



**HAL**  
open science

## Spatial integration of dynamic auditory feedback in electric vehicle interior

Théophile Dupré, Mitsuko Aramaki, Sébastien Denjean, Richard Kronland-Martinet

### ► To cite this version:

Théophile Dupré, Mitsuko Aramaki, Sébastien Denjean, Richard Kronland-Martinet. Spatial integration of dynamic auditory feedback in electric vehicle interior. AES International Conference on Audio for Virtual and Augmented Reality, Aug 2022, Redmond, WA, United States. hal-03844928

**HAL Id: hal-03844928**

**<https://hal.science/hal-03844928>**

Submitted on 5 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spatial integration of virtual sound sources in electric vehicle interior

Théophile Dupré<sup>1,2,\*</sup>, Mitsuko Aramaki<sup>1</sup>, Sébastien Denjean<sup>2</sup>, Richard Kronland-Martinet<sup>1</sup>

<sup>1</sup> Aix-Marseille Univ, CNRS, PRISM, Marseille, France

<sup>2</sup> Automotive Research and Advanced Engineering, Stellantis, Vélizy, France

\* Corresponding author, dupre@prism.cnrs.fr

## Abstract

With the development of electric motor vehicles, the domain of automotive sound design addresses new issues, and is now concerned by creating suitable and pleasant soundscapes inside the vehicle. For instance, the absence of predominant engine sound changes the driver perception of the dynamic of his car. Previous studies proposed relevant sonification strategies to augment the interior sound environment by bringing back vehicle dynamics with synthetic auditory cues. Yet, users report a lack of blending with the existing soundscape. In this study, we analyze acoustical and perceptual spatial characteristics of the car soundscape and show that the spatial attributes of sound sources are fundamental to improve the perceptual coherency of the global environment.

## 1 Introduction

Car industry is facing a major shift from Internal Combustion Engine Vehicles (ICEV) to Battery Electric Vehicles (BEV). Major developments are now focused on BEV. Apart from different dynamic behavior of electric motor, user experience is mainly affected by a different acoustic environment or soundscape due to a different motor generated sound. Even if the environment is quieter, the absence of predominant motor sound may deteriorate the soundscape by unmasking unwanted noises, while no longer informs the driver about the vehicle dynamic nor conveys vehicle characteristics. Driving experience is highly affected and the need for new solutions is expressed by car manufacturer.

For few years, researchers have worked with sonification processes for the so called active sound design, that aimed to bring back the motor dynamic auditory feedback to the driver [1][2]. Integration of this virtual sound source in the interior soundscape augments the perception of the dynamic of the vehicle. We may then consider the cockpit environment as an augmented reality environment. Yet, users report a

lack of blending with the surrounding environment and other perceptual cues. More generally, it raises the question of how we may create a perceptually coherent experience in a multi-modal real environment augmented by virtual sound sources.

In this study, we aimed at investigating the influence of the spatial configuration of sound sources on the perceived coherency of car cabin soundscape in ICEV (Experiment 1) and in BEV (Experiment 2). In particular, we analysed acoustical and perceptual spatial characteristics of a car cabin in ICEV to make assumptions on coherent integration of virtual sonification sources in BEV. The main hypothesis is that if we match the spatial distribution of the ICEV engine sound to design the virtual source in BEV, the resulting soundscape will be more coherent with user expectations and then more accepted.

## 2 Methods

To investigate spatial characteristics on the integration of virtual sound sources in the vehicle, we decided to investigate spatial characteristics of the existing soundscape in ICEV. Especially, the study of the engine sound spatial characteristics will give insights on how to integrate a virtual sound source representing the dynamic of the vehicle in BEV. Indeed, we hypothesize that reproducing the spatial characteristics of a traditional engine sound on the sonification of the dynamic will improve the overall consistency of the soundscape in BEV. The study is divided in five parts :

- measurement of the acoustic field inside the cabin in different dynamic scenarios in ICEV and BEV.
- analysis of spatial acoustic energy distribution, especially the engine sound spatial distribution in ICEV
- construction of a perceptual model of engine sound spatial distribution
- evaluation of the model on the realism of the interior soundscape in ICEV
- application of the model to the integration of vehicle dynamic sonification in BEV and evaluation of the consistency of the interior soundscape.

## 2.1 Measurements and virtual environment

The first step of this study is to measure the acoustic field in the vehicle during specific dynamic scenarios. For practical reasons, the acoustic field has been measured at the passenger viewpoint. We know that the context is important and that in-car experience is multimodal. To increase contextualization of participants and the sense of presence in a real moving vehicle, we constructed a virtual environment, rendering the audio visual environment at the front passenger viewpoint. Measurement procedure and virtual environment rendering setup is described in [3]. Using spherical microphone array, we were able to render the acoustic field accurately from all directions and focus on directional energy analysis.

In practice, perceptual investigations of car experiences in laboratory conditions are difficult to generalize due to the lack of contextualization. At the same time, in-situ experiments are time consuming and experimental conditions are difficult to control. We previously proposed a methodology adapted from urban soundscape planning to overcome this difficulty [3]. It consists of creating a virtual environment based on 3D audio-visual in-situ recordings of usual driving situations where the sense of presence and involvement in a car environment is ensured. Then, solutions may be tested and validated before in-situ implementation. The two experiments presented in this work were conducted in this virtual environment.

## 2.2 Acoustical analysis

From the measurements of the interior of ICEV at constant speed, we computed directional power maps to study the spatial characteristics of the soundfield at the front passenger viewpoint. The power maps illustrated on Fig. 1 are computed with optimized beamforming technique called MVDR algorithm [4]. The measurements are analyzed with time-frequency decomposition to concentrate on engine components and evaluate the spatial distribution of the engine sound. The spectrogram at the top of Fig. 1 clearly shows frequency bands where the engine sound is predominant (i.e. harmonic components). We used constant speed measurements to be able to average in time and get more robust results.

From the maps shown on Fig. 1, the spatial distribution of the engine sound is frequency dependent and acoustic energy comes from all directions around the passenger. This phenomenon may be explained by the fact that sound radiation in the cabin is mainly structure-born and that different structures resonate at specific frequency [5]. However, it is difficult to identify which component of the structure is responsible for the resonances, so we concentrate on constructing approximation models in the next section.

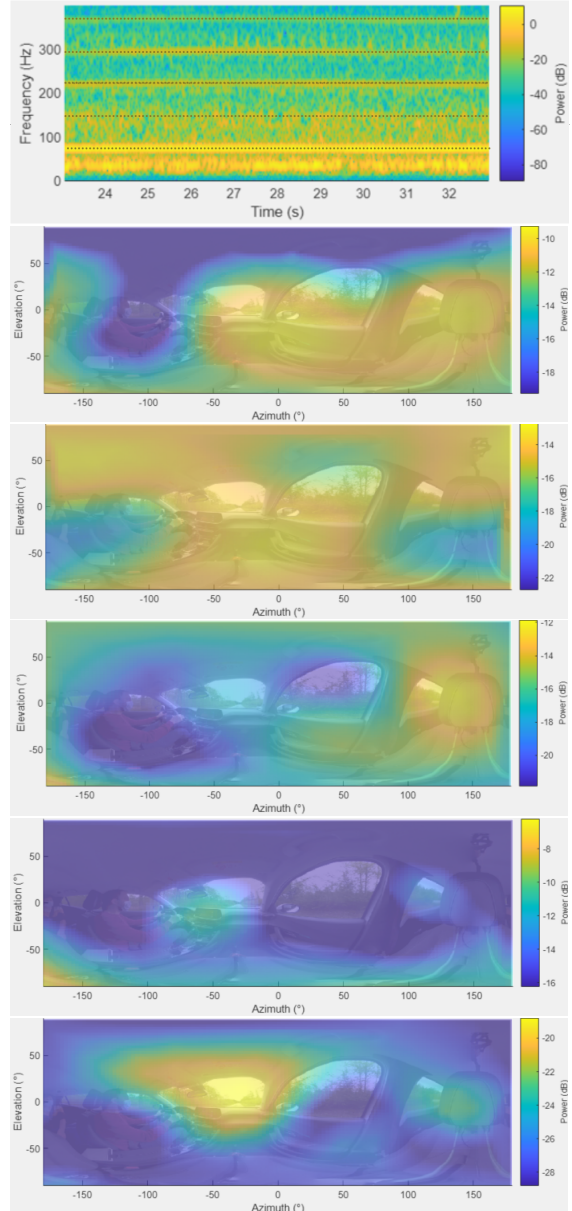


Figure 1: Top : Spectrogram of the omnidirectional component of a measurement in a compact ICEV at 30 km/h in 2nd gear. 2nd top to bottom : Directional power maps of the measurement at 70 Hz, 140 Hz, 210 Hz, 280 Hz and 350 Hz respectively. Dot lines on the spectrogram correspond to the frequency bands used to compute the power maps.

## 2.3 Spatial distribution models

From the acoustical analysis, we constructed two models of engine sound spatial distribution :

- one based on the acoustical analysis;
- one based on perceptual analysis of the acoustical model and on informal listening test.

We showed that the car cabin exhibits located resonances for specific frequencies. In order to model the phenomenon, we chose to filter the omnidirectional component of the engine sound by a complementary filter bank composed of eight bands uniformly distributed with center frequency from 100Hz to 800Hz. The frequency bands correspond to the bandwidth of a traditional four cylinder engine sound. Each band is then spatialized in specific location in front of the listener, distinct from the location of other bands. The idea was not to reproduce exactly the location of the resonances but rather based the model on the reproduction of the spread nature of frequency of the engine sound. The acoustical model is similar to a known technique to extend spatially a sound source based on time-frequency decomposition [6] [7].

From informal listening test and the aforementioned literature, we constructed a model to reproduce the perceptual characteristics of the engine sound. The distinct location of each frequency band results in acoustic energy coming from different directions and informal listening test revealed that the human ear doesn't segregate each frequency band but constructs a coherent and unique sound stream having a higher apparent width than if each frequency band was co-localized. Perceptually, the apparent width of a sound source is related to the inter-aural cross correlation (IACC). To control the IACC, it is possible to control the correlation between physical sound sources (i.e. loudspeakers) placed at distinct locations respectively [8]. Each physical source is associated with an all-pass filter with a random phase. The omnidirectional component of the engine sound is sent through each physical source and filtered to obtain a decorrelated version of the engine sound at each physical source. The resulting sound source at the subject location is an engine sound with an important apparent width. Both models are *a priori* similar in terms of spatial image of the source. Listening tests will validate this hypothesis perceptually and a comparison with the real spatial image of the engine will be performed. In a second experiment, both models will be used to integrate a monophonic vehicle dynamic sonification feedback and the coherency of the global soundscape will be evaluated.

## 3 Perceptual evaluation

A subjective experiment has been conducted to verify that both spatial distribution models improve the realism of the soundscape in ICEV cabin compared to in front punctual source model. A second experiment

with the same protocol has been conducted to assess the impact of both spatial distribution models on soundscape coherency in BEV.

### 3.1 Experiment 1

**Experimental setup:** The listening test was performed in the virtual environment described in section 2.1 and in details in [3]. The visual field is rendered through a Head Mounted Display (HMD) and is not altered for all conditions. The acoustic field is rendered on 42 loudspeakers organized on a 4m diameter sphere around the listener. He is seated on a fixed seat to reproduce a car seat.

**Subjects:** 20 subjects participated in the experiment. They were aged from 21 to 57 years old (mean : 31, std : 5). 10 participants were part of the laboratory and cannot be considered as completely naive listener.

**Stimuli:** The audio-visual measurements used to process the stimuli were captured on a closed road in a compact urban ICEV. Four scenes of five seconds each were selected based on the dynamic of the vehicle : two accelerations, one deceleration and one constant speed. Four *spatializations* of the engine sound were tested :

- Reference (i.e. *Ref*) corresponding to the measured acoustic environment (4th order HOA measurement) in a moving ICEV;
- Condition 0 (i.e. *C0*) corresponding to the the same measurement where the engine sound has been separated and removed using additive analysis estimation model [9]. Then, the omnidirectional component of the extracted engine sound is re-spatialized using the acoustical model described in 2.3.
- Condition 1 (i.e. *C1*) corresponding to the same extraction process but here the omnidirectional component of the engine sound is re-spatialized using the perceptual model described in 2.3.
- Condition 2 (i.e. *C2*) corresponding to the same extraction process but here the omnidirectional component of the engine sound is played back in a single loudspeaker in front of the listener. This condition corresponds to a punctual source.

In total, the experiment was composed of 16 different stimuli ( $4 \text{ scenes} \times 4 \text{ spatializations}$ ).

**Procedure:** The experiment consisted of a paired comparison task based on the perceived realism of the scenes. A total of 24 pairs (6 pairs corresponding to all the combinations between the four *spatializations* for each of the four scenes) was presented in a random order. Each pair A-B was composed of two different *spatializations* of the engine sound of a same scene (i.e.

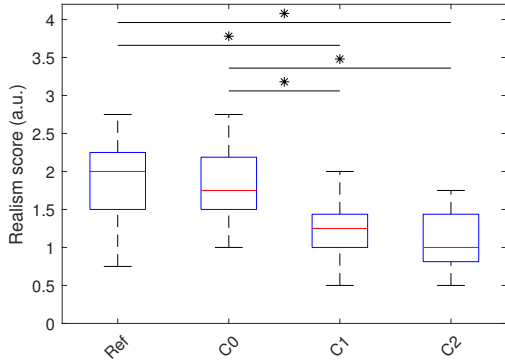


Figure 2: Realism score averaged across scenes for each *spatialization* condition. \* means that two conditions are statistically different ( $p < 0.05$ ).

two different scenes were not directly compared) so that *spatialization* is the only changing factor between A and B. The order of presentation between A and B was random and A and B were always different (i.e. the same stimuli were not compared). Participants were asked to listen to each pair A-B (as often as they wanted) and to choose between A and B (no tie) which one is the most realistic. Participants were explicitly told to focus on the acoustic environment and that there was no difference in the visual environment. The test lasted about 15 min on average.

**Results:** Data were collected into  $4 \times 4$  matrices  $\Delta_p^q$  for each participant  $p$  and for each scene  $q$ . The different *spatializations* are noted in the vector  $S = \{Ref, C0, C1, C2\}$ . Each matrix was completed as follows: if stimulus A ( $S_i$ ) was judged more realistic than stimulus B  $S_j$ , then  $\Delta_p^q(i, j)$  was set to 1; if stimulus B was judged more realistic than stimulus A, then  $\Delta_p^q(j, i)$  was set to 1. Since identical pairs A-A were not tested, the diagonal of was set to 0 by default. Each matrix was converted to construct a one-dimensional subjective scale, i.e. a  $4 \times 1$  vector of scores, by computing the sum value of each column of  $\Delta_p^q$ . The scores (obtained for each participant for each scene) reveal the relative perceived realism for the four *spatializations* ( $S_i, i = 1, 2, 3, 4$ ). Realism scores averaged across participants and scenes are illustrated on Fig. 2.

Statistical analysis showed same results for all scenes independently. Then, the factor *scene* is not considered in further analysis. Data are not normally distributed, non parametric analysis has been used. On the mean score across scenes, Friedman test revealed a significant effect of the *spatialization* factor ( $\chi^2(3) = 18.09, p = 0.0004$ ). Wilcoxon signed rank post-hoc analysis exhibits that *Ref* and *C0* are equivalent and the most realistic. *Ref* is significantly more realistic than *C1* and *C2* ( $p = 0.00084$  and  $p = 0.00034$  respectively). *C0* is also more realistic than *C1* and *C2* ( $p = 0.00033$  and  $p = 0.00009$  respectively). *C1* and *C2* are statistically equivalent.

Then, the results showed that the reference and the acoustical model are equivalent and more realistic than the perceptual model and the punctual source model.

## 3.2 Experiment 2

The second experiment aims to apply the same spatial distribution models to dynamic auditory feedback in BEV and verify if matching the spatial distribution of dynamic auditory feedback in BEV to the ICEV engine sound spatial distribution improves the coherency of the soundscape.

**Experimental setup:** Experimental setup is the same for the second experiment.

**Subjects:** All participants of Experiment 1 also participated in Experiment 2. The session was performed after the first experiment on the same day.

**Stimuli:** The audio-visual measurements used to process the stimuli were captured on a closed road in a compact urban BEV. Four scenes of five seconds each were selected : two accelerations, one deceleration and one constant speed. Here, there is no predominant motor sound. The objective is to add a virtual source to bring back dynamic auditory feedback inside the vehicle. The dynamic auditory feedback is based on the Shepard-Risset illusion or tone and has been successfully used for sonification of the dynamic in BEV [1]. The Shepard-Risset tone is composed of an infinite harmonic comb. Each comb component is modulated in amplitude following a Gaussian window in the frequency domain. In other words, low and high frequency components are quieter than mid frequency components. Increasing or decreasing sweep of each component gives the impression of a forever ascending or descending tone while the spectral centroid of the sound remains the same. This illusion is well suited for the perception of BEV dynamic and avoid annoying high frequencies at high speed or in-audible differences for small dynamic changes. In this experiment, we chose a wide band Gaussian window of 10 octaves, and map the central frequency of the window to the vehicle speed. The comb sweep speed is mapped to the acceleration. The scenes were played back unaltered in the VE and the virtual source was integrated in the soundscape following four different *spatialization* conditions :

- Condition 0 (i.e. *C0*) corresponding to the acoustical model described in section 2.3,
- Condition 1 (i.e. *C1*) corresponding to the perceptual model described in section 2.3,
- Condition 2 (i.e. *C2*) corresponding to a punctual source located in front of the listener,
- Condition 3 (i.e. *C3*) corresponding to a punctual source located at the bottom right of the

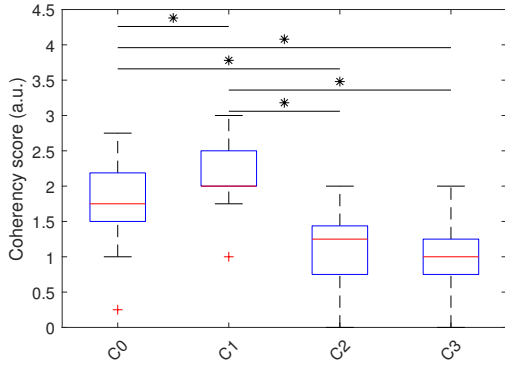


Figure 3: Coherence score averaged across scenes for each *spatialization* condition. '\*' means that two conditions are statistically different ( $p < 0.05$ )

listener, this condition introduces an incoherent location to the source which is supposed to be located at the front of the vehicle.

**Procedure:** The procedure was also a paired comparison task but the comparison criterion was different. Realism cannot be compared in the context of in-existent sound source. Then, participants were asked to judge the coherency of the soundscape in the cabin compared to their day to day experience inside a car. The idea is that if the overall soundscape is coherent, the integration of the dynamic auditory feedback is more likely to be accepted.

**Results:** Data were collected as in the first experiment with the *spatialization* vector being  $S = \{C0, C1, C2, C3\}$ . Statistical analysis showed same results for all scenes independently. Then, the factor *scene* is not considered in further analysis. Data are not normally distributed, non parametric analysis has been used. On the mean score across scenes, Friedman test revealed a significant effect of the *spatialization* factor ( $\chi^2(3) = 27.85, p = 0.000004$ ). Wilcoxon signed rank post-hoc analysis exhibits that *C1* is more coherent than *C0*, *C2* and *C3* ( $p = 0.0295$ ,  $p = 0.000002$  and  $p = 0.000002$  respectively). *C0* is more coherent than *C2* and *C3* ( $p = 0.0009$  and  $p = 0.0003$  respectively). *C2* and *C3* are statistically equivalent. Then, the results showed that the perceptual model creates a more coherent soundscape than the acoustical model. Both punctual source models exhibits a less coherent soundscape.

## 4 Discussion

The perceptual evaluation of engine sound spatial distribution clearly shows the impact of spatial characteristics of the engine sound on the perception of an interior car. First, the task was understood and feasible by most participants because unaltered rendering of measurements (i.e. the reference) was rated

as the most realistic. Without being explicitly told to, subjects focused on the engine sound to discriminate each scene where only spatial characteristics were changing. The acoustical model of spatial distribution was not discriminated from the reference. This validates the proposed model of spatial distribution of the engine sound which did not focus on respecting the position of each resonance but rather on the spatial separation of resonances regardless of their true location. This also means that the spatial separation of resonances constitutes an important perceptual feature of an interior car and that the true location of each resonance is not perceived. Surprisingly, the perceptual model was rated as unrealistic as the punctual source model. The motivation to develop this model was to have a psychoacoustic approach to the phenomenon of spatially separated resonances. Psychoacoustic studies [7] [6] claimed that separate resonances of a same sound results in a spatially extended source. In this case, results show that we perceive more than the extended characteristics of the engine sound and that it is necessary to reproduce the spatial separation of resonances to match the spatial characteristics of a real engine sound.

The second experiment shows interesting results and validate the main hypothesis of the study : matching the spatial distribution of the dynamic auditory feedback to the spatial distribution of the engine sound improves significantly the coherency of the interior soundscape compared to the punctual source models. However, the acoustical model produces a less coherent soundscape than the perceptual model. The spectral content of an traditional engine sound is richer than the dynamic auditory feedback designed in this experiment. The spatial separation between frequency bands in the acoustical model may have introduced segregated sound streams between frequency bands or create the feeling of a moving source in high dynamic scenarios. Regarding the position of the virtual source, subjects did not rate the eccentric position of the source as less coherent than a central position. *A priori*, we could have considered the position detrimental in terms of coherency because we expect the dynamic auditory feedback to come from the front as in a ICEV. The result shows that the spatial extent of the source took over the position of the source.

More generally, a vehicle exhibits a particular structure which creates a structure-born acoustic environment. It requires to reproduce the same phenomenon to integrate a virtual source coherently and naturally in this environment.

## 5 Summary

In this study, we investigated integration strategies of vehicle auditory feedback for BEV. The strategies are based on the spatial analysis of ICEV soundscape and on the reproduction of the spatial distribution of traditional engine noise inside the cabin. Directional

analysis of the acoustic environment at the front passenger viewpoint exhibits a specific distribution of the engine noise. We constructed a frequency model based on the spatial separation of frequency bands to reproduce this spatial distribution and a temporal model aiming at reproducing the perceptual effect of this distribution (a spatially extended source). Perceptual evaluation of both models demonstrates that the spatial characteristics of the engine noise have an impact on the perception of an interior car. The frequency model reproduces the spatial cues of the engine required to reproduce an interior vehicle soundscape as realistic as in-situ measurements. Also, integrating vehicle dynamic auditory feedback with both models improves the perceived coherency of the soundscape with an higher impact of the temporal model that may be due to a difference of spectral content between the designed sound and a traditional engine noise. Reproducing the spatial characteristics of a specific environment to integrate a virtual source create a more coherent and natural soundscape. The results are encouraging for real applications. The temporal model proposed in this paper can be easily implemented in a real vehicle. The next step will be to verify that the results obtained in the virtual environment are reproducible in a real environment. It can validate the virtual environment and our methodology to investigate sound design integration in car interior.

## References

- [1] S. Denjean, R. Kronland-Martinet, V. Roussarie, and S. Ystad, “Zero-emission vehicles sonification strategy based on shepard-risset glissando,” in *International Symposium on Computer Music Multidisciplinary Research*, pp. 709–724, Springer, 2019.
- [2] D. Y. Gwak, K. Yoon, Y. Seong, and S. Lee, “Application of subharmonics for active sound design of electric vehicles,” *The Journal of the Acoustical Society of America*, vol. 136, no. 6, pp. EL391–EL397, 2014.
- [3] T. Dupré, S. Denjean, M. Aramaki, and R. Kronland-Martinet, “Spatial sound design in a car cockpit: Challenges and perspectives,” in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–5, IEEE, 2021.
- [4] B. Rafaely, *Fundamentals of spherical array processing*, vol. 8. Springer, 2015.
- [5] H. B. Huang, X. R. Huang, M. L. Yang, T. C. Lim, and W. P. Ding, “Identification of vehicle interior noise sources based on wavelet transform and partial coherence analysis,” *Mechanical Systems and Signal Processing*, vol. 109, pp. 247–267, 2018.
- [6] T. Pihlajamäki, O. Santala, and V. Pulkki, “Synthesis of spatially extended virtual source with time-frequency decomposition of mono signals,” *Journal of the Audio Engineering Society*, vol. 62, no. 7/8, pp. 467–484, 2014.
- [7] T. Hirvonen and V. Pulkki, “Perception and analysis of selected auditory events with frequency-dependent directions,” *Journal of the Audio Engineering Society*, vol. 54, no. 9, pp. 803–814, 2006.
- [8] G. S. Kendall, “The decorrelation of audio signals and its impact on spatial imagery,” *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, 1995.
- [9] M. Caetano and P. Depalle, “On the estimation of sinusoidal parameters via parabolic interpolation of scaled magnitude spectra,” in *2021 24th International Conference on Digital Audio Effects (DAFx)*, pp. 81–88, IEEE, 2021.