



XX Międzynarodowe Sympozjum Nowości w Technice Audio i Wideo NTAV2024

Wrocław, 17-19 października 2024



Katedra Akustyki,
Multimediów
i Przetwarzania Sygnałów



Polish
Audio Engineering Society

Materiały konferencyjne

ORGANIZATORZY ▪ ORGANIZERS

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

KOMITET NAUKOWY ▪ SCIENTIFIC COMMITTEE

Przewodniczący: Krzysztof Opieliński

Andrzej Brzoska

Andrzej Czyżewski

Andrzej Dobrucki

Tadeusz Kamisiński

Piotr Kleczkowski

Bożena Kostek

Mirosław Meissner

Andrzej Miśkiewicz

Janusz Piechowicz

Adam Pilch

Anna Preis

Tomira Rogala

Ewa Skrodzka

Bogusław Szlachetko

Andrzej Wicher

Jerzy Wiciak

Sławomir Zieliński

Jan Żera

KOMITET ORGANIZACYJNY ▪ ORGANIZING COMMITTEE

Przewodniczący: Przemysław Plaskota (*edycja*)

Sekretarz: Michał Łuczyński (*redakcja*)

Skarbnik: Paweł Dziechciński

Członkowie: Bartłomiej Kruk

Maurycy J. Kin

Piotr Kozłowski

Andrzej Lewandowski

Agnieszka Wielgus

Romuald Bolejko

Zbigniew Świętach

Agnieszka Paula Pietrzak

Teksty referatów zostały opublikowane w formie przygotowanej przez Autorów

Wydawnictwo Polskie Towarzystwo Akustyczne, Oddział w Krakowie
Wrocław 2025

ISBN 978-83-61402-06-0

WSPIERAJĄ NAS



BROADCAST • STUDIO • LIVE • INSTALL

www.caudio.pl



softserve



Wydział Elektroniki,
Fotoniki i Mikrosystemów

SPIS TREŚCI

1. Możliwości wykorzystania pomiarów akustycznych emisji ultradźwiękowej roślin na potrzeby systemów wspomagania decyzji w ogrodnictwie	
Bartłomiej CHOJNACKI, Klara CHOJNACKA, Piotr KSIĄŻEK, Janusz MAZUREK, Wiktoria POTONIEC, Piotr CHOJURA.....	4
2. Dokładność lokalizacji wybranych metod używanych do odtwarzania dźwięku przestrzennego w systemach wirtualnej rzeczywistości	
Ihar BALYKA, Sławomir K. ZIELIŃSKI	6
3. Eksperymentalne źródła wszechkierunkowe na potrzeby rejestracji odpowiedzi impulsowych pomieszczeń	
Maciej SIŁKOWSKI	17
4. Nierównomierne próbkowanie przestrzenne w zagadnieniu estymacji kierunku nadejścia (DOA) fali akustycznej	
Zbigniew ŚWIĘTACH, Bogusław SZLACHETKO, Przemysław PLASKOTA, Bartłomiej KRUK, Michał ŁUCZYŃSKI, Jędrzej SZCZEPANIAK.....	26
5. Wpływ wygrzewania przetworników na parametry subiektywne i obiektywne zestawów głośnikowych	
Tomasz KOPCIŃSKI, Dominika KUCZAK, Bartłomiej KRUK, Jan KUCHARCZYK, Tomasz NOWAK	32
6. Synteza dźwięku przestrzennego z wykorzystaniem zmodyfikowanej bazy HRTF pomiarów wykonanych na Politechnice Wrocławskiej	
Zbigniew ŚWIĘTACH, Przemysław PLASKOTA.....	45

MOŻLIWOŚCI WYKORZYSTANIA POMIARÓW AKUSTYCZNYCH EMISJI ULTRADŹWIEKOWEJ ROŚLIN NA POTRZEBY SYSTEMÓW WSPOMAGANIA DECYZJI W OGRODNICTWIE

**Bartłomiej CHOJNACKI¹,
Klara CHOJNACKA¹,
Piotr KSIAŻEK²,
Janusz MAZUREK³,
Wiktoria POTONIEC¹,
Piotr CHOHURA³**

¹ Katedra Mechaniki i Wibroakustyki, Wydział Inżynierii Mechanicznej i Robotyki, Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

² Katedra Akustyki, Mutlimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska

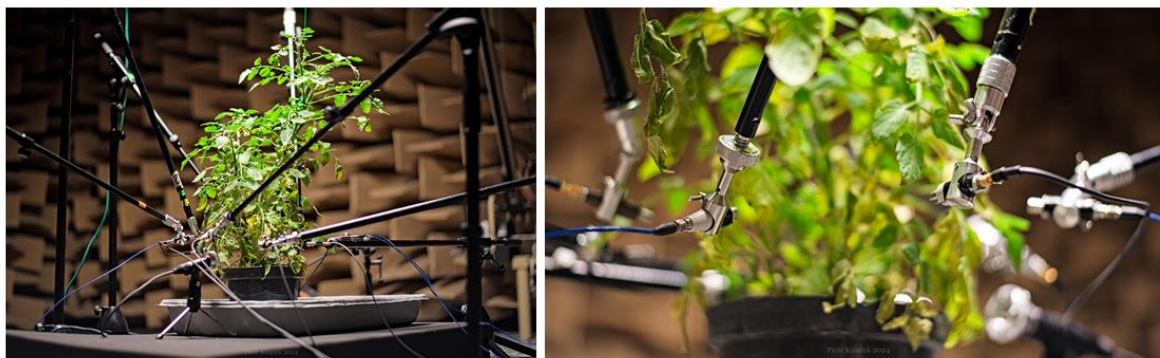
³ Katedra Ogrodnictwa, Wydział Przyrodniczo-Technologiczny, Uniwersytet Przyrodniczy we Wrocławiu

1. WPROWADZENIE

Fitoakustyka (ang. *phytoacoustic*) jest dziedziną akustyki wydzieloną już kilkadziesiąt lat temu, jednak dopiero na skutek niedawnych przełomów w tej dziedzinie zaistniały możliwości zastosowania wiedzy z tego zakresu w przemyśle ogrodniczym [1]. Większość dotychczas prowadzonych badań skupiała się na rejestracji dźwięków wydawanych przez rośliny w fazie wzrostu i rozwoju [2][3]. W wypadku stosowania pobudeń akustycznych, zazwyczaj były to tony proste [4] lub pobudzenie sygnałami muzycznymi lub szumowymi [5]. Odkrycie ultradźwiękowej emisji akustycznej (UEA) od roślin i wskazanie na jej informacyjny charakter [6] umożliwiło prowadzenie nowych badań, zorientowanych na dostarczenie metod pomiarowych na potrzeby m. in. systemów wspomaganie decyzji w ogrodnictwie. Rejestracja UEA od roślin pozwala na zdiagnozowanie stanu rośliny, m. in. w zakresie odwodnienia, ataku szkodnika lub wystąpienia szkód mechanicznych. W niniejszym referacie zaprezentowane zostanie koncepcja badań fitoakustycznych prowadzonych w Akademii Górniczo-Hutniczej w Krakowie oraz wyniki badań pilotażowych pomiarów pomidorów (*Solanum lycopersicum*) w komorze bezchowej AGH.

2. METODA BADAWCZA

W ramach badań pilotażowych UEA wykonano pomiary sadzonki pomidora w komorze bezchowej AGH w Krakowie. Wykorzystując zestaw mikrofonów GRAS 46BE 1/4" (zakres częstotliwości do 80 kHz), badano kierunkowość emisji dźwięku od sadzonki, rozstawiając mikrofony w okręgu o promieniu 30 cm. Mikrofony umieszczone były co 45 stopni na całym obwodzie okręgu. Dodatkowo, wykorzystano dwa mikrofony AviSoft CM16 o maksymalnym zakresie częstotliwościowym do 300 kHz i kartę dźwiękową AviSoft 416H o częstotliwości próbkowania 750 kHz. Fotografię stanowiska badawczego pokazano na rys. 1. Badanie polegało na pomiarze sadzonki pomidora w warunkach umożliwiających normalną vegetację rośliny, poddając ją jednak stresowi wodnemu (wstrzymano podlewanie). Roślina była doświetlana zgodnie z zaleceniami przez 16 godzin na dobę.



Rys. 1. Fotografia stanowiska badawczego w komorze bezechowej do pomiaru kierunkowości ultradźwiękowej emisji akustycznej od sadzonki pomidora

Pomiar trwał 10 dni, rejestrację prowadzono w trybie ciągłym. W uzyskanym materiale odnaleziono impulsy ultradźwiękowe występujące paśmie częstotliwości od 30 do 50 kHz, co potwierdza wstępne założenia analizy UEA roślin tego typu dostępne w literaturze. Wraz ze wzrostem stresu wodnego, wzrastała liczba impulsów rejestrowanych przez aparaturę. W wypadku rośliny zdrowej było to od 4 do 10 impulsów w ciągu godziny, dla rośliny odwodnionej (podczas 7 dnia pomiaru) rejestrowano już ponad 70 impulsów na godzinę.

3. PODSUMOWANIE

Przeprowadzone badania pilotażowe potwierdziły informacyjny i użyteczny charakter UEA pochodzącej od roślin. Dalsze badania kierunkowości dźwięku oraz badania wibracji generowanych przez sadzonkę w trakcie emisji pozwolą na szczegółowe zbadanie genezy tego zjawiska i ustalenie, które elementy rośliny odpowiadają ze UEA. W przyszłości zaplanowano opracowanie metod algorytmicznych pozwalających na precyzyjną diagnostykę stanu rośliny i wykrywanie różnych zagrożeń na podstawie badania ultradźwiękowych emisji akustycznych roślin. Badania tego typu mogą posłużyć do stworzenia systemów wspomagania decyzji dla ogrodników, znacznie przyczyniając się do rozwoju nowych technologii w ogrodnictwie.

LITERATURA

- [1] Ali S, Tyagi A, Park S, Bae H, Understanding the mechanobiology of phytoacoustics through molecular Lens: Mechanisms and future perspectives, *J. Adv. Res.*, [Internet] 2023, . Available from: <https://www.sciencedirect.com/science/article/pii/S2090123223003983>
- [2] Son J-S, Jang S, Mathevon N, Ryu C-M, Is plant acoustic communication fact or fiction?, *New Phytol.*, [Internet] 242 (3)5, 2024, 1876–80,. Available from: <https://doi.org/10.1111/nph.19648>
- [3] Allievi S, Arru L, Forti L, A tuning point in plant acoustics investigation, *Plant Signal. Behav.*, [Internet] 16 (3)8, 2021, . Available from: <https://doi.org/10.1080/15592324.2021.1919836>
- [4] Jeong MJ, Cho J Il, Park SH, Kim KH, Lee SK, Kwon TR, et al., Sound frequencies induce drought tolerance in rice plant, *Pakistan J. Bot.*, 46 (3)6, 2014, 2015–20,.
- [5] Shivanna KR, Phytoacoustics - Plants can perceive ambient sound and respond, *J. Indian Bot. Soc.*, 102 (3)1, 2022, 1–5,.
- [6] Khait I, Lewin-Epstein O, Sharon R, Saban K, Goldstein R, Anikster Y, et al., Sounds emitted by plants under stress are airborne and informative, *Cell*, [Internet] 186 (3)7, 2023, 1328-1336.e10,. Available from: <https://doi.org/10.1016/j.cell.2023.03.009>

DOKŁADNOŚĆ LOKALIZACJI WYBRANYCH METOD UŻYWANYCH DO ODTWARZANIA DŹWIĘKU PRZESTRZENNEGO W SYSTEMACH WIRTUALNEJ RZECZYWISTOŚCI

Ihar BALYKA¹

Sławomir K. ZIELIŃSKI²

¹ Politechnika Białostocka, Wiejska 45A, 15-351 Białystok
ihar.balyka@sd.pb.edu.pl

² Politechnika Białostocka, Wiejska 45A, 15-351 Białystok
s.zielinski@pb.edu.pl

Technologie dźwięku przestrzennego stosowane w systemach rzeczywistości wirtualnej zostały znacząco udoskonalone w ciągu ostatnich kilku dekad, jednak nawet najbardziej zaawansowane techniki wciąż wykazują pewne niedoskonałości w zakresie dokładności lokalizacji i wierności brzmieniowej. Celem niniejszego badania było ocenienie i porównanie dokładności lokalizacji oraz wierności brzmieniowej uzyskiwanej przy użyciu pięciu popularnych technik reprodukcji dźwięku immersyjnego w nowoczesnych systemach rzeczywistości wirtualnej. Przeprowadzone badania miały charakter subiektywnych testów odsłuchowych. Wyniki pokazały, że metoda charakteryzująca się najwyższą dokładnością lokalizacji, jednocześnie cechowała się najniższym stopniem wierności brzmienia dźwięku. Uzyskane wyniki wskazują na potrzebę holistycznej optymalizacji jakości metod stosowanych do odtwarzania dźwięku przestrzennego zarówno pod względem cech przestrzennych jak również brzmieniowych.

1. WPROWADZENIE

Technologie binauralne są obecnie powszechnie stosowane do odtwarzania dźwięku przestrzennego w systemach rzeczywistości wirtualnej, znacząco zwiększając immersję w symulatorach, grach komputerowych oraz aplikacjach służących do oglądania filmów i słuchania muzyki. Chociaż technologie te zostały znacznie udoskonalone w ciągu ostatnich kilku dekad, nawet najnowocześniejsze techniki nadal wykazują pewne niedociągnięcia w zakresie uzyskiwanej dokładności lokalizacji lub wierności brzmieniowej dźwięku [14]. Stąd też, wciąż istnieje potrzeba ich wiarygodnej oceny oraz dalszego udoskonalenia.

W świetle badań dotyczących subiektywnej oceny jakości dźwięku, techniki służące do odtwarzania dźwięku przestrzennego powinny być optymalizowane na dwóch płaszczyznach, mianowicie pod kątem atrybutów dotyczących brzmienia dźwięku jak również pod względem cech przestrzennych [12]. Przykładowo, zgodnie z badaniami przeprowadzonymi przez Rumseya *in*. [18], na ogólną jakość dźwięku zestawów kina domowego wpływa głównie wierność brzmienia dźwięku oraz do pewnego stopnia wierność przestrzenna. Jednoczesna optymalizacja atrybutów brzmieniowych i przestrzennych dźwięku jest jednak zadaniem bardzo trudnym. Na przykład, podczas rejestracji funkcji przenoszenia głowy (ang. *head-related transfer functions*), które są powszechnie wykorzystywane w metodach odtwarzania dźwięku przestrzennego w technice binauralnej, badacze mierzą się z wyzwaniem polegającym na uzyskaniu odpowiedniego kompromisu technicznego pomiędzy dokładnością przestrzenną pomiarów, stosunkiem sygnału do szumu, zniekształceniami nieliniowymi oraz szerokością pasma częstotliwościowego. Z racji ograniczeń fizycznych, w pomiarach tych nadmierna optymalizacja dokładności przestrzennej może wpłynąć negatywnie na parametry rzutujące na wierność brzmienia dźwięku [20].

Celem niniejszej pracy było porównanie pięciu wybranych metod używanych do odtwarzania

dźwięku przestrzennego w systemach wirtualnej rzeczywistości. Na potrzeby badań zastosowano dwa kryteria, mianowicie uzyskiwaną dokładność lokalizacji oraz wierność brzmienia dźwięku. W odróżnieniu od podobnych eksperymentów opisanych w literaturze [11, 15–17], w których wykorzystano wyłącznie statyczny tryb odtwarzania dźwięku, w niniejszych badaniach zastosowano dynamiczną metodę odtwarzania dźwięku binauralnego, uwzględniającą informację o rotacji głowy słuchacza.

Oryginalny wkład niniejszej pracy polega na kompleksowej ocenie pięciu wybranych metod używanych do odtwarzania dźwięku przestrzennego w systemach wirtualnej rzeczywistości opierającej się na dwóch kluczowych aspektach, które nie były wcześniej porównywane. Praca ta dostarcza nowych informacji na temat współcześnie stosowanych technik odtwarzania dźwięku binauralnego w systemach wirtualnej rzeczywistości.

2. POWIĄZANE BADANIA

Wcześniejsze badania opisane w literaturze dotyczyły porównań metod odtwarzania dźwięku przestrzennego w technologii binauralnej z punktu widzenia innych kryteriów niż te, które zostały założone w niniejszej pracy. Na przykład, w pracy autorstwa Catalano [6] porównano metody FMOD oraz Wwise pod względem takich cech jak międzyuszne różnice czasu oraz intensywności. Zgodnie z uzyskanymi wynikami, techniki te okazały się niewiarygodne pod względem reprezentacji międzyusznych różnic czasu. Mankamentem wspomnianych badań był brak porównań percepcyjnych.

Kompleksowe badania metod binauralnego odtwarzania dźwięku przestrzennego z uwzględnieniem atrybutów percepcyjnych zostały opisane w roku 2018 w serii trzech publikacji autorstwa Reardona *i in.* [15–17]. W swoich pracach porównali oni sześć komercyjnie dostępnych technik odtwarzania dźwięku binauralnego pod względem zarówno cech brzmieniowych jak i przestrzennych. Przebadane metody znacznie różniły się percepcyjnie, jednak uzyskane wyniki były mocno uzależnione od rodzaju odtwarzanego sygnału. Uzyskiwana dokładność lokalizacji badanych technik okazała się stosunkowo słaba, szczególnie w przypadku metod wykorzystujących ambisonię pierwszego rzędu. Ich badania zostały kontynuowane przez LaFontaine *i in.* [11]. W roku 2023 wykazali oni, podobnie jak poprzedni badacze, że dokładność lokalizacji w płaszczyźnie poziomej uzyskiwana z użyciem najnowszej generacji technik odtwarzania dźwięku binauralnego wciąż wymaga poprawy. Należy zaznaczyć jednak, iż zarówno badania przeprowadzone przez Raerdona *i in.* [15–17] jak również LaFontaine *i in.* [11] zostały wykonane w trybie statycznego odtwarzania dźwięku binauralnego, z pominięciem ruchów głowy. Zgodnie z literaturą, statyczny sposób odtwarzania dźwięku binauralnego prowadzi do częstych błędów lokalizacji pomiędzy źródłami umieszczonymi z przodu oraz z tyłu słuchacza [5, 10]. Dlatego też w niniejszych badaniach zastosowano dynamiczny tryb odtwarzania dźwięku, uwzględniający informację o rotacji głowy słuchacza.

3. METODA BADAWCZA

Na potrzeby badań wyselekcjonowano pięć powszechnie używanych technik odtwarzania dźwięku binauralnego w aplikacjach wirtualnej rzeczywistości, mianowicie FMOD [8], Unreal Engine [7], Wwise [3], Steam Audio [21] oraz Resonance [9]. Zostały one uszeregowane w tabeli 1 chronologicznie zgodnie z rokiem ich upublicznienia. Należy zauważyć, iż dwie z nich wykorzystują ambisonię trzeciego rzędu, co zgodnie z badaniami Raerdona *i in.* [15–17] powinno zapewnić słuchaczom wyższą precyzję lokalizację źródeł wirtualnych dźwięku. W przypadku metody Steam Audio deweloperzy zastosowali bazę funkcji przenoszenia głowy o akronimie CIPIC [1]. Z kolei w metodzie Resonance Audio wykorzystano bazę SADIE [2]. W przypadku pozostałych porównywanych metod (FMOD, Unreal Engine, Wwise) rodzaj zastosowanych baz funkcji przenoszenia głowy jest nieznanymi.

Dokładny sposób działania porównywanych metod nie jest publicznie ujawniany przez ich deweloperów, jednak podwaliny syntezy binauralnej dźwięku zostały już sformalizowane w drugiej połowie ubiegłego stulecia [4]. Stereofoniczny sygnał binauralny uzyskiwany jest w procesie splotu sygnału monofonicznego z odpowiedziami impulsowymi ludzkich uszu oraz głowy (ang. *head-related impulse response*), co opisuje poniższe równanie:

$$y_c[n] = \sum_{k=0}^{K-1} h_{c,\theta}[n-k] \times x[k], \quad (1)$$

gdzie $y_c[n]$ oznacza wyjściowy sygnał binauralny dla kanału c (lewy oraz prawy) oraz dla n -tej próbki, $h_{c,\theta}[n-k]$ jest odpowiedzią impulsową ucha oraz głowy dla kanału c oraz azymutu θ , natomiast $x[k]$ reprezentuje k -tą próbkę monofonicznego sygnału wejściowego. Liczba K stanowi liczbę próbek w nagraniu wejściowym. Wzór (1) jest zawężony wyłącznie do syntezy nagrań binauralnych w płaszczyźnie poziomej, jednak może być uogólniony poprzez uwzględnienie kątów azymutalnych. W przypadku użycia na wejściu sygnałów ambisonicznych, wspomniane algorytmy dokonują splotu sygnałów wirtualnych głośników z odpowiedziami impulsowymi uszu oraz głowy [13].

Tabela 1. Porównywane metody odtwarzania dźwięku binauralnego.

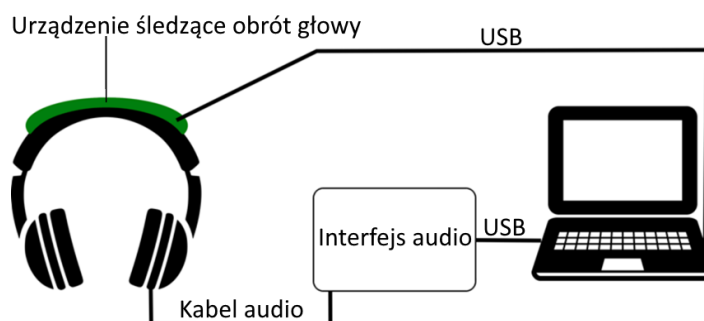
Metoda	Rok upublicznienia	Natywne wsparcie formatu ambisonicznego	Rozwój jest kontynuowany	Baza danych funkcji przenoszenia głowy
FMOD	1995	nie	tak	nieznana
Unreal Engine	1998	tak – rząd 1	tak	nieznana
Wwise	2006	nie	tak	nieznana
Steam Audio	2017	tak – rząd 3	tak	CIPIE (2001)
Resonance Audio	2018	tak – rząd 3	nie (od 2022)	SADIE (2018)

Badania miały charakter odsłuchowych testów subiektywnych. Zostały one przeprowadzone w laboratorium dźwięku przestrzennego Politechniki Białostockiej o poziomie tła akustycznego wynoszącym 22 NR. Na potrzeby badań przygotowano cztery stanowiska badawcze wyposażone w laptopy, interfejsy audio, słuchawki oraz urządzenia do śledzenia rotacji głowy (ang. *head-trackers*). Diagram pojedynczego stanowiska odsłuchowego pokazano na rys. 1. Jako bodziec akustyczny wykorzystano 60-sekundowe monofoniczne nagranie muzyki pop o stosunkowo szerokim widmie częstotliwościowym oraz wyraźnych impulsowych dźwiękach perkusyjnych. Częstotliwość próbkowania wynosiła 44,1 kHz. Jak wspomniano powyżej, w niniejszych badaniach zastosowano dynamiczny tryb odtwarzania dźwięku binauralnego, uwzględniający informację o chwilowej rotacji głowy słuchacza. W tym celu zastosowano urządzenie do śledzenia ruchów głowy, przytwierdzone do słuchawek, o niepewności pomiaru mniejszej niż 1° , czasie odświeżania wynoszącym 100 Hz oraz opóźnieniu nieprzekraczającym 10 ms.

W badaniach przeprowadzono trzy eksperymenty. Celem pierwszego z nich było porównanie statycznego oraz dynamicznego trybu odtwarzania dźwięku wirtualnego. Dodatkowo, w eksperymencie tym zweryfikowano poprawność działania systemu adaptacyjnego odtwarzania dźwięku z urządzeniem do śledzenia ruchów głowy. W drugim eksperymencie porównano techniki odtwarzania dźwięku binauralnego pod kątem dokładności lokalizacji, natomiast celem trzeciego eksperymentu było porównanie metod pod względem wierności brzmienia dźwięku. Badania lokalizacji dźwięku zostały zawężone wyłącznie do płaszczyzny poziomej.

Na potrzeby badań zrekrutowano słuchaczy spośród studentów Wydziału Informatyki Politechniki Białostockiej. W zależności od eksperymentu, liczba słuchaczy biorących udział w badaniach wahała

się od 14 do 45. W pierwszych dwóch eksperymentach wykorzystano interfejs Lexicon Audio wyposażony w 24-bitowe przetworniki cyfrowo-analogowe oraz słuchawki Sennheiser HD 215. Wspomniane słuchawki posiadają impedancję wynoszącą 32 Ω , pasmo przenoszenia 12 Hz – 22 kHz, czułość 110 dB oraz konstrukcję otaczającą ucho, z całkowitymi zniekształceniami harmonicznymi poniżej 0,2 %. W trzecim eksperymencie zastosowano interfejs Focusrite Scarlett 2i2 oraz słuchawki Beyerdynamic DT 770 Pro o impedancji 250 Ω oraz pasmie przenoszenia od 5 Hz do 35 kHz. Różnice metodologiczne pomiędzy wspomnianymi trzema eksperymentami zestawiono w tabeli 2.



Rys. 1. Diagram ilustrujący stanowisko badawcze

Tabela 2. Różnice w liczbie słuchaczy oraz aparaturze wykorzystanej w przeprowadzonych eksperymentach.

Ek-sperymet	Typ badania	Liczba uczest-ników	Interfejs Audio	Model słuchawek
1	Porównanie dynamicznego i statycznego trybu odtwarzania dźwięku	14	Lexicon Audio	Sennheiser HD 215
2	Dokładność lokalizacji	15	Lexicon Audio	Sennheiser HD 215
3	Wierność brzmienia	45	Focusrite Scarlett 2i2	Beyerdynamic DT 770 Pro

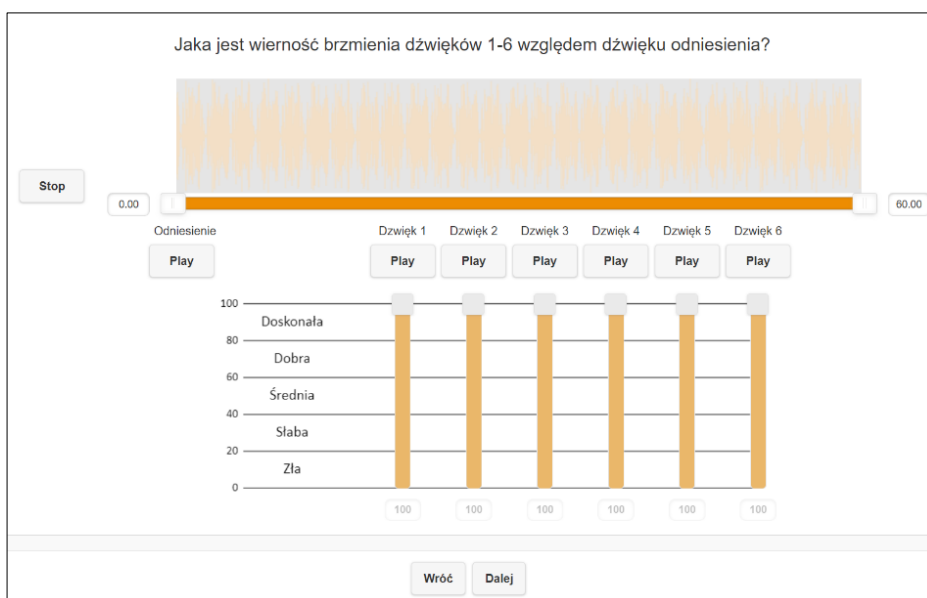
Na potrzeby pierwszych dwóch eksperymentów opracowano aplikację w środowisku Unreal Engine [7]. Aplikacja ta pozwalała słuchaczom na lokalizację wirtualnych źródeł dźwięku. W wirtualnym środowisku użytkownik znajdował się w centrum, będąc otoczony przez 16 kul rozmieszczonych co 22,5°, co zilustrowano na rys. 2. Każda kula była ponumerowana, a głowa z niebieską strzałką wskazywała kierunek spojrzenia słuchacza. Podczas testów, źródło dźwięku było losowo „umieszczane” w jednej z kul. Zadaniem uczestnika badań było wskazanie za pomocą myszy komputerowej kuli, z której dobiegał dźwięk. Gdy użytkownik kliknął kulę, system rejestrował lokalizację kuli oraz wybór użytkownika. Następnie źródło dźwięku „umieszczane” było w nowej lokalizacji. Proces ten powtarzał się, aż wszystkie metody odtwarzania dźwięku zostały przetestowane.

W trzecim eksperymencie porównano wierność brzmieniową wybranych technik z użyciem metody porównania wielu bodźców. W tym celu zaadaptowano publicznie dostępne oprogramowanie WebMUSHRA [19]. Rys. 3 prezentuje interfejs użytkownika. Każdy ze słuchaczy był poproszony, aby odpowiedzieć na następujące pytanie: „Jaka jest wierność brzmienia dźwięków 1–6 względem dźwięku odniesienia?” Odpowiedzi rejestrowane były przez słuchaczy z użyciem suwaków oraz 100-stopniowej skali z pięcioma etykietami słownymi, zilustrowanymi na rys. 3. Uczestnicy badań byli poinstruowani,

iż co najmniej jeden z ewaluowanych dźwięków powinien być oceniony z wartością maksymalną, gdyż dźwięk wzorcowy został losowo dołączony do porównywanych dźwięków. W odróżnieniu od pierwszych dwóch eksperymentów, każdy z porównywanych dźwięków prezentowany był tylko z jednego kierunku, mianowicie z wirtualnego źródła umieszczonego z przodu słuchacza.



Rys. 2. Interfejs graficzny aplikacji do badań dokładności lokalizacji źródeł dźwięku



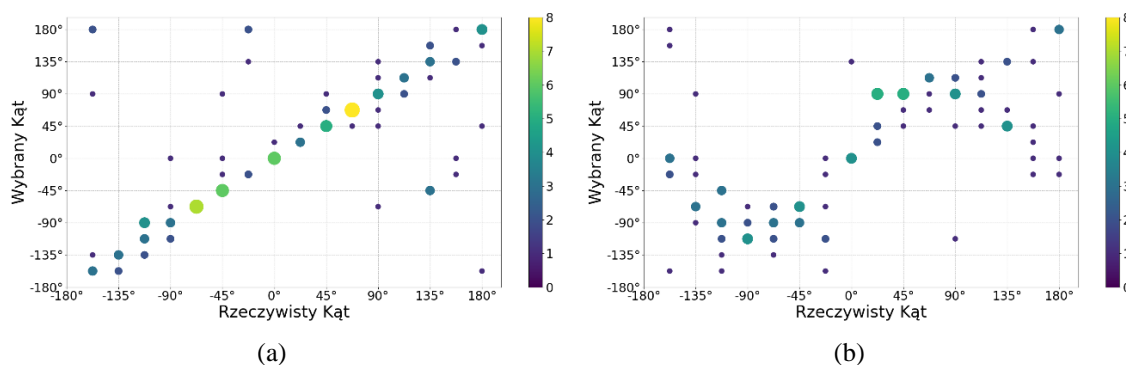
Rys. 3. Interfejs wykorzystany do badania wierności brzmienia dźwięku opracowany w oparciu o oprogramowanie WebMUSHRA [19]

4. WYNIKI

W niniejszym rozdziale zestawiono wyniki uzyskane kolejno w trzech eksperymentach. Wpierw pokazano rezultaty porównania dynamicznej oraz statycznej syntezy dźwięku binauralnego. Następnie przedstawiono wyniki porównania metod odtwarzania dźwięku binauralnego pod kątem dokładności lokalizacji. W ostatnim podrozdziale zilustrowano wyniki porównania wspomnianych metod pod względem wierności brzmieniowej.

4.1 PORÓWNANIE DYNAMICZNEJ ORAZ STATYCZNEJ SYNTEZY DŹWIĘKU BINAURALNEGO

Wyniki uzyskane w pierwszym eksperymencie zilustrowane są na rys. 4. Na rysunku tym pokazano wykresy punktowe lokalizacji w zależności od trybu syntezy dźwięku binauralnego. W przypadku zastosowania dynamicznej techniki odtwarzania dźwięku binauralnego z użyciem urządzenia do śledzenia obrotu głowy, wartości kątów wybranych przez słuchaczy w większości przypadków odpowiadają wartościom kątów zamierzonych w procesie syntezy dźwięku binauralnego, co zilustrowano na rys. 4a. Z kolei w sytuacji statycznej syntezy dźwięku binauralnego zauważyć można liczne pomyłki lokalizacji źródeł dźwięku (rys. 4b). Uzyskane wyniki są zgodne z literaturą [5, 10], z której dobitnie wynika, iż zastosowanie dynamicznej syntezy dźwięku binauralnego skutkuje znaczącą redukcją błędów lokalizacji. Ponadto, otrzymane wyniki stanowią dowód, iż stanowisko badawcze, wyposażone w moduł śledzenia obrotów głowy słuchacza oraz adaptacyjny algorytm syntezy dźwięku binauralnego, zostało opracowane poprawnie.



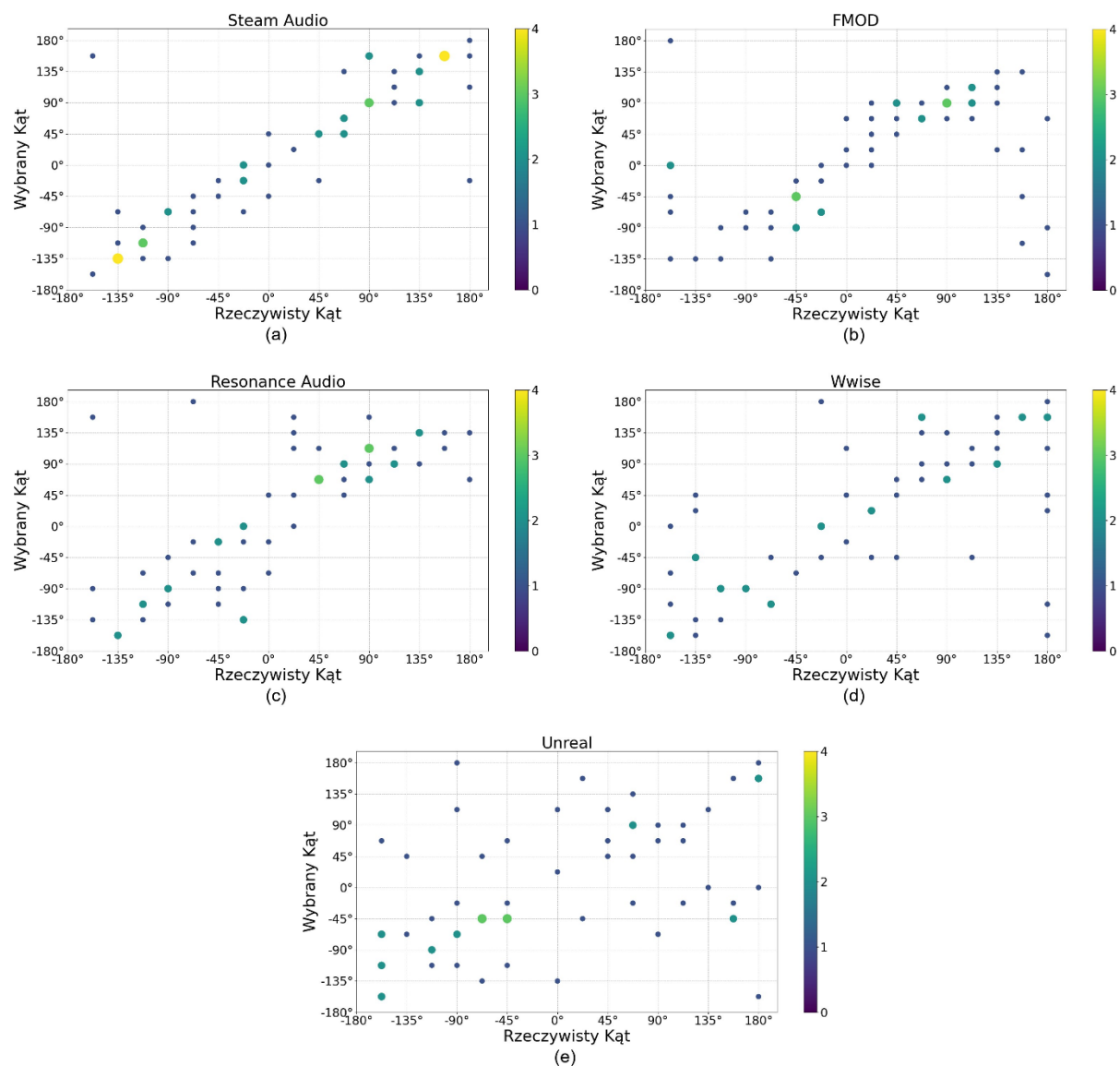
Rys. 4. Wykresy punktowe dokładności lokalizacji w zależności od trybu syntezy dźwięku binauralnego: (a) tryb dynamiczny, (b) tryb statyczny. Kolorem oznaczono liczbę ocen lokalizacji.

4.2 OCENA DOKŁADNOŚCI LOKALIZACJI

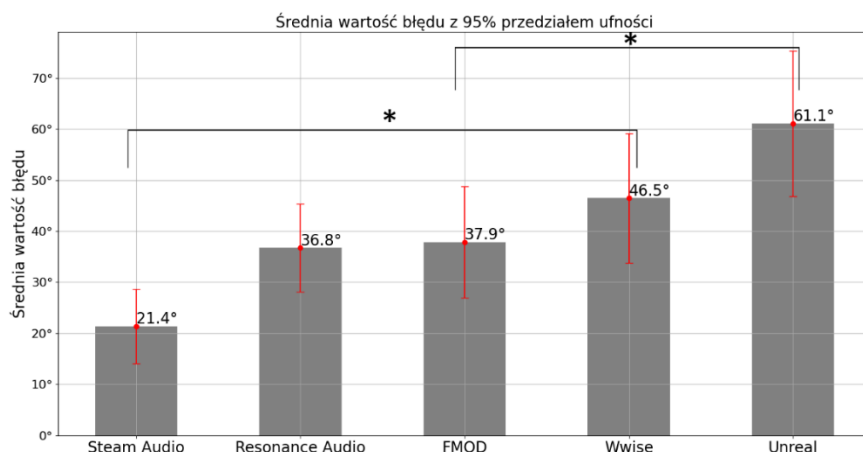
Na rys. 5 przedstawiono wykresy punktowe ilustrujące dokładność lokalizacji z użyciem pięciu wybranych technik odtwarzania dźwięku binauralnego. Należy podkreślić, iż niniejsze badanie zrealizowano wyłącznie w warunkach dynamicznej syntezy dźwięku binauralnego. Porównując uzyskane wyniki, zauważyć należy, iż pierwsza z badanych metod, mianowicie Steam Audio, zapewnia stosunkowo zadowalającą precyzję lokalizacji dźwięku (rys. 5a). Pozostałe metody charakteryzują się dość dużym rozrzutem punktów w znacznej odległości od przekątnej $y = x$, co wskazuje na ich duży błąd lokalizacji.

Chcąc porównać ilościowo różnice w precyzji lokalizacji pomiędzy badanymi technikami, dla każdej z nich obliczono średnią wartość błędu lokalizacji. Uzyskane wyniki zilustrowano na rys. 6. W świetle analizy wariancji otrzymane wyniki różnią się statystycznie ($F(4, 295) = 6,7, p = 3,79 \times 10^{-5}$). Na rys. 6 zaznaczono klamrami z gwiazdką najbliższe sąsiednie przypadki, które różniły się statystycznie zgodnie z testem porównań wielokrotnych Tukeya (*post hoc*). Należy zauważyć, iż najmniejszy błąd lokalizacji, wynoszący średnio $21,4^\circ$, uzyskano dla metody Steam Audio. Z kolei największą średnią wartość błędu otrzymano dla metody Unreal. Uzyskane wyniki są do pewnego stopnia zgodne z oczekiwaniami, biorąc

pod uwagę, iż metoda Steam Audio wykorzystuje ambisonię trzeciego rzędu. Pewnym zaskoczeniem jednak jest to, iż metoda Resonance Audio, również wykorzystująca ambisonię trzeciego rzędu, nie uzyskała tak małej wartości błędu lokalizacji jak metoda Steam Audio.



Rys. 5. Wykresy punktowe dokładności lokalizacji w zależności od metody syntezy dźwięku binauralnego: (a) Steam Audio, (b) FMOD, (c) Resonance Audio, (d) Wwise, (e) Unreal Audio. Kolorem oznaczono liczbę ocen lokalizacji.

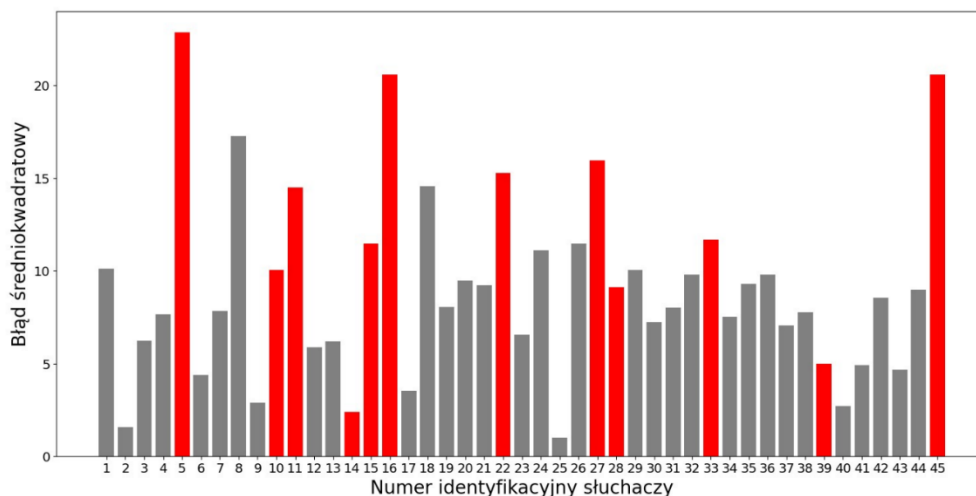


Rys. 6. Średnia wartość błędu lokalizacji w płaszczyźnie poziomej uzyskiwana dla wybranych technik odtwarzania dźwięku przestrzennego

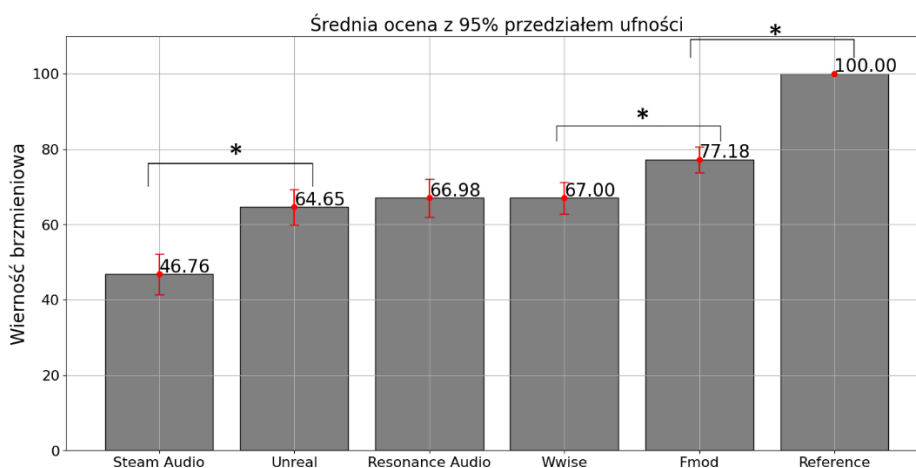
4.3 OCENA WIERNOŚCI BRZMIENIOWEJ

W badaniach uzyskiwanej wierności brzmieniowej wzięło udział łącznie 45 uczestników. Każdy z nich oceniał wszystkie nagrania dwukrotnie, co pozwoliło scharakteryzować stopień powtarzalności ich ocen. Wpierw jednak dla każdego z uczestników zbadano ich zdolność dyskryminacji wykorzystując statystykę F z analizy wariancji obliczonej oddzielnie dla poszczególnych słuchaczy. Biorąc pod uwagę, że wartość F była większa od 0 dla każdego uczestnika, uznano wszystkich słuchaczy jako posiadających wystarczającą zdolność dyskryminacji wierności brzmieniowej. Następnie obliczono stopień powtarzalności dla każdego z uczestników, analizując średni błąd kwadratowy (ang. *Root Mean Square Error* – *RMSE*) ocen uzyskanych dla tych samych nagrań. Otrzymane wyniki zilustrowano na rys. 7. Trzej słuchacze o numerach 5, 16 oraz 45 charakteryzowali się dość dużym błędem średniokwadratowym, wynoszącym ponad 15 punktów względem 100-punktowej skali ocen. Ponadto, dwunastu spośród 45 uczestników, łącznie z wyżej wymienionymi trzema słuchaczami, oceniło wierność brzmieniową oryginalnego dźwięku na mniej niż 100 punktów, a zatem zostali oni uznani za niewiarygodnych. Słuchaczy tych oznaczono kolorem czerwonym na rys. 7. Dane uzyskane od niewiarygodnych słuchaczy zostały wykluczone z dalszej analizy.

Średnie wartości wierności brzmieniowej uzyskanej dla porównywanych metod odtwarzania dźwięku binauralnego zostały przedstawione w kolejności od najmniejszej do największej na rys. 8. Zgodnie z analizą wariancji, różnice między porównywanymi metodami były statystycznie istotne ($F(5, 390) = 70,1, p = 3,24 \times 10^{-52}$). Na rys. 8 klamrami z gwiazdką wyróżniono najbliższe „sąsiednie” metody, które różniły się statystycznie zgodnie z testem porównań wielokrotnych Tukeya (*post hoc*). Zgodnie z wynikami, najgorszą wierność brzmieniową uzyskała metoda Steam Audio. Wynik ten był zaskakujący z tego względu, iż metoda ta zapewniała najlepszą dokładność lokalizacji w poprzednim eksperymencie. Najlepszą wierność brzmieniową uzyskano przy użyciu metody FMOD (pomijając nagranie referencyjne, dla którego, zgodnie z oczekiwaniem, uzyskano maksymalną, 100-punktową wierność brzmieniową).



Rys. 7. Błąd średniokwadratowy (RMSE) ocen dla poszczególnych słuchaczy. Kolorem czerwonym zaznaczono słuchaczy, którzy nie potrafili zidentyfikować ukrytego nagrania wzorcowego.



Rys. 8. Średnia wartość wierności brzmieniowej porównywanych metod odtwarzania dźwięku binauralnego

5. DYSKUSJA

Wyniki niniejszego badania podkreślają znaczne postępy i bieżące wyzwania w technologiach odtwarzania dźwięku przestrzennego w systemach wirtualnej rzeczywistości. Analiza pięciu popularnych technik – Unreal Engine Audio, FMOD, Wwise, Steam Audio i Resonance Audio – ujawniła znaczące różnice w dokładności lokalizacji i wierności brzmieniowej, wskazując na znaczenie dalszych badań i rozwoju w tej dziedzinie. Metoda Steam Audio okazała się najdokładniejszą techniką pod względem lokalizacji, co demonstruje jej potencjalną skuteczność w takich zastosowaniach dźwięku przestrzennego, jak gry komputerowe lub symulatory, chociaż uzyskała najgorsze wyniki w zakresie wierności brzmieniowej.

Obserwowane różnice w wydajności mogą być przypisane odmiennym algorytmom i podejściom, jakie każda technologia stosuje w przetwarzaniu dźwięku. Na przykład, zdolność metody Steam Audio

do wykorzystania ambisonii rzędu trzeciego może tłumaczyć jej uzyskiwaną lepszą dokładność lokalizacji źródła dźwięku. Jednak warto zauważyć, że metoda Steam Audio opiera się na stosunkowo starej bazie danych funkcji przenoszenia głowy CIPIC z 2001 roku [1], co może być jedną z przyczyn jej niskiej wierności brzmienia. Powyższe różnice podkreślają konieczność, aby deweloperzy systemów wirtualnej rzeczywistości brali pod uwagę nie tylko możliwości techniczne metody używanej do odtwarzania dźwięku binauralnego, ale także specyficzne wymagania swoich aplikacji oraz kontekst ich zastosowań.

Wyniki badania potwierdziły kluczową rolę śledzenia obrotów głowy oraz stosowania dynamicznej metody odtwarzania dźwięku binauralnego. Zaobserwowana poprawa lokalizacji wirtualnych źródeł dźwięku, będąca skutkiem zastosowania trybu dynamicznego odtwarzania dźwięku, jest zgodna z wcześniejszymi badaniami w tym zakresie [5, 10]. Wzrost liczby poprawnych odpowiedzi dodatkowo potwierdza konieczność uwzględnienia śledzenia głowy w aplikacjach audio w wirtualnej rzeczywistości.

6. WNIOSKI

Nowoczesne technologie binauralnego odtwarzania dźwięku wykazują wyraźne różnice pod względem dokładności lokalizacji. Metoda charakteryzująca się najwyższą dokładnością lokalizacji, jednocześnie cechowała się najniższym stopniem wierności brzmienia dźwięku, co wskazuje na potrzebę jednoczesnej optymalizacji oraz holistycznej ewaluacji wspomnianych technologii zarówno pod względem cech przestrzennych jak również brzmieniowych. Przeprowadzone eksperymenty dodatkowo potwierdziły, że zastosowanie dynamicznej techniki odtwarzania dźwięku binauralnego, z uwzględnieniem śledzenia obrotów głowy, jest niezbędne w systemach audio, ponieważ pomaga znacząco zmniejszyć średni błąd lokalizacji.

PODZIĘKOWANIA

Autorzy artykułu wyrażają swoją wdzięczność uczestnikom testów odsłuchowych. Przeprowadzone badania uzyskały pozytywną opinię Komisji Etyki Politechniki Białostockiej (protokół nr. 1/2024 z dnia 14.06. 2024 r.). Badania zrealizowano w ramach grantu Politechniki Białostockiej WZ/WI-IIT/5/2023 finansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego.

LITERATURA

- [1] ALGASI V.R., DUDA R.O., THOMPSON D.M., AVENDANO C., *The CIPIC HRTF Database*. In: *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, Mohonk Mountain House, New Paltz, NY, USA, 2001. doi:10.1109/ASPAA.2001.969552.
- [2] ARMSTRONG C., THRESH L., MURPHY D., KEARNEY G., *A perceptual evaluation of individual and non-individual HRTFs: a case study of the SADIE II database*, *Appl. Sci.*, 8(2029), 2018. doi:10.3390/app8112029.
- [3] AUDIOKINETIC INC., *Wwise Documentation*, 2024, <https://www.audiokinetic.com/en/library/edge/>. Dostęp: 23 Wrz. 2024.
- [4] BEGAULT D.R., *3D Sound for Virtual Reality and Multimedia*, Ames Research Center NASA, California, USA (First Edition), 1994.
- [5] BEGAULT D.R., WENZEL E.M., ANDERSON M.R., *Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source*, *J. Audio Eng. Soc.*, 49(10), 2001, pp. 904–916.
- [6] CATALANO G., *Virtual Reality in Interactive Environments: A Comparative Analysis of Spatial Audio Engines*, Oxford, S.A.E. Institute, 2011.
- [7] EPIC GAMES, *Unreal Engine Documentation*, 2024, <https://docs.unrealengine.com/>. Dostęp: 23 Wrz. 2024.
- [8] FIRELIGHT TECHNOLOGIES PTY LTD., *FMOD Engine User Manual 2.03.03 (2024-09-23)*, 2024, <https://www.fmod.com/docs/api>. Dostęp: 23 Wrz. 2024.

- [9] GOOGLE LLC, *Discover Resonance Audio*, <https://resonance-audio.github.io/resonance-audio/discover/overview>. Dostęp: 23 Wrz. 2024.
- [10] HORBACH U., KARAMUSTAFAOGLU A., PELLEGRINI R.S., MACKENSEN P., THEILE G., *Design and Applications of a Data-Based Auralization System for Surround Sound*, In: *Proc. of the 106th Audio Engineering Society Convention*, Paper 4976, Munich, Germany, 1999.
- [11] LAFONTAINE L., PLASSE R., BULLA W., *Binaural renderers accuracy comparison*, In: *Proc. of the 155th Audio Engineering Society Convention*, Paper 173, New York, USA, 2023.
- [12] LINDAU A., *A spatial audio quality inventory for virtual acoustic environments (SAQI)*, In: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, 3-5 April, 2014.
- [13] PATERSON J., LEE H., *3D Audio*, London, Routledge, 2022.
- [14] POURU L., *The Parameters of Realistic Spatial Audio: An Experiment with Directivity and Immersion*, Turku, Bachelor's Thesis, Turku University of Applied Sciences, 2019.
- [15] REARDON G., GENOVESE A.F., ZALLES G., FLANAGAN P., ROGIŃSKA A., *Evaluation of Binaural Renderers: Localization*. In: *Proc. of the 144th Audio Engineering Society Convention*, Paper 9963, Milan, Italy, 2018.
- [16] REARDON G., GENOVESE A.F., ZALLES G., FLANAGAN P., ROGIŃSKA A., *Evaluation of Binaural Renderers: Multidimensional Sound Quality Assessment*. In: *Proc. of the Audio Engineering Society Conference on Audio for Virtual and Augmented Reality*, Redmond, USA, 2018.
- [17] REARDON G., ZALLES G., GENOVESE A., FLANAGAN P., ROGIŃSKA A., *Evaluation of Binaural Renderers: Externalization, Front/Back and Up/Down Confusions*. In: *Proc. of the 144th Audio Engineering Society Convention*, Paper 9989, Milan, Italy, 2018.
- [18] RUMSEY F., ZIELIŃSKI S., KASSIER R., BECH S., *On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality*, *J. Acoust. Soc. Am.*, 118(2), 2005. pp. 968–976, doi: 10.1121/1.1945368
- [19] SCHOEFFLER M., BARTOSCHEK S., STÖTER F.R., ROESS M., WESTPHAL S., EDLER B., HERRE J., *WebMUSHRA – A Comprehensive Framework for Web-based Listening Tests*, *Journal of Open Research Software*, 6(1): 8, 2018. doi:10.5334/jors.187
- [20] SILZLE A., *Quality of Head-Related Transfer Functions – Some Practical Remarks*, In: *Proc. of the 1st ISCA Tutorial & Research Workshop on Auditory Quality of Systems*, Akademie Mount Cenis, Germany, 2003.
- [21] VALVE CORPORATION, *Steam Audio Unreal Engine Integration*, <https://valvesoftware.github.io/steam-audio/doc/unreal/index.html>. Dostęp: 23 Wrz. 2024.

LOCALIZATION ACCURACY OF THE SELECTED METHODS USED FOR SPATIAL AUDIO RENDERING IN VIRTUAL-REALITY SYSTEMS

Spatial audio technologies used in virtual reality systems have undergone significant advancements over the past few decades. However, even the most advanced techniques still demonstrate some shortcomings in terms of localization accuracy and timbral fidelity. The aim of this study was to evaluate and compare the localization accuracy and timbral fidelity of five prevalent techniques utilized for immersive audio reproduction in modern virtual reality systems. The study was conducted as a subjective listening test. The results demonstrated that the method with the highest localization accuracy also exhibited the lowest degree of timbral fidelity. These findings highlight the necessity for a comprehensive optimization of the quality of the methods utilized to render spatial audio, encompassing both spatial and timbral characteristics.

EKSPERYMENTALNE ŹRÓDŁA WSZECHKIERUNKOWE NA POTRZEBY REJESTRACJI ODPOWIEDZI IMPULSOWYCH POMIESZCZEŃ

Maciej SIŁKOWSKI¹

¹Politechnika Gdańska, Gabriela Narutowicza 11/12, 80-233 Gdańsk
s176074@student.pg.edu.pl

W niniejszej pracy porównano różne eksperymentalne źródła wszechkierunkowe, które mają posłużyć jako tańsza i bardziej dostępna alternatywa dla 12-ściennych głośników klasycznie wykorzystywanych w pomiarach akustyki pomieszczeń. Porównane zostały między innymi odgłosy przebijanego balonu, wydźwięk klaskania oraz wystrzał z rewolweru. Na potrzeby eksperymentu w komorze bezekowej zestawiono stanowisko pomiarowe zbudowane z 16 mikrofonów ustawionych w kształcie sfery. Środek sfery zajmowało testowane źródło impulsu. Badania dla wybranych pobudeń dokonywano za pomocą specjalnie opracowanego mechanizmu wykluczającego obecność człowieka, jak i powtórzono pomiary dla odpowiednio kobiety i mężczyzny. Klaskanie wykonywane przez różne osoby po dwa układy dłoni. W sumie wykonano 80 pomiarów.

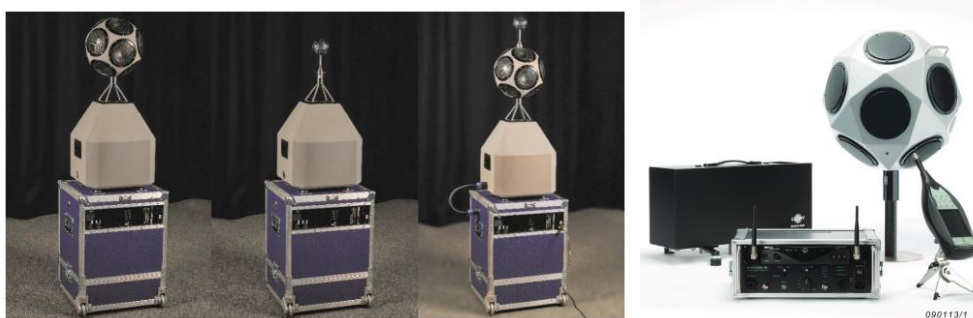
1. WSTĘP

Odpowiedź impulsowa opisuje reakcję pomieszczenia na krótki sygnał dźwiękowy, jest ona jedną z najpowszechniejszych metod badań parametrów akustyki pomieszczeń. Idealne źródło do pomiaru takich wartości powinno być wszechkierunkowe, aby równomiernie propagować dźwięk niezależnie od kąta. W związku z tym ważne jest spełnienie wymagań normy ISO-3382-1 [10], dotyczących wszechkierunkowych źródeł dźwięku, aby uzyskać dokładne pomiary akustyczne.

Podczas pomiarów zachowano zasady BHP i cały personel obecny przy pomiarach był zaopatrzony w ochronniki słuchu.

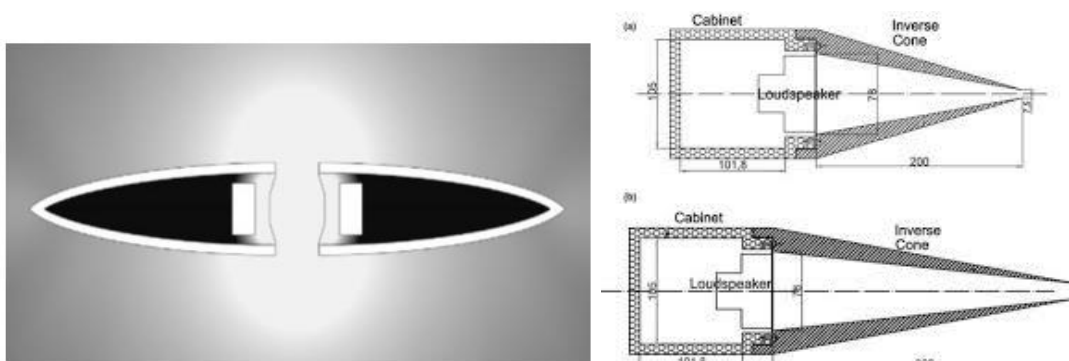
1.2. PRZEGLĄD LITERATURY

W praktyce akustycznej szeroko stosowane są różne typy źródeł dźwięku wszechkierunkowego. Do najczęściej stosowanych należą dwunastościenne systemy głośników przedstawione na Rys.1.



Rys. 1. Kolejno: trójdrożny system głośnikowy[9], dwunastościenny głośnik NTI-Audio DS3[7].

W literaturze również pojawia się koncept odwróconej tuby[1], [9], [13] (Inverse Horn) jak i radiatora pierścieniowego (*ang. ring radiator*)[12], które są przedstawione poniżej na Rys.2. Każde z tych źródeł posiada swoje unikalne zalety jak i ograniczenia. W przypadku odwróconej tuby, dźwięk staje się źródłem punktowym przez zjawisko dyfrakcji fali, gdy szczelina jest mniejsza od jej długości. Prowadzi to do niższej efektywności przy wyższych częstotliwościach. Zmniejszenie szczeliny, aby dostosować układ do tych częstotliwości, skutkuje wzrostem odbić wewnętrznych układu i nieregularnością charakterystyki dźwięku. Radiator pierścieniowy bazuje na podobnym koncepcji, jednakże w tym przypadku szczeliną jest przestrzeń pomiędzy dwoma zbliżonymi do siebie głośnikami. Z uwagi na większą szczelinę, różni się on od odwróconej tuby wzmocnioną głośnością oraz zredukowaną ilością odbić, zachowując jednak te same ograniczenia co tuba.



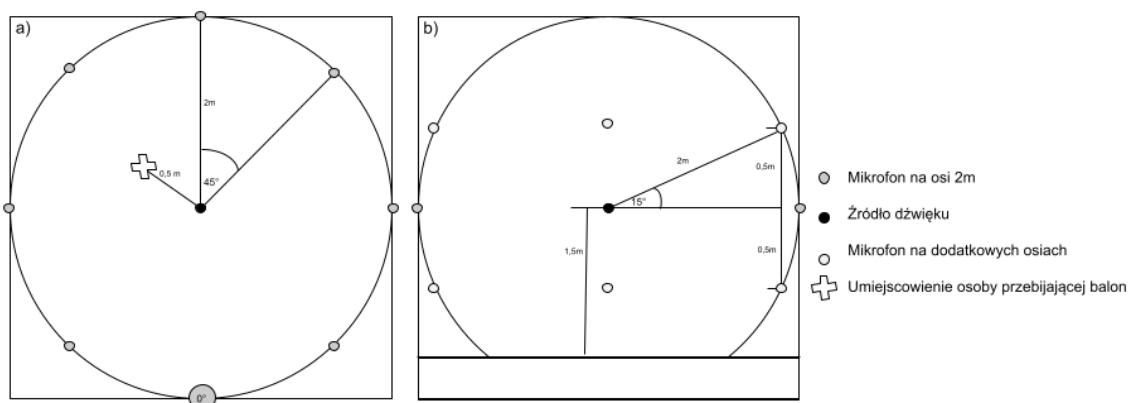
Rys. 2. Po lewej, schemat radiatora pierścieniowego[12]. Po prawej dwa projekty głośnika typu odwróconej tuby[8].

Szukając w literaturze źródeł wszechkierunkowych, możemy również spotkać się z opisem sfery pulsującej oraz źródła monopolowego. Sfera pulsująca funkcjonuje jako obiekt generujący fale akustyczną całą swoją powierzchnią, natomiast źródło monopolowe jest jej wersją istniejącą jedynie w teorii, gdyż w rzeczywistości ma on promień równy 0. W przeciwieństwie do wyżej wymienionych nie są one dostępne na rynku[15], [16]. W mniej profesjonalnych zastosowaniach, często używane są alternatywne źródła dźwięku (proponowane również przez normę ISO 3382-1), takie jak klaskanie[5], strzały z pistoletu hukowego[3] czy też klasyczne głośniki[1].

2. METODOLOGIA POMIAROWA

Wykonanie pomiarów wszechkierunkowości źródeł impulsowych wymagało przygotowania stanowiska pomiarowego w komorze bezchłowej Politechniki Gdańskiej o wymiarach 9x9x4.4m i poziomie

tła 27 dB. W pomiarach wykorzystano szesnaście identycznych mikrofonów Behringer B-5 o charakterystyce wszechkierunkowej i średnicy przetwornika wynoszącej 0,71". Mikrofon ten posiada stosunkowo płaską charakterystyką kierunkową i niski poziom szumów własnych. Mikrofony umieszczono na trzech poziomach wysokości w promieniu 2m od miejsca generowanego impulsu zgodnie z planem widocznym na Rys. 3.



Rys. 3. Plan układu pomiarowego, a) rzut z góry, b) rzut od boku.

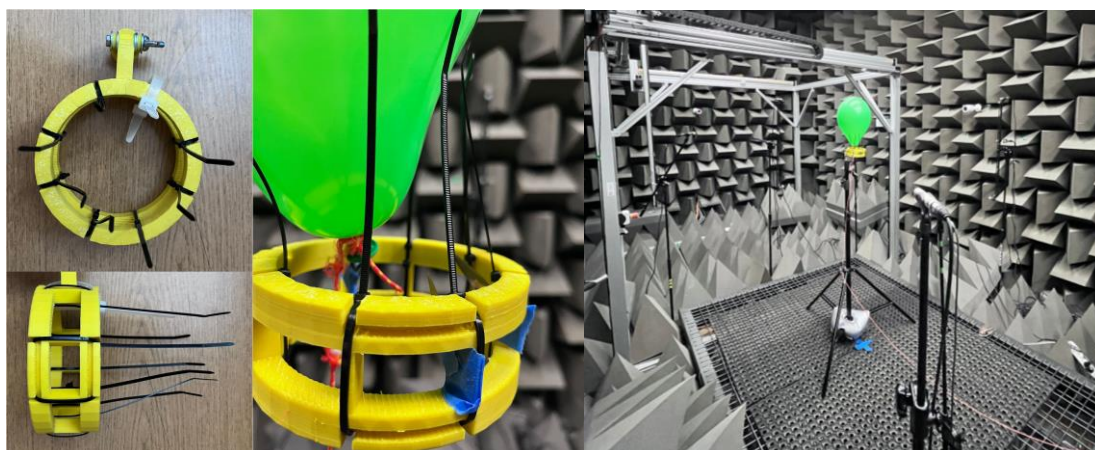
Mikrofony zostały podłączone do konsoli Behringer X32, przez którą były prowadzone nagrania. Początkowo zakładano, że rejestracja zostanie przeprowadzona z próbkowaniem 48kHz/24bit. Jednak zaobserwowano trudności z rejestracją wystrzałów (wg. literatury, impuls wystrzału na otwartej przestrzeni trwa 300us [11]). Zmieniono więc układ na Apogee zapewniającej między innymi lepsze parametry próbkowania sygnału. Do pomiarów rewolweru zatem wykorzystany został zestaw urządzeń Apogee z uwagi na wyższą częstotliwość próbkowania równą 96kHz. Pozwoli to na dokładniejsze zbadanie impulsu. Oprócz tego, w czasie pomiarów obecna była kamera slow motion, która pozwala na dokładniejszą obserwację zachodzących zjawisk. W takim układzie wykonano pomiary widoczne w Tab. 1. Każdy pomiar został wykonany przynajmniej 5 razy, a w trakcie ich trwania zostało zarejestrowane natężenie dźwięku w SPL.

Zgodnie z ISO 3382-1, idealne źródło dźwięku powinno być jak najbardziej wszechkierunkowe, a maksymalne odchylenie kierunkowości, mierzone w polu swobodnym przy użyciu różowego szumu o pasmach oktawowych, powinno mieścić się w określonych granicach (dla różnych częstotliwości wartości te są szczegółowo określone w tabeli ISO 3382-1). Źródło dźwięku musi generować poziom ciśnienia akustycznego wystarczający do uzyskania krzywych zaniku z wymaganym minimalnym zakresem dynamicznym, bez zakłóceń ze strony hałasu tła. Poziom ten powinien być co najmniej 45 dB wyższy od poziomu tła w odpowiednim paśmie częstotliwości.

Typ pomiaru	Ilość próbek [n]	Średnie szczytowe SPL [dB]
Balon Mały na statywie	5	100
Balon Duży na statywie	5	108
Balon Mały Trzymany	5	95
Balon Duży Trzymany	5	107
Rewolwer krótkie naboje na statywie	5	132
Rewolwer długie naboje na statywie	5	140
Rewolwer trzymany długie naboje	5	135
Rewolwer trzymany krótkie naboje	5	140
Klaskanie mężczyzna układ 1	10	115
Klaskanie kobieta układ 1	10	104
Klaskanie mężczyzna układ 2	10	121
Klaskanie kobieta układ 2	10	113

Tab. 1. Tabela przedstawiająca zestawienie wykonanych pomiarów

Pierwszym badanym obiektem były 4 wersje różnych balonów, trzymane przez operatora lub zamocowane na statywie. Sposób mocowania balonu został przedstawiony na Rys. 4. Dla zwiększenia powtarzalności pomiarów balony były nadmuchiwane przy pomocy sprężarki przez określony czas, tak by osiągnęły możliwie najbardziej przybliżone rozmiary. Osoby przebijające balon stały w określonym wcześniej miejscu z wyciągniętymi rękami, tak, aby swoim ciałem jak najmniej przesłonić impuls balonu.



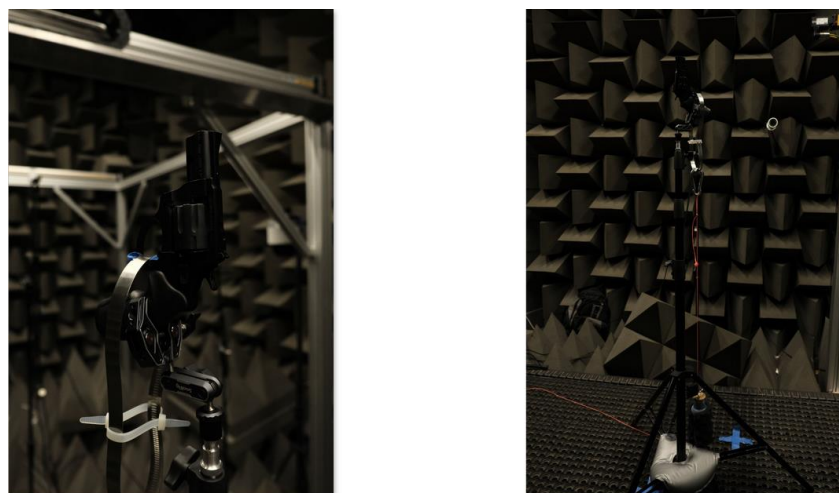
Rys. 4. Prezentacja pomiarów balonu

Klaskanie zostało podzielone na 4 pomiary w zależności od płci osoby klaszczącej oraz układu dłoni, które wzorowane były na najefektywniejszych modelach z literatury [5]. Rezultaty są widoczne na Rys. 5. Przed dokonaniem pomiaru każdy z uczestników badania został przeszkolony z wybranych technik klaskania, w celu zapewnienia jak największej powtarzalności pomiarów.



Rys. 5. Przedstawienie badanych układów klaskania kolejno: układ 1 i układ 2

Badanie rewolweru zostało przeprowadzone na broni osadzonej na statywie jak i trzymanej przez strzelca. Badanie zostało powtórzone dla krótkich i długich nabojów. Statyw został wyposażony w dociążony uchwyt pozwalający na zdalne oddanie strzału, tak aby nie zakłócać pomiarów przedstawionych na Rys. 6. Rewolwer został skierowany w górę z uwagi na charakterystykę pistoletu w planie horyzontalnym, która jest bardziej „płaska” niż jego charakterystyka w planie wertykalnym[3].

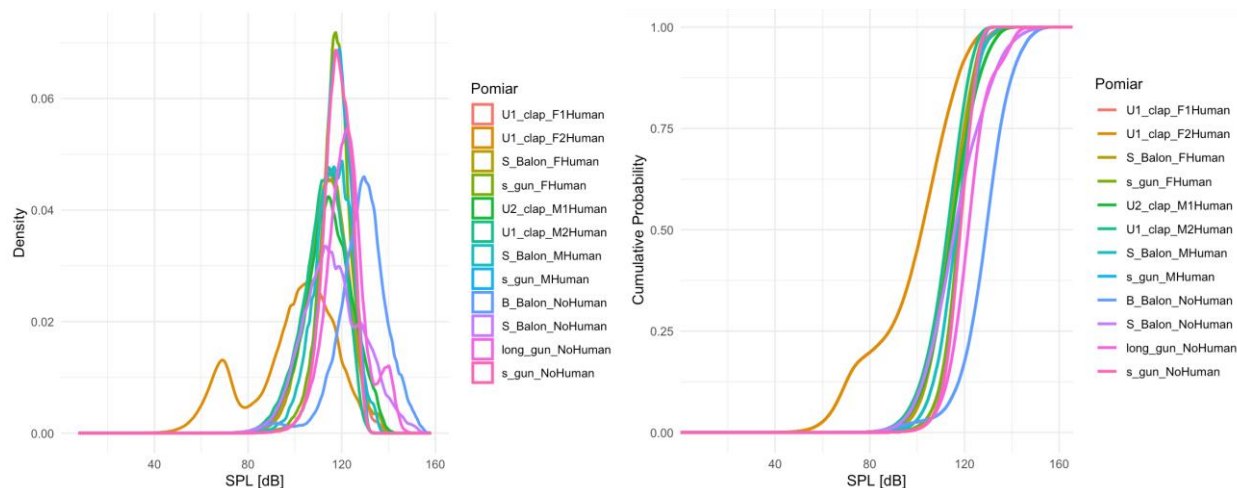


Rys. 6. Prezentacja pomiarów rewolweru

4. WYNIKI

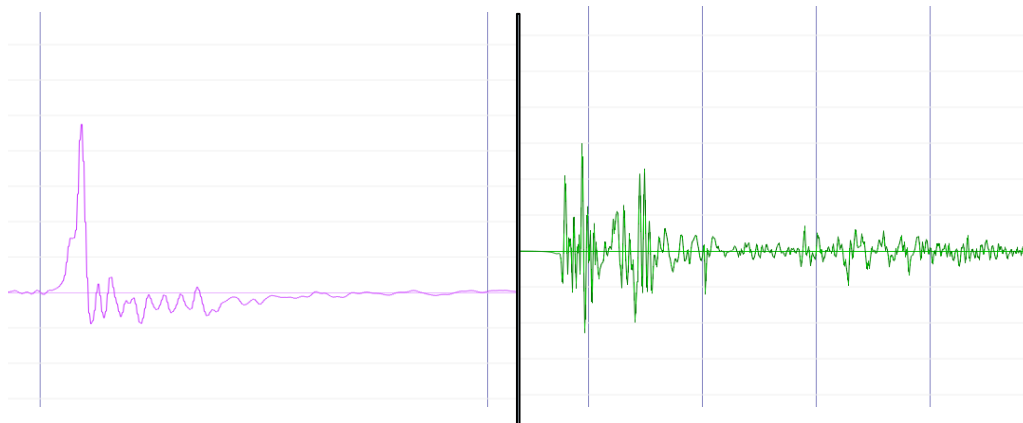
Norma ISO mówi, że źródło wszechkierunkowe powinno spełniać następujący warunek. Maksymalne dopuszczalne odchylenie od średniego poziomu całej płaszczyzny nie może przekraczać podanych wartości [10]. Cechy źródeł impulsowych badanych w tej pracy wraz z ograniczeniami sprzętowymi jednakże sprawiają, że proponowany przez normę pomiar jest trudny do wykonania dla źródeł impulsowych z uwagi na ich naturę. Jako alternatywny wyznacznik użyte zostało odchylenie standardowe charakterystyki częstotliwościowej w zakresie 110-15000 Hz, która została uśredniona pomiędzy

wykonanymi pomiarami każdego typu. Wyniki zostały przedstawione w Tab. 2. widocznej poniżej. Wykresy przedstawione na rys. 9, 10, 11 reprezentują charakterystyki dla konkretnych kątów w środkowej płaszczyźnie dla kolejno rewolweru, balonu i klaskania. Najbardziej powtarzanym pomiarem jest wystrzał z rewolweru, a najmniej klaskanie.



Rys. 7. A) Rozkład normalny wszystkich grup pomiarowych B) Dystrybuanta grup pomiarowych

Pomimo tego, wszystkie pobudzenia impulsowe spełniają warunki bycia głośniejszym o 45 dB od poziomu szumu. Dodatkowo wszystkie pobudzenia mają krótki czas narastania, szczególnie rewolwer, który można potraktować jak deltę Diraca.

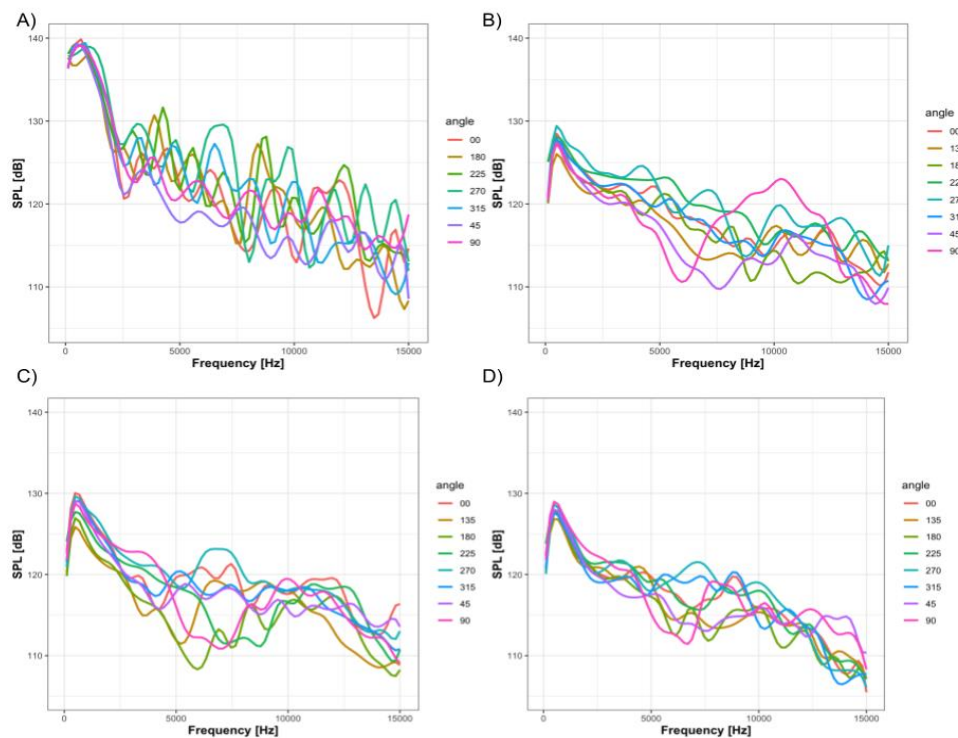


Rys. 8. A) Prezentacja impulsu rewolweru, B) Prezentacja impulsu klasknięcia

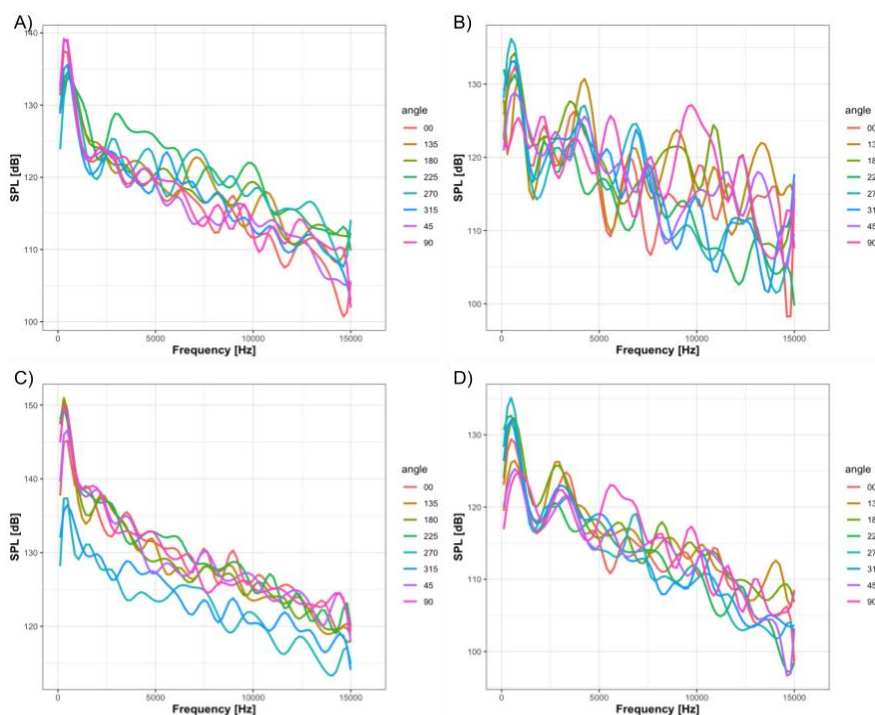
Odchylenie standardowe oraz wykresy są dowodem na dużą rozbieżność pomiędzy poziomami różnych częstotliwości. Największa różnica wynosi około 20 dB dla rewolweru. Mikrofony znajdujące się w kątach jednakże wydają się mieć bardziej spójne wartości, z kilkoma wyjątkami, takimi jak klaskanie, ponieważ osoba klaszcząca zasłania swoim ciałem mikrofon. Szczególnie widoczne jest to dla wyników klaskania mężczyzn, gdzie wartości dla punktu pomiarowego usytuowanego na 180 stopniach mają widocznie niższe wartości SPL. Dla kobiet również jest to zauważalne, ale w mniejszym stopniu. Widocznie inną wartość dla dwóch kątów ma również mały balon przebijany na statywie, jak widać na rys. 10/C. Analiza nagrań kamery slow motion wskazała na zwijanie się gumy z przebitego balonu ze strony mikrofonów usytuowanych na kątach 270° oraz 315°, co mogło wpłynąć na rozchodzenie się fali. Powodem tego mogło być przebijanie ich zawsze tym samym z trzech ostrzy umiejscowionych w uchwycie.

Typ pomiaru	Ilość próbek [n]	Odch. Standardowe [dB]	Średnie szczytowe SPL [dB]
Balon Mały na statywie	5	12.72	100
Balon Duży na statywie	5	11.12	108
Balon Mały Trzymany przez kobietę	5	8.86	95
Balon Mały Trzymany przez mężczyznę	5	8.12	107
Rewolwer krótkie naboje na statywie	5	5.99	132
Rewolwer długie naboje na statywie	5	8.74	140
Rewolwer trzymany przez kobietę krótkie naboje	5	6.31	135
Rewolwer trzymany przez mężczyznę krótkie naboje	5	6.06	140
Klaskanie mężczyzna układ 1	10	8.83	115
Klaskanie kobieta układ 1	10	8.96	104
Klaskanie mężczyzna układ 2	10	10.01	121
Klaskanie kobieta układ 2	10	18.33	113

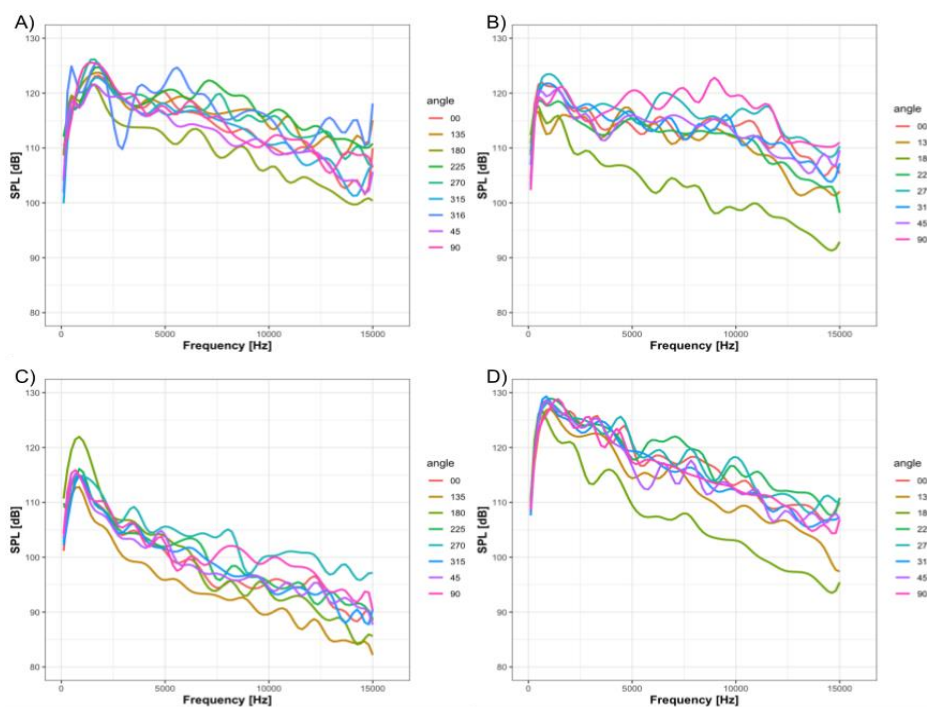
Tab. 2. Tabela przedstawiająca wyniki pomiarów



Rys. 9. Porównanie charakterystyki częstotliwościowej w zależności od kąta dla: A) długie naboje w rewolwerze umieszczonym na statywie B) krótkie naboje z rewolwerem w tej samej konfiguracji C) krótkie naboje w rewolwerze trzymany przez mężczyznę, D) krótkie naboje w rewolwerze trzymany przez kobietę.



Rys. 10. Porównanie wpływu trzymania dużego balonu na uśrednioną charakterystykę częstotliwościową dla: A) dużego balonu pękającego na statywie, B) małego balonu pękającego przez mężczyznę, C) małego balonu pękającego na statywie, D) mały balon pękany przez kobietę.



Rys. 11. Porównanie wpływu płci oraz układu dłoni na charakterystykę kłaśnień: A) kobieta klaszcząc dłońmi w układzie 1, B) mężczyzna klaszczący dłońmi w układzie 1, C) kobieta klaszcząc dłońmi w układzie 2, D) mężczyzna klaszczący dłońmi w układzie 2.

5. WNIOSKI

Powyższe badanie miało posłużyć za porównanie oraz złożenie eksperymentalnych i alternatywnych źródeł impulsowych i ich potencjalnego wykorzystania w pomiarach na potrzeby badań akustyki pomieszczeń. Podsumowując wyniki, kłaśnięcie jest najmniej powtarzalnym źródłem z największym odchyleniem standardowym. Rewolwer hukowy natomiast jest najbardziej powtarzalny, z najmniejszym odchyleniem. O ile wyniki nie dają podstawy do wykorzystania tych źródeł, zachęcają one do dalszych pomiarów i większego zgłębnienia tematu, ponieważ wiele z tych źródeł po odpowiednim dostosowaniu jest w stanie spełnić wymogi norm akustycznych. W planach na rozszerzenie badań znajduje się pomiar w dalszym polu aby zrównoważyć wpływ niskich częstotliwości jak i również porównanie praktyczne w pomiarach sali o znanych parametrach.

LITERATURA

- [1] PAPADAKIS, N.M.; STAVROULAKIS, G.E. Low Cost Omnidirectional Sound Source Utilizing a Common Directional Loudspeaker for Impulse Response Measurements. *Appl. Sci.* 2018, 8, 1703.
- [2] PÄTYNEN J, KATZ BF, LOKKI T. Investigations on the balloon as an impulse source. *J Acoust Soc Am.* 2011 Jan;129(1):EL27-33. doi: 10.1121/1.3518780. PMID: 21302978.
- [3] LAMOTHE, REINA & BRADLEY, J.. (1985). Acoustical characteristics of guns as impulse sources. 13.
- [4] LEISHMAN, TIMOTHY & ROLLINS, SARAH & SMITH, HEATHER. (2006). An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound. *The Journal of the Acoustical Society of America.* 120. 1411-1422. 10.1121/1.2221552.
- [5] PAPADAKIS, N.M.; STAVROULAKIS, G.E. Handclap for Acoustic Measurements: Optimal Application and Limitations. *Acoustics* 2020, 2, 224-245.
- [6] CHRISTENSEN, CLAUS & KOUTSOURIS, GEORGE & RINDEL, JENS. (2013). The ISO 3382 parameters: Can we simulate them? Can we measure them?. *Akustika.* 20.
- [7] NTI Audio, <https://www.nti-audio.com/en/news/new-omnidirectional-sound-source-amplifier>, Data dostępu: 02.09.2024
- [8] Ortiz Montaña, Santiago & Kolbrek, Bjørn & Pedro, Cobo & González, L. & Colina, Carlos. (2014). Point Source Loudspeaker Design: Advances on the Inverse Horn Approach. *Journal of the Audio Engineering Society.* 62. 345-354.
- [9] G. Behler, M. Vorländer. (2018). An active loudspeaker point source for the measurement of high quality wide band room impulse responses. *Proceedings of the Institute of Acoustics, AUDITORIUM ACOUSTICS 2018*
- [10] ISO 3382-1 2009
- [11] ROBERT C. MAHER AND STEVEN R. SHAW, DECIPHERING GUNSHOT RECORDINGS, 2008
- [12] Roland Kruse, Andreas Häußler, Steven van de Par, An omnidirectional loudspeaker based on a ring-radiator, *Applied Acoustics*, Volume 74, Issue 12, 2013, Pages 1374-1377, ISSN 0003-682X,
- [13] Papadakis, N.M.; Stavroulakis, G.E. Review of Acoustic Sources Alternatives to a Dodecahedron Speaker. *Appl. Sci.* 2019, 9, 3705. <https://doi.org/10.3390/app9183705>
- [14] Jaruszewska, Karolina & Rychtarikova, Monika & Wulfrank, Thomas & Jablonska, Joanna & Boeck, Magdalena. (2015). *ArAc - Multibook of Architectural Acoustics.* Energy Procedia. 78. 10.1016/j.egypro.2015.11.096.
- [15] https://www.theregister.com/2006/09/15/jvc_pulsating_sphere_speakers/ Data dostępu: 25.08.2024
- [16] Houterman, T.H.M A design of an omnidirectional sound source used for impulse response measurements

EXPERIMENTAL OMNI-DIRECTIONAL SOURCES FOR CAPTURING ROOM IMPULSE RESPONSES

This work compares various versions of omni-directional sources that are intended to serve as a cheaper and more accessible alternative to the dodecahedron loudspeakers commonly used in room impulse response measurements. Examples of such sources include: balloon bursting, hand clap, gunfire from a revolver and selected multi-speaker configurations. For this purpose a measurement system was set up inside an anechoic chamber, consisting of sixteen microphones arranged in the shape of a cross-section of a sphere, with the tested source positioned at the center. A series of measurements were made on such a system for different sources, i.e. balloons of different sizes and shapes punctured while being held by a human and kept on a stand. Hand claps were recorded in different hand configurations clapped by different people. Gunshots of a revolver were recorded both being held by the user and free-standing on a mount. Speakers were tested in 2 different speaker configurations.

NIERÓWNOMIERNE PRÓBKOWANIE PRZESTRZENNE W ZAGADNIENIU ESTYMACJI KIERUNKU NADEJŚCIA (DOA) FALI AKUSTYCZNEJ

**Zbigniew ŚWIĘTACH,
Bogusław SZLACHETKO,
Przemysław PLASKOTA,
Bartłomiej KRUK,
Michał LUCZYŃSKI,
Jędrzej SZCZEPANIAK**

Katedra Akustyki, Multimediiów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska¹

zbigniew.swietach@pwr.edu.pl
boguslaw.szlachetko@pwr.edu.pl
przemyslaw.plaskota@pwr.edu.pl
bartlomiej.kruk@pwr.edu.pl
michal.luczynski@pwr.edu.pl
jedrzej.szczepaniak@pwr.edu.pl

W artykule zaprezentowano wyniki pomiarów DOA zrealizowanych za pomocą macierzy mikrofonów, w terenie otwartym. Celem przeprowadzonych pomiarów było sprawdzenie możliwości detekcji kierunku nadejścia małych obiektów latających za pomocą mikrofonów rozmieszczonych nierównomiernie w przestrzeni. Wyznaczanie DOA zrealizowano za pomocą typowego beamformingu czasowego, przy czym mikrofony zostały rozmieszczone nierównomiernie, celowo tak aby „przestrzenne” kryterium Nyquista nie było spełnione. Empirycznie sprawdzono, że wzmiankowane kryterium nie ma zastosowania, w przypadku formowania wiązki DOA za pomocą metod czasowych. „Przestrzenne” kryterium Nyquista ogranicza wzajemne odległości pomiędzy mikrofonami, w przypadku formowania DOA za pomocą metod widmowych. Zamieszczone przykłady ilustrują poprawne wyznaczenie DOA przy użyciu beamformingu czasowego oraz mikrofonów rozmieszczonych nierównomiernie w sąsiedztwie mikrofonu referencyjnego.

1. WPROWADZENIE

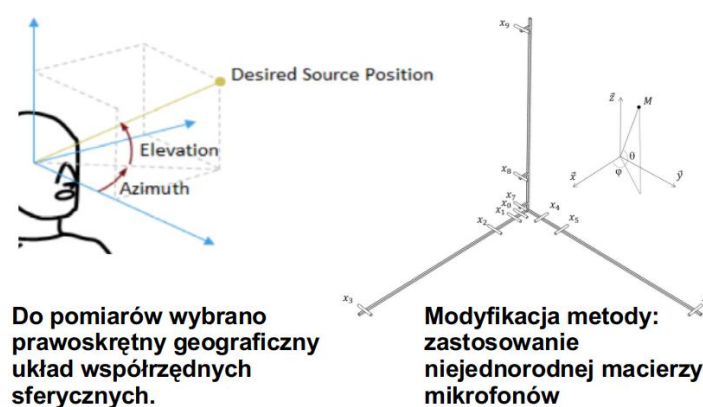
W niniejszej pracy zaprezentowana zostanie metoda wyznaczania kierunku nadejścia fali akustycznej (DOA), gdzie kierunek rozumiany jest jako azymut i elewacja poruszającego się obiektu. Za pomocą macierzy mikrofonów, wzmacniaczy, przetworników ADC (przetworników analogowo-cyfrowych) oraz autorskiego oprogramowania wyznacza się przestrzenną „mapę” chwilowego położenia obiektów latających w przestrzeni. Wszystko to realizowane jest za pomocą typowych urządzeń stosowanych w technice audio, tzn. mikrofonów firmy Superlux oraz 16 kanałowego procesora dźwięku firmy Roland. Dane w formie strumieni PCM (modulacja kodowo-impulsowa) zapisywane są w plikach *.wav i przesyłane do dalszej obróbki w komputerze, za pomocą środowiska obliczeniowego Matlab. Wybór Matlab do dalszego przetwarzania sygnałów podyktowany jest prostotą i przejrzystością kodu źródłowego, będącego plikiem tekstowym. Ponadto sposób zapisu problemu technicznego czy matematycznego w Matlabie jest wręcz intuicyjny oraz niewiele odbiega od zapisu takiego problemu przy użyciu standardowej notacji matematycznej.

Obiektami latającymi, będącymi przedmiotem testów są dwa drony oraz dodatkowo samoloty pasażerskie, przelatujące w odległości około 2-3 kilometrów od macierzy mikrofonów. Wykonano szereg pomiarów fizycznych, zarówno w komorze akustycznej (symulacja fali akustycznej generowanej przez dron na pomocą głośników) jak i w otwartym terenie przy użyciu latającego drona. Zastosowanie technik audio do wyznaczania DOA [1] jest uzupełnieniem technik radarowych wykrywania obiektów latających, przy czym akustyczne DOA umożliwia pasywne wykrywanie np. niewielkich dronów znajdujących się na niskim pułapie i w niewielkiej odległości od potencjalnego celu [3]. W opisanej sytuacji metody radarowe są mało skuteczne.

Ważnym celem omawianego eksperymentu było również zbadanie możliwości wykorzystania macierzy mikrofonów o innej niż jednorodna strukturze przestrzennej [2]. To oznacza nierównomierne próbkowanie przestrzenne i nie było oczywiste, jak takie rozmieszczenie mikrofonów wpłynie na możliwość estymacji DOA. Finalnie okazało się, że stosując metodę estymacji DOA opartą na formowaniu wiązki w dziedzinie czasu (DSB – delay and sum beamforming), ograniczenia dotyczące próbkowania przestrzennego nie znalazły tutaj zastosowania. Było to zgodne z przeprowadzoną przed pomiarami analizą teoretyczną. Wzmiankowane ograniczenia są jednak nadal istotne jeżeli do estymacji DOA używa się metod widmowych [4].

2. REALIZACJA AKUSTYCZNEGO DOA

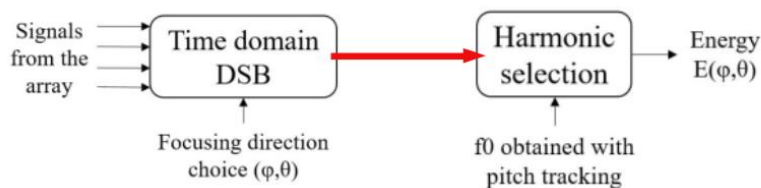
Na rys. 1 pokazano schemat współrzędnych przestrzennych i rozmieszczenie mikrofonów w macierzy pomiarowej.



Rys. 1a, b Układ współrzędnych pomiarowych i macierz mikrofonów (rys. 1a – źródło Matlab help)

Pomiary przeprowadzono przy użyciu 10 mikrofonów rozmieszczonych nierównomiernie w macierzy pomiarowej. Mikrofony numerowane są od zera do dziewięciu, przy czym numer zero oznacza mikrofon referencyjny. Odległości kolejnych mikrofonów od mikrofonu referencyjnego są następujące: $l_1 = 5cm, l_2 = 15cm, l_3 = 90cm$.

Na rys. 2 zamieszczono schemat blokowy realizacji akustycznego DOA. Pierwszy blok odpowiada za formowanie wiązki na podstawie sygnałów otrzymanych z mikrofonów. Wiązka formowana jest za pomocą typowej metody czasowego beamformingu (DSB) [1]. Takie podejście umożliwia stosowanie dowolnej konfiguracji przestrzennej mikrofonów pomiarowych. Nie występuje tutaj ograniczenie związane z próbkowaniem przestrzennym.



Formowanie wiązki

Wyznaczanie energii (mocy)

Rys. 2 Schemat blokowy: blok beamformera i blok wyznaczania energii uformowanej wiązki

Drugi blok na schemacie rys. 2 realizuje wyznaczanie mocy sygnału po przeprowadzonym beamformingu. W pracach [1,4] zaproponowano dwie metody obliczania mocy sygnału nadlatującego obiektu, które w zamierzeniu powinny poprawić stosunek sygnału do szumu (S/N) i tym samym umożliwić detekcję nadlatujących obiektów z np. większej odległości lub w obecności sygnałów zakłócających o większych poziomach mocy. Wzmiankowane metody to HPS (Harmonic Product Spectrum) oraz algorytm SHC (Spectral Harmonic Correlation) będący „rozwinieciem” metody HPS. Powyższe algorytmy nie będą tutaj omawiane szczegółowo, ponieważ w przeprowadzonych przez autorów niniejszego opracowania badaniach okazały się one zupełnie nieprzydatne. W niniejszej pracy wyznaczanie mocy sygnału po beamformingu zrealizowano za pomocą sumowania mocy wszystkich składowych widma w przedziale częstotliwości $f \in (80,3000)Hz$ lub równoważnie za pomocą sumowania mocy próbek sygnału po wykonaniu beamformingu i przeprowadzeniu odpowiedniej filtracji liniowym filtrem pasmowo – przepustowym. Okazało się, że zastosowanie algorytmów HPS lub SHC pogarsza S/N mocy sygnału akustycznego pochodzącego od nadlatującego obiektu.

W ocenie autorów nieprzydatność algorytmów HPS oraz SHC wynika z faktu, że analizowane widma są niestacjonarne, zmieniają się w czasie szybko, w sposób niekontrolowany. Tym samym nie można fizycznie wyznaczyć częstotliwości podstawowej, np. wynikającej z obrotów silnika i ilości łopatek wirnika drona oraz jej harmonicznnych. Wydaje się, że w przypadku wyznaczania DOA nadlatującego obiektu omawiane algorytmy są generalnie mało przydatne. Co więcej, nawet autorzy [1] twierdzący, że ww. algorytmy istotnie poprawiają S/N nie podają żadnego przykładu dla rzeczywistych danych pomiarowych, np. pomiarów w otwartej przestrzeni. W zakresie stosowania algorytmów HPS i SHC przykłady podane w [1] oparte są tylko na przeprowadzonych symulacjach komputerowych.

3. WYNIKI WYZNACZANIA DOA

Wyznaczenie wiązki sygnału po beamformingu dla jednego ustalonego kierunku przestrzennego wymaga przeprowadzenia wielu operacji zmiennoprzecinkowych, co generalnie zajmuje dużo czasu obliczeniowego. W celu zmniejszenia liczby obliczeń przyjęto, że azymut i elewacja będą zmieniać się o 10^{-2} , co daje 324 różnych punktów na półsfery. Przyjęto maksymalną prędkość drona równą $V_d = 200 km/h$ oraz, że w czasie akwizycji pojedynczej ramki sygnału dron może zmienić położenie co najwyżej o $\Delta R = 15m$. W rezultacie maksymalny czas akwizycji wynosi $T = 270ms$. Założono, że dron znajduje się w odległości większej niż $R = 100m$ od macierzy mikrofonów, co oznacza, że względna zmiana położenia drona w czasie akwizycji sygnału jest mniejsza niż 15%. W praktyce z przyczyn technicznych wykonywano pomiary również w odległościach mniejszych niż 100 m, ale wówczas w czasie akwizycji sygnałów dron zawieszony był nieruchomo w danym punkcie sfery.

Częstotliwość próbkowania jest równa $f_s = 48kHz$ i w założeniach miało to umożliwić przesunięcia

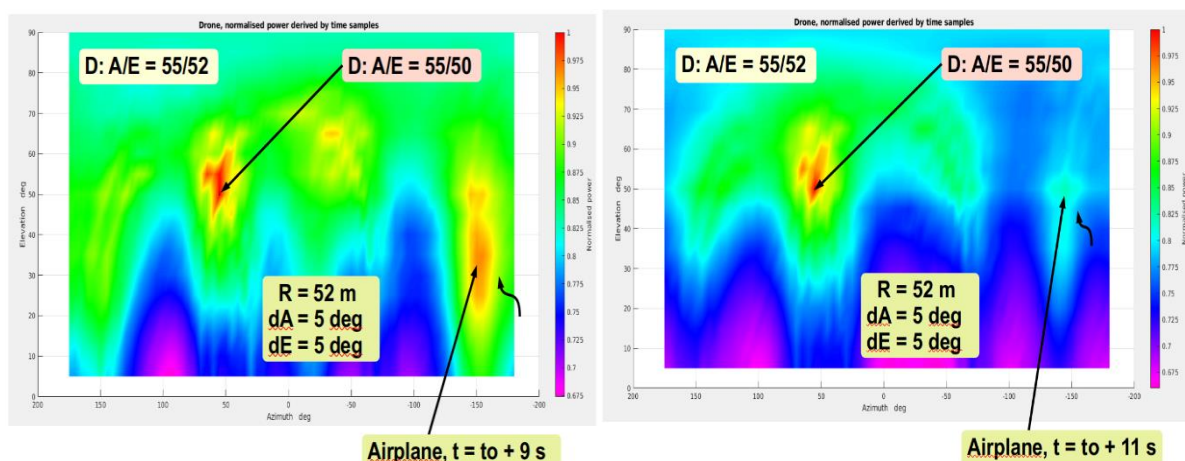
czasowe sygnałów z mikrofonów o co najmniej jedną próbkę, w celu otrzymania sygnału po beamformingu. Okazało się jednak, że wymagane przesunięcia czasowe pomiędzy sygnałami powinny być mniejsze niż $\Delta T = 1/f_s$. Tym samym stało się konieczne zastosowanie interpolacji lub generalnie aproksymacji sygnałów z mikrofonów, aby uzyskać wymagane opóźnienia. Po sprawdzeniu wielu metod przybliżania sygnału na podstawie jego próbek, np. splajny sześciennne, interpolacja trygonometryczna, trygonometryczna aproksymacja średniokwadratowa, zastosowano interpolację w widmie sygnału, ze względu na bardzo dobre własności numeryczne i najszybszy czas obliczeń. Po wykonaniu wzmiankowanej interpolacji ponownie przechodzi się do dziedziny czasu, w celu przeprowadzenia beamformingu.

W trakcie pomiarów używano dwóch dronów. Dron numer 1 to mały, lekki, czterosiłnikowy dron o średnicy około 48 cm po rozłożeniu. Każde śmigło zbudowane jest z trzech łopatek. Dron numer 2 to ciężki, duży, sześciosiłnikowy dron o średnicy około 1.2 m i masie własnej 9.5 kg. Każde śmigło zbudowane jest z dwóch łopatek.



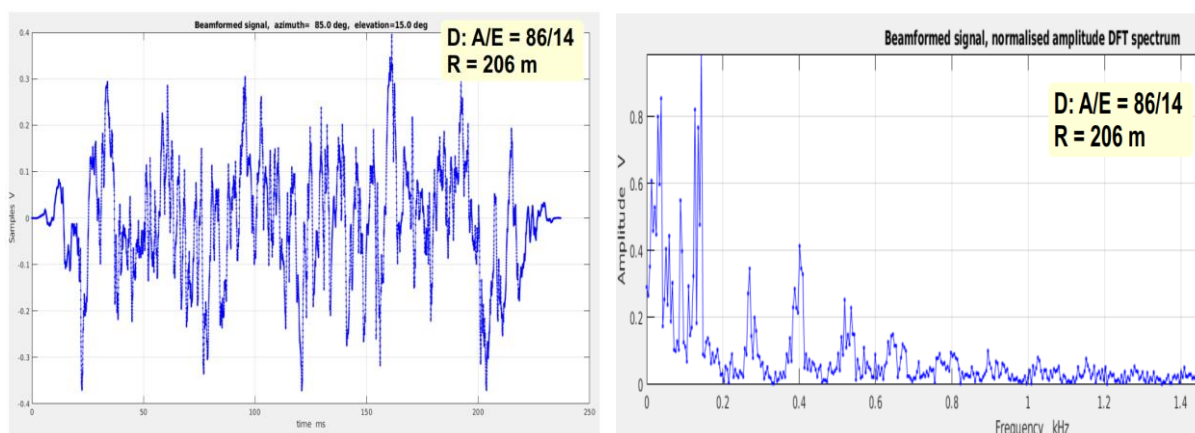
Rys. 3a Macierz mikrofonów (góra), dron numer 2 (dół). 3b Pomiary, trawiaste lotnisko pod Wrocławiem .

Na rys. 4 przedstawiono jeden z wyników eksperymentów przeprowadzonych z wykorzystaniem mobilnych źródeł dźwięku (BSP). Eksperymenty przeprowadzono w środowisku naturalnym, rys. 3b. Poza BSP użytym w eksperymentach pojawiły się zakłócenia pochodzące od przelatujących samolotów pasażerskich. W takich przypadkach macierz umożliwiała określenie DOA obydwu obiektów jednocześnie.



Rys. 4 Mapa przestrzenna rozkładu mocy uformowanych wiązek. Dron znajduje się w położeniu A/E = 55/50 deg, natomiast przelatujący samolot w dwóch położeniach, po 9 i 11 s i odlatuje coraz dalej.

Na rys. 5 zilustrowano uformowaną wiązkę sygnałów po beamformingu dla jednego z wykonanych pomiarów oraz jej widmo amplitudowe. W widmie widać znaczną liczbę harmonicznnych generowanych przez nadlatujący dron. Wzmiankowane harmoniczne wyraźnie wyróżniają się od szumów tła, pomimo że dron numer 2 znajduje się ponad 200 m od macierzy mikrofonów. Częstotliwości ww. harmonicznnych nie są jednak wielokrotnościami częstotliwości obrotów łopatek śmigła, nawet w przybliżeniu. Częstotliwości te zmieniają się również w kolejnych akwizycjach danych. To przypuszczalnie jest przyczyną, że metody HPS oraz SHC nie działają poprawnie.



Rys. 5 Uformowana wiązka i jej widmo amplitudowe. Dron znajduje się w położeniu A/E = 86/14 deg, w odległości 206 m.

4. PODSUMOWANIE

Po przeprowadzeniu dotychczasowych prac można wyciągnąć następujące wnioski:

- 1) Akustyczne DOA działa poprawnie i umożliwia wykrywanie kilku obiektów jednocześnie
- 2) System nie wymaga równomiernego rozmieszczenia mikrofonów, zatem równomierne próbkowanie przestrzenne nie jest konieczne w przypadku stosowania czasowych metod formowania wiązki
- 3) Dla czasowych metod formowania wiązki nie ma zastosowania ograniczenie w postaci „przestrzennego” kryterium Nyquista

Ponadto wydaje się, że metody estymacji mocy sygnału użytecznego z sygnałów otrzymanych w

wyniku beamformingu, bazujące na znajdowaniu częstotliwości charakterystycznych i ich harmonicznych, czyli metody HPS oraz SHC są aktualnie raczej nieprzydatne.

Następnym etapem prac będzie rozważenie płaskiej macierzy mikrofonów do wyznaczania DOA. Płaska macierz mikrofonów z pewnością jest bardziej korzystna w zastosowaniach praktycznych niż macierz przestrzenna, jednak jej użycie wymaga przemyślenia geometrii takiej macierzy. Próby wyznaczania DOA na podstawie omawianych w artykule pomiarów bez użycia mikrofonów umieszczonych na osi „z” niestety prowadziły do niezadowolających rezultatów. Teoretycznie z mikrofonów rozmieszczonych na płaszczyźnie można uzyskać informację o elewacji nadlatującego obiektu, jednak praktyka pokazała, że rozmieszczenie trzech mikrofonów na osiach „ox” oraz „oy” odpowiednio, nie było wystarczające do tego celu. Planowany jest zakup macierzy 16 mikrofonów MEMS rozmieszczonych na planie kwadratu (4 x 4 mikrofony). Wzmiankowana macierz posiada własny system akwizycji sygnałów oraz sterowniki umożliwiające komunikację z komputerem Pc. Zostaną wykonane pomiary w terenie za pomocą takiej macierzy i wówczas okaże się, czy można skutecznie wyznaczać DOA bez użycia mikrofonów rozmieszczonych na osi „oz”.

LITERATURA

- [1] Nathan Itare, Jean-Hugh Thomas, Kosai Raouf, Torea Blanchard Acoustic Estimation of the Direction of Arrival of an Unmanned Aerial Vehicle Based on Frequency Tracking in the Time-Frequency Plane. MDPI Sensors 26.05.2022
- [2] Sedunov, A.; Haddad, D.; Salloum, H.; Sutin, A.; Sedunov, N.; Yakubovskiy, A. Stevens Drone Detection Acoustic System and Experiments in Acoustics UAV Tracking. In Proceedings of the 2019 IEEE International Symposium on Technologies for Homeland Security (HST), Woburn, MA, USA, 5–6 November 2019; pp. 1–7.
- [3] Kloet, N.; Watkins, S.; Clothier, R. Acoustic signature measurement of small multi-rotor unmanned aircraft systems. *Int. J. Micro Air Veh.* 2017, 9, 3–14.
- [4] Zahorian, S.A.; Hu, H. A spectral/temporal method for robust fundamental frequency tracking. *J. Acoust. Soc. Am.* 2008, 123, 4559–4571.

NON-UNIFORM SPATIAL SAMPLING IN THE PROBLEM OF ESTIMATING THE DIRECTION OF ARRIVAL (DOA) OF AN ACOUSTIC WAVE

This work will present a method for determining the direction of arrival of an acoustic wave (DOA), where the direction is understood as the azimuth and elevation of a moving object. Using a matrix of microphones, amplifiers, ADCs (analog-to-digital converters) and proprietary software, a spatial "map" of the momentary position of flying objects in space is determined. The use of audio techniques to determine DOA complements radar techniques for detecting flying objects, with acoustic DOA enabling passive detection of e.g. small drones located at a low altitude and close to a potential target. In the described situation, radar methods are not very effective. An important goal of the discussed experiment was also to investigate the possibility of using microphone arrays with a spatial structure other than homogeneous. This means uneven spatial sampling, and it was not obvious how this microphone placement would affect the ability to estimate DOA.

WPLYW WYGRZEWANIA PRZETWORNIKÓW NA PARAMETRY SUBIEKTYWNE I OBIEKTYWNE ZESTAWÓW GŁOŚNIKOWYCH.

Tomasz KOPCIŃSKI¹
Dominika KUCZAK¹
Bartłomiej KRUK¹
Jan KUCHARCZYK¹
Tomasz NOWAK¹

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

tomasz.kopcinski@pwr.edu.pl

dominika.kuczak@aes.pwr.edu.pl

bartlomiej.kruk@pwr.edu.pl

jan.kucharczyk@aes.pwr.edu.pl

t.nowak@pwr.edu.pl

1. WPROWADZENIE

Zestawy głośnikowe to urządzenia głośnikowe wykorzystujące przetworniki odtwarzające różne pasma częstotliwościowe, uzyskane dzięki zastosowaniu zestawu filtrów RLC, zwanymi zwrotnicą. Przetworniki elektromagnetyczne wykorzystywane w zestawach głośnikowych składają się z układu magnetycznego, kosza oraz układu drgającego kompletnego. Natomiast w skład układu drgającego wchodzi: cewka drgająca, membrana kompletna, resor dolny oraz krążek ochronny (kopułka przeciwpylowa). Cewka drgająca jest nawiniętym w sposób liniowy drutem miedzianym lub CCAW (Cooper Coated Aluminium Wire) na korpusie najczęściej wykonanym z aluminium bądź folii typu kapton. Drut wykorzystywany do nawoju cewek drgających jest drutem termospekialnym z warstwą klejową (tzw. pre-bonding). W celu złączenia zwojów cewki oraz korpusu nawiniętą cewkę wypieka się w piecu. Membrana kompletna składa się z membrany wykonanej najczęściej z papieru lub folii PCV, zawieszona górnego oraz lic połączeniowych. Resor dolny jest rezolitywym elementem sprężystym, centrującym cewkę w szczelinie oraz posiadającym podatność mechaniczną. Wyroby rezolitywowe powstają z wprasowania w wysokiej temperaturze żywicy fenolowej do tkaniny, najczęściej bawełnianej. W kręgach audiofilskich występuje przekonanie, że „wygrzanie” czyli użytkowanie urządzeń głośnikowych przez pewien czas wpływa pozytywnie na parametry obiektywne i subiektywne zestawów głośnikowych. Jako argumenty potwierdzające tę teorię stawia się zmiany starzeniowe zachodzące w przetwornikach, takie jak zmiana podatności zawiesznień, która pozytywnie wpływa na pasmo przenoszenia oraz zniekształcenia nieliniowe przetworników. Kolejnym procesem jest zmiana parametrów cewki drgającej, wynikającej z jej nagrzewania. Zwolennicy teorii wygrzewania stawiają tezy o zwiększeniu głębi, szczegółowości, dynamiki oraz zmniejszeniu zniekształceń. Przeciwnicy natomiast powołują się na brak naukowych dowodów potwierdzających zmiany w parametrach obiektywnych i subiektywnych wynikających z procesu wygrzewania przetworników.

2. PROCES WYGRZEWANIA

Proces wygrzewania przeprowadzono zgodnie z normą PN-EN 60268-5:2021 (PN-EN 60268-5:2021, Urządzenia systemów elektroakustycznych. Część 5: Głośniki, 2021) oraz AES75-2023 (Society, 2023). Na wygrzewane zestawy został podany sygnał szerokopasmowy „Music Noise” (Society, 2023) o napięciu odpowiadającemu połowie maksymalnego napięcia wejściowego U_{MAX} wyznaczonego zgodnie ze wzorem (1). Proces wygrzewania trwał 3 godziny.

$$U_{MAX} = \sqrt{P_{MAX} \cdot Z(1)}$$

Norma AES75-2023 wyznacza sygnał „Music Noise” jako nowy standard dający wyniki bardziej zbliżone do odtwarzania muzyki w porównaniu do sygnału typu program.

W magazynie „Audioholics: Speaker Break-In: Fact or Fiction?” Mark Sanfilipo w artykule „Depending on number of mechanical factors relating to its design, construction and materials composition, a net post-burn in compliance increase on the order of 5% - 10% would not be unusual.” (Sanfilipo, 2005) zwraca uwagę, że części mechaniczne głośnika, takie jak zawieszenie górne oraz dolne zmieniają swoją podatność pod wpływem pracy. Autor w swoich badaniach podkreśla, że proces wygrzewania przetwornika zajmuje kilka sekund i może być zaimplementowany w procesie produkcyjnym. W artykule przedstawiono modele przetwornika przed oraz po procesie wygrzewania wraz z zauważonymi różnicami. Rozbieżność w amplitudzie wynosiła jedynie 0,09 dB na korzyść modelu przetwornika wygrzanego. Zauważono również różnice w maksimach charakterystyk modułów impedancji.

W publikacji Ethana Winer pt. „The Audio Expert: Everything You Need to Know About Audio. Focal Press.” (Winer, 2018) autor wychodzi naprzeciw popularnym mitom audiofilskim i przedstawia stwierdzenie, że po pewnym czasie może dojść do zmian relaksacyjnych w materiałach wykorzystanych w produkcji przetworników. Nie stwierdził on jednak, aby można było usłyszeć różnice.

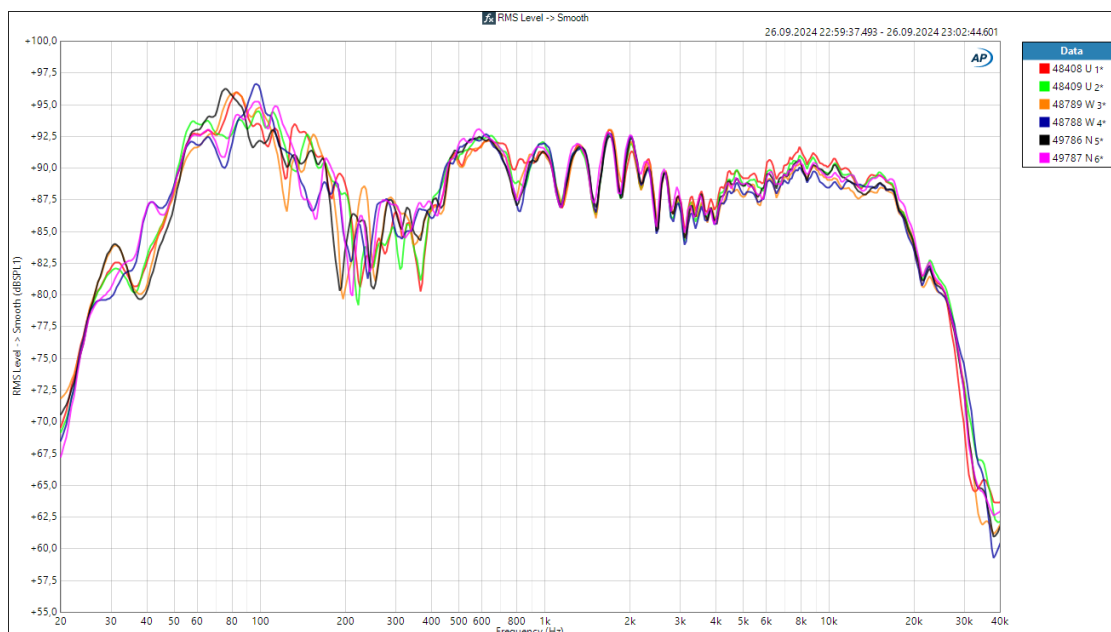
3. METODYKA BADAWCZA

W celach badawczych wykonano serie pomiarów obiektywnych, które szczegółowo omówiono. Następnie wykonano badania subiektywne.

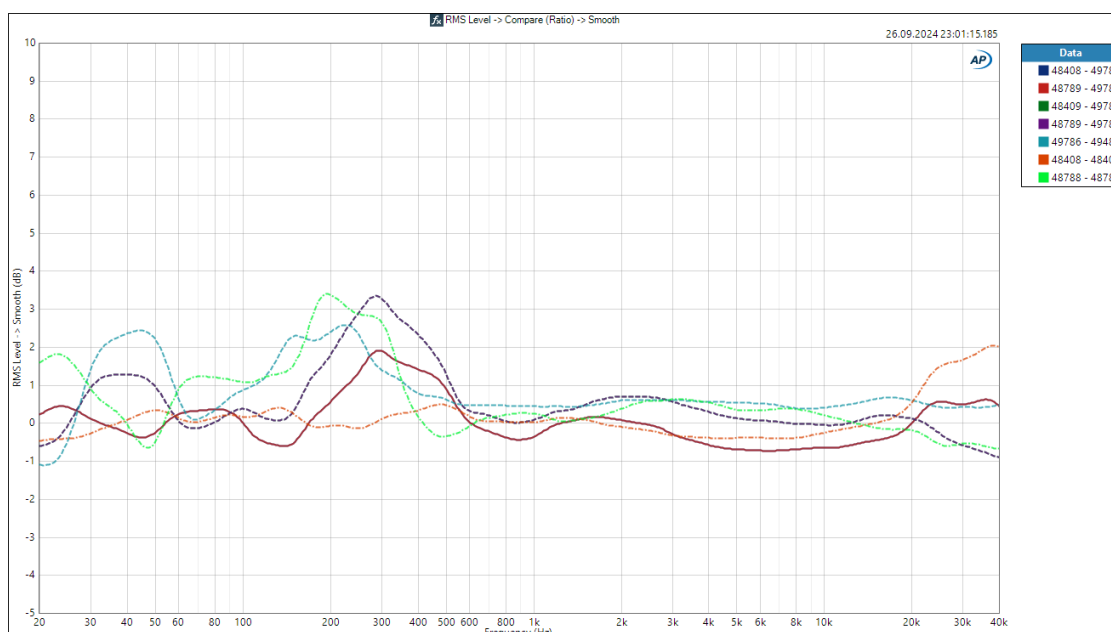
3.1 POMIARY OBIEKTYWNE

Pomiary obiektywne wykonano z wykorzystaniem systemu pomiarowego AudioPrecision APx515 wraz ze wzmacniaczem służącym do pomiarów przetworników elektroakustycznych oraz zestawów głośnikowych produkcji AudioPrecision. Pomiary wykonano umieszczając mikrofon pomiarowy w odległości 1 m od badanych zestawów na wysokości 1,2 m co odpowiadało wysokości przetwornika wysokotonowego znajdującego się w badanych zestawach. Na wyjściu wzmacniacza mocy podany został sygnał Sweep o napięciu 2,83 V RMS. Zestawy mierzono w zakresie częstotliwości od 20 Hz do 40 kHz. (PN-EN 60268-5:2021, Urządzenia systemów elektroakustycznych. Część 5: Głośniki, 2021)

3.2 CHARAKTERYSTYKA CZĘSTOTLIWOŚCIOWA



Rysunek 1 Charakterystyki częstotliwościowe badanych zestawów (wygładzanie 1/12 oktawy)



Rysunek 2 Różnice charakterystyk częstotliwościowych badanych zestawów głośnikowych

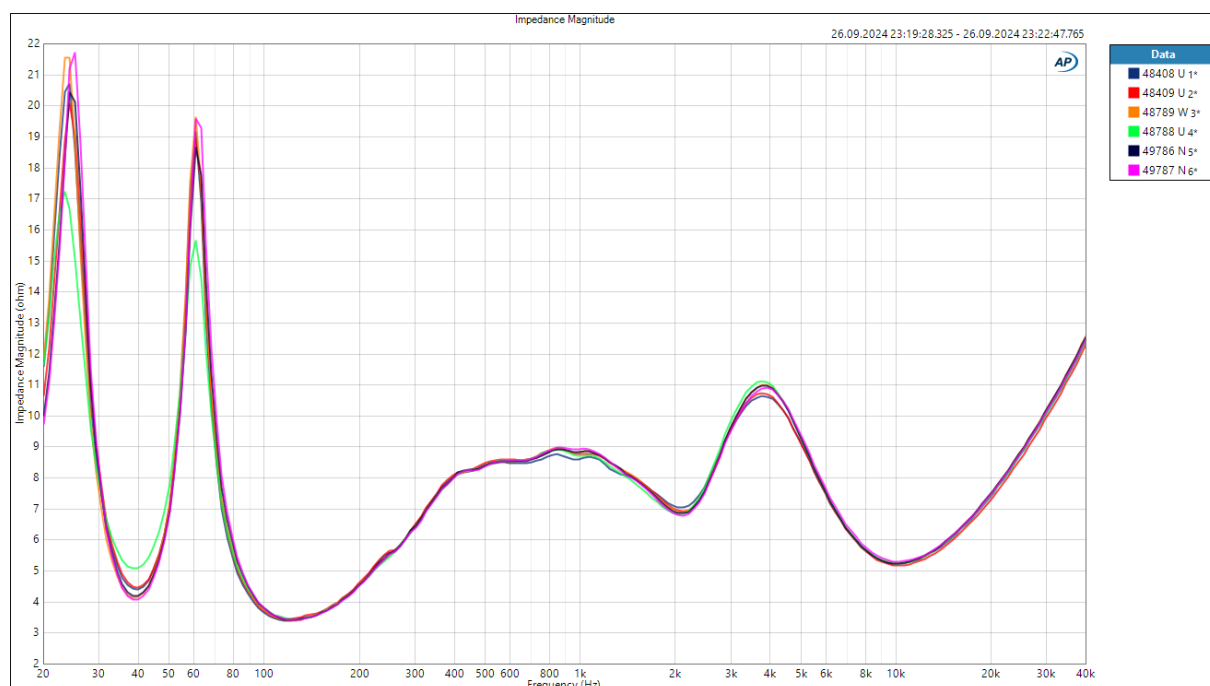
Rys. 1 przedstawia wykresy charakterystyk częstotliwościowych badanych zestawów. Legenda przedstawia numery seryjne urządzeń głośnikowych biorących udział w badaniu. Litery U, W oraz N oznaczają kolejno urządzenia używane, wygrzewane oraz nowe. Charakterystyki częstotliwościowe badanych urządzeń przedstawiono z wygładzaniem 1/12 oktawy. Jak można zauważyć, w charakterystykach badanych urządzeń występują znaczne rozbieżności zwłaszcza w zakresie małych częstotliwości. Tyczy się to również zestawów z tej samej serii o kolejnych numerach seryjnych. Dokładne różnice przedstawia rys. 2. Największe rozbieżności zaobserwowano dla urządzeń niewygrzewanych pochodzących z tej samej serii, posiadające kolejne numery seryjne.

3.3 CHARAKTERYSTYKA MODUŁU IMPEDANCJI

Pomiary modułu impedancji wykonywano doprowadzając do badanego zestawu sygnał sinusoidalny o napięciu 1 V RMS w zakresie od 20 Hz do 40 kHz. Pomiary wykonano w 200 punktach. Znając wartość napięcia na wyjściu wzmacniacza mierzono prąd płynący przez badany zestaw głośnikowy. Korzystając z zależności (2):

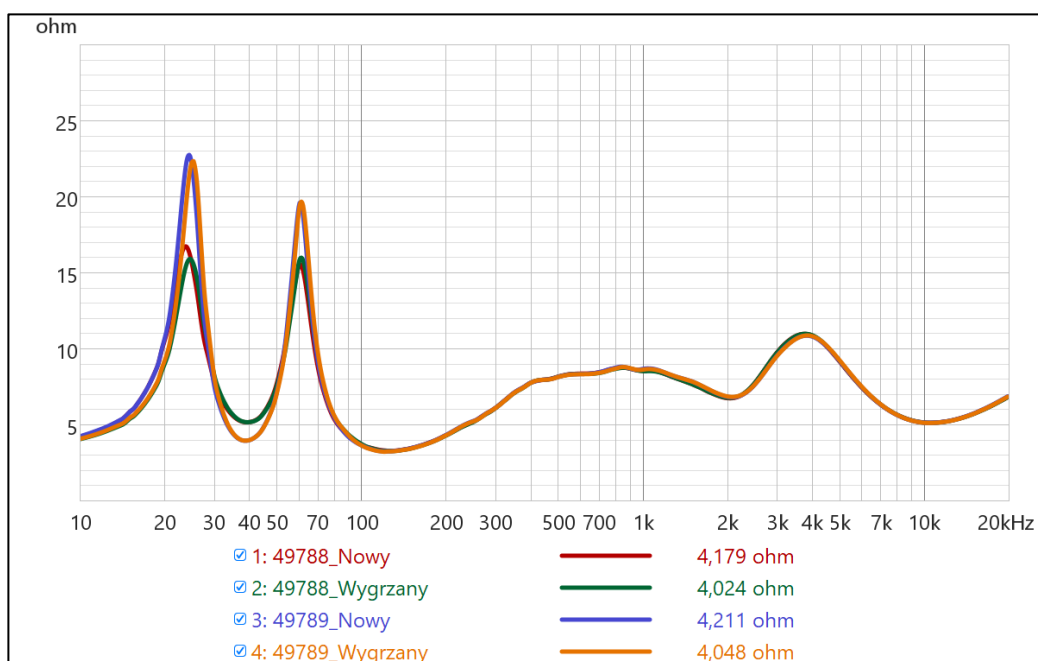
$$Z = \frac{U_f}{I_f} \quad (2)$$

wyznaczano wartość modułu impedancji badanych zestawów w zadanych punktach (częstotliwościach).

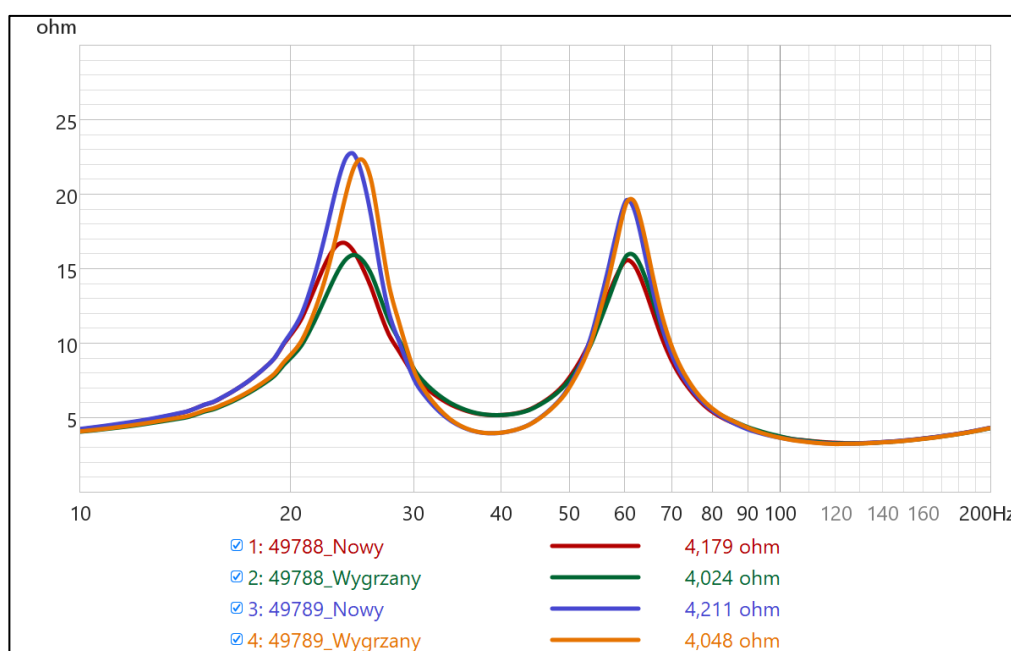


Rysunek 3 Charakterystyki modułów impedancji badanych zestawów

Rys. 3 przedstawia wykres modułów impedancji dla wszystkich badanych urządzeń głośnikowych. Jak można zauważyć, maksima modułów impedancji wynikające z obciążenia membrany portem bass – refleks we wszystkich badanych urządzeniach występują w tym samym miejscu, jednakże maksima wynikające z parametrów układu drgającego przetwornika zdają się być przesunięte względem siebie o kilka herców. Wynika to z rozbieżności produkcyjnych przetworników oraz procesów starzeniowych elementów składowych układu drgającego. Największe różnice w modułach impedancji występują dla zestawów pochodzących z tych samych serii produkcyjnych co potwierdza, że rozbieżności produkcyjne są większe niż wpływ procesu wygrzewania.



Rysunek 4 Charakterystyki modułu impedancji zestawów 49788 oraz 49789 przed oraz po wygrzaniu sygnałem Music-Noise

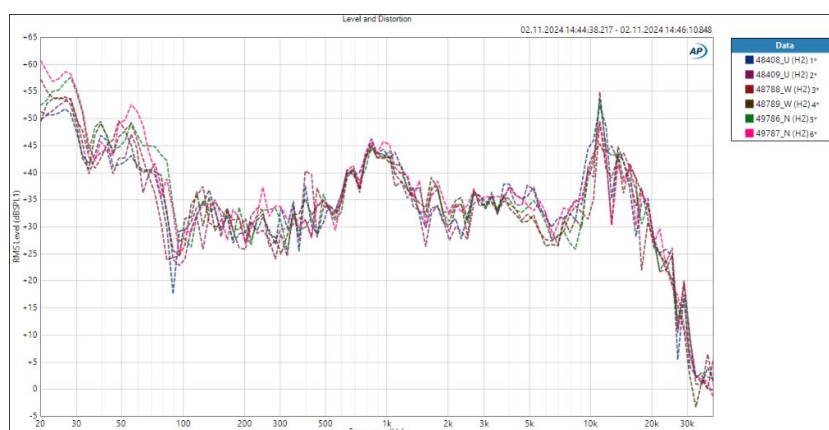


Rysunek 5 Charakterystyki modułu impedancji zestawów 49788 oraz 48789 przed oraz po wygrzaniu sygnałem Music-Noise (zbliżenie na rezonans)

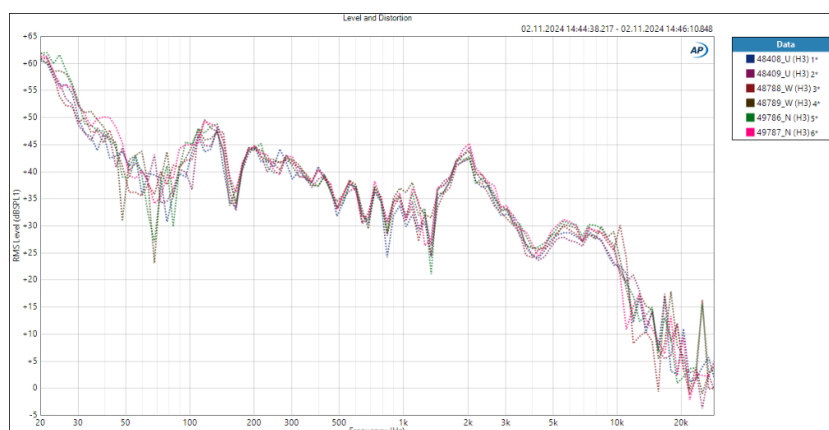
Analizując rys. 4 oraz rys. 5 można stwierdzić, że pomiędzy zestawami z tej samej serii, różniącymi się numerem seryjnym o 1, występują znaczne różnice w wykresach modułu impedancji. Wygrzewanie spowodowało nieznaczne przesunięcie maksimum impedancji przetworników, co pozwala stwierdzić, iż wzrosła podatność zawiesznień. Może to wystąpić poprzez zmianę temperatury gumy, z której zrobione są zawieszania górne, lub poprzez wykruszenie się żywicy fenolowej z resora.

3.4 ZNIEKSZTAŁCENIA HARMONICZNE

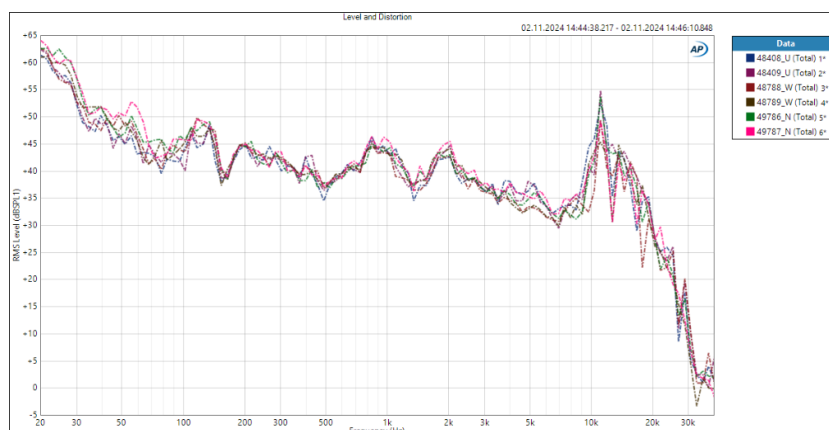
Rysunki 6, 7 i 8 przedstawiają wykresy 2 i 3 harmoniczej oraz całkowite zniekształcenia harmoniczne. Zniekształcenia harmoniczne wynikają z nieliniowości elementów składowych układu drgającego kompletnego, takich jak zawieszenie górne oraz resor (Bruel & Kjaer, 1986). Wartości zniekształceń wszystkich badanych zestawów głośnikowych są ze sobą zbieżne, co pozwala stwierdzić, że wygrzewanie nie wpływa na zniekształcenia harmoniczne.



Rysunek 6 Poziom zniekształceń 2 harmoniczej badanych zestawów



Rysunek 7 Poziom zniekształceń 3 harmoniczej badanych zestawów



Rysunek 8 Całkowity poziom zniekształceń harmonicznych

4. POMIARY SUBIEKTYWNE

4.1 MATERIAŁ TESTOWY

Materiał testowy składał się z fragmentów utworów z trzech gatunków muzycznych:

pop: „Alone” Halsey (0:16-0:31),

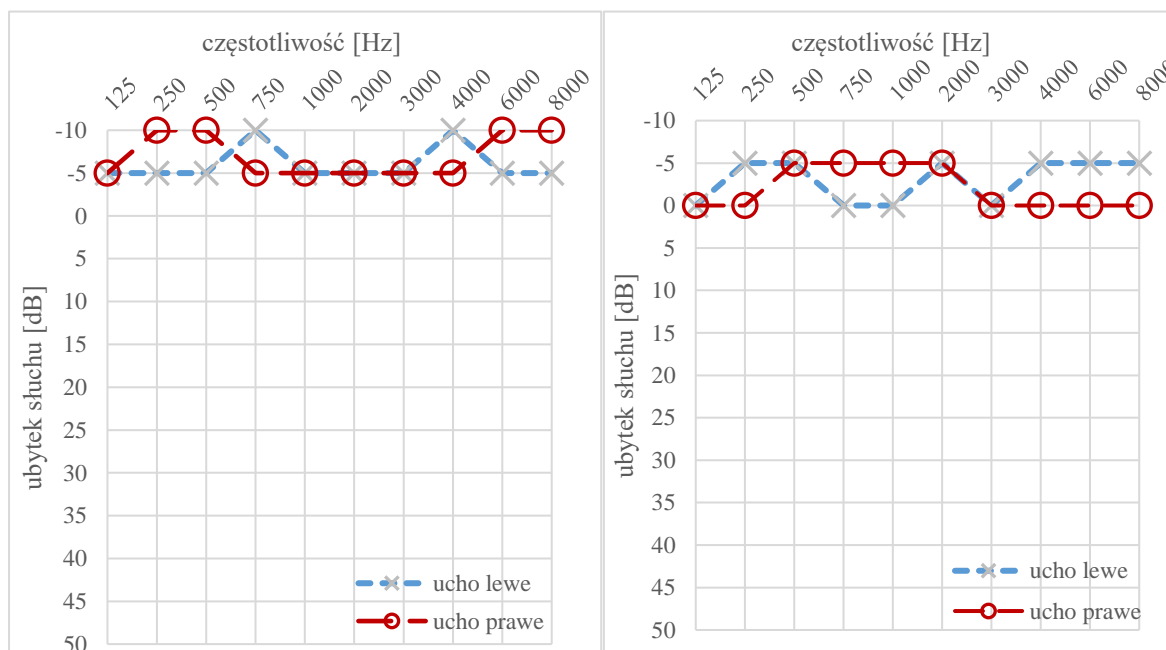
jazz: „Tonight” Oskar’a Peterson’a (0:32-0:47)

rock: „Teardrops” wykonane przez Bring Me The Horizon (0:09-0:24).

Czas trwania każdej z próbek wynosił 15 sekund (ITU, ITU-R BS 562-3), (ITU, ITU-R BS 1284-2), (ITU-T, ITU-T P.913).

4.2 GRUPA BADAWCZA

W badaniach udział brała 11-osobowa grupa ekspercka. Przed przystąpieniem do badań przeprowadzono badania audiometryczne każdego ze słuchaczy. Na rys. 7 przedstawiono dwa z otrzymanych wyników.



Rysunek 9 Wyniki badań audiometrycznych

4.3 SPOSÓB OCENY

Ocenianymi cechami wrażeniowymi były: przejrzystość, dynamika, jasność brzmienia, ocena ogólna.

Do przeprowadzenia oceny w badaniach wykorzystano metodą porównawczą o 7-stopniowej skali CCR (Comparison Category Rating) (ITU, ITU-T P.800). Skalę punktową przedstawiono w tabeli 1.

Tabela 1 Skala ocen

skala ocen	przejrzystość	dynamika	jasność brzmienia	ocena ogólna
+3	dużo bardziej przejrzysty	dużo lepsza	dużo lepsza	dużo lepsza
+2	bardziej przejrzysty	lepsza	lepsza	lepsza
+1	nieznacznie bardziej przejrzysty	nieznacznie lepsza	nieznacznie lepsza	nieznacznie lepsza
0	taki sam	taka sama	taka sama	taka sama
-1	nieznacznie mniej przejrzysty	nieznacznie gorsza	nieznacznie gorsza	nieznacznie gorsza
-2	mniej przejrzysty	gorsza	gorsza	gorsza
-3	dużo mniej przejrzysty	dużo gorsza	dużo gorsza	dużo gorsza

4.4 PRZEPROWADZENIE BADAŃ

W celu przeprowadzenia szczegółowej analizy badanie podzielono na trzy eksperymenty, w których porównywano ze sobą urządzenia:

- nowe (A) z użytkowanymi (B)
- nowe (A) z wygrzewanymi (B)
- użytkowane (A) z wygrzewanymi (B)

Wszystkie testy przeprowadzono w komorze akustycznej na terenie Politechniki Wrocławskiej, spełniającej zalecenia (ITU, ITU-R BS.1116-3, 2015) (AES, 1996). Wykorzystywane zestawy głośnikowe zostały ustawione w odległości czterech metrów. Następnie wyznaczono punkt odsłuchowy na podstawie dokumentów (EBU, EBU Tech. 3276), (EBU, EBU 3286-1997, 1997). Poziom dźwięku, zmierzony miernikiem klasy I, w punkcie referencyjnym wynosił 85 dB.

Testy polegały na ocenie porównawczej, gdzie próbki dźwiękowe były odtwarzane w czterech sekwencjach A-A, B-B, A-B, B-A, gdzie sekwencje A-A oraz B-B to sekwencje kontrolne. Przerwa pomiędzy poszczególnymi próbkami w sekwencji pomiarowej wynosiła 5 sekund, a pomiędzy kolejnymi sekwencjami 15 sekund, w czasie których słuchacze dokonywali oceny (ITU-T, ITU-T P.913, 2014). Przed kolejnymi eksperymentami organizowano dla słuchaczy 20 minutową przerwę.



Rysunek 10 Zdjęcie grupy eksperckiej oraz miejsca ustawienia badanych zestawów

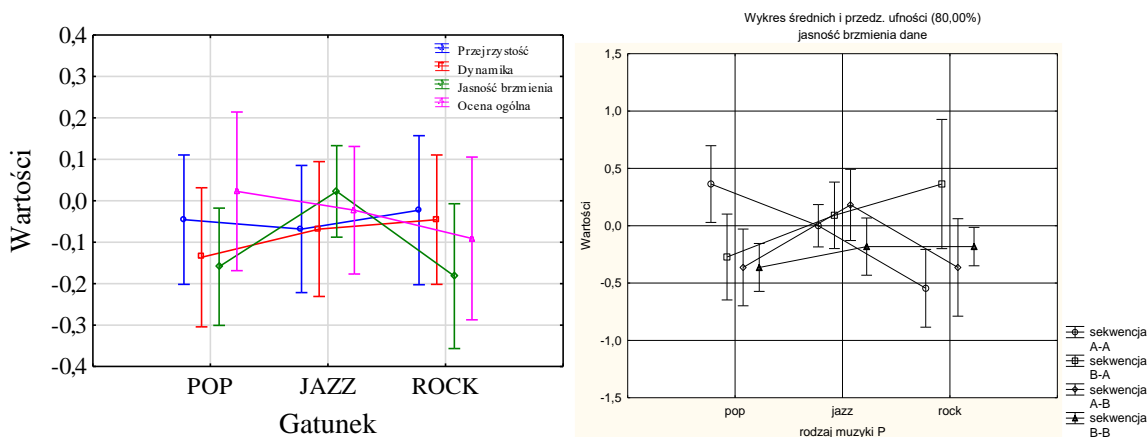
5. ANALIZA OTRZYMANYCH WYNIKÓW

W celu przeprowadzenia analizy wyników przeprowadzono poszczególne miary oraz testy statystyczne, które przedstawiono w postaci tabel i wykresów. Do zobrazowania różnic pomiędzy sekwencjami wyznaczono średnią dla poszczególnych cech wrażeńowych i gatunków muzycznych (rysunki 9-11)

Do szczegółowej analizy przeprowadzono dwa rodzaje testów statystycznych: ANOVA (tab. 3, 5, 7) oraz Levene'a (tab. 2, 4, 6). Pierwszy z nich określa czy wartości średnie poszczególnych zbiorów przyjmują taką samą wartość. Test Levene wyznaczany jest do sprawdzenia jednorodności wariancji. Hipoteza zerowa przyjmuje, iż te wartości są sobie równe. Poziom istotności, czyli wartość poziomu statystycznego, dla obu testów statystycznych wyznaczono na $\alpha=0,05$.

Wszystkie testy statystyczne oraz wykresy zostały wykonane przy użyciu programu STATISTICA 13.3 (Ejaz S. Ahmed), (Breitsohl), (Dario Basso), (Jeffrey N. Rouder).

5.1 ANALIZA DLA PIERWSZEGO POMIARU



Rysunek 11 Wykres średnich z oznaczonym przedziałem ufności 75% dla nowego i używanego urządzenia

Pierwszy wykres przedstawia różnice dla poszczególnych gatunków muzycznych. Słuchacze oceniali próbki muzyki rockowej podobnie we wszystkich cechach wrażeńowych – w przypadku przejrzystości średnia wynosi 0. W przypadku pop'u oceny są nieznacznie gorsze, jednak dla dynamiki otrzymano średnią powyżej 0. Muzyka jazz charakteryzuje się większymi rozbieżnościami w ocenie między poszczególnymi cechami wrażeńowymi. Wszystkie rozbieżności są nieznaczne, a przedział ufności nie przekracza wartości 0,2.

Wykres przedstawiający różnice średnich dla sekwencji są rozbieżne dla poszczególnych układów, jednak ciąg testowy B – B jest skupiony w wartości 0 dla wszystkich ocenianych parametrów. Widoczne rozbieżności są jednak nieznaczne – mogą one wynikać ze stanu emocjonalnego słuchaczy bądź wynikać z nieświadomego oczekiwania różnicy.

Przeprowadzone testy statystyczne wykazały brak istotności statystycznej. To wiąże się z tym, że wariancje są homogeniczne, a różnice w średnich nie występują.

Tabela 2 Test Levene'a jednorodności wariancji dla nowego i używanego urządzenia¹

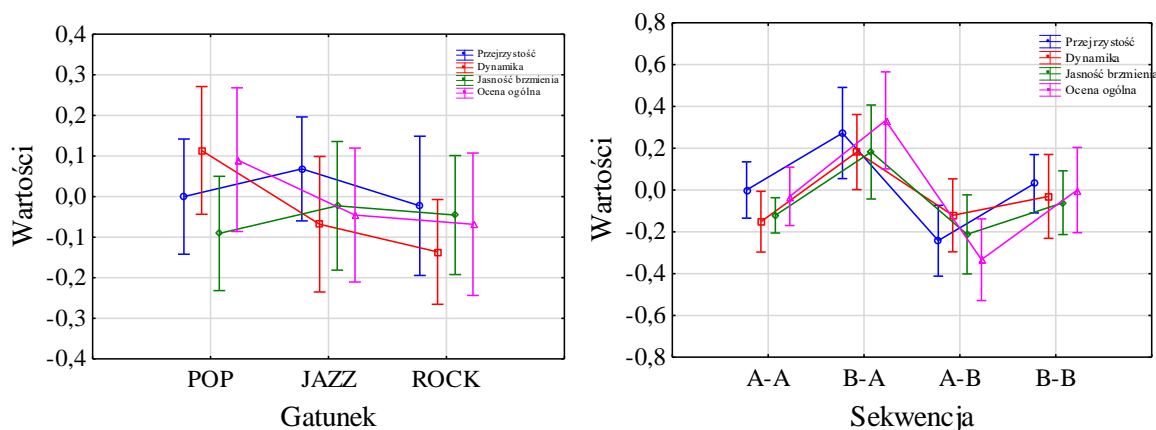
Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Przejrzystość	0,516	2	0,258	65,930	129	0,511	0,505	0,605
Dynamika	0,317	2	0,159	56,057	129	0,435	0,365	0,695
Jasność brzmienia	1,796	2	0,898	45,091	129	0,350	2,569	0,080
Ocena ogólna	1,705	2	0,853	73,464	129	0,569	1,497	0,228

Tabela 3 Analiza wariancji ANOVA dla nowego i używanego urządzenia

Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Przejrzystość	0,045	2	0,023	111,682	129	0,866	0,026	0,974
Dynamika	0,197	2	0,098	109,886	129	0,852	0,116	0,891
Jasność brzmienia	1,106	2	0,553	87,409	129	0,678	0,816	0,444
Ocena ogólna	0,288	2	0,144	137,591	129	1,067	0,135	0,874

5.2 ANALIZA DLA DRUGIEGO POMIARU

Wykresy średnie ukazują, że wartości te są skupione w wartości zerowej. Między gatunkami muzycznymi nie ma widocznych różnic. W przypadku sekwencji można zauważyć, że nowe urządzenie są oceniane nieznacznie lepiej. Sekwencje testowe skupiają się w wartości zero dla wszystkich cech wrażeńowych, tak jak oczekiwano.



Rysunek 12 Wykres średnich z oznaczonym przedziałem ufności 75% dla nowego i wygrzewanego urządzenia

¹ SS – suma kwadratów pomiędzy grupami (SS Efekt); suma kwadratów w obrębie grupy (SS Błąd)
df – ilość stopni swobody
MS – wariancja odnosząca się do zmienności pomiędzy grupami (MS Efekt); wariancja odnosząca się do zmienności w obrębie grup (MS Błąd)
F – test Fishera
p - prawdopodobieństwo

Tabela 4 Test Levene'a jednorodności wariancji dla nowego i wygrzewanego urządzenia

Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Przejrzystość	0,423	2	0,211	52,275	129	0,405	0,522	0,595
Dynamika	1,026	2	0,513	44,161	129	0,342	1,499	0,227
Jasność brzmienia	0,138	2	0,069	54,391	129	0,427	0,163	0,850
Ocena ogólna	0,497	2	0,248	56,775	129	0,440	0,565	0,570

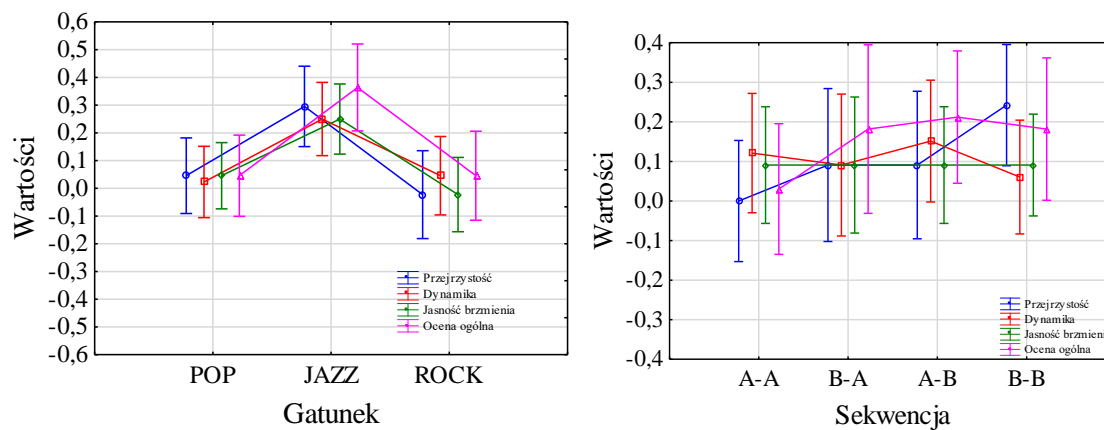
Tabela 5 Analiza wariancji ANOVA dla nowego i wygrzewanego urządzenia

Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Przejrzystość	0,197	2	0,098	91,773	129	0,711	0,138	0,871
Dynamika	1,470	2	0,735	96,409	129	0,747	0,983	0,377
Jasność brzmienia	0,106	2	0,053	92,523	129	0,717	0,074	0,929
Ocena ogólna	0,652	2	0,326	124,341	129	0,964	0,338	0,714

Przeprowadzone testy statystyczne wykazały brak istotności statystycznej. To wiąże się z tym, że wariancje są homogeniczne, a różnice w średnich nie występują

5.3 ANALIZA DLA TRZECIEGO POMIARU

Wykresy średnie ukazują, że wartości te są skupione w wartości zerowej. Między sekwencjami nie ma widocznych różnic. Muzyka jazz nieznacznie odbiega od pozostałych gatunków muzycznych, jednak są to niewielkie wartości.



Rysunek 13 Wykres średnich z oznaczonym przedziałem ufności 75% dla używanego i wygrzewanego urządzenia

Przeprowadzone testy statystyczne wykazały brak istotności statystycznej. To wiąże się z tym, że wariancje są homogeniczne, a różnice w średnich nie występują.

Analiza statystyczna wyników ukazuje, że w przypadku wszystkich testów urządzenia głośnikowe oceniano jednorodnie, a wartości średnie są jednakowe. Otrzymane wykresy wskazują na to, że średnie są skupione w wartości zero i nie przekraczają one różnicy 0,5. Nie znaleziono różnic między ocenami gatunków muzycznych – nie występują bądź są one nieznaczące.

Rozbieżności mogą wynikać z podświadomej chęci wykazania różnic, a wykazane odchyłki efektem złudnego postrzegania różnic.

Tabela 6 Test Levene'a jednorodności wariancji dla używanego i wygrzewanego urządzenia

Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Przejrzystość	0,721	2	0,361	44,142	129	0,342	1,054	0,351
Dynamika	0,529	2	0,264	37,051	129	0,287	0,920	0,401
Jasność brzmienia	0,601	2	0,301	36,119	129	0,280	1,073	0,345
Ocena ogólna	0,642	2	0,321	45,903	129	0,356	0,902	0,408

Tabela 7 Analiza wariancji ANOVA dla używanego i wygrzewanego urządzenia

Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Przejrzystość	2,470	2	1,235	90,046	129	0,698	1,769	0,175
Dynamika	1,379	2	0,689	75,136	129	0,582	1,184	0,309
Jasność brzmienia	1,773	2	0,886	67,136	129	0,520	1,703	0,186
Ocena ogólna	2,970	2	1,485	100,000	129	0,775	1,915	0,151

6. WNIOSKI

Badania obiektywne wykazały występowanie nieznaczących różnic w charakterystykach modułu impedancji analizowanych zestawów. Wynika to ze zwiększenia podatności zawieszonych wykorzystanych przetworników. Pomiarów charakterystyk częstotliwościowych oraz modułów impedancji badanych zestawów wykazały rozbieżności wynikające z procesów technologicznych na liniach montażowych zarówno samych przetworników jak i kompletnych zestawów głośnikowych.

Zmiany parametrów przetworników pod wpływem ich pracy nie powinny różnić się o więcej niż 10% do 15% od parametrów początkowych. Różnice te mierzy się wykonując szereg pomiarów nowego przetwornika, a następnie poddając go pracy przez 24 h z sygnałem typu program skorygowany do pasma przenoszenia badanego przetwornika o napięciu odpowiadającym mocy znamionowej przetwornika. Po ww. czasie przetwornik ponownie poddaje się analogicznym testom i określa się wartości zmian uzyskanych w obu seriach pomiarowych. Jeżeli zmiany są większe niż te zakładane przez normy (+/- 10% - 15%) cała partia produkcyjna przetworników uznawana jest za wadliwą.

Testy subiektywne nie wykazały istotnych różnic pomiędzy badanymi parami zestawów głośnikowych. Potwierdza to badania obiektywne, które wykazały większe różnice pomiędzy zestawami w jednej partii produkcyjnej w stosunku do zmian wprowadzanych przez wygrzewanie. Grupa ekspercka jednoznacznie wskazała na brak zmian w przejrzystości, dynamice, jasności brzmienia czy ocenie ogólnej badanych przewodów zarówno dla gatunków muzycznych jak i sekwencji.

Potwierdzeniem braku wpływu wygrzewania na odczucia subiektywne słuchaczy może być fakt wykazywania zmian w sekwencjach testowych A – A oraz B – B, przy jednoczesnym wskazaniu braku zmian dla sekwencji A – B oraz B – A. Zmiany te mogą być spowodowane podświadomą chęcią wykazania różnic oraz stanu emocjonalnego grupy eksperckiej.

Wygrzewanie zestawów głośnikowych przynosi niewielkie zmiany w parametrach przetworników, co wykazały testy obiektywne (zmiana podatności zawieszonych przetworników), jednakże nie ma wpływu na parametry obiektywne. Większe rozbieżności wykazano dla urządzeń głośnikowych pochodzących z jednej serii produkcyjnej niż dla zestawów nowych, używanych i wygrzewanych.

LITERATURA

- [1] AES. (1996). AES20-1996. AES.
- [2] Breitsohl, H. (n.d.). Beyond ANOVA: An Introduction to Structural Equation Models for Experimental Designs.
- [3] Bruel & Kjaer. (1986). Automatic Quality Testing of Loudspeaker Electroacoustic : Performance. Bruel & Kjaer.
- [4] Dario Basso, L. S. (n.d.). Permutation Tests for Stochastic Ordering and ANOVA.
- [5] EBU. (1997). EBU 3286-1997. EBU.
- [6] EBU. (n.d.). EBU Tech. 3276. EBU.
- [7] Ejaz S. Ahmed, E. R. (n.d.). Absolute Penalty Estimation.
- [8] ITU. (2015). ITU-R BS.1116-3. ITU.
- [9] ITU. (n.d.). ITU-R BS 1284-2. ITU.
- [10] ITU. (n.d.). ITU-R BS 562-3. ITU.
- [11] ITU. (n.d.). ITU-T P.800. ITU.
- [12] ITU-T. (2014). ITU-T P.913. ITU.
- [13] ITU-T. (n.d.). ITU-T P.913. 2014: ITU.
- [14] Jeffrey N. Rouder, C. R. (n.d.). Model comparison in ANOVA.
- [15] PN-EN 60268-5:2021, Urządzenia systemów elektroakustycznych. Część 5: Głośniki. (2021). Polski Komitet Normalizacyjny.
- [16] Sanfilippo, M. (2005). Speaker Break In: Fact or Fiction? Audioholics.
- [17] Society, A. E. (2023). AES75-2023. AES.
- [18] Winer, E. (2018). The Audio Expert: Everything You Need to Know About Audio. Second Edition. Routledge.

THE EFFECT OF TRANSDUCER ANNEALING ON THE SUBJECTIVE AND OBJECTIVE PARAMETERS OF LOUSPEAKERS.

This article presents the results of objective and subjective research into the influence of loudspeaker burn-in on objective parameters of loudspeakers such as impedance, frequency response, harmonic distortion and subjective parameters based on the evaluation of an expert group. The tests were carried out using new, Music Noise annealed and used loudspeakers. Subjective tests were carried out using the CCR method on an expert group of 11 people. In the objective and subjective tests, no differences were observed between the loudspeakers tested due to the annealing process. It can therefore be concluded that the loudspeaker burn-in process does not affect the performance of the loudspeakers. Converging conclusions can be drawn from the subjective measurements. Levene's test showed that the ratings were homogeneous, while ANOVA analysis did not show statistical significance. Only slight differences were observed in the results for musical genres.

SYNTEZA DŹWIĘKU PRZESTRZENNEGO Z WYKORZYSTANIEM ZMODYFIKOWANEJ BAZY HRTF POMIARÓW WYKONANYCH NA POLITECHNICIE WROCŁAWSKIEJ

Zbigniew ŚWIĘTACH, Przemysław PLASKOTA

Katedra Akustyki, Multimediiów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki
i Mikrosystemów, Politechnika Wrocławska

Autor korespondencyjny: **Zbigniew Świętach**, zbigniew.swietach@pwr.edu.pl

W ostatnich latach rośnie zainteresowanie pomiarami Head Related Transfer Functions (HRTF), również z uwagi na możliwość zautomatyzowania pomiarów. Automatyzacja pozwala na wykonanie pomiarów z dużą rozdzielczością przestrzenną dla wielu osób w krótkim czasie. Tak wykonane pomiary mogą być zestawione w bazę danych, którą można następnie wykorzystać w dalszych pracach.

W pracy przedstawiono proces syntezy dźwięku przestrzennego z wykorzystaniem wyników pomiarów HRTF wykonanych na Politechnice Wrocławskiej. Pomiary wykonano dla 56 osób, z wykorzystaniem systemu do szybkiego i automatycznego pomiaru funkcji transmitancji odniesionych do głowy. W pracy przedstawiono również proces rekonstrukcji nieprawidłowo zarejestrowanych sygnałów pomiarowych.

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 [2, 3]. Dzięki zmianom w widmie słuchacz jest w stanie zlokalizować położenie źródła dźwięku w przestrzeni wokół niego. Ponieważ istnieje wiele położen źródła dźwięku w przestrzeni otaczającej słuchacza, istnieje wiele HRTF o kształcie zależnym od położenia źródła dźwięku w tej przestrzeni. Z drugiej strony, znając przebieg HRTF dla różnych położen źródła dźwięku względem słuchacza, możliwe jest stworzenie wrażenia przestrzeni dźwiękowej dla zdarzeń dźwiękowych, które nie zawierają informacji o położeniu w przestrzeni.

Każdy człowiek posiada indywidualny kształt zarówno małżowiny usznej, jak i głowy. W ogólności kształty te są zbliżone u różnych osób, jednak w szczegółach różnice są dość istotne, zwłaszcza jeśli chodzi o przebieg funkcji HRTF [1]. Oznacza to, że każdy człowiek posiada indywidualny zbiór HRTF. Zatem w sytuacji, 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 [4].

2. PROCEDURA POMIAROWA

Pomiar HRTF wykonano z wykorzystaniem metody transmitacyjnej, która jest popularna w cyfrowych systemach pomiarowych [1], [2], [4]. Oprogramowanie pozwala na wykorzystanie wielu sygnałów pomiarowych o szerokim widmie. Długość generowanego sygnału można zmieniać w zakresie od 128 do 8192 próbek, podczas pomiarów używano sygnału o długości 1024 próbek.

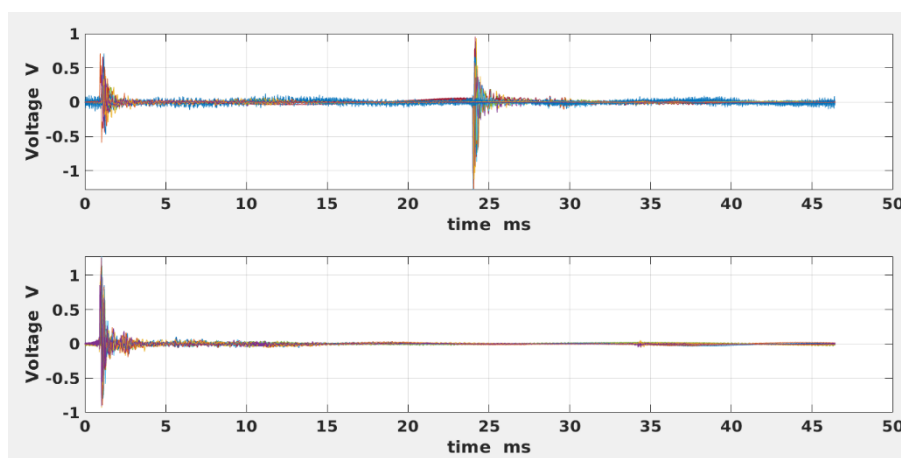
Domyślna częstotliwość próbkowania wynosi 48 kHz. Sygnał pomiarowy jest powtarzany kilka razy, w celu uśrednienia odpowiedzi systemu w dziedzinie czasu, aby w ten sposób poprawić stosunek sygnału do szumu mierzonych sygnałów. Ponieważ odpowiedzi określone w systemie będą wykorzystywane do splotu z rzeczywistymi sygnałami, nie mogą być zbyt długie. Ponadto dłuższe sygnały pomiarowe wydłużają czas oceny.

Procedura pomiarowa składa się z dwóch etapów: 1) pomiaru odpowiedzi referencyjnych i 2) pomiaru standardowych HRTF [5]. Pomiar referencyjny jest wykonywany dla wszystkich punktów pomiarowych określonych przez operatora systemu. Podczas tej procedury mikrofony, głośniki i cały system działają dokładnie tak samo, jak podczas każdego standardowego pomiaru, jedyną różnicą jest brak słuchacza na krześle. Zarówno pomiar HRTF, jak i pomiar referencyjny są przechowywane jako złożone widma. Ostateczne HRTF są obliczane poprzez podzielenie widma słuchacza przez referencyjne widmo „bez słuchacza”. Wykorzystując odpowiedź referencyjną dla każdego punktu pomiarowego, można ograniczyć wiele negatywnych efektów, które obniżają dokładność pomiaru [6]. W szczególności eliminowany jest wpływ charakterystyki częstotliwościowej i kierunkowości głośników oraz mikrofonów. Znacznie zmniejsza się również wpływ pomieszczenia testowego i odbić od elementów urządzenia. Metoda ta różni się od metody wyrównywania w polu swobodnym zaproponowanej np. przez Blauerta [7] lub Møllera i in. [2], ponieważ wykorzystywali oni tylko jeden punkt odniesienia do wyrównywania wszystkich pomiarów. Według najlepszej wiedzy autorów, proponowana metoda stanowi nowatorskie podejście do pomiarów HRTF.

Bezpośrednim wynikiem procesu pomiarowego nie są w rzeczywistości HRTF, ale impulsowe odpowiedzi związane z głową (HRIR). Uzyskane HRIR są splatane z sygnałami źródłowymi w celu wytworzenia dźwięku przestrzennego. HRTF są obliczane jako transformaty Fouriera HRIR.

3. SYNTEZA DŹWIĘKU PRZESTRZENNEGO

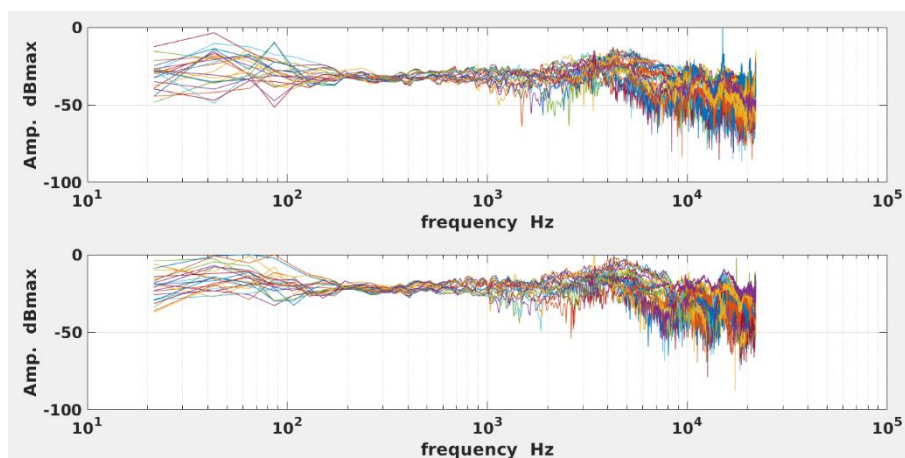
Baza indywidualnych pomiarów własnych HRIR PWr obejmuje pomiary 51 osób o numerach 5 do 56 (cztery pierwsze pomiary okazały się całkowicie nieprawidłowe i zostały usunięte z bazy). Dla każdego użytkownika wykonano 24 pomiary przy azymucie zmieniającym się co 15 deg i 11 pomiarów przy elewacji zmieniającej się co 9 deg, razem daje to liczbę 264 pomiarów dla jednej osoby. Częstotliwość próbkowania w trakcie pomiarów wynosiła 44100 Hz. Sygnałem testowym był chirp wykładniczy o odpowiednio dobranych parametrach. Na rys. 1 pokazano HRIR dla słuchacza 56, przy czym elewacja równa jest 9 deg, a azymut zmienia się od 0, 15, 30, ..., 165 deg oraz od -180, -165, ..., -15 deg.



Rys. 1 Pomiary HRIR (dla ucha lewego i prawego), baza PWr, użytkownik 56.

Na rysunku widać, że na pomiary nakładają się zakłócenia, które przypuszczalnie są przydźwiękami pochodzącymi od sieci energetycznej, ponadto dla azymutów dodatnich w kanale lewym występuje nienaturalne opóźnienie odpowiedzi impulsowych.

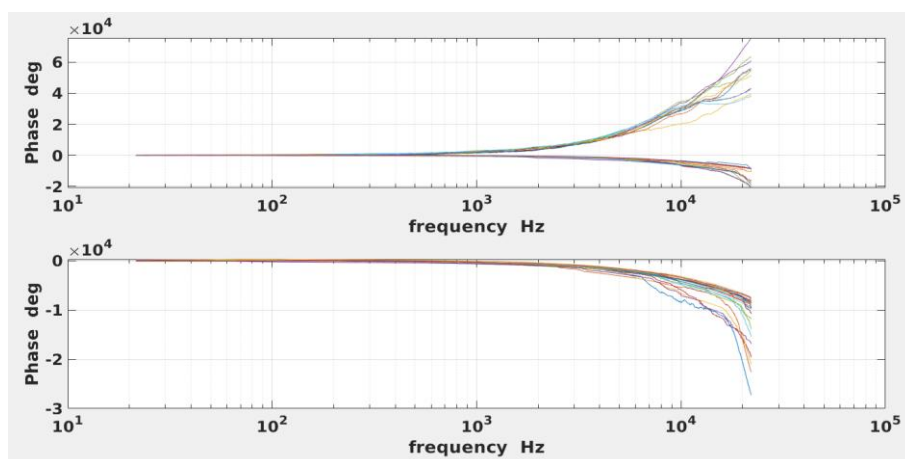
Na rys. 2 i rys. 3 pokazano odpowiednio charakterystyki amplitudowe i fazowe HRTF otrzymane na podstawie HRIR z rys. 1. HRTF obliczono jako DFT (dyskretna transformata Fouriera) z HRIR.



Rys. 2 Charakterystyki amplitudowe HRTF wyznaczone na podstawie HRIR (dla ucha lewego i prawego), baza PWr, użytkownik 56.

W wykreślonych widmach amplitudowych występują zakłócenia o częstotliwościach około 60 i 90 Hz. Przypuszczalnie pochodzą one od nie do końca poprawnie przeprowadzonego ekranowania. Na wykresach widać również, że charakterystyki amplitudowe w sąsiedztwie częstotliwości Nyquista $f_q = f_s/2$, gdzie $f_s = 44100$ Hz, przyjmują istotne wartości, których nie można zaniedbać, co sugeruje, że w trakcie pomiarów dobrano niepoprawnie analogowe filtry antyaliasingowe. Ww. przypuszczenia zostaną zweryfikowane później.

Na wykresach w kanale prawym widać ponadto zakłócenia wąskopasmowe w otoczeniu częstotliwości około 4.5 kHz i 18 kHz. Na podstawie wyników badań nad fizjologią słuchu przyjmuje się, że przestrzenna lokalizacja źródeł dźwięku jest możliwa w zakresie częstotliwości od kilkuset Hz do około 8 kHz. Tym samym zakłócenia powyżej 8 kHz można odfiltrować za pomocą filtru liniowego, natomiast zakłócenie wąskopasmowe o częstotliwości 4.5 kHz można usunąć jedynie przez ręczną korektę charakterystyk częstotliwościowych w otoczeniu ww. częstotliwości.



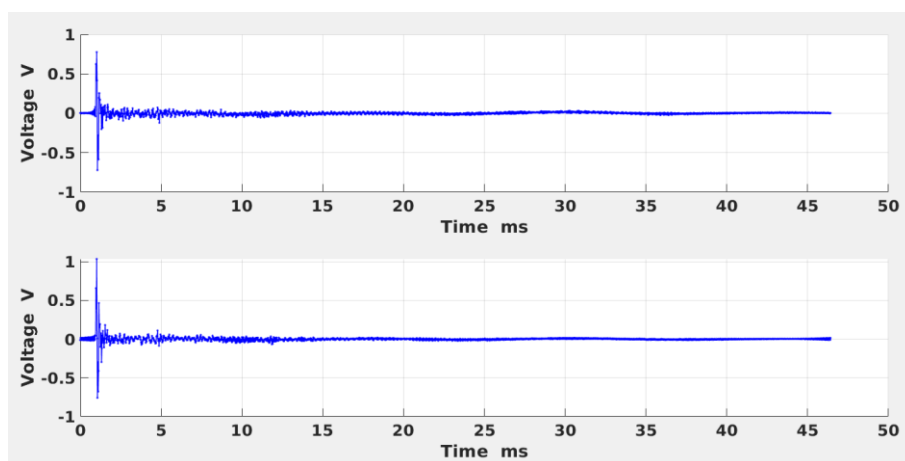
Rys. 3 Charakterystyki fazowe HRTF wyznaczone na podstawie HRIR (dla ucha lewego i prawego), baza PWr, użytkownik 56.

Na wykresach widać, że dla azymutów dodatnich w kanale lewym charakterystyka fazowa jest niepoprawna. Taki przebieg charakterystyk fazowych sugeruje ujemny czas opóźnienia grupowego od około 10 kHz aż do częstotliwości Nyquista, a to implikuje brak przyczynowości odpowiedzi impulsowej.

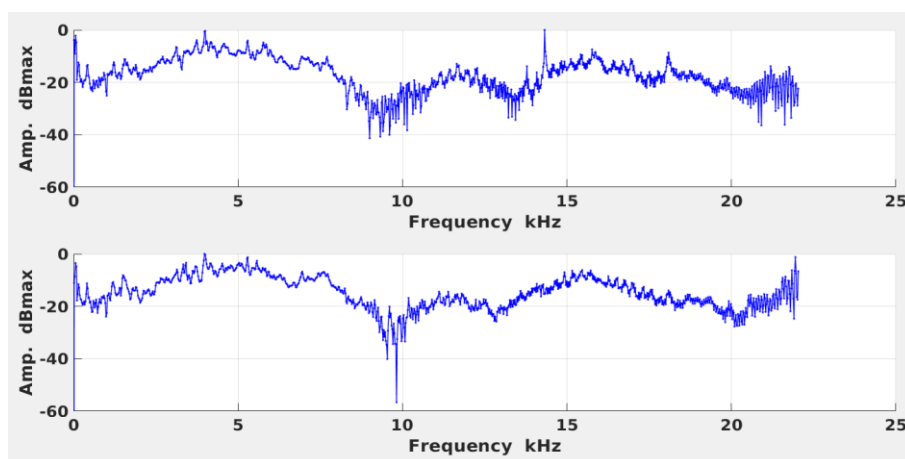
Baza pomiarów HRIR Pwr wymaga korekty. Niestety błędów spowodowanych aliasingiem nie można wyeliminować metodami post-processingowymi. Należy zatem wykonać wszystkie inne konieczne korekty bazy HRIR i na koniec sprawdzić, czy efekt dźwięku przestrzennego jest dla słuchacza zadowalający.

Dla użytkownika o numerze 56 zostały sprawdzone wszystkie HRIR, a zauważone nieprawidłowości zanotowano. Podobnie postąpiono jeszcze z kilkoma losowo wybranymi użytkownikami rozważanej bazy pomiarów. W każdym przypadku zaobserwowano podobne nieprawidłowości, które zostaną omówione poniżej, na przykładzie pomiarów wykonanych dla użytkownika 56.

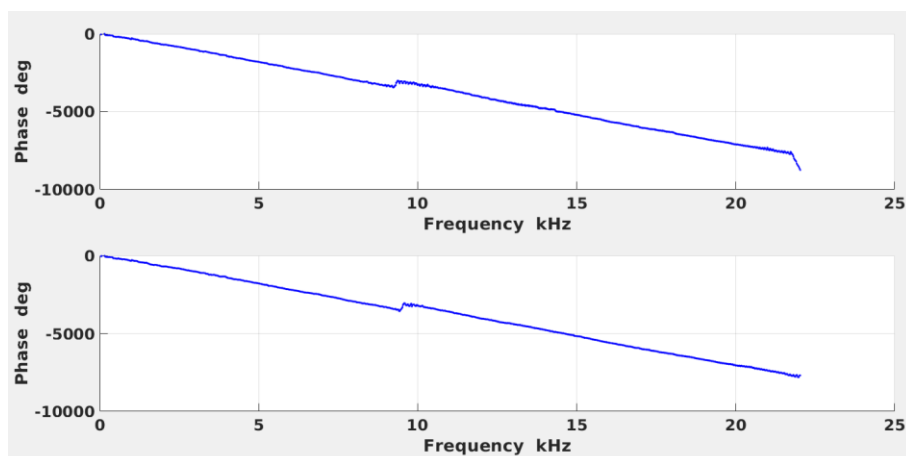
Na rys. 4 przedstawiono HRIR dla azymutu 0 i elewacji 45 deg. Widać przydźwięk pochodzący od nieprawidłowego ekranowania. Na rys. 5 oraz rys. 6 zamieszczono charakterystyki amplitudowe i fazowe HRTF wyznaczone jako DFT z ww. HRIR. W obydwu kanałach można zaobserwować przydźwięk sieciowy, pochodzący przypuszczalnie od nieprawidłowego ekranowania. Ponadto w kanale lewym widać wąskopasmowe zakłócenie przy częstotliwości około 14 kHz.



Rys. 4 HRIR (dla ucha lewego i prawego), użytkownik 56, elewacja 0, azymut 45 deg.



Rys. 5 HRTF (dla ucha lewego i prawego) charakterystyka amplitudowa, użytkownik 56, elewacja 0, azymut 45 deg.

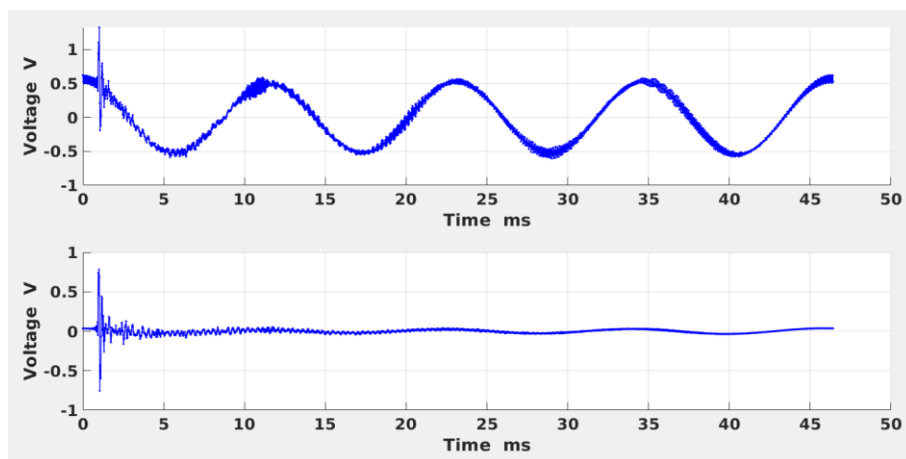


Rys. 6 HRTF (dla ucha lewego i prawego) charakterystyka fazowa, użytkownik 56, elewacja 0, azymut 45 deg.

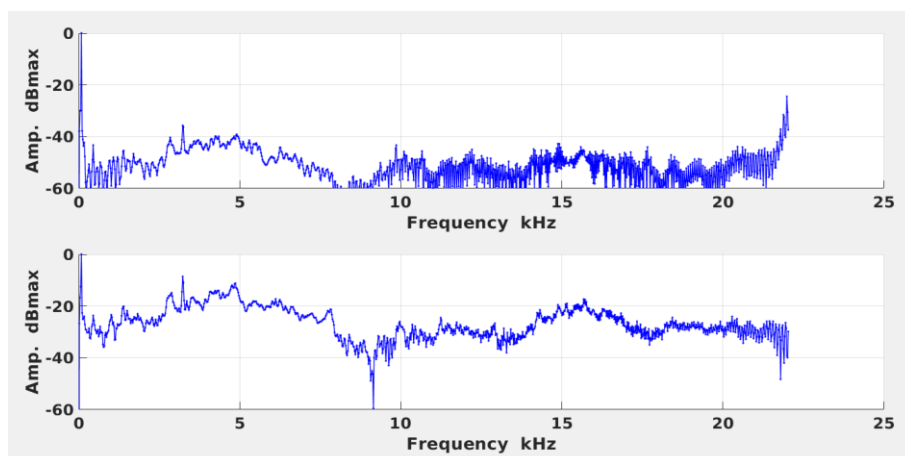
Na wykresach widać skok fazy w okolicy 9 kHz. Przypuszczalnie jest to spowodowane przełączeniem generatora sygnałowego DDS na nowy zakres syntezy częstotliwości. Ten typ nieprawidłowości pomiarów jest typowy dla części pomiarów HRIR bazy Pwr.

Na rys. 7 przedstawiono HRIR dla azymutu 0 i elewacji 27 deg. Widać olbrzymi przydźwięk pochodzący od nieprawidłowego ekranowania w kanale lewym i duży przydźwięk w kanale prawym. Na rys. 8 oraz rys. 9 zamieszczono charakterystyki amplitudowe i fazowe HRTF wyznaczone jako DFT z ww. HRIR. W obydwu kanałach można zaobserwować przydźwięk sieciowy, pochodzący przypuszczalnie od nieprawidłowego ekranowania. W lewym kanale przydźwięk sieciowy jest dominujący.

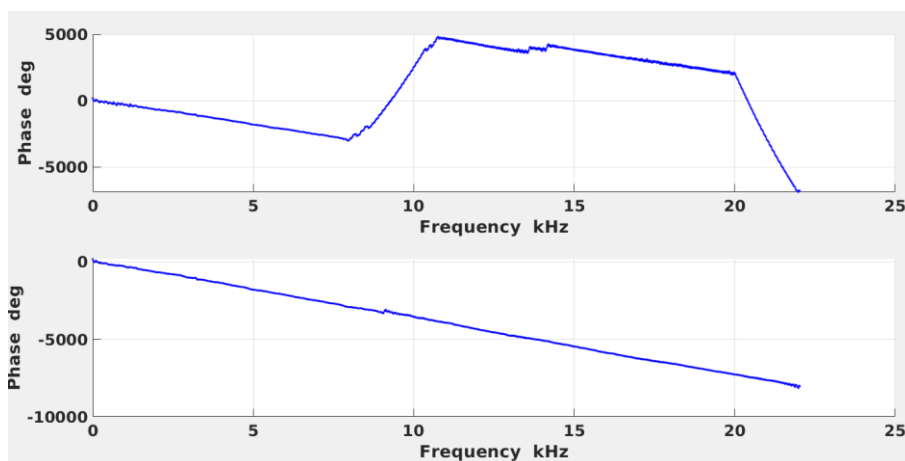
Na wykresach widać skok fazy w okolicy 9 kHz. Przypuszczalnie jest to spowodowane przełączeniem generatora sygnałowego DDS na nowy zakres syntezy częstotliwości. Ponadto widać w kanale lewym niepoprawny wykres charakterystyki fazowej. Widoczne „podbicie” charakterystyki fazowej jest konsekwencją złego ekranowania i przeniku zakłóceń sieciowych do sygnału pomiarowego. Ten typ nieprawidłowości dość często występuje w pomiarach HRIR bazy Pwr.



Rys. 7 HRIR (dla ucha lewego i prawego), użytkownik 56, elewacja 0, azymut 27 deg.



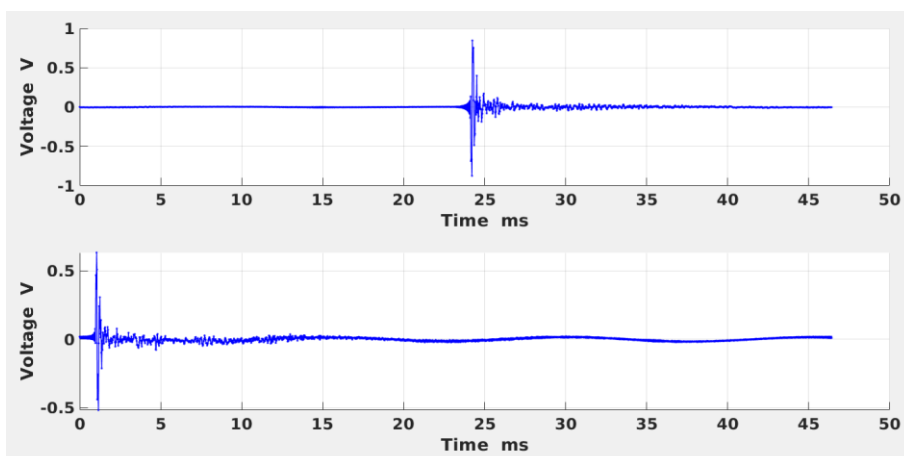
Rys. 8 HRTF (dla ucha lewego i prawego) charakterystyka amplitudowa, użytkownik 56, elewacja 0, azymut 27 deg.



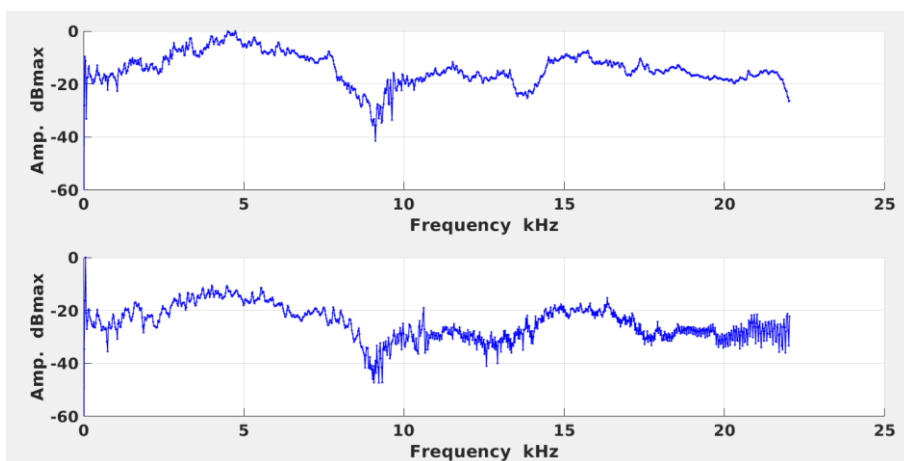
Rys. 9 HRTF (dla ucha lewego i prawego) charakterystyka fazowa, użytkownik 56, elewacja 0, azymut 27 deg.

Na rys. 10 przedstawiono HRIR dla azymutu 15 deg i elewacji 36 deg. Widać nienaturalne opóźnienie HRIR w kanale lewym oraz przydźwięk sieciowy w kanale prawym. Na rys. 11 oraz rys. 12 zamieszczono charakterystyki amplitudowe i fazowe HRTF wyznaczone jako DFT z ww. HRIR. W kanale prawym można zaobserwować przydźwięk sieciowy, pochodzący przypuszczalnie od nieprawidłowego ekranowania. W lewym kanale poziom przydźwięku sieciowego jest mniejszy.

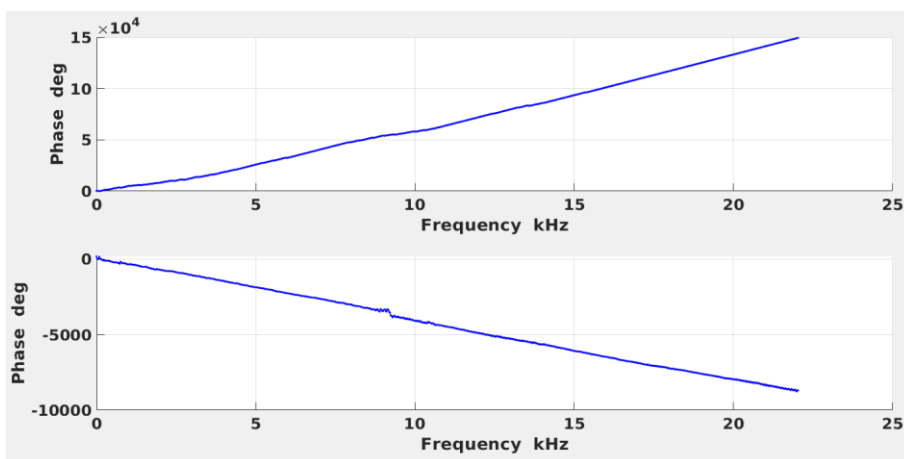
Na wykresie kanału prawego widać skok fazy w okolicy 9 kHz. Przypuszczalnie jest to spowodowane przełączeniem generatora sygnałowego DDS na nowy zakres syntezowanej częstotliwości. W kanale lewym widać niepoprawny wykres charakterystyki fazowej. Taka charakterystyka fazowa implikuje ujemny czas opóźnienia grupowego w całym zakresie częstotliwości od zera do $f_s/2$. W konsekwencji HRTF w kanale lewym jest nieprzyczynowe i tym samym niefizyczne. To jest błąd kardynalny bazy HRIR PWr, który bezwzględnie musi zostać naprawiony. Bez naprawienia ww. błędu baza HRTF PWr jest nieprzydatna do realizacji dźwięku przestrzennego. Ten typ nieprawidłowości niestety występuje w wielu pomiarach HRIR bazy Pwr.



Rys. 10 HRIR (dla ucha lewego i prawego), użytkownik 56, elewacja 36 deg, azymut 15 deg.



Rys. 11 HRTF (dla ucha lewego i prawego) charakterystyka amplitudowa, użytkownik 56, elewacja 36 deg, azymut 15 deg.



Rys. 12 HRTF (dla ucha lewego i prawego) charakterystyka fazowa, użytkownik 56, elewacja 15 deg, azymut 36 deg.

Naprawę bazy HRIR PW_r rozpoczęto od eliminacji nienaturalnych opóźnień, rys. 10. Po żmudnej analizie HRIR użytkownika 56 i kilku innych użytkowników, stwierdzono, że ww. opóźnienia są wynikiem błędu programowania. Rozwiązanie problemu okazało się proste. Liczba próbek HRIR jest równa 2048. Naprawa bazy HRIR ww. przypadkach polegała na zamianie kolejności próbek. Próbkę o indeksach 1024-2048 zostały przesunięte na początek ciągu próbek, natomiast próbki o indeksach 0-2047 zostały przesunięte na koniec ciągu próbek. Okazało się jednak, że charakterystyki fazowe HRTF dla zmodyfikowanych ciągów próbek HRIR są nadal istotnie różne od charakterystyk fazowych dla ciągów próbek niezmodyfikowanych. Ponownie po przeprowadzeniu analizy charakterystyk fazowych zidentyfikowano przyczynę. Po zmianie znaku zmodyfikowanych ciągów próbek, tzn. po pomnożeniu całego ciągu przez -1 otrzymano poprawne charakterystyki fazowe.

Eliminacja przydźwięków sieciowych i wąskopasmowych zakłóceń o częstotliwościach leżących powyżej 8 kHz została przeprowadzona za pomocą filtracji filtrem FIR o zerowej fazie. W takim przypadku zniekształcenia sygnału użytecznego są najmniejsze. Niemniej jednak ww. filtr jest nieprzyczynowy i posiada nieskończoną liczbę próbek. Taki filtr można zastosować jedynie w sytuacji, gdzie wszystkie filtrowane sygnały zostały uprzednio zapisane na nośnik pamięci. Ponadto do filtracji można użyć jedynie skończonej liczby próbek odpowiedzi impulsowej, zatem biorąc M próbek położonych symetrycznie względem zera (M nieparzyste) aproksymuje się w sensie najmniejszych kwadratów nieskończoną w czasie dyskretnym odpowiedź impulsową filtru.

Dla filtru o zerowej fazie transmitancja zespolona na okręgu jednostkowym jest równa jego charakterystyce amplitudowej. Jako charakterystykę amplitudową filtru wybrano okno Tukeya w dziedzinie częstotliwości, równanie (1).

$$H(f) = \begin{cases} 0 & f \in [0, f_1] \\ \sin^2\left(\frac{\pi(f-f_1)}{2(f_2-f_1)}\right) & f \in (f_1, f_2) \\ 1 & f \in [f_2, f_3] \\ \cos^2\left(\frac{\pi(f-f_3)}{2(f_4-f_3)}\right) & f \in (f_3, f_4) \\ 0 & f \in [f_4, f_q] \end{cases} \quad (1)$$

Częstotliwości charakterystyczne projektowanego filtru są równe $f_1=120$ Hz, $f_2=300$ Hz, $f_3=8$ kHz, $f_4=10$ kHz. Częstotliwość Nyquista $f_q = f_s/2$ wynosi $f_q=22.05$ kHz, gdzie $f_s=44.1$ kHz.

Odpowiedź impulsowa filtru o zerowej fazie została wyznaczona analitycznie oraz dla porównania numerycznie. Wyprowadzenie analitycznego wzoru na współczynniki odpowiedzi impulsowej jest koncepcyjnie proste lecz niestety wymaga przeprowadzenia długiego ciągu żmudnych obliczeń. Niżej podano wynik końcowy obliczeń bez wyprowadzeń.

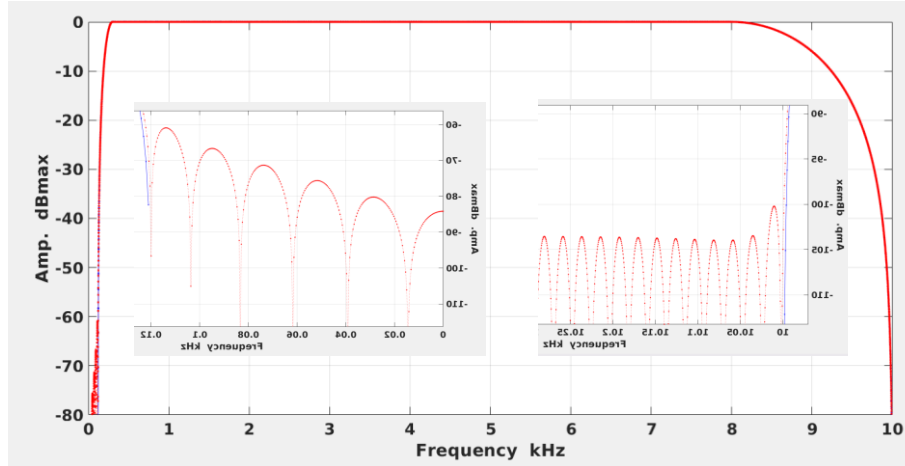
$$h(n) = \frac{2}{f_s} \int_0^{f_s/2} H(f) e^{j2\pi n \frac{f}{f_s}} df = \begin{cases} \frac{f_4+f_3-f_2-f_1}{f_s} & n = 0 \\ h^I(n) - h^{II}(n) & n = 1, 2, \dots \end{cases} \quad (2)$$

$$h^I(n) = \frac{\sin(\pi n \frac{f_2+f_1}{f_s}) \cos(\pi n \frac{f_2-f_1}{f_s})}{\pi n [4 \cdot n^2 (\frac{f_2-f_1}{f_s})^2 - 1]}, \quad h^{II}(n) = \frac{\sin(\pi n \frac{f_4+f_3}{f_s}) \cos(\pi n \frac{f_4-f_3}{f_s})}{\pi n [4 \cdot n^2 (\frac{f_4-f_3}{f_s})^2 - 1]}$$

Odpowiedź impulsowa ww. filtru jest parzysta, tzn. $h(-n) = h(n)$. Eksperymentalnie ustalono, że aproksymacja odpowiedzi impulsowej (2) za pomocą $M=2001$ próbek jest w rozważanym przypadku wręcz wzorowa. Aproksymację transmitancji widmowej (1) przeprowadzono za pomocą DTFT (discrete time Fourier transform), wzór (3).

$$H(f) = \sum_{m=-M_0}^{M_0} h_m e^{-j2\pi m \frac{f}{f_s}} = h_0 + 2 \sum_{m=1}^{M_0} h_m \cos(2\pi \cdot m \frac{f}{f_s}), \quad M_0 = \frac{M-1}{2} \quad (3)$$

Siatka częstotliwości testowych jest dowolna, tutaj przyjęto $f_k = \frac{k}{2K} f_s, k = 0, 1, 2, \dots, K - 1$, gdzie $K = 50021$ jest liczbą pierwszą. Próbki h_m wyznaczono ze wzoru (2). Wykresy charakterystyk amplitudowych obliczonych na podstawie wzorów (1) oraz (3) zamieszczono na rys. 13.



Rys. 13 Charakterystyki amplitudowe projektowanego filtru FIR o zerowej fazie. Częstotliwości charakterystyczne filtru wynoszą $f_1=120$ Hz, $f_2=300$ Hz, $f_3=8$ kHz, $f_4=10$ kHz oraz $f_q=22.05$ kHz, $f_s=44.1$ kHz. Linia niebieska – charakterystyka teoretyczna, wzór (1), linia czerwona – charakterystyka praktyczna, wzór (3).

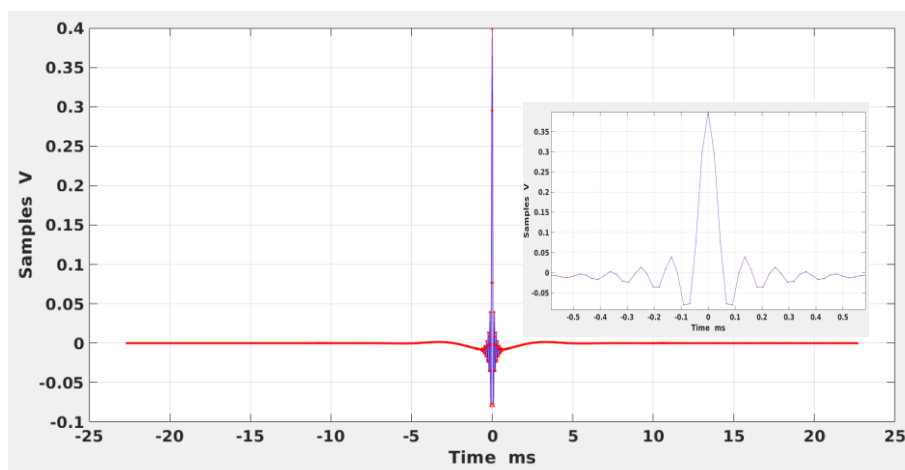
W pasmie przepustowym $f \in (f_2, f_3)$ nierównomierność charakterystyki amplitudowej jest mniejsza niż 0.009 dB. W dolnym pasmie zaporowym $f \in (0, f_1)$ tłumienie jest nie mniejsze niż 60 dB. W górnym pasmie zaporowym $f \in (f_4, f_q)$ tłumienie jest nie mniejsze niż 110 dB. Ponadto w pasmie częstotliwości $f \in (f_1 + \Delta f, f_4 - \Delta f)$, $\Delta f = 10$ Hz charakterystyka fazowa jest równa zero (ten wykres nie jest zamieszczony).

Przeprowadzono również numeryczne obliczenie M współczynników odpowiedzi impulsowej filtru o zerowej fazie za pomocą DFT (Discrete Fourier Transform) realizowanej przy użyciu procedury FFT dostępnej w Matlabie. W tym celu wybrano N częstotliwości rozmieszczonych na równoodległej siatce $f_k = \frac{k}{2N} f_s, k = 0, 1, 2, \dots, N - 1$ i ze wzoru (1) obliczono $H_k = H(f_k)$. Następnie utworzono wektor $\hat{H} = [H_0, H_1, \dots, H_{N-2}, H_{N-1}, 0, H_{N-1}, H_{N-2}, \dots, H_1]$ o $2N$ elementach. Ponieważ $H_k^* = H_k$ (gwiazdka oznacza sprzężenie zespolone), zatem

$$\hat{h}_n = \frac{1}{f_s} \sum_{k=0}^{2N-1} H_k e^{-j2\pi \frac{k \cdot n}{2N}}, n = 0, 1, 2, \dots, 2N - 1 \quad (4)$$

$$h = [h_{2N-M_0}, \dots, h_{2N-1}, \hat{h}_0, \hat{h}_1, \dots, \hat{h}_{M_0}]$$

Do obliczeń przyjęto $N=22050$, wówczas M próbek o indeksach $m = -M_0, -M_0 + 1, \dots, -1, 0, 1, \dots, M_0 - 1, M_0$ wyznaczane za pomocą wzorów (2) oraz (4) są w praktyce identyczne, moduł błędu bezwzględny jest mniejszy niż 8.04 pV. Na rys. 14 przedstawiono wykresy odpowiedzi impulsowych, otrzymanych za pomocą wzorów (2) i (4).



Rys. 14 Odpowiedzi impulsowe projektowanego filtra FIR o zerowej fazie. Linia niebieska – odpowiedź teoretyczna, wzór (2), linia czerwona – odpowiedź wyznaczona numerycznie, wzór (4).

Filtrację HRIR bazy PWr przeprowadzono realizując liniowy splot dyskretny każdej odpowiedzi impulsowej złożonej z $N=2048$ próbek z $M=2001$ próbkami odpowiedzi impulsowej (2). Ze względu na brak przyczynowości filtra o zerowej fazie, jako wynik operacji splotu należy przyjąć próbki o numerach

$$n = \frac{M+1}{2}, \dots, N - \frac{M-1}{2} \quad (5)$$

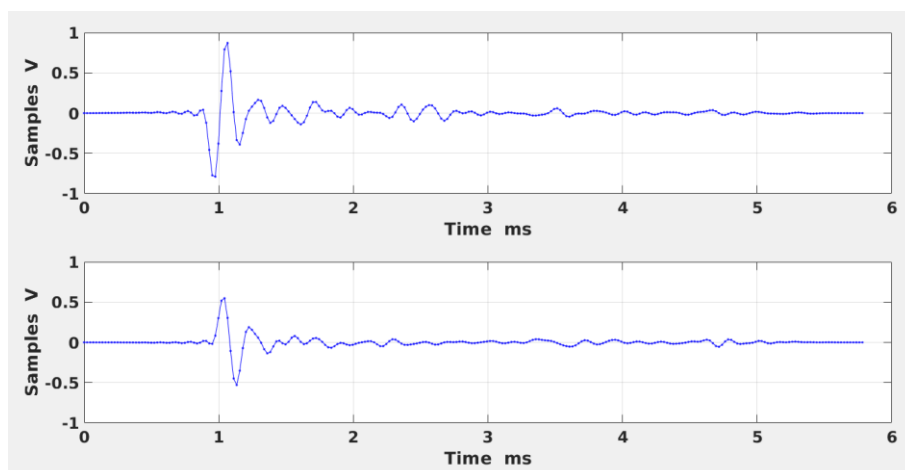
razem N próbek, przy czym numeracja próbek po filtracji równa jest $n=1,2,\dots,N+M-1$.

Energię sygnału oblicza się na podstawie pierwszych 1024 próbek HRIR po przeprowadzeniu filtracji i wyborze próbek zgodnie ze wzorem (5). Energię zawartą w pozostałych próbkach pomija się, ponieważ z fizycznych przyczyn powinna być ona zanedbywalnie mała. W praktyce w pozostałych 1024 próbkach zgromadzona energia jest wynikiem kumulacji błędów obliczeń numerycznych i błędów wynikających z przeprowadzonych pomiarów, a nie zjawisk fizycznych. Stwierdzono eksperymentalnie, że w próbkach HRIR o numerach $N_1=26$ do $N_2=281$ zawarte jest nie mniej niż 95% energii sygnału, dla każdego położenia przestrzennego i dla każdego słuchacza. W celu wyeliminowania efektu Gibbsa, który powstanie nieuchronnie na skutek obcięcia ciągu próbek do $N_0=256$ elementów, wymnożono ww. 256 próbek HRIR przez funkcję okna czasowego będącego oknem Tukey'a, wzór (6).

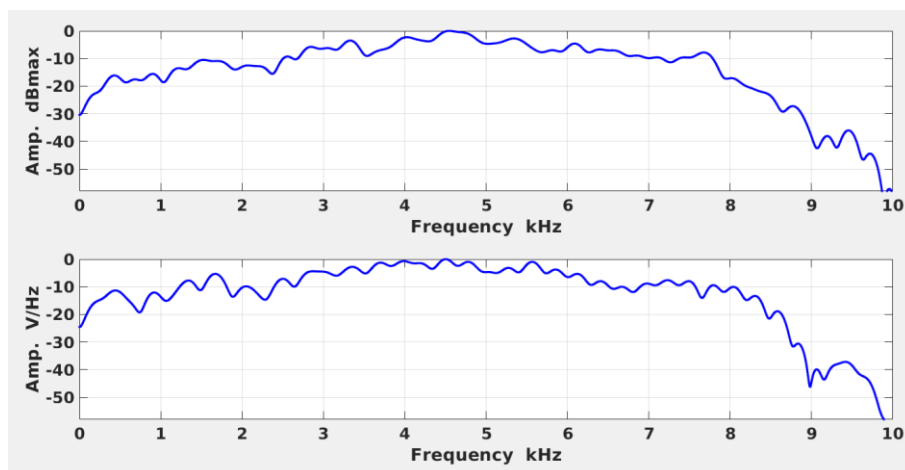
$$w_0(n) = \begin{cases} \sin^2\left(\frac{\pi n}{2n_2}\right) & n \in [1, n_2 - 1] \\ 1 & n \in [n_2, n_3] \\ \cos^2\left(\frac{\pi(n-n_3)}{2(n_4-n_3)}\right) & n \in [n_3 + 1, N_0] \end{cases} \quad (6)$$

gdzie $n = 1, 2, \dots, N_0$, $n_2 = 12$, $n_3 = 225$, $n_4 = 257$.

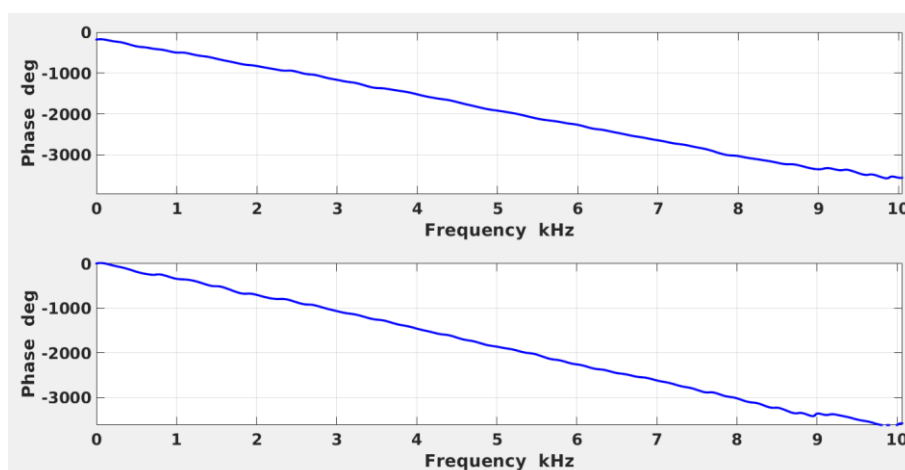
Na rys. 15 przedstawiono HRIR dla azymutu 15 deg i elewacji 36 deg, po przeprowadzeniu eliminacji nienaturalnych opóźnień (rys. 10), filtracji filtrem FIR o zerowej fazie, skróceniu ciągu wynikowego HRIR do 256 próbek i okienkowaniu ciągu oknem czasowym (6). Na rys. 16 oraz rys. 17 zamieszczono charakterystyki amplitudowe i fazowe HRTF wyznaczone jako DTFT na siatce częstotliwości f_k (akapit poniżej wzoru (3)) z HRIR po dokonaniu wszystkich ww. korekt.



Rys. 15 Naprawione HRIR (dla ucha lewego i prawego), użytkownik 56, elewacja 36 deg, azymut 15 deg.

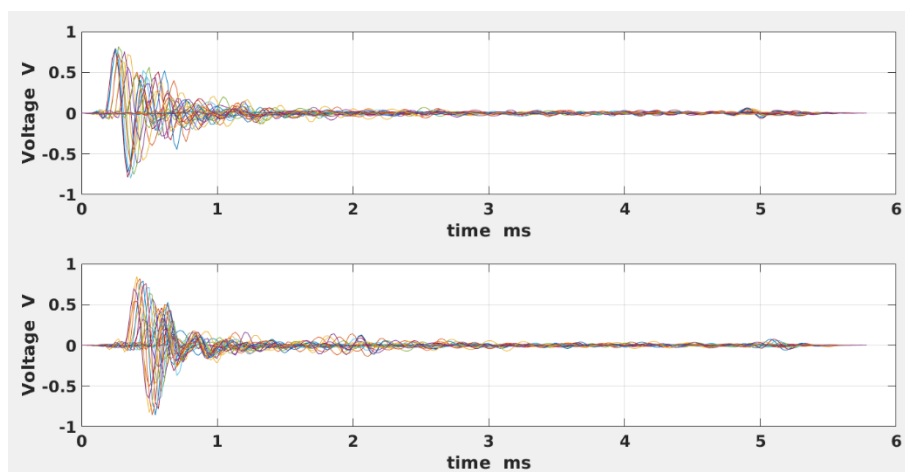


Rys. 16 Naprawione HRTF (dla ucha lewego i prawego) charakterystyka amplitudowa, użytkownik 56, elewacja 36 deg, azymut 15 deg.



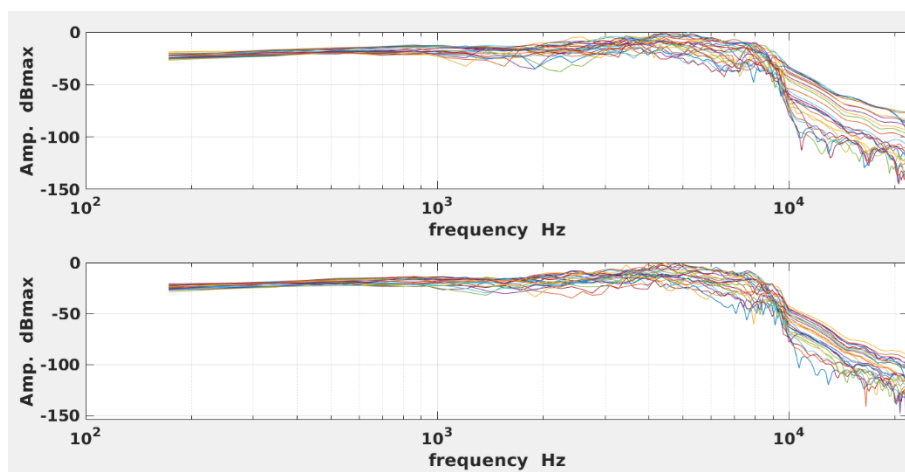
Rys. 17 Naprawione HRTF (dla ucha lewego i prawego) ch.fazowa, użytkownik 56, elewacja 36 deg, azymut 15 deg.

Na rys. 18 pokazano naprawione HRIR dla użytkownika 56, przy czym elewacja równa jest 9 deg, natomiast azymut zmienia się od 0, 15, 30, ..., 165 deg oraz od -180, -165, ..., -15 deg.

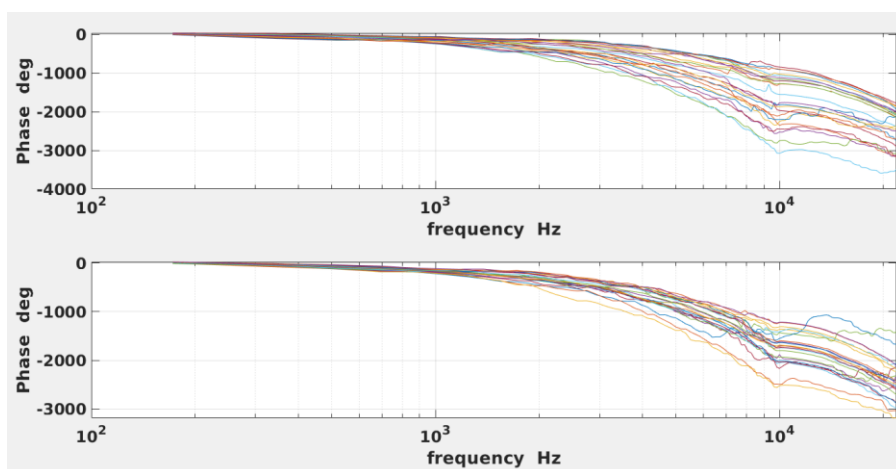


Rys. 18 Pomiary HRIR (dla ucha lewego i prawego), baza PWr po naprawie, użytkownik 56.

Na rys. 19 i rys. 20 pokazano odpowiednio charakterystyki amplitudowe i fazowe HRTF otrzymane na podstawie HRIR z rys. 18. HRTF obliczono jako DFT (dyskretna transformata Fouriera) z HRIR.



Rys. 19 Charakterystyki amplitudowe HRTF wyznaczone na podstawie HRIR (dla ucha lewego i prawego), baza PWr po naprawie, użytkownik 56.



Rys. 20 Charakterystyki fazowe HRTF wyznaczone na podstawie HRIR (dla ucha lewego i prawego), baza PWr po naprawie, użytkownik 56.

4. PODSUMOWANIE

Naprawiona baza HRIR PWr została zapisana jako zmienna języka Matlab. Struktura bazy HRIR PWr jest następująca:

HRIR_PWR.mat – plik Matlaba zawierający bazę HRIR
hrirdata – zmienna strukturalna Matlaba zawarta w ww. pliku
hrirdata.info – słowny opis zmiennych i wymiarów zmiennych w bazie HRIR
hrirdata.Nhv – zmienna o wymiarach 264 x 2 zawierająca azymuty i elewacje kolejnych pomiarów
hrirdata.hrir – zmienna strukturalna zawierająca dane pomiarowe wszystkich słuchaczy
hrirdata.hrir.pK – tablica o wymiarach 264 x 2 x 256 zawierająca dane pomiarowe K – tego słuchacza, K=5,6,...,56

LITERATURA

- [1] C. I. Cheng, G. H. Wakefield, Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space, *J. Audio Eng. Soc.*, Vol. 49, No 4, April 2001, pp. 231-249.
- [2] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen, *Head Related Transfer Functions of Human subjects*, *J. Audio Eng. Soc.*, Vol 43, No 5, May 1995, pp. 300-321.
- [3] P. Plaskota, *Research of acoustical impedance of human skin*. *Vibrations in Physical Systems*. 2019, vol. 30, nr 1, art. 2019143, s. 1-6
- [4] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen, *Binaural Technique: Do We Need Individual Recordings?*, *J. Audio Eng. Soc.*, Vol 44, No 6, June 1996, pp. 451-469.
- [5] P. Pruchnicki, P. Plaskota, HRTF Automatic Measuring System, *Archives of Acoustics*. 2008 vol. 33, 1, 19-25.
- [6] P. Plaskota, P. Pruchnicki, Practical aspects of using HRTF measuring device, *Archives of Acoustics*, 2006, vol. 31, no 4, suppl., pp. 439-444.
- [7] J. Blauert, *Spatial Hearing*, The MIT Press, Massachusetts 1997.

SPATIAL SOUND SYNTHESIS USING A MODIFIED HRTF DATABASE BASED ON MEASUREMENTS TAKEN AT THE WROCLAW UNIVERSITY OF TECHNOLOGY

In recent years, there has been a growing interest in head-related transfer function (HRTF) measurements, partly due to the potential for automation. Automation enables high-resolution spatial measurements to be taken quickly from many people. These measurements can then be compiled into a database for use in further work.

This paper presents a spatial sound synthesis process based on HRTF measurements conducted at the Wrocław University of Technology. These measurements were taken from 52 individuals using a system designed for the rapid, automated measurement of HRTFs. The paper also describes how to reconstruct incorrectly recorded measurement signals.