






# Analysis by synthesis of engine sounds for the design of dynamic auditory feedback of electric vehicles

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

<sup>1</sup> Aix-Marseille Univ, CNRS, PRISM, 31 Chemin J. Aiguier, 13402 Marseille Cedex 20, France

<sup>2</sup> Technology Office, Stellantis, Route de Gisy, 78140 Vélizy-Villacoublay, France

Received 16 December 2022, Accepted 19 June 2023

**Abstract** – In traditional combustion engine vehicles, the sound of the engine plays an important role in enhancing the driver's experience of the vehicle's dynamics, and contributes to both comfort and safety. However, with the development of quieter electric vehicles, drivers no longer receive this important auditory feedback, and this can lead to a less satisfying acoustic environment in the vehicle cabin. To address this issue, sonification strategies have been developed for electric vehicles to provide similar auditory feedback to the driver, but feedback from users has suggested that the sounds produced by these strategies do not blend seamlessly with the other sounds in the vehicle cabin. This study focuses on identifying the key acoustic parameters that create a sense of cohesion between the synthetic sounds and the vehicle's natural soundscape, based on the characteristics of traditional combustion engine vehicles. Through analyzing the time and frequency of the noises produced by combustion engine vehicles, the presence of micro-modulations in both frequency and amplitude was identified, as well as resonances caused by the transfer of sound between the engine and the cabin. These parameters were incorporated into a synthesis model for the sonification of electric vehicle dynamics, based on the Shepard-Risset illusion. A perceptual test was conducted, and the results showed that the inclusion of resonances in the synthesized sounds significantly enhanced their naturalness, while micro-modulations had no significant impact.

**Keywords:** Active sound design, Automotive, Analysis, Synthesis

## 1 Introduction

The automotive industry is undergoing a major transformation from Internal Combustion Engine Vehicles (ICEV) to Battery Electric Vehicles (BEV). This transition is not only altering the dynamic behavior of vehicles but also impacting the user experience by changing the acoustic environment or soundscape [1, 2]. This is due to the different powertrain-generated sounds produced by BEV, which can lead to an altered soundscape even if the environment is quieter. The absence of the previously predominant motor sound can result in unwanted noises becoming more noticeable [3], while also depriving the driver of important information about the vehicle's dynamics [4] and characteristics [5, 6]. This has a significant impact on the overall driving experience and has prompted car manufacturers to seek new solutions to address these challenges.

For a 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 [7]. 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 [8]. The same principle has been proposed in BEV [9]. Subharmonic generation is used to create a machine-like sound [10, 11]. 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 [12]. Maunder proposed to capture the electric motor vibration with an accelerometer and enhanced, tuned and re-played it in real-time in the cabin [13]. Adaptive design has also been proposed to adapt the auditory feedback timbre depending on driver's emotion [14] and driving style [15]. 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 gives the impression of pitch variation without variation of spectral content [16].

The current focus of active sound design has been on designing the sound itself and ensuring that it accurately represents vehicle dynamics. However, some users have

\*Corresponding author: [dupre@prism.cnrs.fr](mailto:dupre@prism.cnrs.fr)

reported a lack of integration with their surroundings and other perceptual cues. The integration of sound in the environment has not been thoroughly studied except for loudness considerations [17]. To address this issue, the cockpit environment can be considered as an augmented reality environment, where virtual auditory sources must be seamlessly integrated into the real environment to be accepted by users. Neidhardt et al. have explained that the acoustical properties of the virtual element must match those of the real environment [18] and the internal reference developed by people from their everyday listening experience [19]. The environment may also require specific auditory characteristics for sources such as loudness, timbre, width, or location. For instance, in the case of a BEV interior soundscape, Cao et al. have studied the dynamic auditory feedback loudness based on the expected loudness of the engine in ICEV [17]. They found that integrating dynamic feedback with the same loudness variation as the engine in ICEV can improve pleasantness. However, it was compared with no feedback only. Furthermore, we wonder if the specific configuration of the car cabin may require certain timbre characteristics of the virtual source to be consistently integrated. It is also worth noting that the substitution of the engine noise by active sound design may lead to certain expectations by the users.

The aim of this study is to explore how to effectively integrate dynamic auditory feedback into the interior soundscape of electric vehicles, with a specific focus on timbre aspects. To remove the influence of the spatial aspect of the acoustic environment, it was chosen to study monophonic sounds. The virtual source in question is an auditory feedback system that provides information on the vehicle's dynamic, intended to be comparable to the engine sound in ICEV. The integration will be considered consistent if the virtual source blends well with the environment and if the overall acoustic space meets users' expectations. Users' expectations can be influenced by their driving experience in ICEV soundscape generally, as well as by the characteristics of the engine sound specifically. Using an analysis/synthesis approach, relevant timbre-related characteristics of the engine sound will be identified and matched to the virtual source to achieve the integration of dynamic feedback in BEV.

To achieve this goal, in-car acoustic scenes and vehicle dynamic parameters  $V(t)$  were recorded during controlled driving scenarios in ICEV (cf. Sect. 2.1). These recordings were then decomposed in the time-frequency domain to identify and model potentially relevant characteristics of the engine sound (cf. Sects. 2.2–2.4). By extracting the corresponding model parameters  $\Theta$ , the measurements were re-synthesized with the vehicle dynamic data, and auditory feedback for BEV [16] were designed based on relevant ICEV sound features (cf. Sect. 3). Finally, the re-synthesis and auditory feedback models based on  $\Theta$  were evaluated through a listening test to assess their perceptual impact on reproduced scenes in terms of realism in ICEV and naturalness in BEV (cf. Sect. 4). A summary of the methodology used in the study is illustrated on Figure 1.

## 2 Engine sound analysis

### 2.1 In-situ recordings

To analyze the characteristics of the engine sound and the cockpit acoustic environment, binaural scenes were recorded with the Head Acoustics artificial head HMS-IV in different moving vehicles. For technical reasons, the artificial head was placed on the passenger seat and a person was driving the vehicle. Binaural signals are reduced to a monophonic signal by averaging the left and right channels. Synchronously, dynamic parameters of the vehicle noted  $V(t) = [v(t), dv(t)/dt, \omega(t)]$ , with the vehicle speed  $v(t)$  in km/h, the acceleration  $dv(t)/t$  in  $m/s^{-2}$  and the engine speed  $\omega(t)$  in  $min^{-1}$ , are recorded from the vehicle sensors.

The following dynamic scenarios have been recorded on a closed flat asphalted road: accelerations, decelerations without braking and various constant speed scenes. The measurements were repeated in two mid-range compact vehicles.

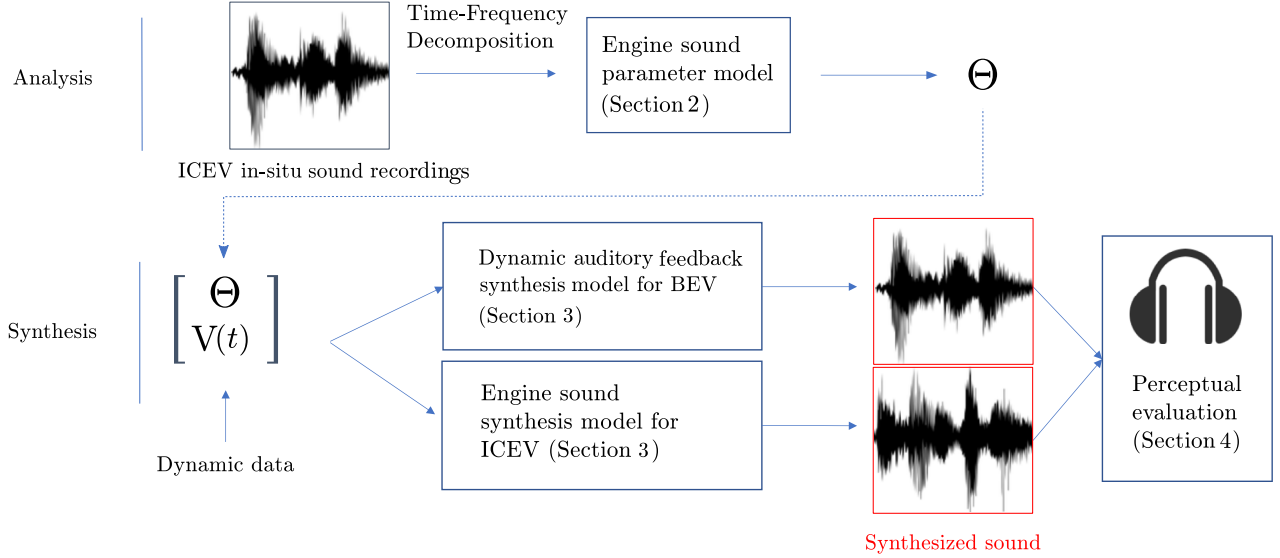
Traditional ICEV acoustic environment is composed of three main acoustic sources. At low and medium speeds, engine sound and low frequency tire-road contacts noise are predominant. At higher speed, wide-band aerodynamic noise tends to mask engine sound [21]. This paper will mainly focus on the analysis of the engine sound.

### 2.2 Harmonic analysis

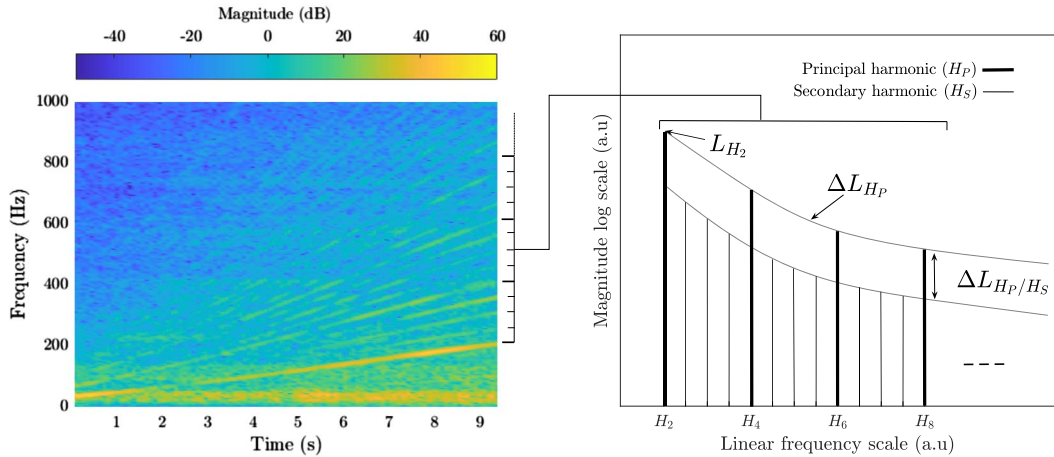
Engine sound is a harmonic sound that results from the periodic combustion process that takes place inside the engine cylinders. The frequency of these harmonics called partials is determined by the engine speed  $\omega$  and the number of cylinders. In a four-cylinder engine, for example, each cylinder undergoes combustion every two rotations, resulting in the slowest periodic process that gives rise to the fundamental partial of the sound. By convention, the harmonic corresponding to one engine rotation is denoted as  $H_1$ , hence, the fundamental partial is denoted as  $H_{0.5}$  because it has a frequency twice lower.  $f_{0.5}$  is the frequency of  $H_{0.5}$  and the fundamental frequency. It is related to the engine speed  $\omega$  as follows:

$$f_{0.5}(\omega) = 0.5 \frac{\omega}{60}. \quad (1)$$

Figure 2 (left) illustrates a spectrogram of a four-cylinder engine sound in acceleration. The partials are noted  $H_n$  with  $n = \{0.5, 1, 1.5, \dots, N\}$  and  $N$  the last partial considered.  $H_{0.5}$  is not visible on Figure 2 because it is masked by road/tire noise. The first visible partial is  $H_2$ , the fourth harmonic of  $H_{0.5}$ , but also the fundamental partial of a second periodic process: the repetition of combustions from one cylinder to the next. Partial at multiple frequencies of  $H_2$  are called principal harmonics noted  $H_p$  because they are more present at low engine speed. Partial at multiple frequencies of  $H_{0.5}$  apart from  $H_p$  are called secondary harmonics noted  $H_s$ . The magnitude of each partial  $H_n$  depends on the engine speed  $\omega$ .



**Figure 1.** Diagram of the analysis/synthesis approach used in this study (ICEV: Internal Combustion Engine Vehicle, BEV: Battery Electric Vehicle).



**Figure 2.** Left: Spectrogram of a four-cylinder compact vehicle in acceleration in second gear. The measurement is a monophonic reduction of a binaural recording at the passenger viewpoint. Right: Schematic representation of engine harmonic relative magnitude at a given engine speed adapted from [20].

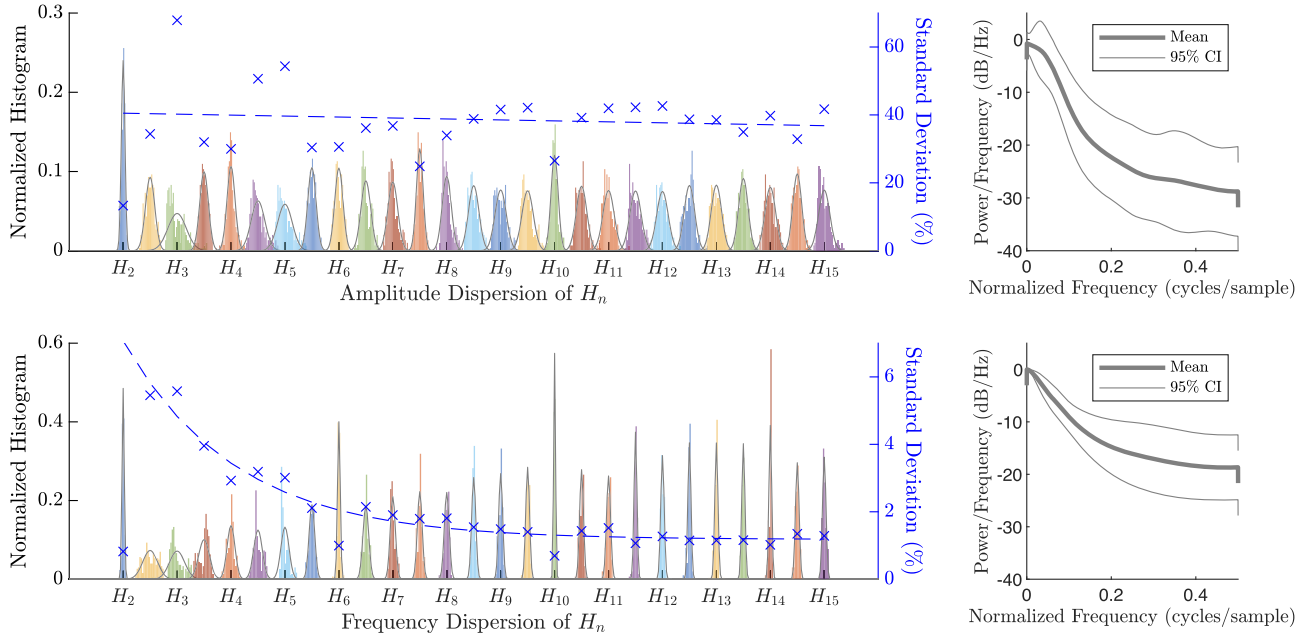
Sciabica proposed an empirical model of partial magnitude variations in acceleration [20] as illustrated on Figure 2 (right). The model is composed of three parameters:

- $L_{H_2}$  expressed in dB/oct represents the magnitude variation of  $H_2$  with respect to  $\omega$ ,  $(L_{H_2})_{\omega_0}$  being the magnitude of  $H_2$  in dB at the minimum engine speed  $\omega_0$ .  $L_{H_2}$  has an influence on the engine presence;
- $\Delta L_{H_p/H_s}$  expressed in dB/oct corresponds to the attenuation of  $H_s$  relative to  $H_p$  with respect to  $\omega$ ,  $(\Delta L_{H_p/H_s})_{\omega_0}$  being the attenuation in dB at the minimum engine speed  $\omega_0$ .  $\Delta L_{H_p/H_s}$  has an impact on engine roughness;
- $\Delta L_{H_p}$  expressed in dB/oct is the attenuation of  $H_p$  regardless of  $\omega$ . For example, if  $\Delta L_{H_p} = -3$  dB/oct and  $L_{H_2} = 70$  dB,  $L_{H_6} = 64$  dB because  $H_2$  and  $H_6$

are separated by 2 octaves, the attenuation of  $-6$  dB holds for any engine speed as  $\Delta L_{H_p}$  is independent of  $\omega$ .  $\Delta L_{H_p}$  has an influence on engine brightness.

### 2.3 Modulations

In the spectrogram of Figure 2 (left), amplitude modulations are present on each partial  $H_n$  in the engine sound. From the Short Time Fourier Transform (STFT) of an engine sound measurement in the cabin, the instantaneous amplitude and frequency of each partial  $H_n$  are extracted at each time frame. Several algorithms have been proposed to extract sinusoidal components from a harmonic plus noise signal [22–25] to name a few. Here, a closed-source software



**Figure 3.** Top left: Amplitude dispersion of  $H_n$  from its mean expressed in percent starting at  $H_2$  and fitted Gaussian distributions extracted from a constant speed measurement of 9.3 s (corresponding to 937 time frames). Blue crosses indicate the standard deviation of the corresponding amplitude dispersion. Top right: Mean and 95% confidence interval of the power spectral densities of all amplitude dispersions estimated with autoregressive models of order 20. Bottom left: Frequency dispersion of  $H_n$  from its expected value deduced from engine speed expressed in percent starting at  $H_2$  and fitted Gaussian distributions extracted from a constant speed measurement of 9.3 s (corresponding to 937 time frames). Blue crosses indicate the standard deviation of the corresponding frequency dispersion. Bottom right: Mean and 95% confidence interval of the power spectral densities of all frequency dispersions estimated with autoregressive models of order 20. Data are from a constant speed measurement in a compact urban vehicle.

called Additive developed by IRCAM based on [22, 26] was used. The STFT was calculated with the following parameters: 20 ms Blackmann window, 10 ms hop size and a Fast Fourier Transform (FFT) length of 8192 bins at 44,100Hz. A rather long time window is necessary to have a better resolution in the frequency domain and discriminate each partial.

Figure 3 illustrates the amplitude (top) and frequency (bottom) dispersion histograms of each partial  $H_n$ . The histograms are fitted with Gaussian distributions. The data have been extracted from a constant speed measurement of 9.3 s on a compact urban vehicle (corresponding to 937 time frames). Amplitude dispersion of partial  $H_n$  noted  $A_n$  is the amplitude deviation in percent from its mean. Based on the fitted distributions illustrated on Figure 3 (top left), the dispersion is assumed to be normally distributed and the standard deviation constant along all partials.

The amplitude modulations  $A_n$  are then defined as:

$$A_n \sim \mathcal{N}(0, \sigma_A^2), \quad (2)$$

with  $\sigma_A$  the standard deviation for all partials  $H_n$ . The power spectral density for all partials, estimated with an autoregressive model of order 20, presents a low-pass profile with cut-off frequency noted  $f_A$  (cf. Fig. 3, top right). The normalized cut-off frequency must be interpreted with the STFT sampling frequency  $f_{\text{STFT}} = 100$  Hz. The correlation between dispersions noted  $r_A$  was computed to account for correlated modulations that could be perceived

In the measure used as example on Figure 3,  $\sigma_A = 40\%$  except for  $H_2$ ,  $f_A = 0.1$  cycles/sample (10 Hz) and  $r_A = 0.1$ . Therefore, the amplitude varies by  $\pm 3$  dB, relatively slowly and the modulations between partials are weakly correlated.

Frequency dispersion of  $H_n$  noted  $F_n$  is the deviation in percent to its expected value deduced from the engine speed ( $f_n = 2nf_{0.5} = n\omega/60$ ). Based on the fitted distributions illustrated on Figure 3 (bottom left), the dispersion is also normally distributed with a low-pass power spectral density of cut-off frequency noted  $f_F$  (cf. bottom right). The means are always null meaning that the engine sound is purely harmonic but the standard deviations decrease exponentially with partial order  $n$ . The frequency modulations  $F_n$  is then noted:

$$F_n \sim \mathcal{N}(0, \sigma_F^2(n)), \quad (3)$$

with  $\sigma_F(n)$  the standard deviation depending on  $n$ . In the example of Figure 3,  $6\% > \sigma_F(n) > 1.5\%$  except for  $H_2$ ,  $f_F = 0.1$  cycles/sample (10 Hz) and the correlation  $r_f$  between partial dispersions is 0.3. A deviation of 6% corresponds to a semitone so the frequency modulations should be audible and relatively slow. Modulations between partials are more correlated than amplitude modulations.

## 2.4 Formants

The spectrogram on Figure 2 exhibits static resonances at 40 Hz, around 200 Hz and 400 Hz, at 550 Hz and 750 Hz.

The previous empirical engine model does not account for these resonances that are characteristics of the car structure. From a signal point of view, the vibroacoustic transfer from the powertrain to the cabin and the diffusion inside the cabin filters the engine sound (i.e. the source) as in a source/filter model. Sciabica proposed to assimilate these resonances to speech formants and to identify them by vocal imitation [20]. He showed that the formants impact the perception of the vehicle especially in dynamic situations where the energy moves from one formant to another and alters the perceived timbre. It means that formants could be key parameters to integrate interior car sounds. It is hypothesized in this study that a significant role is played by these formants in designing a realistic engine sound inside the cabin and that these formants must be complied with by any virtual source that should be integrated consistently in BEV.

### 3 Engine sound and dynamic feedback synthesis

To synthesize engine sound in ICEV and dynamic feedback in BEV, an additive signal model is used. The sound  $s(t)$  is composed of a sum of  $M$  sinusoidal components (i.e. partials) at varying frequencies  $f_m(t)$  and amplitudes  $a_m(t)$ :

$$s(t) = \sum_{m=1}^M a_m(t) \cos \Phi_m(t), \quad (4)$$

with

$$\Phi_m(t) = \Phi_m(t-1) + \frac{2\pi}{f_s} f_m(t), \quad (5)$$

where  $f_s$  is the sampling frequency,  $\phi_m(0) = 2\pi u_m$  is the initial phase value and  $u_m$  is a random uniform draw from 0 to 1. The engine sound and the dynamic feedback sound differ by the number of partials  $M$  and their associated amplitude and frequency coefficients  $a_m(t)$  and  $f_m(t)$ , which are parameterized by vehicle dynamic data  $V(t)$  (cf. Sects. 1 and 2.1).

#### 3.1 Engine sound

To synthesize the measures, or synthesize new engine sounds, only the engine speed  $\omega(t)$  is required as input. From equation (1), the frequency  $f_n(t)$  of each partial  $n \in [0.5, 1, 1.5, 2, \dots, N]$  is deduced:

$$f_n(t) = \frac{n}{60} \omega(t). \quad (6)$$

The amplitude  $a_n(t)$  in decibels is given by the following relationships:

$$(a_n(t))_{\text{dB}} = L_{H_2} \frac{\omega(t)}{\omega_0} + (L_{H_2})_{\omega_0} + \Delta L_{H_p} \frac{f_n(t)}{f_2(t)}, \quad (7)$$

if  $n \in [2, 4, 6, \dots, N]$  (i.e. principal harmonics), and

$$(a_n(t))_{\text{dB}} = \Delta L_{H_p} \frac{f_n(t)}{f_2(t)} + (L_{H_2} + \Delta L_{H_p/H_s}) \frac{\omega(t)}{\omega_0} + (L_{H_2})_{\omega_0} + (\Delta L_{H_p/H_s})_{\omega_0}, \quad (8)$$

if  $n \in [2.5, 3, 3.5, 4.5, 5, \dots, N - 0.5]$  (i.e. secondary harmonics). The amplitude of the first harmonics  $a_{0.5}(t)$ ,  $a_1(t)$  and  $a_{1.5}(t)$ , not accounted in the model of Sciabica [20] and masked on Figure 2, is set to  $-15$  dB compared to  $a_2(t)$ . This value is based on measurements with less road/tyre noise at low frequency where their estimation is possible.

#### 3.2 Dynamic auditory feedback

The dynamic auditory feedback sound for BEV is based on the Shepard-Risset tone which uses the circularity in pitch perception to give the auditory illusion of a forever ascending or descending tone [27]. Figure 4 illustrates the Shepard Risset tone as applied to the dynamic feedback design. The tone is composed of sinusoidal components (i.e. partials) each separated by an octave forming an harmonic comb, swept at a given speed  $v_s$ . The amplitude of each partial is determined by a raised cosine function that covers the desired frequency range (i.e. number of octaves  $L$ ). This illusory infinite ascending/descending tone enables to represent accelerations with significant pitch variations for an unlimited range of speeds [16].

We note  $f_n$  the frequency of partial  $n$ , the following relation gives the amplitude  $a_n$  of partial  $n$ :

$$a_n(f_n) = \begin{cases} \frac{1}{2} \left[ 1 - \cos \left( 2\pi \frac{\log f_n - \log \frac{F_c}{2^{L/2}}}{L \log 2} \right) \right] & \text{if } \frac{F_c}{2^{L/2}} < f_n < 2^{L/2} F_c, \\ 0 & \text{else} \end{cases}, \quad (9)$$

where  $F_c$  is the central frequency of the spectral window and  $L$  the window width in octaves (i.e. the desired frequency range). When one partial goes outside the frequency range, another is generated at the other end of the spectrum.

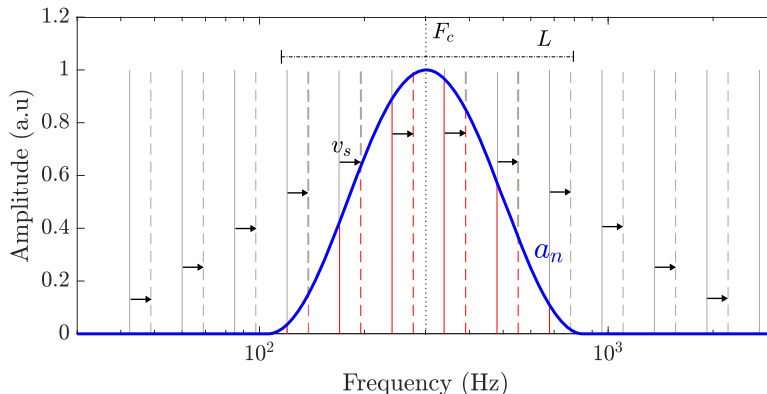
To convey vehicle dynamic information, the sweep speed  $v_s(t)$  is mapped to the vehicle speed  $v(t)$  and acceleration  $dv(t)/dt$ :

$$v_s(t) = \begin{cases} \frac{1}{2\sqrt{v(t)}} \frac{dv(t)}{dt} & \text{if } v > 1, \\ 0 & \text{else.} \end{cases} \quad (10)$$

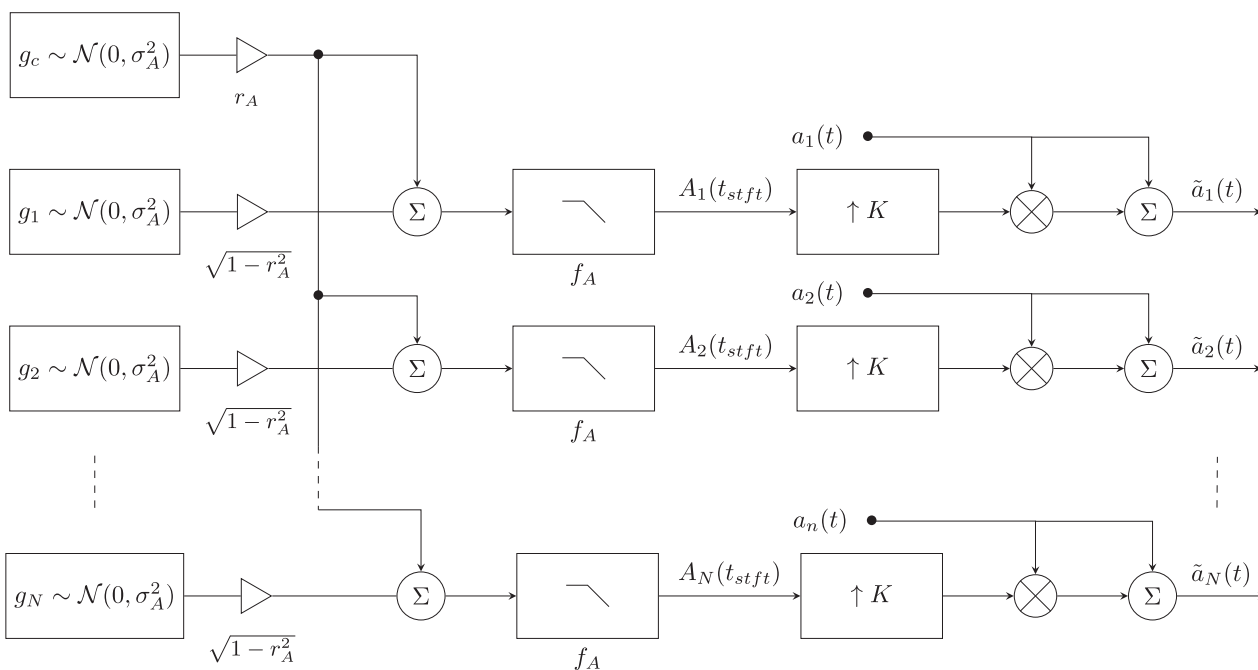
The function  $v_s(t)$  is constrained to be zero for speeds lower than 1 km/h in order to prevent it from diverging to infinity. The central frequency of the window  $F_c(t)$  is mapped to the vehicle speed:

$$F_c(t) = F_c^{\min} \left( \frac{F_c^{\max}}{F_c^{\min}} \right)^{v(t)/v_{\max}}, \quad (11)$$

where  $F_c^{\min}$  and  $F_c^{\max}$  are the minimum and maximum central frequency respectively, and  $v_{\max}$  is the maximum vehicle speed. See [28] for more details. Denjean et al. proposed to enrich the spectral content of the sound by



**Figure 4.** Schematic illustration (adapted from [28]) of the Shepard-Risset illusion as applied to the design of dynamic auditory feedback for BEV.  $a_n$  is the shape of the window,  $F_c$  its central frequency,  $L$  its width and  $v_s$  the swept speed of the harmonic comb in red.



**Figure 5.** Signal flow diagram to account for the amplitude modulations described in Section 2.3. The same algorithm is used to process the frequency modulations with the appropriate parameters  $\sigma_F$ ,  $r_F$ ,  $f_F$ .

adding sinusoidal components between the partials to form chords and allow the creation of more varied timbres [16]. The choice of the chords gives the frequency relation between partials. With the frequency  $f_n$  of each partial, the amplitude  $a_n$ , the dynamic feedback is computed based on equations (4) and (5).

### 3.3 Modelling modulation and formants

To account for the modulations described in Section 2.3, it requires to add a stochastic component to the amplitude and frequency values  $a_n(t)$  and  $f_n(t)$ . The algorithm is illustrated on Figure 5. The following paragraph describes the algorithm to generate the modulated amplitude

components  $\tilde{a}_n(t)$ . The same procedure must be applied to generate the modulated frequency components  $\tilde{f}_n(t)$  by replacing the parameters  $\sigma_A$ ,  $r_A$ , and  $f_A$  by  $\sigma_F$ ,  $r_F$ , and  $f_F$ , respectively. To simplify the model of frequency dispersion, the exponential decay of  $\sigma_F(n)$  is ignored, then  $\sigma_F(n) = \text{cst}$  with  $\text{cst}$  being the asymptotic value of dispersion (cf. Fig. 3, bottom left).

The sequences generated by each Gaussian random generator  $g_n$  are weighted by another sequence from the common generator  $g_c$  to take into account the correlation defined by the coefficient  $r_A$  as detailed in [29]. The weighted sequences are then filtered by a low pass filter at cut-off frequency  $f_A$  to match the power spectral density illustrated on Figure 3 (top right). At this stage the

sequences  $A_n(t_{\text{stft}})$  are sampled at the analysis frequency  $f_{\text{stft}}$  with time index  $t_{\text{stft}}$ . Then, the sequences are upsampled by a factor  $K = f_s/f_{\text{stft}}$  to be combined to the the initial amplitude components  $a_n(t)$  as follows:

$$\tilde{a}_n(t) = a_n(t)(1 + A_n(t)). \quad (12)$$

Similarly,  $F_n(t)$  are combined to  $f_n(t)$  to produce the frequency modulations  $\tilde{f}_n(t)$ :

$$\tilde{f}_n(t) = f_n(t)(1 + F_n(t)). \quad (13)$$

To account for the formants, the resonances observed on the measurements are modelled. Formal identification through spectral envelope estimation is difficult because the spectral width of a formant can be similar to the width of a partial, hence difficult to separate. Then, peak filters are manually tuned to match each resonance. All filters are combined to form a unique FIR filter noted  $h$  which is convolved with the synthesized engine sound or dynamic feedback.

## 4 Perceptual evaluation

The synthesis models proposed in Section 3 have been perceptually evaluated in terms of realism for the ICEV engine sounds, and in terms of naturalness for BEV sounds. Due to the COVID-19 pandemic restrictions at the time of the experiment, the listening test has been performed online.

### 4.1 Participants

One hundred and twenty nine participants took part in the listening test. Forty six participants were rejected from the analysis since they did not complete the test or skip a part of the test. Two others participants were rejected since they self-reported to have performed the test on loudspeakers instead of using headphones as requested in the instructions. Participants were mainly part of the Stellantis company or the PRISM laboratory. The following data on participant profile were collected: age, gender, expertise in acoustic and automotive, driving experience.

### 4.2 Stimuli

The measurements described in Section 2.1 were recorded inside two different ICEV, noted  $M_1$  and  $M_2$ , and for two dynamics, i.e. acceleration and deceleration. The accelerations were performed in second gear with full throttle opening. The decelerations were also performed in second gear and without using the brake pedal. An additive analysis was conducted by using the Additive software (cf. Sect. 2.3), leading to the decomposition of the recorded signals into harmonic plus noise signals. The engine sounds of four seconds were synthesized with the model described in Section 3.1 with parameters estimated manually to perceptually match the measurements gathered in Table 1 and  $N = 25$  (i.e. 50 partials). The dynamic profiles corresponded to an engine speed from 1860 to 4115  $\text{min}^{-1}$  and

from 1140 to 3390  $\text{min}^{-1}$  in acceleration for  $M_1$  and  $M_2$  respectively and from 2720 to 2140  $\text{min}^{-1}$  in deceleration for both vehicles. The sounds were then combined with an aerodynamic and road/tyre noise extracted from the measurements (noise part of the additive analysis). Two noise levels were considered: the high level corresponds to the actual noise level in the measurement and the low level is fixed 6 dB lower, making the engine sound more present. Five conditions were evaluated:

- the reference condition (i.e.  $C_0$ ) is the engine model from [20] and presented in Section 3.1;
- the low anchor condition (i.e.  $c_0$ ) is a degraded version of the reference where only the principal harmonics  $H_p$  are synthesized, leading to a presumably non-realistic engine sound. This condition is an anchor as usually included in MUSHRA test [30];
- the condition  $C_1$  is the reference engine sound with modulations described in Section 2.3 (modulation parameters are gathered in Tab. 1);
- the condition  $C_2$  is the reference engine sound with formants described in Section 2.4 and displayed on Figure 6;
- the combined condition  $C_1 + C_2$  is the reference engine sound with modulations and formants.

The BEV sounds are synthesized with the model described in Section 3.2 with  $L = 7$  octaves,  $F_c^{\text{min}} = 60$  Hz,  $F_c^{\text{max}} = 500$  Hz,  $v_{\text{max}} = 130$  km/h and a duration of four seconds. To simulate two different vehicles noted  $E_1$  and  $E_2$ , two different chords were defined to generate the dynamic auditory feedback: a major chord for  $E_1$  creating a consonant timbre and an augmented chord for  $E_2$  creating a more dissonant timbre. Similar to ICEV sounds, two dynamics (acceleration from 30 km/h to 60 km/h and deceleration from 65 km/h to 55 km/h) as well as two noise levels were synthesized. Four conditions were evaluated:

- the condition  $C_0$  is the dynamic auditory feedback presented in Section 3.2;
- the condition  $C_1$  is the dynamic auditory feedback with modulations described in Section 2.3;
- the condition  $C_2$  is the dynamic auditory feedback with formants described in Section 2.4;
- the combined condition (i.e.  $C_1 + C_2$ ) is the dynamic auditory feedback with modulations and formants.

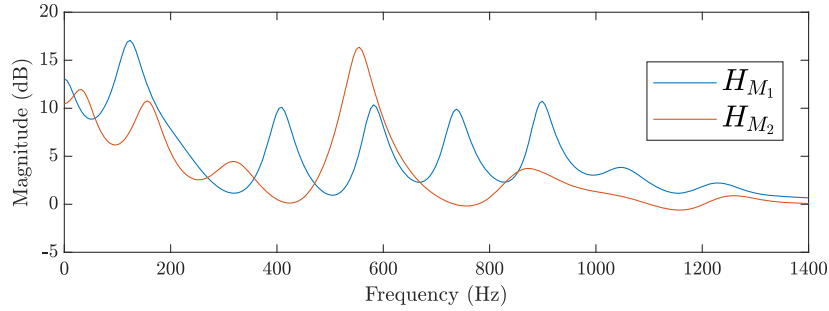
Here, there is no anchor condition because no assumption can be made on a presumably more artificial dynamic feedback. All stimuli have been equalized in loudness with the corresponding reference measurement through informal listening sessions.

### 4.3 Task and procedure

The participants were instructed to use headphones and adjust the volume to a comfortable level relative to a reference sound calibrated at the same level as the stimuli. A MUSHRA-like interface in French was developed based on the webMushra framework created by AudioLabs

**Table 1.** Synthesis parameters for ICEV  $M_1$  and  $M_2$ , and for BEV  $E_1$  and  $E_2$ .

	$L_{H_2}$	$(L_{H_2})_0$	$\Delta L_{H_p}$	$\Delta L_{H_p/H_s}$	$(\Delta L_{H_p/H_s})_0$	Chord	$\sigma_f$	$r_F$	$\sigma_a$	$r_A$	$f_A/f_F$	Formants
$M_1$	2	0	-7	3.6	-15	-	0.02	0.3	0.4	0.1	0.1	$h_{M_1}$
$M_2$	2	0	-8	3.6	-20	-	0.02	0.3	0.4	0.1	0.1	$h_{M_2}$
$E_1$	-	-	-	-	-	Major	0.04	0.3	0.38	0.1	0.05	$h_{M_1}$
$E_2$	-	-	-	-	-	Augmented	0.04	0.3	0.38	0.1	0.05	$h_{M_2}$

**Figure 6.** Magnitude responses  $H$  of estimated car structure resonances  $h$  called formants estimated on the measurement of vehicles  $M_1$  and  $M_2$ .

Erlangen [31]. The experiment consisted of two sessions: the first session focused on evaluating the realism of engine sounds in ICEV, while the second session aimed to assess the naturalness of dynamic auditory feedback in BEV.

The first session (ICEV sounds) comprised four trials, each consisting of 10 stimuli to evaluate (40 stimuli in total). Within each trial, the 10 stimuli represented the five conditions (e.g.,  $c_0$ ,  $C_0$ ,  $C_1$ ,  $C_2$ , and  $C_1 + C_2$ ) at two different noise levels, corresponding to a specific vehicle and dynamic. The four trials covered the two vehicles and two dynamics. Participants were instructed to imagine themselves behind the steering wheel of an ICEV and rate the realism of each of the 10 presented engine sounds on a graduated scale from 0 to 100. They were informed that a rating of 0 meant “not realistic at all” and 100 meant “quite realistic.” It was also specified that the engine was in second gear and the vehicle speed ranged between 30 km/h and 40 km/h.

The second session focused on evaluating the BEV dynamic auditory feedback and included four trials, each containing eight stimuli to evaluate (32 stimuli in total). The eight stimuli corresponded to the four conditions at two different noise levels, for a specific vehicle and dynamic. The four trials covered the two BEV vehicles and two dynamics. In the beginning, participants were introduced to the concept of dynamic feedback in BEV. The translated instructions in French stated: “Now imagine yourself behind the steering wheel of an electric vehicle. Since there is no sound from the engine, perceiving the vehicle’s dynamics can be more challenging. This is why it can be interesting, for comfort and safety reasons, to recreate auditory feedback from the motor inside the cabin. The sounds you will hear in this session are intended to fulfill this role. You have to evaluate, on a graduated scale, whether they seem natural or artificial.” Participants were also provided with information about the vehicle speed (ranging between 30 km/h

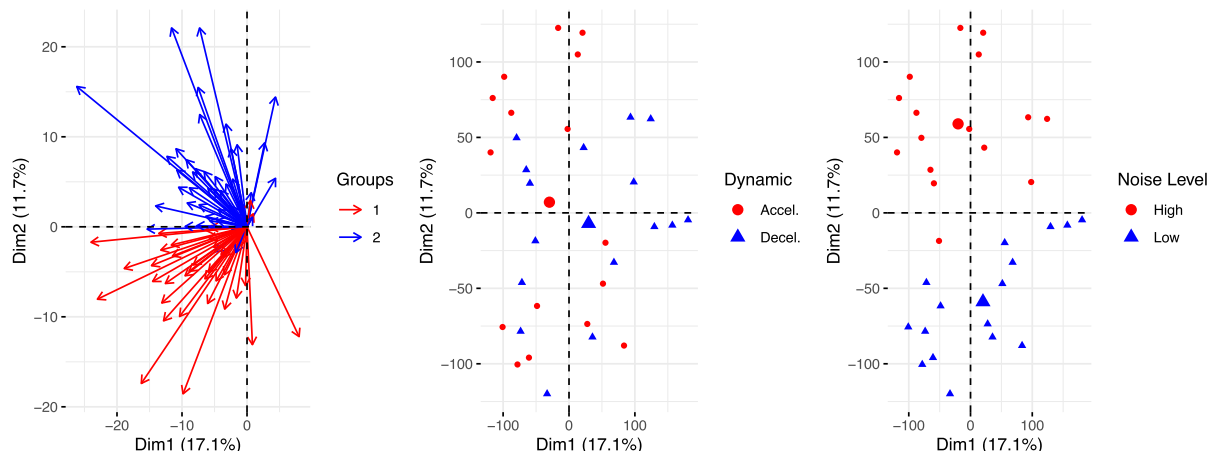
and 40 km/h), the scale boundaries (0 indicating artificial and 100 indicating natural), and were instructed to respond intuitively. In this case, the terms “artificial” and “natural” were used due to the lack of a realistic reference. Letowski [32] introduced these terms to assess sound quality of audio system. He defined naturalness as “the perceptual similarity between an auditory image produced by a given sound and a generalized conceptual image residing in the memory of the listener and used as a point of reference”. Later, the construct “natural – artificial” was found to be elicited by verbal descriptions in the evaluation of spatial audio quality in different studies [33, 34]. Even though the terms were originally used for spatial audio, they evaluate the degree to which the proposed soundscape corresponds to the expected soundscape by the listener. As mentioned by Neidhart et al. [18], in order to improve the integration of a virtual source, it must match the internal reference developed by listeners. Furthermore, informal listening tests often involved comments about the artificial quality of sounds and the perception that the overall soundscape was not natural, aligning with Letowski’s definition. The term “plausible” is also commonly used to evaluate augmented auditory environments [35]. However, in this study, “plausible” was not used to avoid confusion regarding the possibility of electric motors producing these sounds, which is clearly not the case.

For both sessions, each stimulus was evaluated once. Presentation orders of trials and stimuli in each trial were randomized.

#### 4.4 Statistical analyses

Participants’ ratings were collected for each sound (ICEV and BEV sounds). To evaluate differences in the participants’ ratings, a clustering method based on the projection of variables onto latent components was used [36].





**Figure 7.** Principal component analysis of ICEV centered data. Left: Projection of participants on the first two dimensions of the principal component analysis, each arrow corresponds to the ratings of one participant, the color corresponds to his group formed by decomposition on latent components. Center and right: Projection of stimuli on the first two dimensions, each point corresponds to one stimulus. Point color indicates the factors Dynamic (center) and Noise (right), larger points are the centroid of the corresponding set of stimuli.

The ratings for each stimulus were treated as variables. The data were centered but not standardized, as the ratings were expressed on the same scale. The number of clusters,  $K$ , was determined by identifying the point where the explained variance showed less increase between  $K + 1$  and  $K$  clusters compared to  $K$  and  $K - 1$  clusters. Additionally, a Principal Component Analysis (PCA) was conducted with the ratings as variables to demonstrate the effective separation of groups on the primary plane of the PCA.

Statistical analysis was then performed on the ratings within each cluster. First, linear mixed-effects models were fitted to the data, with *Participants* as a random effect and *Transformation* (for ICEV:  $c_0, C_0, C_1, C_2$ , and  $C_1 + C_2$ ; for BEV:  $C_0, C_1, C_2$ , and  $C_1 + C_2$ ), *Dynamic* (acceleration and deceleration), *Vehicle* (for ICEV:  $M_1, M_2$ ; for BEV:  $E_1, E_2$ ), and *Noise* (high and low levels) as fixed effects. Mixed models were used to account for between-subject variability. The statistical significance of the random effect was evaluated by comparing the corresponding fixed-effect-only model with the mixed-effects model (using a  $\chi^2$  test). Secondly, analysis of variance (ANOVA) was conducted on the fitted models. Effects of significant factors were not considered when the effect size ( $\eta^2$  in [37]) was lower than 0.02 with a confidence interval that included 0. Post-hoc analysis (paired  $t$ -tests with Bonferroni correction) revealed statistical differences between the levels of *each factor*.

To further examine the influence of *Transformation* levels independently of the other factors, the ratings were transformed by ranking the stimuli for each trial. For example, in a trial involving deceleration and the  $M_1$  vehicle, if a participant rated transformation  $C_1$  with a high level of noise as the most realistic and transformation  $C_0$  with a low level of noise as the least realistic, the rank scores for  $C_1$  and  $C_0$  would be 10 and 1, respectively. The rank scores were then averaged across trials and *Noise* levels to obtain a

general rank score for each transformation for each participant. This transformation of the data into a non-parametric framework was done to avoid applying parametric models to non-normally distributed data, with effect sizes smaller than those of other factors (see Sect. 4.5). Statistical analysis was then conducted to confirm the results obtained from the parametric models (using Friedman tests), and post-hoc analysis revealed statistical differences between the levels of *Transformation*.

## 4.5 Results

### 4.5.1 Realism of ICEV sounds

The clustering method employed involved choosing  $K = 2$  clusters. Figure 7 (left) illustrates the projection of participants and stimuli onto the first two dimensions of the principal component analysis, clearly separating participants along these dimensions. Analyzing the stimuli, the first dimension tends to separate them based on the *Dynamic* (Fig. 7, center), while the second dimension distinctly separates them based on the *Noise* (Fig. 7, right). Subsequent analyses were conducted separately for each group.

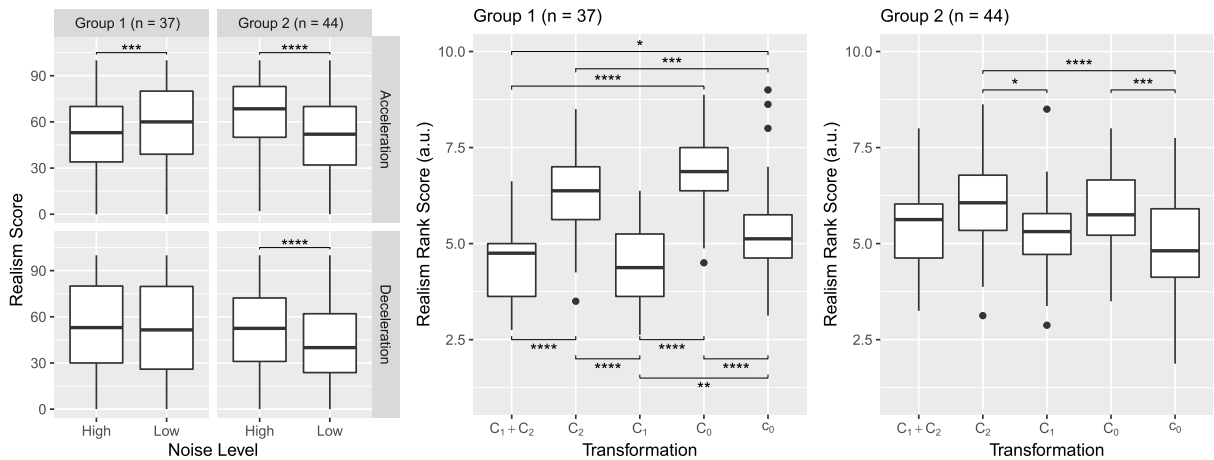
The random effect of *Participant* was found to be significant in both groups ( $p < 0.05$ ). ANOVA on the fitted linear mixed-effects models for groups 1 and 2 are summarized in Table 2. With the exception of *Noise* in group 1, all factors in both groups were significant ( $p < 0.05$ ). The interaction between *Dynamic* and *Noise* was also significant. The main effects (according to the threshold on effect size) are indicated by bold values in Table 2 and displayed in Figure 8 (left).

The results suggest that participants in group 1 primarily evaluated realism in terms of *Transformation*, while participants in group 2 also focused on *Dynamic* and *Noise*, rating acceleration higher than deceleration ( $p < 0.0001$ ) and preferring a higher level of noise ( $p < 0.0001$ ).

**Table 2.** Results of the analysis of variance on realism of ICEV sounds for groups derived from the clustering analysis.

Factor	Group 1 ( $n = 37$ )		Group 2 ( $n = 44$ )	
	$F$ -score ( $p$ -value)	Effect size [CI]	$F$ -score ( $p$ -value)	Effect size [CI]
Transformation	<b>47.46 (2.2e-16)</b> ****	<b>0.12 [0.09, 1]</b>	<b>14.22 (2.0e-11)</b> ****	<b>0.03 [0.02, 1]</b>
Dynamic	4.38 (3.65e-2)*	3e-3 [0, 1]	<b>119.65 (2.2e-16)</b> ****	<b>0.07 [0.05, 1]</b>
Vehicle	4.58 (3.24e-2)*	3e-3 [0, 1]	5.7 (1.7 e-2)*	3e-3 [0, 1]
Noise	1.11 (2.92e-1)	–	<b>134.38 (2.2e-16)</b> ****	<b>0.07 [0.05, 1]</b>
Dynamic $\times$ Noise	8.63 (3.35e-3)**	6e-3 [0, 1]	7.05 (8.0e-3)**	4e-3 [0, 1]

Note. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ . CI: One sided 95% confidence interval. The main effects (according the the threshold on effect size) are indicated by bold values.



**Figure 8.** Left: Realism score of the ICEV sounds for the groups separated by the clustering analysis. Only significant factors *Dynamic* and *Noise* are displayed. Center and right: Rank analysis of ICEV sound realism compared by *Transformation* conditions. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

The influence of *Transformation* was examined separately using rank scores to identify any significant differences in realism independent of other factors. The results are presented in Figure 8 (center and right). Similar trends were observed in both groups, although more differences between conditions were apparent in group 1, where participants placed greater emphasis on *Transformation*. In group 1, all conditions were statistically different ( $p < 0.05$ ), except between  $C_1$  and  $C_1 + C_2$ , and between  $C_0$  and  $C_2$ . Therefore,  $C_0$  and  $C_2$  were rated as the most realistic, followed by  $c_0$ , while the least realistic sounds were associated with  $C_1$  and  $C_0 + C_2$ . In group 2, the following conditions showed statistical differences ( $p < 0.05$ ):  $c_0$  and  $C_0$ ,  $c_0$  and  $C_2$ , and  $C_1$  and  $C_2$ . In terms of sound transformations, the presence of modulations decreased the realism of the sounds, whereas applying formants helped maintain realism. The differences between participants in group 1 and group 2 were more pronounced in relation to the impact of these factors.

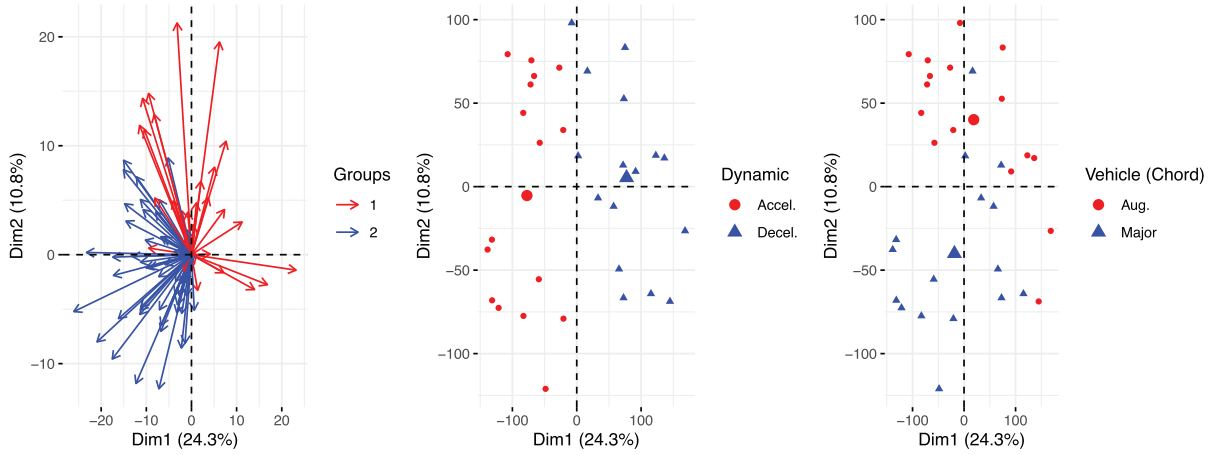
#### 4.5.2 Naturalness of BEV sounds

The ratings of the participants were subjected to a clustering method with  $K = 2$  clusters (Fig. 9, left), revealing contrasting trends between the defined groups. The composition of these groups differed from those observed in the ICEV analysis, with 31 participants not belonging to the same group.

Analyzing the stimuli, the first dimension was primarily explained by the *Dynamic* (Fig. 9, center), while the second dimension was influenced by the *Vehicle* (Fig. 9, right).

The random effect of *Participant* was found to be significant in both groups ( $p < 0.05$ ). Further analysis using an ANOVA on the linear mixed-effects models yielded the results summarized in Table 3. Figure 10 illustrates the distribution of ratings among the groups.

The rank analysis indicated that in group 1, the *Transformation* did not significantly influence the participants' ratings regarding the natural character of the sounds

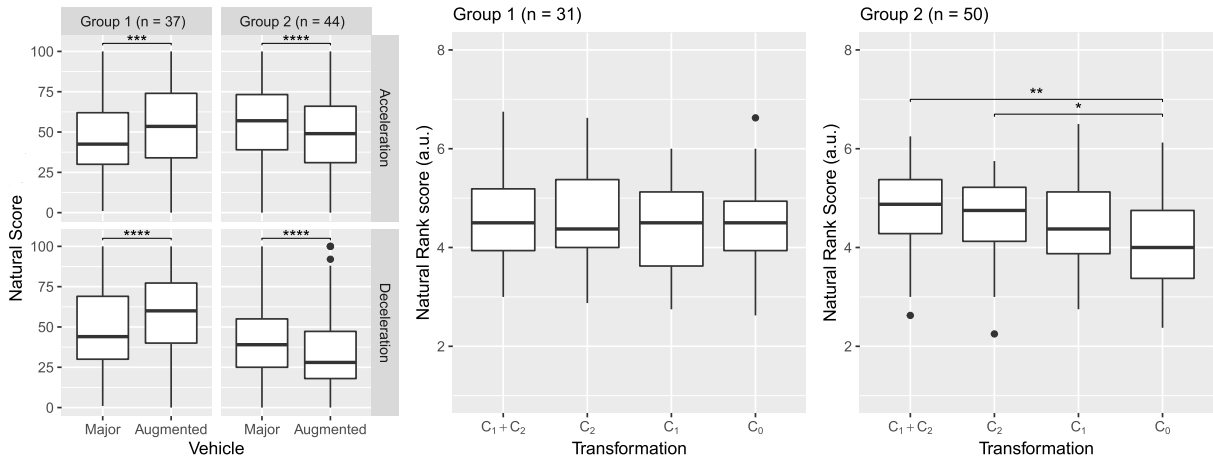


**Figure 9.** Principal component analysis of BEV centered data. Left: Projection of participants on the first two dimensions of the principal component analysis, each arrow corresponds to the ratings of one participant, the color corresponds to his group formed by decomposition on latent components. Center and right: Projection of stimuli on the first two dimensions, each point corresponds to one stimuli. Point color indicates the factors *Dynamic* (center) and *Vehicle* differentiated by the chord of the dynamic feedback (right), larger points are the centroid of the corresponding set of stimulus.

**Table 3.** Results of the analysis of variance on naturalness of BEV sounds for groups derived from the clustering analysis.

Factor	Group 1 ( $n = 31$ )		Group 2 ( $n = 50$ )	
	$F$ -score ( $p$ -value)	Effect size [CI]	$F$ -score ( $p$ -value)	Effect size [CI]
Transformation	5.42 (1.1e-3)**	0.02 [0, 1]	<b>8.73 (1.0e-5)****</b>	<b>0.02 [0.01, 1]</b>
Dynamic	7.49 (6.3e-3)**	7e-3 [0, 1]	<b>305.9 (2.2e-16)****</b>	<b>0.17 [0.14, 1]</b>
Vehicle	<b>38.33 (8.9e-10)****</b>	<b>0.04 [0.02, 1]</b>	<b>51.58 (1.1e-12)****</b>	<b>0.03 [0.02, 1]</b>
Noise	0.54 (4.6e-1)	-	1.69 (1.9e-1)	-

Note. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ . CI: One sided 95% confidence interval. The main effects (according to the threshold on effect size) are indicated by bold values.



**Figure 10.** Left: Naturalness score of the BEV sounds for the groups separated by the clustering analysis. Only significant factors *Dynamic* and *Vehicle* are displayed. Center and right: Rank analysis of BEV sound naturalness compared by *Transformation* conditions. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

(Fig. 10, center). Instead, participants in group 1 predominantly focused on the *Vehicle*, resulting in higher scores for the augmented chord ( $p < 0.0001$ ) across both dynamics. Conversely, participants in group 2 rated both  $C_2$  and  $C_1 + C_2$  as more natural than  $C_0$  ( $p < 0.05$ ), highlighting the importance of formants for the naturalness aspect of the sounds (Fig. 10, right). Additionally, participants in group 2 placed greater emphasis on the *Dynamic* ( $p < 0.0001$ ) and considered sounds in acceleration as more natural than those in deceleration. They also assigned less significance to the Risset chord but rated the major chord as more natural ( $p < 0.0001$ ).

#### 4.5.3 Post experiment reports

At the end of the experiment, feedback was collected through a questionnaire, which included a free comment field. Information regarding the participants, such as age, expertise, driving experience, and familiarity with BEV, did not explain the differences in ratings. A total of 36 free comments were collected. Among these, 17 comments mentioned the difficulty in some cases to differentiate between the sounds. Three comments mentioned the challenge of evaluating the stimuli without being behind the wheel or inside the vehicle, as the context of the test was not representative of a real driving or vehicle experience. Additionally, one participant noted a mismatch between their expectation of a quiet environment in a BEV and the introduction of a new sound source, which led to lower ratings. The remaining 15 comments were not relevant in explaining the results, as they either expressed opinions about the aesthetic aspects of the proposed sounds or the test procedure itself.

## 5 Discussion

The results indicated that the *Transformation* had a significant effect on the perception of realism/naturalness for both ICEV and BEV sounds. In the case of ICEV sounds, both groups showed similar trends, but group 1 exhibited more differences between conditions. The low anchor  $c_0$  was rated as less realistic than the reference sound  $C_0$ . This suggests that a dense harmonic comb is important for characterizing a realistic engine sound. Surprisingly, the inclusion of modulations (conditions  $C_1$  and  $C_1 + C_2$ ) decreased the perceived realism compared to the reference  $C_0$  even though these modulations were present in the measurements. This observation may be explained by the fact that these modulations might not be perceived or might be perceived differently in conditions closer to reality, where other auditory cues such as spatial cues or multimodal cues (vision, vibration, etc.) play a more important role. The experimental setup, which did not fully replicate real-world conditions, may have led participants to focus on these modulations more than they would have otherwise. Another possible explanation for the lower ratings of modulations (conditions  $C_1$  and  $C_1 + C_2$ ) is that the modulations might be slightly overestimated for partials higher than  $H_2$  because they do not stand out significantly from

the background noise (cf. the spectrogram in Fig. 2). Consistently, the modulations of the most prominent partial,  $H_2$ , were lower than those of higher partials. Consequently, the resulting sounds could have been assessed as less realistic than the anchor  $c_0$ . Interestingly, sounds containing formants (condition  $C_2$ ) were evaluated as realistic as the reference sound which indicates that this attribute keeps the realistic aspect of sounds. However, the presence of formants did not improve the condition  $C_1 + C_2$  compared to  $C_1$ . In the context of the experiment, refining the model did not improve the realism.

Concerning the BEV sounds, the presence of formants (conditions  $C_2$  and  $C_1 + C_2$ ) improved the naturalness of the sounds in group 2. Initially, the designed auditory feedback did not possess any characteristics of a sound originating from a vehicle. When the sound is filtered by the resonances of the car structure, it aligns with the acoustic properties of the environment and shares common characteristics with other sources perceived inside the vehicle. This blending of the sound with the environment may contribute to its naturalness. However, the formants used in the BEV sounds matched those of ICEV. Despite this mismatch, the presence of formants still improved the naturalness of the BEV sounds. This finding suggests that participants' expectations might lie in the variation of timbre with dynamics rather than the identification of resonances present in other sources. When the harmonic comb varies with the dynamics of the vehicle, it interacts with the resonances and affects the timbre. To further investigate this hypothesis, a comparison with formants corresponding to those of a BEV would be interesting. Furthermore, the ratings for sounds including modulations did not differ from those without modulations. This indicates that these timbre micro-variations do not significantly contribute to the naturalness of BEV sounds, unlike the case with ICEV sounds.

Participants in both sessions of group 2 consistently rated the sounds during acceleration as more realistic/natural compared to deceleration. This finding is supported by the PCA results (cf. Figs. 7 and 9), where most participants are positioned on the left side of the main plane, corresponding to higher scores for acceleration. Higher ratings for acceleration can be attributed to several factors. Firstly, the engine reference model (cf. Sect. 2.2) was initially developed for studying the perception of accelerating sounds [20]. Therefore, its applicability to decelerating sounds has not been evaluated. Secondly, the dynamic profile used during acceleration aligns more closely with typical driving situations, such as full-throttle acceleration, compared to the deceleration profile without braking. This suggests that the consistency of the dynamic profile with expected driving scenarios contributes to the perceived realism/naturalness of the sounds.

In the case of ICEV, participants evaluated sounds with a high level of tyre/road noise as more realistic, whereas this trend was not observed for BEV. The high level of tyre/road noise corresponded to the measured level in the reference recordings (as explained in Sect. 2.1), indicating that participants perceived the actual signal-to-noise ratio between the engine sound and road/tyre noise as more

realistic. However, in the case of BEV sounds where no reference existed, this level of tyre/road noise was no longer influential in the evaluation of naturalness. In contrast to ICEV, which possess specific vehicle features, the types of BEV were arbitrarily associated with sounds characterized by different chords. Group 1 participants found the augmented chord more natural, while group 2 participants evaluated the major chord as more natural. From a perceptual standpoint, the chords introduced changes in the timbre of the sound. This observation suggests that the expected timbre related to naturalness depends on individuals, potentially influenced by specific experiences such as musical background. These findings highlight the potential of adaptive design strategies based on users.

In both experimental sessions, the raw data exhibited a large inter-subject variability, as evidenced by the significant effect of the random factor *Participant* in the linear mixed-effects models. Further analysis using clustering techniques revealed the presence of two distinct groups of participants. This indicates that participants employed different strategies when evaluating the stimuli, assigning varying degrees of importance to certain factors and even providing contrasting ratings. The collected metadata on participants, such as age, gender, and their interest or experience in acoustics or the automotive field, did not account for the observed differences in evaluation strategies. Overall, the task of evaluating realism and naturalness proved challenging for the participants, as indicated by their post-experiment reactions. The results indicated that participants paid more attention to factors such as *Dynamic*, *Vehicle*, and *Noise* rather than the *Transformation* aspect involving formants and modulations. This suggests that subtle variations in timbre have minimal impact on the perceived realism and naturalness of sounds within the interior car context. Instead, consistent dynamic variations, noise levels, and expected timbre (e.g., the choice of vehicle chord) play a more significant role in shaping participants' perception.

In summary, the findings of this study provide valuable insights for refining strategies in the integration of dynamic auditory feedback within the cabin environment. The results indicate that modulations do not enhance the naturalness of dynamic feedback in BEV, suggesting that these specific timbre variations are not relevant parameters for achieving a consistent environment. Also, the presence of formants, which characterize the resonances of the vehicle, preserved the realism of ICEV sounds and even increased the naturalness of BEV sounds. These considerations lead us to claim that the integration problem may be related to other properties of the sound, such as spatial features and features induced by the environment rather than by the sound.

## 6 Conclusion

In this study, the focus was on investigating the impact of timbre-related features observed in the interior soundscape measurements of ICEV, particularly the engine sound, on the integration of dynamic auditory feedback in BEV.

The combustion process in the powertrain of ICEV involves random amplitude and frequency micro-modulations of the engine sound's harmonic components. These micro-modulations were modeled as slow normally distributed random variations in amplitude and frequency for each harmonic component. It was hypothesized that these micro-modulations could enhance the naturalness of dynamic feedback by introducing non-deterministic components.

Additionally, the vibroacoustic transfer from the engine compartment to the car cabin results in resonant filtering. These resonances were modeled as a combination of peak filters. The assumption was that matching the acoustic properties of the interior car environment would improve the naturalness of dynamic feedback by aligning the sound characteristics with the surrounding environment.

However, for ICEV sounds, it was found that refining the engine sound model with micro-modulations did not contribute to improving the perceived realism. Subsequently, the study evaluated the perceived naturalness of dynamic auditory feedback in BEV. In this context, the presence of resonances was found to improve the naturalness of the sound, indicating the importance of matching the characteristics of the surrounding environment. These results suggest that, rather than focusing solely on subtle timbre variations, other aspects of the sound, such as spatial aspects, may have a larger impact on achieving a consistent integration of virtual sources in the car cabin.

To further explore this, a similar study was conducted on the spatial integration of virtual sounds using a multi-sensory environment, including Virtual Reality devices [38]. This approach aimed to improve the contextualization and the feeling of being immersed in a realistic car scene. The obtained results from this study align with the aforementioned considerations, leading to the proposal of integration strategies based on the spatial conformations of virtual sources within the car cabin [39].

## References

1. F. Doleschal, J.L. Verhey: Pleasantness and magnitude of tonal content of electric vehicle interior sounds containing subharmonics. *Applied Acoustics* 185 (2022) 108442.
2. M. Mündler, C.-C. Carbon: Howl, whirr, and whistle: The perception of electric powertrain noise and its importance for perceived quality in electrified vehicles. *Applied Acoustics* 185 (2022) 108412.
3. G. Goetchius: Leading the charge – the future of electric vehicle noise control. *Sound & Vibration* 45, 4 (2011) 5–8.
4. S. Denjean, V. Roussarie, R. Kronland-Martinet, S. Ystad, J.-L. Velay: How does interior car noise alter driver's perception of motion? Multisensory integration in speed perception, in *Acoustics 2012*, Nantes, France, April, 2012.
5. J.-F. Sciabica, M.-C. Bezat, V. Roussarie, R. Kronland-Martinet, S. Ystad: Towards the timbre modeling of interior car sound, in *Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, Denmark, May 18–22, 2009.
6. V. Roussarie, F. Richard, M.-C. Bezat: Perceptive qualification of engine sound character: validation of auditory attributes using analysis-synthesis method, in *Proceedings of the CFA/DAGA*, Strasbourg, France, March 22–25, 2004.

7. M. Bodden: Principles of active sound design for electric vehicles, in INTER-NOISE and NOISE-CON Congress and Conference Proceedings, vol. 253, Institute of Noise Control Engineering, 2016, pp. 7700–7704.
8. R. Schirmacher, Active design of automotive engine sound, in Audio Engineering Society Convention 112, Audio Engineering Society, 2002.
9. D. Swart, A. Bekker, J. Bienert: The subjective dimensions of sound quality of standard production electric vehicles. *Applied Acoustics* 129 (2018) 354–364.
10. D.Y. Gwak, K. Yoon, Y. Seong, S. Lee: Application of subharmonics for active sound design of electric vehicles. *Journal of the Acoustical Society of America* 136, 6 (2014) EL391–EL397.
11. Y. Cao, H. Hou, Y. Liu, L. Tang, Y. Li: Engine order sound simulation by active sound generation for electric vehicles. *SAE International Journal of Vehicle Dynamics, Stability, and NVH* 4 (2020) 151–164.
12. F. Doleschal, H. Rottengruber, J.L. Verhey: Influence parameters on the perceived magnitude of tonal content of electric vehicle interior sounds. *Applied Acoustics* 181 (2021) 108155.
13. M. Maunder, Experiences tuning an augmented power unit sound system for both interior and exterior of an electric car. *SAE Technical Paper* 2018-01-1489, 2018.
14. K.-J. Chang, G. Cho, W. Song, M.-J. Kim, C.W. Ahn, M. Song: Personalized EV driving sound design based on the driver’s total emotion recognition. *SAE International Journal of Advances and Current Practices in Mobility* 5, 2 (2023) 921–929.
15. R. Schramm, J. de Kruiff, R. Doerfler, J. Merkt, P. Kampmann, F. Walter: AI in automotive audio: approaching dynamic driving sound design, in Audio Engineering Society Conference: AES 2022 International Automotive Audio Conference, Audio Engineering Society, 2022.
16. S. Denjean, R. Kronland-Martinet, V. Roussarie, S. Ystad: Zero-emission vehicles sonification strategy based on shepard-risset glissando, in International Symposium on Computer Music Multidisciplinary Research, Springer, 2019, pp. 709–724.
17. Y. Cao, H. Hou, Y. Liu, Y. Li, S. Wang, H. Li, C. Zhang: Sound pressure level control methods for electric vehicle active sound design. *SAE International Journal of Vehicle Dynamics, Stability, and NVH* 5 (2021) 205–226.
18. A. Neidhardt, C. Schneiderwind, F. Klein: Perceptual matching of room acoustics for auditory augmented reality in small rooms – literature review and theoretical framework. *Trends in Hearing* 26 (2022).
19. C. Kuhn-Rahloff: Realitätstreue, Natürlichkeit, Plausibilität: Perzeptive Beurteilungen in der Elektroakustik. PhD thesis, TU, Berlin, Germany, 2012.
20. J.-F. Sciabica: Caractérisation acoustique et perceptive du bruit moteur dans un habitacle automobile. PhD thesis, Aix-Marseille 1, 2011.
21. F. Richard, F. Costes, J.-F. Sciabica, V. Roussarie: Vehicle acoustic specifications using masking models, in INTER-NOISE and NOISE-CON Congress and Conference Proceedings, vol. 2007, Institute of Noise Control Engineering, 2007, pp. 3153–3162.
22. R. McAulay, T. Quatieri: Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34, 4 (1986) 744–754.
23. B.G. Quinn: Estimation of frequency, amplitude, and phase from the DFT of a time series. *IEEE Transactions on Signal Processing* 45, 3 (1997) 814–817.
24. S. Provencher: Estimation of complex single-tone parameters in the DFT domain. *IEEE Transactions on Signal Processing* 58, 7 (2010) 3879–3883.
25. M. Caetano, P. Depalle: On the estimation of sinusoidal parameters via parabolic interpolation of scaled magnitude spectra, in 24th International Conference on Digital Audio Effects (DAFx), IEEE, 2021, pp. 81–88.
26. X. Serra, J. Smith: Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal* 14, 4 (1990) 12–24.
27. R.N. Shepard: Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America* 36, 12 (1964) 2346–2353.
28. S. Denjean: Sonification des véhicules électriques par illusions auditives: étude de l’intégration audiovisuelle de la perception du mouvement automobile en simulateur de conduite. PhD thesis, Aix-Marseille, 2015.
29. W.M. Hartmann, Y.J. Cho: Generating partially correlated noise – a comparison of methods. *Journal of the Acoustical Society of America* 130, 1 (2011) 292–301.
30. B. Series: Method for the subjective assessment of intermediate quality level of audio systems, International Telecommunication Union Radiocommunication Assembly, 2014.
31. M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, J. Herre: webMUSHRA – a comprehensive framework for web-based listening tests. *Journal of Open Research Software* 6, 1 (2018) 8.
32. T. Letowski, Sound quality assessment: concepts and criteria, in Audio Engineering Society Convention 87, Audio Engineering Society, 1989.
33. J. Berg, F. Rumsey, In search of the spatial dimensions of reproduced sound: verbal protocol analysis and cluster analysis of scaled verbal descriptors, in Audio Engineering Society Convention 108, Audio Engineering Society, 2000.
34. J. Berg, F. Rumsey: Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction, in 109th AES Convention, Los Angeles, CA, USA, September 22–25, 2000.
35. A. Lindau, S. Weinzierl: Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica* 98, 5 (2012) 804–810.
36. E. Vigneau, E. Qannari: Clustering of variables around latent components. *Communications in Statistics-Simulation and Computation* 32, 4 (2003) 1131–1150.
37. R. Bakeman: Recommended effect size statistics for repeated measures designs. *Behavior Research Methods* 37, 3 (2005) 379–384.
38. T. Dupré, S. Denjean, M. Aramaki, R. Kronland-Martinet: Spatial sound design in a car cockpit: Challenges and perspectives, in 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), IEEE, 2021, pp. 1–5.
39. T. Dupré, S. Denjean, M. Aramaki, R. Kronland-Martinet: Spatial integration of dynamic auditory feedback in electric vehicle interior, in Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference, Audio Engineering Society, 2022.