# SPATIAL AUDIO PARAMETERS AND LISTENER IMMERSION: AN EMPIRICAL STUDY OF SOUND LOCALIZATION, IMPULSE RESPONSE, AND EQUALIZATION EFFECTS

**Anish Nair**[1] **Luhee Hyunkyung Shin**[1]
**Zehao Jiang**[1] **Yueh-Kao Wu**[1]
[1] School of Music, Georgia Institute of Technology

### ABSTRACT

This research delves into the profound influence of spatial sound quality on auditory perception and immersion, addressing the dearth of empirical understanding in this domain. Spatial audio effects are pivotal in modern audio technology and media content, yet the concept of "immersion," commonly associated with spatial audio, lacks a robust empirical foundation. Our primary objective is to establish a concrete connection between "immersion" and "presence," as defined by [1], by systematically manipulating core spatial audio parameters. In this experiment, we meticulously manipulate three key spatial audio parameters: Localization, Impulse Response, and Equalization. These parameters significantly impact auditory experiences by reproducing audio signals in specific spaces and adjusting responses to environmental factors. Our study seeks to identify which parameters exert the most influence on enhancing auditory experiences and how they relate to immersion. Through perceptual assessments and physiological measurements, we aim to gain deeper insights into user perception and cognitive responses. Anticipated outcomes promise to enhance our understanding of audio content perception and inform improvements in audio technology for multimedia and audio production. This research advances the empirical understanding of the interplay between spatial audio parameters, human perception, and immersion, potentially enhancing audio experiences across various media formats.

## 1. INTRODUCTION

The recent emergence of virtual reality (VR) products has led to rapid advancements in spatial audio rendering technology, providing consumers with new audio experiences primarily evaluated under the concept of "immersion." Immersion is considered a core feature of spatial audio, which is grounded in the human auditory system's ability to perceive the spatial characteristics of sound waves, making it a crucial element in creating immersive audio experiences. This research is motivated by the increasing interest in understanding the impact of spatial audio quality on the auditory perception and cognitive experiences of individuals. Specifically, it focuses on how personal perceptions and cognitions are shaped and aims to enhance the scientific understanding of the true effects of essential spatial audio effects, which are integral to modern audio technology and media content. The objective of this study is to bridge the gap between the theoretical concept of "immersion" and the practical application of spatial audio technology.

### 1.1 Theoretical Background

In the context of our research on spatial audio and auditory perception, several key studies provide foundational insights. Firstly, the concept of cognitive "immersion" in the realm of spatial audio is rooted in the framework by Haywood and Cairns (2005). Haywood and Cairns decompose immersion into three aspects: temporal perception absence, loss of perception of the real world, and engagement with task environments. This framework aligns well with our experimental design and serves as a fundamental concept in discussing immersion in this study, closely related to the concepts of "flow," "cognitive absorption," and "presence" [2]. Consequently, it has been adopted as a central element in the discussion of immersion in this experiment and forms the basis of the experimental design. Additionally, our experiment employs the concept of "presence" defined by Cummings et al. (2015) to provide various environments for audio listening experiences. Cummings' theory suggests that spatial audio parameters, such as localization, impulse response, and equalization, play a significant role in enhancing user immersion in audio content. To understand the specific impact of spatial audio parameters on immersion, we integrate an experimental approach by manipulating these spatial audio parameters, including changes in sound localization, impulse response, and equalization, to investigate their effects on participants' perceptual responses.

### 1.2 Literature Review

Recent studies have explored the impact of spatial audio on listener experiences and emphasized the role of audio technology. Research in the field of listener perception and immersive responses to spatial audio has shown that audio enhancement in VR applications corresponds to users' immersive experiences [3]. This confirms the potential for audio interaction to enrich user experiences in environments like VR, where user experience is emphasized. Spatial audio's role is most prominently highlighted in the gaming industry, where it enhances user immersion by creating a realistic auditory environment during gameplay [4]. Furthermore, spatial audio contributes to improving experiences in virtual meetings conducted through VR and complements visual depth in scenarios like panoramic live broadcasts [5]. It also holds potential for information delivery in environments with limited visual stimuli or for individuals with visual impairments who rely heavily on auditory stimuli [6]. However, research on users' cognitive responses to spatial audio is relatively scarce [7], highlighting the need for fundamental research to understand the relationship between spatial audio's immersiveness and users' cognition.

### 1.3 Hypothesis

Spatial audio plays a crucial role in providing users with an immersive auditory experience and contributes to various fields such as gaming, virtual reality, and film. This research is based on the hypothesis that spatial audio offers a more immersive experience compared to standard stereo audio and investigates the cognitive foundations of the impact of spatial audio. This hypothesis is grounded in the assumption that by manipulating spatial audio parameters, including sound location, impulse response, and equalization, participants' auditory perceptual experiences and cognitive responses related to immersion can be enhanced [8]. To confirm this hypothesis, the study will measure participants' cognitive and emotional responses and quantitatively evaluate the effects of spatial audio. This approach will contribute to a better understanding of the practical potential of spatial audio technology for enhancing user experiences.

## 2. METHOD

We divide the features of audio into the three features: localization, impulse responses, and equalizations. We expect that those three features may impact the effect of spatial audio differently from the audience perspective, and hence we examine their effects separately by collecting subjective reactions or responses to audio of different spatial features or qualities from the experiment participants.

### 2.1 Audio Parameters Setting and Pre-processing

All of our testing audio wave files are mixed under the standard of Dolby Atmos provided in Logic Pro. We will neither mix the audio with Dolby 5.1 or Dolby 7.1 because the experiment would be conducted with headphones. To map the spatial audio to the binaural earphones, we will use Apple binaural spatial audio render to map the spatial information into binaural channels. Below are what we would do in DAW of Logic Pro to control the spatial parameters when mixing. Note the for each feature, we provide two

| Song Name | Genre | Excerpt |
|---|---|---|
| The Marsh Marigold's Song | Choir | `0:00-0:20` |
| Whiptails | EDM | `4:00-4:27` |
| Corine | Jazz | `0:34-0:54` |

**Table 1**. List of songs, their genres, and the time of the excerpt chosen

different levels of mixing, and hence all features are tested in a binary manner. For simplicity, one can view the two levels as a true level, where the parameter is present, and a false level, where the parameter feature is absent in the mixing. Figure 1 provides sample manipulation for each parameters.

#### *2.1.1 Localization*

Stereo Localization: Sources of sound are only localized on a single dimension. That is the regular mixing standard for CDs or music streaming services. we are using left/right to adjust this parameter.
Spatial Localization: Sources of sound are localized in two dimensions. We will establish several presets of positions of sources that we think would be the best spatial mixing setting for each song.

#### *2.1.2 Impulse Response*

Dynamic Impulse Response: As our testing songs contain 4 different stems, we want to simulate the listening experience in theaters. we would use reverb and delay effects to approximate the impulse response of the theater: we are using the preset parameters provided in the reverb effect plug-in that simulate the reverb in the theater.
Static Impulse Response: Bypass the reverb and the delay plug-ins.

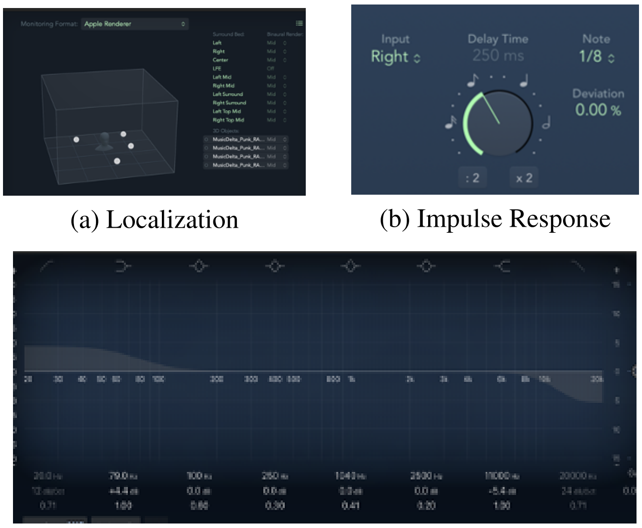
#### *2.1.3 Equalizer*

Static EQ: Maintaining a constant Equalization profile throughout
Dynamic EQ: Varying the equalization according to the distance of localization. For example, if the source is far away, then we will lower the magnitudes on high frequencies. Otherwise, if the source is near, we can boost the magnitudes on high frequencies and cut-off the ones on low frequencies.

### 2.2 Stimuli

In this experiment, we investigate how participant's immersive level varies as they are exposed to audio with different spatial mixings. Participants will listen to 24 roughly 20 seconds music excerpts — 8 versions of 1 excerpt for total 3 excerpts —, which covers a range of songs in different genres and different frequency ranges. This list of songs is as shown in table 1

This selection of songs and excerpts cover both instrumental and vocal parts of the songs with different ranges, styles, and genres. Such selection allows us to examine if




(a) Localization

(b) Impulse Response

(c) Equalizer

**Figure 1**. The demonstration of the exact manipulation of the parameters in this experiment.

the spatial audio quality has a different effect on the listening experiences as the content of the songs varies. The 20-second length minimizes the boredness of repetition of the same excerpt. For each feature we investigate, we mix the excerpt audio in 1 of the 2 levels discussed above. We provide 8 different versions for each excerpt, with each two versions of the same excerpt differ by at least 1 of the 3 features in spatial mixing, and hence 8 versions cover all possible combinations. Each participant will listen to all stimuli in no particular order and complete a questionnaire regarding their immersiveness. Through this within-subject design, we can discover if differences in spatial features provide different listening experiences.

### 2.3 Participants

For this experiment, we have $N = 11$ participants, 7 males and 4 females, age range between 19-23. 9 participants have prior experience with music production or sound engineering. All participants complete the experiment for course credits.

### 2.4 Experimental Conditions

The experiment is conducted in-person and individually for each participant due to the resources and time available for this project. The experiment is fully conducted using each participant's personal laptop. However, since the audio quality is important for this experiment, all participants are required to use the same headphones we provide to listen to the stimuli. For this experiment, we use the headphone, Presonus HD7. We unfortunately do not have the resources to provide speakers instead of headphones, and future investigation can endorse speakers if available. All participants complete their experiment session during daytime in a quiet room so that the surrounding does not heavily disturb the participants. The location and time are, however, not the same for all participants. This is partly due to the resources available for this project, and we believe that time and location have rather only random and limited effects on the results.

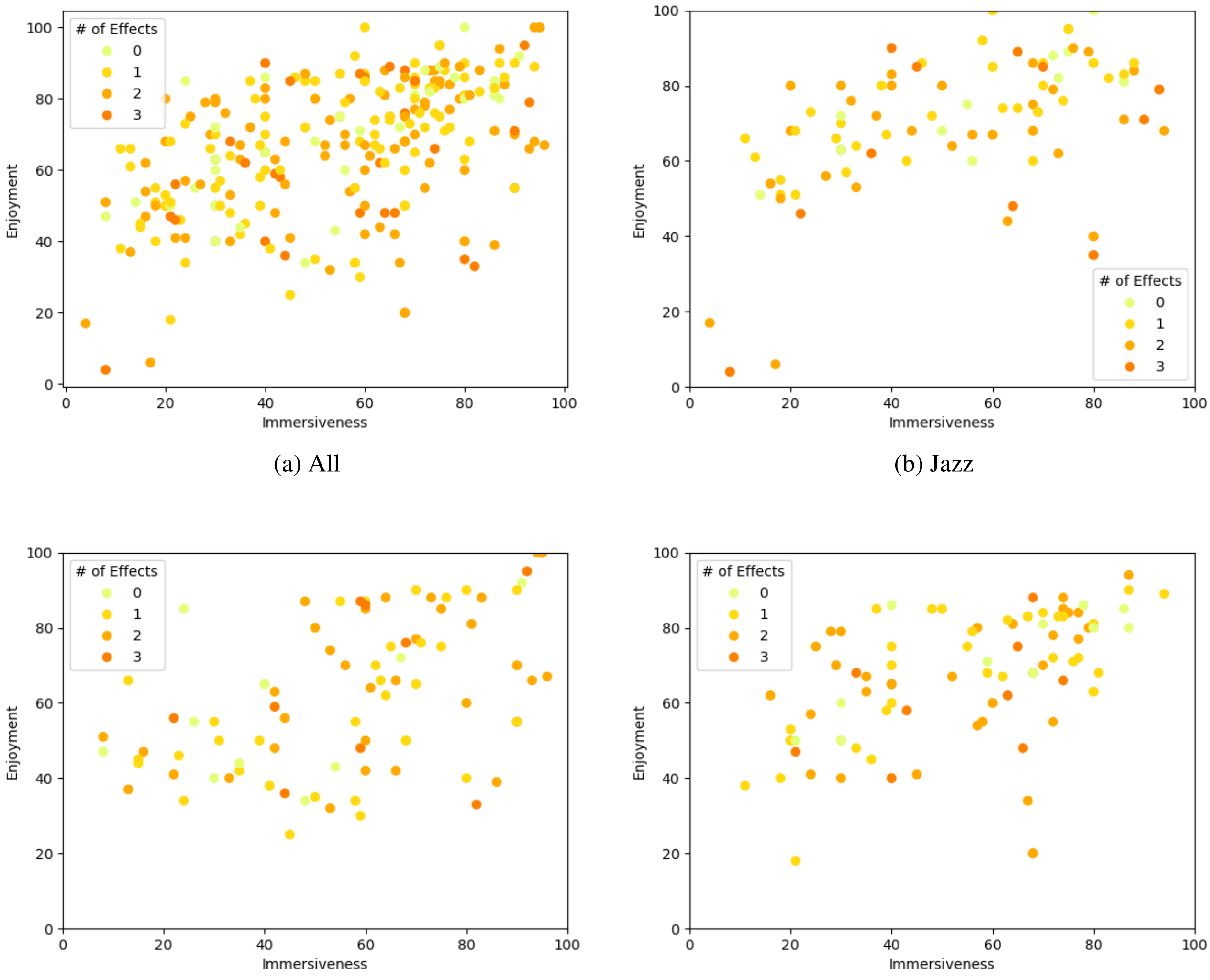
### 2.5 Conducting and Instructions

During the experiment, participants are first shown a pre-questionnaire that collects participants' background information and prior experience with music engineering. This allows us to later investigate if participants' backgrounds have any non-random impact on the results. After that, participants will adjust the volume of the device and prepare to listen to 3 excerpts in the order choir first, EDM second, and Jazz the last. For each excerpt, all 8 different spatially mixed versions will be provided and tested in the same order. For each version, participants will have the play button and the audio timeline at the top of the screen. At the bottom, three questions related to their respective immersiveness are shown on the left, and 3 sliders between 0 and 100 to answer the questions on the right. The question list can be found in the appendix and is the same across all versions and all excerpts. Those questions assess the participants' engagement and awareness of the surroundings while the excerpts are in play, with 100 end correlates with the most immersive side and the 0 end the least. Participants are allowed to adjust the sliders during and after each version until they are satisfied before moving to the next version of audio. This setup allows us to collect responses that directly compare the listening experiences of different versions of excerpts from the same participants, thus minimizing the impact of any personal bias from the participants, since they will have the same bias toward all stimuli.

## 3. RESULT

The results of the experiment are presented in table 2. We have 11 participants to attend our experiment. All the values presented on the table are accumulated by calculating the median value among all the answers for every question collected from 11 participants. In Table 2, there is no prominent trend to tell that the audio processing variants of localization, impulse response, and equalizer are really improving either the immersiveness or the enjoyment. However, if we reorganize the scores by songs with different numbers of audio processing effects applied, which are provided in figure 3, there are obvious trends for EDM excerpts that the immersiveness and enjoyment increase as the numbers of audio processing effects increase, but not for other genres. This suggests that the effectiveness of the audio processing techniques may be genre-specific. Further analysis is needed to confirm this hypothesis and to identify the specific characteristics of EDM that make it more responsive to immersiveness and enjoyments to these techniques.

Next, we analyze the influences on a single audio processing variant. First, we compute the median among all




(a) All (b) Jazz

(c) EDM (d) Choir

**Figure 2**. The scatter plot for quantitative evaluation of immersiveness (X-axis) and enjoyment (y-axis). The color of the dots represent the numbers of the audio processing techniques applied during mixing.

the scores with localization audio processing during mixing. The median score of enjoyment is 67.0 with localization and 73.0 without localization. The median score of immersiveness is 57.0 and 59.5 without localization. It surprisingly turned out to be counter-intuitive and opposite to our expectations. Despite our hypothesis of localization audio processing to enhance the immersive experience, the median scores for both immersiveness and enjoyment were actually lower with localization enabled. This suggests that our implementation of the localization audio processing may be hindering, rather than aiding, the user's perception of the audio. In contrast, the median scores for both immersiveness and enjoyment were higher when using impulse response and equalizer audio processing compared to without. This suggests that these processing techniques may be effective in enhancing the user's perception of the audio. The comparison is provided in table 3.

Therefore, we calculate the p-value for different audio processing variants of the same song. Here, our null hypothesis is that "All three audio processing variants cannot increase the immersiveness and the enjoyment of the music." Only the result from the enjoyment test of EDM song can reject the null hypothesis with the p-value $= 0.039$ but not other genres ( p-values $> 0.05$). This suggests that the limited sample size of our study (only 11 participants) might be insufficient to detect statistically significant differences for the other genres and audio processing variants. While the Choir song's enjoyment test showed a p-value rejecting the null hypothesis, further investigation with a larger and more diverse participant pool is necessary to confirm this finding and draw more robust conclusions about the effectiveness of different audio processing techniques for enhancing immersiveness and enjoyment across various musical genres.

| Local | IR | EQ | Imm. | Enjoy |
|---|---|---|---|---|
| ✓ | ✓ | ✓ | 57.79 | 62.48 |
| ✓ | ✓ | | 50.36 | 62.88 |
| ✓ | | ✓ | **58.23** | 63.91 |
| ✓ | | | 50.24 | 64.52 |
| | ✓ | ✓ | 54.58 | **70.79** |
| | ✓ | | 56.12 | 68.00 |
| | | ✓ | 52.58 | 66.03 |
| | | | 52.52 | 68.73 |

**Table 2**. The table of immersiveness and enjoyment result. If the boxes are checked that means the audio processing effects listed above are applied in the testing audio clips

| Effects | Applied | Imm. | Enjoy |
|---|---|---|---|
| Localization | ✓ | 57 | 67 |
| | X | **59.5** | **73** |
| Impulse Response | ✓ | **60.0** | **71.0** |
| | X | 56.5 | 68.0 |
| Equalizer | ✓ | **60** | 68 |
| | X | 57 | **70** |

**Table 3**. The accumulated results of immersiveness and enjoyment by 3 different audio processing effects. Median scores were calculated for all participants to assess the impact of applying each audio processing effect compared to not applying it.

## 4. CONCLUSION

### 4.1 Summary

Our study explored the intricacies of how spatial audio impacts a listeners immersiveness and enjoyment to a certain song based on the methodical examination of three distinct spatial audio parameters which includes localization, equalization, and impulse response. Our findings present a new nuanced narrative that challenges some preconceived notions all while affirming others. In the realm of localization, which is fundamental to the spatial dimension of audio, our expectations for universally heightened immersion encountered unexpected results. This was evident from the median scores for immersiveness and enjoyment, which were unexpectedly lower with localization enabled, challenging our initial hypothesis. Our data from our study suggests that the application of localization parameters did not consistently amplify listener immersion or enjoyment, suggesting a more complex relationship between spatial audio and perceptual experience than we previously understood. This finding prompts us to further investigate the nuances of spatial localization in audio. In contrast, impulse response and equalization parameters demonstrated a clear positive impact, which affirms their effect in enriching spatial audio experiences. These effects were pronounced in our genre specific responses, with EDM and Choir excerpts showing significant positive linear trends of immersiveness and enjoyment when equalization and impulse response parameters were switched on, pointing out to genre specific spatial audio strategies.

### 4.2 Discussion

In our exploration of spatial audio's impact in a listeners immersion and enjoyment, we gained valuable insight of how spatial audio parameters had a interesting impact on different musical genres. We have learned the specifics of what makes spatial audio spatial, the parameters of spatial audio, and their distinct impacts on the auditory experience. This includes the realization that localization may not completely enhance immersion and enjoyment as proved, and also the positive effects both equalization and impulse response has in shaping the listener's perception and enjoyment. In addition, our group has learned how to conduct live experiments and conduct, gather, and facilitate data that can be used for academic research. While our approach was comprehensive, there remain areas for enhancement identified by our group, and several insightful critiques and observations our professor have provided to refine our interpretation and methodologies which can further support our approach to analyzing results. In terms of methodological considerations, our experimental setup was meticulously designed, featuring 8 different versions (including variations in localization, equalization and impulse response) of 3 different songs.The in-person nature of the experiment, where participants used their laptops and assigned headphones provided by our group, added a commendable level of detail and control to our study setup. Although, since the participants were using their personal laptops, the variability in laptop audio and the personal setting of sound levels by the participants could create inconsistencies in the audio experience. Our professor's insight of the inclusion of 'enjoyment' as a quantitative measure as opposed to 'quality' raises a interesting point. It suggests that immersion and enjoyment might be interconnected in a way that differs from the traditional understanding of audio quality, prompting more a reevaluation of how these elements can be measured in spatial audio research. In terms of our data presentation and analysis, our initial data table and scatterpot did not effectively communicate the trends and our significant findings. A more detailed presentation which includes the medians and distributions across variables would better illustrate these trends. Lastly, our presentation does not clearly define the possibilities or levels of each variable, especially equalization which encompasses multiple variables. The binary representation of variables like equalization and reverb for instance might oversimplify the complex nature of spatial audio. Finding a better 'scale' to represent these complex parameters would better help this experiment. In our study, we must acknowledge the limited sample size which consisted of 11 participants. A small sample size in this study reduces the statistical power, this means that our ability to find a true effect that these parameters will effect the immersiveness and enjoyment of spatio audio will be limited. With a larger sample size, these findings can be more robustly confirmed. Secondly, the effect size which includes the amount of parameters used to determine spatial audio immersiveness and the limited amount of genres used for this study can effect the overall interpretation of our results.

### 4.3 Future Works

For future research in the field of spatial audio and cognition, the exploration of spatial audio immersion should extend beyond our genre constraints of EDM, Jazz, and Choir. There may be genres or types of music which might be better suited for spatial audio since that genre might have a larger frequency bandwidth than other genres. For instance, cinematic music might have a different effect in immersion than Jazz or EDM. Exploring genre diversity is important for understanding how different musical styles with different frequency bandwidths interact with spatial audio parameters. In the topic of spatial audio parameters, impulse response, localization, and equalization are three of the largest variables that defines the characteristics of spatial audio. There might be other confounding variables that might directly affect the immersiveness of spatial audio such as the listener's individual auditory processing abilities, environmental acoustics, or the psychological state of the listener. Finally, introducing a dimension of visual stimuli that correspond to spatial audio will have a effect on immersion as well. Incorporating this variable into this realm of research could provide us with a holistic understanding of immersive experiences.

## 5. REFERENCES

[1] N. Haywood and P. Cairns, "Engagement with an interactive museum exhibit," in *People and computers XIX—The bigger picture: Proceedings of HCI 2005*. Springer London, 2006, pp. 113–129.

[2] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *International journal of human-computer studies*, vol. 66, no. 9, pp. 641–661, 2008.

[3] T. Potter, Z. Cvetković, and E. De Sena, "On the relative importance of visual and spatial audio rendering on vr immersion," *Frontiers in Signal Processing*, vol. 2, p. 904866, 2022.

[4] M. Rieger and D. Zingler, "3d audio: The next step in realism for gaming," *ACM Games*, vol. 1, no. 1, p. Article 12, 2023.

[5] H. Zhang, J. Wang, Z. Li, and J. Li, "Design and implementation of two immersive audio and video communication systems based on virtual reality," *Electronics*, vol. 12, no. 5, p. 1134, 2023.

[6] L. Fialho, J. Oliveira, A. Filipe, and F. Luz, "Soundspace vr: spatial navigation using sound in virtual reality," *Virtual Reality*, vol. 27, pp. 397–405, 2023.

[7] L. Buck, M. Flores Vargas, and R. McDonnell, "2022 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Trinity College Dublin, 2022.

[8] S. Mehrotra, W.-g. Chen, and Z. Zhang, "Interpolation of combined head and room impulse response for audio spatialization," in *2011 IEEE 13th International Workshop on Multimedia Signal Processing*, 2011, pp. 1–6.