akusmonium* została przetestowana przez czwórkę kompozytorów – stworzyli on muzykę elektroniczną specjalnie dla skonstruowanego systemu głośnikowego. Podczas rezydencji, koncertów oraz testów instalacji pojawiło się kilka subiektywnych spostrzeżeń dotyczących tej formy odsłuchów nagrań:

1. Osiągnięto dużą przestrzenność nagrań dzięki dowolnemu położeniu w przestrzeni każdej ścieżki utworu.
2. Uzyskano dużą szczegółowość nagrania i możliwość realizacji niuansów dźwiękowych trudnych do wyeksponowania w tradycyjnym miksie stereo. Dzięki odtwarzaniu ścieżek w konkretnych głośnikach bądź ich grupach łatwiej było uniknąć maskowania poszczególnych pasm niż ma to miejsce w przypadku klasycznego miksu stereo.
3. Uzyskano mocno słyszalny ruch wirtualnych źródeł w obrębie instalacji umożliwiający osiągnięcie zaawansowanych efektów dźwiękowych.

10.4. System odsłuchowy w instalacji *Pokój do słuchania*

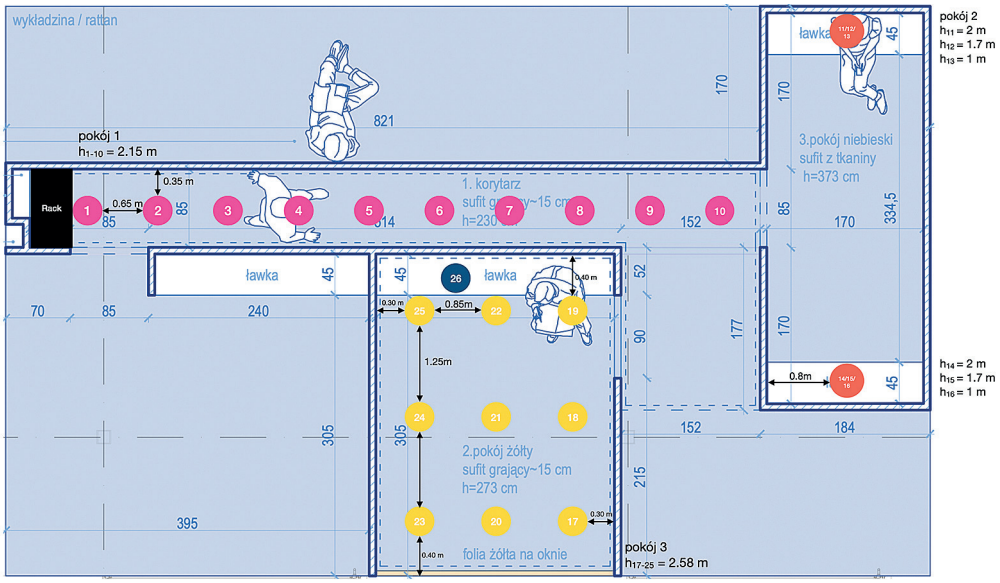
10.4.1. Idea i cel

Z inicjatywy kompozytorki Hani Rani oraz studia architektury Zmir w Warszawskim Pawilonie Architektury ZODIAK powstała instalacja artystyczna podejmująca tematykę relacji architektury i muzyki. W ramach przedsięwzięcia zaprojektowano specjalną drewnianą konstrukcję o powierzchni ponad 60 m^2 oraz skomponowano specjalną dla tej przestrzeni ścieżkę dźwiękową. Stworzona forma składała się z trzech oddzielnych pomieszczeń, w których wybrzmiewały różne warstwy kompozycji. Częścią stworzonej konstrukcji była instalacja dźwiękowa zaprojektowana konkretnie na potrzeby wystawy. Projekt systemu dźwiękowego obejmował trzy różne konstelacje głośnikowe dla poszczególnych pomieszczeń.

Założonym celem całego przedsięwzięcia było zaprojektowanie wielokanałowego systemu dystrybucji dźwięku przestrzennego wraz z zaprogramowaniem systemu sterującego opartego na dostępnej i otwartej bibliotece programistycznej Spat.5, a także przetestowanie jej w kontekście konstruowania nieregularnego systemu wielogłośnikowego. Na podstawie testów i prototypów założono wytworzenie autorskiego systemu pozwalającego na odtwarzanie oraz miks przestrzenny muzyki i efektów dźwiękowych z możliwością zastosowania w przyszłości w innych, podobnych realizacjach.

10.4.2. Instalacja wielogłośnikowa

Zaprojektowana instalacja dźwięku przestrzennego całościowo składa się z 25 monitorów bliskiego pola Genelec 8010A oraz jednego głośnika subbasowego Genelec 7350A. Zaprojektowany system jest zarządzany globalnie za pomocą aplikacji zaprojektowanej w środowisku programistycznym max/MSP. Każde z pomieszczeń podzielono na trzy oddzielne konstelacje, w których odtwarzany był różny materiał dźwiękowy. Na rysunku 6 zaprezentowano schemat przedstawiający rozmieszczenie głośników w poszczegól-



Rys. 10.6. Schemat przedstawiający pozycję głośników w instalacji

nych pomieszczeniach. Warto zwrócić uwagę na to, że znaczącym warunkiem definiującym układ głośników były założenia projektu architektonicznego powstałego przed projektem instalacji wielokanałowej.

W pokoju 1 głośniki (rys. 10.6, różowe koła oznaczone numerami 1–10) podwieszono pod sufitem w układzie liniowym. Wszystkie głośniki znajdowały się na wysokości 2,15 m i były odsunięte od ściany o 0,35 m. Głośniki oddalono od siebie o 0,65 m, a konstelację docelowo zasłonięto cienkim papierem. Wysokość wynikała z projektu architektonicznego, oddalenie głośników od ścian pozwoliło na uniknięcie dodatkowych rezonansów. Ze względu na małą szerokość pomieszczenia zdecydowano, aby głośniki umieścić regularnie na środku sufitu. W pokoju 2 ze względu na brak możliwości montażu głośników na suficie stworzono dwa pionowe systemy liniowe na dwóch przeciwległych ścianach (rys. 10.6, pomarańczowe koła oznaczone numerami 11–16). Układ głośników na obydwu ścianach jest analogiczny – 3 głośniki w jednej linii znajdowały się na trzech różnych wysokościach. Ustawienie głośników w centralnych punktach ścian pozwoliło na uniknięcie dodatkowych rezonansów. Wykorzystanie ich większej liczby w obrębie jednej ściany umożliwiło na dalszym etapie programowanie ruchu wirtualnych źródeł. Głośniki docelowo zasłonięto deskami o szerokości 7 cm, między którymi znajdowały się szczeliny o szerokościach 3–4 cm. Dylatacje te miały zapobiec nadmiernemu tłumieniu sygnałów. W pokoju 3 głośniki (rys. 10.6, żółte koła oznaczone numerami 17–25) podwieszono pod sufitem na wysokości 2,58 m i utworzono matrycę 3 × 3. Głośniki, podobnie jak w pokoju 1, zasłonięto cienkim papierem. Celem tego układu

Rys. 10.7. Ustawienie głośników w instalacji

było zrealizowanie regularnej matrycy umożliwiającej przestrzenny miks muzyki w obrębie dogłośnionej przestrzeni. Pod ławką w pokoju 3 znajdował się także głośnik niskotonowy (głośnik 26). Rzeczywiste ustawienie głośników w każdym pomieszczeniu przedstawiono na rys. 10.7. Analogicznie do przypadku instalacji *Akusmonium* sygnał do głośników dystrybuowany był za pośrednictwem protokołu Dante między komputerem sterującym a interfejsem (Focusrite Rednet 3) oraz ADAT między interfejsem a preampami (4 × Behringer ADA8200).

10.4.3. Projekt instalacji dźwiękowej w kontekście założeń programowych oraz konstrukcyjnych *Pokoju do słuchania*

Forma projektu wielokanałowej instalacji została podyktowana wieloma czynnikami. Podstawowym założeniem systemu było, aby dźwięk odtwarzany w konstrukcji stanowił integralną część formy architektonicznej, a słuchacz nie odczuwał obecności pojedynczych źródeł dźwięku. Kompozycja dźwiękowa była jednym, trwającym 1 h 15 min utworem, której poszczególne ścieżki zostały podzielone na oddzielne pomieszczenia. Istotne pozostawało, by poszczególne warstwy przydzielone do konkretnego pomieszczenia w nim wyeksponować i nadać specjalny, dźwiękowy charakter temu pomieszczeniu. Przy założeniu natomiast, że każde pomieszczenie jest oddzielną warstwą, a nie częścią kompozycji, oczekiwano, by dochodziło do kontrolowanych przesłuchów między pomieszczeniami. Tym sposobem słuchacz miał mieć poczucie odsłuchu utworu z trzech różnych perspektyw z zachowaniem kontekstu całościowej narracji.

Oprócz założeń merytorycznych i programowych równie istotne były aspekty konstrukcyjne. W pomieszczeniach nie dało się zastosować adaptacji akustycznej, a mię-

dzy pomieszczeniami nie występowały przegrody. To wpłynęło na liczbę oraz gęstość umieszczenia głośników. Zadecydowano, aby dogłośnić przestrzeń jak największą liczbą źródeł dźwięku, przez co pojedyncze źródło mogło emitować niższy poziom dźwięku. Dzięki temu częściowo odseparowano od siebie poszczególne warstwy muzyczne oraz zmniejszono wpływ akustyki pomieszczeń na odtwarzany materiał dźwiękowy. Ważne było również to, że cała instalacja elektroakustyczna powinna zostać ukryta. W pokoju 2 zasłonięcie głośników deskami ze szczelinami miało znikomy wpływ na pasmo przenoszenia głośnika. W pokoju 1 i 3 zasłonięcie głośników cienkim papierem zarówno wpłynęło na tłumienie sygnału w wysokim rejestrze, jak i obniżało poczucie ruchu animowanych ścieżek dźwiękowych. Duża liczba źródeł dźwięku w pokoju 1 i 3, jak również perforacje między kolejnym arkuszami papieru pozwoliły częściowo zniwelować ten problem. Dla projektu instalacji kluczowym był również charakter kompozycji dźwiękowej oraz jej podział między poszczególnymi pomieszczeniami (ten aspekt zostanie dalej rozwinięty w podrozdz. 10.4.5).

10.4.4. Aplikacja sterująca

Do obsługi instalacji dźwiękowej zaprojektowano aplikację w środowisku programistycznym max/MSP, opartą na bibliotece Spat.5 [1] zawierającej narzędzia do obsługi wielokanałowych systemów dźwiękowych. Aplikacja stanowiła kompleksowe narzędzie umożliwiające odtwarzanie, panoramowanie i miksowanie ścieżek zawartych w kompozycji oddzielnie dla każdego pomieszczenia. Najbardziej rozbudowanym elementem aplikacji był moduł odpowiadający za rozkład przestrzenny kompozycji. W ramach modułu przy wykorzystaniu biblioteki Spat.5 stało się możliwe zaimplementowanie fizycznej pozycji poszczególnych głośników, dobór algorytmu pozwalającego na rozkład źródeł w przestrzeni, nadanie właściwości akustycznych każdemu źródłu oddzielnie oraz programowanie torów ruchu.

10.4.5. Algorytm panoramujący

Instalacja dźwiękowa w *Pokoju do słuchania* nie zachowuje żadnej regularności ani nie tworzy zamkniętych obszarów. Ponadto założeniem wystawy było, aby uczestnicy mogli swobodnie poruszać się w przestrzeni. Takie kryteria uniemożliwiały zdefiniowanie stałej pozycji odsłuchowej. W systemie głośnikowym wykorzystano algorytm KNN (K-Nearest Neighbor) [10] zawarty w bibliotece Spat.5. To metoda oparta na DBAP Distance based amplitude panning) [7], w której do określenia adekwatnego wzmocnienia brany jest pod uwagę parametr odległości wirtualnego źródła od konkretnego głośnika, i na tej podstawie obliczane wzmocnienie w konkretnych kanałach. W tej metodzie nie zakłada się stałej pozycji słuchacza i nie generuje *sweet spot* jak w przypadku VBAB [1], WFS [1] czy ambisonii [4]. Pozwala ona na efektywne programowanie ruchu wirtualnego źródła

w instalacjach głośnikowych opartych na matrycach o nieregularnych ustawieniach. Zastosowanie metody KNN w odróżnieniu od DBAP umożliwia dodatkowo kontrolowanie następujących parametrów:

1. *Neighbors*, czyli maksymalnej liczby głośników jednocześnie odpowiadających za panoramowanie konkretnego wirtualnego źródła dźwięku w przestrzeni dogłośnienia [10].
2. *Spread*, czyli rozszerzenia parametru *Neighbors* określającego proporcje wzmocnienia w głośnikach odpowiadających za panoramowanie konkretnego źródła. W przypadku maksymalnej wartości parametru wszystkie głośniki mają wzmocnienie na tym samym poziomie bez względu na pozycję źródła. W przypadku minimalnej wartości tylko głośnik znajdujący się najbliżej wirtualnego źródła odpowiada za jego nagłośnienie [10].
3. *Max distance*, czyli maksymalnej odległości wirtualnego źródła od pozycji głośnika, w której jest ono nagłośnione [10].

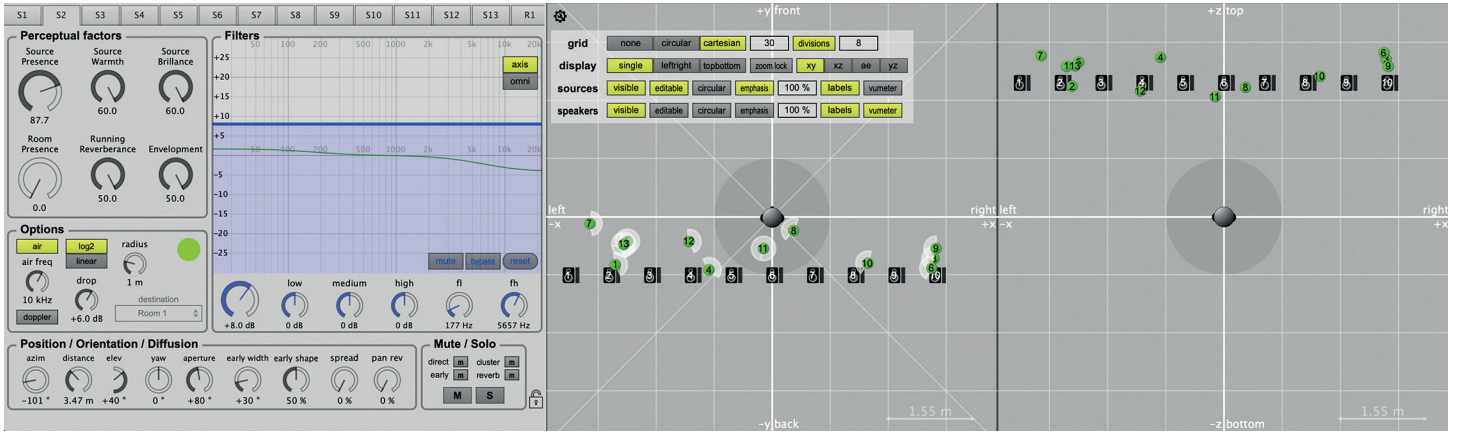
10.4.6. Mix, adaptacja, rozłożenie przestrzeni głośnikowej oraz animacja ścieżek dźwiękowych w pomieszczeniach

Miksowanie oraz rozłożenie ścieżek dźwiękowych w przestrzeni wielogłośnikowej zrealizowano na dwóch etapach. Na pierwszym z nich założono wykorzystanie instalacji prototypowej zbudowanej w Laboratorium Art & Science Poznańskiego Centrum Superkomputerowo-Sieciowego. Prototyp instalacji został przygotowany w jednym – dużym i otwartym pomieszczeniu, bez podziału poszczególnych konstelacji głośnikowych na osobne pomieszczenia. Poszczególne konstelacje zachowywały docelowe wymiary i pozycje ustawienia głośników. Na rysunku 10.8 przedstawiono zdjęcie tej prototypowej instalacji. Nie można było na tym etapie uwzględnić wszystkich aspektów, takich jak wpływ rezonansów pomieszczenia na odtwarzane ścieżki czy głośności przesłuchów warstw między poszczególnymi pomieszczeniami. Przygotowano również wstępne ustawienia poziomów głośności poszczególnych ścieżek, parametry korekcji pasmowej, pozycje w przestrzeni poszczególnych źródeł oraz ich tory ruchu.

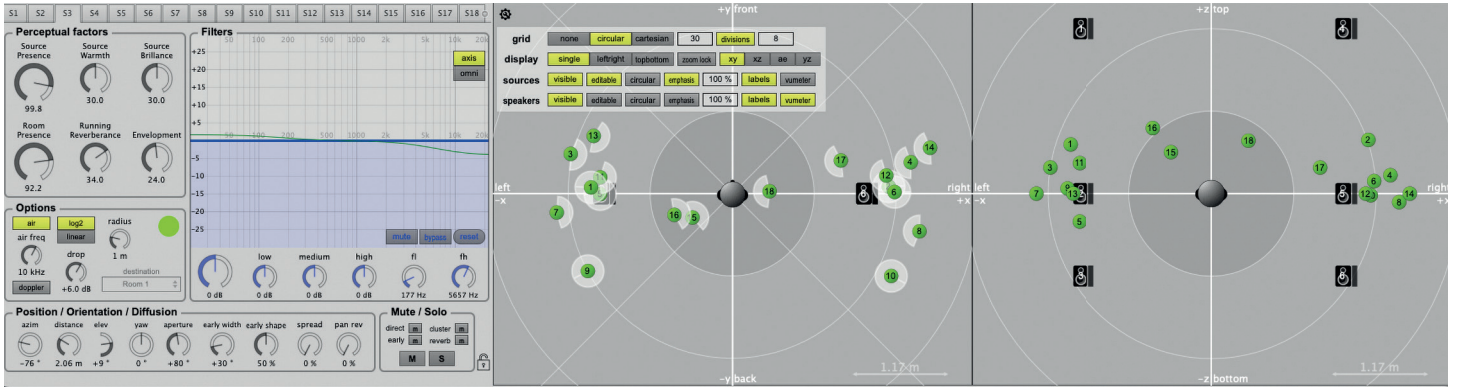
Drugi etap miksu warstwy dźwiękowej zrealizowano w docelowej lokalizacji. Na tym etapie ścieżka dźwiękowa została skorygowana o warunki akustyczne, uwzględniono także kwestie przesłuchu między pomieszczeniami. W kontekście kształtowania brzmienia i budowania sceny dźwiękowej w każdym z pomieszczeń stosowano odmienne kryteria. W pokoju 1 warstwa dźwiękowa składała się głównie z warstw oscylujących gatunkowo wokół muzyki ambient i nagrań terenowych. Warto zwrócić uwagę, że pomieszczenie to nie miało żadnych przystanków, w których odbiorcy byliby zachęceni do zatrzymania się i wsłuchania w muzykę. Ze względu na te kryteria w pomieszczeniu ustawiono parametr *Neighbors* = 10 (każde wirtualne źródło słyszalne w każdym z 10 głośników), parametrem dywersyfikującym obecność wirtualnego źródła w konkretnym punkcie pokoju był natomiast *Spread* ustawiony w przedziale od 65 do 90%. Przy założeniu, że uczestnicy będą się poruszać w po-

Rys. 10.8. Zdjęcie przedstawiające prototypową instalację dźwiękową:
na kratownicy znajduje się symulacja pokoju 1, stanowisko realizatora umieszczone w środku instalacji z pokoju 2, po prawej stronie widoczny jest fragment instalacji z ostatniego pokoju

mieszczeniu, ruch wirtualnych źródeł został realizowany głównie w osi *X* (w linii głośników) przy bardzo małej prędkości. Dodatkowy ruch źródeł w osi *Y* wykorzystano jako narzędzie dywersyfikujące amplitudę sygnałów, co nadało warstwie muzycznej naturalność oraz organiczność. Dzięki bibliotece Spat.5 jest możliwe indywidualne modelowanie właściwości akustycznych wirtualnego pomieszczenia oraz jego wpływu na brzmienie poszczególnych źródeł. Pokój 1 symulowano jako pomieszczenie o kubaturze 12 355 m^3. Aby zapewnić czytelność i klarowność dźwięków muzycznych wpływ sygnału imitującego właściwości pomieszczenia ustalono na niskim poziomie. W sygnale wyjściowym dominował sygnał wejściowy reprezentowany przez parametr *source presence* na poziomie ok. 80% względem sygnału zawierającego efekt imitacji wirtualnego pomieszczenia reprezentowanego przez parametr *room presence* na poziomie ok. 0–30%. Na rysunku 10.9 przedstawiono przykładowe parametry wpływu wirtualnego pomieszczenia na ścieżkę muzyczną. W przypadku nagrań terenowych parametr *source presence* był na podobnym poziomie, lecz zwiększono wartość parametru *room presence* w granicach 60%. Zabieg ten umożliwił znaczne oddalenie dźwięków terenowych i utworzył z nich tło dla dźwięków muzycznych. Było to szczególnie istotne w czasie nagrań terenowych deszczu, które generowały problemy z fazą podczas przechodzenia między liniami dogłośnienia poszczególnych głośników. Pod względem barwy we wszystkich ścieżkach za pomocą filtra górnoprzepustowego usunięto częstotliwości



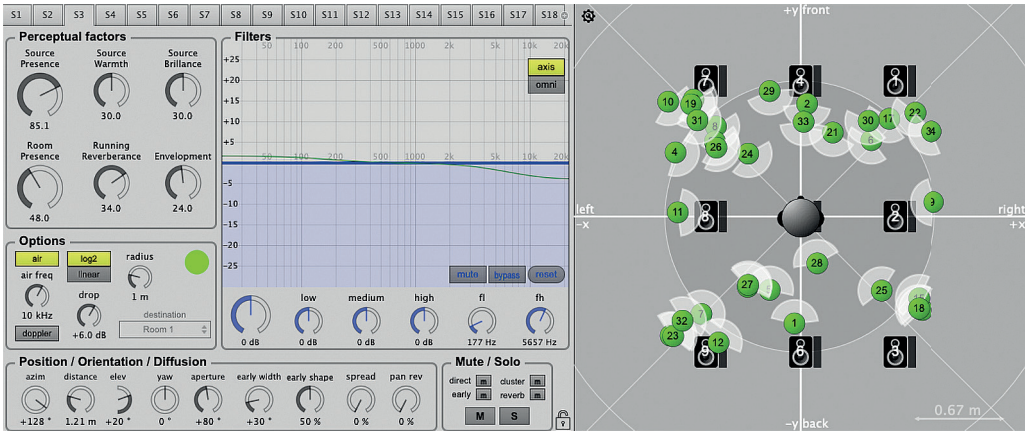
Rys. 10.9. Przykładowe parametry wirtualnego źródła w pokoju 1



Rys. 10.10. Przykładowe parametry wirtualnego źródła w pokoju 2

poniżej 50 Hz (uwzględnienie charakterystyki przeniesienia monitorów studyjnych Genelec 8010A). W pomieszczeniu zauważono rezonans w granicach 200 Hz – został również usunięty za pomocą filtra pasmowo-zaporowego.

W pokoju 2 warstwa muzyczna składała się głównie z punktowych dźwięków odegranych w technice *pizzicato* na skrzypcach i wiolonczeli oraz odegranych w dynamice *piano* pojedynczych nut *staccato* na pianinie. Dodatkową ścieżką dźwiękową w tej warstwie były efekty ambientowe, takie jak nagranie zegara oraz szeptów. Ze względu na układ konstelacji głośników i punktowy charakter poszczególnych warstw dźwiękowych głównym efektem brzmieniowym w tym pomieszczeniu było imitowanie wczesnych odbić na przeciwległej ścianie od pozycji wirtualnego źródła. Zabieg ten pozwolił na wytworzenie słuchowego wrażenia przebywania w znacznie większym pomieszczeniu niż pomieszczenie rzeczywiste przy jednoczesnym zachowaniu klarowności poszczególnych warstw dźwiękowych. Uzyskano to dzięki imitacji wirtualnego pomieszczenia o kubaturze 15 000 m^3 przy wysokich parametrach *source presence* oraz *room presence* – w granicach 90% i zmniejszonym parametrze *running reverbeance* odpowiedzialnym za obecność w sygnale wyjściowym późnych odbić i pogłosu. Parametr *neighbors* został tak ustawiony, aby wirtualne źródło było odtwarzane jednocześnie w trzech głośnikach przy parametrze *spread* wynoszącym 90%. Na rysunku 10.10 przedstawiono przykładowe parametry wirtualnego źródła. Przy rozpatrywanych ustawieniach wytworzono słuchowe wrażenie pojawienia się źródła na jednej ścianie, a efektów związanych z wybrzmieniem wirtualnego pomieszczenia na drugiej. Ruch poszczególnych wirtualnych źródeł zaprogramowany został głównie na osi *X* i *Z*. Ruch na osi *Y* został wykorzystany, aby zdywersyfikować w czasie głośność źródeł. Zmiany położenia w czasie realizowane były w taki sposób, aby słuchacz miał możliwość ich śledzenia. W pomieszczeniu zastosowano korekcję filtrem pasmowo-zaporowym w celu usunięcia rezonansów pomieszczenia.



Rys. 10.11. Przykładowe parametry wirtualnego źródła w pokoju 3

W pokoju 3 warstwa muzyczna składała się głównie z długich dźwięków odgrywanych na skrzypcach oraz pojedynczych akordów w partii pianina. W jedynej z warstw występowały tutaj kanały basowe i subbasowe. Dodatkowym tłem dla głównego motywu muzycznego były warstwy dźwiękowe z pozostałych pomieszczeń. Ze względu na dużą liczbę ścieżek dźwiękowych w tym pokoju tym jako jedynym skonstruowano instalację głośnikową umożliwiającą ruch wirtualnego źródła w dwóch płaszczyznach. Kluczowe dla tego pomieszczenia było uzyskanie efektu ciągłego i nieregularnego ruchu wirtualnych źródeł przy zachowaniu spójności obrazu dźwiękowego, dlatego parametry *neighbors* i *spread* zostały ustawione tak, aby pojedyncze wirtualne źródło było słyszalne we wszystkich dziewięciu głośnikach na niskim i równym poziomie jako tło – tylko w punktach aktualnego położenia poziom dźwięku wzrastał. W pokoju 3 jedna ściana była oknem, co wpłynęło na zwiększenie czasu pogłosu. Właściwości akustyczne pomieszczenia znacząco obniżały klarowność obrazu dźwiękowego oraz utrudniały lokalizację poszczególnych źródeł dźwięku. Aby zminimalizować ten efekt, zmniejszono wykorzystanie sztucznego pogłosu przez imitację wirtualnego pomieszczenia o kubaturze 2 000 m^3 przy małej obecności wczesnych oraz późnych odbić w sygnale wyjściowym. Parametr *room presence* dla wszystkich źródeł był na poziomie 80%, a parametr *room source* 40–50 %. Poziom wyjściowy warstw dźwiękowych z pozostałych pomieszczeń zmniejszono o ok. 8 dB względem głównych ścieżek. W pomieszczeniu zastosowano korekcję filtrem pasmowo-zaporowym w celu usunięcia rezonansów pomieszczenia. W kanale LFE odtworzono wszystkie warstwy muzyczne z zastosowanym filtrem dolnoprzepustowym o odcięciu 80 Hz, dla całej matrycy sufitowej natomiast zastosowano odwrotny filtr górnoprzepustowy. Na rysunku 10.11 przedstawiono przykładowe parametry wirtualnego źródła w pokoju 3.

10.4.7. Wnioski

W czasie prototypowania i ekspozycji systemu odsłuchowego w instalacji *Pokój do słuchania* wyciągnięto następujące wnioski:

1. Biblioteka Spat.5 stanowi rozbudowane narzędzie oferujące szereg funkcjonalności umożliwiających konstruowanie, prototypowanie oraz testowanie różnych rozwiązań dotyczących projektowania wielogłośnikowych systemów odsłuchowych, co ma zastosowanie w realizacji niekonwencjonalnych instalacji, w szczególności wykorzystywanych przez artystów.
2. Adekwatnym rozwiązaniem było wykorzystanie dużej liczby źródeł punktowych gęsto pokrywających przestrzeń odsłuchową. Dzięki temu stało się możliwe: zniwelowanie wpływu akustyki pomieszczenia na odtwarzane treści, kontrolowanie przesłuchów między poszczególnymi pomieszczeniami, osiągnięcie satysfakcjonującego efektu w kontekście słyszalności ruchu źródeł.
3. Adekwatne okazało się również wykorzystanie algorytmu DBAP do rozkładu wirtualnych źródeł dźwięku w przestrzeni wielogłośnikowej. Zastosowanie tej metody pozwoliło na wytworzenie przestrzeni odsłuchowej nieposiadającej jednego

punktu *sweet spot* przy jednoczesnym zachowaniu odczucia przestrzenności nagrania dla wielu odbiorców.

4. W pokoju 1 zauważono problemy związane z nakładaniem się faz sygnałów pochodzących z poszczególnych głośników. Efekt ten destruktywnie wpływał na odsłuch dźwięków w wyższych rejestrach Częściowo udało się to rozwiązać przez zastosowanie zwiększenia parametru *room presence*. Kolejnym krokiem w usunięciu zaistniałej niedoskonałości mogłoby być oddalenie głośników od pozycji odsłuchowej. W kontekście rozpatrywanej realizacji i z uwagi na projekt konstrukcji nie można było rozwiązać problemu całościowo.
5. W pokoju 1 i 3 ze względu na zasłonięcie głośników cienkim papierem częściowo utracono efekt ruchu wirtualnych źródeł. Problem częściowo rozwiązano przez pozostawienie niewielkich dylatacji między poszczególnymi kartonami papieru. Aby jeszcze bardziej zniwelować ten negatywny efekt, należy wykonać serię perforacji w papierze bądź wykorzystać inny materiał o ziarnistej strukturze.

10.5. Podsumowanie

Rozwiązania związane z dźwiękiem przestrzennym i immersyjnym w ostatnich latach stają się coraz popularniejsze. Przedstawione w rozdziale przykłady są prototypowymi przedsięwzięciami, w których wykorzystane zostały znaczne zasoby techniczne. Rozwiązania tego typu mogą być dobrą wskazówką dla instytucjonalnych placówek kultury w kontekście niekonwencjonalnych form adaptacji przestrzennych efektów dźwiękowych w realizacjach teatralnych, wystawowych czy koncertowych. W kolejnych pracach i realizacjach tego typu ograniczonych do wykorzystania słuchawek autorzy chcą zwrócić uwagę na bardziej dostępne rozwiązania dźwięku przestrzennego, takie jak dźwięk wolumetryczny czy rozwiązania typu 6-DOF [2], natomiast w pracach i realizacjach z jednoczesnym wykorzystaniem słuchawek i głośników – na narzędzia oparte na hybrydowych przestrzeniach odsłuchowych, takich jak Hybrid Audio Diffusion Systems [11]. Zastosowanie tego typu innowacji może przyczynić się do lepszej popularyzacji i upowszechnienia immersyjnych metod odsłuchu dźwięku przy ograniczeniu nakładu technicznego.

Na podstawie przedstawionych przykładów wypracowane rozwiązania w obrębie systemu odsłuchowego w instalacjach *Akusmonium* oraz *Pokój do słuchania* można traktować jako uniwersalne i przetestowane środowisko do prowadzenia zaawansowanych realizacji dźwiękowych w autorskich i nieregularnych konstelacjach wielogłośnikowych.

Obydwie instalacje opisane w niniejszym rozdziale zostały zaprojektowane w Laboratorium Art & Science oraz Laboratorium Działu Nowych Mediów Poznańskiego Centrum Superkomputerowo-Sieciowego.

Instalacja *Akusmonium* wyprodukowano w ramach festiwalu muzyki elektronicznej Fazma we współpracy ze stowarzyszeniem OKS. Zespół producencki tworzyli S. Dembski, M. Kaca, J. Skorupa

Projekt *Pokój do słuchania* to działanie autorstwa: Hani Rani, studia architektury Zmir (Ł. Palczyński, J. Szeliga), K. Janasa, I. Łysiuka.

Bibliografia

[1] Carpentier T., Noisternig M., Warusfel O., *Twenty years of Ircam Spat: looking back, looking forward*, 41st International Computer Music Conference, Denton 2015.

[2] Ciotucha T., Rumiński A., Żernicki T., Mróz B., *Evaluation of Six Degrees of Freedom 3D Audio Orchestra Recording and Playback using multi-point Ambisonic interpolation*, 2021; https://www.aes.org/e-lib/browse.cfm?elib=21052

[3] Fielder J., *A History of the Development of Multichannel Speaker Arrays for the Presentation and Diffusion Acousmatic Music*, Austin 2016; https://www.jonfielder.com/uploads/1/2/3/0/12308331/fielder_pennycookcomp_multichannelarrays.pdf

[4] Gerzon M., *What is wrong with quadraphonics*, 1974; https://www.michaelgerzonphotos.org.uk/ambi-sonics.html [dostęp: 28.10.2022].

[5] Głowiak M., Skorupa J., *Content production guidelines: Ambisonic recordings and postproduction*, Immersify Guidelines & Reports, 2020; https://immersify.eu/home/guidelines-reports/ambisonic-sound-production/

[6] Głowiak M., Skorupa J., *Produkcja dźwięku immersyjnego. Praktyczne metody i zastosowania dźwięku ambisonicznego wyższego rzędu do tworzenia produkcji audiowizualnych VR/360°*, w: *Postępy badań w inżynierii dźwięku i obrazu. Nowe trendy i zastosowania technologii dźwięku wielokanałowego oraz badania jakości dźwięku*, K.J. Opieliński (red.), Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2021.

[7] Lossius T., Baltazar P., Hogue T., *DBAP – Distance-Based Amplitude Panning*, Proceedings of the International Computer Music Conference, Montreal 2009.

[8] Williams S., *Osaka Expo '70: The promise and reality of a spherical sound stage*, w: *Insonic*, Karlsruhe 2015; https://oro.open.ac.uk/48743/1/inSonic2015_Williams_rev2.pdf

[9] https://cycling74.com [dostęp: 28.10.2022].

[10] https://doc.flux.audio/en_US/spat_revolution_doc/Spatialisation_Technology_Panning_Algorithms.html [dostęp: 28.10.2022].

[11] https://ars.electronica.art/planetb/en/perceiving-augmented-sound-field/[dostęp: 15.08.2023].

Słowa kluczowe: dźwięk przestrzenny, instalacje artystyczne, instalacje wielogłośnikowe.

Projektowanie i tworzenie systemów zarządzających w niekonwencjonalnych instalacjach dźwięku przestrzennego

Instalacje dźwięku przestrzennego od wielu lat stanowią przestrzeń współpracy między artystami zajmującymi się dźwiękiem oraz inżynierami. W przeszłości tego typu działania prowadziły do nowatorskich

rozwiązań w dziedzinie konstruowania nowych systemów dystrybucji dźwięku przestrzennego znacząco wykraczających poza ówczesne trendy. W rozdziale omówiono dwa autorskie projekty instalacji wielogłośnikowych przeznaczonych do prezentacji kompozycji przestrzennych: *Akusmonium* (nawiązujące do historycznego *Acousmonium*) oraz instalację głośnikową z *Pokoju do słuchania* – autorskiej wystawy kompozytorki Hani Rani oraz studia architektury Zmir. Na podstawie tych przykładów omówiona zostanie problematyka związana z projektowaniem i konstruowaniem systemów komputerowych umożliwiających zarządzanie sygnałem w instalacjach wielogłośnikowych.

Design and development of management systems in unconventional spatial audio installation

The publication discusses two multi-speaker installation dedicated to the presentation of spatial compositions *Akusmonium* referring to the historical *Acousmonium* and the loudspeaker installation from the *Room for Listening* by composer Hani Rani and the Zmir studio of architecture. On the basis of two examples, we discussed issues related to the construction as well as design of computer systems that allow signal management in multi-speaker installations.

11. Synteza dźwięku przestrzennego z wykorzystaniem zindywidualizowanych pomiarów HRTF

Zbigniew Świętach, Przemysław Plaskota

Politechnika Wrocławska,
Wydział Elektroniki, Fotoniki i Mikrosystemów,
wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław

11.1. Wprowadzenie

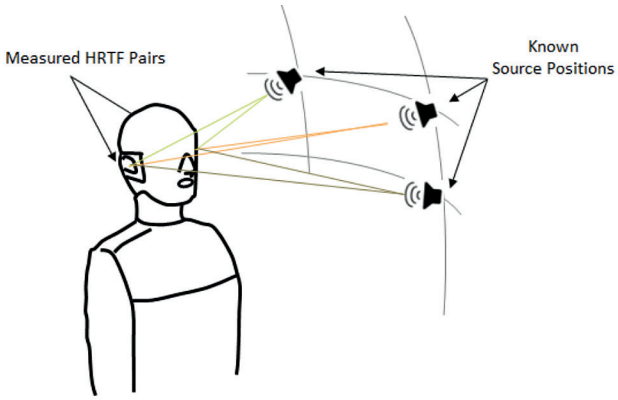
Funkcja transmitancji odniesiona do głowy (ang. *head-related transfer function* – HRTF) jest reprezentacją wpływu układu akustycznego tworzonego przez małżowinę uszną, głowę i tors ludzki na widmo sygnału akustycznego docierającego do ucha słuchacza [1]. Na zniekształcenie widma sygnału akustycznego wpływa zarówno kształt, jak i struktura materiałowa tkanek, z jakich zbudowana jest głowa [7], [9]. Dzięki zmianom w widmie słuchacz jest w stanie zlokalizować położenie źródła dźwięku w przestrzeni wokół niego. Ponieważ istnieje wiele położeń źródła dźwięku w przestrzeni otaczającej słuchacza, istnieje wiele HRTF o kształcie zależnym od położenia źródła dźwięku w tej przestrzeni. Z drugiej strony, jeśli znany jest przebieg HRTF przy różnych położeniach źródła dźwięku względem słuchacza, można stworzyć wrażenia przestrzeni dźwiękowej zdarzeń dźwiękowych niezawierających informacji o położeniu w przestrzeni.

Głowa każdego człowieka, a także małżowina uszna mają indywidualny kształt. Ogólnie są one zbliżone pod u różnych osób, ale w szczegółach różnice są dość istotne, zwłaszcza jeśli chodzi o przebieg funkcji HRTF [1]. Oznacza to, że każdy człowiek ma indywidualny zbiór HRTF. W sytuacji zatem, w której chcemy stworzyć u słuchacza wrażenie lokalizacji źródła dźwięku w określonym punkcie otaczającej go przestrzeni, konieczne jest zastosowanie indywidualnych funkcji HRTF, tzn. wyznaczonych (zmierzonych) dla konkretnej osoby [6].

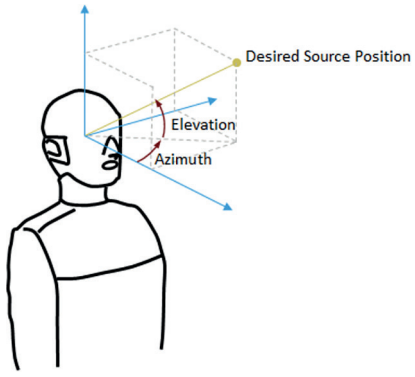
11.2. Podstawowe koncepcje

W czasie pomiarów HRTF źródła dźwięku testowego zostają umieszczone na sferze wokół słuchacza [1] znajdującego się w jej środku (rys. 11.1). Typowa przestrzenna siatka pomiarowa jest równoodległa, co oznacza, że azymut i elewacja zmieniają się w stałych odstępach odpowiednio w przedziałach (–180°,180°) oraz $(-\alpha,\alpha)$, gdzie $0° < \alpha < 90°$. Sygnałem testowym jest zazwyczaj sygnał typu chirp lub MLS.

W niniejszym rozdziale przyjęto tzw. geograficzny układ współrzędnych biegunowych, co oznacza, że azymut jest dodatni w I i II ćwiartce układu (prawoskrętny układ



Rys. 11.1. Pomiary HRTF (dokumentacja Matlaba [3])



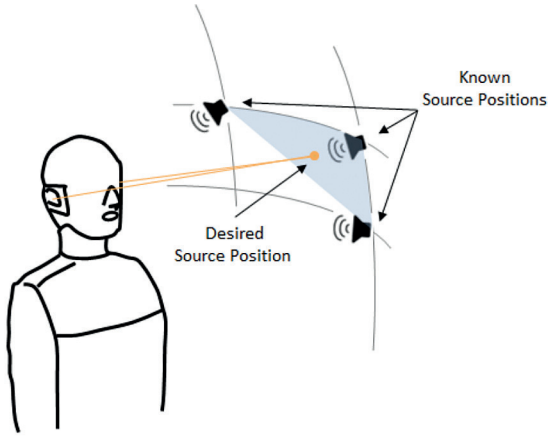
Rys. 11.2. Położenie źródła dźwięku w „geograficznym”
układzie współrzędnych biegunowych (dokumentacja Matlaba [3])

współrzędnych). Elewację mierzy się od płaszczyzny równika przechodzącej przez głowę słuchacza, gdzie oś *OZ* układu skierowana jest w górę (rys. 11.2). Omawiany układ współrzędnych często znajduje zastosowanie w pracach dotyczących tematyki HRTF.

Poza tym wykorzystano dwie bazy pomiarów HRTF: bazę ARI udostępnioną publicznie przez Instytut Badań Akustycznych z Austrii [5] i własną bazę pomiarów HRTF wykonanych w Katedrze Akustyki i Multimediów Politechniki Wrocławskiej, zwaną dalej bazą PWr [2]. Metodyka pomiarów HRTF zrealizowanych w celu utworzenia bazy PWr przedstawiona została w pracach [2], [10]. Praktyczne wykorzystanie wymienionych wcześniej baz HRTF do realizacji dźwięku przestrzennego poruszającego się obiektu wymaga zastosowania interpolacji ze względu na to, że siatki pomiarowe mają zbyt małą rozdzielczość przestrzenną. Omawianą sytuację zilustrowano na rys. 11.3.

Idea realizacji dźwięku przestrzennego na podstawie danego sygnału monofonicznego z wykorzystaniem pomiarów HRTF jest dobrze opisana w literaturze przedmiotu, np. [1]. HRTF jest reprezentowana przez skończony ciąg liczbowy, który jednocześnie reprezentuje próbki HRIR (ang. *head-related impulse response*). Dźwięk przestrzenny otrzymuje się przez splecenie próbek HRIR z próbkami danego sygnału monofonicznego. Jakość otrzymanego sygnału wynikowego rozumiana w potocznym znaczeniu, czyli to, czy źródło dźwięku przemieszcza się w sposób ciągły od punktu do punktu, czy można zlokalizować jego położenie przestrzenne i czy dźwięk nie jest zniekształcony, zależy jednak od wielu szczegółów.

W praktyce bardzo istotne są przyjęte założenia o sposobie wirtualnego przemieszczania się dźwięku, długości wybranych ramek czasowych, metodzie przechodzenia (przepływu danych) między kolejnymi ramkami czasowymi i zastosowanie odpowiednich algorytmów obliczeniowych.

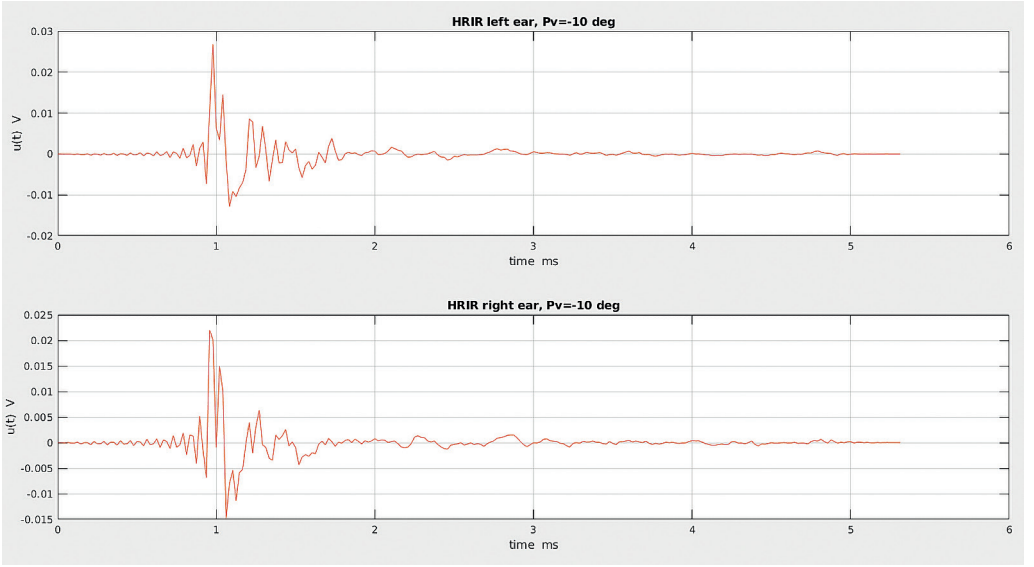


Rys. 11.3. W wymaganym położeniu niezbędna jest interpolacja, ponieważ pomiary HRTF nie są dostępne (dokumentacja Matlaba [2])

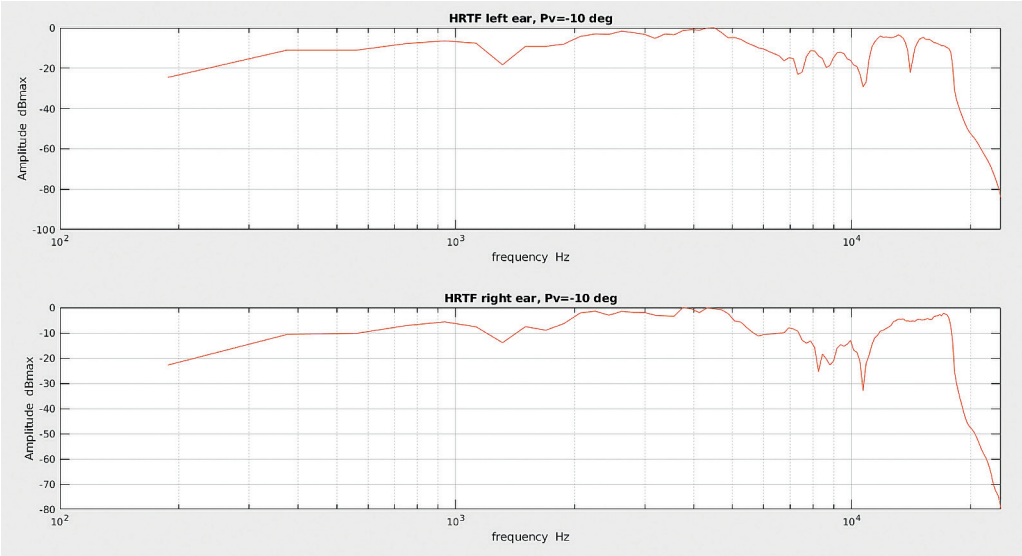
Za cel przyjęto napisanie autorskich procedur obliczeniowych realizujących dźwięk przestrzenny za pomocą HRTF. Procedury powinny być uniwersalne, z możliwością łatwej adaptacji do wymaganych zastosowań. Oznacza to na przykład możliwość zadania trajektorii poruszania się źródeł dźwięku czy zastosowania dowolnych filtrów dyskretnych FIR lub IIR. Wzmiankowane procedury (funkcje) zostały zrealizowane w środowisku obliczeniowym Matlab ze względu na prostotę zapisu formuł matematycznych i przejrzystość kodu programu, który jest zwykłym plikiem tekstowym. Ponadto sposób zapisu problemu matematycznego w Matlabie jest bardzo intuicyjny i niewiele odbiega od zapisu takiego problemu w standardowej notacji matematycznej. Matlab dysponuje także olbrzymią liczbą gotowych funkcji matematycznych, które można zastosować przy pisaniu własnych funkcji.

11.3. Analiza baz pomiarów HRTF: ARI i PWr

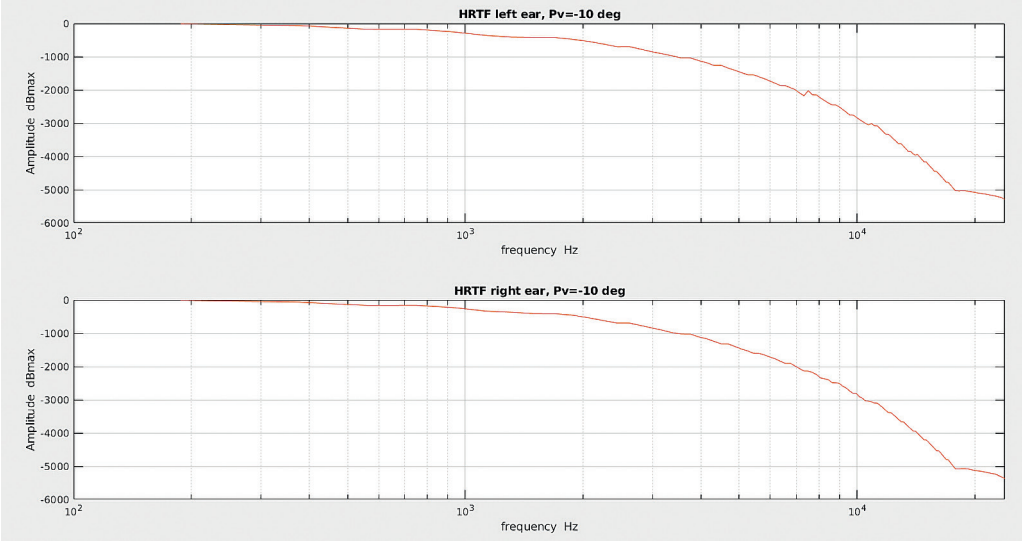
Bazę ARI udostępnił Instytut Badań Akustycznych z Austrii [5]. Rozdzielczość siatki azymutu wynosi 2,5°, a elewacji 5°. Baza jest dostępna w Matlabie w formie odpowiednich zmiennych środowiska Matlab. Szczegółowy sposób tworzenia HRTF dla tej bazy nie jest znany, na podstawie opisu zamieszczonego na stronach internetowych tego instytutu można jednak wnioskować o metodzie jej utworzenia. Wykonano pomiary



Rys. 11.4. Przykładowe odpowiedzi impulsowe HRIR z bazy ARI



Rys. 11.5. Charakterystyki amplitudowe przykładowych transmitancji HRTF z bazy ARI

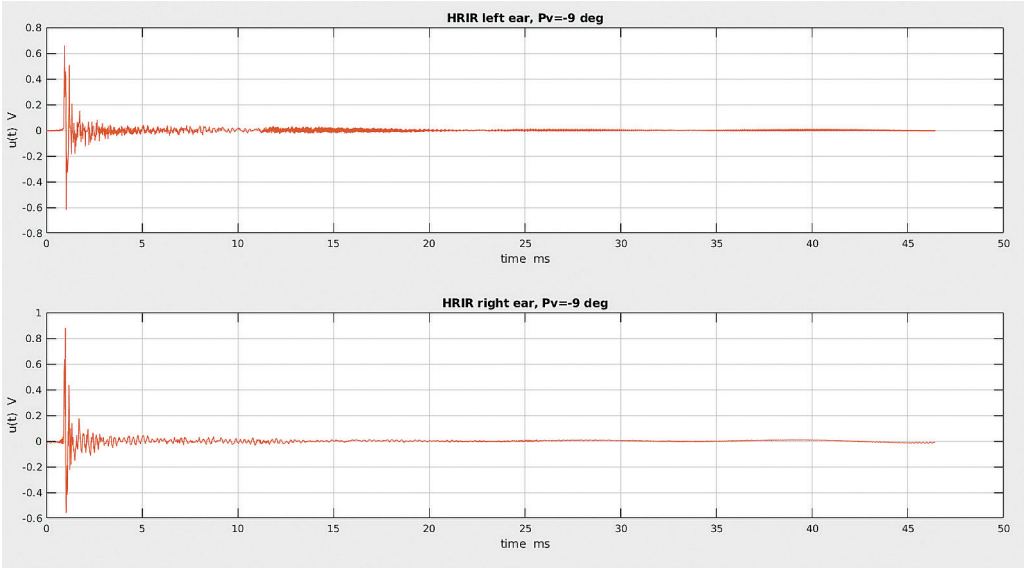


Rys. 11.6. Charakterystyki fazowe przykładowych transmitancji HRTF z bazy ARI

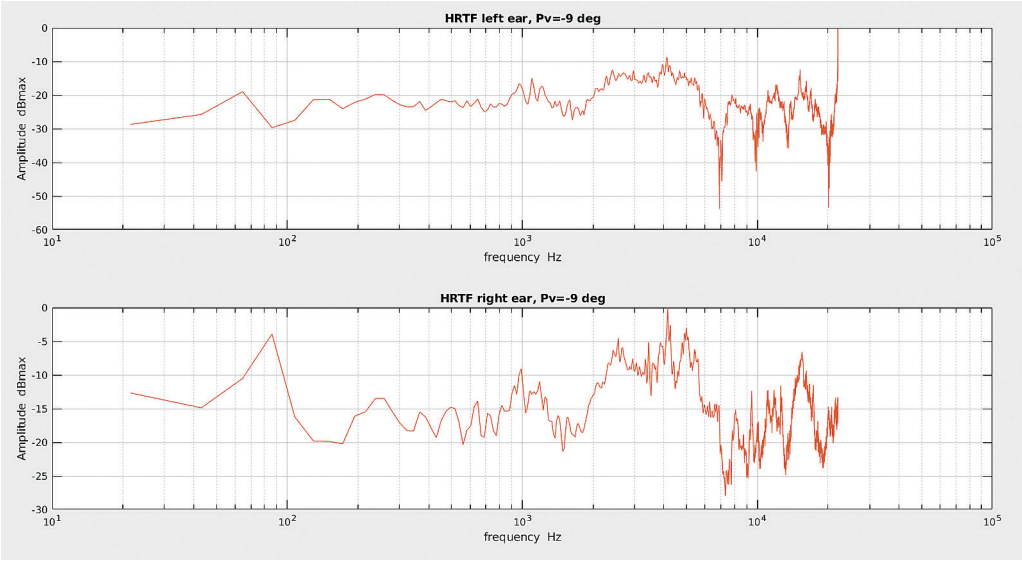
HRTF dla kilkuset osób. Wyniki pomiarów uśredniono. Następnie dla każdego punktu pomiarowego obliczono korelację wzajemną między pomiarami HRTF dla lewego i prawego ucha. W rezultacie wyznaczono średni czas opóźnienia propagacji fali akustycznej między lewym i prawym uchem $\tau_0 = n_0 / f_s$ jako czas, po którego upływie korelacja osiąga maksimum; gdzie n_0 – liczba opóźnionych próbek, f_s – częstotliwość próbkowania [8]. Uśrednione HRTF reprezentowane na płaszczyźnie z przez wielomiany argumentu z^{-1} skrócono do 256 współczynników. Potem zastąpiono wymienione wielomiany ich minimalnofazowymi odpowiednikami w iloczynie z jednomianem z^{-n_0}; gdzie n_0 – opóźnienie propagacji fali akustycznej wyrażone w wielokrotności okresu próbkowania. Opisane operacje prowadzą do zastąpienia oryginalnych uśrednionych HRTF ich skróconymi i zmodyfikowanymi równoważnikami, przy czym równoważność rozumiana jest tutaj nie jako matematyczny ekwiwalent, lecz odsłuchowy. Oznacza to, że według autorów bazy ARI ani dokładność lokalizacji źródła dźwięku, ani jakość dźwięku nie zmieni się w sposób zauważalny dla słuchacza, jeżeli zamiast zmierzonych HRTF zastosuje się zmodyfikowane HRTF. Skrócenie HRTF do 256 liczb znacząco zmniejsza czas obliczeń splotu dyskretnego.

Na rysunku 11.4 pokazano przykładowe odpowiedzi impulsowe HRIR dla azymutu 0° i elewacji –10°.

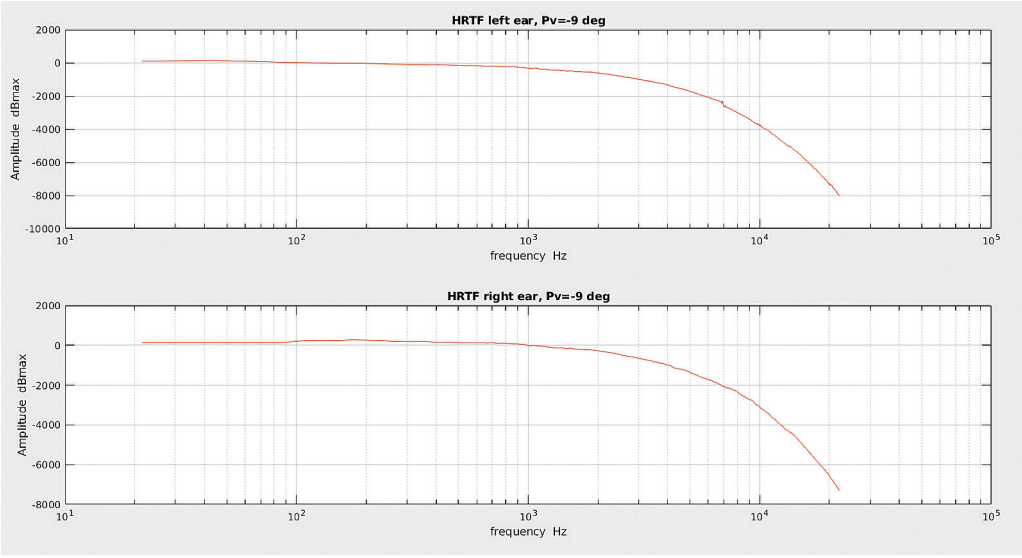
Na rysunkach 11.5 i 11.6 pokazano charakterystyki amplitudowe i fazowe przykładowych transmitancji HRTF dla azymutu 0° i elewacji –10°. Charakterystyki amplitudowe zostały znormalizowane względem wartości maksymalnej.



Rys. 11.7. Przykładowe odpowiedzi impulsowe HRIR z bazy PWr



Rys. 11.8. Charakterystyki amplitudowe przykładowych transmitancji HRTF z bazy PWr



Rys. 11.9. Charakterystyki fazowe przykładowych transmitancji HRTF z bazy PWr

Na podstawie analizy rys. 11.4–11.6 można sądzić, że baza ARI została starannie przygotowana. Obwiednia odpowiedzi impulsowych powoli narasta od zera do wartości maksymalnej i następnie powoli maleje, praktycznie do zera. Nie obserwuje się żadnych skoków wartości oraz nieoczekiwanych oscylacji, co mogłoby świadczyć o niepoprawnej akwizycji sygnałów pomiarowych. Charakterystyki amplitudowe przyjmują istotne obliczeniowo wartości w przedziale częstotliwości 0,2–16,0 kHz. Poza tym przedziałem wzmocnienie amplitudy maleje monotonicznie do wartości –80 dB_{max} przy częstotliwości Nyquista równej 24,0 kHz. Charakterystyki fazowe są monotonicznie malejącymi funkcjami częstotliwości, w zasadzie nie obserwuje się skoków wartości fazy. W przedziale o długości kilkuset herców położonym nieco powyżej częstotliwości 7,0 kHz można zaobserwować, że charakterystyka fazowa jest tam funkcją rosnącą.

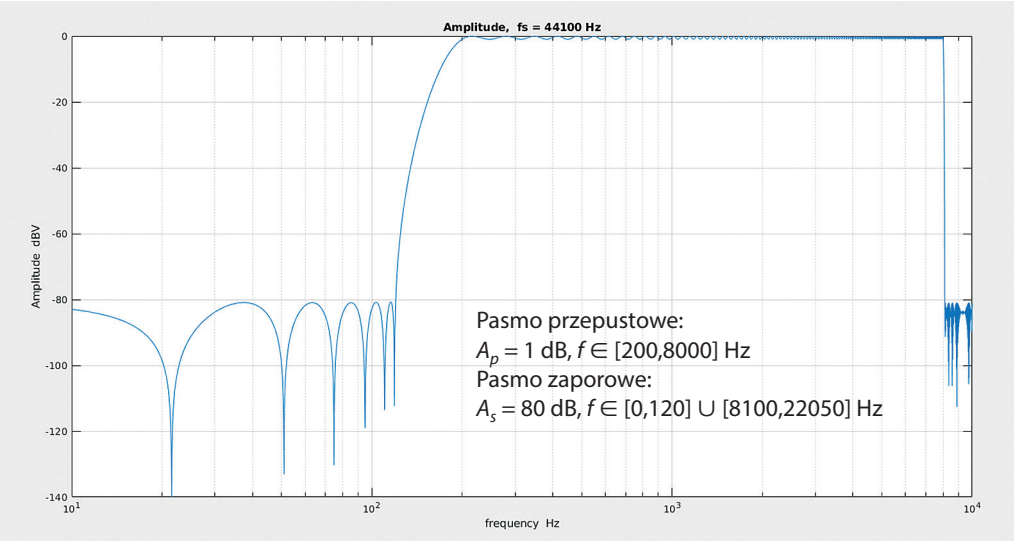
Omawiana baza pomiarów HRTF będzie bazą referencyjną służącą do testowania autorskich procedur obliczeniowych realizujących dźwięk przestrzenny za pomocą HRTF. Można będzie również porównywać dźwięk przestrzenny zrealizowany za pomocą bazy ARI oraz bazy PWr.

W przypadku bazy PWr rozdzielczość siatki azymutu wynosi 15°, a elewacji 9°. HRTF bazy PWr zbudowane są z 2048 współczynników. Baza zawiera HRIR otrzymane bezpośrednio na podstawie pomiarów. Wykonano wówczas pomiary HRTF, HRIR dla kilkudziesięciu osób. Wyników pomiarów nie uśredniano, co oznacza, że są to pomiary indywidualne. Można zatem zrealizować dźwięk przestrzenny na podstawie pomiarów HRTF indywidualnego słuchacza i następnie przetestować otrzymany dźwięk przy pomocy tego samego słuchacza (w rozpatrywanym przypadku słuchaczem był jeden z autorów tego rozdziału). Takie postępowanie eliminuje błąd metody, który nieuchronnie powstaje wówczas, gdy używa się uśrednionych HRTF. W kontekście dalszych rozważań należy nadmienić, że baza PWr od chwili jej utworzenia nie była nigdy wcześniej wykorzystywana do realizacji dźwięku przestrzennego. Z tego powodu konieczne było sprawdzenie, czy HRIR bazy PWr są wyznaczone prawidłowo, czyli z poprawną akwizycją sygnałów, tak jak zostało to omówione wcześniej dla bazy ARI. Na rysunku 11.7 pokazano przykładowe odpowiedzi impulsowe HRIR dla azymutu 0,0° i elewacji –9,0°.

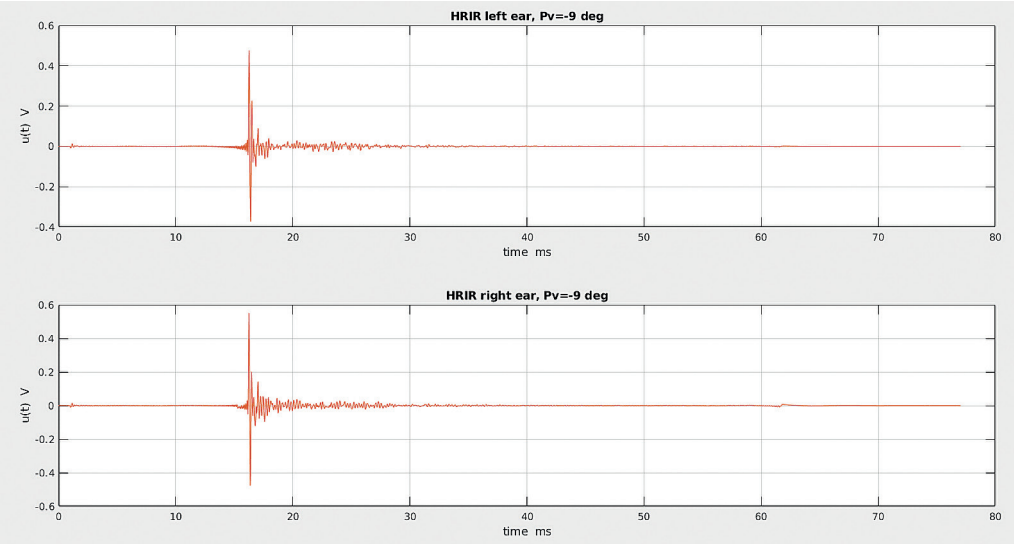
Obwiednia odpowiedzi impulsowych przedstawiona na rys. 11.7 bardzo szybko narasta od zera do wartości maksymalnej (praktycznie jest to skok wartości) i następnie szybko maleje. Dodatkowo na wykresach widać szybkozmienny sygnał zakłócający o charakterze szumowym. Pod koniec czasu pomiaru sygnał szumowy oscyluje na poziomie nieco poniżej 10 mV, HRIR nie maleje zatem do zera w przedziale pomiarowym [0, 46,5] ms. Na wykresie dla kanału prawego widać również wolnozmienny sygnał zakłócający o okresie $T = 20$ ms pochodzący z sieci energetycznej.

Na rysunkach 11.8 i 11.9 zamieszczono charakterystyki amplitudowe i fazowe przykładowych transmitancji HRTF dla azymutu 0,0° i elewacji –9,0°. Charakterystyki amplitudowe zostały znormalizowane względem wartości maksymalnej.

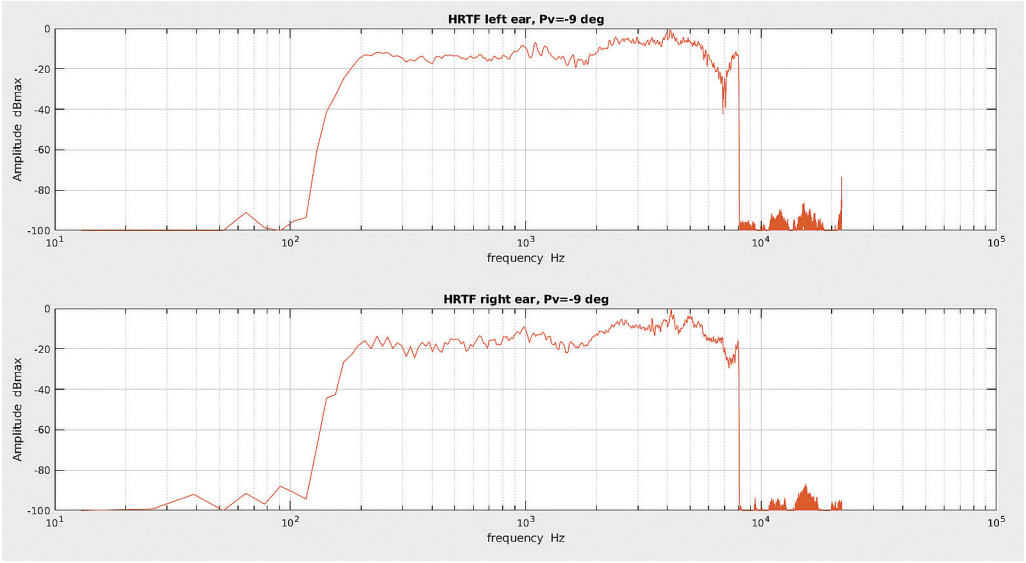
W założeniu widma amplitudowe powinny przyjmować istotne obliczeniowo wartości w przedziale częstotliwości 0,2–8,0 kHz. Poza tym przedziałem wzmocnienie ampli-



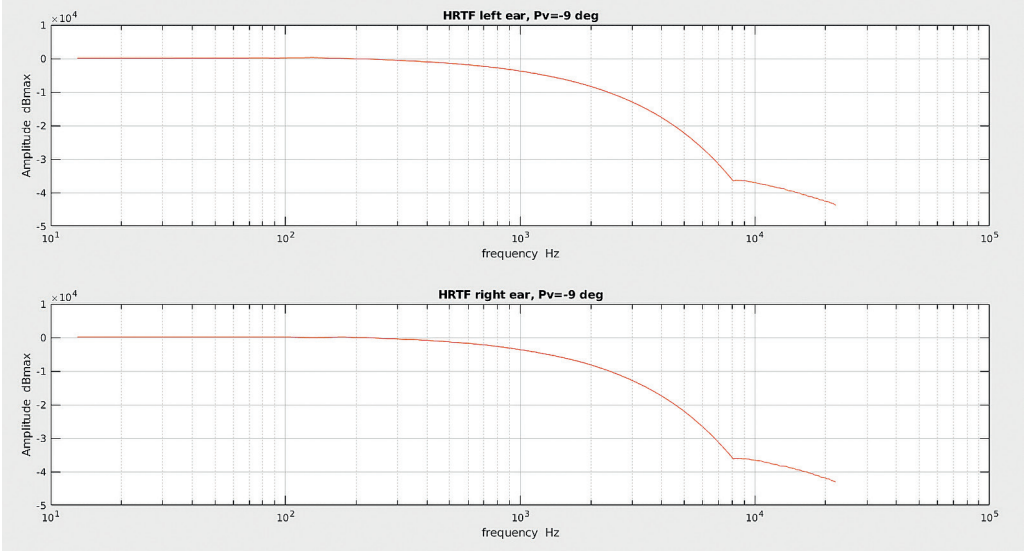
Rys. 11.10. Charakterystyka amplitudowa zaprojektowanego filtru FIR



Rys. 11.11. Przykładowe odpowiedzi impulsowe HRIR z bazy PWr
po przeprowadzonej filtracji



Rys. 11.12. Charakterystyki amplitudowe przykładowych transmitancji HRTF z bazy PWr po filtracji



Rys. 11.13. Charakterystyki fazowe przykładowych transmitancji HRTF z bazy PWr po filtracji

tudy powinno maleć monotonicznie do pomijalnie małych wartości, np. do –80 dB_{max} przy częstotliwości Nyquista równej 22,05 kHz.

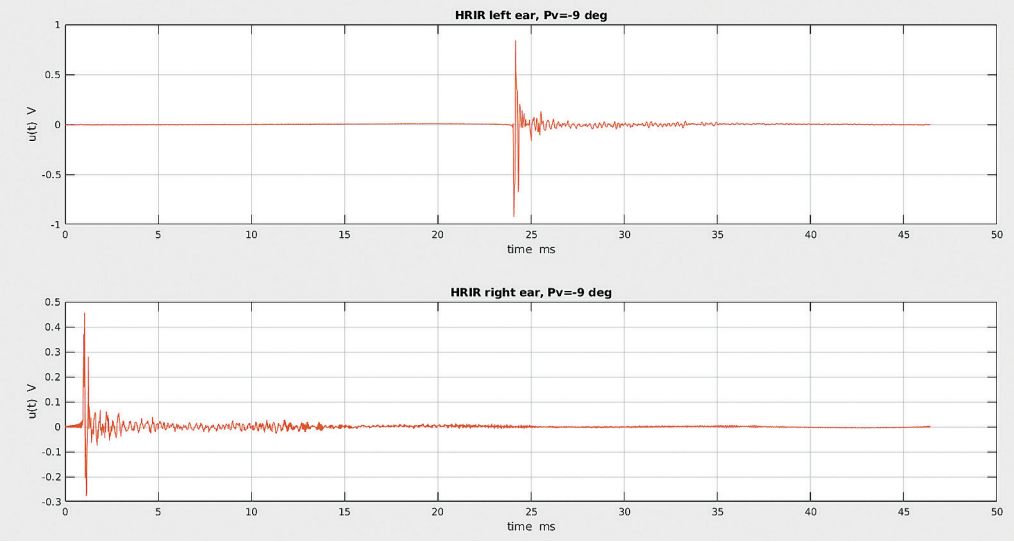
Charakterystyki fazowe są monotonicznie malejącymi funkcjami częstotliwości, nie obserwuje się skoków wartości fazy – z wyjątkiem niewielkiego otoczenia częstotliwości 7,0 kHz.

Na podstawie analizy wykresów charakterystyk amplitudowych można przypuszczać, że nie zostało dostatecznie dobrze ograniczone pasmo sygnału analogowego przed przetwornikiem ADC (brak filtru lub niepoprawnie działający filtr antyaliasingowy). Jest prawdopodobne również, że wystąpiły pewne błędy czy nieścisłości w przetwarzaniu sygnału spróbkowanego lub w samej metodyce pomiarów, lub i tu, i tu.

Jedyne co można było zrobić, to odfiltrować rozpatrywane HRIR za pomocą odpowiednio dobranego filtru dyskretnego. Z wykorzystaniem II algorytmu Remeza został zaprojektowany filtr FIR o liniowej fazie, zbudowany z 1351 współczynników. Charakterystykę amplitudową filtru FIR zamieszczono na rys. 11.10.

Na rysunku 11.11 pokazano odpowiedzi impulsowe HRIR dla azymutu 0,0° i elewacji –9,0° po przeprowadzeniu filtracji za pomocą zaprojektowanego filtru FIR. Zakłócenia sieci energetycznej i zakłócenia szybkozmienne zostały w znacznym stopniu wyeliminowane.

Na rysunkach 11.12 i 11.13 zamieszczono natomiast charakterystyki amplitudowe i fazowe transmitancji HRTF dla azymutu 0,0° i elewacji –9,0° po przeprowadzeniu filtracji. Charakterystyki amplitudowe zostały znormalizowane względem wartości maksymalnej. Charakterystyki amplitudowe przyjmują istotne obliczeniowo wartości



Rys. 11.14. Odpowiedzi impulsowe HRIR z bazy PWr; azymut 45°, elewacja –9°

w przedziale częstotliwości 0,2–8,0 kHz. Poza tym przedziałem wzmocnienie amplitudy maleje monotonicznie do wartości poniżej –80 dB_{max}.

Charakterystyki fazowe są monotonicznie malejącymi funkcjami częstotliwości, niewielkie fluktuacje fazy, które można zauważyć na rys. 11.9, zostały teraz wyeliminowane.

W dalszych testach bazy HRTF PWr wykazano, że dla azymutu z przedziału (0°, 180°) i każdej elewacji został popełniony błąd przy wyznaczaniu HRIR. Zamieszczone na rys. 11.14 odpowiedzi impulsowe HRIR lewego kanału są raczej nieprzydatne do celów syntezy dźwięku przestrzennego. Nie można aktualnie ustalić, na którym etapie akwizycji sygnałów pomiarowych powstał błąd.

Propozycja przesunięcia HRIR w lewym kanale o pewną liczbę próbek nie prowadzi do rozwiązania problemu. Nie wiadomo, o ile próbek przesunąć HRIR, a to kluczowe dla poprawnej lokalizacji źródła dźwięku, ponieważ względne przesunięcie HRIR kanału lewego i prawego jest skutkiem opóźnienia propagacji fali akustycznej między lewym i prawym uchem słuchacza. Co więcej, po przeprowadzeniu takiego sztucznego przesunięcia HRIR, nie byłoby takie oczywiste, czym należy uzupełnić brakujące próbki przesuniętej odpowiedzi impulsowej.

W rezultacie przeprowadzonych testów bazy PWr stwierdzono, że porównania syntezowanego dźwięku przestrzennego można wykonywać jedynie dla azymutu zawartego w przedziale [–180°, 0°].

11.4. Przegląd gotowych funkcji Matlaba dotyczących sygnałów audio

Przed przystąpieniem do tworzenia własnych procedur syntezy dźwięku przestrzennego został wykonany przegląd gotowych rozwiązań dostępnych w Matlabie. W podstawowej licencji Matlaba otrzymuje się następujące funkcje służące do operacji na plikach dźwiękowych:

- *audioinfo* – informacja o strukturze danego pliku dźwiękowego. Funkcja obsługuje wiele formatów audio oraz dźwięk w formatach wideo typu mp4, m4v, avi.
- *audioread* – odczyt danego pliku dźwiękowego i zmiana formatu na zmienną Matlaba. Funkcja obsługuje wiele formatów audio.
- *audiowrite* – zapis danego pliku dźwiękowego ze zmiennej Matlaba na wybrany format audio. Funkcja obsługuje wiele formatów audio.
- *audiodevinfo* – informacja o urządzeniach audio dostępnych w środowisku Matlab. Zazwyczaj tymi urządzeniami są karty dźwiękowe lub instrumenty (format danych midi).
- *audioplayer* – odtwarzanie dźwięku zapisanego w formie zmiennej Matlaba za pomocą karty dźwiękowej lub innego urządzenia wyjściowego, np. przetwornika DAC. Jest to systemowa funkcja obiektowa Matlaba.

- *audiorecorder* – nagrywanie dźwięku z urządzenia wejściowego (np. mikrofonu) w formacie zmiennej Matlaba. Jest to systemowa funkcja obiektowa Matlaba o niemodyfikowalnych metodach.
- *sound* – odtwarzanie dźwięku zapisanego w formie zmiennej Matlaba za pomocą karty dźwiękowej lub innego urządzenia wyjściowego, np. przetwornika DAC. Funkcja ta korzysta z funkcji obiektowej audioplayer.
- *soundsc* – działa jak sound, przy czym dodatkowo przeprowadzane jest automatyczne skalowanie poziomów wzmocnień

Z wymienionych funkcji jako niezbędne minimum wykorzystano audioread, audiowrite oraz sound. Po zakupie licencji do Toolboxów Matlaba Audio System Toolbox i DSP System Toolbox możliwości operacji na plikach dźwiękowych zostają rozszerzone o kolejne funkcje Matlaba, czyli:

- *audioPlayerRecorder* –nagrywanie dźwięku z urządzenia wejściowego (np. mikrofonu) w formacie zmiennej Matlaba i jednoczesne odtwarzanie dźwięku.
- *audioDeviceReader* – nagrywanie dźwięku np. z mikrofonu w formacie zmiennej Matlaba.
- *dsp.AudioFileReader* – odczyt danych z pliku audio.
- *audioDeviceWriter* – odtwarzanie dźwięku (zmiennej Matlaba) na kartę dźwiękową.
- *dsp.AudioFileWriter* – zapis danych z formatu zmiennej Matlaba na wybrany format audio.
- *dsp.FIRfilter* – obiekt realizujący dyskretną filtrację wielomianową o liniowej fazie.
- *dsp.SignalSink* – obiekt buforujący i zarządzający danymi filtrowanymi przez dsp. FIRfilter.

Dodatkowo w ramach Toolboxu Audio System Toolbox udostępniana jest w Matlabie baza HRTF ARI. Przytoczone wcześniej funkcje to wyłącznie systemowe funkcje obiektowe, w których kody źródłowe zastosowanych w nich metod są ukryte przed użytkownikami, co w połączeniu z dostępną, a w szczegółach niepełną dokumentacją, nie pozwoliło na swobodne operowanie wymienionymi funkcjami.

11.5. Autorskie procedury wyznaczania dźwięku przestrzennego

W kontekście wykorzystania HRTF dostępne są w Audio Systemie Toolbox jedynie dwa proste przykłady, które nie mogą zostać użyte (zaadaptowane) do praktycznego tworzenia dźwięku przestrzennego. Co więcej, ich analiza połączona ze szczegółowym zrozumieniem działania wymagała od autorów niniejszego rozdziału napisania funkcji testowych, w których konieczne było m.in. wykonanie od podstaw sposobu indeksowania próbek, indeksowania ramek czasowych, zapętlanie ramek, przekazywanie warunków początkowych dla filtru FIR po przejściu do kolejnej ramki. W rezultacie zbudowano całą strukturę danych i sterowania niezbędną do praktycznego przeprowadzenia filtracji

i otrzymania dźwięku przestrzennego. Tym samym wykorzystanie gotowych obiektowych systemowych funkcji Matalaba stało się zbędne, bo zbyt zawiłe i nie do końca przewidywalne. W konsekwencji autorzy w swoich programach wykorzystują tylko funkcje Matlaba audioread, audiowrite, sound, filter i fft.

Problemem, który w naturalny sposób pojawił się przy tworzeniu dźwięku przestrzennego, jest interpolacja HRTF dla położeń przestrzennych, gdzie nie zostały wykonane pomiary fizyczne. W Audio Systemie Toolbox dostępna jest funkcja interpolacyjna interpolateHRTF – zaimplementowano w niej dwa algorytmy: bilinear interpolation i VBAP (ang. *vector base amplitude panning*). Pierwszy algorytm (opisany w publikacji [4]) wykorzystuje prostą interpolację dwuliniową, algorytm działa numerycznie poprawnie, ale jakość interpolacji jest niezadowalająca. Drugi algorytm wymaga wyznaczania macierzy odwrotnych o wymiarach 3×3, co w przypadkach praktycznych prowadzi często do źle uwarunkowanych obliczeń i wyznaczania macierzy, która jest prawie osobliwa lub numeryczne jest macierzą osobliwą [11]. W testach przeprowadzonych przez autorów rozpatrywany algorytm często prowadził do przedwczesnego zakończenia obliczeń na przykład z informacją o dzieleniu przez zero.

Autorzy podjęli się także przeglądu zagadnień związanych z reprodukcją dźwięku przestrzennego, w tym również dotyczących interpolacji przestrzennej [12]. Artykuł temu poświęcony, opublikowany niedawno, obecnie jest w trakcie analizy – przypuszczalnie jej rezultaty zostaną wykorzystane w dalszych pracach. Temat interpolacji przestrzennej pozostaje bowiem otwarty i w chwili obecnej nie został jeszcze zaimplementowany w autorskich programach syntezy dźwięku przestrzennego.

We wspomnianych programach autorskich przyjęto, że każde źródło dźwięku lub grupa źródeł dźwięku jest obiektem reprezentowanym przez nagranie monofoniczne. Takie obiekty przemieszczają się w dowolny sposób po sferze wokół słuchacza. Ze względu na brak interpolacji aktualnie przemieszczanie następuje skokowo po węzłach siatki pomiarowej. To ograniczenie powinno zostać usunięte w kolejnych pracach.

Ruch obiektu zdefiniowany jest w macierzy **D** o czterech kolumnach i o liczbie wierszy odpowiadającej liczbie punktów przestrzennych, w których zlokalizowano źródło dźwięku. Pierwsza i druga kolumna zawierają odpowiednio informacje o azymucie i elewacji. W trzeciej podana natomiast została odległość obiektu od słuchacza, a w czwartej – chwilowa prędkość obiektu. Obecnie nie wykorzystuje się informacji o odległości w celu zapewnienia odpowiedniego tłumienia sygnału, co oznacza na przykład, że obiekt oddalający się od słuchacza po pewnej trajektorii powinien być słyszalny coraz ciszej. W najprostszej koncepcji można założyć wykładnicze tłumienie sygnału wraz z odległością, dobrać/przyjąć odpowiednie współczynniki tłumienia i przeprowadzić syntezę dźwięku przestrzennego. Takie postępowanie nie uwzględnia jednak zależności tłumienia sygnału od częstotliwości, co może mieć znaczenie w sytuacjach praktycznych – zagadnienie pozostaje na dziś otwarte.

Dzięki programowi wyznacza się przemieszczenie między dwoma sąsiednimi punktami sfery *A* i *B*, zgodnie z danymi zawartymi w macierzy **D**. Następnie – średnią prędkość obiektu na tym odcinku i średni czas przemieszczania się obiektu między punktami.

Liczba próbek (długość) ramki czasowej w punkcie *A* jest obliczana przez zaokrąglenie iloczynu czasu przemieszczania i częstotliwości próbkowania. Ramka czasowa zostaje splatana z HRTF odpowiadającym punktowi *A* odpowiednio dla lewego i prawego ucha. Potem cała procedura powtarza się dla punktów *B* i *C*. W ostatnim punkcie przyjmuje się, że długość ramki czasowej jest taka jak dla przedostatniego punktu. Opisywany algorytm umożliwił definiowanie ramek czasowych o zmiennej długości odpowiadającej średniemu czasowi przemieszczania się źródła dźwięku między kolejnymi punktami przestrzennymi.

Jeżeli liczba próbek dźwięku w sygnale monofonicznym jest mniejsza niż całkowita liczba próbek w ramkach czasowych, to program w odpowiedni sposób „zapętla" oryginalny sygnał tak, że efektywna liczba próbek jest nieograniczona.

Do splatania ramek sygnału z odpowiednim HRTF wykorzystano funkcję Matlaba filter. Jest to funkcja wbudowana (tzn. funkcja skompilowana przez producenta środowiska Matlab, plik binarny). Na podstawie szybkości działania wykorzystuje ona algorytmy FFT, a splot dyskretny obliczany jest pośrednio przez operacje na widmach sygnałów. Zamiast tej funkcji można użyć wbudowanej funkcji fft i napisać odpowiedni program w Matlabie. Daje to takie same rezultaty, użycie funkcji filter jest jednak wygodniejsze.

W celu zapewnienia wrażenia ciągłości ruchu obiektu tylko dla pierwszej ramki danych obliczenia splotu rozpoczynają się dla zerowych warunków początkowych. W przypadku kolejnych ramek danych obliczenia splotu z nowymi HRTF rozpoczynają się z wykorzystaniem warunków początkowych pochodzących ze splotu wyznaczonego w poprzedniej ramce. W przeciwnym razie po przeprowadzeniu syntezy w wynikowym dźwięku słyszalne byłyby nienaturalne efekty: na początku każdej ramki czasowej dźwięk powoli mógłby narastać od zera do pewnego średniego poziomu, a następnie być gwałtownie ucięty na końcu ramki.

Omawiane funkcje Matlaba napisane przez autorów umożliwiają ponadto przeprowadzenie dowolnej filtracji przetwarzanych sygnałów. Jeżeli zajdzie taka konieczność, można filtrować oryginalny sygnał monofoniczny indywidualnie dla każdej ramki, można filtrować dane HRTF oraz sygnał wynikowy otrzymany po splocie sygnału monofonicznego z danym HRTF. Nie ma żadnych ograniczeń co do wymaganych filtrów – mogą być to filtry FIR o liniowej fazie lub nieliniowej fazie oraz filtry wymierne IIR o dowolnej wymaganej aproksymacji charakterystyk widmowych, np. filtry Czebyszewa lub Cauera.

Co więcej, omawiane programy umożliwiają również odfiltrowanie z oryginału monofonicznego dźwięku o efektywnym pasmie, np. 0,2–8,0 kHz, w którym człowiek jest w stanie rozpoznać lokalizację źródła od reszty sygnału dźwiękowego. Po przeprowadzeniu syntezy dźwięku przestrzennego można złożyć wynikowy dźwięk z tą „resztą" sygnału oryginalnego. W efekcie nie są tracone dolne i górne części pasma akustycznego oryginalnego sygnału monofonicznego.

Opisane w rozdziale autorskie programy syntezy dźwięku przestrzennego nie są programami typu: pulpit mikserski audio. Nie mają one interfejsu graficznego, dane wejściowe dla każdego obiektu przygotowuje się w formie odpowiedniej macierzy. Jednocześnie syntezowany jest dźwięk przestrzenny dla jednego obiektu. Oczywiście tak otrzymane

dźwięki przestrzenne można złożyć w jeden plik dźwiękowy według zadanych reguł filtracji (nie było to jednak celem przeprowadzonych testów). Można natomiast, jeśli byłaby taka potrzeba, utworzyć na podstawie omawianych programów jeden duży obiekt, który za pomocą dostępnego w środowisku Matlab oprogramowania konwertuje się do formatu VST audio plugin rozpoznawanego przez DAW (Digital Audio Workstation). W tym sensie przedstawione programy mogą rozszerzyć możliwości realizacji dźwięku kompatybilnych z VST audio plugin DAW.

11.6. Przykłady realizacji dźwięku przestrzennego

Za pomocą omówionych wcześniej programów otrzymano 16 przykładowych realizacji dźwięku przestrzennego, w tym 8 realizacji statycznego źródła dźwięku i 8 realizacji przemieszczającego się źródła dźwięku. Ze wszystkich przykładowych realizacji dźwięku przestrzennego 8 zostało otrzymanych na podstawie bazy HRTF ARI, a pozostałe 8 na podstawie bazy HRTF PWr.

Statyczne źródła dźwięku rozmieszczone są co 45° na półokręgu, tzn. dla azymutu równego 0°, –45°, –90°, -135°, –180°. W każdym punkcie źródło emituje dźwięk przez ok. 2 s. Syntezę dźwięku wykonano dla elewacji równej 0° i 45°.

Ruchome źródła dźwięku przemieszczają się w przedziale azymutu od 0° do –180° z korkiem co –15°. Całkowity czas przemieszczania się źródła dźwięku wynosi 3–12 s. Tak jak poprzednio syntezę dźwięku wykonano dla elewacji równej 0° i 45°.

Jako pliki źródłowe wykorzystano cztery monofoniczne sygnały:

- dźwięk lecącego helikoptera, czas nagrania ok. 5,9 s;
- brzęczenie lecącej muchy, czas nagrania ok. 1,3 s;
- kroki osoby poruszającej się w pomieszczeniu z pogłosem (np. pusta hala, kościół), czas nagrania ok. 12,2 s;
- żeńskim głosem wypowiadana fraza „pada deszcz", czas nagrania ok. 1,1 s.

11.7. Podsumowanie

Na podstawie zrealizowanych przykładów nie można jednoznacznie stwierdzić, w konsekwencji użycia która baza HRTF umożliwi otrzymanie lepszych rezultatów pozycjonowania źródła dźwięku w przestrzeni otaczającej słuchacza. Jakość dźwięku i lokalizacja źródła dźwięku jest podobna w obu przypadkach. Wynik jest ciekawy, ponieważ przy zindywidualizowanych pomiarach HRTF można by oczekiwać, że zastosowanie bazy HRTF PWr będzie prowadzić do lepszych rezultatów, np. w lokalizacji obiektu dźwiękowego. Tego jednak w praktyce nie stwierdzono. Autorzy przypuszczają,

że prawdopodobnie wynika to ze zbyt małej bazy próbek dźwięku monofonicznego użytych do testów. Być może udałoby się znaleźć takie próbki dźwięku monofonicznego, dla których różnica w zastosowaniu omawianych baz HRTF byłaby widoczna, z jedn znacznym wskazaniem na bazę prowadzącą do uzyskania lepszych wyników wirtualizacji.

Drugim powodem braku jednoznacznych wyników testów jest przypuszczalnie brak przeprowadzonej interpolacji. Przy korzystaniu z interpolacji i wirtualnej zmianie położenia obiektu o np. 1° lub 2° dźwięk zmieniałby się płynnie od punktu do punktu. W takim scenariuszu przypuszczalnie dałoby się stwierdzić, która z rozpatrywanych baz HRTF prowadzi do lepszych wyników w kontekście użycia interpolacji przestrzennej. Tym samym problem interpolacji przestrzennej i włączenie procedur interpolacji do oprogramowania pisanego przez autorów będzie jednym z kierunków dalszych badań.

A kolejnym takim zagadnieniem będzie ponowny pomiar HRTF metodą podaną w pracach [1], [6], w kilku punktach przestrzennych i porównanie otrzymanych wyników z danymi zamieszczonymi w bazie HRTF PWr. Dzięki temu można się upewnić, czy HRIR z bazy PWr są faktycznie nieprzetworzonymi odpowiedziami impulsowymi dla ustalonego słuchacza. Po wykonaniu takich pomiarów (np. azymut zmieniający się z krokiem 10° i ustalona elewacja np. 0) stanie się możliwe porównanie wyników otrzymanych HRIR z wynikami z bazy ARI oraz ponowne przeprowadzenie syntezy dźwięku przestrzennego. Autorzy mają nadzieję, że tak uzyskane wyniki będą bardziej miarodajne, co pozwoli stwierdzić, która z baz HRTF jest bardziej przydatna do syntezy dźwięku przestrzennego.

Przy założeniu, że nowe pomiary kontrolne bazy PWr zostaną wykonane bez istotnych uchybień czy błędów, można przetworzyć pomiary analogicznie – tak jak to zostało wykonane w przypadku bazy ARI. W ten sposób zostanie otrzymany zbiór HRIR rzeczywistych, tzn. nieprzetworzonych odpowiedzi impulsowych, oraz zbiór pomiarów przetworzonych zbudowany z uwzględnieniem wielomianów minimalnofazowych i estymowanego czasu opóźnienia wzajemnego między lewym i prawym uchem słuchacza. W ten sposób będzie możliwe przeprowadzenie syntezy dźwięku przestrzennego za pomocą obydwu zbiorów HRIR i porównanie wyników. Pozwoli to sprawdzić w praktyce, czy przetworzone HRTF o mniejszej liczbie próbek mogą zastąpić rzeczywiste HRTF bez straty jakości realizowanego dźwięku.

Bibliografia

[1] Cheng C.I., Wakefield G.H., *Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space*, „Journal of the Audio Engineering Society" 2001, Vol. 49, No. 4, s. 231–249.

[2] Dobrucki A., Plaskota P., Pruchnicki P., Pec M., Bujacz M., Strumillo P., *Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with*

Visually Impaired and Sighted Individuals, „Journal of the Audio Engineering Society" 2010, Vol. 58, No. 9, s. 724–738.

[3] https://www.mathworks.com/help/matlab/ [dostęp: 09.09.2022].

[4] Freeland F.P., Biscainho L.W.P., Diniz P.S.R., *Interpolation of Head-Related Transfer Functions (HRTFs): A Multi-Source Approach*, 12th European Signal Processing Conference, Vienna, 2004, s. 1761–1764.

[5] Majdak P., Mihocic M., „HRTF-Database"; https://www.oeaw.ac.at/isf/das-institut/software/hrtf-database [dostęp: 09.09.2022].

[6] Møller H., Sørensen M. F., Hammershøi D., Jensen C. B., *Binaural Technique: Do We Need Individu-al Recordings?*, „Journal of the Audio Engineering Society" 1996, Vol. 44, No. 6, s. 451-469.

[7] Møller H., Sørensen M.F., Hammershøi D., Jensen C.B., *Head Related Transfer Functions of Human subjects*, „Journal of the Audio Engineering Society" 1995, Vol. 43, No. 5, s. 300–321.

[8] Pausch F., Doma S., Fels J., *Hybrid multi-harmonic model for the prediction of interaural time differences in individual behind-the-ear hearing-aid-related transfer functions*, „Acta Acustica" 2022, Vol. 34, s. 1–23.

[9] Plaskota P., *Research of Acoustical Impedance of Human Skin*, „Vibrations in Physical Systems" 2019, Vol. 30, No. 1.

[10] Plaskota P., Stasiak J., *Baza danych zawierająca wyniki pomiarów HRTF w formacie XML. Postępy akustyki*, Polskie Towarzystwo Akustyczne, Gliwice 2017.

[11] Pulkki V., *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*, „Journal of the Audio Engineering Society" 1997, Vol. 45, Iss. 6, s. 456–466.

[12] Rafaely B., Tourbabin V., Habets E., Ben-Hur Z., Lee H. et al., *Spatial audio signal processing for binaural reproduction of recorded acoustic scenes – review and challenges*, „Acta Acustica" 2022, Vol. 6, s. 1–19.

Słowa kluczowe: dźwięk przestrzenny, HRTF, HRIR, Matlab.

Synteza dźwięku przestrzennego z wykorzystaniem zindywidualizowanych pomiarów HRTF

W ostatnich latach rośnie zainteresowanie wykorzystaniem *head related transfer functions* (HRTF) do syntezy dźwięku przestrzennego w środowisku wirtualnym. Synteza dźwięku przestrzennego z wykorzystaniem HRTF pozwala na umieszczenie źródła dźwięku w przestrzeni otaczającej słuchacza bez konieczności wykorzystania wielokanałowego systemu odtwarzania dźwięku.

Niniejszy rozdział poświęcono syntezie dźwięku przestrzennego opartej na spersonalizowanych pomiarach HRTF wykonanej przede wszystkim przy użyciu środowiska obliczeniowego Matlab – dobrze przystosowanego do rozwiązywania różnorodnych problemów inżynierskich. Autorzy wykorzystali spersonalizowaną bazę pomiarów HRTF. Przeprowadzono syntezę wirtualnie poruszającego się źródła dźwięku jako nieruchomego źródła dźwięku – próba adaptacyjnego dostosowania długości sąsiadujących ramek czasowych między kolejnymi pozycjami źródła dźwięku nie przyniosła jednak jeszcze zadowalających rezultatów. Będzie to przedmiotem intensywnych badań w przyszłości.

Spatial Sound synthesis using individualised HRTF measurements

In recent years, there has been a growing interest in using Head Related Transfer Functions (HRTFs) to synthesise surround sound in a virtual environment. Spatial sound synthesis using HRTF allows a sound source to be placed in the space surrounding the listener without the need for a multi-channel audio playback system.

This paper deals with synthesis of spatial sound based on personalized measurements of HRTF (Head Related Transfer Functions). The main work has been done using Matlab computing environment that is well adapted to solve manifold engineering problems. Authors used the personalized base of HRTF measurements. Synthesis of the virtually moving sound source as the still sound source has been performed, however an attempt that was made to adaptively adjust the length of adjacent time frames between successive positions of the sound source did not achieve satisfactory results yet. It will be extensively studied in the future.

12. Adaptacyjny system kształtowania wiązki w polu bliskim oparty na uczeniu maszynowym

Agnieszka Wielgus

Politechnika Wrocławska,
Wydział Elektroniki, Fotoniki i Mikrosystemów,
wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław

12.1. Wprowadzenie

Zagadnienie kształtowania wiązki (ang. *beamforming*) pojawia się w wielu rzeczywistych systemach: bezprzewodowej komunikacji, radarach, sonarach, macierzach mikrofonów i anten [4], [5], [7], zarówno po stronie nadawczej, jak i odbiorczej. W systemach, w których należy odebrać sygnał pożądany, bardzo często mamy do czynienia z niepożądanym sygnałem lub sygnałami zakłócającymi. Jeżeli pasmo częstotliwości sygnału pożądanego pokrywa się częściowo lub w pełni z pasmem częstotliwości sygnału zakłócającego, zastosowanie klasycznych metod filtracji częstotliwości (ang. *classical temporal filtering*) [12] bazujących na wzmacnianiu bądź tłumieniu konkretnych składowych nie jest możliwe. W takich przypadkach wykorzystuje się tzw. filtrację przestrzenną (*spatial filtering*) [8], [15], która z wykorzystaniem zjawiska interferencji sygnałów umożliwia rozdzielenie sygnałów pożądanego i zakłócającego.

Zastosowanie metod filtracji przestrzennej wymaga wykorzystania macierzy sensorów. W przypadku problemu odbioru sygnału z zakłóceniami mającymi źródło poza obszarem, z którego pochodzi sygnał pożądany, wykorzystywana jest macierz mikrofonów. Każdy z mikrofonów traktowany jest wówczas jako filtr (o skończonej bądź nieskończonej odpowiedzi impulsowej). Jego współczynniki powinny zostać tak wyznaczone, aby spełnić zadane kryterium opisujące jakość systemu (np. maksymalizacja poziomu sygnału pożądanego, minimalizacja poziomu zakłóceń czy ich eliminacja).

Rozpatrywane podejście jest metodą dobrze znaną i doczekało się wielu modyfikacji [6], [9[, [11]. Jeżeli jednak położenie mówcy ulega zmianie, to odpowiedź systemu (war-

tość funkcji celu) również może ulec znaczącej zmianie. Na jakość systemu ma wpływ zatem nie tylko konfiguracja macierzy i jej parametry, lecz także położenie mówcy.

W dotychczasowych pracach dotyczące optymalizacji macierzy mikrofonów koncentrowano się głównie na doborze odpowiedniej konfiguracji macierzy, jej rozmiaru oraz współczynników filtrów, wag poszczególnych sygnałów. W publikacji [3] zaprezentowano metodę umożliwiającą jednoczesne wyznaczanie położenia mikrofonów i współczynników filtru. W tym celu przyjęto, że długość filtru FIR jest nieskończona i wówczas można zahamować wpływ zmiany położenia mikrofonów na parametry filtru. W konsekwencji takiego podejścia uzyskuje się znacznie lepsze wyniki niż w przypadku klasycznych konfiguracji macierzy.

Następnie wykazano, że zwiększanie rozmiaru macierzy nie zawsze jest optymalnym podejściem [10], [17]. Zastosowanie technik przerzedzania (ang. *thinning technique*) może znacząco poprawić jakość systemu [17]. Zgodnie z tą techniką aktywne pozostają jedynie wybrane mikrofony (lub anteny) – sygnały z pozostałych mikrofonów są pomijane podczas kształtowania wiązki.

Na podstawie otrzymanych rezultatów zaproponowano opisaną w tym rozdziale metodę, w której wykorzystuje się technikę przerzedzania dużej macierzy mikrofonów i doboru współczynników poszczególnych filtrów, tak aby macierz ta adaptowała się do poruszającego się źródła sygnału pożądanego (mówcy) oraz była odporna na zmiany systemu (np. awarię części mikrofonów). W tym podejściu ma zastosowanie metoda uczenia maszynowego, czyli tzw. uczenie ze wzmocnieniem (ang. *reinforcement learning*) [16]. Metody bazujące na uczeniu maszynowym zyskały w ostatnich latach olbrzymią popularność w wielu obszarach badawczych, również w kształtowaniu wiązki [1], [2], [8], [13], [14]. Rozważane zagadnienie polega na zaprojektowaniu adaptacyjnego systemu macierzy mikrofonów, która będzie w taki sposób dostosowywała zbiór aktywnych w danej chwili mikrofonów, żeby wyjście systemu (jego odpowiedź) było jak najbardziej zbliżone do pożądanej zgodnej z normą L_2.

12.2. Sformułowanie problemu

Dane jest środowisko, w którym występuje sygnał pożądany i sygnał zakłócający. Zakresy częstotliwości tych sygnałów pokrywają się, ale ich źródła mają różne położenie w przestrzeni. Przyjmuje się założenie, że pasmo sygnałów jest szerokie, a pole bliskie.

Dana jest prostokątna macierz mikrofonów o rozmiarze $\mathbf{N}$ ($n \times m$ równomiernie rozmieszczonych mikrofonów ze środkiem geometrycznym w r_c). Sygnały trafiające do mikrofonów są próbkowane w sposób synchroniczny, następnie kierowane na wejście filtru FIR rzędu L (zakłada się, że każdy element macierzy jest filtrem rzędu L o skończonej odpowiedzi impulsowej. Dodatkowo, że w danej chwili aktywnych może być wybrana liczba mikrofonów (1, 2, …, N)).

Zdefiniowano pożądaną odpowiedź systemu $G_d(r, L, f)$, gdzie: r – położenie źródła dźwięku, f – częstotliwość. Położenie źródła dźwięku nie jest stałe i może ulegać zmianom (np. mówca przemieszcza się po sali). Pożądana odpowiedź systemu zależy zatem od aktualnego położenia mówcy. Dla N-elementowej macierzy mikrofonów funkcja transmitancji i-tego aktywnego mikrofonu w polu bliskim to:

$$A_i(r, f)=\frac{1}{\|r-r_i\|}e^{-j2\pi f\|r-r_i\|/c} \tag{12.1}$$

gdzie c – prędkość dźwięku w powietrzu, r_i – położenie i-tego mikrofonu.

Zgodnie z [3] odpowiedź częstotliwościowa filtru FIR przyjmie postać:

$$H_i(h, f, L)=h_i^T d_0(f) \tag{12.2}$$

gdzie:

$$h_i=\left[h_i(0), h_i(1), \ldots, h_i(L-1)\right]^T, h_i \in R^L \tag{12.3}$$

$$d_0(f)=\left[1, e^{\frac{-j2\pi f}{f_s}}, \ldots, e^{\frac{-j2\pi f(L-1)}{f_s}}\right] \tag{12.4}$$

Informacja o opóźnieniu grupowym zawarta jest w funkcji transmitancji systemu.

Dla danego rozmieszczenia mikrofonów i ich liczby odpowiedź systemu może być wyznaczona z następującego równania:

$$G(r, f)=\sum_{i=1}^{N} H_i(h, L, f)A_i(r, f)=A^T(r, f)H(h, L, f) \tag{12.5}$$

Zaprojektowanie macierzy mikrofonów, stanowiące zasadniczy temat niniejszego rozdziału, oznacza konieczność określenia, które mikrofony w danej chwili powinny być aktywne, oraz wyznaczenia współczynników filtru FIR, tak aby aktualna odpowiedź systemu (wyjście) była jak najbliższa pożądanemu (zgodnemu z normą) L_2. Funkcję kosztu w przypadku ustalonej pozycji mówcy r opisuje się wzorem:

$$E(H)=\frac{1}{\|\Omega\|}\int_{\Omega}\sigma(r, f)\left|A^T(r, f)H(h, L, f)-G_d(r, L, f)\right|^2 dr\,df \tag{12.6}$$

gdzie $\sigma(r, f)$ – dodatnia funkcja wagi, Ω – dziedzina przestrzenno-częstotliwościowa składająca się z obszaru przepustowego Ω_P i obszaru zaporowego Ω_s, czyli $\Omega = \Omega_P \cup \Omega_s$. Obszar zaporowy oznacza pewien ograniczony fragment płaszczyzny oraz zakres częstotliwości sygnałów, których źródło ma położenie na wskazanym obszarze. Sygnały te powinny zostać stłumione. Sygnał pochodzący z obszaru przepustowego (o zdefiniowanych częstotliwościach) nie powinien ulec tłumieniu.

Wartość funkcji celu zależy zarówno od współczynników filtru FIR, jak i zbioru aktywnych mikrofonów (ich położenia). W ogólności wartości współczynników filtrów FIR są uzależnione od wybranego rzędu filtru. W pracy [3] autorzy zauważyli, że war-

tość funkcji celu nie rośnie wraz ze wzrostem rzędu filtru. Na tej podstawie określona została granica wydajności systemu. A na podstawie eksperymentu numerycznego wykazano, że wraz ze wzrostem rzędu filtru wartość funkcji celu zbiega szybko do minimum. W przypadku dostatecznie wysokiego rzędu filtru wartość funkcji celu nie ulega poprawie. Dla ustalonej długości filtru może to zatem oznaczać konieczność równoczesnego określenia zbioru aktywnych mikrofonów i współczynników filtru.

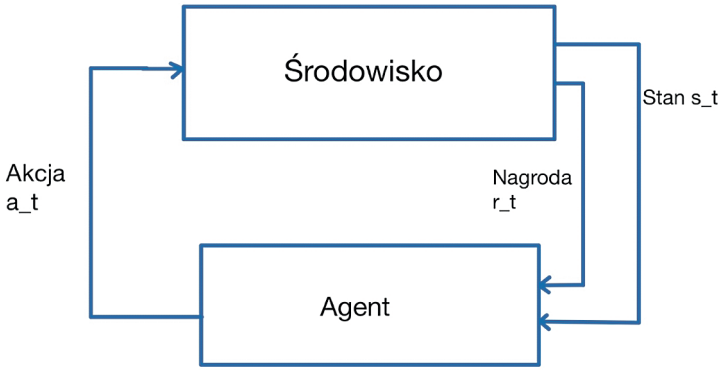
Niech $\boldsymbol{\lambda} = (r_1, r_2, ..., r_N)$ to wektor przechowujący położenie poszczególnych mikrofonów, a $\boldsymbol{\lambda_a} = (r_{a1}, r_{a2}, ..., r_{aM})$ – wektor położeń aktywnych mikrofonów, przy czym $M \leq N$. Rozpatrywany w rozdziale problem może zostać zdefiniowany formalnie w postaci:

$$\min_{\lambda a,\, h \in R^{ML}} E\left(h, \lambda_a, r\right) \tag{12.7}$$

gdzie $\boldsymbol{\lambda_a}$ – zmienna decyzyjna dla danego $\boldsymbol{r}$ (położenia mówcy). W przypadku ustalonego $\boldsymbol{\lambda_a}$ funkcja (12.7) jest wypukła, ale ze względu na wektor $\boldsymbol{\lambda_a}$ – niewypukła.

12.3. Algorytm rozwiązania

W celu rozwiązania zdefiniowanego problemu zaproponowano metodę uczenia maszynowego, czyli tzw. uczenie ze wzmocnieniem. Podstawowy schemat uczenia ze wzmocnieniem modelowany jest jako proces decyzyjny Markova, występują w nim: zbiór stanów S określających środowisko i agenta, zbiór akcji A, które może podjąć agent, prawdopodobieństwo przejścia ze stanu s do stanu s' po wykonaniu akcji a $P_a(s, s')$, nagroda $R_a(s, s')$ po wykonaniu akcji a polegającej na przejściu ze stanu s do stanu s'. Zastosowano w tym rozdziale Q-learning – metodę uczenia przez wzmocnienie niewymagającą modelu środowiska/systemu, co oznacza pominięcie zbioru prawdopodobieństw $P_a(s, s')$. W zaimplementowanym podejściu są agent, zbiór stanów S, a także zbiór możliwych



Rys. 12.1. Schemat algorytmu uczenia ze wzmocnieniem

do wykonania w stanie *s* akcji. Przez wykonanie akcji $a \in A$ agent dokonuje przejścia ze stanu *s* do stanu *s'*. Za wykonanie określonej akcji agent otrzymuje nagrodę (karę) w postaci liczbowej. Ogólny schemat uczenia ze wzmocnieniem przedstawiono na rys. 12.1.

Celem agenta jest maksymalizacja nagrody bądź minimalizacja kary. Agent dokonuje tego na podstawie wiedzy o nagrodzie, którą może uzyskać za znalezienia się w kolejnym stanie – uwzględnia możliwe przyszłe nagrody. W tym celu w algorytmie występuje macierz **Q**, która przechowuje informacje na temat korzyści z wykonania danej akcji w konkretnym stanie. Wartości macierzy w danym kroku aktualizowane są zgodnie ze wzorem:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a \left(Q(s_{t+1}. a_t) \right) - Q(s_t, a_t) \right] \tag{12.8}$$

gdzie $Q(s_t,a_t)$ – wartość Q w aktualnym stanie, $\gamma \max_a (Q(s_{t+1}\cdot a_t))$ – maksymalna nagroda (minimalna kara), która może być uzyskana podczas przejścia ze stanu s_{t+1} do kolejnego stanu, r_t – nagroda za wykonanie akcji a_t w stanie s_t, parametry $\alpha,\gamma \in (0,1)$. Algorytm kończy działanie po uzyskaniu przez system stanu końcowego nazywanego celem.

W analizowanym przypadku przyjęto następujące założenia:

- stan s_t – konfiguracja mikrofonów w chwili *i* – zbiór aktywnych mikrofonów;
- akcja a_t – włączenie/wyłączenie danego mikrofonu;
- macierz **Q**: $s \times$ – przechowująca „nagrody" związane w wyborem akcji a_t dla stanu s_t;
- nagroda r_t – zmiana jakości systemu, tj. różnica między kryteriami aktualnym a następującym;
- aktualizacja macierzy **Q** – przypisanie nagrody za przejście do innego stanu zgodnie ze wzorem (12.8).

Proces uczenia przebiega następująco:

1. Określ stan początkowy – wszystkie mikrofony są aktywne.
2. Określ cel – przy użyciu algorytmu metaheurystycznego wyznacz rozwiązanie suboptymalne dla danej konfiguracji mikrofonów (wszystkie mikrofony są równomiernie rozmieszczone na obszarze prostokąta, szukamy zbioru aktywnych mikrofonów), przyjmij uzyskaną wartość kryterium za cel (uzyskanie dokładnie takiego samego rozwiązania może nie być możliwe podczas uczenia, przyjęto zatem możliwość błędu na pewnym akceptowalnym poziomie).
3. Ustal liczbę iteracji uczenia.
4. Rozpocznij uczenie agenta – macierzy mikrofonów.
5. Dokonaj weryfikacji uzyskanego rozwiązania.

12.4. Eksperyment numeryczny

W tym podrozdziale przedstawione zostaną wyniki eksperymentu numerycznego. Algorytmy zaimplementowano w środowisku Matlab. Przyjęta instancja testowa w rozpatrywanym przypadku to obszar Ω zdefiniowany w [3] – obszar przepustowy:

$$\Omega_p = \{(r, f): -0{,}4\,\text{m} \le |x| \le 0{,}4\,\text{m},\ y = 0\,\text{m},\ 0{,}5\,\text{kHz} \le f \le 1{,}5\,\text{kHz}\}$$

oraz obszar zaporowy:

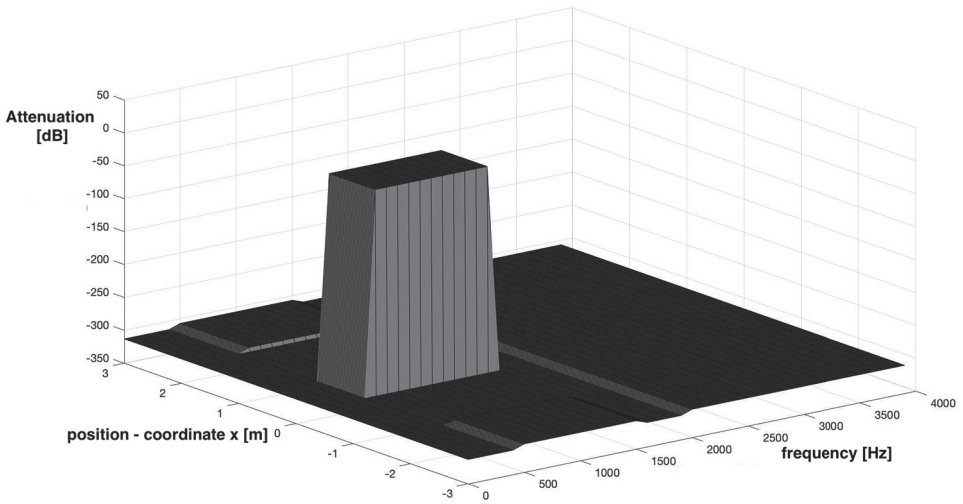
$$\Omega_s = \{(r, f): -3{,}0\,\text{m} \le |x| \le 3{,}0\,\text{m},\ y = 0\,\text{m},\ 2{,}0\,\text{kHz} \le f \le 4{,}0\,\text{kHz}\} \cup \{(r, f): 1{,}8\,\text{m} \le |x| \le 3{,}0\,\text{m},\ y = 0\,\text{m},\ 0{,}5\,\text{kHz} \le f \le 1{,}5\,\text{kHz}\} \cup \{(r, f): -3{,}0\,\text{m} \le |x| \le -1{,}8\,\text{m},\ y = 0\,\text{m},\ 0{,}5\,\text{kHz} \le f \le 1{,}5\,\text{kHz}\}.$$

Pożądaną odpowiedź systemu w obszarze przepustowym można opisać wzorem:

$$G_d(\lambda, r, f) = \mathrm{e}^{-j\pi f\left(\frac{\|r - r_c\|}{c} + \frac{L-1}{2}T\right)}, \tag{12.9}$$

gdzie $r_c = \sum_{i=1}^{M} r_i / M$ – środek geometryczny macierzy mikrofonów dla rozmieszczenia $\boldsymbol{\lambda}_a$, c = 340,9 m/s – prędkość dźwięku w powietrzu. W przypadku badanej instancji problemu pożądaną odpowiedź systemu przedstawiono na rys. 12.2 (przekrój dla jednej wartości y).

Częstotliwość próbkowania przyjęto zgodnie z twierdzeniem Nyquista równą 8 kHz. Pasmo sygnału ograniczone jest do 4 kHz, podobnie jak w pracy [3]. Pasmo to jest wystarczające dla sygnału mowy. Parametr $\sigma(r,f) = 1$. Długość filtru przyjęto równą 40, podobnie jak w pracy [3]. Założono, że mikrofony mogą być rozmieszczone na płaszczyźnie 2D. Za cel – kryterium STOP przyjęto wartość kryterium w rozwiązaniu uzyskanym przez algorytm symulowanego wyżarzania zaproponowany w publikacji [17]. Algorytm



Rys. 12.2. Pożądana odpowiedź systemu

uczenia zatrzymuje się, kiedy wartość kryterium dla wyuczonej macierzy nie jest gorsza o więcej niż 10% od rozwiązania dostarczonego przez symulowane wyżarzanie (SA). Podejście takie umożliwia znalezienie rozwiązania lepszego niż to dostarczone przez SA. Może być ono postrzegane jako metoda typu popraw.

Obszary przepustowy i zaporowy zostały zdyskretyzowane (rozdzielczość częstotliwościowa to 0,1 kHz), a przestrzeń zdyskretyzowana co 0,02 m. W pracy [3] wykazano, że w przypadku analizowanej instancji problemu optymalne jest nierównomierne rozmieszczenie pięciu mikrofonów w następujących punktach: $\lambda^* = \{(0{,}1\ 4975), (0{,}1\ 4736), (0{,}1\ 4139), (0{,}1\ 2585), (0{,}09)\}$, w konsekwencji wartość funkcji celu równa się –52,76 dB.

Eksperyment numeryczny przeprowadzono dla macierzy składającej się z 25 mikrofonów w konfiguracji 5 × 5 mikrofonów rozmieszczonych równomiernie na obszarze $x = -0{,}1, 0{,}1]$; $y = [0{,}9, 1{,}5]$.

Gdy wszystkie mikrofony są aktywne, otrzymuje się wartość kryterium równą –16,88 dB. Kryterium uzyskane przez zastosowanie algorytmu bazującego na symulowanym wyżarzaniu (SA) [17] to: –41,72 dB. Zastosowano 100 iteracji (epok) uczenia, kryterium po uczeniu to –41,78 dB. Uzyskaną odpowiedź systemu przedstawiono na rys. 12.3.

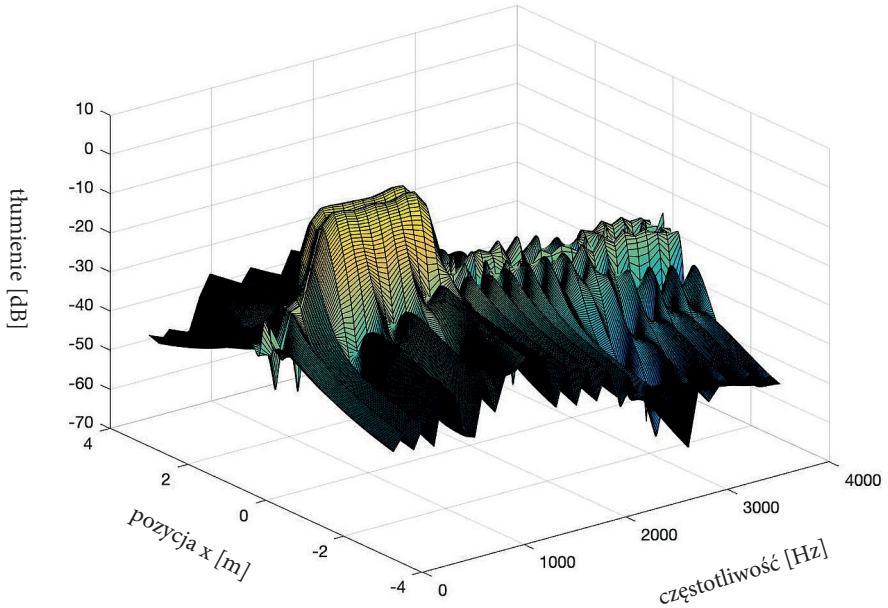
Następnie sprawdzono, czy macierz potrafi zaadaptować się w przypadku zmiany dostępności mikrofonów. Uruchomiono ponownie algorytm SA dla pełnej macierzy. Uzyskana wartość kryterium –38,85 dB – różni się od poprzedniej ze względu na pewną losowość w przeszukiwaniu zbioru rozwiązań cechującą symulowane wyżarzanie. Później dokonano uczenia (na pełnej macierzy) i ze zbioru dostępnych mikrofonów usunięto losowo jeden z mikrofonów. Na podstawie wyuczonej macierzy algorytm zaproponował rozwiązanie o wartości kryterium –37,62 dB (odpowiedź przedstawiono na rys. 12.4).

Kolejno usunięto możliwość wyboru trzech losowych mikrofonów. Eksperyment przeprowadzono analogicznie do przypadku z jednym usuniętym mikrofonem. Kryterium z SA: –41,27 dB, 100 iteracji uczenia, wyłączona możliwość włączenia trzech losowych mikrofonów, na podstawie wyuczonej macierzy **Q** wyznaczone zostaje rozwiązanie, kryterium: –40,10 dB. Odpowiedź systemu przedstawiono na rys. 12.5.

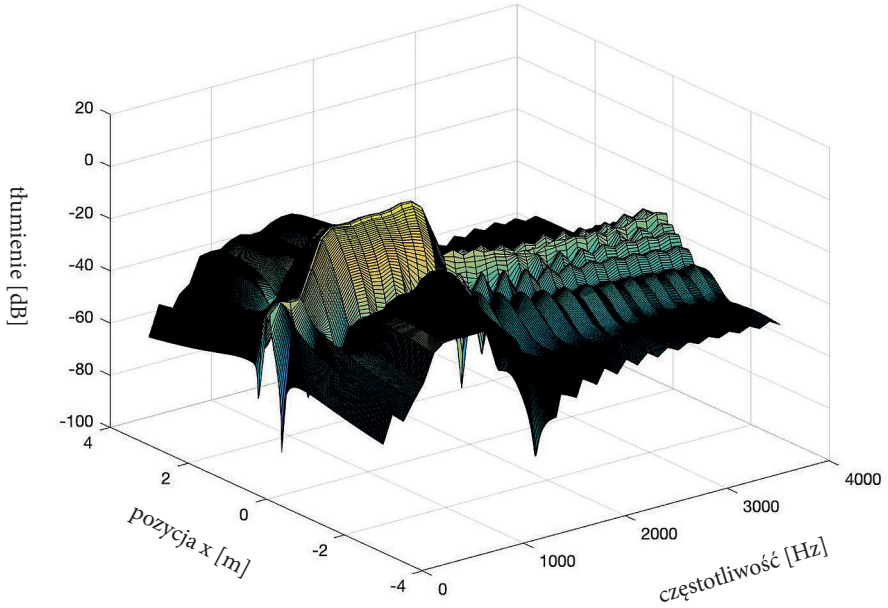
Wielokrotne uruchamianie testów w przypadku niedostępnych losowych mikrofonów (jednego i trzech) potwierdziło rezultaty przedstawione na rys. 12.4 i rys. 12.5, tzn. algorytm znajdował rozwiązanie charakteryzujące się bardzo dobrą jakością. Dalsze zwiększanie liczby niedostępnych mikrofonów (dla pięciu niedostępnych mikrofonów: –40,46 dB, dla ośmiu: –40,12 dB, dla dziesięciu: –37,62 dB, a dla 15: –35,15 dB) również znacząco nie wpłynęło na jakość dostarczanego rozwiązania. Sygnały pochodzące z obszarów zabronionych były tłumione w dużym stopniu. Metoda bazująca zatem na uczeniu ze wzmocnieniem może być wykorzystywana w systemach, w których należy uwzględnić możliwość awarii/niedostępności mikrofonów i zapewnić systemowi odporność.

Następnie wzięto pod uwagę przypadki, w których źródło sygnału się przemieściło. Początkowo obszar przepustowy uległ zmianie zgodnie ze wzorem:

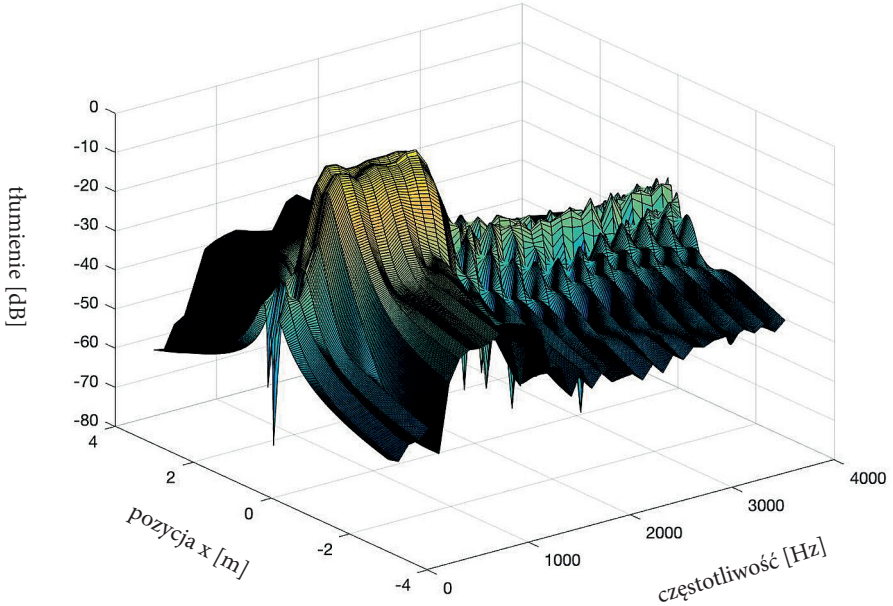
$$\Omega_p = \left\{ (r, f) : -0{,}3\,\text{m} \le |x| \le 0{,}5\,\text{m},\ y = 0\,\text{m},\ 0{,}5\,\text{kHz} \le f \le 1{,}5\,\text{kHz} \right\}$$



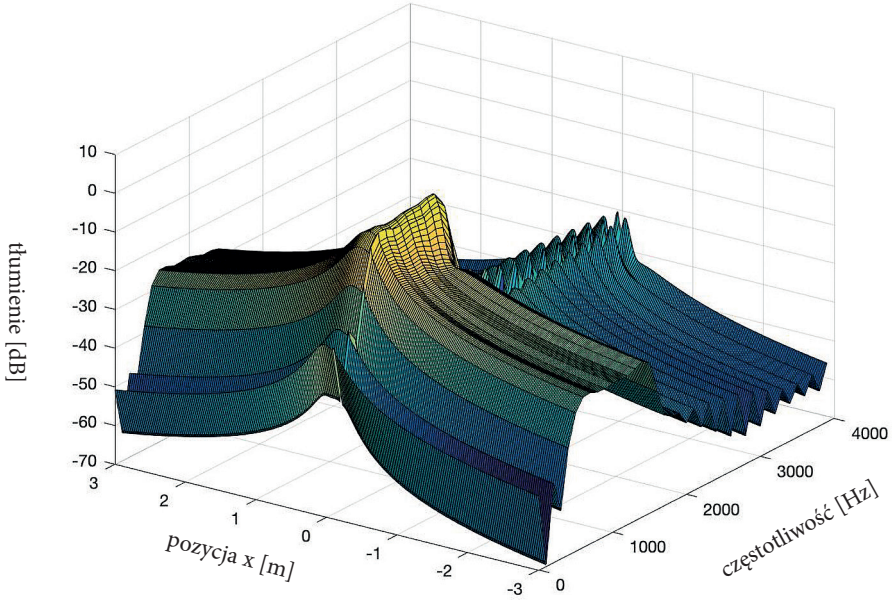
Rys. 12.3. Odpowiedź systemu – możliwość wyboru wszystkich mikrofonów



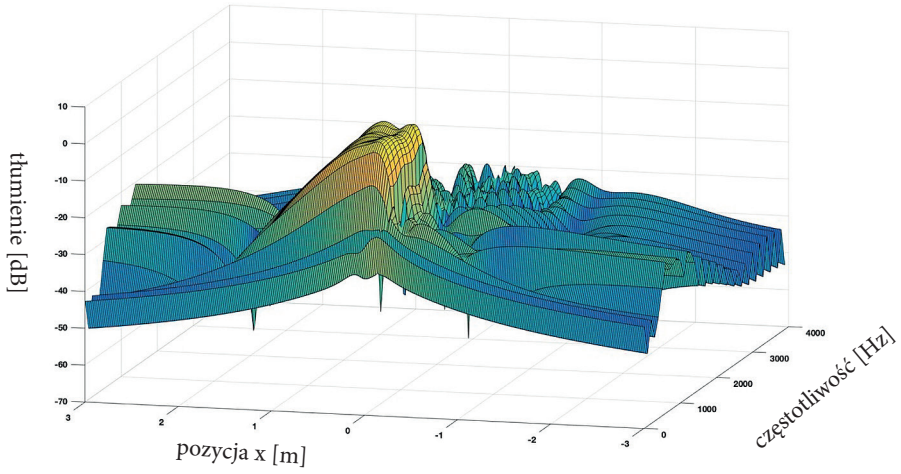
Rys. 12.4. Odpowiedź systemu – losowy mikrofon jest niedostępny



Rys. 12.5. Odpowiedź systemu – trzy losowe mikrofony są niedostępne



Rys. 12.6. Odpowiedź systemu w przypadku, w którym zmieniło się położenie źródła sygnału: $-0{,}3\text{m} \leq |\, x \,| \leq 0{,}5$ m



Rys. 12.7. Odpowiedź systemu w przypadku, w którym zmieniło się położenie źródła sygnału: $-0{,}1\,\mathrm{m} \le |x| \le 0{,}7\,\mathrm{m}$

System został nauczony dla przypadku z pierwotnym położeniem źródła sygnału pożądanego. Następnie sprawdzono, czy tak wyuczona macierz zaadaptuje się do zmiany położenia źródła sygnału. Uzyskano następującą odpowiedź systemu (por. rys. 12.6) – charakterystyka częstotliwościowo-przestrzenna systemu uległa zmianie, zaobserwować można dostosowanie charakterystyki do przesuniętego obszaru przepustowego. Charakterystyka z jednej strony (w usuniętym obszarze) jest jednak znacznie bardziej stroma. Kolejno przesunięto obszar przepustowy tak, aby różnica między obszarem, dla którego macierz została wyuczona była większa:

$$\Omega_p = \{(r, f): -0{,}1\,\mathrm{m} \le |x| \le 0{,}7\,\mathrm{m},\ y = 0\,\mathrm{m},\ 0{,}5\,\mathrm{kHz} \le f \le 1{,}5\,\mathrm{kHz}\}$$

Odpowiedź systemu przedstawiono na rys. 12.7. Również w tym przypadku można zauważyć, że charakterystyka częstotliwościowo-przestrzenna jest bardziej stroma w usuniętym obszarze w porównaniu do obszaru, dla którego system został wyuczony. W obszarze dodanym do pierwotnego charakterystyka jest znacznie mniej stroma, ale nadal ma kształt wskazujący na dostosowanie się macierzy do poruszającego się mówcy.

12.5. Podsumowanie

W rozdziale zbadano możliwość wykorzystania algorytmu uczenia ze wzmocnieniem do konfiguracji macierzy mikrofonów. W badaniach eksperymentalnych wykazano, że przedstawiona metoda może być stosowana do ustalania konfiguracji macierzy. Dodat-

kowo raz wyuczony system adaptuje się szybko do zmian (np. braku możliwości wykorzystania w macierzy wcześniej wybranych mikrofonów). Zastosowane podejście wymaga jednak dalszych analiz w przypadku zmiany położenia źródła sygnału pożądanego. Na podstawie wstępnych rezultatów, jakie uzyskano, można wnioskować, że system próbuje adaptować się do zmian, ale kształt, tj. znaczna niesymetryczność uzyskanej charakterystyki (odpowiedzi systemu), wskazuje na to, że problem wymaga kontynuacji badań. W tym celu sprawdzone zostanie podejście bazujące na sieci neuronowej.

Bibliografia

[1] Almeida N.C., Fernandes M.A.C., Neto A.D.D, *Beamforming and power control in sensor arrays using reinforcement learning*, „Sensors" 2015, Vol. 15, No. 3, s. 6668–6687.

[2] Barhoush M. et al., *Localization-Driven Speech Enhancement in Noisy Multi-Speaker Hospital Environments Using Deep Learning and Meta Learning*, „IEEE/ACM Transactions on Audio, Speech, and Language Processing" 2022, Vol. 31, s. 670–683.

[3] Feng Z.G., Yiu K.F.C., Nordholm S.E., *Placement design of microphone arrays in near-field broadband beamformers*, „IEEE Transactions on Signal Processing" 2012, Vol. 60, No. 3, s. 1195–1204.

[4] Griffith D.A., *Spatial filtering*, Springer, Berlin–Heidelberg 2003.

[5] Johnson D.H., Dudgeon D.E., *Array Signal Processing: Concepts and Techniques*, Simon & Schuster, New York 1992.

[6] Le Son P., Phan, *Irregular microphone array design for broadband beamforming*, „Signal Processing" 2022, Vol. 193, Iss. C.

[7] Liu W., Weiss S., *Design of frequency invariant beamformers for broadband arrays*, „IEEE Transactions on Signal Processing" 2008, Vol. 56, No. 2, s. 855–860.

[8] Kayser C., Kujawski A., Sarradj E., *A trainable iterative soft thresholding algorithm for microphone array source mapping*, Proceedings of the CD of the 9th Berlin Beamforming Conference, Berlin, Germany, 2022.

[9] Kennedy R.A., Ward D.B., Abhayapala T.D., *Nearfield beamforming using radial reciprocity*, „IEEE Transactions on Signal Processing" 1999, Vol. 47, No. 1, s. 33–40.

[10] Mingjie G., Yiu K.F.C., Nordholm S., *On the sparse beamformer design*, „Sensors" 2008, Vol. 18, No. 10.

[11] Nordebo S., Claesson I., Nordholm S., *Weighted chebyshev approximation for the design of broadband beamformers using quadratic programming*, „IEEE Signal Processing Letters" 1994, Vol. 7, No. 1, s. 103–105.

[12] Oppenheim A.V., Willsky A.S., Nawab S.H., *Signals & System*, Prentice-Hall, Hoboken 1996.

[13] Salvati D., Drioli C., Foresti G.L., *On the use of machine learning in microphone array beamforming for far-field sound source localization*, IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2016.

[14] Tarafder P., Choi W., *Deep Reinforcement Learning-Based Coordinated Beamforming for mmWave Massive MIMO Vehicular Networks*, „Sensors" 2023, Vol. 23, No. 5, s. 2772.

[15] Veen B.D. vaan, Buckley K.M., *Beamforming: A versatile approach to spatial filtering*, „IEEE ASSP Magazine" 1988, Vol. 5, No. 2, s. 4–24.

[16] Watkins Ch.J.C.H., *Learning from Delayed Rewards*, praca doktorska, King's College, Cambridge 1989.
[17] Wielgus A., Szlachetko B., *A Simulation of Thinning of Microphone Array in Near-field Broadband Beamformers*, „Vibrations in Physical Systems" 2021, Vol. 32, No. 2.

Słowa kluczowe: kształtowanie wiązki, uczenie maszynowe, optymalizacja.

Adaptacyjny system kształtowania wiązki w polu bliskim oparty na uczeniu maszynowym

Niniejszy rozdział dotyczy zagadnienia możliwości wykorzystania uczenia maszynowego w kształtowaniu wiązki w polu bliskim i sygnału szerokopasmowego (mowy ludzkiej). Celem ukształtowania wiązki było wzmocnienie sygnału pożądanego (np. sygnału mówcy) przy jednoczesnym zminimalizowaniu poziomu zakłóceń. Pasmo częstotliwości sygnału pożądanego oraz zakłócającego pokrywa się w pewnym zakresie, zastosowanie filtracji częstotliwościowej nie jest zatem możliwe. Żeby rozwiązać rozpatrywany problem skonstruowano algorytm oparty na tzw. uczeniu ze wzmocnieniem oraz zbadano jego efektywność. Pod uwagę wzięty był również przypadek, w którym zbiór dostępnych mikrofonów ulega zmianie (część mikrofonów nie jest dostępna). W przeprowadzonym eksperymencie numerycznym wykazano, że zaproponowane podejście dostarcza rozwiązań charakteryzujących się wysoką jakością i system po procesie uczenia potrafi adaptować się do zmian dostępnych konfiguracji macierzy. Dodatkowo zbadano wpływ zmiany położenia mówcy na jakość rozwiązania. Na podstawie symulacji udowodniono, że zaproponowany system dostosowuje się do zmian położenia mówcy.

Adaptive near-field broadband beamforming system based on machine learning

This work deals with the problem of near-field broadband beamforming. Since the issue of the number and position of individual microphones is of great importance in the design process of broadband beamforming, we try to propose an adaptive beamforming system in which the number and position of microphones depend on the current speaker's position (we assume that the speaker can move in a restricted, previously defined area). For this purpose, we consider a big rectangular microphone matrix, in which any number of microphones can be active depending on the speaker's position. The applied method is based on reinforcement learning (RL) – a machine learning methodology in which agent learns how to maximize returns or achieve the given goal through the system of rewards and punishments. RL has close connections to both adaptive control and optimization. We try to learn our system to adapt to changing speaker's positions by changing the set of active microphones and its filters' coefficients so that the output of the system is as close as possible to the optimal output in the sense of the l_2 norm.

13. Algorytm uprzestrzenniający sygnały dźwiękowe oparty na przesunięciach fazowych

Kamil Zimny, Teresa Makuch

Akademia Górniczo-Hutnicza w Krakowie,
Wydział Inżynierii Mechanicznej i Robotyki,
al. Adama Mickiewicza 30, 30-059 Kraków

13.1. Wprowadzenie

Słyszenie przestrzenne jest właściwością układu słuchowego, wykorzystywaną do wielu zadań, m.in. orientacji w przestrzeni, szacowania wymiarów pomieszczenia i parametrów źródeł dźwięku. W podstawowych modelach słuchowych je opisujących zakłada się, że układ słuchowy wykorzystuje międzyuszne różnice między poziomami dźwięku (ang. *Interaural Level Difference* – ILD) i czasem dotarcia fali dźwiękowej (ang. *Interaural Time Difference* – ITD) do określenia położenia źródła [7]. Subiektywne wrażenie przestrzenności sygnału jest jednak czymś więcej niż tylko lokalizacją źródła dźwięku – dlatego takie wyjaśnienie okazuje się niewystarczające. Wynika to z tego, że mechanizmy za nie odpowiedzialne nie zostały jeszcze dokładnie przebadane.

13.1.1. Badania Griesingera

Problematyką słyszenia przestrzennego zajmował się David Griesinger – badacz i twórca legendarnych algorytmów pogłosowych znanych pod szyldem marki Lexicon®. Niektóre z jego prac dotyczyły sal koncertowych [3], [4]. Wiele z nich charakteryzowało się właściwymi parametrami związanymi z rozchodzeniem się fali akustycznej i pogłosowością (np.: czas pogłosu, parametry klarowności, przejrzystości). Mimo tego nie gwarantowały one optymalnych wrażeń przestrzenności dźwięku podczas odsłuchu – zadaniem Griesingera była poprawa warunków odsłuchowych rozpatrywanych pomieszczeń. Zapro-

ponował w związku z tym inne podejście do tematu słyszenia przestrzennego, bazujące na opracowanym nowym modelu słuchowym: według Griesingera układ słuchowy jest czuły na koherencję (spójność fazową) między składowymi harmonicznymi w sygnałach złożonych i to ona bezpośrednio wpływa na subiektywne wrażenie przestrzenności [2], co zostało wstępnie potwierdzone w badaniach jednego z autorów tego rozdziału [6]. W ogólności faza składowych sygnału może być też odpowiedzialna za wrażenia związane z barwą dźwięku, zakolorowaniem i klarownością wrażenia wysokości [5].

13.1.2. Algorytmy wpływające na przestrzenność sygnałów dźwiękowych

Możliwość kontrolowania przestrzenności sygnałów dźwiękowych jest istotnym zagadnieniem w szeroko pojętym przetwarzaniu sygnałów. Można wymienić wiele algorytmów, które wpływają na przestrzenność sygnałów, lecz zazwyczaj opierają się one na modelowaniu zjawisk fizycznych naturalnie wpływających na to wrażenie. W implementacjach (algorytmach) tego typu liczne parametry umożliwiają precyzyjną kontrolę symulowanych procesów. Na przykład algorytmy symulujące zjawisko pogłosu zazwyczaj pozwalają na kontrolowanie m.in. czasu pogłosu, opóźnienia pierwszego odbicia czy parametrów symulowanego pomieszczenia.

W praktyce jednak problematyczne okazuje się ich przełożenie na subiektywne wrażenie przestrzenności dźwięku. Co więcej, modelowanie zjawisk fizycznych oprócz wpływu na przestrzenność sygnału zazwyczaj modyfikuje także inne właściwości dźwięku, tj. barwę, głośność, obwiednię sygnału – przekładające się na przykład na klarowność i przejrzystość. W niektórych zastosowaniach takie efekty mogą okazać się niepożądane.

13.2. Założenia projektowe

Z uwzględnieniem wcześniejszych uwag postawiono sobie za cel zaprojektowanie algorytmu uprzestrzenniającego sygnały dźwiękowe z wykorzystaniem modyfikacji fazowych. Od strony przetwarzania sygnałów rozpatrywano dwa podejścia – wykonywanie operacji w dziedzinie czasu (z wykorzystaniem filtrów) lub w dziedzinie częstotliwości (operacje na widmie sygnału). Na obecnym etapie rozwoju algorytmu przyjęto drugą z koncepcji ze względu na mniejszą złożoność implementacji i co się z tym łączy: szybsze uzyskanie efektów. Na dalszym – rozważana jest implementacja trybu modyfikacji fazowych przy użyciu filtrów o skończonej odpowiedzi impulsowej (ang. *Finite Impulse Response* – FIR) z powodu liniowości przetwarzania fazy oraz ich stabilności.

Aby oddziaływać na wrażenie przestrzenności, modyfikacje fazowe muszą być wprowadzane w sposób ciągły. Dlatego założono, że w pierwszej kolejności sygnał wejściowy

jest poddawany procesowi segmentacji (okienkowania), tak aby umożliwić późniejszą rekonstrukcję za pomocą metody *overlap-add* [1]. W ten sposób modyfikacje fazowe są wprowadzane dla każdego okna oddzielnie, co umożliwia uzyskanie ciągłych zmian tego parametru sygnału.

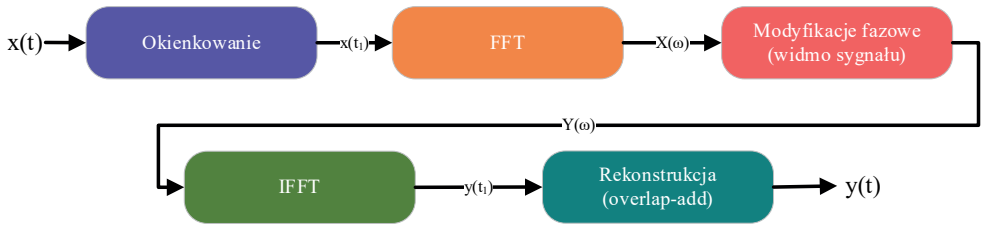
W przeprowadzonych przez autorów wstępnych badaniach naukowych wykazano, że wprowadzanie losowych modyfikacji fazowych (zgodnie z rozkładem równomiernym) oddziałuje na wrażenie pogłosowości [6]. Aby dokładniej zbadać to zjawisko, przyjęto, że zastosowanie algorytmu umożliwi wprowadzanie przesunięć fazowych na kilka sposobów, m.in. losowo z uwzględnieniem różnych rozkładów czy zgodnie z przebiegami funkcji matematycznych. Modyfikacjom będą podlegały składowe harmoniczne sygnałów lub wybrane pasmo częstotliwości. Dlatego początkowo algorytm zaoferuje wiele parametrów odpowiadających dostępnym trybom działania. Docelowo po przebadaniu różnych sposobów wprowadzania modyfikacji liczba parametrów zostanie ograniczona w celu ułatwienia obsługi programu.

Algorytm ma posłużyć za narzędzie badawcze do testowania wpływu różnego typu modyfikacji fazowych na subiektywne wrażenie przestrzenności. Z tego względu istotną kwestią jest możliwość działania algorytmu także w czasie rzeczywistym. Dlatego opracowano także wersję algorytmu w postaci wtyczki programowej (VST). Dzięki niej możliwe będzie jego testowanie w cyfrowych stacjach roboczych przez producentów i realizatorów dźwięku. To kolejny powód, by na obecnym etapie prac nie decydować się na wykorzystanie filtracji FIR – ze względu na wprowadzane opóźnienia.

13.2.1. Implementacja

Algorytm został zaimplementowany przy użyciu środowiska Matlab – jego schemat blokowy przedstawiono na rys. 13.1.

Zgodnie z przedstawionym schematem sygnał wejściowy w pierwszej kolejności poddawany jest procesowi okienkowania. Aby późniejsza rekonstrukcja sygnału była wykonywana poprawnie, zastosowano okno Hamminga z nakładkowaniem 75%, co



Rys. 13.1. Schemat blokowy z zaznaczonym kierunkiem transmisji sygnału w algorytmie uprzestrzenniającym

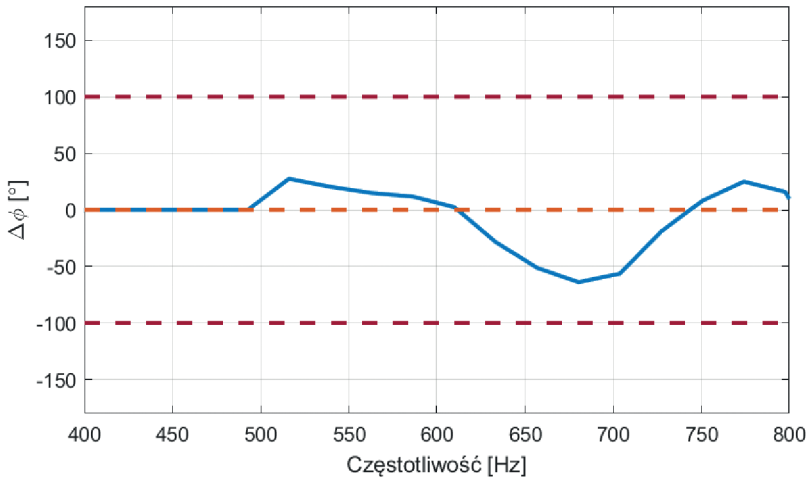
przekłada się na płynność przejść między oknami, a tym samym na brak dodatkowych (niepożądanych) zniekształceń. Długość okna będąca kompromisem między częstotliwością zmian fazy a pozostawieniem sygnału bez słyszalnych zniekształceń została ustalona na 1024 próbki (częstotliwość próbkowania 44 100 Hz). Na dalszym etapie rozwoju algorytmu założono wprowadzenie mniejszego rozmiaru okien, co wymaga wprowadzenia także dodatkowych ograniczeń odnośnie do wartości modyfikacji fazy, tak aby uniknąć jej skokowych zmian – może to powodować słyszalne zniekształcenia.

Okienkowany sygnał jest poddawany Szybkiej Transformacji Fouriera (ang. *Fast Fourier Transform* – FFT) o liczbie punktów równej lub wyższej od rozmiaru okna. Parametr liczby punktów decyduje o dokładności (rozdzielczości) działania algorytmu w dziedzinie częstotliwości, przy czym jeśli byłby on mniejszy niż rozmiar okna, późniejsza poprawna rekonstrukcja sygnału nie jest możliwa. Z kolei górnym ograniczeniem liczby punktów zdaje się być jedynie dostępna moc obliczeniowa komputera.

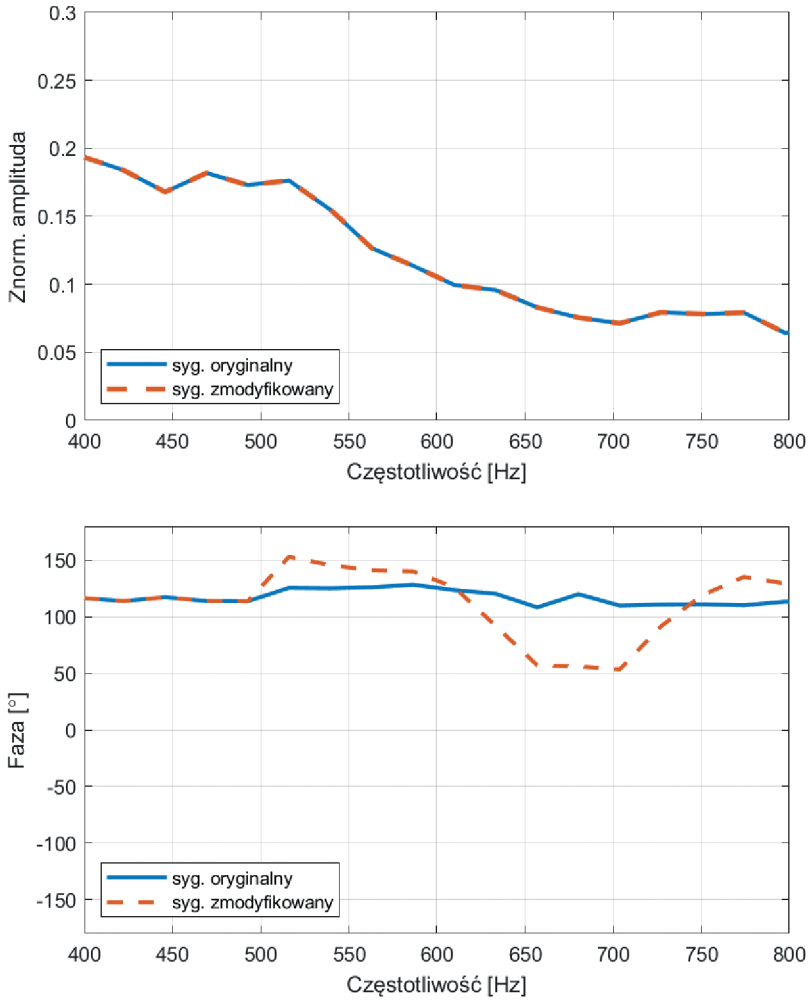
W dalszej kolejności wprowadza się modyfikacje fazowe sygnału bazujące na widmie amplitudowo-częstotliwościowym. Na tym etapie zaimplementowano ich wykonywanie:

- w poszczególnych pasmach częstotliwości,
- dla składowych harmonicznych sygnału.

W algorytmie wykorzystano mechanizm rozpoznawania składowych harmonicznych sygnału, który określa daną składową widma jako harmoniczną, jeśli jej amplituda przekracza ustalony próg oraz jeśli jej częstotliwość jest wielokrotnością częstotliwości podstawowej (z określonym marginesem dokładności). Przesunięcia fazowe mogą być wprowadzane do jednej bądź kilku harmonicznych.



Rys. 13.2. Przykład wprowadzonych losowo przesunięć fazowych do sygnału przetwarzanego przez algorytm (rozkład równomierny)



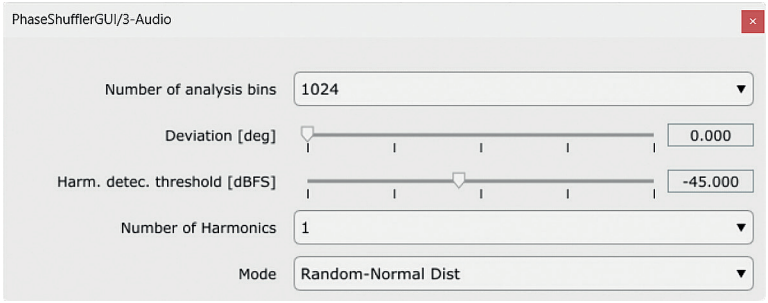
Rys. 13.3. Fragment widma amplitudowego (na górze; krzywe pokrywają się) i fazowego (na dole) sygnału poddanego działaniu algorytmu

Dotychczas w algorytmie zaimplementowano modyfikowanie fazy sygnału w sposób losowy z uwzględnieniem rozkładów: równomiernego, normalnego, *T*-studenta, Poissona. Jako następny krok zaplanowane zostało wprowadzenie modyfikacji zgodnie z przebiegami funkcji matematycznych, w tym funkcji okresowych (np. sinus). Przykład działania algorytmu z zastosowaniem losowych przesunięć fazy (rozkład równomierny) w paśmie częstotliwości 500–800 Hz przedstawiono na rys. 13.2.

Po modyfikacji fazy sygnał zostaje przeniesiony do dziedziny czasu z wykorzystaniem odwrotnej szybkiej transformaty Fouriera (ang. *Inverse Fast Fourier Transform* – IFFT).

Jeśli rozdzielczość analizy FFT była wyższa od długości okna, sygnał zostaje wydłużony o próbki o wartości około zerowej, które zostają dalej usunięte. Finalnie następuje rekonstrukcja sygnału zgodnie z założeniami metody *overlap-add* – poszczególne okna układane są z identycznym nakładkowaniem, jak podczas analizy FFT, a wartości nachodzących na siebie próbek są sumowane. W wyniku operacji uzyskiwany jest sygnał, w którym jedynym zmodyfikowanym parametrem pozostaje jego faza (rys. 13.3).

Algorytm uprzestrzenniający sygnał jako wtyczka programowa



Rys. 13.4. Interfejs graficzny algorytmu – wtyczka programowa VST2

Dzięki wykorzystaniu środowiska Matlab po zainstalowaniu odpowiednich dodatków (tzw. *toolboxów*) możliwe było opracowanie wersji algorytmu jako wtyczki programowej w standardzie VST2. Na podstawie zaimplementowanych funkcjonalności programu opracowano interfejs graficzny umożliwiający kontrolę kilku wybranych parametrów (rys. 13.4).

Na podstawie parametru *number of analysis bins* określa się rozdzielczość częstotliwościową analizy – im wyższa, tym dokładniej wyznaczane są częstotliwości składowych. Suwak *deviation* odpowiada za maksymalny (±) kąt zmiany fazy sygnału. Dalej suwak *harm. detect. threshold* określa próg poziomu sygnału (w dBFS), powyżej którego składowa sygnału może zostać uznana za jego harmoniczną. Następnie parametr *number of harmonics* określa, ile harmonicznych sygnału ma podlegać przetwarzaniu. Parametr *mode* odpowiada za typ wprowadzanych modyfikacji fazowych (zmiany losowe według wybranego rozkładu lub według funkcji).

13.3. Podsumowanie

W niniejszym rozdziale przedstawiono algorytm uprzestrzenniający sygnały dźwiękowe przez wprowadzanie modyfikacji fazowych, dzięki któremu możliwe jest kontrolowanie wrażenia przestrzenności sygnału niezależnie od innych jego parametrów. Zaprezento-

wany algorytm posłuży przede wszystkim za narzędzie do prowadzenia dalszych badań naukowych nad wpływem modyfikacji fazowych (różnego typu) na subiektywne wrażenia przestrzenności sygnału. W tym celu został on wyposażony w szereg parametrów umożliwiających szybkie zmiany sposobu przetwarzania sygnału. Ponadto opracowano go także w formie wtyczki programowej VST, dzięki czemu znajdzie zastosowanie w cyfrowych stacjach roboczych, tym samym stanie się możliwe sprawdzenia jego działania w praktyce.

Na obecnym etapie wszystkie założenia projektowe związane z algorytmem zostały pomyślnie zrealizowane. W zależności od toku rozwoju badań naukowych planowane jest doposażenie algorytmu w dodatkowe tryby pracy oraz parametry. Docelowo po określeniu najbardziej efektywnych sposobów działania algorytmu liczba parametrów będzie zredukowana do minimum, aby ułatwić użytkownikom jego obsługę. Ma to szczególne znaczenie w kontekście opracowywanej wtyczki programowej, która dzięki temu może stać się przydatnym narzędziem w pracy producentów muzycznych i realizatorów dźwięku.

Bibliografia

[1] Allen J.B., Rabiner L.R., *A Unified Approach to Short-Time Fourier Analysis and Synthesis*, Proceedings of the IEEE, Vol. 65, No. 11, 1977.

[2] Griesinger D., *Phase Coherence as a Measure of Acoustic Quality*, cz.1, *The Neural Mechanism*, Proceedings of 20th International Congress of Acoustics (ICA), 2010.

[3] Griesinger D., *Phase Coherence as a Measure of Acoustic Quality*, cz. 3, *Hall Design*, Proceedings of 20th International Congress of Acoustics (ICA), 2010.

[4] Griesinger D., *The importance of the direct to reverberant ratio in the perception of distance, localization, clarity and envelopment*, convention paper No. 7724 presented at 126th AES Convention in Germany, 2009.

[5] Laitinen M.-V., Disch S., Pulkki V., *Sensitivity of Human Hearing to Changes in Phase Spectrum*, „Journal of the Audio Engineering Society" 2013, Vol. 61, No. 11, s. 860–877.

[6] Makuch T., Kleczkowski P., *Wpływ modyfikacji fazy sygnału na percepcję pogłosowości w nagraniach audio*, w: *Advances in Audio Engineering and Psychoacoustics*, A. Król-Nowak (red.), Wydawnictwa AGH, Kraków 2022.

[7] Moore B.C.J., *Wprowadzenie do psychologii słyszenia*, Wydawnictwo Naukowe PWN, Warszawa–Poznań 1999.

Słowa kluczowe: przetwarzanie sygnałów, widmo fazowe, modyfikacje fazowe, słyszenie przestrzenne.

Algorytm uprzestrzenniający sygnały dźwiękowe oparty na przesunięciach fazowych

Większość obecnie stosowanych algorytmów uprzestrzenniających sygnały dźwiękowe opiera się m.in na dodawaniu pogłosu, powtórzeń sygnału, wprowadzaniu opóźnień. Zazwyczaj oferują one szereg parame-

trów odnoszących się do właściwości modelowanego procesu. Parametry te jednak słabo odnoszą się do subiektywnego wrażenia przestrzenności sygnału. Co więcej, w przeprowadzonych badaniach naukowych wskazuje się na złożoność mechanizmów odpowiedzialnych za percepcję przestrzenności. Zgodnie z nimi układ słuchowy jest wrażliwy na zależności fazowe między harmonicznymi sygnału, a to w większym stopniu przekłada się na poczucie przestrzenności.

Przedmiotem niniejszego rozdziału jest opis algorytmu uprzestrzenniającego sygnały dźwiękowe, który działa z wykorzystaniem modyfikacji fazowych. W pierwszej kolejności sygnał poddawany jest okienkowaniu, a następnie dla każdego okna wykonywana jest transformata Fouriera. W kolejnym kroku wykonuje się modyfikacje fazowe składowych harmonicznych sygnału lub w pasmach częstotliwości. W zaproponowanym algorytmie wykorzystanych zostaje kilka metod modyfikacji fazy – losowa z wykorzystaniem kilku rozkładów oraz z wykorzystaniem funkcji matematycznych, w tym funkcji okresowych. Finalnie zmodyfikowany sygnał jest przenoszony z powrotem do dziedziny czasu przez odwrotną transformację Fouriera i rekonstruowany przy użyciu metody *overlap-add*.

Sound spatialization algorithm based on phase shifting

Most of the currently used signal spatialization algorithms are based on adding reverberation, signal repetitions, delays. Usually, they offer a number of parameters corresponding to the parameters of the modelled process. However, these parameters poorly relate to the subjective impression of the signal spatiality. Scientific research indicates that much more complex mechanisms are responsible for perception of the spatiality of signals. According to them, the auditory system is sensitive to phase dependencies between the harmonics of the signal, which translates to a greater extent into the sense of spatiality.

For this reason, the subject of this paper is description of the spatialization algorithm based on phase modifications. They are performed on the phase spectrum of the windowed signal obtained using the Fourier transform. Then the phase modifications are applied to certain harmonics in a given frequency band. The proposed algorithm describes several methods of phase modification: random with the use of several distributions, and with the use of mathematical functions, including periodic functions. Then the modified signal is transferred back to the time domain by inverse Fourier transform and overlap-add reconstruction.

Indeks nazwisk autorów

W kolejnej monografii z cyklu „Postępy badań w inżynierii dźwięku i obrazu" przedstawiamy czytelnikom zagadnienia z obszaru akustyki dotyczące pomiarów, przetwarzania, klasyfikacji i oceny jakości sygnałów audio-wideo w kontekście aktualnych osiągnięć naukowo-badawczych w tym zakresie. Książka zawiera 13 obszernych rozdziałów, opracowanych przez polskich akustyków z różnych ośrodków naukowo-badawczych we współpracy z kilkoma firmami.

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 Oddziału Wrocławskiego Polskiego Towarzystwa Akustycznego.

Wydawnictwa Politechniki Wrocławskiej
są do nabycia w sprzedaży wysyłkowej:
zamawianie.ksiazek@pwr.edu.pl

ISBN 978-83-7493-258-5