XXI Międzynarodowe Sympozjum Nowości w Technice Audio i Wideo NTAV2026

Wrocław, 15-17 października 2026

Katedra Akustyki, Multimediów i Przetwarzania Sygnałów

PTA

PAN
POLSKA AKADEMIA NAUK

Polish Audio Engineering Society

Materiały konferencyjne

ORGANIZATORZY ▪ ORGANIZERS

Katedra Akustyki, Multimediów i Przetwarzania Sygnałów
Wydział Elektroniki, Fotoniki i Mikrosystemów
Polskie Towarzystwo Akustyczne Oddział we Wrocławiu
Polska Sekcja Audio Engineering Society
Polska Akademia Nauk
Politechnika Wrocławska

KOMITET NAUKOWY ▪ SCIENTIFIC COMMITTEE

Przewodniczący: prof. dr hab. inż. Krzysztof J. Opieliński

- dr hab. inż. Bogusław Szlachetko
- prof. dr hab. Andrzej Brzoska
- prof. dr hab. inż. Andrzej Czyżewski
- prof. dr hab. inż. Andrzej Dobrucki
- prof. dr hab. inż. Tadeusz Kamisiński, prof. AGH
- Prof. dr hab. inż. Piotr Kleczkowski
- prof. dr hab. inż. Bożena Kostek
- prof. dr hab. inż. Mirosław Meissner
- prof. dr hab. inż. Andrzej Miśkiewicz
- dr hab. inż. Janusz Piechowicz, prof. AGH
- dr hab. inż. Adam Pilch,
- prof. dr hab. Anna Preis,
- prof. dr hab. Ewa Skrodzka, prof. UAM
- dr hab. Andrzej Wicher, prof. UAM
- prof. dr hab. inż. Jerzy Wiciak
- dr hab. inż. Sławomir Zieliński, prof. PB
- prof. dr hab. inż. Jan Żera

KOMITET ORGANIZACYJNY ▪ ORGANIZING COMMITTEE

Przewodniczący:	Przemysław Plaskota *(korekty)*
Sekretarz:	Michał Łuczyński *(redakcja)*
Skarbnik:	Dorota Młynarczyk
Członkowie:	Bartłomiej Kruk
	Paweł Dziechciński
	Piotr Kozłowski
	Maurycy Kin
	Romuald Bolejko
	Andrzej Lewandowski
	Paweł Górski
	Agnieszka Paula Pietrzak
	Tomasz Nowak
	Tomasz Kopciński
	Piotr Książek

WSPIERAJĄ NAS



Fundacja Wspierania
Rozwoju Radiokomunikacji
i Technik Multimedialnych



SVANTEK



SONITUS
akustyka wnętrz • ochrona przeciwdźwiękowa



Wydział Elektroniki,
Fotoniki i Mikrosystemów

Program sympozjum NTAV2026, 14-17 października 2026, Wrocław

v 3.0

Środa, 14.10.2026

08:00	Dni Akustyki, bud. C-16
09:00	
09:20	
09:40	Komora Akustyczna A0.2
10:00	**Otwarcie Dni Akustyki**
10:20	Produkcja dźwięku
10:40	Prezentacja możliwości syntezatorów dźwięku i instrumentów elektronicznych
11:00	
11:20	
11:40	Marcin Kaczmarek
12:00	Próba do koncertu z warsztatami
12:20	
12:40	Agata Zemla
13:00	
13:20	Koncert odbędzie się w piątek, 16.10.2026 g.19.00 (obowiązują zapisy)
13:40	
14:00	
14:20	
14:40	Komora Akustyczna A0.2
15:00	**Seminarium Akustyków PWr**
15:30	Prezentacje uczestników Seminarium Dni Akustyki
15:40	
16:00	
16:20	
16:40	
17:00	Networking Rozmowy w kuluarach
17:30	
17:40	**Zakończenie Dni Akustyki**
18:00	Posiedzenie prezydium Zarządu PS AES Przyjaciele i Znajomi
18:20	
18:40	
19:00	**Posiedzenie Zarządu PS AES Przyjaciele i Znajomi (ul. Grunwaldzka 98)**
20:00	
21:00	
22:00	

Czwartek, 15.10.2026

08:00	Recepcja	Sympozjum, bud. D-21. s 201	
09:00		Otwarcie konferencji	
09:20		Referat plenarny T. Żernicki	
09:40		Zdjęcie grupowe	
10:00		Referaty sekcyjne E. Łukasik	T. Nowak
10:20			R. Bolejko
10:40			M. Kin
11:00		Przerwa	
11:20		Referaty sekcyjne W. Mickiewicz	A.P. Pietrzak
11:40			M. Lewandowski
12:00			A. Jasińska
12:20		Przerwa	
12:40		Referaty sekcyjne B. Kostek	P. Dziechciński
13:00			B. Mróz
13:20			A. Białczyk
13:40			
14:00	Sesja	Obiad	
14:20			
14:40			
15:00		Referat plenarny M. Lewandowski	
15:20	Profil	Referaty sekcyjne A. Dobrucki	A. Dobrucki
15:40			E. Łukasik
16:00			B. Kostek
16:20			B. Kostek
16:40		Przerwa	
17:00	Zakończenie	Referaty sekcyjne A. Dobrucki	W. Mickiewicz
17:20			M. Jasiński
17:40			S. Weyna
18:00			S. Weyna
18:20			
18:40			
19:00		**Bankiet Przyjaciele i znajomi (ul. Grunwaldzka 98)**	
20:00			
21:00			
22:00			

Piątek, 16.10.2026

08:00	Recepcja	Sympozjum, bud. D-21. s 201		Konkurs, bud. C-16
09:00		Referat plenarny D. Majerek		
09:20				
09:40		Przerwa		Komora Akustyczna A0.2
10:00		Referaty konkursowe K. Opieliński	J. Kucharczyk	Konkurs nagrań Komora Akustyczna Technopolis
10:20			J. Warzych	
10:40			T. Kopciński	
11:00		Przerwa		
11:20		Referaty konkursowe B. Szlachetko	I. Balyka	
11:40			M. Różycki	
12:00			P. Książek	
12:20		Przerwa		
12:40		Referaty konkursowe A.P. Piertrzak	K. Sterniuk	bud. D-21 s. 102
13:00			P. Kojs	Grupa Poziom dźwięku na wydarzeniach rozrywkowych
13:20			M. Ładyżyński	
13:40			P. Kluzowicz	
14:00		Obiad		Przerwa obiadowa
14:20				
14:40				
15:00		Referat plenarny Ł. Januszkiewicz		Obrady komisji konkursowych
15:20		Informacje Grupa PDnWR P. Górski, P. Plaskota		
15:40				
16:00		Wyniki konkursów Zamknięcie konferencji		
16:20				
16:40		Spacer dźwiękowy wokół Kampusu Politechniki Wrocławskiej J. Dumanowski (start: Komora Akustyczna)		
17:00				
17:20				
17:40				
18:00				
18:20				
18:40				Komora Akustyczna A0.2
19:00		Przejście do Teatru Sztampa		Koncert Agata Zemla
20:00		Spektakl Teatr Sztampa (bud. C-13, 3. piętro)		
21:00				
22:00				

Sobota, 17.10.2026

	Warsztaty, bud. C-16		
	Komora Akustyczna A0.2		A0.1
10:00	System wspomagania akustyki	Targi pracy Wystawa	
10:20			
10:40			
11:00	Montaż taśmy Wprowadzenie		Montaż taśmy analogowej praca indywidualna
11:20			
11:40	Solfeż barw		
12:00			
12:20			
12:40	Przerwa		
13:00	Dźwięk immersyjny		
13:20			
13:40			
14:00	Przerwa obiadowa		
14:20			
14:40			
15:00	Ambisonia jako alternatywa dla dźwięku obiektowego Odsłuchy nagrań w systemie dźwięku immersyjnego	Targi pracy Wystawa	Solfeż barw praca indywidualna
15:20			
15:40			
16:00			
16:20			
16:40			
17:00			
17:20			
17:40			
18:00			
18:20			
18:40			
19:00	**Koncert Canti Spazialzizati (Żyjnia, ul. Świdnicka 2-4)**		
20:00			
21:00			
22:00			

SPIS TREŚCI

Tomasz Żernicki

Architektura agentów AI w produkcji muzycznej: sterowanie DAW, uczenie się preferencji producenta i weryfikacja działań 7

Marcin Lewandowski

Współczesne przetworniki audio A/C i C/A: architektury, algorytmy i miary jakości 8

Dariusz Majerek

Jak AI uczy się słyszeć i widzieć 9

Łukasz Januszkiewicz

Agentic AI for Audio: From Large Audio Models to Synthetic Acoustic Worlds 10

Tomasz Nowak

Dual plane acoustic lens measurements 11

Romuald Bolejko, Stanisław Kuna

Wykorzystanie głośnika wstęgowego jako ustrój pochłaniający dźwięk 12

Anna Frączkowska, Maurycy Kin, Piotr Kozłowski

Badanie skuteczności działania indywidualnych ochronników słuchu i wpływ ich na percepcję barwy dźwieku. 13

Agnieszka Paula Pietrzak

Profile akustyczne śpiewu chóralnego, rozrywkowego i amatorskiego w kontrolowanym zadaniu śpiewania długo utrzymywanych samogłosek 14

Marcin Lewandowski, Bartosz Kisły

Śledzenie partytury w czasie rzeczywistym z wykorzystaniem algorytmów typu DTW oraz uczenia maszynowego 15

Alicja Jasińska, Maurycy Kin

Badanie zdolności echolokacyjnych człowieka w identyfikacji barier architektonicznych. . 16

Paweł Dziechciński, Jan Kucharczyk

Pomiary mocy wyjściowej wzmacniaczy przeznaczonych do systemów rozgłoszeniowych 17

Bartłomiej Mróz

Ambisonia wyższych rzędów w każdej przeglądarce: kontenerowy stos do transmisji 360° na żywo 18

Andrzej Białczyk

DOLBY ATMOS i dźwięk nowej generacji (NGA) W TVP 19

Andrzej Dobrucki, Grzegorz Matusiak

Kierunkowość głośnika omnipolarnego 21

Mateusz Maciejewski, Ewa Łukasik

Modele percepcyjne i reprezentacje wyuczone: porównanie jakości tradycyjnych oraz neuronowych kodeków audio 22

Andrzej Czyżewski, Bożena Kostek, Henryk Skarżyński

Model akustyczno-elektrycznego przetwarzania sygnału mowy dla potrzeb predykcji efektów terapii częściowej głuchoty 23

Dorota Gutowska, Bożena Kostek

Estymacja wieku na podstawie mowy z wykorzystaniem dostrojonych wektorów reprezentacji mówcy 25

Witold Mickiewicz, Kaja Kosmenda

Odpowiedzi impulsowe pomieszczenia na potrzeby technik dźwięku immersyjnego – rejestracja i przetwarzanie. 27

Maciej Jasiński, Jan Żera

Modelowanie niepewności pomiarowej przy rejestracji odpowiedzi impulsowej z wykorzystaniem mikrofonu ambisonicznego 30

Stefan Weyna

Dźwięk jako efekt strukturalnych zaburzeń przepływów strumieniowych. Kształtowanie wektorowych pól dźwięku i hałasu 32

Stefan Weyna

Od fali akustycznej do struktur przepływowo-akustycznych 33

Daniel Górski

Automatyczna ocena jakości mowy klonowanej 35

Jan Kucharczyk, Bartłomiej Kruk

Wpływ złożoności lingwistycznej mowy na skuteczność rozpoznawania syntetycznego głosu przez ludzi 36

Jakub Warzych, Michał Łuczyński

Automatyczna detekcja deepfake'ów audio w sieciach telekomunikacyjnych z wykorzystaniem uczenia maszynowego na obiektywnych parametrach akustycznych..... 37

Tomasz Kopciński, Kacper Panek, Bartłomiej Kruk

Ocena zmian percepcji słuchowej realizatora dźwięku w trakcie długotrwałej pracy w warunkach wysokiego poziomu ciśnienia akustycznego 38

Ihar Balyka, Sławomir K. Zieliński

Non-instrusive PEAQ approximation using deep neural networks: a comparison of single-task and multi-task approaches 39

Mikołaj Różycki

Śledzenie ruchu gałek ocznych z wykorzystaniem metod uczenia maszynowego 41

Piotr Książek

Modelowanie stanu zimującej rodziny pszczelej na podstawie pomiarów wibroakustycznych 42

Miłosz Bakiera, Marcin Sutor, Kamil Sterniuk, Bożena Kostek

Bezkontaktowa estymacja aktywności oddechowej i sercowej na podstawie nagrań wideo mikroruchów klatki piersiowej podczas zadań w języku ojczystym i obcym 43

Przemysław Kojs

Analiza metod symulacji organów piszczałkowych 45

Michał Ładyżyński, Franciszek Malewski, Adam Piwoński, Jakub Kostrzewa

Aplikacja internetowa do treningu słuchania krytycznego 47

Piotr Kluzowicz, Bartłomiej Kruk

Porównanie parametrów obiektywnych mikrofonów wielkomembranowych oraz analiza powtarzalności ich parametrów między egzemplarzami 49

GENEROWANIE, EDYCJA I

WYKŁAD PLENARNY 1

CZWARTEK 15.10.2026

GODZ. 9:20 – 10:00

ARCHITEKTURA AGENTÓW AI W PRODUKCJI MUZYCZNEJ: STEROWANIE DAW, UCZENIE SIĘ PREFERENCJI PRODUCENTA I WERYFIKACJA DZIAŁAŃ

Tomasz Żernicki
My3Daudio Ventures
Autor korespondencyjny: **Tomasz Żernicki**, *tomasz.zernicki@my3daudio.com*

Czy AI może sterować studiem, uczyć się preferencji producenta i weryfikować własne działania? Wykład łączy przegląd aktualnego stanu wiedzy z propozycją innowacyjnej architektury agentów AI wspierających produkcję muzyczną.

Tematy:
•Sterowanie DAW i wtyczkami przez agentów AI.
•Pamięć intencji i preferencji producenta.
•Weryfikacja rezultatów i bezpieczne cofanie zmian.
•Granice pomiarów technicznych i oceny estetycznej.

Program:
•Przegląd badań — możliwości, ograniczenia i otwarte problemy.
•Architektura AI DAW OS — kontekst, pamięć, procedury i współpraca agentów.
•Demonstracje — korekta brzmienia, pomiary audio i porównania odsłuchowe.

Wykład jest skierowany do producentów muzycznych, realizatorów dźwięku, badaczy i twórców technologii audio. Podczas wykładu słuchacze dowiedzą się, jak budować system AI interpretujący intencje, wykonujący operacje w DAW i sprawdzający ich skutki w pętli zwrotnej.

WYKŁAD PLENARNY 2

CZWARTEK 15.10.2026

GODZ. 15:00 – 15:40

WSPÓŁCZESNE PRZETWORNIKI AUDIO A/C I C/A: ARCHITEKTURY, ALGORYTMY I MIARY JAKOŚCI

Marcin Lewandowski
Instytut Radioelektroniki i Technik Multimedialnych, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
Autor korespondencyjny: **Marcin Lewandowski**, *marcin.lewandowski@pw.edu.pl*

Współcześnie o jakości przetwornika nie decyduje tylko odpowiednio zaprojektowana pętla modulatora czy dokładne sterowanie źródłami napięciowymi lub prądowymi, ale skuteczne działanie systemów dynamicznego dopasowania elementów, kontrola zniekształceń międzysymbolowych (ISI), redukcja jitter'a, stopień tłumienia tętnienia zasilania i separacja tych zakłóceń od toru sygnałowego czy kwantyzacja czasu w stopniach impulsowych. W literaturze te dynamiczne i nieliniowe procesy są nazywane nieliniowymi zjawiskami 2. rzędu. Jako bazę do porównania przetworników wybrano ponad 230 źródeł literaturowych i wykonano eksperymenty numeryczne dla każdego z omawianych mechanizmów przetwarzania. Zestawiono ze sobą podział architektur, parametry techniczne (DR, SNDR, THD+N, FoM) i algorytmy kształtowania szumu i widma niedopasowania, modulacje impulsowe wyjściowych stopni mocy oraz implementacje modulatorów na układach programowalnych.

WYKŁAD PLENARNY 3

PIĄTEK 16.10.2026

GODZ. 9:00 – 9:40

JAK AI UCZY SIĘ SŁYSZEĆ I WIDZIEĆ

Dariusz Majerek
Katedra Matematyki Stosowanej, Wydział Matematyki i Informatyki Technicznej, Politechnika Lubelska
Autor korespondencyjny: **Dariusz Majerek**, *d.majerek@pollub.pl*

Słowa kluczowe: uczenie samonadzorowane, reprezentacje multimodalne, analiza audio i wideo, uczenie kontrastowe, modele świata

Przez ostatnią dekadę postęp w analizie audio i wideo napędzały zbiory z ręcznymi etykietami: ImageNet, AudioSet, Kinetics. Ten model wyczerpał się z powodów praktycznych. Anotacja minuty nagrania to godziny pracy eksperta, etykiety bywają niespójne, a wytrenowany model zna wyłącznie zamknięty słownik klas i słabo przenosi się między domenami. Jednocześnie danych nieoznaczonych mamy praktycznie bez ograniczeń. Odpowiedzią jest uczenie samonadzorowane, w którym model tworzy sobie zadanie z samych danych: maskuje fragment sygnału i uczy się go przewidywać, rozpoznaje, że dwa widoki przedstawiają to samo, albo dopasowuje nagranie do opisu. Zamiast rozwiązywać konkretne zadanie, buduje ogólną reprezentację, którą dostraja się później na małym zbiorze z etykietami. Wystąpienie omawia pięć metod reprezentatywnych dla tego nurtu. wav2vec 2.0 uczy dyskretnych jednostek mowy i pozwala zbudować system rozpoznawania mowy z minutami transkrypcji zamiast tysięcy godzin. CLIP i CLAP łączą obraz oraz dźwięk z językiem naturalnym, umożliwiając klasyfikację z otwartym słownikiem. DINO poprzez samodestylację wytwarza cechy, w których segmentacja obiektów pojawia się samoistnie, choć nikt jej nie uczył. JEPA idzie krok dalej i przewiduje nie piksele, lecz reprezentację brakującego fragmentu, co prowadzi do modeli świata zdolnych planować działanie.

WYKŁAD PLENARNY 4

PIĄTEK 16.10.2026

GODZ. 15:00 – 15:40

AGENTIC AI FOR AUDIO: FROM LARGE AUDIO MODELS TO SYNTHETIC ACOUSTIC WORLDS

Łukasz Januszkiewicz
SoftServe Sp. z o.o.
Autor korespondencyjny: **Łukasz Januszkiewicz**, *ljanu@softserveinc.com*

Generative AI is rapidly transforming the audio domain, enabling systems that can not only recognize speech and sounds, but also generate, simulate, and reason about complex acoustic environments. This talk explores the evolution of audio AI, from foundation models and audio-language systems to emerging agentic architectures capable of planning, creating, and evaluating audio content autonomously. Particular attention will be given to synthetic data, which is becoming a key enabler for training next-generation speech and audio intelligence systems. The presentation will introduce the concept of agentic audio dataset synthesis, where AI agents collaborate to design realistic scenarios, generate conversations, and simulate acoustic spaces. The talk will conclude with an introduction to the concept of synthetic acoustic worlds and their potential role in the development of future audio AI systems.

Paper ID: **1**

DUAL PLANE ACOUSTIC LENS MEASUREMENTS

Tomasz Nowak
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Tomasz Nowak**, *t.nowak@pwr.edu.pl*

Słowa kluczowe: Transducers, electroacoustics, measurements, directivity

The paper presents the ongoing research on a lens that allows to control the shape of the acoustic wave front in two planes independently. The technical solution is innovative in relation to similar ideas present in the literature. Measurements were made in the experimental part. The research is based on a designed sample lens for a specific isodynamic transducer, which is characterized by an approximately homogeneous velocity distribution over the surface of the diaphragm, making it a source of a flat acoustic wave. The task of the lens was to give the flat wavefront a specific curvature by dividing it into channels of different lengths, achieving a distribution of delays in the cells of the output matrix. The lens was made using 3D printing technology. Measurements showed the possibility of independent influence on shaping the wavefront in two planes of the lens. The results in the far field confirm this indirectly, by characterizing the source in terms of directivity as a function of frequency, comparing a transducer with and without a lens. Obtained spatial characteristics are not possible without achieving a curved acoustic wave front. The transducer with a lens functions like a section of a pulsating ellipsoid.

Paper ID: **2**

WYKORZYSTANIE GŁOŚNIKA WSTĘGOWEGO JAKO USTRÓJ POCHŁANIAJĄCY DŹWIĘK

Romuald Bolejko*, Stanisław Kuna
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Romuald Bolejko**, *romuald.bolejko@pwr.edu.pl*

Słowa kluczowe: ustroje pochłaniające dźwięk, głośnik wstęgowy, przetworniki akustyczne

W artykule przedstawiono próbę wykorzystania głośnika wstęgowego do budowy ustroju pochłaniającego dźwięk. Idea polega na wykorzystaniu głośnika elektrodynamicznego, który jest przetwornikiem odwracalnym, do zamiany energii akustycznej na energię elektryczną i ostatecznie do jej rozproszenia w układzie elektrycznym i częściowo mechanicznym głośnika. Wy-korzystano głośnik wstęgowy, który charakteryzuje się bardzo lekką membraną w postaci wstęgi aluminiowej, a tym samym jest elementem o stosunkowo małej impedancji mechanicznej. W pracy omówiono model takiego ustroju pochłaniającego dźwięk opracowany na podstawie analogii elektro-mechano-akustycznej, metodykę wyznaczenia parametrów układu zastępczego ustroju oraz rozszerzoną analizę wpływu parametrów głośnika na jego zdolność pochłaniania dźwięku. Między innymi wykazano możliwość uzyskania dużych wartości współczynnika pochłaniania dźwięku (powyżej 0,6) w zakresie małych i średnich częstotliwości. Omówiono wpływ impedancji elektrycznej na zaciskach głośnika na wartość i kształt częstotliwościowej charakterystyki pochłaniania dźwięku oraz ograniczenia takiego ustroju dźwiękochłonnego ze względu na parametry mechaniczne i elektryczne głośnika. Wykazano, że głównym ograniczeniem w uzyskaniu układu o silnych właściwościach dźwiękochłonnych jest niska stała Bl oraz masa wstęgi głośnika. W artykule przedstawiono także wyniki wstępnych pomiarów współczynnika pochłaniania dźwięku rozpatrywanego ustroju dźwiękochłonnego uzyskane metodą transmisyjną w tak zwanej rurze Kundta.

Paper ID: **3**

BADANIE SKUTECZNOŚCI DZIAŁANIA INDYWIDUALNYCH OCHRONNIKÓW SŁUCHU I WPŁYW ICH NA PERCEPCJĘ BARWY DŹWIEKU.

Anna Frączkowska, Maurycy Kin*, Piotr Kozłowski
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Maurycy Kin**, *maurycy.kin@pwr.edu.pl*

Słowa kluczowe: ochronniki słuchu, barwa dźwięku, tłumienność wtrąceniowa

W artykule przedstawiono wyniki badań dotyczących skuteczności działania indywidualnych ochronników słuchu. Badania przeprowadzono w dwóch etapach. W pierwszym etapie dokonano pomiarów tłumienia metodą obiektywną, z wykorzystaniem sztucznego fantomu ludzkiej głowy, oraz metodą subiektywną. W drugim etapie zbadano wpływ ochronników na zmiany w postrzeganej barwie dźwięku. Wyniki wykazały, że wartości tłumienności wtrąceniowej uzyskane na fantomie były zgodne z wynikami badań subiektywnych, opartych na audiometrii tonowej z użyciem zatyczek i bez nich. Wykazano również, że zatyczki do uszu mają znaczący wpływ na postrzeganie barwy dźwięku, co podważa zasadność stosowania zatyczek podczas procedur związanych z produkcją, realizacją i inżynierią dźwięku.

Paper ID: **4**

PROFILE AKUSTYCZNE ŚPIEWU CHÓRALNEGO, ROZRYWKOWEGO I AMATORSKIEGO W KONTROLOWANYM ZADANIU ŚPIEWANIA DŁUGO UTRZYMYWANYCH SAMOGŁOSEK

Agnieszka Paula Pietrzak
Instytut Radioelektroniki i Technik Multimedialnych, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
Autor korespondencyjny: **Agnieszka Paula Pietrzak**, *agnieszka.pietrzak@pw.edu.pl*

Słowa kluczowe: śpiew chóralny, śpiew rozrywkowy, śpiew amatorski, analiza akustyczna głosu, intonacja

Celem pracy było porównanie profili akustycznych śpiewu chóralnego, rozrywkowego i amatorskiego w kontrolowanych warunkach. Analizę przeprowadzono na podstawie zrealizowanych 800 nagrań długo utrzymywanych samogłosek wykonanych przez 20 śpiewaków: 7 chóralnych, 5 rozrywkowych i 8 amatorów. Każda osoba miała za zadanie zaśpiewać pięć samogłosek na ośmiu wysokościach od C4 do C5. Z nagrań wyznaczono parametry opisujące kontrolę wysokości, stabilność fonacji, wibrato, oraz charakterystyki widmowe i rezonansowe. Ponieważ styl był czynnikiem międzyosobniczym, porównania prowadzono przede wszystkim na poziomie median dla poszczególnych śpiewaków. Najwyraźniejsze różnice dotyczyły kontroli wysokości. Mediana bezwzględnego błędu intonacyjnego wyniosła 8,1 centa w grupie chóralnej, 19,7 centa w rozrywkowej i 25,2 centa w amatorskiej. Grupa chóralna wykazywała także mniejszą zmienność i dryf częstotliwości podstawowej oraz wyższy HNR. Shimmer i parametry wibrata różnicowały grupy słabiej, przy znacznym nakładaniu wyników poszczególnych osób. Cechy widmowe wskazywały na koncentrację widma w zakresie niższych częstotliwości w śpiewie chóralnym.

Paper ID: **5**

ŚLEDZENIE PARTYTURY W CZASIE RZECZYWISTYM Z WYKORZYSTANIEM ALGORYTMÓW TYPU DTW ORAZ UCZENIA MASZYNOWEGO

Marcin Lewandowski*, Bartosz Kisły
Instytut Radioelektroniki i Technik Multimedialnych, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
Autor korespondencyjny: **Marcin Lewandowski**, *marcin.lewandowski@pw.edu.pl*

Słowa kluczowe: śledzenie partytury, transformer, przetwarzanie dźwięku

W niniejszym artykule przedstawiono porównawcze metod śledzenia partytury w czasie rzeczywistym oraz zaproponowano nowy system oparty na modelu transformer. Badano cztery systemy: Online Time Warping, detektor wizualny CYOLO-SB+A, system oparty na sieciach CNN (HeurMiT) oraz proponowany model wykorzystujący podwójne enkodery do reprezentacji typu „piano roll" i CQT z dopasowaniem za pomocą iloczynu skalarnego. W eksperymentach przeprowadzonych na zbiorze danych MAESTRO oceniono dokładność, opóźnienie oraz odporność na zmiany tempa, transpozycję wysokości dźwięku, szumy i błędy wykonawcy. Model oparty na transformatorze osiągnął 98% dokładności przy opóźnieniu wynoszącym 213 ms oraz wykazał wysoką odporność na zakłócenia.

Paper ID: **6**

BADANIE ZDOLNOŚCI ECHOLOKACYJNYCH CZŁOWIEKA W IDENTYFIKACJI BARIER ARCHITEKTONICZNYCH.

Alicja Jasińska, Maurycy Kin*
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Maurycy Kin**, *maurycy.kin@pwr.edu.pl*

Słowa kluczowe: echolokacja, lokalizacja dźwięku, komfort akustyczny

W środowisku zabudowanym różnorodne bariery fizyczne znacznie utrudniają samodzielną orientację i poruszanie się, zwłaszcza osobom niewidomym i niedowidzącym. Podczas gdy osoby widzące opierają się głównie na bodźcach wzrokowych, osoby niedowidzące wykorzystują ludzką echolokację do tworzenia obrazów przestrzennych otoczenia poprzez analizę odbitych sygnałów akustycznych. Badania wykazują, że utrata słuchu wywołuje kompensacyjną reorganizację neuroplastyczną, sprawiając, że osoby niedowidzące są bardziej wrażliwe na bodźce lokalizacji binauralnej, takie jak międzyuszne różnice poziomu (ILD). Podczas gdy wcześniejsze badania skupiały się głównie na ogólnym poziomie ciśnienia akustycznego, wysokości dźwięku lub parametrach czasowych, rola barwy dźwięku — określanej przez strukturę spektralną — pozostaje niedostatecznie zbadana jako kluczowy atrybut percepcyjny służący do identyfikacji przeszkód przestrzennych. W niniejszej pracy przeanalizowano sposób, w jaki zmiany barwy dźwięku spowodowane pochłanianiem akustycznym, geometrią pomieszczenia i przeszkodami fizycznymi wpływają na postrzeganie i wykrywanie barier architektonicznych.

Paper ID: **7**

POMIARY MOCY WYJŚCIOWEJ WZMACNIACZY PRZEZNACZONYCH DO SYSTEMÓW ROZGŁOSZENIOWYCH

Paweł Dziechciński*, Jan Kucharczyk
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Paweł Dziechciński**, *pawel.dziechcinski@pwr.edu.pl*

Słowa kluczowe: wskaźnik transmisji mowy, STIPA, systemy rozgłoszeniowe, elektroakustyczne wzmacniacze mocy

Jednym z elementów projektu systemu rozgłoszeniowego jest określenie maksymalnej wartości skutecznej niezniekształconego sygnału, który może zostać doprowadzony do urządzenia głośnikowego. Jak to pokazano w pracy, w przypadku niektórych współczesnych konstrukcji, opierając się na wynikach pomiarów znormalizowanymi metodami, obliczenie takiej wartości dla sygnału mowy z akceptowalną niepewnością, jest niemożliwe. W związku z tym, w pracy zaproponowano prostą metodę pomiaru mocy wyjściowej, która dawałaby projektantom jednoznaczną informację o rzeczywistych możliwościach wzmacniacza mocy dla sygnału mowy.

Paper ID: **8**

AMBISONIA WYŻSZYCH RZĘDÓW W KAŻDEJ PRZEGLĄDARCE: KONTENEROWY STOS DO TRANSMISJI 360° NA ŻYWO

Bartłomiej Mróz*
Katedra Systemów Multimedialnych, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska
Autor korespondencyjny: **Bartłomiej Mróz**, *bartlomiej.mroz@pg.edu.pl*

Słowa kluczowe: ambisonia wyższych rzędów; transmisja na żywo; wideo 360°; MPEG-DASH; renderowanie binauralne; dźwięk przestrzenny; przeglądarki internetowe; interoperacyjność

Transmisja na żywo wideo 360° z dźwiękiem ambisonicznym wyższych rzędów wymagała dotąd złożenia zestawu luźno powiązanych narzędzi, z których każde konfiguruje się inaczej i psuje się inaczej. Praca przedstawia kontenerowy stos, rozwinięty z systemu użytego do transmisji koncertu na żywo w 2022 r. i przebudowany tak, aby mógł zostać łatwo uruchomiony przez innych.

Stos przyjmuje sygnał w protokole RTMP lub SRT, transkoduje go do MPEG-DASH z dźwiękiem Opus o liczbie kanałów do 25 i renderuje binauralnie w przeglądarce z uwzględnieniem ruchów głowy słuchacza, bez potrzeby instalowania jakiejkolwiek aplikacji po stronie odbiorczej. Działa na skromnym sprzęcie: wdrożenie referencyjne obsługiwane jest przez Mac Mini z 2012 r., a stos uruchamia się również na Raspberry Pi 4.

Uruchomienie systemu we wszystkich głównych przeglądarkach wymagało zidentyfikowania kilku nieudokumentowanych wcześniej błędów interoperacyjności, z których każdy zgłoszono do właściwego projektu: rozbieżności w stosowaniu przesunięcia prezentacji, gdzie trzy silniki przeglądarek są niezgodne same ze sobą w zależności od ścieżki odtwarzania; błędu w dash.js, przez który odtwarzacz trwale przestaje pobierać segmenty; nieaktualnej kategorii sesji audio w WebKit, wyciszającej odtwarzanie na iOS bez żadnego sygnału błędu; oraz regresji w koderze AAC w FFmpeg, która usunęła jedyny układ kanałów zdolny przenieść ambisonię trzeciego rzędu.

Praca obejmuje architekturę systemu, pomiary stojące za decyzjami projektowymi oraz trudność diagnozowania usterek, które każdy pomiar raportuje jako poprawne działanie.

Paper ID: **9**

DOLBY ATMOS I DŹWIĘK NOWEJ GENERACJI (NGA) W TVP

Andrzej Białczyk
TVP S.A
Autor korespondencyjny: **Andrzej Białczyk**, *a.bialczyk@wp.pl*

Słowa kluczowe: Dolby Atmos, TVP, Telewizja Polska

Telewizja Polska S.A. (TVP S.A.) jest jednym z największych nadawców publicznych w Europie. TVP produkuje programy telewizyjne i obsługuje 16 głównych kanałów w jakości HD. Od 2003 roku TVP nada-je jednocześnie dźwięk stereofoniczny oraz 5.1, przy czym oba formaty zapewniają kreację dźwięku w przestrzeni dwuwymiarowej.

W czerwcu 2018 roku TVP 1 HD wyprodukowała i przeprowadziła pierwszą na świecie transmisję na ży-wo z dźwiękiem przestrzennym Dolby Atmos w naziemnej telewizji cyfrowej w standardzie DVB-T. Dźwięk Dolby Atmos został zakodowany przy użyciu kodowania E-AC3+JOC (Joint Object Coding). Dzięki zastosowaniu Dolby Atmos dźwięk może docierać do odbiorcy nie tylko z przodu i z tyłu ale także z góry i z dołu. Daje to telewidzowi wrażenie uczestnictwa w wydarzeniu i obecności na stadionie lub sali koncertowej

Podczas UEFA Euro 2020 oraz Mistrzostw Świata FIFA 2022 uruchomiono tymczasowy kanał UHD-TVP 4K. Po raz pierwszy zastosowano wówczas kodowanie NGA (New Generation Audio) w standardzie AC-4. System ten umożliwia kreację dźwięku w trójwymiarowej przestrzeni jak również zapewnia widzom sper-sonalizowane wrażenia dźwiękowe. Dźwięk był transmitowany w formacie 5.1.4+4D (5.1.4+D+D+D+D). w którym D oznacza dialog.

Telewidzowie mogli wybrać prezentację z komentarzem polskim TVP, komentarzem anglojęzycznym pro-dukowanym przez organizatora mistrzostw (UEFA/(FIFA), komentarzem radiowym stanowiącym audiode-skrypcję lub czysty dźwięk ze stadionu, bez komentarza. Ponadto telewidzowie mogli dostosować poziom głośności komentarza względem dźwięku tła ze stadionu, zgodnie z własnymi preferencjami.

Chociaż NGA jest przede wszystkim rekomendowane dla usług UHD, podjęto również decyzję o zastoso-waniu NGA także w głównych kanałach HD: TVP 1 HD, TVP 2 HD, TVP Kultura HD. Dolby Atmos zako-dowany w AC-4 był produkowany i nadawany przez TVP Kultura HD podczas transmisji na żywo z Mię-dzynarodowego Konkursu Pianistycznego im. Fryderyka Chopina w 2021 i 2025 roku. Koncert laureatów Konkursu Chopinowskiego był nadawany w 2025 roku jednocześnie w TVP Kultura HD i TVP 1 HD z dźwiękiem Dolby Atmos 5.1.4 zakodowanym w AC-4.

W 2022 roku TVP Kultura transmitowała 16 Międzynarodowy Konkurs Skrzypcowy im. Henryka Wie-niawskiego także z dźwiękiem Dolby Atmos 5.1.4. Wcześniejsze wyżej wymienione wydarzenia muzyczne były produkowane w natywnym formacie 5.1.4. Tym razem dźwięk Dolby Atmos był upmixowany z na-tywnie produkowanego dźwięku 5.1. Telewidzowie mogli porównać wrażenia podczas odsłuchu stereo, 5.1 i automatycznie wytworzonego upmixu 5.1.4.

Podczas mistrzostw Europy UEFA Euro 2024 oraz Mistrzostw Świata FIFA 2026 Telewizja Polska nada-wała mecze w Dolby Atmos w naziemnej telewizji cyfrowej w TVP 1 HD, TVP 2 HD. Format nadawania 5.1.4+D+D+D umożliwiał wybór komentarza w języku polskim lub angielskim oraz tylko dźwięku stadio-nowego bez komentarza Możliwa była także regulacja poziomu dźwięku komentarza w stosunku do tła stadionu.

Podczas Mistrzostw Świata FIFA 2026 TVP w internecie udostępniała trzy równoległe, niezależne transmi-sje z dźwiękiem stereo: Video z miksem audio z komentarzem polskim, Video z miksem audio z komenta-rzem angielskim, Video tylko z dźwiękiem stadionowym, bez komentarza,

W DVB-T2 w jednym strumieniu audio znajdowały się trzy prezentacje zawierające dźwięk stadionowy 5.1 i dialogi. Po stronie odbiorczej po wyborze prezentacji wybrany dialog (komentarz) był miksowany z tłem – dźwiękiem stadionowym. Przełączanie prezentacji i wybór języka było natychmiastowe, bez zau-ważalnego opóźnienia i utraty fragmentu video. Przełączanie między prezentacjami nie miało wpływu na oglądany obraz. Mogło być dokonywane nawet podczas sytuacji podbramkowej.

Podczas nadawania Dolby Atmos i AC-4 w DVB-T/DVB-T2 okazało się, że na polskim rynku znajduje się wiele telewizorów i dekoderów w których producenci w rózny sposób ustalają priorytety wyboru ścieżki audio. Priorytetem w jednych telewizorach był MPEG audio w innych AC-4. Istotny był wybór podstawo-wego języka podczas konfiguracji odbiornika. Ważne było oznaczenie deskryptorów języka dla poszcze-gólnych strumieni audio i prezentacji, przyporządkowanie PIDów i kolejność umieszczenia w tablicy PMT odpowiednich danych.

Najwięcej problemów występowało podczas transmisji AC-4 w serwisach HD. Największym problemem był zmultiplikowany błąd oprogramowania w telewizorach Smart TV wielu producentów. Telewizory te wybierały AC-4 jako dźwięk podstawowy i próbowały go dekodować pomimo tego, ze nie miały wbudo-wanych dekoderów AC-4. W efekcie podczas odbioru programu TVP 1 lub TVP 2 pojawiała się cisza.

Po doświadczeniach z nadawaniem AC-4 grupa robocza składająca się z inżynierów Dolby, Samsung i TVP przygotowała dokument roboczy: Rekomendację dla nadawców, operatorów sieci nadawczych i pro-ducentów odbiorników telewizyjnych dotyczący priorytetów audio dla przypadków z wykorzystaniem strumienia AC-4 w sieciach telewizji cyfrowej DVB.

Paper ID: **10**

KIERUNKOWOŚĆ GŁOŚNIKA OMNIPOLARNEGO

Andrzej Dobrucki* [1], Grzegorz Matusiak [2]
1) Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska,
2) JRC Group S.A
Autor korespondencyjny: **Andrzej Dobrucki**, *andrzej.Dobrucki@pwr.edu.pl*

Słowa kluczowe: Głośnik omnipolarny, promieniowanie dźwięku, charakterystyka kierunkowości

W ostatnim czasie popularność zdobywają urządzenia głośnikowe, tzw. omnipolarne, które ze względu na swoją konstrukcję mają dookólną charakterystyką kierunkowości w płaszczyźnie poziomej. Jeśli głośnik taki umieszczony jest na wysokości uszu siedzących słuchaczy, emitowany poziom dźwięku jest niezależny od kierunku. W niniejszej pracy rozpatrywane jest urządzenie składające się z dwóch głośników umieszczonych membranami „face to face" i promieniujących w fazie. Głośniki umieszczone są w kulistej obudowie, a dźwięk jest emitowany na zewnątrz przez wąską, pierścieniową szczelinę w płaszczyźnie między głośnikami. W ten sposób wymuszona jest dookólna charakterystyka kierunkowości w płaszczyźnie szczeliny niezależnie od częstotliwości. Celem pracy było określenie charakterystyki kierunkowości w płaszczyźnie pionowej – prostopadłej do szczeliny. Przeprowadzono odpowiednie obliczenia stosując rozkład prędkości pasa promieniującego na harmoniki sferyczne. Spodziewano się maksimum charakterystyki w płaszczyźnie poziomej. Okazało się jednak, że maksimum występuje w kierunku prostopadłym do tej płaszczyzny. Przeprowadzono dyskusje i zaproponowano wyjaśnienie tego zjawiska.

Paper ID: **11**

MODELE PERCEPCYJNE I REPREZENTACJE WYUCZONE: PORÓWNANIE JAKOŚCI TRADYCYJNYCH ORAZ NEURONOWYCH KODEKÓW AUDIO

Mateusz Maciejewski, Ewa Łukasik*
Politechnika Poznańska, Instytut Informatyki
Autor korespondencyjny: **Ewa Łukasik**, *ewa.lukasik@cs.put.poznan.pl*

Słowa kluczowe: neuronowe kodeki audio; percepcyjna jakość dźwięku; MUSHRA; obiektywne metryki jakości audio; ViSQOL; UTMOS; DAC; EnCodec; MP3; AAC

Artykuł przedstawia porównanie jakości tradycyjnych kodeków audio, opartych na opracowanych eksperymentalnie modelach percepcyjnych, z rozwiązaniami wykorzystującymi nowoczesne paradygmaty neuronowego kodowania dźwięku, bazujące na wyuczonych reprezentacjach ukrytych. Podczas gdy tradycyjne algorytmy jawnie modelują ludzki układ słuchowy w celu usunięcia zbędnych informacji psychoakustycznych, kodeki neuronowe optymalizują struktury głębokiego uczenia bezpośrednio na dużych zbiorach danych w celu znalezienia wysoce efektywnych reprezentacji. W ramach przeprowadzonego eksperymentu porównano dwa kodeki neuronowe (EnCodec oraz DAC) z dwoma rozwiązaniami tradycyjnymi (MP3 oraz AAC). W badaniu wykorzystano zdalny test odsłuchowy MUSHRA (ITU-R BS.1534) obejmujący siedem gatunków muzycznych, przeprowadzony z udziałem zróżnicowanej grupy 24 słuchaczy w ich codziennych warunkach akustycznych. Wyniki te zestawiono z algorytmiczną oceną obiektywną przy użyciu metryk jakości ViSQOL oraz UTMOS. Uzyskane rezultaty jednoznacznie wskazują, że oba kodeki neuronowe osiągnęły istotnie wyższe oceny postrzeganej jakości przy znacznie niższej przepływności bitowej w porównaniu do klasycznych kodeków percepcyjnych.

Paper ID: **12**

MODEL AKUSTYCZNO-ELEKTRYCZNEGO PRZETWARZANIA SYGNAŁU MOWY DLA POTRZEB PREDYKCJI EFEKTÓW TERAPII CZĘŚCIOWEJ GŁUCHOTY

Andrzej Czyżewski, Bożena Kostek*, Henryk Skarżyński
Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska
Autor korespondencyjny: **Bożena Kostek**, *bozena.kostek@pg.edu.pl*

Wprowadzenie

Częściowa głuchota jest specyficzną formą ubytku słuchu, w której pacjent zachowuje sprawne słyszenie w paśmie niskoczęstotliwościowym, a wysokie częstotliwości, kluczowe dla zrozumienia mowy, są niesłyszalne lub silnie zniekształcone. Terapia EAS (Electric-Acoustic Stimulation), rozwijana od dwóch dekad przez zespół prof. Henryka Skarżyńskiego, łączy wzmocnienie akustyczne aparatem słuchowym w paśmie zachowanym z elektryczną stymulacją implantu ślimakowego w paśmie utraconym. Skuteczność terapii zależy od precyzyjnego dopasowania parametrów obu składowych do konkretnego pacjenta, co w praktyce klinicznej wymaga wielotygodniowego procesu iteracyjnego. Możliwość wcześniejszej predykcji sposobu, w jaki pacjent będzie słyszał po implantacji, oraz porównania różnych scenariuszy parametrycznych bez angażowania pacjenta, ma zarówno wartość kliniczną (poradnictwo przedoperacyjne), jak i badawczą (testowanie nowych schematów kodowania sygnału).

Cel pracy

Opracowanie modelu matematycznego oraz jego implementacji programowej, umożliwiających symulację wrażenia słuchowego pacjenta EAS w warunkach kontrolowanych, na podstawie jego rzeczywistego audiogramu.

Model i implementacja

Zaproponowany model realizuje w jednej pipelinie dwa równoległe tory przetwarzania sygnału mowy: tor akustyczny, którego charakterystyka zależy od audiogramu pacjenta i modeluje resztkowy słuch, oraz tor elektryczny, realizujący symulację stymulacji z implantu ślimakowego. Tor akustyczny to kaskada 11 filtrów pasmowych Butterwortha 4. rzędu ze wzmocnieniami kompensacyjnymi wyliczanymi z audiogramu, uzupełniona o kompresor dynamiki i filtr odcinający dla słuchu resztkowego. Tor elektryczny implementuje wokoder szumowy (CIS) lub tonowy o konfigurowalnej liczbie 5–22 kanałów, z uwzględnieniem zjawiska spread of excitation oraz kompresora front-end AGC. Oba tory są miksowane liniowo suwakiem A/E symulującym kliniczne dopasowanie balansu. Całość zaimplementowano jako pojedynczy plik HTML uruchamiany w przeglądarce, bez zależności serwerowych i bez potrzeby instalacji.

Rozszerzenia modelu

Względem wersji bazowej opracowano cztery grupy nowych funkcjonalności.

(1) Moduł kliniczny obejmuje osadzoną w oprogramowaniu, zanonimizowaną bazę 85 rzeczywistych audiogramów kandydatów EAS z Instytutu Fizjologii i Patologii Słuchu (wiek 8–90 lat, mediana 57 lat), zawierającą także dane audiometrii słownej zgodnie z metodologią Skarżyńskiego. Baza umożliwia szybkie iterowanie po realistycznych profilach ubytku słuchu podczas eksperymentów odsłuchowych i pokazów dydaktycznych.

(2) Moduł automatycznej ekstrakcji audiogramu z pliku PNG lub JPG wykrywa ramkę wykresu z ciągłych czarnych krawędzi (niezależnie od danych) i wybiera krzywą przewodnictwa powietrznego wg kryterium ciągłości. Rozwiązanie jest odporne na typową kontaminację zrzutów klinicznych (przyciski interfejsu, znaczniki maskowania) i skraca ścieżkę od gabinetu audiologa do symulacji.

(3) Moduł transformatowo-neuromorficzny udostępnia cztery transformaty widmowe (FFT, DCT, FWHT oraz DWT z falką Daubechies-5) jako alternatywę dla banku filtrów wokoderowego. Dodatkowo, opcjonalne kodowanie neuromorficzne typu Time-to-First-Spike, z dwoma pełnymi biofizycznymi modelami neuronu (Leaky Integrate-and-Fire oraz Izhikevich, oba realizowane jako pełne symulacje ODE), otwiera model na badania nad implantami nowej generacji opartymi na hardware neuromorficznym (Intel Loihi, IBM TrueNorth). Wskaźnik zysku energetycznego pozwala oszacować oszczędności obliczeniowe w porównaniu z klasycznym bankiem filtrów.

(4) Zaimplementowano trzy jawne tryby pracy transformaty w torze EAS: postprocesor, zastąpienie wokodera (klinicznie najistotniejsze; pozwala porównać implant klasyczny z hipotetycznym implantem neuromorficznym w warunkach współistnienia z resztkowym słuchem akustycznym) oraz pominięcie całego toru EAS (do izolowanego testu samej transformaty).

Walidacja

Poprawność numeryczna wszystkich transformat została zweryfikowana metodą round-trip (błąd od $3 \cdot 10^{-16}$ do $4 \cdot 10^{-13}$, perfekcyjna rekonstrukcja w precyzji double). Kodowanie TTFS obu modeli neuronu osiąga RMSE round-trip ~1,5 %. Algorytm ekstrakcji audiogramu z PNG, testowany na rzeczywistych zrzutach klinicznych, ma dokładność ±3 dB.

Zastosowania

Obszary zastosowań obejmują predykcję wyników terapii dla konkretnego pacjenta przedimplantacyjnie, poradnictwo pacjenta i jego rodziny, edukację specjalistów, generowanie kontrolowanego materiału do testów odsłuchowych (HINT, AzBio, SRT) oraz badania nad nowymi schematami kodowania mowy w implantach ślimakowych.

Paper ID: **13**

ESTYMACJA WIEKU NA PODSTAWIE MOWY Z WYKORZYSTANIEM DOSTROJONYCH WEKTORÓW REPREZENTACJI MÓWCY

Dorota Gutowska, Bożena Kostek*
Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska
Autor korespondencyjny: **Bożena Kostek**, *bozena.kostek@pg.edu.pl*

Słowa kluczowe: Wiek biologiczny i chronologiczny, estymacja wieku, biometria głosu, osadzenia mówcy (x-wektory), MAE (Mean Absolute Error; średni błąd bezwzględny), uczenie głębokie.

Kontekst i motywacja badawcza - rozwój metod uczenia głębokiego rewolucjonizuje obszar estymacji wieku na podstawie sygnału mowy. Dokładna ocena wieku z głosu odgrywa kluczową rolę w weryfikacji tożsamości, bezpieczeństwie, kryminalistyce oraz interakcji człowiek-komputer (systemy HMI (Human-Machine Interface), ASR (Automatic Speech Recognition), IVR (Interactive Voice Response). Głównym wyzwaniem w tej dziedzinie jest **rozbieżność między wiekiem chronologicznym a wiekiem biologicznym** wyrażanym w cechach akustycznych głosu. Na percepcję wieku wpływa szereg czynników, takich jak tempo mowy, użycie słownictwa, spontaniczność wypowiedzi, chrypka oraz zmiany jakości głosu. Ocena wieku przez słuchaczy ma charakter nieliniowy i ulega błędom subiektywnym – ludzie mają tendencję do zawyżania wieku osób młodszych oraz zaniżania wieku osób starszych.

Metodologia i wyniki badań - w pracy zbadano, czy wykorzystanie dostrojonych wektorów reprezentacji mówcy (speaker embeddings), pierwotnie przeznaczonych do weryfikacji tożsamości mówcy, można skutecznie zaadaptować do zadania predykcji wieku.

-Ekstrakcja cech (feature extraction): wstępne przetwarzanie sygnału przeprowadzono z użyciem modułu Wav2Vec2FeatureExtractor, który przyjmuje surowe próbki dźwiękowe i generuje ich reprezentacje. Normalizacja do zerowej wartości średniej i jednostkowej wariancji (zero-mean unit variance) ujednolica sygnał wejściowy dla modeli uczenia głębokiego i przekształca surowy przebieg w postać oczekiwaną przez model. Moduł ten jest zintegrowany z implementacją każdego z ocenianych modeli osadzeń.

-Architektura modeli i dostrajanie: wykorzystano głębokie modele reprezentacji mowy Wav2Vec2, WavLM oraz UniSpeech, dostrojone (fine-tuned) na zbiorze VoxCeleb1. Każdy z testowanych modeli, trenowany do zadania rozróżniania mówców, przekształca wypowiedzi o zmiennej długości w osadzenia o stałym wymiarze, określane jako x-wektory (x-vectors). Dostrajanie przeprowadzono w środowisku S3PRL, traktując modele jako ekstraktory cech wyposażone w głowicę x-vector uwzględniającą tożsamość mówcy.

-Konfiguracja eksperymentu: modele trenowano na zbiorze VoxCeleb1. Model Wav2Vec2 wstępnie wytrenowano na mowie próbkowanej z częstotliwością 16 kHz i dostrojono do zadania weryfikacji mówcy z wykorzystaniem benchmarku SUPERB (Speech processing Universal PERformance Benchmark). Model WavLM poprawia skuteczność w zadaniach innych niż rozpoznawanie mowy (ASR) dzięki odszumianiu sygnału i przewidywaniu oryginalnej mowy w zamaskowanych fragmentach; wykorzystany w eksperymentach model wavlm-base-plus-sv wstępnie wytrenowano na mowie 16 kHz i dostrojono z głowicą X-Vector przy użyciu kontrastowej funkcji straty na poziomie mówcy (speaker contrastive loss). Modele UniSpeech, tj. unispeech-sat-base-plus-sv oraz unispeech-sat-large-sv, wstępnie wytrenowano na mowie 16 kHz z kontrastową funkcją straty na poziomie wypowiedzi i mówcy.

-Model WavLM–ECAPA: dodatkowo model wavlm-large zastosowano jako ekstraktor cech z głowicą ECAPA-TDNN w celu wytrenowania własnego modelu osadzeń mówcy.

-Odwzorowanie regresyjne: do mapowania osadzeń mówcy na wiek zastosowano estymatory: LASSO, Ridge oraz klasyczną regresję liniową (LinearRegression). Ich zastosowanie umożliwiło porównanie dostrojonych modeli w jednolitych warunkach.

-Ewaluacja: uzyskane reprezentacje oceniono na zbiorze AgeVoxCeleb. Do oceny modeli wykorzystano średni błąd bezwzględny (MAE). Miara ta jest łatwa w interpretacji, ponieważ wyraża średni błąd w jednostkach zmiennej docelowej (latach), a dzięki liniowemu traktowaniu błędów pojedyncze skrajne predykcje, wynikające ze zróżnicowania cech głosu, nie zaburzają nadmiernie oceny modelu.

Główny wniosek z przeprowadzonych eksperymentów: konfiguracja **WavLM–ECAPA–LASSO** osiągnęła najniższy średni błąd bezwzględny (**MAE**). Wykazano, że dostrojone osadzenia mówcy skutecznie odwzorowują cechy akustyczne związane z wiekiem, ograniczając wariancję predykcji.

Paper ID: **14**

ODPOWIEDZI IMPULSOWE POMIESZCZENIA NA POTRZEBY TECHNIK DŹWIĘKU IMMERSYJNEGO – REJESTRACJA I PRZETWARZANIE.

Witold Mickiewicz*, Kaja Kosmenda
Katedra Inżynierii Systemów, Sygnałów i Elektroniki, Wydział Elektryczny,
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Autor korespondencyjny: **Witold Mickiewicz**, *witold.mickiewicz@zut.edu.pl*

Słowa kluczowe: dźwięk immersyjny, pogłos splotowy, odpowiedź impulsowa pomieszczenia, natężenie dźwięku

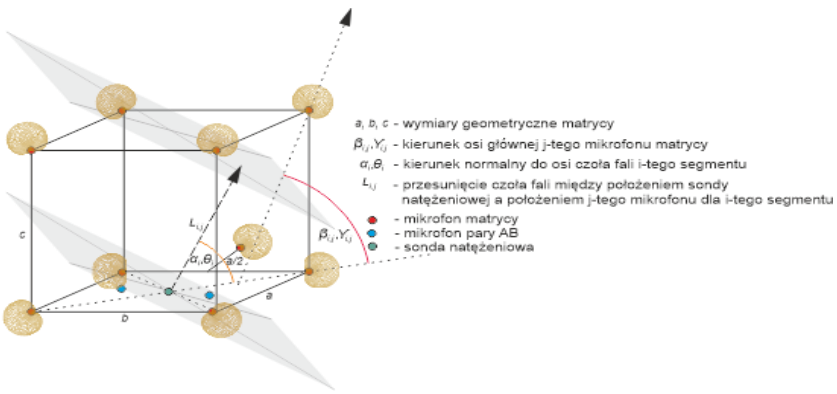
1. WPROWADZENIE

Technologia dźwięku immersyjnego daje możliwość reprodukcji dźwięku z zachowaniem jego pełnych atrybutów przestrzennych. Dotyczy to zarówno lokalizacji źródeł w przestrzeni 3D wokół słuchacza jak i przestrzennego odczuwania efektów pogłosowych związanych z akustyką lokalizacji, w jakiej umieszczone są poszczególne źródła. W dalszej części rozważań ograniczono się do wykorzystania technologii dźwięku immersyjnego do realizacji wysokiej jakości nagrań muzyki akustycznej. W takim przypadku dąży się do realizacji fonogramów, które odtwarzane w nagłośnieniowym systemie immersyjnym (np. system głośnikowy w konfiguracji 7-1-4) utworzą w świadomości słuchacza wrażenia tożsame z tymi, jakie odczuwa w sali koncertowej. Aby osiągnąć ten cel, wiele badań poświęconych jest obecnie optymalizacji układów matryc mikrofonowych zarówno do wykonywania pełnych nagrań jak i rejestracji odpowiedzi impulsowych do procesorów splotowych. W literaturze opisywane są eksperymenty polegające na realizacji nagrań muzyki klasycznej wykonywanych jednocześnie za pomocą wielu układów mikrofonowych, a następnie porównywane są subiektywne odczucia dotyczące reprodukcji wrażeń barwowych i przestrzennych [1]. W konkluzjach takich badań nie ma zazwyczaj jednoznacznych ocen, które z użytych konfiguracji są najlepsze. Często mówi się również, że optymalny wybór zależy od rodzaju muzyki i warunków akustycznych nagrania. W tej materii zawsze mamy do czynienia z subiektywnymi preferencjami reżysera dźwięku i słuchacza. Wydaje się więc słuszne twierdzenie, że umożliwienie słuchaczowi wyboru, w jakiej wersji omikrofonowania ma być utworzone jego nagranie, może być atrakcyjne i pożądane. Współczesna technologia może to umożliwić. Wraz z ideą dźwięku immersyjnego pojawiła się bowiem technologia dźwięku obiektowego, np. format DolbyAtmos. Przed erą dźwięku obiektowego systemy dźwięku przestrzennego bazowały na idei kanałowej. Oznaczało to, że fonogram zawierał określoną liczbę sygnałów kanałowych wygenerowanych u producenta nagrania, umożliwiających osiągnięcie zamierzonych efektów dźwiękowych tylko przy reprodukcji w systemie głośnikowym, dla którego zostało ono zrealizowane. Jednym z nowych aspektów technologii obiektowej jest kodowanie nagrania z wykorzystaniem metadanych, dzięki czemu sygnały dostarczane ostatecznie do głośników w systemie odbiorcy końcowego są generowane (renderowane) po stronie słuchacza w zależności od tego, jakim systemem głośnikowym dysponuje. Tak więc w systemie dźwięku obiektowego pojawia się po stronie odbiorcy aplikacja do odtwarzania dźwięku wyposażona w silnik DSP, korzystająca ze strumienia danych audio

oraz metadanych i generująca w czasie rzeczywistym sygnały kanałowe przeznaczone dla głośników w systemie odbiorcy. W prezentowanym artykule przedstawiono pomysł kodowania nagrań muzyki akustycznej umożliwiający zmianę subtelnych walorów brzmieniowych przestrzennych poprzez wybór po stronie słuchacza techniki mikrofonowej zastosowanej przy nagraniu.

2. IDEA, EKSPERYMENT I PRZETWARZANIE

W proponowanym podejściu potrzebne wielokanałowe immersyjne odpowiedzi impulsowe pomieszczenia syntezowane są na podstawie zarejestrowanych w rzeczywistym pomieszczeniu dwukanałowych impulsowych odpowiedzi ciśnieniowych i natężeniowych oraz informacji o strukturze matrycy mikrofonowej i rodzaju użytych mikrofonów. W dalszej części artykułu przedstawiono kroki postępowania dla prostopadłościennej matrycy 9-mikrofonowej i mikrofonów kardioidalnych. W dostępnej na Wydziale Elekrycznym ZUT w Szczecinie dużej sali audytoryjnej (auli) zarejestrowano jednocześnie odpowiedzi impulsowe pomieszczenia za pomocą prostopadłościennej matrycy 9-mikrofonowej z mikrofonami kardioidalnymi umieszczonymi w każdym narożniku pod odpowiednim kątem. Ponadto odpowiedzi zarejestrowano za pomocą pary mikrofonów dookólnych w konfiguracji AB oraz za pomocą sondy natężeniowej 3D Microflown. Sygnały pomiarowe typu sweep emitowano przez głośnik umieszczany w różnych lokalizacjach względem matrycy. Do syntezy immersyjnych odpowiedzi kanałowych wykorzystano inteligentną metodę segmentacji odpowiedzi impulsowych opisaną w [2]. Dzięki informacji kierunkowej z sondy natężeniowej każdy segment czasowy odpowiedzi z mikrofonów AB przed umieszczeniem w wyjściowych odpowiedziach kanałowych modyfikuje się amplitudowo w zależności od orientacji czoła fali względem orientacji danego mikrofonu kardioidalnego oraz przesuwa w czasie zgodnie z opóźnieniem wynikającym z przestrzennego usytuowania mikrofonów w narożnikach matrycy o konkretnych wymiarach geometrycznych. Pokazano to schematycznie na rys.1



Rys. 1. Model geometryczny układu matryca mikrofonowa-pole akustyczne wykorzystany do syntezy odpowiedzi impulsowych pomieszczenia

3. WNIOSKI

Wstępne testy odsłuchowe potwierdzają przydatność metody do kreowania wrażeń immersyjnych i upmixingu nagrań dwukanałowych. Poznanie szczegółowych właściwości metody wymaga jeszcze dalszych badań. Zaproponowany model umożliwia łatwą parametryzację użytych mikrofonów (również w zakresie zmienności właściwości kierunkowych w funkcji częstotliwości) oraz bardziej zaawansowane wyznaczanie opóźnień poprzez zastosowanie zmiennych w funkcji częstotliwości kształtów czoła fali.

LITERATURA

[1] ECHO PROJECT https://apl-hud.com/echo/
[2] MICKIEWICZ W., KOSMENDA K., *Creation of immersive experience for headphone listening using directional room impulse responses for BRIR synthesis.* W: W: International Journal of Electronics and Telecommunications Tom: 71, Zeszyt: 4, 2025.

Paper ID: **15**

MODELOWANIE NIEPEWNOŚCI POMIAROWEJ PRZY REJESTRACJI ODPOWIEDZI IMPULSOWEJ Z WYKORZYSTANIEM MIKROFONU AMBISONICZNEGO

Maciej Jasiński*, Jan Żera
Instytut Radioelektroniki i Technik Multimedialnych, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
Autor korespondencyjny: **Maciej Jasiński**, *maciej.jasinski@pw.edu.pl*

Słowa kluczowe: niepewność pomiarowa, odpowiedź impulsowa, akustyka pomieszczeń, mikrofon ambisoniczny, matryca mikrofonowa

1. WPROWADZENIE

Zagadnienie niepewności pomiarowej w akustyce pomieszczeń jest problemem złożonym, ponieważ wpływ na wynik pomiaru ma nie tylko jakość wykorzystanego układu pomiarowego, ale także m.in. precyzja i powtarzalność ustawień źródła dźwięku, czy też warunki środowiskowe w pomieszczeniu. Sposób wyznaczania niepewności pomiarowej w przypadku powszechnie wykorzystywanych mikrofonów pomiarowych o pojedynczej membranie jest tematyką znaną w literaturze i opisaną m. in. w pracach Lundeby i in. [1], a także bardziej szczegółowo w pracy Witewa [2]. Biorąc pod uwagę rodzaj pomiaru, sposób określania niepewności pomiarowej czasu pogłosu T20 i T30 został również określony w normie PN-EN ISO 3382-1 [3]. Natomiast we współczesnych technikach pomiarowych, gdzie wykorzystywane są systemy komputerowe obliczające odpowiedź impulsową na podstawie pobudzenia tonem przestrajanym lub sygnałem pseudolosowym MLS [4], analizę niepewności pomiarowej powinno się przeprowadzić zgodnie z instrukcjami Przewodnika Wyrażania Niepewności Pomiaru [5]. Niniejsza praca przedstawia rozszerzenie budżetu niepewności pomiarowej dla przetwornika wielomikrofonowego, w postaci mikrofonu ambisonicznego lub sferycznej matrycy mikrofonowej [6].

2. METODYKA

Przedstawiona analiza budżetu niepewności oparta jest na metodzie Witewa [2]. Uwzględniono m.in. wpływ kierunkowości źródła dźwięku, czynniki związane z kalibracją systemu, parametrami źródła dźwięku, czy też warunkami środowiskowymi w pomieszczeniu. Dla jednomembranowego mikrofonu pomiarowego, typowy wynik niepewności złożonej określonej na podstawie pełnego budżetu niepewności jest rzędu 0,6 dB i odnosi się do amplitudy odbić zarejestrowanej odpowiedzi impulsowej. W celu uzyskania budżetu niepewności pomiarowej pozwalającego na uwzględnienie wpływu mikrofonu ambisonicznego, bądź też sferycznej matrycy mikrofonowej, zmodyfikowano metodę Witewa poprzez uwzględnienie czynników związanych z kolejnymi kapsułami mikrofonowymi stanowiącymi matrycę, a także wzięto pod uwagę wpływ zaburzenia pola akustycznego wprowadzanego ogólnie przez matrycę mikrofonową. Zaproponowano zależność wyznaczania niepewności pomiaru odpowiedzi impulsowej matrycą mikrofonową według wzoru (1).

$$u_{c_n} = \sqrt{n * u_c^2 + u^2\left(b_{Mic-field}\right)}, \quad (1)$$

gdzie: u_{c_n} – niepewność złożona matrycy n-mikrofonowej, n – liczba mikrofonów stanowiących matrycę, u_c – niepewność złożona pomiaru odnosząca się do pojedynczego mikrofonu, $u(b_{Mic\text{-}field})$ - zakłócenie pola akustycznego spowodowane obecnością matrycy mikrofonowej.

3. WYNIKI I WNIOSKI

Oszacowanie niepewności przeprowadzono dla dwóch przypadków. W pierwszym przyjęto, że wszystkie mikrofony tworzące matrycę są mikrofonami pomiarowymi, których niepewność złożona wynosi tylko 0,6 dB. W drugim przypadku przyjęto wartość odpowiadającą mikrofonom niższej jakości, zakładając w przybliżeniu dwukrotnie większą wartość składowych niepewności, czyli 1,1 dB. Jak można się spodziewać, w obu przypadkach uzyskane wartości dla całej matrycy wskazują na wzrost niepewności złożonej wraz ze zwiększaniem liczby mikrofonów tworzących matrycę. Dla matrycy 4-mikrofonowej uzyskano odpowiednio 1,2 lub 2,3 dB, natomiast dla matrycy 19-mikrofonowej 2,7 lub 5,0 dB, w zależności od przyjętej niepewności pojedynczego mikrofonu.

Opisane podejście analizy niepewności pomiarów akustycznych wykonywanych z wykorzystaniem mikrofonów ambisonicznych, może być stosowane w przypadku wszelkich zastosowań wykorzystujących matryce mikrofonowe przy rejestracji poziomu ciśnienia dźwięku.

LITERATURA

[1] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of measurements in room acoustics," *Acta Acustica united with Acustica,* vol. 81, no. 4, pp. 344-355, 1995.

[2] I. Witew, "Measurements in room acoustics: Uncertainties and influence of the measurement position."

[3] *PN-EN ISO 3382-1:2009 Akustyka -- Pomiar parametrów akustycznych pomieszczeń -- Część 1: Pomieszczenia specjalne*, PKN, 2009.

[4] *PN-EN ISO 18233:2006 Akustyka -- Zastosowanie nowych metod pomiarowych w akustyce budynku i pomieszczeń*, PKN, 2006.

[5] G. U. Miar, *Ewaluacja danych pomiarowych - Przewodnik wyrażania niepewności pomiaru*. 2008.

[6] M. Jasiński and J. Żera, "Application of different types of microphones in room impulse response measurements," in *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2023: IEEE, pp. 1-6.

Paper ID: **16**

DŹWIĘK JAKO EFEKT STRUKTURALNYCH ZABURZEŃ PRZEPŁYWÓW STRUMIENIOWYCH. KSZTAŁTOWANIE WEKTOROWYCH PÓL DŹWIĘKU I HAŁASU

Stefan Weyna
Katedra Mechaniki Konstrukcji, Wydział Techniki Morskiej i Transportu, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Autor korespondencyjny: **Stefan Weyna**, *stefan.weyna@zut.edu.pl*

Słowa kluczowe: akustyka wektorowa, wizualizacja dźwięku, optymalizacja akustyczna.

Autor podejmuje próbę reinterpretacji natury dźwięku w świetle hydrodynamicznych opisów zawartych w analogach mechano-akustycznych (J. Lighthill, A. Powell, M.E. Goldstein, Howe M.C.), w nowej, wirowej teorii dźwięku (Vortex Sound Theory, VST) i graficznych analizach wyników badań wektorowych struktur dźwiękowych (Acoustical Imaging). Istotny potencjał tej pracy nie jest kierowany wyłącznie na „nową teorię dźwięku" lecz ma podkreślić rolę eksperymentalnych badań wektorowych jako źródła nowej wiedzy akustycznej sugerując koncepcję przejścia od akustyki parametrycznej do akustyki strukturalnej, w której dźwięk jest nie tylko efektem falowym, ale również efektem strukturalnych zaburzeń przepływu. Uzasadnienie tak stawianej tezy Autor widzi w udokumentowanych wynikach badań eksperymentalnych prowadzonych własną metodą SI_AOD+AI (Sound Intensity, Acoustic Orthogonal Decomposition + Acoustical Imaging) prowadzonych z użyciem sond natężeniowych i laserowych technik bezinwazyjnych PIV i LDA dostosowanych do badań zaburzonych przepływów akustycznych. W artykule przedstawiona jest metodologia badań SI_AOD+AI wykorzystana do optymalizacji źródeł dźwięku i hałasu we wnętrzu i na wylocie falowodów akustycznych z otwartym końcem będących synonimem przemysłowych systemów typu HVAC. Autor podkreśla zalety wektorowej wizualizacji akustycznej stosowane w naukowych analizach wyników badań i przemysłowych pracach inżynierskich. Wskazuje na metodę SI_AOD+AI jako efektywne i wiarygodne narzędzie do prowadzenia walidacji wyników modelowania numerycznego.

Paper ID: **17**

OD FALI AKUSTYCZNEJ DO STRUKTUR PRZEPŁYWOWO-AKUSTYCZNYCH

Stefan Weyna
Katedra Mechaniki Konstrukcji, Wydział Techniki Morskiej i Transportu, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Autor korespondencyjny: **Stefan Weyna**, *stefan.weyna@zut.edu.pl*

Klasyczna akustyka od ponad stu lat opisuje dźwięk przede wszystkim jako falę ciśnienia rozchodzącą się w ośrodku sprężystym. Opis ten okazał się niezwykle skuteczny zarówno w teorii, jak i w praktyce technicznej. Umożliwił rozwój elektroakustyki, diagnostyki hałasu, architektury akustycznej, telekomunikacji oraz nowoczesnych metod pomiarowych. Jednocześnie jednak dominujący model akustyki liniowej doprowadził do utrwalenia uproszczonego obrazu zjawisk akustycznych, w którym podstawowym nośnikiem informacji stał się skalarny parametr poziomu ciśnienia akustycznego SPL, zaś sama fala akustyczna zaczęła być traktowana niemal jako byt samodzielny, oderwany od złożonej dynamiki przepływów fizycznych.

Współczesna hydrodynamika i aeroakustyka pokazują jednak, że rzeczywiste procesy generacji dźwięku są znacznie bardziej złożone. W wielu układach przepływowych źródłem emisji akustycznej okazują się niestabilności hydrodynamiczne, bifurkacje, zrywanie wirów, reorganizacje struktur przepływu oraz przejścia do turbulencji. Mechanizmy te nie mają charakteru czysto skalarnego, lecz przestrzenny, dynamiczny i strukturalny. Znane z mechaniki płynów struktury koherentne - wiry podkowiaste, struktury typu hairpin, oscylacyjne komórki modalne czy niestabilności Hopfa - mogą być interpretowane nie tylko jako zjawiska przepływowe, ale również jako rzeczywiste organizatory emisji akustycznej.

W tym kontekście tradycyjne utożsamianie dźwięku wyłącznie z falą ciśnienia może okazać się opisem niepełnym. Fala akustyczna pozostaje oczywiście poprawnym rozwiązaniem matematycznych równań ruchu ośrodka, lecz być może opisuje jedynie projekcję bardziej złożonych procesów przepływowo-strukturalnych zachodzących w rzeczywistych układach fizycznych. Dźwięk można wtedy rozumieć nie tylko jako propagację zaburzenia, ale jako manifestację utraty stabilności przepływu i pojawiania się uporządkowanych struktur dynamicznych w materii.

Nowe możliwości interpretacji dźwięku pojawiają się szczególnie dzięki rozwojowi metod graficznego obrazowania wyników badań - Acoustical Imaging. W przeciwieństwie do klasycznych metod opartych wyłącznie na pomiarze SPL, obrazowanie akustyczne umożliwia rekonstrukcję przestrzennej organizacji pola akustycznego. Struktury wcześniej ukryte za pojedynczym parametrem energetycznym zaczynają ujawniać swoją geometrię, symetrię, koherencję i modalną organizację. Pola akustyczne przestają przypominać anonimowe rozkłady amplitud, a zaczynają wykazywać cechy charakterystyczne dla złożonych układów przepływowych podobne do tych, jakie badane są w mechanice płynów.

W tym sensie Acoustical Imaging może dla akustyki pełnić rolę podobną do tej, jaką techniki wizualizacji przepływów odegrały wcześniej w hydrodynamice (platformy obliczeniowe CFD). Obrazy nie stają się jedynie ilustracją wyników pomiarowych, lecz są nowym narzędziem poznania fizycznego. Możliwość obserwacji struktur przepływowo-akustycznych może prowadzić do stopniowej zmiany paradygmatu przechodząc od akustyki parametrycznej ku akustyce strukturalnej.

Jedną z prób uporządkowania tego nowego sposobu myślenia może być rozwijana nowa teoria dźwięku Vortex Sound Theory (VST), traktująca dźwięk jako zjawisko nierozerwalnie związane z dynamiką struktur wirowych i destabilizacją przepływów. W takim ujęciu źródłem akustyki nie jest już wyłącznie lokalna kompresja i dekompresja ośrodka, lecz cała organizacja przepływu, jego niestabilności, sprzężenia modalne, procesy przejściowe i krytyczne. Teoria ta nie musi negować

klasycznej teorii falowej, lecz może ją rozszerzać, podobnie jak współczesna mechanika turbulencji rozszerzyła dawny opis przepływów laminarnych.

Konsekwencje takiego podejścia mogą wykraczać poza samą fizykę akustyki. Jeżeli dźwięk jest częścią bardziej złożonych struktur przepływowo-dynamicznych, wówczas jego oddziaływanie na organizmy żywe nie musi ograniczać się wyłącznie do słuchu. Drgania, mikroprzepływy, modulacje ciśnienia i organizacja przestrzenna pól mogą wpływać na mechanorecepcję, układ somatosensoryczny, percepcję wibracyjną oraz inne formy biologicznej detekcji dynamicznych zmian środowiska. Dźwięk przestaje być wtedy wyłącznie „tym, co słyszymy", a staje się szczególną formą oddziaływania dynamicznej materii na organizmy żywe.

Być może współczesna akustyka znajduje się obecnie w podobnym momencie, w jakim mechanika płynów znajdowała się przed epoką nowoczesnej wizualizacji przepływów. Tak jak obserwacja struktur turbulentnych zmieniła rozumienie przepływu, tak rozwój obrazowania akustycznego może stopniowo zmienić sposób rozumienia samego dźwięku. Nie oznacza to konieczności odrzucenia klasycznej akustyki, lecz możliwość jej rozszerzenia o wymiar strukturalny, przestrzenny i hydrodynamiczny.

W takim ujęciu przyszłość akustyki może należeć nie tylko do analizy fal, lecz również do analizy organizacji przepływowo-akustycznych struktur materii.

Paper ID: **18**

AUTOMATYCZNA OCENA JAKOŚCI MOWY KLONOWANEJ

Daniel Górski
Politechnika Warszawska
Autor korespondencyjny: **Daniel Górski**, *daniel.gorski345@gmail.com*

Słowa kluczowe: przetwarzanie dźwięku, uczenie maszynowe, synteza mowy, klonowanie głosu, automatyczna ocena jakości, sieci neuronowe

Generowanie mowy za pomocą sztucznej inteligencji osiągnęło poziom, w którym odróżnienie jej od ludzkiego głosu bywa dużym wyzwaniem. Pojawia się zatem potrzeba stworzenia narzędzi, które automatycznie i obiektywnie ocenią jakość takich nagrań. Celem projektu było zaprojektowanie i weryfikacja systemu do oceny mowy klonowanej pod kątem jej naturalności, zrozumiałości oraz podobieństwa do oryginalnego mówcy. W ramach projektu wykorzystano 6000 ujednoliconych próbek audio - 50 nagrań referencyjnych 40 osób z bazy VCTK oraz ich klony wygenerowane przez modele XTTS-V2 i OmniVoice. System trenowano na etykietach z trzech algorytmów: NISQA (ocena ogólna MOS), STOI połączonym z DTW (zrozumiałość) oraz ECAPA-TDNN (tożsamość mówcy). Model sieci neuronowej, bazujący na architekturze Wav2Vec 2.0, analizował próbki na dwa sposoby: przewidywał MOS bezpośrednio z nagrania, a do oceny STOI i ECAPA-TDNN porównywał oryginał z próbką sklonowaną. Wytrenowany model najskuteczniej przewiduje ocenę MOS, osiągając korelację r=0.548 i błąd MAE na poziomie 0.236. Dla STOI oraz tożsamości (SIM) system uzyskał MAE wynoszące odpowiednio 0.080 i 0.104 (przy r=0.304 i r=0.368). Ciekawą obserwacją z etapu tworzenia zbioru było to, że model NISQA ocenił klony OmniVoice wyżej niż mowę prawdziwych ludzi (np. ocena ciągłości 4.60 w porównaniu do 4.54 dla człowieka), preferując jej cyfrową sterylność.

Paper ID: **19**

WPŁYW ZŁOŻONOŚCI LINGWISTYCZNEJ MOWY NA SKUTECZNOŚĆ ROZPOZNAWANIA SYNTETYCZNEGO GŁOSU PRZEZ LUDZI

Jan Kucharczyk*, Bartłomiej Kruk
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Jan Kucharczyk**, *jan.kucharczyk@pwr.edu.pl*

Słowa kluczowe: synteza mowy, zero-shot TTS, klonowanie głosu, audio deepfake

Rozwój modeli syntezy mowy typu zero-shot sprawia, że odróżnienie głosu sztucznego od naturalnego staje się wyzwaniem dla ludzkiego słuchu. Większość dotychczasowych badań skupia się na ocenie mowy ciągłej, pomijając wpływ samej struktury i złożoności materiału językowego na procesy percepcyjne słuchacza. Niniejsza praca analizuje, jak poziom złożoności lingwistycznej wypowiedzi wpływa na zdolność ludzi do detekcji głosów syntetycznych. W eksperymencie wykorzystano materiał dźwiękowy z bazy mowy polskiej, nagranej w komorze bezechowej, podzielony na trzy poziomy złożoności: logatomy, pojedyncze słowa oraz pełne zdania. Próbki te sklonowano za pomocą współczesnych modeli syntezy mowy. W teście odsłuchowym badani oceniali autentyczność nagrań oraz poziom pewności swojej decyzji. Badanie pozwala ustalić, czy wyższa złożoność językowa angażuje zasoby poznawcze i maskuje mikro-artefakty syntezy.

Paper ID: **20**

AUTOMATYCZNA DETEKCJA DEEPFAKE'ÓW AUDIO W SIECIACH TELEKOMUNIKACYJNYCH Z WYKORZYSTANIEM UCZENIA MASZYNOWEGO NA OBIEKTYWNYCH PARAMETRACH AKUSTYCZNYCH

Jakub Warzych* [1], Michał Łuczyński [2]
1) Katedra Telekomunikacji i Teleinformatyki, Wydział Informatyki i Telekomunikacji, Politechnika Wrocławska
2) Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Jakub Warzych**, *263858@student.pwr.edu.pl*

Słowa kluczowe: audio deepfake, klonowanie głosu, uczenie maszynowe, kanały telekomunikacyjne, GSM, VoLTE, VoIP

Rozwój generatywnej syntezy mowy sprawia, że sztucznie wygenerowane głosy stają się trudne do odróżnienia przez człowieka. Badania percepcyjne wykazują, że skuteczność detekcji mowy syntetycznej w środowisku telekomunikacyjnym wynosi jedynie 54,8%, przy czym w kanale szerokopasmowym (VoLTE) spada poniżej poziomu losowego (44,0%). Niniejsza praca prezentuje system automatycznej klasyfikacji mowy, stanowiący rozszerzenie badań nad obiektywną analizą sygnałową w kanałach GSM, VoLTE i VoIP. Do wytrenowania modeli uczenia maszynowego wykorzystano zbiór obiektywnych parametrów akustycznych wyekstrahowanych z nagrań mowy naturalnej oraz syntetycznej. Wektor cech objął m.in. zmienność częstotliwości podstawowej, jitter, shimmer, HNR oraz środek ciężkości widma w różnych warunkach transmisji. Wyniki klasyfikacji zestawiono bezpośrednio z subiektywnymi ocenami słuchaczy oraz modelami bazowymi opartymi na cechach MFCC i Wav2Vec2. Badania te miały na celu weryfikację potencjału aplikacyjnego przyjętej metody w detekcji mowy syntezowanej w środowisku telekomunikacyjnym.

Paper ID: **21**

OCENA ZMIAN PERCEPCJI SŁUCHOWEJ REALIZATORA DŹWIĘKU W TRAKCIE DŁUGOTRWAŁEJ PRACY W WARUNKACH WYSOKIEGO POZIOMU CIŚNIENIA AKUSTYCZNEGO

Tomasz Kopciński*, Kacper Panek, Bartłomiej Kruk
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Tomasz Kopciński**, *tomasz.kopcinski@pwr.edu.pl*

Słowa kluczowe: Tymczasowe przesunięcie progu słyszenia (TTS), standardy bezpiecznego słuchania, realizator dźwięku, koncert, poziom ciśnienia akustycznego

Wpływ długotrwałej ekspozycji na wysoki poziom ciśnienia akustycznego podczas koncertów stanowi jedno z kluczowych wyzwań w pracy realizatora dźwięku. Przedstawione badania skupiają się na analizie tymczasowego przesunięcia progu słyszenia (ang. Temporary Threshold Shift – TTS) u osób poddanych działaniu hałasu o zróżnicowanej dynamice i czasie trwania. Podstawę teoretyczną analizy stanowią fizjologiczne mechanizmy zmęczenia słuchu oraz standardy bezpiecznego słuchania Światowej Organizacji Zdrowia (WHO).

Część eksperymentalną przeprowadzono podczas kilkunastu koncertów w Strefie Kultury Studenckiej Politechniki Wrocławskiej oraz Narodowym Forum Muzyki. Pomiary poziomu ciśnienia akustycznego na stanowisku FOH realizowano za pomocą kalibrowanego miernika klasy 1 Svantek SVAN977A. Stan narządu słuchu badano bezpośrednio przed i po wydarzeniach przy użyciu wdrożonej mobilnej aplikacji audiometrycznej w trybie samobadania.

Wyniki wskazują na bezpośrednią zależność między profilem akustycznym wydarzenia a stopniem zmęczenia słuchu. Podczas koncertów rockowych i alternatywnych, przy poziomach dźwięku rzędu 90 dB(A), odnotowano istotne przesunięcia progu słyszenia. Zmiany te koncentrowały się w paśmie 2000–6000 Hz, osiągając w skrajnych przypadkach blisko 10 dB w okolicy 3–4 kHz. Odnotowano również dużą zmienność osobniczą w podatności na zmęczenie słuchowe. Wnioski wskazują na konieczność ciągłego monitorowania poziomu ciśnienia akustycznego w czasie rzeczywistym.

Paper ID: **22**

NON-INTRUSIVE PEAQ APPROXIMATION USING DEEP NEURAL NETWORKS: A COMPARISON OF SINGLE-TASK AND MULTI-TASK APPROACHES

Ihar Balyka*, Sławomir K. Zieliński
Faculty of Computer Science, Białystok University of Technology, Białystok, Poland
Corresponding author: **Ihar Balyka**, *ihar.balyka@pb.edu.pl*

Keywords: Objective audio quality assessment, PEAQ, multi-task learning, deep neural networks, cascade architecture

1. INTRODUCTION

The Perceptual Evaluation of Audio Quality (PEAQ) method [1, 4] is a widely adopted standard for objective assessment of audio quality, particularly in the development and evaluation of audio codecs. It remains the only internationally standardized procedure for objective sound-quality estimation. However, it requires both the original and the degraded signals, which is often impractical. For exam-ple, it cannot be used in such applications as automated audio restoration or real-time sound quality monitoring. This motivates the development of a no-reference PEAQ approximation which works without the original audio.

A previous study has applied machine learning to approximate a full-reference method using a no-reference model [2], but significant room for improvement remains. A promising direction is multi-output learning [3], where auxiliary targets guide the model. In the basic version of PEAQ, the final Objective Difference Grade (ODG) is derived from eleven Model Output Variables (MOVs) and a Distortion Index (DI). Training a model to predict them in addition to ODG may help it learn the underlying perceptual relationships and improve the accuracy of ODG prediction.

The aim of this study was to create a non-intrusive method that approximates PEAQ without the original audio. This study compares different approaches: multi-task models and a single-output model predicting only ODG, using the results of the basic version of PEAQ as the baseline.

2. EXPERIMENTS AND RESULTS

The dataset was built using 367 unique music tracks, each from a unique musician. Every track was converted to mono and trimmed in duration to 7 seconds. Degradations included lossy coding with various codecs, additive white or pink noise, and packet loss simulation. Audio was converted to log-mel spectrograms and scored by the basic version of PEAQ.

First, we trained two models across four backbone sizes (0.5M, 2M, 4M, 8M parameters): a single-output model predicting only ODG, and a flat multi-task model predicting MOVs, DI, and ODG sim-ultaneously. The single-output model outperformed the flat multi-task model, indicating that auxiliary supervision through intermediate PEAQ variables does not improve ODG prediction.

We hypothesized that too many targets introduce noise and distract the model from the primary ODG target. Therefore, for the 4M backbone we trained two cascade variants. In both, the shared feature vector first predicts MOVs; then the shared vector and predicted MOVs are concatenated to predict DI; finally, all that data is concatenated to predict ODG. In the dynamic-weight cascade, all weights started at 1, and the weights of MOVs and DI were linearly decreased to 0. The idea was that MOVs might provide context early in training, while later the model should focus only on ODG.

To assess our models we used the PEAQ conformance set (BS1387) [4], CoreSV14 [5], ODAQ [6], the IIT-JMU-UGM Audio Dataset [7,8], and a proprietary set of degraded audio files. Table 1 presents a comparison of the PEAQ baseline method, the best single-output model, and the best multi-

output model. Compared with the PEAQ baseline, the developed models achieve higher performance on two datasets, comparable performance on one dataset, and lower performance on two datasets, including the PEAQ conformance set. These results indicate that, on the subjective datasets, the performance of our model is generally comparable to that of PEAQ, despite using only degraded signals at its input.

Table 1. Pearson correlation coefficients of the best-performing models with 4M backbone

	BS1387	CoreSV14	ODAQ	IIT-JMU-UGM	Proprietary set
PEAQ BASELINE	0.93	0.52	0.87	0.43	0.59
SINGLE-OUTPUT	0.43	0.85	0.47	0.41	0.65
MULTI-OUTPUT	0.36	0.83	0.53	0.14	0.38

3. SUMMARY

We developed and evaluated non-intrusive deep learning models that approximate PEAQ scores without requiring a reference signal. On three of the five datasets selected for evaluation, the single-output model achieved performance that was comparable to, or exceeded, that of the PEAQ method. Our method can be used in situations where the original audio is unavailable, such as real-time audio quality monitoring and automated audio restoration. The single-output model proved to be the most effective, as dynamic weighting of intermediate PEAQ variables did not yield statistically significant improvements. Overall, using MOVs and DI as auxiliary targets adds complexity without providing meaningful gains in ODG prediction.

REFERENCES

[1] THIEDE T.V., TREURNIET W.C., BITTO R., SCHMIDMER C., SPORER T., BEERENDS J.G., & COLOMES C., *PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality.* Journal of The Audio Engineering Society, 48 (2000): 3-29.

[2] KASPERUK A., ZIELINSKI S.K., *Non-intrusive method for audio quality assessment of lossy-compressed music recordings using convolutional neural networks*, International Journal of Electronics and Telecommu-nications, 70 (2024), 331-339.

[3] WU C. W., WILLIAMS P. A., WOLCOTT W., *A multitask teacher-student framework for perceptual audio quality assessment*, 29th European Signal Processing Conference (EUSIPCO). IEEE, 2021, 396-400.

[4] ITU-R BS.1387-2, *Method for objective measurements of perceived audio quality*, Geneva, International Telecommunication Union, 2023.

[5] CoreSV, Public multiformat listening test @ 96 kbps, July 2014, [Online]. Available: http://listening-test.coresv.net/index.htm

[6] TORCOLI M., WU C. W., DICK S., WILLIAMS P. A., HALIMEH M. M., WOLCOTT W., HABETS E. A. P., *ODAQ: Open Dataset of Audio Quality*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

[7] MUMTAZ D., JENA A., JAKHETIYA V., NATHWANI K., GUNTUKU S. C., *Transformer-based quality assessment model for generalized user-generated multimedia audio content*, Proc. INTERSPEECH, 2022, 674–678.

[8] MUMTAZ D., JAKHETIYA V., NATHWANI K., SUBUDHI B. N., GUNTUKU S. C., *Nonintrusive per-ceptual audio quality assessment for user-generated content using deep learning*, IEEE Transactions on In-dustrial Informatics, 18 (2022), 7780–7789.

Paper ID: **23**

ŚLEDZENIE RUCHU GAŁEK OCZNYCH Z WYKORZYSTANIEM METOD UCZENIA MASZYNOWEGO

Mikołaj Różycki
Politechnika Warszawska
Autor korespondencyjny: **Mikołaj Różycki**, *01149614@pw.edu.pl*

Słowa kluczowe: estymacja spojrzenia, interfejs człowiek-komputer, sieci neuronowe, mechanizm uwagi krzyżowej, złożoność obliczeniowa

Rozwój interfejsów człowiek-komputer wymaga precyzyjnych i przystępnych metod śledzenia punktu spojrzenia (ang. Point of Gaze, PoG) z użyciem standardowych kamer. Celem przeprowadzonych badań jest analiza porównawcza hierarchicznych architektur sieci neuronowych w zadaniu estymacji spojrzenia, ze szczególnym uwzględnieniem wpływu rozdzielczości wejściowej (64x64 vs 224x224) oraz rodzaju wyjścia modelu (współrzędne 2D na ekranie vs wektor kierunku w przestrzeni 3D).

Zaproponowane modele wykorzystują sieć EfficientNet-B0 do ekstrakcji cech z obrazów oczu i twarzy, dokonując ich fuzji z danymi geometrycznymi, w tym punktami kluczowymi z biblioteki MediaPipe oraz maską face grid. W wariancie o wyższej rozdzielczości (224x224) zastosowano moduły SPPF oraz mechanizm cross-attention, który dynamicznie wiąże orientację twarzy z informacją wizualną płynącą z oczu.

Eksperymenty prowadzone są na wycinkach otwartych baz danych, w tym zbiorach GazeCapture oraz ETH-XGaze. Dotychczasowe wyniki pokazują, że zaawansowana architektura w wariancie predykcji 2D osiąga średni błąd euklidesowy na poziomie 1,28 cm. Z kolei w wariancie estymacji wektora kierunku uzyskano średni błąd kątowy rzędu 3,73°. Co szczególnie istotne, autorskie rozwiązanie osiąga wyniki dokładności estymacji zbliżone do obecnego stanu wiedzy, oferując przy tym kilkukrotnie mniejszą złożoność obliczeniową oraz wagową sieci.

Paper ID: **24**

MODELOWANIE STANU ZIMUJĄCEJ RODZINY PSZCZELEJ NA PODSTAWIE POMIARÓW WIBROAKUSTYCZNYCH

Piotr Książek
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Piotr Książek**, *piotr.ksiazek@pwr.edu.pl*

Słowa kluczowe: Uczenie Maszynowe, Bioakustyka, Modelowanie Stanu, Dźwięk

Zimujące rodziny pszczele z samej natury procesu zawiązywania kłębu zimowego emitują sygnały wibroakustyczne. W trakcie trwania zimy pszczoły moderują temperaturę kłębu zimowego poprzez poruszanie skrzydełkami oraz mięśniami tułowia, co generuje dźwięki słyszalne przez człowieka. Referat dotyczy metody modelowania sygnału dźwiękowego rejestrowanego wewnątrz ula w okresie zimowym za pomocą technik uczenia maszynowego ze szczególnym względem na modelowanie stanu rodziny pszczelej przy użyciu modeli reprezentacji typu MMVAE. Wykazano wysoką zdolność predykcji zmian temperatury wewnątrz ula na podstawie analizy dźwięków zarejestrowanych mikrofonem zamontowanym w wylotku. Przedstawione badania rozszerzają aspekt monitorowania stanu rodziny pszczelej na okres zimowy i tworzą podstawę do dalszych eksperymentów dotyczących budowy reprezentacji cyfrowych dla rodzin pszczelich.

Paper ID: **25**

BEZKONTAKTOWA ESTYMACJA AKTYWNOŚCI ODDECHOWEJ I SERCOWEJ NA PODSTAWIE NAGRAŃ WIDEO MIKRORUCHÓW KLATKI PIERSIOWEJ PODCZAS ZADAŃ W JĘZYKU OJCZYSTYM I OBCYM

Miłosz Bakiera, Marcin Sutor, Kamil Sterniuk*, Bożena Kostek
Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska
Autor korespondencyjny: **Kamil Sterniuk**, *kamil.sterniuk@pg.edu.pl*

Słowa kluczowe: język obcy; wysiłek poznawczy; estymacja częstości oddechu; obserwacja aktywności fizjologicznej; analiza obrazów wideo z wykorzystaniem metody Eulerowskiego wzmacniania wideo

1. WPROWADZENIE

Bezkontaktowe monitorowanie aktywności fizjologicznej może umożliwić ocenę zmian zachodzących podczas wykonywania zadań poznawczych bez ingerencji w zachowanie uczestnika. Większość metod wizyjnego pomiaru tętna wykorzystuje zmiany fotometryczne odsłoniętej skóry, natomiast obserwacja ubranego torsu eliminuje tę metodę pomiaru i pozostawia przede wszystkim subtelne ruchy mechaniczne powierzchni odzieży. Jednocześnie ruch klatki piersiowej wywołany oddychaniem ma znacznie większą amplitudę niż ruch związany z pracą serca. Komponenty mechaniczne aktywności sercowej są znane z badań balisto- i sejsmokardiograficznych [2], natomiast częstość oddechu może być wyznaczana również na podstawie subtelnych ruchów rejestrowanych przez kamerę [3].
Celem badania była ocena, czy wieloskalowa analiza mikroruchów ubranego torsu, oparta na etapach przetwarzania wywodzących się z Eulerian Video Magnification (EVM) [4], umożliwia estymację częstości oddechu i rytmu serca oraz odtworzenie zmian fizjologicznych występujących podczas zadań wykonywanych w języku ojczystym i obcym.

2. MATERIAŁ I METODY

Przeanalizowano 29 sesji pomiarowych z udziałem uczestników. Scenariusz obejmował naprzemienne fazy relaksacji i zadania językowe wykonywane w języku polskim (L1) i angielskim (L2), różniące się poziomem trudności oraz blok czytania ze zrozumieniem. Aktywność fizjologiczną rejestrowano równocześnie przy użyciu kamery Phantom VEO-E 310L skierowanej na ubrany tors (640×480 pikseli, 24 kl./s), pasa oddechowego oraz piezoelektrycznego czujnika tętna podłączonych do systemu PowerLab (1000 Hz) oraz radaru mikrofalowego, pracującego jako dodatkowy tor bezkontaktowy. Pomiar radarowy oddechu stanowi sprawdzone podejście alternatywne wobec pasa kontaktowego [1].
Materiał wideo przetwarzano bezstratnie. Proces przetwarzania obejmował redukcję przestrzenną, analizę wielu skal obrazu, filtrację czasową oraz analizę składowych głównych, prowadzącą do utworzenia sygnałów kandydatów. Ze względu na brak wspólnego wyzwalacza sprzętowego, osie czasu poszczególnych urządzeń odtwarzano w postprodukcji. Zbiór podzielono na 15 sesji przeznaczonych do doboru parametrów metody i 14 niezależnych sesji walidacyjnych, według uczestników. Jako główne miary zastosowano średni błąd bezwzględny (MAE) oraz współczynnik korelacji z sygnałem referencyjnym. Wyniki porównywano także z predyktorem stałym opartym na medianie wartości w obrębie sesji.

3. WYNIKI

Częstość oddechu była wiarygodnie odtwarzana z ruchu ubranego torsu. Dla 20-sekundowych okien zadaniowych na zbiorze walidacyjnym uzyskano wartość błędu MAE równą 1,575 oddechu/min oraz współczynnik korelacji (r=0,792) z pomiarem referencyjnym. Błąd predyktora stałego wynosił w tym samym przypadku 2,329 oddechu/min, a metoda wizyjna osiągała lepszy wynik w 13 z 14 sesji. Wydłużenie przedziału analizy wyraźnie zwiększało dokładność. Dla całych bloków zadaniowych MAE spadał do 0,324 oddechu/min, a współczynnik korelacji wzrastał do (r=0,987); metoda była wówczas lepsza od predyktora stałego we wszystkich 14 sesjach walidacyjnych.
Zarejestrowany materiał obejmował zmienność częstości oddechu pomiędzy fazami relaksacji i blokami zadaniowymi, dzięki czemu skuteczność metody wizyjnej oceniano nie tylko w warunkach spoczynkowych, lecz również podczas wykonywania zadań.
W przypadku rytmu serca nie uzyskano analogicznej skuteczności. MAE estymaty z wideo wynosił 7,39 uderzenia/min wobec około 2,85 uderzenia/min dla predyktora stałego. Analizy sygnału z pasa oddechowego oraz radaru wskazały na obecność komponentu mechanicznego związanego z aktywnością serca, jednak jego stosunek do tła był niewystarczający do stabilnego śledzenia częstości serca.

4. DYSKUSJA I WNIOSKI

Uzyskane wyniki pokazują wyraźną granicę zastosowania analizy mikroruchów ubranego torsu. Ruch oddechowy zachowuje amplitudę i spójność przestrzenną wystarczające do dokładnej estymacji częstości oddechu, szczególnie przy dłuższych oknach czasowych. Co istotne, sygnał wizyjny odtwarza nie tylko wartość bezwzględną częstości oddechu, ale również jej zmiany występujące pomiędzy warunkami zadaniowymi.
W przypadku aktywności sercowej obecność składowej mechanicznej nie przełożyła się na użyteczną dokładność estymacji. Wynik wskazuje, że ograniczeniem nie jest wyłącznie czułość fotometryczna kamery, lecz także niski stosunek ruchu związanego z pracą serca do przestrzennie skorelowanego tła ruchowego oraz przenoszenie drgań przez warstwę odzieży. Liniowe wzmocnienie zmian obrazu nie poprawia tego stosunku, ponieważ wzmacnia zarówno składową użyteczną, jak i zakłócenia.
Badanie potwierdziło zatem przydatność wideo ubranego torsu do bezkontaktowego monitorowania oddychania i zmian oddechowych związanych z wykonywanym zadaniem, natomiast w badanych warunkach nie potwierdza możliwości wiarygodnego śledzenia tętna. Dalsze prace powinny obejmować wyższą częstotliwość rejestracji, referencję EKG, kontrolę geometrii i właściwości odzieży oraz metody zwiększające stosunek sygnału mechanicznego do tła zamiast jedynie zwiększać jego amplitudę.

LITERATURA

[1] CZYŻEWSKI A., KOSTEK B., KUROWSKI A. i in., Algorithmically improved microwave radar monitors breathing more acurrate than sensorized belt, Scientific Reports, 12, 2022, 14412.
[2] INAN O.T., MIGEOTTE P.-F., PARK K.-S. i in., Ballistocardiography and Seismocardiography: A Review of Recent Advances, IEEE Journal of Biomedical and Health Informatics, 19(4), 2015, 1414–1427.
[3] WANG W., DEN BRINKER A.C., Algorithmic insights of camera-based respiratory motion extraction, Physiological Measurement, 43(7), 2022, 075004.
[4] WU H.-Y., RUBINSTEIN M., SHIH E. i in., Eulerian Video Magnification for Revealing Subtle Changes in the World, ACM Transactions on Graphics, 31(4), 2012, 1–8.

Paper ID: **26**

ANALIZA METOD SYMULACJI ORGANÓW PISZCZAŁKOWYCH

Przemysław Kojs
Politechnika Wrocławska
Autor korespondencyjny: **Przemysław Kojs**, *przemek.kojs@gmail.com*

Słowa kluczowe: organy piszczałkowe, synteza dźwięku, modelowanie fizyczne, sampling, analiza sygnału audio, uczenie maszynowe

Symulacja organów piszczałkowych jest procesem niezwykle złożonym, obejmującym m.in. odwzorowanie charakterystyki poszczególnych piszczałek. Na przestrzeni ostatnich dekad opracowano wiele metod cyfrowej symulacji organów, opartych m.in. na technikach syntezy dźwięku, samplowaniu oraz modelowaniu fizycznym, a współcześnie również wykorzystujących metody uczenia maszynowego. Celem referatu jest syntetyczne przedstawienie rozwoju tej dziedziny, ze szczególnym uwzględnieniem zmian zachodzących wraz z ogólnym wzrostem dostępnych możliwości obliczeniowych. Omówione zostaną wybrane, reprezentatywne podejścia do generowania dźwięku organowego, ich założenia, możliwości oraz ograniczenia. Istotną część pracy stanowi porównanie rezultatów wybranych metod symulacji z nagraniami rzeczywistych organów piszczałkowych. Analiza przeprowadzona zostanie z wykorzystaniem obiektywnych metryk porównania sygnałów audio. Pozwoli to na ocenę stopnia odwzorowania wybranych cech rzeczywistego instrumentu przez poszczególne podejścia oraz wskazanie obszarów, w których omawiane metody symulacji nadal wykazują istotne ograniczenia.

Cel wystąpienia

Celem wystąpienia jest przedstawienie i porównanie trzech grup metod symulacji brzmienia organów piszczałkowych: metod opartych na samplingu, modelowaniu fizycznym oraz syntezie dźwięku. Metody te zostaną ocenione pod kątem percepcyjnego realizmu uzyskanego brzmienia, określonego za pomocą wybranych metryk obiektywnych. Analiza pozwoli wskazać kompromis pomiędzy realizmem brzmienia oferowanym przez poszczególne metody a ich wymaganiami obliczeniowymi. Podstawę analizy stanowić będzie referencyjne nagranie rzeczywistej piszczałki organowej, zestawione z brzmieniami uzyskanymi za pomocą poszczególnych metod symulacji komputerowej.

Plan wystąpienia

1. Wstęp o działaniu organów piszczałkowych (2 min)
2. Historia rozwoju dziedziny (3 min)
3. Problematyka prezentowanego zagadnienia (8 min)
   a. Omówienie wybranych metryk obiektywnego porównania percepcji dźwięku
      i. Widmowa odległość logarytmiczna (LSD)
      ii. Współczynniki mel-cepstralne (MFCC)
      iii. Zbieżność widmowa (SC)
   b. Synteza addytywna
   c. Jednowymiarowy model falowodowy
   d. Wielowymiarowy model fizyczny
   e. Sampling
4. Podsumowanie wyników, dyskusja (2 min)

1. WSTĘP O DZIAŁANIU ORGANÓW PISZCZAŁKOWYCH

Rozdział ma na celu przedstawienie złożoności organów piszczałkowych w kontekście zasad wytwarzania dźwięku przez różne typy piszczałek. W instrumencie wyróżnia się piszczałki labialne oraz stroikowe, których materiał wykonania, kształt i wymiary bezpośrednio wpływają na charakterystykę generowanego dźwięku. Piszczałki o zbliżonej konstrukcji grupowane są w tzw. głosy organowe. W obrębie danego głosu ich brzmienie powinno być możliwie spójne i jednolite, podczas gdy poszczególne głosy powinny zachowywać odmienny charakter brzmieniowy.

2. HISTORIA ROZWOJU DZIEDZINY

Technik związanych z próbami odtworzenia brzmienia organowego można doszukiwać się już w rozwiązaniach stosowanych w organach Hammonda, natomiast początki właściwej symulacji cyfrowej przypadają na lata 70. XX wieku. Stosowane metody rozwijały się wraz ze wzrostem możliwości obliczeniowych komputerów, dlatego techniki reprezentatywne dla kolejnych etapów rozwoju symulatorów pozostają ściśle związane z dostępną w danym okresie mocą obliczeniową. Przedstawione zostaną również wyniki analizy literatury dotyczącej szerzej rozumianej symulacji instrumentów dętych, pozwalające określić aktualny stan rozwoju dziedziny oraz wskazać potencjalne kierunki dalszych badań.

3. PROBLEMATYKA PREZENTOWANEGO ZAGADNIENIA

Referat skupia się na porównaniu różnych technik symulacji procesu wytwarzania dźwięku przez reprezentatywne piszczałki należące do czterech grup: labialnych pryncypałowych, labialnych fletowych, labialnych smyczkujących oraz stroikowych. Omówione zostaną metody oparte na syntezie addytywnej, jednowymiarowym modelu falowodowym, wielowymiarowych modelach fizycznych oraz samplingu. Wyniki uzyskane za pomocą poszczególnych metod zostaną porównane z nagraniem referencyjnym z wykorzystaniem obiektywnych metryk porównania sygnałów dźwiękowych: widmowej odległości logarytmicznej (LSD), współczynników mel-cepstralnych (MFCC) oraz zbieżności widmowej (SC).

4. PODSUMOWANIE WYNIKÓW, DYSKUSJA

Ostatnia część referatu obejmie podsumowanie przedstawionych zagadnień i uzyskanych wyników, a także wskazanie najważniejszych wniosków. Przewidziany zostanie również czas na pytania oraz dyskusję.

Paper ID: **27**

APLIKACJA INTERNETOWA DO TRENINGU SŁUCHANIA KRYTYCZNEGO

Michał Ładyżyński*, Franciszek Malewski, Adam Piwoński, Jakub Kostrzewa
Politechnika Warszawska
Autor korespondencyjny: **Michał Ładyżyński**, *michal.ladyzynski@gmail.com*

Słowa kluczowe: krytyczne słuchanie, lokalizacja źródła dźwięku, Web Audio API, HRTF, barwa dźwięku, aplikacja internetowa

1. WPROWADZENIE I CEL

Krytyczne słuchanie stanowi jedną z podstawowych kompetencji w pracy z dźwiękiem, zarówno w badaniach nad akustyką pomieszczeń, jak i w projektowaniu przestrzennych scen dźwiękowych oraz realizacji i miksowaniu materiału muzycznego [1], [2]. Istotnym elementem kształcenia w tym obszarze jest rozwijanie umiejętności identyfikowania słyszalnych zmian parametrów sygnału i powiązania ich z wartościami ilościowymi, np. częstotliwością modyfikowanego pasma, różnicą poziomu lub położeniem źródła dźwięku [3]. Celem pracy było opracowanie internetowej aplikacji umożliwiającej trening tych umiejętności w trzech obszarach: percepcji barwy, poziomu oraz lokalizacji źródła dźwięku. Aplikacja działa bez konieczności instalacji i umożliwia elastyczny dobór materiału dźwiękowego oraz poziomu trudności ćwiczeń.

2. KONCEPCJA I IMPLEMENTACJA APLIKACJI

Aplikację zaimplementowano z wykorzystaniem framework'u Django [4] (Python [5]) oraz standardowych technologii webowych. Kluczową decyzją projektową było odseparowanie przetwarzania sygnału od serwera. Serwer odpowiada za generowanie widoków i udostępnianie niemodyfikowanych próbek audio, natomiast przetwarzanie dźwięku odbywa się lokalnie w przeglądarce z użyciem Web Audio API [6]. Dzięki takiej strukturze zmiana parametrów próbki lub przełączenie próbki daje natychmiastową reakcję, a serwer nie musi wykonywać kosztownych obliczeń, co ułatwia obsługę wielu użytkowników jednocześnie.

W zadaniach porównawczych istotne jest, aby jedyną różnicą między próbkami był badany parametr. W niniejszej pracy zaimplementowano dwa mechanizmy ograniczające występowanie artefaktów. Zastosowano bezszwowe zapętlanie wykorzystujące przejście krzyżowe (ang. *crossfade*). Wybrany fragment jest normalizowany z zapasem dynamiki, a w modułach głośności dodatkowo zabezpieczony limiterem. Drugi mechanizm to płynne przełączanie próbek w architekturze dwóch ścieżek: przy przełączeniu bieżąca ścieżka jest wyciszana, a nowa równocześnie narasta w regulowanym czasie (domyślnie 15 ms), co pozwala wyeliminować kliknięcia.

System obejmuje pięć niezależnych modułów ćwiczeniowych. W module identyfikacji częstotliwości zadaniem użytkownika jest wskazanie pasma, w którym wprowadzono wzmocnienie. Wzmocnienie może wynosić od 0,5 dB do 12 dB, a szerokość modyfikowanego pasma od 1/3 do 2 oktaw, przy czym wartość ta jest przeliczana na dobroć filtru Q. Drugi moduł wykorzystuje procedure XAB do porównania charakterystyk częstotliwościowych próbek. Charakterystyka sygnału jest w tym przypadku kształ-

towana za pomocą pary filtrów górno- i dolnoprzepustowych wybieranych z zestawu ośmiu zdefiniowanych pasm, od pełnego pasma do zakresu 250–6500 Hz.

Dwa kolejne moduły dotyczą percepcji głośności. W module AB użytkownik wskazuje próbkę o większej głośności, przy czym różnica poziomów jest losowana w zadanym zakresie i z określonym krokiem. Dostępne są predefiniowane poziomy trudności od *Easy* do *Expert*. W module szacowania różnicy głośności użytkownik wybiera z siatki odpowiedzi wartość różnicy poziomu dźwięku wyrażoną w decybelach. Konfigurowalne są zakres wartości, krok, liczba dostępnych odpowiedzi oraz kierunek zmiany.

Piąty moduł służy do treningu lokalizacji przestrzennej. Zadaniem użytkownika jest wskazanie azymutu źródła dźwięku w zakresie od −90° do +90°. Przestrzenne położenie źródła jest realizowane z użyciem węzła *PannerNode* i modelu HRTF. Poziom trudności można regulować przez zmianę dopuszczalnego promienia błędu, a jako miara wyników rejestrowany jest m.in. średni błąd uzyskany w trakcie sesji.

Każdy moduł oferuje wizualizację przebiegu fali, wybór fragment nagrania, pracę z pojedynczą próbką dźwiękową lub próbkami losowanymi z określonego zestawu. Zbierana jest informacja o liczbie odpowiedzi poprawnych i błędnych oraz najdłuższej serii. Możliwy jest eksport wyników do pliku .csv.

3. PODSUMOWANIE

W ramach niniejszej pracy opracowano internetową aplikację do treningu krytycznego słuchania obejmującą ćwiczenia związane z percepcją barwy, różnic poziomu oraz lokalizacji przestrzennej źródła dźwięku. Przetwarzanie sygnału realizowane jest lokalnie w przeglądarce z wykorzystaniem Web Audio API, co umożliwia szybką zmianę parametrów ćwiczeń i ogranicza obciążenie warstwy serwerowej. Zastosowane mechanizmy płynnego zapętlania i przełączania próbek ograniczają powstawanie artefaktów mogących wpływać na ocenę słuchową, a możliwość konfiguracji parametrów trudności oraz rejestracji wyników pozwala dostosować ćwiczenia do różnych scenariuszy treningowych i dydaktycznych. Kierunki dalszego rozwoju to adaptacyjny dobór trudności, możliwość wygrywania własnych nagrań oraz badania z udziałem użytkowników.

LITERATURA

[1]H. Fastl i E. Zwicker, *Psychoacoustics: Facts and models*, 3. wyd., t. 22. w Springer Series in Information Sciences, vol. 22. Heidelberg: Springer Berlin, 2006. doi: https://doi.org/10.1007/978-3-540-68888-4.
[2]S. Elmosnino, „A Review of Literature in Critical Listening Education”, *Journal of the Audio Engineering Society*, t. 70, s. 328–339, maja 2022, doi: 10.17743/jaes.2022.0004.
[3]B. Dudziński, „Zastosowanie algorytmów adaptacyjnych do samodzielnej nauki solfeżu barwy”, praca magisterska, Wydział Reżyserii Dźwięku, Uniwersytet Muzyczny Fryderyka Chopina, Warszawa, 2026.
[4]Django Software Foundation, „Django documentation | Django documentation”, Django Project. Dostęp: 29 września 2026. [Online]. Dostępne na: https://docs.djangoproject.com/en/6.1/
[5]Python Software Foundation, „Welcome to Python.org”, Python.org. Dostęp: 29 września 2026. [Online]. Dostępne na: https://www.python.org/doc/
[6]W3C Audio Working Group, „Web Audio API 1.1”. Dostęp: 29 września 2026. [Online]. Dostępne na: https://www.w3.org/TR/webaudio-1.1/

Paper ID: **28**

PORÓWNANIE PARAMETRÓW OBIEKTYWNYCH MIKROFONÓW WIELKOMEMBRANOWYCH ORAZ ANALIZA POWTARZALNOŚCI ICH PARAMETRÓW MIĘDZY EGZEMPLARZAMI

Piotr Kluzowicz*, Bartłomiej Kruk
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów, Wydział Elektroniki, Fotoniki i Mikrosystemów, Politechnika Wrocławska
Autor korespondencyjny: **Piotr Kluzowicz**, *277746@student.pwr.edu.pl*

Słowa kluczowe: mikrofony wielkomembranowe, analiza porównawcza parametrów obiektywnych, tolerancja produkcyjna i kontrola jakości

W pracy przedstawiono analizę porównawczą parametrów obiektywnych mikrofonów wielkomembranowych, obejmującą oryginalne konstrukcje renomowanych producentów, ich odpowiedniki oraz dostępne na rynku repliki. Celem badań było określenie różnic pomiędzy poszczególnymi modelami na podstawie wybranych parametrów elektroakustycznych, a także ocena stopnia podobieństwa ich właściwości do mikrofonu referencyjnego. W ramach badań przeprowadzono pomiary wybranych parametrów obiektywnych, umożliwiające porównanie charakterystyk badanych mikrofonów w tych samych warunkach pomiarowych.

Drugim aspektem pracy była analiza powtarzalności parametrów w obrębie tego samego modelu mikrofonu. Porównano wyniki uzyskane dla kilku egzemplarzy poszczególnych modeli, zwracając szczególną uwagę na występujące między nimi rozbieżności. Pozwoliło to ocenić wpływ różnic produkcji na końcowe właściwości elektroakustyczne mikrofonów.

Przeprowadzone badania umożliwiają ocenę zarówno różnic pomiędzy poszczególnymi konstrukcjami, jak i rzeczywistej powtarzalności parametrów mikrofonów tego samego modelu.