Spatial Integration of Dynamic Auditory Feedback in Electric Vehicle Interior

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

Spatial Integration of Dynamic Auditory Feedback in Electric Vehicle Interior

THÉOPHILE DUPRÉ,[1,2] *AES Student Member,* **SÉBASTIEN DENJEAN,**[2] **MITSUKO ARAMAKI,**[1]
(dupre@prism.cnrs.fr) (sebastien.denjean@stellantis.com) (aramaki@prism.cnrs.fr)
AND RICHARD KRONLAND-MARTINET,[1]
(kronland@prism.cnrs.fr)

[1] *Aix-Marseille Univ, CNRS, PRISM, Marseille, France*
[2] *Automotive Research and Advanced Engineering, Stellantis, Vélizy, France*

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.

0 Introduction

The 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, user experience is mainly affected by a different acoustic environment or soundscape due to a different powertrain generated sound [1, 2]. Even if the environment is quieter, the absence of predominant motor sound may deteriorate the soundscape by unmasking unwanted noises [3], while no longer informs the driver about the vehicle dynamic [4] nor about vehicle characteristic [5, 6]. The driving experience is highly affected and the need for new solutions is expressed by car manufacturers [7, 8].

For few years, researchers have worked with sonification processes for the so called active sound design, that aimed to bring back the dynamic auditory feedback to the driver [9]. Originally in ICEV, active sound design consisted in enhancing the engine sound signature by synthesizing corresponding engine harmonic content inside the cabin through the audio system to modify the vehicle perception [10]. The same principle has been proposed in BEV [11]. Subharmonic generation is used to create a machine-like sound [12, 13]. It has been shown by Doleschal et al. that it creates a more pleasant soundscape by masking other noise sources and merging the normal electric motor sound [14]. Maunder et al. proposed to capture the electric motor vibration with an accelerometer and enhance, tune and replay it in real-time in the cabin [15]. Adaptive design have also been proposed to adapt the auditory feedback timbre depending on driver's emotion [16] and driving style [17]. Denjean et al. studied the influence of engine sound feedback on the perception of motion [4]. They noted that the absence of gear in BEV powertrain involves less frequency variation for the same dynamic variation. To overcome this limitation, they proposed to use the Shepard-Risset illusion that give the impression of pitch variation without variation of spectral content [18].

So far, active sound design has been mainly focused on the design of the sound itself and the correct transmission of the vehicle dynamics. To the best of the author's knowledge, the integration of the sound in the environment has not been studied except for loudness considerations. Yet, some users report a lack of blending with the surrounding environment and other perceptual cues. The integration of this virtual sound source in the interior soundscape augments the perception of vehicle motion. We may then consider the cockpit environment as an augmented reality environment. In this context, auditory source must be integrated seamlessly into a real environment in order to be accepted by user. As explained by Neidhardt et al., the acoustical properties of the virtual element must match with the real environment [19] and be in agreement with an internal reference develop by people from their listening experience in everyday life [20]. The environment may require specific

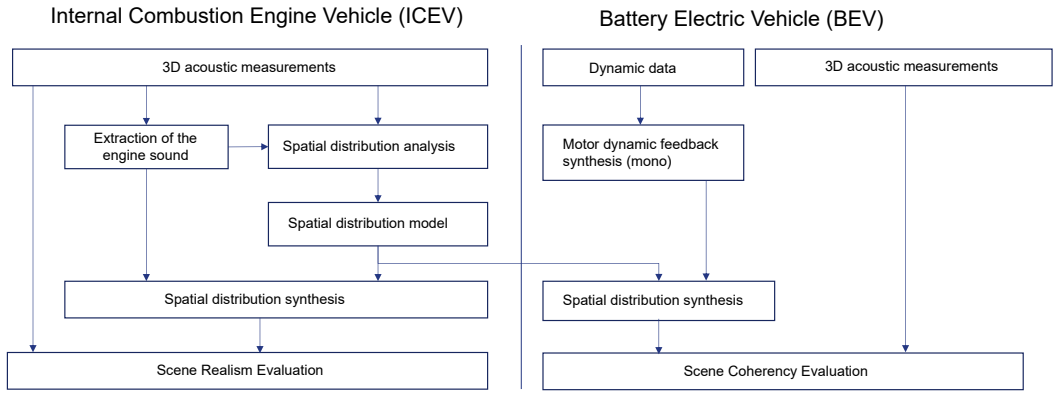

Fig. 1. General diagram of the method. Analysis and modeling of the engine sound spatial distribution is evaluated perceptually by resynthesis of the scene. Engine distribution model is applied to dynamic feedback in BEV and the overall coherency is evaluated.

auditory appearance of sources, i.e. source loudness, width or location. As a counter example, an unsuccessful integration would be to include a sound source in a reverberant environment, without applying the appropriate reverberation on the added source. The resulting soundscape would be perceived as incoherent because the source would not merge with the surrounding environment. In the case of BEV interior soundscape, Cao et al. studied the dynamic auditory feedback loudness based on the expected loudness of the engine in ICEV [21]. They reported an improvement of the pleasantness when a dynamic feedback is integrated with the same loudness variations as the engine in ICEV. However, it was compared with no feedback only. The car cabin also exhibits a particular configuration that might require specific spatial characteristics of the virtual source to be coherently integrated. Substitution of the engine noise by the active sound design may also involve a certain expectation by the users.

In this study, the authors analysed acoustical and perceptual spatial characteristics of a car cabin and especially the engine noise in ICEV to make assumptions on coherent integration of virtual dynamic feedback 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 the user expectations and then more accepted. Section 1 explains the methodology that we used to test this hypothesis. The virtual environment where the experiments were conducted is presented in Section 2. Then, the spatial distribution analysis and modeling is detailed in Section 3. Finally, the hypothesis is tested in Section 4 and the results are discussed in Section 5.

1 Method

The objective of the study is to integrate coherently a virtual source in the interior soundscape of electric vehicles. Coherent integration is achieved if the virtual source match user expectation of how the source should sound in this environment. Here, the virtual source is an auditory feedback on vehicle dynamic and take the place of the engine sound in ICEV. User expectation may be influenced by the characteristics of ICEV soundscape in general and by the engine sound in particular. Then, the strategy is to identify perceptually relevant spatial characteristics of the engine sound and matches these spatial characteristics in the integration of a dynamic feedback in BEV.

Figure 1 presents each step of the study. First, 3D acoustic scenes have been recorded in both ICEV and BEV in driving situations as well as vehicle dynamic parameters (i.e. speed, acceleration, throttle opening and engine speed if relevant). These recordings are used to analyse spatial features of ICEV soundscape and later reproduce the overall driving scenes. Then, engine spatial distribution models are derived from the scene analysis. Separation of the engine sound and the background noise allows to independently apply spatial models to the engine omnidirectional channel and combined the resulted engine sound to the original background noise to compare reproduced scene realism with the original recordings. In parallel, dynamic data are used to parameterize the auditory feedback synthesis and match the BEV recording dynamic. Spatial distribution models derived from engine distribution in ICEV are applied to integrate spatially the auditory feedback in the recorded scenes. Scene coherency is then evaluated because no real reference exists in this case. The term coherency has been preferred to the term plausibility, usually used in the evaluation of virtual and augmented acoustic environments [22], to avoid the evaluation of the plausibility that the electric motor produces such auditory feedback. The idea is that if the overall experience is coherent, the integration of the dynamic auditory feedback is more likely to be accepted.

Context plays an essential role in the perception of a soundscape, the perception of the sounds or the sound sources can differ depending on how the stimuli are presented [23, 24, 25]. It is important to ensure that comparable cognitive processes are involved in a laboratory experiment as in the same experiment in-situ [23, 26]. This concept is called ecological validity and was first developed for the study of visual perception by Brunswik [27] and then Gibson [28]. In the field of auditory cognition, this notion was developed by Gaver [29, 30] and extended by the notion of *everyday listening*, ”the perception of sound-

producing events", as opposed to *musical listening*, "the perception of the sound itself". In a car cabin, several of our senses are involved in order to create a representation of the scene we are living, mainly vision, hearing, proprioception and vibration sensation. The objective is to recreate an ecological experimental environment (i.e. representative of the real environment that we simulate) allowing to test our hypothesis and to favour an "everyday listening" that will allow to generalize the results to a real situation. We previously proposed a methodology [31] adapted from urban soundscape planning [32] to capture, analyze and reproduce real driving scenes. 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 is ensured. Indeed, it has been shown that combining different modalities helps improve the sense of presence in a Virtual Environment (VE) [33, 34, 35]. Then, solutions can be tested and validated before in-situ implementation. The two experiments presented in this article were conducted in this virtual environment and the scene analysis was based on the recordings. Next section will detail the technical aspects of in-situ recordings and virtual environment construction.

2 Cabin environment rendering

2.1 In-situ recordings

In-situ recordings involve measurements of the acoustic field, the visual field and dynamic parameters of the vehicle. The acoustic field is recorded using a spherical microphone array composed of 32 microphones (*Eigenmike32*). Spherical microphone arrays capture the sound field at a point in 3D, which can be analyzed using beamforming processing [36] and accurately reproduced in all directions using specific processing such as the ambisonic framework. [37]. The visual field is recorded by a 3D camera (*Insta360 Pro*) composed of 6 optic lenses on the azimutal plane, each having a 6k resolution. Reconstruction processing of the visual field is done by Insta360 professional stitching software allowing 360° 3D videos with 4k resolution (*Insta360 Stitcher*). Vehicle speed, engine speed for ICEV measurements and throttle opening are recorded simultaneously. The camera is placed on the front passenger seat at head height with the microphone just over it as in Figure 2. Time and orientation alignments of the recordings are done by hand claps. For technical reasons, it is not possible to record at the driver viewpoint while someone is driving. Also, the VE is a passive environment. In order to maximize the consistency of the experience and improve the sense of presence, the VE simulates the experience of the front passenger.

The measurements were conducted on a straight closed road in order to record controlled scenarios and limit cybersickness due to turns. The road was 3 meters wide and was surrounded by lawns, trees and bushes. Due to the limited length of the road, the maximum speed was 100 km/h. We chose to record two B segment vehicles (one thermal and one electric) because it is the most common segment in Europe [38]. The two models were very close in geometry (constructed on the same platform) to guarantee comparable acoustic responses. Several dynamic scenarios were recorded : accelerations, constant speeds and decelerations at different gear ratios (2nd, 3rd and 4th, only for the thermal vehicle) and different speeds (from 20km/h to 100km/h).

2.2 Virtual environment

We constructed a VE to render driving situations previously recorded in-situ. A Virtual Reality (VR) headset is used to render the visual field (*Oculus Quest*). Ambisonic

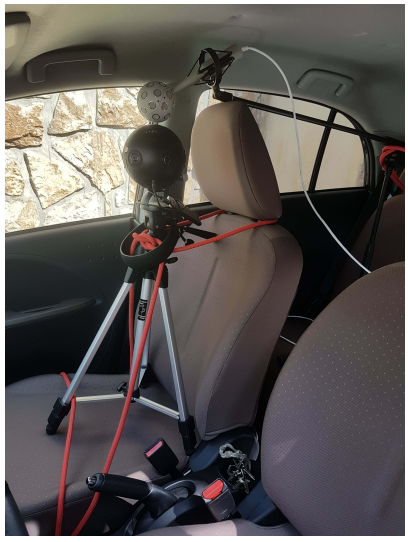

Fig. 2. Experimental setup to record 3D video and 3D audio from passenger seat viewpoint in driving situations

Fig. 3. Multichannel rendering system

channels are decoded on 42 loudspeakers (*Genelec 8020C*) distributed on a sphere. The ambisonic format was chosen for two reasons. First, it is known that ambisonics is adequate when immersion and envelopment are important [23] at the expense of localization performance. Source localization is less important than immersion and envelopment because we intend to evaluate the quality of the acoustic environment in a global way. Furthermore, it has been shown that the appropriate reproduction system depends on the sound material [39]. Here the acoustic environment is an indoor environment composed of a multitude of sound sources that surround the listener; it has been shown that 3D ambisonics is most appropriate in this situation [39]. The setup is placed in an anechoic chamber to avoid any influence of the room on the reconstruction of the sound field. An illustration of the setup is shown on Figure 3. A Unity application has been developed to play back 360° video recordings in the VR headset. A modular synthesizer developed by Stellantis [40] is used to synthesize the dynamic auditory feedback. Each module parameter can be parametrized by a function of the vehicle dynamic data (speed, acceleration, throttle opening, etc). A Max/MSP patch controls the recorded scene and the synthesizer. Spatialization of virtual sources is controlled in real-time by IRCAM Spat Max Toolbox [41]. Video and audio synchronization is done via the OSC protocol.

3 Spatial distribution analysis and models

The car interior is a complex acoustic environment. Sounds perceived inside the cabin mainly comes from outside. At low speeds, engine noise and road/tire noise are dominant, while aerodynamic noise tends to dominate at high speeds. The geometrical configuration its variety of materials, as well as the transmission from the source location to the interior, make it difficult to predict the acoustic behavior inside the vehicle. These sources cannot be considered as point sources. For example, tire/road noise comes from all four wheels and the transmission within the vehicle is partly carried by the structure and is radiated by the entire vehicle structure. The same remarks can be made about engine noise. The analysis aims to qualitatively characterize the spatial distribution of the engine noise energy and to develop approximation models of the phenomenon.

3.1 Spatial distribution analysis

Spherical array microphone recording allows to analyse spatial information and to deduce directions of arrival of the energy at the point of measurement. Beamforming methods mix the signals from physical receivers to create virtual receivers in given directions with a specific directivity pattern [42]. By scanning in every direction, an acoustic power map can be drawn that illustrates the energy distribution around the measurement point. In this analysis, Minimum Variance Distorsionless Response (MVDR) beamformer has been chosen to analyze the spatial distribution of the engine sound. MVDR beamformer is a widely used optimized beamformer designed to reject uncorrelated noise while maintaining unity gain in the look direction [43]. The beampattern is optimized to specifically reject region where noise is present. It is suited for noisy environment, hence an interior car.

The measurements are analyzed with time-frequency decomposition to concentrate on each engine component. Figure 4 (bottom) shows the Short Time Fourier Trans-

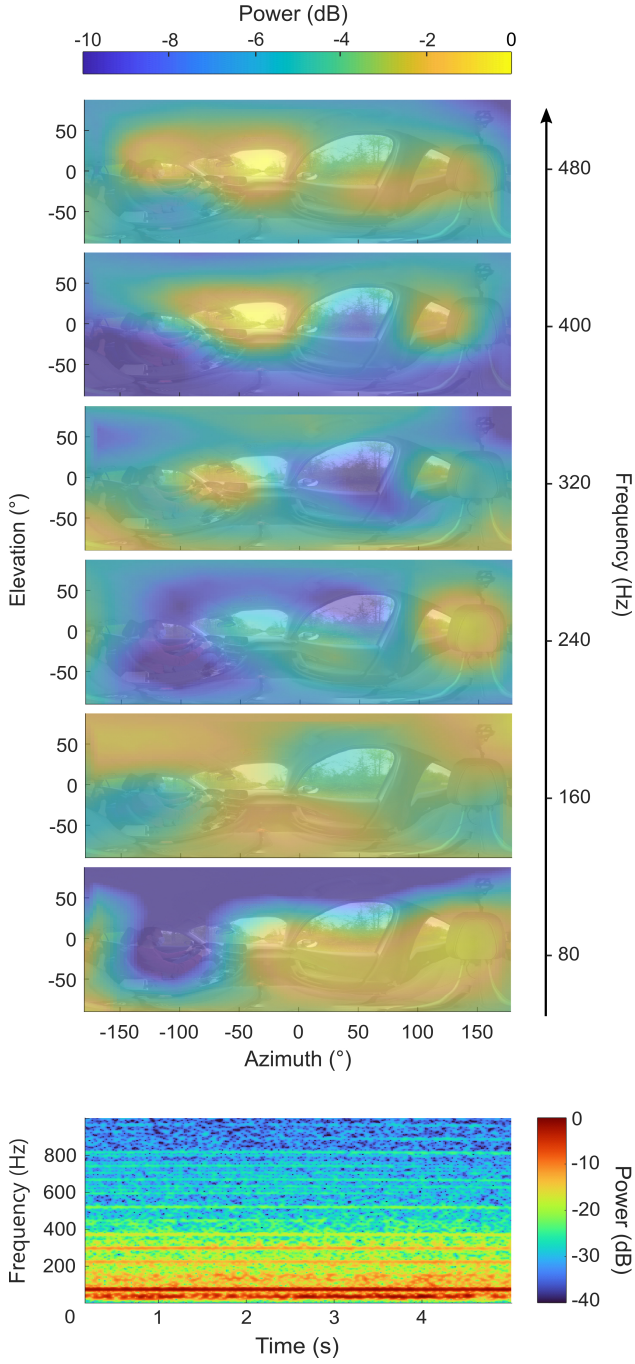

Fig. 4. Bottom: Magnitude spectrogram of the STFT of an ICEV HOA measurement in 2nd gear at 30 km/h at front passenger viewpoint (only the omnidirectional component is shown). Top: Power maps obtained with MVDR beamforming computed on the same measurement. All HOA channels are used to maximize spatial resolution. Each map correspond to a frequency band containing an engine harmonic.

form (STFT) magnitude spectrogram of the omnidirectional component of ICEV HOA recordings in 2nd gear at 30 km/h. The engine noise is harmonic and proportional to engine speed. Frequency resolution needs to be high to be able to discriminate each harmonic component (window length is here 4096 samples at 48 kHz). At low frequency, the components are clearly dominant. They are also stationary at constant speed. Hence, power maps for each frequency band and each time frame are computed and time averaged to mitigate noise. Each map is normalized by the maximum power. 360° image of the car cabin at the front passenger viewpoint is overlaid on the power maps to localize power maximum and minimum.

Maps corresponding to frequency bands containing engine harmonics are illustrated on Figure 4 (top). It is clear that engine spatial distribution is frequency dependent and varies greatly even for small frequency shifts. Intuitively, one could expect the energy to come mostly from the front where the engine is located. Under 300 Hz, where most of the engine sound energy is concentrated, the energy comes consecutively from the left side, the bottom right and from behind. Similar maps for different engine configurations and other dynamic scenarios have been obtained. However, time averaging is necessary to limit noise, hence the choice of a stationary measurement to present the results. Maps at frequency bands where no strong engine harmonic is present exhibit a more diffuse sound field where no strong resonance can be located. It emphasizes the specificity and the steadiness of the engine spatial distribution. This phenomenon may be explained by the transmission of the sound from the engine compartment to the cabin. Engine vibrations are transmitted by the mechanical structure from the engine to the cabin wall. Then, cabin walls radiate at their specific eigen frequencies. Transfer path analysis could give further information on the phenomenon and may be able to predict the presence and location of each resonances [44, 45] but it is out of the scope of the paper.

3.2 Spatial distribution models

From the listener perspectives, spatial separation of engine resonances does not produce segregated auditory streams leading to several sources but it involves binaural differences. Perceived apparent width of a source is known to be related to the Interaural Cross Correlation (IACC) [46]. Perceptual impact of this spatial distribution could be a large apparent source width in the cabin. Two models presented on Figure 5 have been developed to test this hypothesis. The first approach tends to reproduce the spatial distribution described above and the second approach tends to reproduce the possible perceptual effect. Both models exhibit the same architecture. A monophonic source signal S is sent through a filter bank to compute N secondary sources S_n. Each secondary source S_n is then spatially located apart from others. Depending on the location of each source relatively to others, IACC at the listener's ears varies and the apparent width of the source signal S can be adjusted.

The first approach is inspired by the power map analysis and called the frequency approach because the source signal is filtered by a complementary filter bank. The idea is to reproduce the spatial scattering of resonances inside the cabin without necessarily matching the location of each resonance. Then, the source signal S is splitted into N complementary frequency bands to maintain the same energy noted S_n^f. We used the complementary filter bank from IRCAM Spat MAX toolbox [47] based on cascaded complementary IIR filters [48]. Number of bands N and corresponding central frequencies f_c must be chosen to cover the source bandwidth. In the case of an engine sound, most the energy is concentrated under 500 Hz so we chose to split the source signal into $N = 8$ bands with $f_c = \{100, 150, 200, 250, 300, 400, 500, 700\}$. This approach is similar to a known technique to extend spatially a sound source based on time-frequency decomposition [49, 50].

The second approach is called the temporal approach because it is based on temporal decorrelation of the source signal S. It aims at reproducing the expected perceptual effect of the physical phenomenon. It involves deriving N uncorrelated secondary sources S_n^t from the source signal S. We used all-pass filters with random and uncorrelated phase response [51] to produce uncorrelated secondary sources. The filters are designed in the frequency domain: magnitude response is set to unity for all filters and the phase responses are randomly set from a Gaussian distribution. Then, FIR coefficients of each filter are computed by means of the inverse FFT of the combine magnitude and random phase response. The length of the FIR depends on the resolution in the frequency domain and can affect the transient and the timbre coloration of the signal. Engine noise or dynamic feedback in BEV are continuous sounds with smooth variations. Hence, filters' length is not critical and it is more important to concentrate on low timbre coloration. In the following experiments, we chose $N = 4$ uncorrelated secondary sources S_n^t, computed with 500 coefficients FIR filters. Informal listening tests confirmed no critical timbre coloration.

Spatial scattering may be obtained by several means. First, physical sources (i.e. loudspeakers) can be placed at the location of each secondary source. Alternatively, source location can be encoded in a specific format and later decoded for a specific physical source layout. In this study, recordings were encoded in the spherical harmonic domain. Then, we chose to encode the secondary sources of the frequency approach in the spherical harmonic domain and decode simultaneously both the synthesized part and the background noise of the recordings. In order to test the influence of another scatter methods and be closer to the technological constraints of an interior car, the temporal approach used direct physical loudspeaker at the position of the secondary sources. The position of each secondary source used in the following perceptual evaluation for both approaches is listed in Table 1. In the frequency model, the secondary sources have been scattered in the front hemisphere to reproduce the large separation observed on the measurements but to avoid a perception of a source located

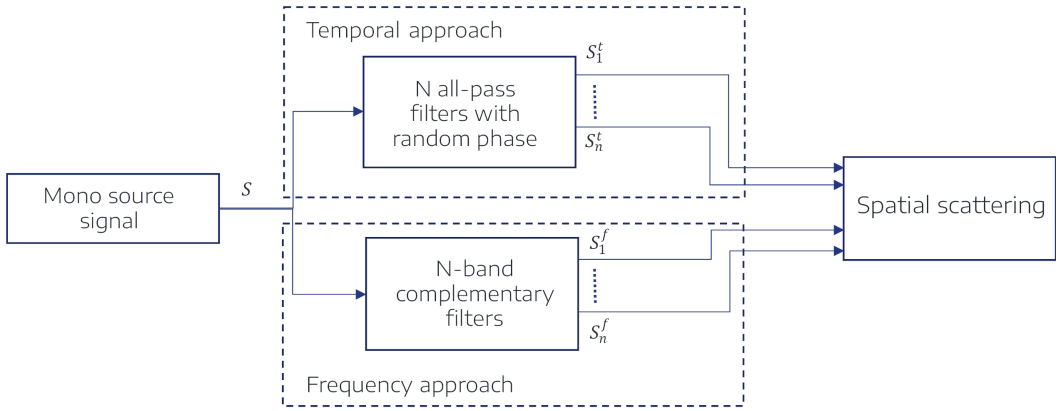

Fig. 5. Diagram of the spatial models to reproduce engine spatial distribution.

Table 1. Position of the secondary sources for each spatial model of engine spatial distribution.

S_n	Frequency model			Temporal model		
	Azi	Elev	Dis	Azi	Elev	Dis
S_1	70	30	1	30	30	1
S_2	-30	-30	1	-30	30	1
S_3	30	30	1	-30	-30	1
S_4	-70	-30	1	30	-30	1
S_5	-10	30	1			
S_6	50	-30	1			
S_7	-50	30	1			
S_8	10	-30	1			

Note. Azimuth (Azi) and Elevation (Elev) are expressed in degree and Distance (Dis) is expressed in meter. Origin of the coordinate system is the head of the passenger, 0° azimuth and 0° elevation correspond to the front of the passenger. Positive azimuthal and elevation coordinates are to the right and above the passenger.

in the back. In the temporal model, the position of sources have been manually tuned to match the measurements.

4 Perceptual evaluation

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 (Experiment 1). Both models will then be used to integrate a monophonic vehicle dynamic auditory feedback and the coherency of the global soundscape will be evaluated (Experiment 2).

4.1 Experiment 1 : ICEV scene realism evaluation

4.1.1 Experimental setup:

The listening test was performed in the virtual environment described in section 2 and in details in [31]. The subject was seated at the center of the setup on a fixed seat to reproduce a car seat.

4.1.2 Participants:

19 participants participated in the experiment (11 men, 8 women). They were aged from 21 to 57 years old (mean: 31, std: 5). 13 reported to drive at least once a month and 9 of them reported to drive several times a week. 8 participants were part of the laboratory team. Although they were not aware of the protocol and objective of the experiment, they cannot be considered as completely naive listeners, they will be referred as expert in the analysis. Participants were informed that they could stop the experiment at any time without justification and that they could rest whenever they wished.

4.1.3 Stimuli:

The audio-visual measurements used to process the stimuli were captured on a closed road in a compact urban ICEV as explained in Section 2.1. They resulted in 4th order HOA recordings and 4K resolution 360° video recordings. Four scenes of five seconds each were selected based on the dynamic of the vehicle: two accelerations, at low and high speed (2nd gear from 30 to 65 km/h and 4th gear from 60 to 80 km/h respectively), one constant speed (3rd gear at 50 km/h) and one deceleration (4th gear from 95 to 80 km/h). The scenes are referred as Acceleration 1, Acceleration 2, Stabilized and Deceleration respectively on Figure 6. Two accelerations were selected because we expected to have more differences between conditions due the engine noise being more present. The duration of the stimuli is limited to 5 seconds due to the length of the road available without turns, avoiding cybersickness. The engine noise was extracted from the measurement using the additive model described in [52]. The model was applied to each ambisonic channel independently. The engine sound is composed of harmonic components and we supposed that they emerged from the background noise, composed of tyre/road noise and aerodynamic noise. Background noise ambisonic channels were decoded on the loudspeaker sphere to ensure an authentic spatial reproduction of the sound field apart from the engine components. Four spatial configurations were compared to reproduce the engine noise spatial distribution:

- the Reference (i.e. C_0) corresponds to the unaltered 4th order ambisonic channels extracted from the measurement decoded on the loudspeaker sphere. The resulted sound field (background noise + engine noise) is identi-

cal to the original measurement, hence a reproduction of reality in the virtual environment;
- the Condition 1 (i.e. C_1) corresponds to the frequency model applied to the omnidirectional component extracted from the measurement;
- the Condition 2 (i.e. C_2) corresponds to the temporal model applied to the omnidirectional component;
- the Condition 3 (i.e. C_3) corresponds to a point source configuration applied to the omnidirectional component. The source is placed in front of the listener (position (0,0,1) in the coordinate system described in Table 1).

In total, the experiment was composed of 16 different stimuli (4 scenes $\times$ 4 spatial configurations).

4.1.4 Procedure:

The experiment consisted of a paired comparison task based on the perceived realism of the scenes. We chose a pairwise comparison protocol because our goal is to compare different models, direct comparison allows participants to focus more easily on the differences between models. Each pair A-B was composed of two different spatial configurations of the engine for a same scene (i.e. two different scenes were not directly compared) so that the spatial configuration 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). A total of 24 pairs were presented in a random order. Participants were asked to listen to each pair A-B (as many times as they wanted) and to choose between A and B (no tie) which one feels more realistic. Before the session began, the experimenter explained to the participant that the notion of realism refers to the similarity with his or her own experience of the automotive environment and that he or she should adopt a global listening approach (i.e., everyday listening) and not focus on the details. They were explicitly told that there was no difference in the visual environment. Participants were free to ask questions before or during the session. The session lasted 20 min on average.

4.1.5 Results:

Data were collected into 4×1 score vector S_p^q for each scene p and each participant q. Each vector value $S_p^q(i)$, is associated with one spatial configuration noted C_i with $i = \{0,1,2,3\}$. Each vector was completed as follow: if stimulus $A(C_i)$ was judged more realistic than stimulus $B(C_j)$, $S_p^q(i)$ was incremented by 1; if stimulus $B(C_j)$ was judged more realistic than stimulus $A(C_i)$, $S_p^q(j)$ was incremented by 1. In other words, $S_p^q(i)$ revealed the number of times spatial configuration C_i has been judged more realistic than any other configuration for scene p and participant q. Since no repetition were presented, $S_p^q(i) \in \{0,1,2,3\}$. This subjective scale is interpreted as the realism score associated with each spatial configuration C_i.

Analysis regarding expert and naive listeners revealed no statistical differences in their ratings of each spatial configuration as no interaction were found between the spatial configurations and the participant expertise (repeated measures ANOVA: $F(3,52) = 2.10$, $p > 0.1$).

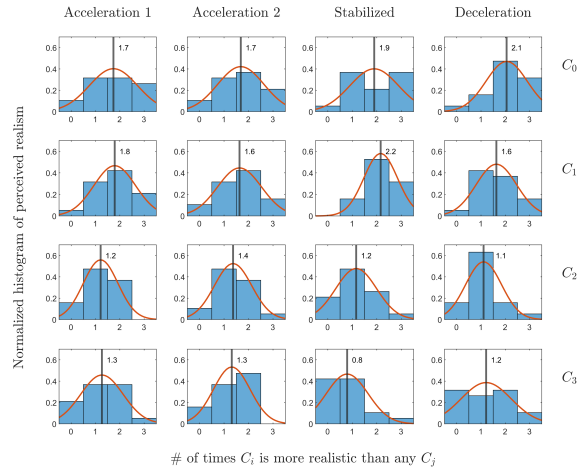

Fig. 6. Normalized histograms of the scene realism comparison between spatial configurations C_i for each scene. Red curves are Gaussian distributions fitted to the histograms. Thick vertical lines correspond to the mean of each distribution.

Figure 6 shows normalized histograms of $S_p^q(i)$. Gaussian fitting was computed to estimate a realism score for each spatial configuration and scene. From direct observations, results are consistent across scenes, C_0 and C_1 are perceived as more realistic than C_2 and C_3 independently of the scene. Surprisingly, the stabilized scene exhibits more variance between conditions followed by the deceleration scene. Non parametric statistic analysis of each scene demonstrates an influence of spatial configurations C_i on the realism score in case of stabilized and deceleration dynamic scenes (Friedman tests: $\chi^2(3) = 16.93$, $p < 0.001$ and $\chi^2(3) = 8.84$, $p < 0.05$ respectively). For both acceleration scenes, spatial configuration condition shows no statistical difference (Friedman tests: $\chi^2(3) = 3.9$, $p > 0.1$ and $\chi^2(3) = 1.05$, $p > 0.1$ respectively). Even if differences are not statistically significant in acceleration scenes, a tendency to improve realism with spatial configurations C_0 and C_1 is present in all scenes. We chose to mariginalize the score over the scenes and further investigate the differences between spatial configurations: $S^q(i) = \frac{1}{4}\sum_p S_p^q(i)$.

Figure 7 shows boxplot of S^q. Statistical analysis confirmed a significant effect of the spatial configurations C_i (Friedman test: $\chi^2(3) = 18.09$, $p < 0.001$). Wilcoxon signed rank post-hoc analysis exhibits that C_0 and C_1 do not differ from each other and are perceived as the most realistic. C_0 is significantly more realistic than C_2 and C_3 ($p < 0.001$ and $p < 0.001$ respectively). C_1 is also more realistic than C_2 and C_3 ($p < 0.001$ and $p < 0.001$ respectively). C_2 and C_3 do not differ. Then, the results showed that the reference and the frequency model are perceived as equally realistic and more than the temporal model and the point source model.

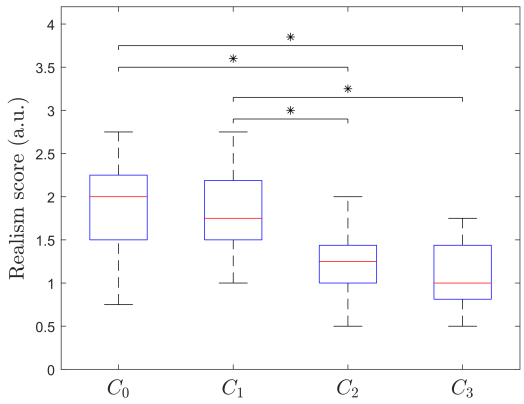

Fig. 7. Box plot of the scene realism score averaged over the scenes for each spatial configuration C_i. $*$ indicates a statistical difference ($p < 0.05$) between two spatial configuration.

4.2 Experiment 2: BEV scene coherency evaluation

The second experiment aims to apply the same spatial configurations to integrate dynamic auditory feedback in BEV and verify if matching the spatial distribution of the ICEV engine noise spatial distribution improves the coherency of the experience inside the virtual environment.

4.2.1 Experimental setup:

Experimental setup is the same as in Experiment 1 (see Section 4.1.1).

4.2.2 Participants:

All participants of Experiment 1 also participated in Experiment 2. The session was performed after the first experiment, the same day. Participants were encouraged to rest between the sessions.

4.2.3 Stimuli:

The audio-visual measurements used to process the stimuli were captured on the same closed road as in the previous experiment but in a compact urban BEV (see Section 2.1 for details). They resulted in 4th order HOA recordings and 4K resolution 360° video recordings. Four scenes of five seconds each were selected with the same dynamic profile as in the first experiment: two accelerations, at low and high speed, one constant speed and one deceleration (referred as Acceleration 1, Acceleration 2, Stabilized and Deceleration respectively on Figure 8). Four scenes of five seconds each were selected with the same dynamic profile as in the first experiment: two accelerations (from 30 to 65 km/h and from 65 to 100 km/h), one constant speed (at 50 km/h) and one deceleration (from 95 to 85 km/h). The scenes are referred as Acceleration 1, Acceleration 2, Stabilized and Deceleration respectively on Figure 8.

Here, there is no predominant motor sound. The objective is to add a virtual source to bring back dynamic auditory feedback inside the vehicle captured by the measurements. The dynamic auditory feedback is based on the Shepard-Risset illusion or tone and has been successfully used for sonification of dynamic in BEV [18]. The Shepard-Risset tone is composed of an infinite harmonic comb. Each comb component is modulated in amplitude by a cosine window in the frequency domain [53]. The following relation gives the amplitude a of the component:

$$a(f) = \begin{cases} \frac{1}{2}\left[1 - \cos\left(2\pi \frac{\log f - \log \frac{F_c}{2^{L/2}}}{L \log 2}\right)\right] & \text{if } \frac{F_c}{2^{L/2}} < f < 2^{L/2} F_c \\ 0 & \text{else} \end{cases} \quad (1)$$

where f is the frequency of the component, F_c is the central frequency of the window and L the window width in octaves. Low and high frequency components are quieter than mid frequency components. An 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 window of $L = 10$ octaves, and map the central frequency F_c to the vehicle speed. Each octave is composed three components corresponding to a major chord. The comb sweeping speed is mapped to the acceleration.

The HOA recordings were decoded on the loudspeaker sphere and ensured an authentic spatial reproduction of the acoustic field measured in the cabin. The virtual source was integrated in the environment following four spatial configurations:

- the Condition 1 (i.e. C_1) corresponding to the frequency model described in section 3.2;
- the Condition 2 (i.e. C_2) corresponding to the temporal model described in section 3.2;
- the Condition 3 (i.e. C_3) corresponding to a point source located in front of the listener (position (0,0,1) in the coordinate system described in Table 1);
- Condition 4 (i.e. C_4) corresponding to a point source located at the bottom right of the listener (position (45,-30,1) in the coordinate system described in Table 1). This location corresponds to direction of the front passenger closest loudspeaker in the real cabin.

The models are the same as the previous experiment except for the C_0 due to the absence of real reference in BEV. C_4 has been added to study the influence of the location of the virtual source. Also, this location corresponds to the existing solution used in real vehicle to integrate the dynamic feedback; no spatial aspect is considered resulting in a source perceived as coming from the closest loudspeaker.

4.2.4 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,

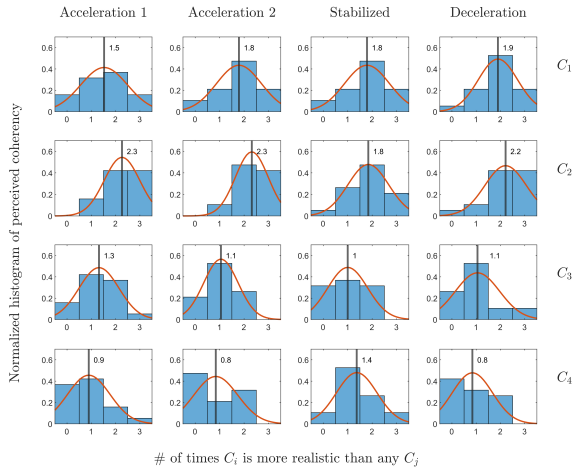

Fig. 8. Normalized histograms of the scene coherency comparison between spatial configurations C_i for each scene. Red curves are Gaussian distributions fitted to the histograms. Thick vertical lines correspond to the mean of each distribution.

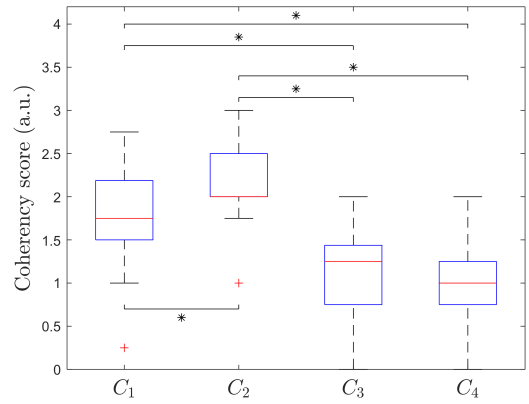

Fig. 9. Box plot of the scene coherency score averaged over the scenes for each spatial configuration C_i. $*$ indicates a statistical difference ($p < 0.05$) between two spatial configuration.

participants were asked to judge the coherency of the experience in the cabin compared to their day to day experience inside a car and their expectation. We explained this notion to the participants as follows: "As in the previous experiment, you will have to compare different scenes inside a vehicle but here the vehicle is electric, and we added sound feedback to replace the engine sound. As no real reference exists, we ask you to evaluate the most coherent scene based on your everyday experience of a car environment, i.e. which scene matches the most your expectation of a car environment". The session lasted 20min on average.

4.2.5 Results:

Data were collected in vector S_p^q for each scene p and each participant q, as in the first experiment. Each vector value $S_p^q(i)$, is associated with one spatial configuration noted C_i with $i = \{1,2,3,4\}$ and is interpreted as the coherency score of each spatial configuration C_i.

As in the first experiment, no differences were found between the ratings of expert and naive listeners (repeated measures ANOVA: $F(3,52) = 2.56$, $p > 0.05$).

Figure 8 shows normalized histograms of S_p^q. We estimated the coherency score as the mean of the fitted Gaussian distribution. Direct observations reveal that coherency is robust to change of dynamic except for C_2 and C_4 in the stabilized scene. This observation is confirmed by non parametric statistical analysis. The influence of the spatial configuration C_i is statistically significant in both acceleration scenes and the deceleration scene (Friedman tests: $\chi^2(3) = 14.51$, $p < 0.01$, $\chi^2(3) = 18.95$, $p < 0.001$ and $\chi^2(3) = 18.28$, $p < 0.001$ respectively) but not for the stabilized scene (Friedman test: $\chi^2(3) = 7.24$, $p > 0.05$). However, the same tendency appears for all scenes. As in the previous experiment, the coherency are marginalized over the scenes to investigate the differences between spatial configurations.

Figure 9 shows a boxplot of the mean coherency score for each spatial configuration. Still, Friedman test demonstrates an strong influence of the spatial configurations on the coherency ($\chi^2(3) = 27.85$, $p < 0.001$). Wilcoxon signed rank post-hoc analysis exhibits that C_2 is more coherent than C_1, C_3 and C_4 ($p < 0.001$, $p < 0.001$ and $p < 0.001$ respectively). C_1 is more coherent than C_3 and C_4 ($p < 0.001$ for both conditions). C_3 and C_4 do not differ statistically. Then, the results showed that the temporal model creates a more coherent experience than the frequency model. Both point source configurations create a less coherent experience.

5 Discussion

First, the task in Experiment 1 was understood and feasible by participants since unaltered rendering of measurements (i.e. C_0) was rated as the most realistic. Results of Experiment 1 clearly show the impact of spatial characteristics of the engine noise on the perception of an interior car. Without being explicitly mentioned, participants focused on the engine noise to discriminate the scenes in which only its spatial characteristics differed.

The spatial configuration based on the frequency model (i.e. C_1) was not discriminated from the reference (i.e. C_0). This result validates the proposed model of spatial frequency distribution of the engine noise which focused on the spatial scattering of resonances regardless of their actual location. It also means that the spatial separation of resonances constitutes an important perceptual feature of an interior car and that matching the position of each resonance does not impact the realism of the scene.

Surprisingly, the spatial configuration based on the temporal model (i.e. C_2) was rated as unrealistic just as the point source configuration (i.e. C_3). This model was motivated by a perceptual approach aiming at evaluating the need of spatially separated resonances. Actually, studies [50] [49] claimed that spatially scattered frequency bands of a same sound only results in a spatially extended source.

Hence, spatializing the engine noise as an extended source with a temporal model could have led to the same perceptual effect. However, it appears that such approach is not sufficient to match our expectation of a realistic interior scene. Binaural differences might be expected regarding the frequency content of the engine noise (low frequencies should be separated from higher frequency for example). Also, the radiation of the engine noise in the car interior is composed of a variety of physical sources with their own frequency content and specific locations. Since the temporal model involves the grouping of all sources as a single source, this method may not render correctly this reality and therefore our expectations.

Experiment 2 shows interesting results and validates the main hypothesis of the study: matching the spatial distribution of the dynamic auditory feedback to the spatial distribution of the engine noise significantly improves the coherency of the interior soundscape compared to both point source configurations (i.e. C_3 and C_4). Note that regarding the positions of the point source, participants did not rate the eccentric position of the source (i.e. C_4) as less coherent than the central position (i.e. C_3). However, we could have considered the position detrimental in terms of coherency since we expect the dynamic auditory feedback to come from the front as in an ICEV (the position of the eccentric source was at 45° on the right, which corresponds to the location of door loudspeaker in a car). The obtained results mean that the spatial extent of the virtual source takes over the position of the source and that the engine spatial distribution has a key role in the expected characteristics of an interior car soundscape. However, this result can also be discussed with respect to technological considerations. Ambisonic is know to introduce a localization blur, the differences of position of the two point sources may have been hard to discriminate by participants.

Finally, in contrast to Experiment 1, the temporal approach (i.e. C_2) has been preferred to the frequency approach (C_1). This result can be explained by the fact that the dynamic feedback synthesized in Experiment 2 has a more discrete spectral density than a traditional engine noise (i.e. frequency components are farther away from each other). Unlike the engine noise composed of a multitude of sources, the proposed auditory feedback is a single source and its spatial scattering distributed over frequency bands in the frequency approach may be perceived as separated auditory streams, leading to a less coherent soundscape. Furthermore, we found that the difference of coherency between C_1 and C_2 is more important in acceleration scenes and absent in the constant speed scene (c.f. Figure 8). This result indicates that, in case of high dynamic scenario, the coherency could be even more degraded due to the perception of moving sources when each separated stream moves from one frequency band to another.

6 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. The spatial configurations should be tested on other vehicle types. The impact of interaction (i.e. in a driving situation) with the vehicle should be investigated. It could validate the virtual environment and our methodology to investigate sound design integration in car interior environment. The virtual environment can be further augmented with vibration modulations. The improvement of the physical reconstruction of the environment can benefit the ecological validity of the experimental setup. The investigation of the impact of multimodal context on the human perception and behaviour could be one of the major challenges for designers and researchers in the domain of VR/AR and immersive technologies.

Concerning the dynamic auditory feedback, it would be interesting to investigate the influence of more diverse sounds and their impact on the perception of the vehicle to give more control to sound designers. More generally, this study opens new perspectives on designing the vehicle soundscape. With the development of autonomous vehicle, the car environment will become more interactive and it is crucial to ensure a good integration of all new sources within the existing environment.

7 ACKNOWLEDGMENT

The authors would like to thank François Kolaczek and Emmanuelle Diaz from the Automotive Research and Advanced Engineering department at Stellantis for his technical help during the recording sessions and for her advice on the experimental protocol and the statistical analysis respectively. They also would like to thank Marcelo Caetano from CNRS for his help configuring the additive model.

8 REFERENCES

[1] F. Doleschal and J. L. Verhey, "Pleasantness and magnitude of tonal content of electric vehicle interior sounds containing subharmonics," *Applied Acoustics*, vol. 185, p. 108442 (2022 Jan.), doi:https://doi.org/10.1016/j.apacoust.2021.108442.

[2] M. Münder and C.-C. Carbon, "Howl, whirr, and whistle: The perception of electric powertrain noise and its importance for perceived quality in electrified vehicles," *Applied Acoustics*, vol. 185, p. 108412 (2022 Jan.), doi:https://doi.org/10.1016/j.apacoust.2021.108412.

[3] G. Goetchius, "Leading the charge—the future of electric vehicle noise control," *Sound & Vibration*, vol. 45, no. 4, pp. 5–8 (2011 Apr.).

[4] S. Denjean, V. Roussarie, R. Kronland-Martinet, J.-L. Velay, *et al.*, "How does interior car noise alter driver's perception of motion? Multisensory integration in speed perception," presented at the *Acoustics 2012* (2012 Apr.).

[5] J.-F. Sciabica, M.-C. Bezat, V. Roussarie, R. Kronland-Martinet, and S. Ystad, "Towards the timbre modeling of interior car sound," presented at the *15th International Conference on Auditory Display* (2009 May).

[6] V. Roussarie, F. Richard, and M. Bezat, "Perceptive qualification of engine sound character; validation of auditory attributes using analysis-synthesis method," *Proceedings of the CFA/DAGA* (2004 Mar.).

[7] S. Denjean, *Sonification des véhicules électriques par illusions auditives: étude de l'intégration audiovisuelle de la perception du mouvement automobile en simulateur de conduite*, Ph.D. thesis, Aix-Marseille Université (2015).

[8] S. Denjean, J.-L. Velay, R. Kronland-Martinet, V. Roussarie, J.-F. Sciabica, and S. Ystad, "Are electric and hybrid vehicles too quiet for drivers?" presented at the *Inter-Noise 2013*, pp. 3081–3090 (2013 Sep.).

[9] M. Bodden, "Principles of Active Sound Design for electric vehicles," presented at the *INTER-NOISE and NOISE-CON congress*, vol. 253, pp. 7700–7704 (2016 Aug.).

[10] R. Schirmacher, "Active design of automotive engine sound," presented at the *Audio Engineering Society Convention 112* (2002 May).

[11] D. Swart, A. Bekker, and J. Bienert, "The subjective dimensions of sound quality of standard production electric vehicles," *Applied Acoustics*, vol. 129, pp. 354–364 (2018 Jan.), doi:https://doi.org/10.1016/j.apacoust.2017.08.012.

[12] 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 Nov.), doi:https://doi.org/10.1121/1.4898742.

[13] Y. Cao, H. Hou, Y. Liu, L. Tang, and Y. Li, "Engine Order Sound Simulation by Active Sound Generation for Electric Vehicles," *SAE International Journal of Vehicle Dynamics, Stability, and NVH*, vol. 4, no. 10-04-02-0011, pp. 151–164 (2020 Feb.), doi:https://doi.org/10.4271/10-04-02-0011.

[14] F. Doleschal, H. Rottengruber, and J. L. Verhey, "Influence parameters on the perceived magnitude of tonal content of electric vehicle interior sounds," *Applied Acoustics*, vol. 181, p. 108155 (2021 Oct.), doi:https://doi.org/10.1016/j.apacoust.2021.108155.

[15] M. Maunder, "Experiences Tuning on Augmented Power Unit Sound Sytem for Both Interior and Exterior of an Electric Car," *SAE Technical Paper 2018-01-1489* (2018 Jun.), doi:https://doi.org/10.4271/2018-01-1489.

[16] K.-J. Chang, G. Cho, W. Song, M.-J. Kim, C. W. Ahn, and M. Song, "Personalized EV Driving Sound Design Based on the Driver's Total Emotion Recognition," Tech. rep. (2022 Jun.), doi:https://doi.org/10.4271/2022-01-0972.

[17] R. Schramm, J. de Kruiff, R. Doerfler, J. Merkt, P. Kampmann, and F. Walter, "AI in Automotive Audio: Approaching Dynamic Driving Sound Design," presented at the *Audio Engineering Society Conference: AES 2022 International Automotive Audio Conference* (2022 Jun.).

[18] S. Denjean, R. Kronland-Martinet, V. Roussarie, and S. Ystad, "Zero-emission vehicles sonification strategy based on shepard-risset glissando," presented at the *International Symposium on Computer Music Multidisciplinary Research*, pp. 709–724 (2019 Mar.), doi:https://doi.org/10.1007/978-3-030-70210-6_46.

[19] A. Neidhardt, C. Schneiderwind, and F. Klein, "Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms-Literature Review and Theoretical Framework," *Trends in Hearing*, vol. 26, p. 23312165221092919 (2022 May), doi:https://doi.org/10.1177/23312165221092919.

[20] C. Kuhn-Rahloff, *Realitätstreue, Natürlichkeit, Plausibilität: Perzeptive Beurteilungen in der Elektroakustik*, Ph.D. thesis, TU Berlin, Germany (2012).

[21] Y. Cao, H. Hou, Y. Liu, Y. Li, S. Wang, H. Li, *et al.*, "Sound Pressure Level Control Methods for Electric Vehicle Active Sound Design," *SAE International Journal of Vehicle Dynamics, Stability, and NVH*, vol. 5, no. 10-05-02-0014, pp. 205–226 (2021 Mar.), doi:https://doi.org/10.4271/10-05-02-0014.

[22] A. Lindau and S. Weinzierl, "Assessing the plausibility of virtual acoustic environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810 (2012), doi:https://doi.org/10.3813/AAA.918562.

[23] S. J. Schlecht and E. A. Habets, "Spatial audio quality evaluation: comparing transaural, ambisonics and stereo," in *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 53–59 (Montréal, Canada) (2007 Jun.).

[24] C. Tarlao, J. Steffens, and C. Guastavino, "Investigating contextual influences on urban soundscape evaluations with structural equation modeling," *Building and Environment*, vol. 188, p. 107490 (2021 Jan.), doi:https://doi.org/10.1016/j.buildenv.2020.107490.

[25] C. Tarlao, D. Steele, and C. Guastavino, "Assessing the ecological validity of soundscape reproduction in different laboratory settings," *Plos one*, vol. 17, no. 6, p. e0270401 (2022 Jun.), doi:https://doi.org/10.1371/journal.pone.0270401.

[26] C. Guastavino, "Validité écologique des dispositifs expérimentaux," *D. Dubois (éd.), Le Sentir et le Dire. Concepts et méthodes en psychologie et linguistique cognitives, Paris, L'Harmattan (Coll. Sciences Cognitives)* (2009).

[27] E. Brunswik, *Perception and the representative design of psychological experiments* (Univ of California Press, 1956), doi:https://doi.org/10.1525/9780520350519.

[28] J. J. Gibson, *The ecological approach to visual perception: classic edition* (Psychology press, 2014 Dec.), doi:https://doi.org/10.4324/9781315740218.

[29] W. W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecological psychology*, vol. 5, no. 4, pp. 285–313 (1993), doi:https://doi.org/10.1207/s15326969eco0504_2.

[30] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29 (1993), doi:https://doi.org/10.1207/s15326969eco0501_1.

[31] T. Dupré, S. Denjean, M. Aramaki, and R. Kronland-Martinet, "Spatial Sound Design in a Car Cockpit: Challenges and Perspectives," presented at the *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–5 (2021 Sep.), doi:https://doi.org/10.1109/I3DA48870.2021.9610910.

[32] J. Y. Hong, J. He, B. Lam, R. Gupta, and W.-S. Gan, "Spatial audio for soundscape design: Recording and reproduction," *Applied sciences*, vol. 7, no. 6, p. 627 (2017 Jun.), doi:https://doi.org/10.3390/app7060627.

[33] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 6, pp. 603–616 (1997 Dec.), doi:https://doi.org/10.1162/pres.1997.6.6.603.

[34] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, and A. Kobayashi, "Evaluating the importance of multisensory input on memory and the sense of presence in virtual environments," presented at the *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pp. 222–228 (1999 Mar.), doi:https://doi.org/10.1109/VR.1999.756955.

[35] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa, "Do multisensory stimuli benefit the virtual reality experience? A systematic review," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1428–1442 (2020 Jul.), doi:https://doi.org/10.1109/TVCG.2020.3010088.

[36] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical Microphone Array Beamforming," in I. Cohen, J. Benesty, and S. Gannot (Eds.), *Speech Processing in Modern Communication: Challenges and Perspectives*, chap. 11, pp. 281–305 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), doi:https://doi.org/10.1007/978-3-642-11130-3_11.

[37] J. Daniel and S. Moreau, "Further Study of Sound Field Coding with Higher Order Ambisonics," in *Audio Engineering Society Convention 116* (2004 May), URL `http://www.aes.org/e-lib/browse.cfm?elib=12789`.

[38] JATO, "H1 2022: Europe by Segments," https://www.jato.com/h1-2022-europe-by-segments/ (accessed Jan. 9, 2023).

[39] C. Guastavino and B. F. Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1105–1115 (2004 Aug.), doi:https://doi.org/10.1121/1.1763973.

[40] G. Desoeuvre, F. Richard, V. Roussarie, and M. C. Bezat, "Hartis, a re-synthesis tool for vehicles sound design," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3247 (2008), doi:https://doi.org/10.1121/1.2933514.

[41] T. Carpentier, "A new implementation of Spat in Max," presented at the *15th Sound and Music Computing Conference (SMC2018)*, pp. 184–191 (2018 Jul.).

[42] B. Rafaely, *Fundamentals of spherical array processing*, vol. 8 (Springer Berlin, Heidelberg, 2015 Feb.), doi:https://doi.org/10.1007/978-3-662-45664-4.

[43] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 158–170 (2009 Jun.), doi:https://doi.org/10.1109/TASL.2009.2024731.

[44] C. Colangeli, J. Lanslots, S. Paillasseur, K. Janssens, and L. Lamotte, "Sound Source Localization Analysis in the Combustion Cycle of ICE Powertrains," presented at the *DAGA* (2018 Mar.).

[45] M. V. van der Seijs, D. de Klerk, and D. J. Rixen, "General framework for transfer path analysis: History, theory and classification of techniques," *Mechanical Systems and Signal Processing*, vol. 68, pp. 217–244 (2016 Feb.), doi:https://doi.org/10.1016/j.ymssp.2015.08.004.

[46] R. Mason, T. Brookes, and F. Rumsey, "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," *The Journal of the Acoustical Society of America*, vol. 117, no. 3, pp. 1337–1350 (2005 Mar.), doi:https://doi.org/10.1121/1.1853113.

[47] T. Carpentier, "Spat˜: a comprehensive toolbox for sound spatialization in Max," *Ideas Sonicas*, vol. 13, no. 24, pp. 12 – 23 (2021), URL `https://hal.science/hal-03356292`.

[48] A. Favrot and C. Faller, "Complementary N-band IIR filterbank based on 2-band complementary filters," *Proc. Intl. Works. on Acoust. Echo and Noise Control (IWAENC)* (2010).

[49] 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 Aug.), doi:https://doi.org/10.17743/jaes.2014.0031.

[50] 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 Sep.).

[51] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Mu-*

sic Journal, vol. 19, no. 4, pp. 71–87 (1995), doi:https://doi.org/10.2307/3680992.

[52] M. Caetano and P. Depalle, "On the Estimation of Sinusoidal Parameters Via Parabolic Interpolation of Scaled Magnitude Spectra," presented at the *2021 24th International Conference on Digital Audio Effects (DAFx)*, pp. 81–88 (2021 May), doi:https://doi.org/10.23919/DAFx51585.2021.9768277.

[53] R. N. Shepard, "Circularity in judgments of relative pitch," *The journal of the acoustical society of America*, vol. 36, no. 12, pp. 2346–2353 (1964 Dec.), doi:http://dx.doi.org/10.1121/1.1919362.

THE AUTHORS

Théophile Dupré

Sébastien Denjean

Mitsuko Aramaki

Richard Kronland-Martinet

Théophile Dupré received the french engineer's degree in electrical engineering and signal processing from INSA Lyon in 2018 and the M.S. degree in acoustic, signal processing and computer science applied to music from IRCAM in Paris in 2019. He is currently a PhD fellow in PRISM laboratory in Marseille and in Stellantis Automotive Research and Advanced Engineering department in Paris. His research interests are in spatial audio and sound design in augmented and virtual reality with a focus on electric vehicle applications.

●

Sebastien Denjean earned his M.S. degree from the Ecole Centrale Marseille, France, in 2010 and his Ph.D. degree in acoustics from the Aix-Marseille University, France, in 2015 for his work on in-car sonification for electric cars. He is currently working at Stellantis in the Automotive Research and Advanced Engineering department, on multimodal interfaces and sound enhancement.

●

Mitsuko Aramaki received the Ph.D. degree from Aix-Marseille University, Marseille, France, in 2003, for her work on analysis and synthesis of impact sounds using physical and perceptual approaches. She is currently Director of Research at the National Center for Scientific Research (CNRS). Since 2017, she is head of the "Perception Engineering" team at the laboratory PRISM "Perception, Representations, Image, Sound, Music". Her research mainly focuses on sound modeling, perceptual and cognitive aspects of timbre, and multimodal interactions in the context of virtual/augmented reality.

●

Richard Kronland-Martinet has a background in theoretical physics and acoustics. He got the "Doctorat d'Etat ès Sciences" degree (habilitation) in 1989 from Aix Marseille University, France, for his pioneer work on analysis and synthesis of sounds using time-frequency and time-scale (wavelets) representations. He is currently Director of Research at the National Center for Scientific Research (CNRS), and head of the Interdisciplinary laboratory PRISM (Perception, Representations, Image, Sound, Music). His primary research interests are in analysis and synthesis of sounds with a particular emphasis on high-level control of synthesis processes. He published more than 250 journal articles and conference proceedings in this domain. He recently addressed new scientific challenges linked to the semantic description of sounds and to their synthesis control based on sound invariants, using an interdisciplinary approach associating signal processing, physics, perception and cognition.