ISSET 2025

Polish Audio Engineering Society

20th International Symposium
on Sound Engineering and
Tonmeistering 2025

20th International Symposium on Sound Engineering and Tonmeistering 2025

Poznań, 8.10-10.10.2025

Book of Abstracts (in order of presentation)

ASCOE: VR Platform for Investigating Spatial Hearing and Cross-Sensory Integration in 6DoF Environment

Paweł Perkowski[1], Bartłomiej Mróz[1], Przemysław Danowski[2]
[1] *Department of Multimedia Systems, Gdańsk University of Technology, Gabriela Narutowicza 11/12, Gdańsk, Poland*
[2]*Department of Sound Engineering, The Chopin University of Music, Okólnik 2, Warszawa, Poland*

E-mail: s180464@student.pg.edu.pl

ASCOE (Auditory Spatial & Cognitive Observation Environment) is a virtual reality (VR) platform designed for controlled studies of auditory localization and visuo-auditory integration in dynamic six degrees of freedom (6DoF) environment. Precise spatial audio reproduction is a fundamental component of immersive VR systems, particularly in applications requiring accurate auditory localization, such as perceptual research, sensorimotor rehabilitation, and training simulations. Despite substantial progress in spatial audio rendering and head-tracking technologies, empirical studies on auditory perception typically utilize three degrees of freedom: yaw, pitch and roll (rotations). VR environments allowing for the translatory movements, which together with rotations facilitate six degrees of freedom, remain a relatively unexplored area of study. This is primarily due to the lack of accessible platforms that support precise experimental control over both auditory and visual stimuli in dynamic, user-controlled scenes. In this paper, we present a dedicated standalone VR application, designed for conducting controlled experiments on spatial hearing and cross-modal integration in 6DoF while maintaining compatibility with standard VR head-mounted displays (HMDs).

The VR application was implemented in Unity, which served as the main development environment, employing the Wwise audio engine (developed by Audiokinetic Inc.) for sound management and Atmoky trueSpatial to enable dynamic spatial audio rendering. A visual overview of the implemented VR environment is provided in Figure 1. It supports the presentation of parametrically defined spatial sound sources, real-time manipulation of intermodal cue congruence, and structured trial-based data collection with sub-frame synchronization between auditory and visual stimuli. The platform enables investigation of perceptual metrics such as localization accuracy, response latency, and adaptation under conditions of audiovisual divergence. We demonstrate the utility of the platform via a pilot study exploring the effects of visual displacement on auditory localization in a dynamic 6DoF scenario. By reducing technical overhead and providing a reproducible experimental framework, the proposed platform facilitates systematic investigation of auditory localization, sensory integration, and perceptual plasticity in immersive environments. It is intended as a research tool for both auditory scientists and VR developers, bridging the gap between perceptual research and interactive 3D audio system design.

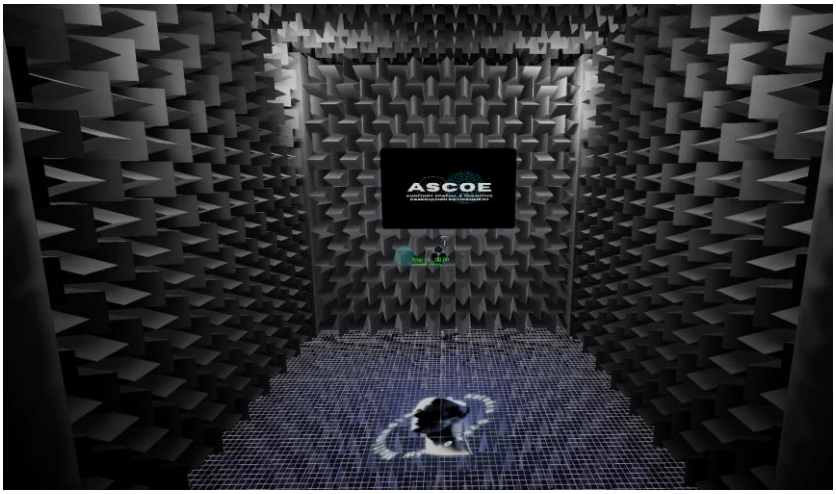

Figure 1: Visualization of the virtual reality environment developed for the experimental procedure.

References

1. T. Huisman, A. Ahrens, E. MacDonald, *Frontiers in Virtual Reality* 2021, 2.
2. F. Zotter, M. Frank, Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality, Springer International Publishing, 2019.
3. J. Paterson, H. Lee, 3D Audio, Routledge, Abingdon Oxon, 2021.

Methods for assessing the passive attenuation of hearing protection devices

Tomasz Kopciński[1], Bartłomiej Kruk[1], ...
[1]*Department of Acoustics, Multimedia and Signal Processing, Poitechnika Wrocławska ul. Wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland*

E-mail: tomasz.kopcinski@pwr.edu.pl

This article presents detailed measurements of the passive attenuation characteristics of earplugs used by sound engineers, industrial workers and everyday users, such as public transport passengers or those wishing to reduce annoying ambient noise. In an era of growing awareness of hearing hygiene and the risks of excessive exposure to high sound pressure levels, the appropriate selection of hearing protection measures is of particular importance.

For sound engineers, especially those working in concert and studio environments, the key criterion is not only effective noise reduction and hearing protection but also maintaining the most linear attenuation characteristics possible. Non-linear attenuation could lead to perceptual distortion and make it difficult to assess the proportions of the sound mix. On the other hand, for factory workers and people exposed to prolonged environmental noise, it is often more important to maximise sound reduction in specific frequency bands, e.g. in the low frequencies generated by machines.

As part of the research, the presented earplugs were subjected to measurements of their attenuation characteristics in the frequency domain. Both subjective methods, based on tonal audiometry performed on a group of test subjects, and objective methods using an artificial ear equipped with measuring microphones were used. This allowed for comparable results that reflect both the physical attenuation properties of the tested elements and the actual user perception.

The measurement results made it possible to select earplug models with the flattest possible attenuation characteristics across a wide frequency range, making them particularly useful in professional applications. At the same time, the analyses revealed significant differences in the effectiveness of individual solutions – from models with clearly enhanced attenuation in the high frequency range to those that provide a more even reduction across the entire acoustic band.

In summary, the measurement data obtained can serve as a basis for informed choices regarding hearing protection and audio equipment by people exposed to noise. These results also indicate the need for further research into the development of solutions that will effectively protect hearing while minimising interference with natural sound perception.

The Use of Sound in Tinnitus Management and Psychoacoustic Assessment

Barbara A.Kłos, Andrzej Wicher,
Department of Acoustics, Faculty of Physics and Astronomy, Adam Mickiewicz University in Poznań, Poland

E-mail: barbara.klos@amu.edu.pl

Tinnitus is the perception of sound without an external source, typically caused by abnormal neural activity in the auditory pathway. It affects millions of people worldwide and can significantly reduce their quality of life. Subjective tinnitus is only perceived by the patient and the description of symptoms is the only clue [1]. Diagnosis includes audiologic evaluations to assess hearing and psychoacoustic testing to define characteristics such as frequency, intensity, and maskability by comparison with external acoustic stimuli [2]. In addition, diagnostic procedures often include standardized questionnaires, such as the Tinnitus Handicap Inventory (THI), which assess the impact of tinnitus on daily functioning and emotional well-being [1]. Although there is no universally accepted treatment, the auditory cortex remains an important therapeutic target due to its role in conscious sound perception [3]. Current approaches aim to reduce tinnitus by enhancing external sounds or disrupting tinnitus-related neural activity.

In this project, specialized software was used to diagnose tinnitus in individual patients. The Tinnitus Handicap Inventory (THI) was administered to assess its subjective impact. The first phase of the study aimed to identify correlations between psychoacoustic test results and THI scores. Our results indicate that higher Minimum Masking Level (MML) scores are associated with greater tinnitus-related annoyance, particularly within the emotional subscale of the THI. Integrating subjective responses with objective measures. In the second phase, we will estimate tinnitus pitch by adaptively matching an acoustic stimulus to the patient's perception. Although pitch matching is not currently recommended as a routine clinical diagnostic tool, it remains essential for the development of individualized sound-based therapies. We will then evaluate a personalized notched music intervention that removes a frequency band centered on the patient's tinnitus pitch. This method aims to reduce tinnitus distress by promoting lateral inhibition and reducing hyperactivity in the auditory cortex [4].

This research aims to validate a non-invasive, inexpensive and accessible therapy for tinnitus that could support more personalized and effective treatment strategies. A key component of the study is the use of sound to determine the psychoacoustic characteristics of tinnitus, such as its perceived pitch, which are essential for tailoring sound-based therapies. The results have potential implications for both clinical practice and future research in tinnitus management, particularly in refining therapeutic approaches that utilize auditory stimulation.

References

1. A.R. Møller, B. Langguth, D. De Ridder, T. Kleinjung, *Textbook of Tinnitus*, Springer, New York, 2011.
2. F.A.B. Suzuki, F.A. Suzuki, E.T. Onishi, N.O. Penido, Braz. *J. Otorhinolaryngol*. **2018**, *84*, 583-590.
3. C. Pantev, H. Okamoto, H. Teismann, *Front. Syst. Neurosci*. **2012**, *6*, 50.
4. S.Y. Kim, M.Y. Chang, M. Hong, S.-G. Yoo, D. Oh, M.K. Park*, Auris Nasus Larynx* **2017**, *44*, 528-533.

Questionnaire survey on noise in primary school

Dominika A. Zagórska[1], Ewa B. Skrodzka[2], Natalia Czajka[1], Piotr H. Skarżyński[1,3]
[1] *Teleaudiology and Screening Department, World Hearing Center, Institute of Physiology and Pathology of Hearing, Warsaw/Kajetany, Poland*
[2] *Department of Acoustics, Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*
[3] *Institute of Sensory Organs, Kajetany, Poland*

E-mail: d.zagorska@ifps.org.pl

According to a well-known definition, noise is “any unwanted sound that may be a nuisance or harmful to health or increase the risk of an accident at work” [1]. On the other hand, our attitude to a given sound is also important. Defining a particular sound as noise depends on how much it is preferred by us. For some people, for instance, music from the radio is a desirable sound, but at the same time for someone else it may already become noise. Noise does not have to be associated only with loud sounds - even quiet sounds can be annoying.

Noise pollution is one of the most significant threats to the health and well-being of people around the world. It negatively impacts our health. Being in a noisy environment can affect our auditory and nervous systems, interfere with our ability to concentrate, and disrupt our sleep [2]. We encounter noise every day. The dynamic development of today's world makes noise unavoidable. Both adults and children are affected by being in a noisy environment. Children spend most of their time at school, where it is well known that the environment is very noisy. Noise negatively impacts cognitive function, making it difficult to focus and work. Chronic exposure to noise impacts children's academic performance and achievement in school in a negative way [3].

It is also worth remembering that the definition of noise may have different meanings for children than for adults. So it is important to check whether children perceive a noisy environment in the same way as adults. Are they aware that noise is harmful? Does noise interfere with children's daily activities? To find out the answers to these questions, ask students directly. In addition to objective measurements of the equivalent average sound level A, the subjective view of students on the noise problem in their schools is very important. For this purpose, appropriately constructed, age-matched student questionnaires can be used. The answers to the questions in the survey will allow to check how the youngest define noise, what they associate it with and whether they report difficulties in functioning in a noisy environment.

The study focused on the subjective evaluation of noise and checked the awareness of elementary school students and staff, the dangers of being in a noisy environment. The study was conducted through questionnaires prepared for students in grades 1-8 in elementary schools and employees.

The aim of this study was to investigate the perception of noise by primary school students and staff using surveys. The study focused on children's attitudes toward noise, to see if it bothers them, and to see how they identify the noise phenomenon.

An accessible and understandable questionnaire for school-age children was constructed to examine the awareness of noise perception. In addition, a similar survey was constructed for teachers and school staff. Student surveys were divided into three age groups. The surveys for students were intended to check issues such as: acoustic conditions in schools, subjective assessment of the noise level, checking how students define noise, determining the noisiest place, the problem of attention and communication disorders among students, determining in what acoustic conditions students spend their time after school, checking what noise sources dominate in a given school, checking students' awareness of the negative effects of noise, checking whether students suffer from symptoms such as headache and fatigue after being in a noisy environment. The employee survey addressed similar issues, with the addition of a subjective assessment of exposure to noise from various sources.

A test group of 2001 people was collected. Based on the survey results many dependencies were observed e.g the loudest place turned out to be the corridor for all study groups, in Polish primary schools, according to students' subjective assessment, it is too loud and the acoustic conditions are unfavorable, schools are dominated by internal noise resulting from the activity of people staying inside.

The survey turned out to be a good tool for examining the problem of noise in the school environment. Students

define noise as loud noises, “something loud’ which differs from the definition known in the literature, used by adults. The students’ age influences the formulation of the definition of noise, the choice of the loudest place, the choice of the preferred resting place, and the evaluation of communication disruption. Moreover, school noise interferes with employees' ability to perform their daily duties. Unfavorable acoustic conditions in classrooms cause voice fatigue for teachers, and the evaluation of employees' exposure to noise from students depends on the job position they hold.

References

1. E. Ozimek, *Dźwięk i jego percepcja*. *Aspekty fizyczne i psychoakustyczne.* PWN, Warszawa, 2018
2. M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, S. Stansfeld (2014). Auditory and non-auditory effects of noise on health. The Lancet, 383(9925), 1325-1332.
3. D. Augustyńska, J. Radosz, *Bezpieczeństwo Pracy: nauka i praktyka Hałas w szkołach (2) - wpływ hałasu szkolnego na uczniów i nauczycieli oraz jego profilaktyka* **2009**, (10), 8-10.

Statistical Analysis and Synthesis of Polish Dysarthric Speech

Tomasz Piernicki[1], Bożena Kostek[2]
[1]*Multimedia Systems Department,*
Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gabriela Narutowicza 11/12, Gdańsk, Poland
[2]*Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications and Informatics,*
Gdańsk University of Technology, Gabriela Narutowicza 11/12,Gdańsk, Poland

E-mail: tomasz.piernicki@pg.edu.pl

This paper presents a statistical analysis of Polish dysarthric speech, encompassing a classification of dysarthria subtypes, comparative studies with normative speech datasets, and alignment with existing medical literature [1]. Key acoustic and prosodic features that distinguish dysarthric speech from typical speech are identified and visualized using dimensionality reduction techniques, such as Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) [2, 3]. Statistical tests are used to validate the significance of features across different types of dysarthria.

Building on this analysis, we propose a novel method for automatic dysarthric speech synthesis that preserves the distinguishing characteristics of clinically recognized subtypes [1]. Each subtype is modeled using a tailored set of acoustic parameters to capture its unique speech patterns. As an alternative approach, we also explore the feasibility of leveraging voice cloning to transfer the features of the dysarthric speech to the synthesized normative speech [4][5]. The block diagram of the proposed approach is depicted in Figure 1.

Evaluation of system performance includes objective metrics, such as PESQ-MOS [6], as specified in the ITU-T Recommendation P.563 [7], and a novel dysarthria-specific speech quality measure based on acoustic deviations. A comparative evaluation of these methods highlights the challenges and potential for modeling and synthesizing dysarthric speech with clinical applicability.

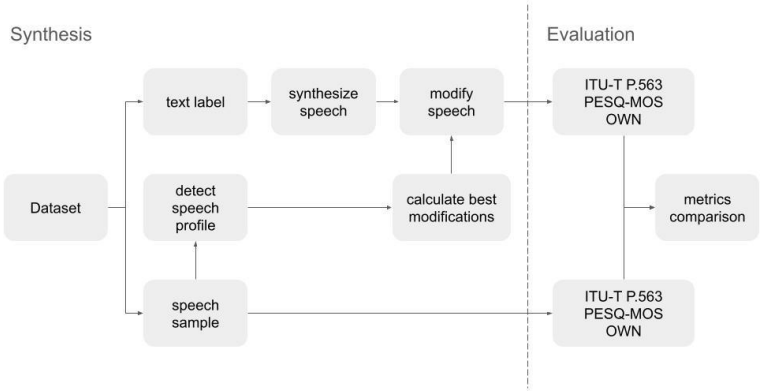

Figure 1: Block diagram of the proposed testing framework

References

1. F.L. Darley, A.E. Aronson, J.R. Brown, *Journal of Speech and Hearing Research* 1969, **12**, 246–269.
2. I.T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, 2002.
3. L. van der Maaten, G. Hinton, *Journal of Machine Learning Research* 2008, **9**, 2579–2605.
4. Coqui, *Coqui TTS: Open-Source Text-to-Speech Toolkit*. https://github.com/coqui-ai/TTS (accessed July 2025).
5. S. Brachmański, M. Kin, P. Kozłowski, Quality assessment of synthetic speech, International Journal of Electronics and Telecommunications, **71**, 3., 1-6, 2025, doi: 10.24425/ijet.2025.153612.
6. A.W. Rix, J.G. Beerends, M.P. Hollier, A.P. Hekstra, *IEEE Transactions on Speech and Audio Processing* 2001, **10**, 745–758.
7. ITU-T, *Recommendation P.563: Single-ended method for objective speech quality assessment in narrow-band telephony applications*, International Telecommunication Union, 2004.

Integrating Acoustic Music Understanding and Large Language Models for Music Captioning

Mateusz Zieleziński[1], Ewa Łukasik[1]
[1]*Faculty of Computing and Telecommunication., Poznan University of Technology, Piotrowo 2, Poznań, Poland*

E-mail: mateusz.zielezinski@hotmail.com

Recent developments in deep learning have substantially advanced the capabilities of artificial intelligence in multimodal content analysis. Architectures such as Convolutional Neural Networks (CNNs), WaveNet, and transformers have demonstrated state-of-the-art performance in tasks involving image recognition, raw audio modeling, and sequence-to-sequence learning. These breakthroughs have enabled robust cross-modal learning, wherein models are trained to associate and generate content across different modalities, such as audio and text.

Within this context, automatic music captioning has emerged as a specialized task that involves generating natural language descriptions of musical pieces. Unlike image captioning, which operates on spatially static and semantically consistent visual data, music captioning must contend with the temporal, abstract, and often subjective nature of musical content. Describing music requires capturing complex attributes such as instrumentation, genre, mood, and tempo—features that are not only temporally distributed but also semantically nuanced.

Furthermore, while general-purpose models trained on Automatic Audio Captioning (AAC) tasks can describe everyday sounds, they exhibit limited transferability to music captioning due to the structural and perceptual complexity of musical compositions. This necessitates the development of domain-specific models capable of extracting high-resolution audio features and generating context-aware textual representations. Music captioning thus represents a challenging intersection of audio signal processing and natural language generation, requiring tailored approaches that go beyond existing captioning paradigms.
This challenges are being addressed by designing, developing, and evaluating novel encoder-decoder models that bridge the gap between complex musical audio and descriptive natural language.

The motivation of this thesis was to explore and develop a new type of solution for the music captioning task. One of the goals was to mitigate the problem of non-availability of musical data samples by employing a new architecture capable of extracting more valuable information from the original data sample, which could be processed by the model during training to correctly map musical features with language tokens. Another goal of the experiments was to check whether a smaller, in the parameter-wise sense, model could generate more meaningful results. This research direction was directly motivated by evidence that current foundation models suffer from computational inefficiencies due to their massive parameter counts, hinting that model downsizing could be not only feasible but potentially beneficial for music understanding tasks.

The paper describes the approach leveraging the state-of-the-art MERT [1] architecture as a powerful audio encoder to generate rich, high-quality musical embeddings. Two distinct architectures are proposed and evaluated: MERT+BART [2] and MERT+T5 [3]. These models utilise a custom adapter, equipped with a cross-attention resampling mechanism, to effectively map the MERT embeddings into the token space of the respective language model decoders. The models were trained on a curated subset of the LP-MusicCaps [4] dataset. Furthermore, a summarisation module using the Mistral-7B large language model was developed to synthesise cohesive captions. A quantitative and subjective evaluation is then performed, highlighting the successes and shortcomings of the presented models, along with comparison of results obtained at different stages of the work.

References

1. Li, Yizhi, et al. "Mert: Acoustic music understanding model with large-scale self-supervised training." arXiv preprint arXiv:2306.00107 (2023).
2. Lewis, Mike, et al. "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." *arXiv preprint arXiv:1910.13461* (2019).
3. Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." *Journal of machine learning research* 21.140 (2020): 1-67.
4. Doh, SeungHeon, et al. "Lp-musiccaps: Llm-based pseudo music captioning." *arXiv preprint arXiv:2307.16372* (2023).

Perception of linear and non-linear reverberation

Łukasz Błasiński[1], Jan Felcyn[1], Jędrzej Kociński[1]
[1]*Department of Acoustics, Faculty of Physics and Astronomy, Adam Mickiewicz University, 61-614 Poznań, Poland*

E-mail: lukasz.blasinski@amu.edu.pl

This study investigated the perception of reverberation under conditions of linear and nonlinear acoustic energy decay. The analysis combined room-acoustic measurements with psychoacoustic listening experiments to examine how subjective impressions of reverberation length relate to objective acoustic parameters. Particular emphasis was placed on identifying both similarities and differences in perception between linear and nonlinear decay. The experimental material was derived from more than 600 impulse responses recorded in 28 concert halls with diverse acoustic properties, which subsequently served as the basis for listening tests involving over 100 participants. These results provide new insights into the subjective evaluation of room acoustics and contribute to the broader understanding of reverberation perception and may inform future research in room acoustics.

Keywords: room acoustics, reverberation, perception of reverberation

Educational computer game supporting the skills with in timbre solfege

Paulina Bielesz,
Institute of Music, University of Silesia, ul. Bielska 62, Cieszyn, Poland

E-mail: paulina.bielesz@us.edu.pl

An educational computer game based on the issues inherent in timbre solfege. An innovative tool was created, a computer environment based on the Unity game engine, working with the FMOD sound engine, which allows shaping the tone of the sound. The game operating with a coherent, composed world of sounds suppose to be also a didactic tool. Supporting the developing of skills in the field of timbre solfege, having all features the computer game should have.

A computer game named Sound Jobs is an educational game that has all the features of a computer game, i.e. graphics, storyline and competition, which are not available in the programmes already existing on the market. While working on the project I tried to facilitate and adapt the exercises in the field of timbre solfege in the virtual world. The aim of the study is a making a prototype this type of a game.

The game includes exercises likes: identifyng equalization, distinguishing sound dynamics, distortion, reverb and delay. Exercises are carried out on the example of pink noise, pieces of music, voice over. The game has two modes. The first is the learning mode and the second is the game mode.

Timbre solfege is so important for sound engineers because they should possess analytical listening skills and they must be sensitive listeners. Hearing training leads to higher quality of sound projects, raises the quality of work in a recording studio, allows for the conscious use of tools for sound edition, such as: sound equalizer, compressor, various types of sound effects. A sound engineer has to work on sound layers properly, so that the listener feels comfortable.

The plot is based on a true story. The time of the game was set in the 70s, when there are no mobile phones. 60 % of the population have landline phones. Phone calls are very expensive, especially a long distance once, they cost a lot of money. Hardly anyone can afford them. There are hackers who can break into the telephone network. They are able to make free phone calls. They can call anyone even the president.

In this game the player is a young talented hacker in the age of technological development who earns money on making illegal phone calls using proprietary equipment. It is possible by a player listening skills. In order to make a phone calls, it is necessary to dial the number after which the sound motif is heard. If the player recognizes the specific timbre the phone call can be made.

Figure 1: Game logo

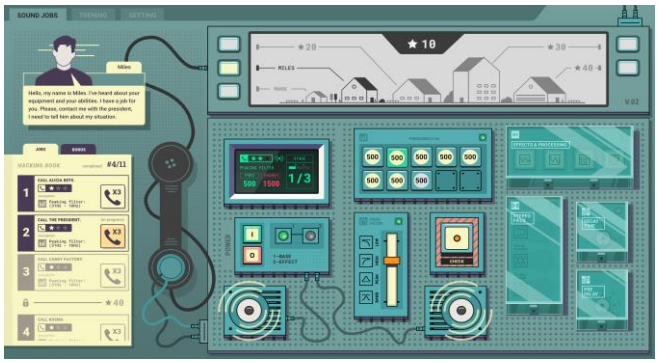

Figure 2: Game screen

Figure 3: Game screen in Unity Engine

References

1. Rogala T., Solfeż Barwy, [w:] Sztuka Słuchania, red. Tomira Rogala, Warszawa 2015
2. Letowski T., Miśkiewicz A., Development of technical listening skills for sound quality assessment. *Proceedings of Inter-Noise'95, Newport Beach*, 917-920, 1995
3. Ryan Henson Creighton, Unity 4.x Game Development by Example Beginner's Guide, 2013
4. Horachek D., Creating E-Learning Games with Unity, Develop your own 3D e-learning game using gamification, systems design, and gameplay programming techniques, 2014, BIRMINGHAM – MUMBAI

Acoustic Analysis Of Selected Homographs For Speech Recognition Systems

Dominik Lentas[1], Michał Łuczyński[2]
[1]*Faculty of Information and Communication Technology, Wrocław University of Science and Technology, wyb. Stanisława Wyspiańskiego 2, 50-370 Wrocław, Poland*
[2]*Department of Acoustics, Multimedia and Signal Processing, Faculty of Electronics, Photonics and Microsystems, Wrocław University of Science and Technology, wyb. Stanisława Wyspiańskiego 2, 50-370 Wrocław, Poland*

E-mail: michal.luczynski@pwr.edu.pl

This paper presents an acoustic analysis of selected homographs in the context of automatic speech recognition (ASR) systems. The study focuses on the Polish words “Dania” (the country) and “dania” (meals), which, despite identical spelling, differ subtly in pronunciation. These differences pose challenges for ASR systems, especially when context is unavailable.

The methodology includes time-frequency analysis, MFCC (Mel-Frequency Cepstral Coefficients) extraction, and classification using a Support Vector Machine (SVM) algorithm. A custom audio database was created using recordings from ten speakers, followed by manual segmentation and normalization of samples. Spectrograms and formant trajectories were analyzed to identify phonetic distinctions, particularly the presence of the semi-vowel [j] in “Dania”.

A subjective listening test involving 27 participants was conducted to assess human recognition accuracy. Results showed an average recognition rate of 58%, indicating significant ambiguity. In contrast, the machine learning model achieved up to 79% accuracy with randomly stratified data and 95% accuracy when tested on the same samples used in the subjective test.

The findings suggest that MFCC-based classification combined with SVM is a promising approach for distinguishing homographs in speech, outperforming human listeners in controlled conditions. Limitations include the small dataset and variability in speaker articulation. The study highlights the importance of phonetic exception handling in ASR systems and proposes extending the method to other homographic pairs.

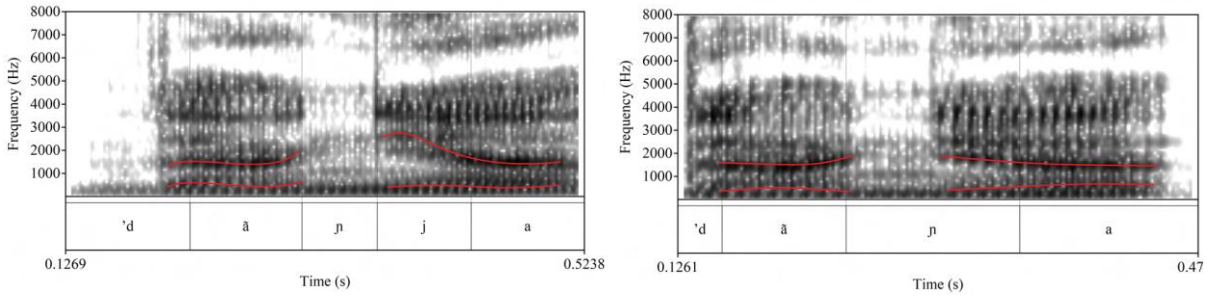

Figure 1: Example spectrogram showing formant trajectories for the word “Dania” (left) and “dania” (right) pronounced by speaker G.

References

1. Abbas, A. W., Ahmad, N., & Ali, H. (2012). Pashto Spoken Digits Database for the Automatic Speech Recognition Research. 18th International Conference on Automation and Computing, 1–5.
2. Crystal, D. (2008). A Dictionary of Linguistics and Phonetics. Wiley.
3. Dhingra, S. D., Nijhawan, G., & Pandit, P. (2013). Isolated Speech Recognition Using MFCC and DTW. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy, 2, 4085–4092.
4. Mohammed, R. A., Ali, A. E., & Hassan, N. F. (2019). Advantages and Disadvantages of Automatic Speaker Recognition Systems. Journal of Al-Qadisiyah for Computer Science and Mathematics, 11(3), 21–30.

From Exhibition Hall to Concert Venue: Improving Sound Insulation and Acoustic Adaptation

Łukasz Błasiński[1], Michał Gałuszka[2], Piotr Pękala[2]
[1]*Department of Acoustics, Faculty of Physics and Astronomy, Adam Mickiewicz University, 61-614 Poznań, Poland*
[2]*Akustix Ltd., Wiosny Ludów 54, 62-081 Przeźmierowo, Poland*

E-mail: m.galuszka@akustix.pl

This paper presents the design and implementation of measures aimed at significantly improving the acoustic insulation of a partition between two exhibition halls that also serve as concert venues. This was a particularly challenging task due to the buildings' large volume, structural constraints and the need to maintain their functional flexibility. Despite these difficulties, the insulation targets were achieved, enabling the two halls to be used simultaneously without risking sound transmission. At the same time, the interiors were comprehensively adapted acoustically to meet the requirements of popular music with sound reinforcement. The achieved acoustic parameters, such as reverberation time, speech intelligibility indices and sound field uniformity, are within the optimal ranges and provide excellent conditions for the planned use of the halls. This presentation will summarise the design challenges, the technical solutions employed, and the measurement results obtained.

Keywords: room acoustics, sound insulation, acoustic adaptation, multipurpose halls

Analysis of the Influence of Pop Filters on the Frequency Response of Condenser Microphones

Bartłomiej Kruk[1], Tomasz Kopciński[1]
[1]*Wrocław University of Science and Technology*

E-mail: bartlomiej.kruk@pwr.edu.pl

Modern sound recording techniques, both in voice-over and music studios, rely heavily on condenser microphones. They are valued for their high sensitivity, wide frequency response, and natural reproduction of vocal timbre. For this reason, they constitute an essential tool for voice-over artists, vocalists, and sound engineers. However, their high sensitivity also brings certain limitations—particularly in voice recordings. These microphones are especially susceptible to so-called plosive sounds, which are produced by a sudden burst of air during the articulation of consonants such as “p,” “b,” “d,” or “t.” Excessive air impact on the microphone diaphragm can lead to nonlinear distortions that degrade the quality of the recorded material.

A commonly used solution to this problem is the pop filter, designed to disperse the energy of the airflow reaching the microphone while exerting minimal influence on the sound’s timbre. Although they may appear to be a simple accessory, their construction, shape, and the material used can significantly alter the frequency response of the recorded signal. In practice, this means that a pop filter, while protecting the recording from plosive distortions, simultaneously affects the spectrum of the recorded voice by modifying certain frequency bands.

The purpose of this article is to investigate the extent to which different types of pop filters influence the frequency response of a microphone. The results of measurements performed on several popular filter models, representing the most commonly used designs, are presented. This analysis provides a deeper understanding of how the choice of filter impacts the final sonic outcome and what trade-offs must be considered by sound engineers when working in a studio environment.

Transforming Office Spaces into Piano Practice Rooms: Acoustic Insulation and Adaptation Challenges

Łukasz Błasiński[1], Michał Gałuszka[2], Piotr Pękala[2]
[1]*Department of Acoustics, Faculty of Physics and Astronomy, Adam Mickiewicz University, 61-614 Poznań, Poland*
[2]*Akustix Ltd., Wiosny Ludów 54, 62-081 Przeźmierowo, Poland*

E-mail: m.galuszka@akustix.pl

The presentation describes the process of transforming standard office spaces into practice rooms designed for acoustic pianos. As part of the project, four such practice rooms were created, each meeting the demanding requirements of both airborne and structure-borne sound insulation between adjacent spaces. In addition, the internal acoustic adaptation of each room was carried out to achieve parameters appropriate for piano practice, including balanced reverberation time and sound clarity. The task posed significant technical challenges due to the limitations of the original office structure, yet the final results fully satisfied the acoustic goals. The paper will present the step-by-step transformation process, the materials and construction methods applied, as well as measurement results confirming the achieved performance. The experience gained may serve as a practical reference for similar future projects involving the conversion of non-purpose-built spaces into high-quality music practice environments.

Keywords: room acoustics, sound insulation, practice rooms, acoustic adaptation

Evaluation of Acoustic Conditions during the Recording of Logatome and Sentence Lists

Bartłomiej Kruk[1], Przemysław Plaskota[1], Tomasz Kopciński[1]
[1]*Wrocław University of Science and Technology*

E-mail: bartlomiej.kruk@pwr.edu.pl

Sentence and logatome lists are used in studies of communication channel quality. Both bidirectional channels, such as telephony or internet communicators, and unidirectional channels, such as broadcast radio or internet streaming services, can be examined. Due to the specific features of national spoken languages, research on a given communication channel should be conducted in as many languages as possible. The lists can also be used to study speech intelligibility in public address systems, speech recognition systems, and audio recording authentication systems.

Databases of recorded sentence and logatome lists for the Polish language do exist, but they were created in the 1970s. For this reason, they do not meet modern quality requirements in terms of signal parameters and the acoustic characteristics of the recording environments. The available resources are affected by noise, reverberation, and low signal dynamics. These features hinder their effective use. Therefore, a new database of logatome and sentence list recordings is planned, using contemporary technologies that ensure high recording quality.

The article presents an analysis of the impact of the type of microphone stand and its setup on the quality of recorded audio. Three microphones were used during the recordings, so both the mounting method and the type of stands had to be optimized. The study included measurements of microphone frequency responses in various spatial configurations. Particular attention was given to diffraction and interference of acoustic waves on microphone mounting elements. These phenomena can significantly affect the quality of the recorded signal. The research demonstrated differences in the properties of recorded signals depending on the mounting method of the microphone and its position relative to the sound source.

VR Interface versus WEB Application in Sound-Localization Tests - Localization Errors and User Experience

Grzegorz Rusinek[1], Agnieszka Paula Pietrzak[1]
[1]*Institute of Radioelectronics and Multimedia Technology, Warsaw University of Technology, ul. Nowowiejska 15/19, Warsaw, Poland*

E-mail: agnieszka.pietrzak@pw.edu.pl

Accurate and efficient assessment of spatial hearing abilities increasingly depends on immersive and accessible test platforms. In this study a head-mounted virtual-reality (VR) interface was compared with a conventional browser-based (WEB) application in a sound-localisation task that presented broadband noise bursts binaurally, with sound sources arranged along horizontal and vertical plane. Thirty-three normal-hearing adults completed both conditions and subsequently rated each interface on intuitiveness, comfort, perceived accuracy and interaction speed. Mean absolute error (MAE) served as the objective metric. In the horizontal plane VR yielded a significantly lower MAE than WEB. The vertical plane showed the opposite tendency, although the difference did not reach significance. Thus, VR consistently improved performance where precise azimuth-related pointing dominated, whereas elevation accuracy may have been lowered by the use of a non-personalized HRTF. Subjective data converged on a strong user preference for VR. Ratings were higher for VR on all four criteria and qualitative comments emphasised “natural controller aiming” for VR interface and “cursor precision problems” in the WEB version. The study demonstrates that VR interface provides a viable, user-friendly alternative to browser formats for psychoacoustic evaluation, offering benefits in intuitive control and horizontal spatial fidelity.

Directivity Vector Fields in Ambisonics training of Acoustic Volumetric Rendering model

Jakub Wasilewski[1], Piotr Cenda[1], Maria Peńsko[1], Tomasz Woźniak[1], Łukasz Januskziewicz[1]
[1]*SoftServe Poland Sp. z o.o., Jaworska 11-13, 53-612 Wrocław, Poland*
jwasi@softserveinc.com, pcend@softserveinc.com

Recently there is a growing effort in the development of advanced Machine Learning (ML) models for accurate representation of sound field in enclosed acoustic spaces and acoustic digital twins. Some of the approaches, like Neural Acoustic Fields (NAF) [1] or Acoustic Volumetric Rendering (AVR) [2] have been proven to be very effective in reproduction of non-spatial Room Impulse Responses (RIRs) [3]. However, those methods were not developed with a view to reproduction of Spatial RIRs (SRIRs) or Ambisonics RIRs and therefore require additional feature extraction for training the models capable of inferring directional information.

In this paper, we show a novel approach to training ML models for accurate Ambisonics RIRs reproduction based on Directivity Vector Fields (DVF), and we compare this method to similar ones such as Intensity Vector (IV) approach [4]. Our method comprises of feeding the training pipeline with energy intensity distribution for individual transmitter-receiver pairs allowing for massive parallelization of computation and validation based on DVF divergence maps for the whole enclosed acoustic space. Utilization of DVF divergence maps allows for acoustic field representation that focuses on representing directivity field distribution of real and virtual field sources.

We present a comparison of proposed methods with results of non-directional aware training methods presented in our previous work [3]. The results achieved in this work are evaluated based on the objective metrics rather than subjective comparison, which will be the subject of further work.

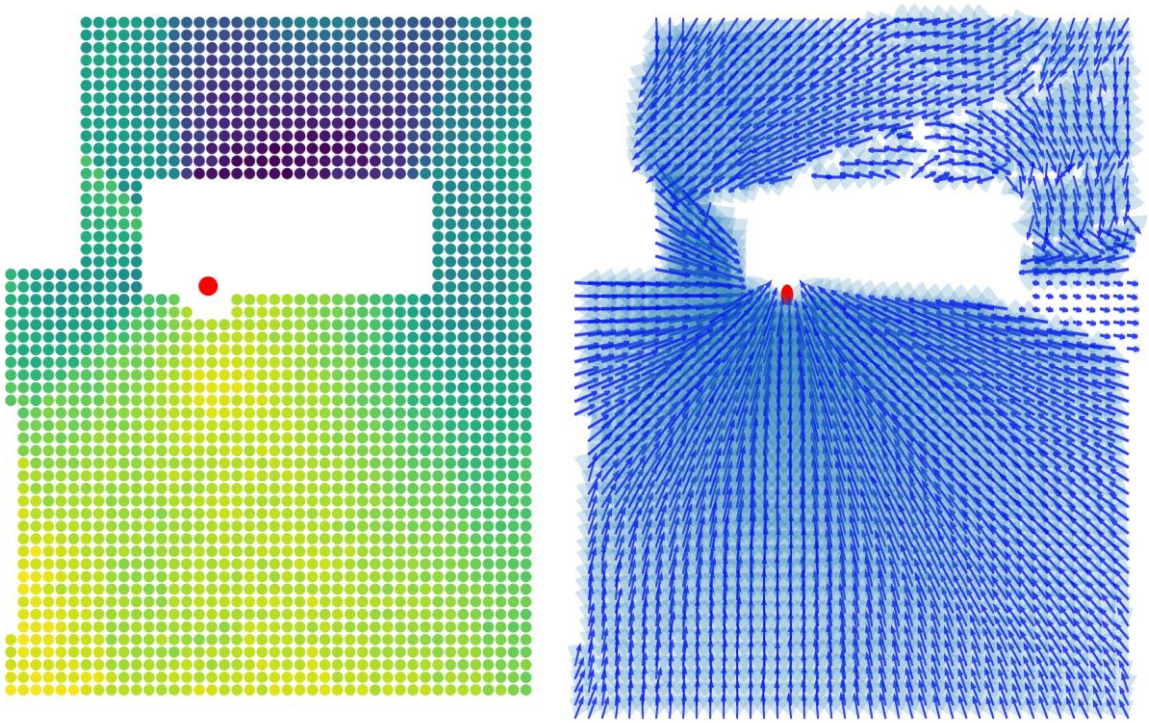

Figure 1: Loudness map (left) and directivity field map (right). Source at red dot.

References

1. Luo, A., et al. Learning neural acoustic fields. Advances in Neural Information Processing Systems, 35, pp. 3165-3177, 2022.
2. Lan, Z., Zheng, C., Zheng, Z., Zhao, M. *Acoustic Volume Rendering for Neural Impulse Response Fields*, arXiv preprint arXiv:2411.06307, 2024.
3. Januszkiewisz Ł., Cenda P., *Neural 3D Audio Renderer for acoustic digital twin creation*, Presented at the 158th Convention, 2025 May 22-24, Warsaw, Poland
4. Ick, C., Wichern, G., Masuyama, Y., Germain, F., Roux, J.L., *Direction-Aware Neural Acoustic Fields for Few-Shot Interpolation of Ambisonic Impulse Responses*, arXiv preprint arXiv:2505.13617, 2025.

Benchmarking Widely Adopted Immersive Audio Formats Through HOA-Based Choral Production

Bartłomiej Mróz[1]
[1]*Department of Multimedia Systems, Gdańsk University of Technology, Gabriela Narutowicza 11/12, Gdańsk, Poland*

E-mail: bartlomiej.mroz@pg.edu.pl

This study investigates the immersive capture, rendering, and evaluation of choral music using higher-order ambisonics (HOA) and various commercial 3D audio formats. A newly commissioned piece – Deus Ex Machina by Jakub Neske, performed by the award-winning Academic Choir of Gdańsk University of Technology – was recorded in an experimental 3D spatial setup, with choir sections arranged across both horizontal and vertical axes around the listener. The primary soundfield was recorded with a third-order ambisonic microphone, enhanced by sectional spot microphones, and mixed in seventh-order ambisonics for accurate spatial detail.

The HOA master served as the reference for format conversion into Dolby Atmos, Auro-3D, Sony 360 Reality Audio, and the emerging IAMF/Eclipsa Audio format. A custom 42-channel decoding array, based on the geometry of the pentakis icosidodecahedron, enabled consistent downmixing into channel- and object-based formats. Experimental renderings highlight each format's capacity – or limitations – in reproducing elevation and full-sphere envelopment.

Because of different loudspeaker setups and hardware needs across these formats – like Auro-3D with its overhead "Voice of God" channel and Sony's use of bottom-layer speakers — directly comparing loudspeakers was not practical. Instead, all tests used each system's official binaural output. To provide a reliable reference, the original seventh-order ambisonics mix was binauralized with the magnitude least-squares (magLS) decoding algorithm, which is widely recognized as the standard for perceptually accurate ambisonic-to-binaural rendering. The perceptual quality of each binaural version was assessed using the Combined Audio Quality Model developed at the University of Oldenburg, which combines objective acoustic metrics with psychoacoustic models to predict how listeners perceive audio fidelity.

To support both public engagement and reproducibility, all versions of the recording—including the HOA master, rendered formats, and binaural versions—are freely accessible online through streaming services and open repositories. The project provides a new framework for critically evaluating immersive audio systems with vertically-rich, acoustically complex content such as choral music.

References

1. E. Pfanzagl-Cardone, The Art and Science of 3D Audio Recording, 1st ed. 2023. ed., Cham: Imprint: Springer, 2023.
2. J.-H. Flesner, T. Biberger and S. D. Ewert, "Subjective and Objective Assessment of Monaural and Binaural Aspects of Audio Quality," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, p. 1112–1125, July 2019.
3. P. Danowski, Decoding Ambisonics to Dolby Atmos using beamforming and spherical array of objects., 2023.

Immersive recording of vocal ensemble on location – practical considerations

Kaja Kosmenda[1], Witold Mickiewicz[2]
[1, 2] *Department of Systems, Signals and Electronics Engineering, Faculty of Electrical Engineering, West Pomeranian University of Technology in Szczecin, al. Piastów 17, 70-310 Szczecin, Poland*

E-mail: kaja_kosmenda@zut.edu.pl

Throughout the recent decades, audio reproduction has advanced significantly, driven by changes and inovations in both recording techniques and playback systems. [1,2] The world has progressed from monophonic and stereophonic formats to ground-breaking surround sound and now culminated in immersive audio systems, which currently stand as the most advanced form of spatial sound reproduction. Following the trends and with a grand passion for music the authors have managed to create an immersive audio laboratory for 7.1.4 listening system within the faculty. However after the inital set up the first critical issue was identified. Following installation, in order to fully utilize the new space most important task was the acquisition of a multichannel audio material that could have been played back in our new laboratory. This quickly proved to be a challange for effective use of the new facility.

There is a pronounced lack of publicly available databases containing genuine multichannel audio samples. This limitation affects not only immersive audio systems, but even surround configurations, exceeding five channels, remain difficult to access. While several streaming platforms provide immersive audio in cooperation with formats such as Dolby Atmos or Sony 360 Reality Audio, these sources are not well suited for research purposes. [3] Testing revealed that many of the additional channels in such material are duplicated rather than independently mixed, and the overall sound quality was below expectations. Moreover, because the underlying encoding technologies are proprietary, they cannot be fully validated and are therefore unsuitable for in-depth investigation of immersive audio. The primary objective of our future work is to develop authentic immersive audio tailored for headphone playback; however, this requires first establishing a deeper understanding of system behavior in real-world loudspeaker-based environments.

In the search for immersive audio samples and a deeper understanding of multichannel production techniques, we identified the work of Lee, H. [4-6] on 3D microphone arrays as a valuable reference. After carefully researching his concepts, we designed a custom microphone array tailored specifically for our 7.1.4 playback system. As our team frequently collaborates with the university's ensembles, including bands and choirs, we had the opportunity to deploy this array during a recording session in a church. The results proved highly satisfactory, yielding authentic multichannel material suitable for further research in the laboratory.

In this work, we aim to present not only the concept behind the design of a 3D microphone array, but also the practical and technical challenges associated with system setup, including the integration of multiple microphones, cabling, and audio equipment. Given the growing popularity of immersive audio and the increasing accessibility of immersive laboratory environments, our objective is to make this method more widely known as a reliable and adaptable solution for generating multichannel content. At present, a stereo mixdown of the recorded performance is publicly available [7,8], while future work will focus on adapting the recordings for immersive binaural playback on streaming platforms. A key advantage of using multiple microphones is the ability to directly compare and evaluate the perceptual differences between stereo, surround, and immersive reproduction.

The original arrays proposed by Lee comprised of a wide selection of microphones, combining both omnidirectional and cardioid types to capture spatial detail. In our implementation, we opted for a more streamlined design. Three omnidirectional microphones were positioned at the front of the array, corresponding to the three primary loudspeakers in a 7.1.4 configuration (Left, Right, and Center). The remaining microphones were cardioids, with the upper level pointed toward the ceiling to capture reflections and enhance the perception of added height level. [Fig 1]

Several approaches were explored for processing the recorded material. In the first method, each microphone signal was directly assigned to its corresponding loudspeaker channel. This approach, effectively treating each microphone as a dedicated bed for playback, provided a faithful reproduction of the recorded space. In the second method, the signals were treated as objects. Using the integrated Dolby Atmos renderer within Pro Tools [9], we experimented with repositioning the sources in the virtual sound field. This resulted in a dynamic allocation of signals across multiple loudspeaker channels, producing a noticeably different spatial impression compared to the first approach.

The results demonstrate that a custom-designed 3D microphone array can provide high-quality, authentic multichannel recordings well suited for immersive audio research. The approach proved not only effective in practice, but also adaptable to different playback configurations, offering a reliable method for generating material that supports both experimental studies and future applications such as binaural rendering for streaming platforms, which was exactly what the authour needed to achieve.

Figure 1: 3D microphone array during recording

References

1. Roginska, A.; Geluso, P. (Eds.) Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio, 1st ed.; Routledge: London, UK, 2018.
2. Pfanzagl-Cardone, E. *The Art and Science of 3D Audio Recording*; Springer: Cham, Switzerland, 2023
3. Kosmenda K, Mickiewicz W. Immersive Two-Channel Recordings Based on Personalized BRIRs and Their Applications in Industry. *Applied Sciences*. 2024; 14(24):11724.
4. Lee H (2021) Multichannel 3D microphone arrays: a review. J Audio Eng Soc 69(1/2):5–26.
5. Lee H, Johnson D (2021) 3D microphone array recording comparison (3D-MARCo): objective measurements.
6. Lee H, Gribben C (2014) Effect of vertical microphone layer spacing for a 3D microphone array. J Audio Eng Soc 62(12):870–884
7. https://sonuscantus.azurewebsites.net/
8. https://push.fm/fl/uc5u5a9q
9. https://www.avid.com/resource-center/dolby-atmos-renderer

Leveraging Vowel Characteristics for Multi-channel Signal Decorrelation and Reverberation

Michele Pizzi[1], Bartłomiej Mróz[2]
[1]*Independent Researcher, Italy*
[2]*Department of Multimedia Systems, Gdańsk University of Technology, Gabriela Narutowicza 11/12 Street, Gdańsk, Poland*

E-mail: info@pizzimusic.com

This paper proposes a novel audio decorrelation approach that leverages the acoustic features of vowels by combining velvet noise-based decorrelation methods with parametric modeling techniques for estimating vowel filters from recorded speech sounds. This method enables capturing the timbral qualities of the voice and applies them within audio decorrelation and artificial reverberation tools. The result is a new audio processing technique inspired by the human voice, offering audio professionals an expanded palette of sonic possibilities.

The proposed implementation utilizes velvet noise as the input for digital filters that model vowel timbre. This allows users to capture the resonant structures of vowels from recorded speech or singing, enabling them to shape the frequency content of a multi-channel effect potentially using their own voice as a source. Rather than seeking to replace or enhance existing velvet noise-based decorrelation or artificial reverberation techniques, the primary aim is to create a novel immersive audio effect inspired by choir singing and the acoustics of the vocal tract. Each audio channel is processed by filtering two independent velvet noise sequences to synthesize two distinct vowels, resulting in a speech-like signal reminiscent of vocal fry. This synthesized signal is then convolved with the original audio, yielding a multi-channel output. The integration of velvet noise and speech synthesis offers extensive control over processing parameters, fostering creative exploration and sonic experimentation.

This paper presents the development of a novel audio effect that integrates velvet noise decorrelation methods with speech synthesis techniques. The key stages and challenges encountered during the design process are described, and examples of the tool's creative applications are provided. The paper presents the results of the audio analysis and listening tests, while also suggesting potential directions for future research.

Enhancing Polish Medical Speech Recognition: Applying Knowledge Distillation to Whisper ASR Models

Szymon Zaporowski[1,2], Bożena Kostek[2]
[1]*Multimedia Systems Department, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233, Gdańsk, Poland*
[2]*Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233, Gdańsk, Poland*

E-mail: szyzapor@pg.edu.pl

Automatic Speech Recognition (ASR) systems have transformed various fields, including healthcare, by enabling efficient transcription of medical conversations. Although models like OpenAI's Whisper [1] are quite versatile, they often struggle with specialized, less common languages like Polish, especially in medical settings where precision is crucial. This paper focuses on knowledge distillation techniques to adapt Whisper models for Polish medical speech data. A simplified block diagram of this training method is shown in Figure 1.

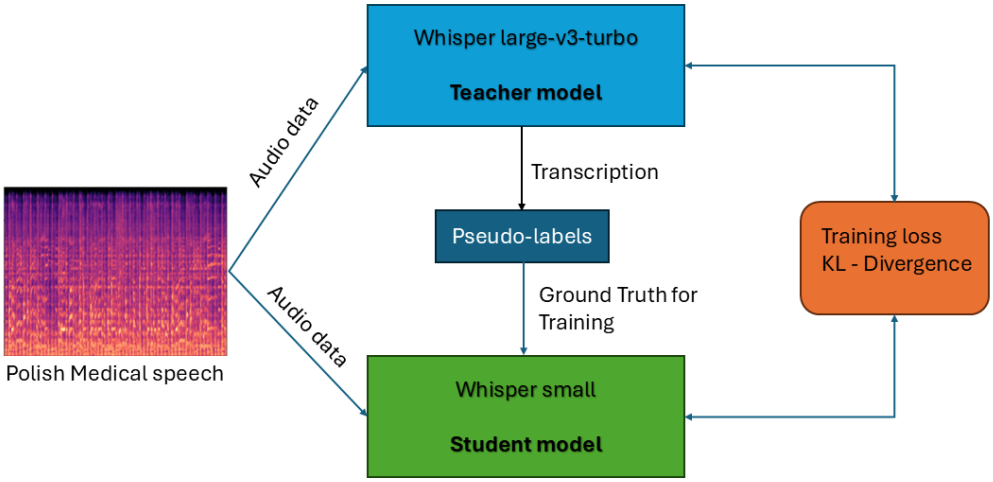

Figure 1: Block diagram of the training method for knowledge distillation using models from the Whisper family.

The vast majority of approaches to the problem of medical speech processing in Polish point to the lack of sufficient training data and focus on attempts at classical fine-tuning or fine-tuning using Low-Rank Adaptation (LORA) [3-5] .The approach presented in the article addresses the issue of insufficient data by employing knowledge distillation based on one of the largest available ASR models from the Whisper family and creating automatic transcription that serves as training labels for a smaller model from the same family. By transferring knowledge from a large, teacher Whisper model to a smaller, student version, we achieve better efficiency and performance on domain-specific datasets comprising Polish medical consultations and terminology. The paper presents various approaches to distillation, both using attempts at fine-tuning and utilising different techniques derived from LORA [6], such as AdaLORA [7], DORA [8], MoELoRA [9].

Experimental results show reductions in Word Error Rate (WER) and increased robustness in noisy environments, making practical deployment in Polish healthcare settings feasible.

The approach outlined in this article for training ASR models can be applied not only in the medical field but also in other domains where vocabulary is poorly transcribed by standard models, making the solution more universal.

References

1. A. Radford, J.W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," arXiv preprint, arXiv:2212.04356, **2022**. Available online: https://arxiv.org/abs/2212.04356
2. G. Hinton, O. Vinyals, J. Dean, "Distilling the Knowledge in a Neural Network," arXiv preprint, arXiv:1503.02531, **2015**. Available online: https://arxiv.org/abs/1503.02531
3. M. Zielonka, W. Krasiński, J. Nowak, P. Rośleń, J. Stopiński, M. Żak, F. Górski, A. Czyżewski, A survey of automatic speech recognition deep models performance for Polish medical terms, **2023** Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), pp. 19–24. https://doi.org/10.23919/SPA59660.2023.10274442
4. K.Pondel-Sycz, et. al, A comparative study of deep End-to-End Automatic Speech Recognition models for doctor-patient conversations in Polish in a real-life acoustic environment, **2025** International Journal of Electronics and Telecommunications 71, no. 3: [online]. https://ijet.ise.pw.edu.pl/index.php/ijet/article/view/10.24425-ijet.2025.153609
5. T. Xu, K. Huang, P. Guo, Y. Zhou, L. Huang, H. Xue, L. Xie, "Towards Rehearsal-Free Multilingual ASR: A LoRA-based Case Study on Whisper," in: Proceedings of Interspeech **2024**, pp. 2534–2538. https://doi.org/10.21437/Interspeech.2024-1953
6. E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," arXiv preprint, arXiv:2106.09685, **2021**. Available online: https://arxiv.org/abs/2106.09685
7. Q. Zhang, M. Chen, A. Bukharin, N. Karampatziakis, P. He, Y. Cheng, W. Chen, T. Zhao, "AdaLoRA: Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning," arXiv preprint, arXiv:2303.10512, **2023**. Available online: https://arxiv.org/abs/2303.10512
8. S.-Y. Liu, C.-Y. Wang, H. Yin, P. Molchanov, Y.-C. F. Wang, K.-T. Cheng, M.-H. Chen, "DoRA: Weight-Decomposed Low-Rank Adaptation," arXiv preprint, arXiv:2402.09353, **2024**. Available online: https://arxiv.org/abs/2402.09353
9. T. Luo, J. Lei, F. Lei, W. Liu, S. He, J. Zhao, K. Liu, "MoELoRA: Contrastive Learning Guided Mixture of Experts on Parameter-Efficient Fine-Tuning for Large Language Models," arXiv preprint, arXiv:2402.12851, **2024**. Available online: https://arxiv.org/abs/2402.12851