HAL
open science

A perceptual evaluation of numerical errors in acoustic FEM simulation for sound quality applications

Giorgio Pulvirenti, Nicolas Totaro, Etienne Parizet

▶ **To cite this version:**

Giorgio Pulvirenti, Nicolas Totaro, Etienne Parizet. A perceptual evaluation of numerical errors in acoustic FEM simulation for sound quality applications. Applied Acoustics, 2023, 207, 10.1016/j.apacoust.2023.109295 . hal-04082703

HAL Id: hal-04082703

https://hal.science/hal-04082703v1

Submitted on 27 Apr 2023

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

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

Applied Acoustics 207 (2023) 109295

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

A perceptual evaluation of numerical errors in acoustic FEM simulation for sound quality applications

Giorgio Pulvirenti, Nicolas Totaro *, Etienne Parizet

Laboratoire Vibrations Acoustique, INSA-Lyon, 25 bis, av. Jean Capelle, 69621 Villeurbanne Cedex, France

ARTICLE INFO

Article history:
Received 3 September 2022
Received in revised form 18 January 2023
Accepted 21 February 2023
Available online 23 March 2023

Keywords:
Finite elements
Sound quality
Transfer function

ABSTRACT

Sound perception is a key aspect in many industrial applications where the acoustic comfort of the final user is relevant. In the case of automotive industry, an important effort is put into prediction of sound quality, which in turn necessitates detailed description of the acoustic problem. The complexity of boundaries and constraints in the typical engineering problems requires the use of numerical methods for this description, and the Finite Elements Method (FEM) is one of the most spread. Numerical methods in general and FEM in particular, can be rather accurate if the simulated model is fine enough. However, this usually means a high computational load, with an associate long computation time and (or) very high hardware requirements. These costs need to be diminished to make future design process faster and more efficient. Several ways to address this problem have been proposed and successfully used, usually based on the idea of making more efficient the mathematical formulation and solution of the simulation. Another option is to consider limits of human hearing system to reduce the accuracy requirements of FE meshes. The basic idea is that, since humans are not able to detect very small differences between sounds, a simulation can have a certain degree of numerical error without affecting the perception. Despite the concept being so simple, its development is rather complex since it requires two complicated fields (numerical analysis and sound perception) to merge into one unique solution. Outcome cannot be forecast in a simple manner, and introductory studies are required to form a solid base on which future research can be built. This paper analyses the sensitivities of human perception with respect to some numerical parameters typical of a FE model for exterior acoustic problems. The concept of adaptive mesh is taken into account and the Perfectly Matched Layer (PML) is used to ensure the Sommerfeld radiation condition, thus introducing a set of parameters to be analyzed. Several sounds are obtained by convolving a source signal with several simulated Transfer Functions (TF). These TFs represent the same acoustic path but are obtained through FE models with different parameters. Psychoacoustic metrics are used to check how the changes in FE models affect the sound properties. Lastly, jury tests are performed to assess how these differences modify human perception.

© 2023 Elsevier Ltd. All rights reserved.

1. Introduction

In many industrial applications, acoustic comfort of the final user is critical. This is mainly due to two different reasons. The first one is related to acoustic pollution. European Union showed that noise is one of the most dangerous pollutant actors for public health, being related to high discomfort and, as a consequence, to a significant increase of cardiac problems [1]; therefore, it is important to reduce the annoyance and the acoustic energy of noises that pollute the environment. The second reason is important in the case of products that are purchased by private customers. Intuitively, a potential customer may discard a product not considered satisfying enough. In such cases, engineers more appropriately speak about product sound quality [2].

Automotive industry forms a perfect example for this kind of applications. Road vehicles need to be comfortable for customers, and also need to emit low level noise. Several other factors contribute to make this example even more complicated. First, vehicle need to be detected by pedestrians, in order to avoid accidents. As a consequence, emitted noise should not be too low, which is a clear trade off with acoustic pollution. Such a trade off is taken into account by most recent European regulation [3]: vehicles must radiate a level of acoustic energy higher than a minimum and lower than a maximum to pass the test. Some research has already been carried out to define sounds that allow easy identification of the vehicle without excessively increasing the discomfort of

* Corresponding author.

https://doi.org/10.1016/j.apacoust.2023.109295
0003-682X/© 2023 Elsevier Ltd. All rights reserved.

pedestrians [4]. Furthermore, many sources are present in vehicles, and acoustic transfer paths are several and complicated due to the complex geometry involved.

The importance of sound quality in such a context is crucial, and it brings several challenges. First of all, sound quality generally needs a detailed description of the acoustic problem. In engineering applications, such a description may pose a significant issue to designers, because of the degree complexity usually involved. Even neglecting non-linearity, insurmountable troubles for analytical descriptions depends on complicated geometries, boundaries, constraints, inhomogeneity and material properties that are part of the average engineering problem. Numerical methods are usually adopted to overcome these difficulties.

In this context, to make sound quality design decisions, engineers create sound files from simulated, measured, or synthesized data for listening purpose. This process is called auralization [5]. In the early stage of development and in case of auralization aim, the use of simulated sounds is critical. The strategy is to define a mathematical model, accounting for all the non-negligible phenomena involved in the acoustic transfer path (TP), from the source to the listener. Several recent applications of auralization are discussed in literature; for instance, [6–9]. Even if modern simulations tools are incredibly powerful, they still face some limitations. The bitter endpoint of their capabilities is the requirement of high computational power to produce finely tuned models. Researchers proposed many different techniques to tackle this issue (see, for example, [10–15]). Many of these methodologies are strictly mathematical, meaning that the use of a particular algorithm allows to reduce the calculation time by reducing the size of data-sets, by increasing the efficiency of the calculation process, or by rewriting the original problem in a simplified manner, for instance.

Those above studies proposed promising ways to exploit auralization techniques exclusively relying on simulations for sound quality prediction. Making such techniques robust, reliable and efficient requires a deep comprehension of the impact of simulation errors on acoustic perception and constitutes a necessary further step. Such knowledge can give indications on how to adequately set parameters of the mathematical models and, eventually, even to optimize them. For example, the following studies focused on this issue, contributing with essential results.

In 2011, Nykänen et al. [16] investigated the impact of frequency resolution and spectrum smoothing of binaural TFs into perception. The test case was the contribution of engine sounds to interior sounds of a truck. They performed tests to compare artificial head recordings to modified binaural TFs. Their findings showed that a resolution of 4 Hz or higher and a smoothing with maximum 1/96 octave moving average filters were comparable to artificial head recordings. Trollé et al. confirmed this frequency resolution value for vibroacoustic applications [17].

Nykänen, together with the same researchers of the previous work, continued their research by publishing further results [18] in 2013. This time the objective was to evaluate the maximum changes in frequency resolution and smoothing to preserve preference rating. They found that a much higher degree of degradation for binaural TFs was possible. They found a value of 32 Hz for frequency resolution, and 1/24 octave bandwidth filters or 63 Hz absolute bandwidth filters for smoothing operations. Even though these values preserved preference ratings, the degradation comes together with a loss of consistency between participants of jury tests. Furthermore, results were not general since the modifications affected only TFs, and not the source. Final recommendations included the need for variance analysis when repeating a similar approach.

In 2018, Aujogue and Parizet further extended scientific knowledge about the accuracy of TFs for auralization purposes [19] in the case of TFs from input forces to acoustic pressure. Furthermore, the modifications were different for diverse area of the audible spectrum. At low frequency (i.e. low modal density), the study focused on shifting the eigenfrequencies towards higher or lower frequencies by modifying the poles of the TFs. At mid and high frequencies, the main changes were smoothing techniques, consistently with previous studies. Results were similar to the ones of Nykänen et al. Besides, they found that at low frequency ranges, shift of the resonance frequency could easily be detected, in part due to the strong harmonic structure of the source.

The results in literature about the perception of TF's errors are consistent and suggest not only that models might represent an efficient supporting tool for auralization even in presence of errors since the introduced modifications of TFs does not affect too much people's perception.

The most significant limits of the previously mentioned studies is that they introduced modifications in experimental TFs that doe not represent errors typical of numerical simulations. On the contrary, simulations present specific errors that might peculiarly affect perception. Understanding how people perceive and react to these errors is a crucial step towards the definition of thresholds between acceptable and unacceptable compromises in the use of these virtual methods.

This research covers this gap in knowledge, focusing on the influence of simulation parameters in the perception of auralized sounds. The human hearing system presents some filtering effects that prevent people from perceiving all the details of acoustic inputs. Therefore, depending on the task, it should be possible to simulate acoustic signals with low accuracy while retaining all the attributes necessary to perform the given assignment. Instead of developing additional mathematical tools to improve simulation performances, the focus switches to the production of "fine enough" models, which are expected to be much coarser than those needed following the first strategy. In detail, this paper studies some FE models based on meshes whose number of nodes per wavelength is purposefully and significantly lower than the most commonly accepted rule of thumbs. Indeed, if very coarse mesh can provide satisfactory results for certain specific applications, why not use them? The consequences will be faster computations and the same perception evaluation of sounds. Obviously, the resulting auralized sounds need, at any rate, to keep essential aims-dependant psychoacoustic properties. It is impossible to define what "fine enough" means without considering the specific application; this paper provides some hints on what can be considered sufficient in the case of similarities perception (namely, when people need to identify two acoustic signals as similar or different).

The research focuses on simulated transfer functions. The main reason to study in detail the TP modelling via simulation is the evident flexibility of this approach. As mentioned, it allows engineers to analyze different design configurations without the need for prototypes, taking into account only the TPs of interests. Evaluation of potential TP modifications would become simple even for complex geometries, thanks to simulation capabilities. Furthermore, assuming a description of the source being available as a set of monopoles (or other simple sources), either with a testing or a modelling approach, the coupling with the TPs would be possible relatively straightforwardly. All of this would already be possible in the early stages of development, with significant advantages for the process. The success of simulations in industrial applications is a proof of such strengths, and a crucial reason to bring improvements to the field.

The present paper is organized as follows. Section 2 provides a description of the numerical models used to calculate the TFs and an introduction to the simulation technique used to solve the models themselves, with a focus on the error they introduce into the TFs. Results of the simulation are compared in Section 3, from a

mathematical perspective, to show the behaviour of the numerical errors in practical cases. The sound quality aspects are discussed in Section 4; data are collected via jury tests and analysed using several statistical techniques.

2. Numerical model

Many families of transfer paths have importance in automotive engineering. Among them, one has some peculiar characteristics that make it a good fit for this research: the set of TPs between the tires (sources) and the pedestrians (receivers). Indeed, the tires are a critical source of noise and will be even more in the early futures due to the introduction of hybrid and electric engines. Furthermore, these TPs do not present any obstacles whose physical modelling may introduce extra approximations. The acoustic problem is further simplified by considering a (virtual) acoustic quarter-vehicle model, shown in Fig. 1, on a rigid baffled plane. No other phenomena are modeled, in particular the model is completely rigid and no relative rotation between the parts is considered. This model has to be seen as a study case whose simplifications allow to consider only the errors due to the numerical techniques used to solve the differential equations. Four monopolar sources and four receivers are used, for a total of sixteen TPs. The monopolar sources are modeled using the software Actran [20] as P-sources, namely a spherical monopolar source with an acoustic pressure amplitude of 1 Pa at a distance of 1 m.

The spatial locations of sources and receivers are identified on a Cartesian reference system. Its origin is located in a position that exploits symmetry properties of the mock-up. This latter has only one symmetry plane, which is vertical and includes the axis of the wheel. The intersection of this plane with the ground and the plane comprising the front panel of the box is the origin O of the Cartesian coordinate system used hereby, see Fig. 1. The z-axis is vertical and positively oriented towards the top. The x-axis is parallel to the tire axis and is positive oriented towards the direction opposite to the mock-up itself. Fig. 1 shows the sources' locations with yellow spheres; Table 1 summarizes their coordinates in meter. The receivers are shown in Fig. 1 by the red points; their coordinates are presented in Table 2. A shows the multiview ortographic projections of this mock-up to give a more accurate vision to the readers.

Table 1
Monopole locations' coordinates.

Mono-pole	X [m]	Y [m]	Z [m]
S+	−0.204	0.29334	0.174
G+	−0.204	0.1	0.0065
G-	−0.204	−0.1	0.0065
S-	−0.204	−0.29334	0.174

Table 2
Receiver locations' coordinates.

Receiver	X [m]	Y [m]	Z [m]
*P*10	1	0	1.2
*P*20	2	0	1.2
*P*11	1	1	1.2
*P*21	2	1	1.2

From now on, the TPs will be indicated by the two letters associated to the monopole-microphone position followed by the two numbers associated to the receiver-source position, e.g. TP S + 21 is the transfer function between the mono-pole located at position S+ and receiver at position 21. The transfer functions associated to these TPs are calculated via FE with perfectly matched layers (PML). The following subsections briefly discuss these two methods and some associated errors: approximation, pollution, and reflection.

2.1. The finite element method

The FEM for Helmholtz problem consists in solving the following equation [21]:

$$\left[-k^2\mathbf{M} - ik\mathbf{C} + \mathbf{K}\right]\mathbf{p} = \mathbf{f}, \tag{1}$$

where $\mathbf{M}, \mathbf{C}, \mathbf{K}$, and $\mathbf{f}$ are, respectively, the acoustic mass, damping, stiffness matrices and the boundary vector; k is the wavenumber, i is the imaginary unit, and the vector $\mathbf{p}$ represents the pressure at the nodes. The matrices and vector $\mathbf{M}, \mathbf{C}, \mathbf{K}$, and $\mathbf{f}$ are functions of the physics of the problem and of polynomial shape functions. Different types of polynomial functions can be used: the Lagrangian

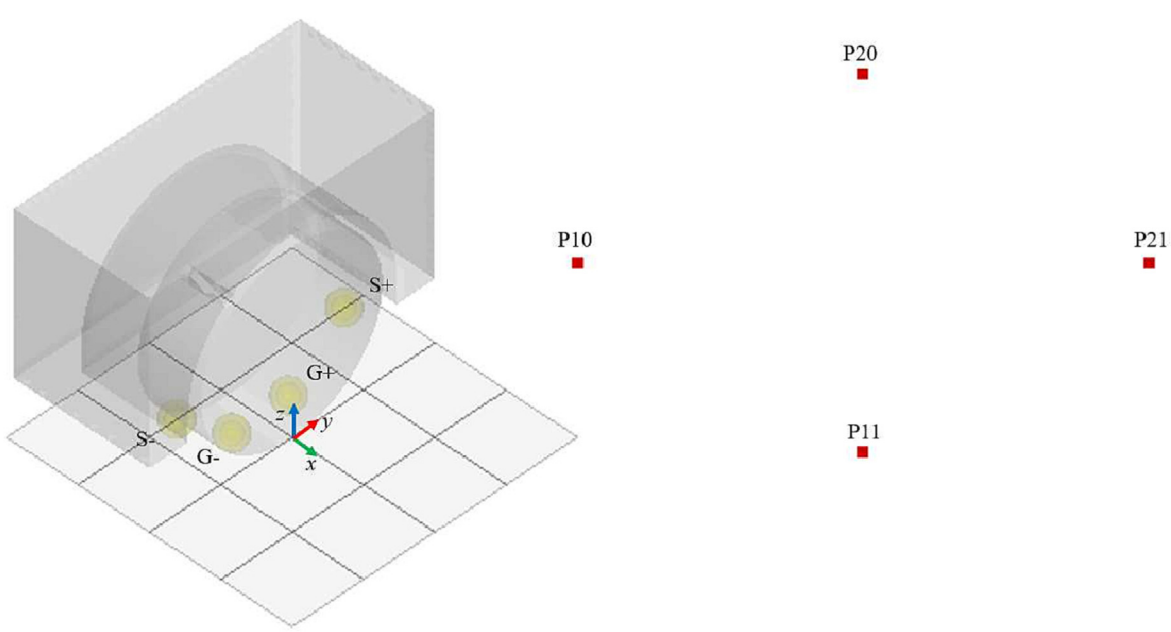

Fig. 1. Monopole (yellow spheres) and receiver (red points) positions; x-(green) y-(red) and z-(blue) axis centered in the origin, at the contact between the tire and the ground. The rigid baffle plane (shown as a grid) extends to the infinite.

polynomials are common, and even the Legendre polynomials can be found [22]. The shape functions serve to interpolate the approximated value of the acoustic pressure inside each element, given the values at each node of the element itself; for these reasons, the order of the polynomial $\mathfrak{p}_{FE}$ is here called order of interpolation of the shape functions. The latter parameter is chosen by the user, and it is usually the same across all the elements, but not always: see, for instance, [23].

2.1.1. Numerical errors in the FEM

Given an acoustic wave propagating in a homogeneous and isotropic medium, when the speed of sound c is independent from frequency, the linear dispersion relation $\frac{\omega}{k} = c$ holds [24], where ω is the angular frequency. When discretizing the propagative phenomenon via FEM, the linear dispersion relation does not hold anymore, leading to spurious dispersion. An introduction of this phenomenon is provided in [25]. An extension exists in the literature that leads to the same numerical dispersion relation for the cylindrical and the spherical cases [26]. However, since these errors are the basis of the next part of this paper, it makes sense to introduce them, which is the purpose of this section. Results in this section are not obtained via FE software, but via analytical simulations whose parameters have been set up according to the literature [25–28].

Regarding the case of planar waves in the one-dimensional case, given a regular mesh of elements with size $\mathfrak{h}$, the matrix of the FEM leads to a set of algebraic equations at each node that provides the following dispersion relation for the computed wavenumber $k_{\mathfrak{C}}$ [25,27,29]:

$$k_{\mathfrak{C}}\mathfrak{h} = \arccos\left[\frac{1-(k\mathfrak{h})^2/3}{1-(k\mathfrak{h})^2/6}\right], \tag{2}$$

which is valid for $k\mathfrak{h} \leqslant \sqrt{12}$, equivalent to about two nodes per wavelength; above this threshold the approximated waves become evanescent (see [29], Section 3.2s). The finite element wavenumber $k_{\mathfrak{C}}$ produces a different phase velocity $\frac{\omega}{k}$ for the Galerkin FE solution $\mathscr{S}_{FE}$ with respect to the exact solution $\mathscr{S}_E$. To better understand the mathematical behaviour of the FE solution, another function $\mathscr{S}_I$ is introduced: it is the linear interpolant of the exact solution $\mathscr{S}_E$ at the finite element nodes. The three functions $\mathscr{S}_E, \mathscr{S}_I$, and $\mathscr{S}_{FE}$ can be written as:

$$\mathscr{S}_E = \Re\left(e^{ikx}\right) \qquad \forall x \in \Omega, \tag{3}$$

$$\mathscr{S}_I = \Re\left(e^{ikn\mathfrak{h}}\right) \qquad \forall n\mathfrak{h} \in \Omega_{\mathfrak{C}}, \tag{4}$$

$$\mathscr{S}_{FE} = \Re\left(e^{ik_{\mathfrak{C}}n\mathfrak{h}}\right) \qquad \forall n\mathfrak{h} \in \Omega_{\mathfrak{C}}, \tag{5}$$

where Ω is the continuous domain and $\Omega_{\mathfrak{C}}$ represents the set of all the nodal coordinates of the FE mesh. Fig. 2 shows the real part of $\mathscr{S}_E$ (back, continuous line), $\mathscr{S}_I$ (red, dashed line on dots) and $\mathscr{S}_{FE}$ (blue, continuous line on dots) in the case of a purely tonal planar wave; the spatial resolution is $k\mathfrak{h} \simeq 1.0472$, that is, six elements per wavelength. The spatial coordinate, normalized per the wavelength, is the abscissa. In this example, $k\mathfrak{h}$ is independent of frequency f, and as a consequence $k_{\mathfrak{C}}\mathfrak{h}$ too. Therefore, results in Fig. 2 would hold at any wavenumber, but the spatial coordinate would have a different value in absolute terms (six wavelengths at frequency $f_1 \gg f_2$ are shorter than six wavelengths at f_2). Decreasing the product $k\mathfrak{h}$ would make $\mathscr{S}_{FE}$ closer to $\mathscr{S}_I$, and $\mathscr{S}_I$ itself closer to $\mathscr{S}_E$.

At this point, it is useful to introduce the H^1-seminorm $\|\cdot\|^*_{H^1}$, defined as the norm of the first distributional derivative of its argument; also, the space $V_{\mathfrak{C}}$ is the space of all the piecewise polynomial functions of order $\mathfrak{p}_{FE}$ defined over the FE volume (that

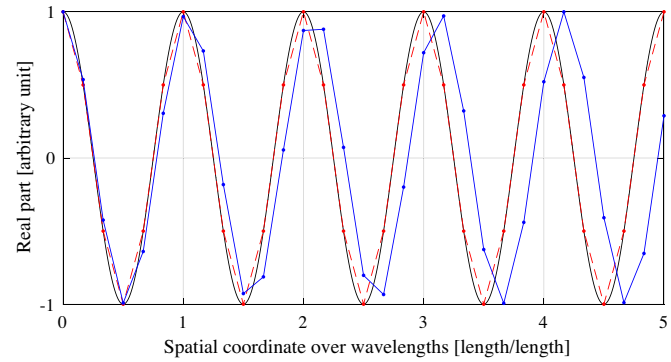

Fig. 2. The exact solution of Helmholtz equation $\mathscr{S}_E$ (black continuous line) and its interpolant with linear picewise polynomials $\mathscr{S}_I$ (red, dots on dashed line) propagate with the same wavenumber k, here in the one dimensional case. The Galerkin FE approximation $\mathscr{S}_{FE}$ (blue, dots on continuous line) propagates with a different wavenumber $k_{\mathfrak{C}}$.

comprises $\mathscr{S}_{FE}, \mathscr{S}_I$, and $\mathscr{S}_E$). The minimum error (in $\|\cdot\|^*_{H^1}$) between $\mathscr{S}_E$ and any element $\mathscr{S}_{V_{\mathfrak{C}}} \in V_{\mathfrak{C}}$ is the one between $\mathscr{S}_E$ and $\mathscr{S}_I$, which in literature is often called best approximation error [28]. On the other hand, the error between $\mathscr{S}_{FE}$ and $\mathscr{S}_I$ is due to the phase delay, and it builds up in space; this part of the error is usually called pollution error [25]. Ihlenburg and Babuska showed that there exists a supremum for the error of the Galerkin FE solution that comprises two parts: the best approximation error and another part related to the pollution error. They first find the supremum for the linear Galerkin FEM [29], and then extended the result to the Galerkin FEM with polynomial shape functions of generic order $\mathfrak{p}_{FE}$ [28]:

$$\|err\|^*_{H1} \leqslant C_1\left(\frac{k\mathfrak{h}}{2\mathfrak{p}_{FE}}\right)^{\mathfrak{p}_{FE}} + C_2 Lk\left(\frac{k\mathfrak{h}}{2\mathfrak{p}_{FE}}\right)^{2\mathfrak{p}_{FE}} \tag{6}$$

where C_1 and C_2 are problem dependent constants, L is the characteristic length of the problem. It appears clear from this formula that $\mathfrak{p}_{FE}$ plays a significant role in controlling the pollution error at high wavenumbers. It comes with no surprise that several articles are present in literature that deal with high order schemes for FE acoustics [30,31]

In higher dimensions, the situation is similar. Historically, analysis of the two dimensional Helmholtz equation [32,33] showed that the results of the above simple cases well explain the error in more complicated cases. Thereafter, rigorous investigations of the 2D and 3D problems appeared [34,35] and, finally, Ainsworth proposed a generalization in the multi-dimensional case [27]. The general behaviour in higher dimension does not change, but the FE solution exhibits an anisotropic behaviour. Furthermore, the pollution error becomes not avoidable, not even with the so-called stabilized method [36]. A clarifying introduction to the anisotropy issue is available in [25].

It is interesting to understand the effect of dispersion in case of reflected wave with an example. A train of stationary plane waves with the spectrum of a white noise propagates towards a rigid infinite plane with a direction inclined of 45°, Fig. 3 (a). A regular mesh of square elements discretizes the domain ($\mathfrak{h} = 1\,\mathrm{m}$), and an ideal boundary condition on the truncated boundary absorbs all the reflected waves (such boundary condition does not exist in reality: this scenario is ideal), Fig. 3 (b). Point $P^{(1)}_{FE}$ is reached by the train of planar waves and by its reflection on the rigid wall; the distance covered by the train that directly reaches $P^{(1)}_{FE}$ is half of the one covered by the reflected waves. Since the angle between the waves and the faces of the square FEs is 45°, the formula for $k_{\mathfrak{C}}\mathfrak{h}$ is [25]:

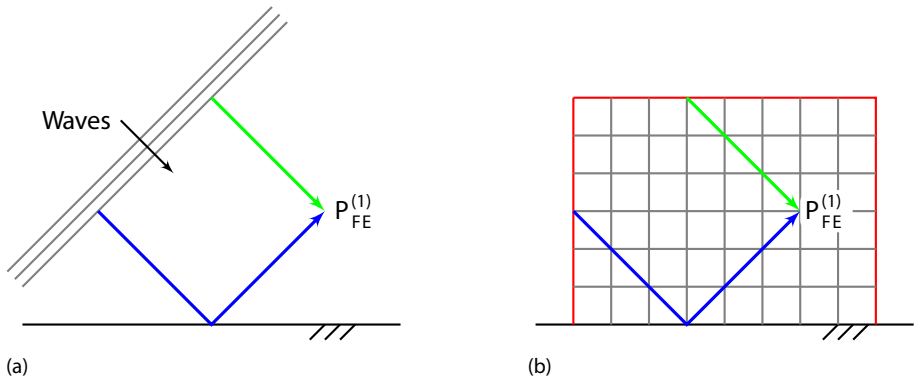

Fig. 3. A train of planar waves inciding on a rigid plane (a) is discretized with a regular FE mesh of square elements (b); $\mathfrak{h}$ is 1 m.

$$k_{\mathfrak{C}}\mathfrak{h} = \sqrt{2}\arccos\left[\frac{1-(k\mathfrak{h})^2/6}{1+(k\mathfrak{h})^2/12}\right]. \tag{7}$$

Fig. 4 shows amplitude (a) and phase (b) of the spectra of the exact solution and of the Galerkin FE solution at point $P_{FE}^{(1)}$. The effect of the pollution error is clear: in the spectrum, points of destructive interference shift towards higher frequencies. The higher the frequency, the more pollution error plays a role, as expected. The reason for this shift is simple: the two FE waves hit $P_{FE}^{(1)}$ with similar amplitudes, but with different phases in comparison with the exact solutions, causing this phenomenon. The phase presents the same frequency shift than the amplitude. These results are obtained with $\mathfrak{h} = 1$ m, but similar results would derive with any comparable setup.

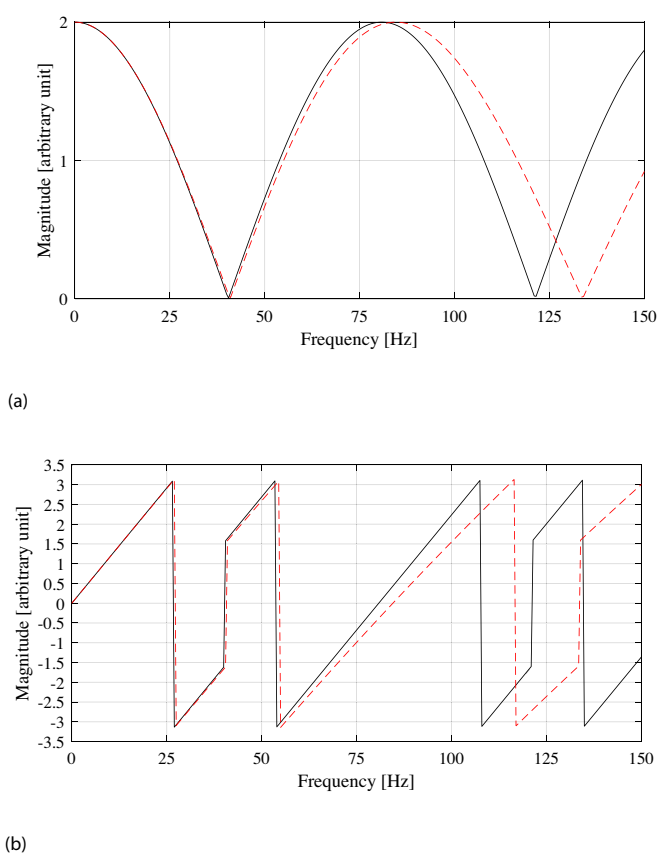

Fig. 4. Pressure amplitude (a) and phase (b) at point $P_{FE}^{(1)}$ (Fig. 3) calculated assuming a white noise source: exact (black continuous line) and Galerkin FE (dashed red line) solutions.

2.2. The PML

The solution of the Helmholtz problem is given by outgoing (scattered by the geometry and radiated by sources) and incoming (originating at infinity) waves. However, in the case of unbounded problems, the incoming waves do not have any physical meaning: they need to be removed. The Sommerfeld radiation condition provides a simple mathematical statement that ensures only outgoing waves are present in the solution. It can be written as [37]:

$$\lim_{\|\mathbf{x}\|\to+\infty} \|\mathbf{x}\|^{\frac{d-1}{2}}\left(\frac{\partial}{\partial\|\mathbf{x}\|} - ik\right)u(\mathbf{x}) \to 0 \tag{8}$$

in which the operator $\|\cdot\|$ is the Euclidean norm. Eq. 1 is derived without taking into account the Sommerfeld radiation condition, that is, for bounded acoustic problems; the introduction of this condition introduces non trivial complications.

First of all, in the FEM, the discretization is on the volume; hence, an unbounded geometrical domain requires an infinite number of elements, which is not a feasible solution. Therefore, the computational domain needs to be truncated at an appropriate artificial boundary Γ_A properly introduced, which divides the original unbounded domain into two regions: a bounded computational domain Ω_{FE} discretized with FE and an infinite residual region $\mathscr{R} = \mathbb{R}^d \setminus \Omega_{FE}$, with d dimension of the space. Subsequently, some techniques acting on Γ_A that satisfy the Sommerfield radiation condition are in charge of ensuring a solution to the problem. Intuitively, the most straightforward solution would be to use some boundary condition in the form of Eq. 8, but this one produces spurious reflections that seriously affect the quality of results (see [38], Section 3.3 and 10.3). Because of this reason, some alternative techniques that emulate the Sommerfeld radiation conditions without introducing spurious reflections are necessary. Accurate reviews of the most common methodologies for modelling the infinite region exterior to the artificial boundary are available in literature [39,40]. Here the perfectly matched layers (PML, [41]) are used.

Berenger originally introduced the concept of PML [42,43] for electromagnetic waves (that satisfy as well the Helmholtz equation). The idea is to add an exterior layer of finite thickness at an artificial interface such that outgoing plane waves are absorbed before reaching the outer layer truncation boundary. The mathematical formulation of the PML is such that the following properties hold [40]:

(1) no reflection occurs at Γ_A;
(2) the decay of the solution inside the layer is exponential.

There exist several formulation of the PML; for instance, the so-called split formulation of the PML method, originally developed in two dimensions and Cartesian coordinates [41]. An unsplit formulation exists too [44]. The interface in rectilinear Cartesian coordinates allows a tight fit around elongated objects, but can also be reformulated in spherical and other general curvilinear coordinates (see, for instance, [20]).

The PML converges to perfect wave absorption as the thickness of the layer is increased [45]. However, a compromise between a thin layer which requires a rapid variation of the absorption parameters and a thick layer which requires more elements is required [46]. Due to the dissipative nature of the PML, the role of the engineer is to choose an appropriate thickness that allows to dissipate enough energy so that the reflection error does not affect excessively the solution inside the FE volume. Therefore, the only two requirements for an effective PML are:

(1) truncated FE boundary Γ_{FE} located far enough from the sources, to reproduce a realistic field on Γ_{FE} itself;
(2) thickness of the PML enough to ensure a sufficient dissipation of energy, so to avoid excessive reflection error.

Under these conditions, PML offers some advantages in comparison with other techniques. First of all, it allows to treat properly non-homogeneous exterior domain. Secondly, No requirements on the shape of Γ_{FE} are present, allowing the use of any convex shape wrapped tightly around a limited conventional FEM domain [39]; the latter characteristic is extremely useful at high frequencies, because the engineers can make use of small computational volumes. However, since the PML does not appropriately propagate the acoustic field towards infinity (as this is the case, for instance, of PML), an integral formulation must be enforced to fulfil this aim. Several formulations, usually based on free field Green's functions, are available [47]: Lighthill's acoustic analogy [48], the Curle's formulation [49], the Ffwocs Williams and Hawkings formulation (FW-H, [50]) and Kirchhoff methods [51]. The FW-H formulation is the most widely used among these since it can take into account for noise generation effects due to moving surfaces immersed in a turbulent flow. An important feature of the FW-H technique is that it allows to place the truncated boundary relatively close to the sources [52]. As a consequence, the FE volume can be rather small, especially at high frequencies.

2.3. The reference model

The PML and the FEM are used to build the numerical models for this research thanks to the software Hypermesh [53] and Actran [20]. As a first step, an accurate model that will serve as a reference for the rest of this paper is build. This is called the *reference numerical model.*

First, the surface geometry is meshed with Hypermesh and refined in Actran, resulting in a surface mesh Ω_{start}. Quadratic serendipity elements (that is, without central node) are used.

The FE volume is modeled through the Exterior Acoustic component of Actran, which permits to automatize the creation of an adaptive mesh. A FE volume is created around the surface mesh Ω_{start}; the thickness of this volume t_{FE} is defined by the user. Then a PML mesh is automatically obtained on top of the extrusion itself. This tool allows using adaptive techniques. Table 3 shows the adaptive frequency bands chosen for the simulations. At low frequencies the bands are very narrow when compared with bands at higher frequencies. This is because the volume thickness is determined by λ_{max}, while the elements size by λ_{min}. The ratio between λ_{min} and λ_{max} must be kept high enough to avoid big volumes with small elements, which in turn would lead to high computational costs. This ratio changes faster at low frequencies than at high ones, explaining the higher density of adaptive frequency bands at low frequencies. For instance, $\lambda \simeq 17$ m at 20 Hz and $\lambda \simeq 7$ m at 52 Hz, resulting in a ratio between $\lambda_{min}/\lambda_{max}$ of about 0.41 over 32 Hz. A band between 20 Hz and 100 Hz would already lead to a ratio of about 0.2, and the resulting volume would require too many elements to be computed in an acceptable time. On the other hand, $\lambda \simeq 0.25$ m at 1500 Hz, and $\lambda \simeq 0.17$ m at 2000 Hz, with a ratio of about 0.68 spread over a band of 500 Hz, which is even higher than the one of the lowest frequency band.

Given the maximum and minimum wavelengths per adaptive frequency band λ_{max} and λ_{min}, the user must define (among others):

(1) gradient factor;
(2) FE volume thickness t_{FE} (absolute or with respect to λ_{max});
(3) minimum number of elements per λ_{min};
(4) PML thickness t_{PML} with respect to λ_{max}.

The gradient factor (GF) represents the maximum change in size between two adjacent finite elements; this parameter is left equal to the standard value (=2) throughout this whole research. The FE thickness has a clear impact on the computational size of the model, and different values also redistribute the nodes in a different way. The other two parameters affect the quality of the result as discussed in the previous section. For the reference numerical model, the chosen parameters per adaptive bandwidth are:

(1) FE volume thickness t_{FE} equal to 1 (maximum) wavelength;
(2) 6 quadratic elements per minimum wavelength;
(3) gradient factor equal to 2;
(4) PML thickness t_{PML} equal to 1 (maximum) wavelength.

Fig. 5 shows a cross-section of the mesh along the plane *xz* for the frequency band between 1504 and 2000 Hz, and also highlights t_{FE} and t_{PML}. In the figure, the FE volume is in red, while the PML is in grey; the geometry of the acoustic quarter-vehicle model is shown as well.

Environmental characteristics are set as following. Speed of sound is calculated by linearly interpolating results of [54] with a relative humidity of 50%, a temperature of 25°C and an environmental pressure of 101325 Pa. Air density is calculated according to the CIPM-2007 equation of state for moist air [55] with same setting used for speed of sound, and neglecting change of CO2 concentration with respect to dry air. A damping ratio of 10^{-3} has been set for the air. Simulations have been run in the frequency band between 20 Hz and 2000 Hz with a frequency step of 4 Hz. The latter one has been chosen in accordance with results found in literature. In particular, Nykänen et al. showed that binaural TFs with a

Table 3
Boundaries of the adaptive frequency bands.

Frequency bands [Hz]						
20–52	56–100	104–252	256–500	504–1000	1004–1500	1504–2000

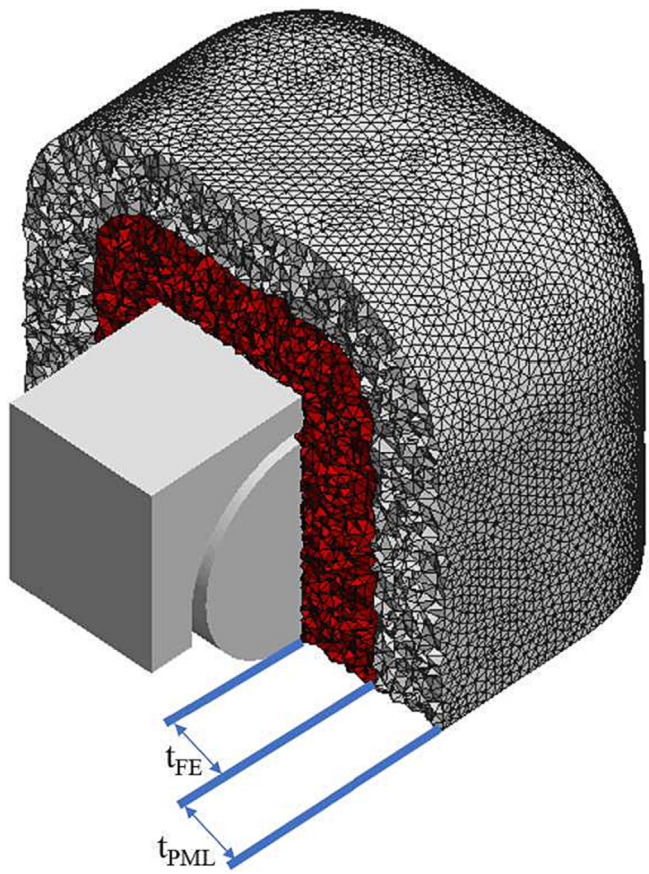

Fig. 5. A cross-section of the mesh of the reference model along the plane *xz* for the frequency band between 1504 and 2000 Hz. The FE volume is in red, while the PML is in grey; the geometry of the acoustic quarter-vehicle model is also shown. The thicknesses t_{PML} and t_{PML} are indicated.

resolution of 4 Hz are perceived as similar to artificial head recordings [16]. This result was later confirmed by Trollé et al. in the case of vibroacoustic applications [17]. The software used to run the simulations was Actran [20]; a convergence analysis has been previously carried out to ensure the correctness of the results; for sake of conciseness, this study is not illustrated in this paper.

2.4. Design of experiment

The next step requires to properly prepare and produce the others simulations. Recalling that the main purpose is to evaluate the perception of the numerical errors, a strategy must be defined including the definitions of:

(1) *factors*, namely those parameters that have a major influence on the accuracy of simulations;
(2) *levels*, namely the values assumed for each *factors*;
(3) *effects*, the qualitative or quantitative metrics used to derive conclusions.

Regarding *factors*, a summary of results from previous sections can provide a direction to the most appropriate choices. The order of polynomial shape functions $\mathfrak{p}_{FE}$ and the average size of the finite elements $\mathfrak{h}$ play a significant role in controlling approximation and pollution error. This two parameters are related together through the nodal resolution of the mesh $\mathfrak{N} = \frac{\lambda_{min}}{\mathfrak{h} \cdot \mathfrak{p}_{FE}}$ where λ_{min} is relative to the maximum frequency, and $\mathfrak{h} \cdot \mathfrak{p}_{FE}$ is the product of $\mathfrak{h}$ and $\mathfrak{p}_{FE}$. Two other key *factors* are the thickness of the PML and the thickness of the FE volume t_{FE}.

Another crucial aspect regards the strategy to test the error introduced by these *factors* and their *levels*. Since there is a lack of results in this field, the most reasonable approach is a *full factorial experiment*. In the present case, it consists in choosing some *levels* for each *factor*. Then, all the possible combinations of the *levels* across all *factors* are used to define the simulations.

Two *levels* L for each *factor* are chosen, giving raise to a so-called 2^n factorial experiment[1], where *n* is the number of *factors*. For each *factor*, *level* L = 2 represents the most accurate solution, with *level* L = 1 being a poorer choice. The *reference numerical model* has all the *factors* at *level* L = 2, thus being the most accurate simulation[2]. On the other hand, *level* L = 1 should be poor enough to introduce errors that make the simulated sounds perceptually different from the reference one. However, the difference should not be excessive; indeed, jury tests are conceived to analyse similarities (see next section), and the introduction of too different sounds might affect final results. Participants may use such extremely different sounds as the reference for the score "very different", consequently rating as "very similar" all the others (see chapters 4 for more details about the meaning of these scores). This fact would obviously be counterproductive. In the following, L[·] indicates the *level* of the generic *factor* ·.

The values of the *factors* at each *level* chosen for creating the numerical models are the following:

(1) L[$\mathfrak{N}$]=1 corresponds to a nodal resolution of four nodes per minimum wavelength, while L[$\mathfrak{N}$]=2 corresponds to twelve nodes per minimum wavelength;
(2) L[$\mathfrak{p}_{FE}$]=1 corresponds to an order of interpolation of the FE shape functions equal to 1 (linear FE) and L[$\mathfrak{p}_{FE}$]=2 corresponds to an order of interpolation equal to 2 (quadratic finite elements);
(3) L[$\frac{t_{FE}}{\lambda_{max}}$]=1 corresponds to a FE thickness equal to 0.4 times the largest wavelength and L[$\mathfrak{N}$]=2 corresponds to a FE thickness equal to one maximum wavelength;
(4) L[$\frac{t_{PML}}{\lambda_{max}}$]=1 corresponds to a PML thickness equal to 0.6 times the largest wavelength and L[$\frac{t_{PML}}{\lambda_{max}}$]=2 corresponds to a PML thickness equal to one maximum wavelength.

A remark about the nodal resolution of quadratic and linear models deserves special attention: models with L[$\mathfrak{N}$]=2 have six elements per minimum wavelength if they are quadratic (namely, if L[$\mathfrak{p}_{FE}$]=2) and twelve elements per wavelength if they are linear (L[$\mathfrak{p}_{FE}$]=1); models with L[$\mathfrak{N}$]=1 have two elements per minimum wavelength if are quadratic and four elements per wavelength if are linear. Each model comprises only one type of FE. The latter two values are rather low, and are used to introduce the numerical error into the TFs. Table 4 summarizes these details; in the following, the subscripts *min* and *max* are omitted where this does not create confusion. In the end, $2^4 = 16$ models have been created, and details of their *levels* for each *factor* are available in Table B.9.

To create realistic sounds, impulse response functions (IRFs) are calculated via inverse fast Fourier Transform (ifft function in Matlab); given the 4 Hz frequency step of the FRFs, the IRFs length is 0.25 s. Then, each of these IRFs is convoluted with a tire noise signal of 3 s recorded in real conditions; in detail, the signal is obtained from a much longer laboratory time recording of a 195/65R15 tire rolling on a drum equipped with a rough surface. The test is a coast-down from 120 to 20 km/h (constant decelera-

[1] The notation 2^k, common in literature, is not utilized to avoid confusion with the wave-number *k*

[2] In the theory of Design Of Experiment (DOE), each experimental run is called "test", see [56], page 1; here this nomenclature is not used, in order to avoid confusion; the word "simulation" is used instead.

Table 4
Factors and their levels L.

Factor	L = 1	L = 2
$\mathfrak{N} = \frac{\lambda_{min}}{\mathfrak{h} \cdot \mathfrak{p}_{FE}}$	4	12
$\mathfrak{p}_{FE}$	1	2
$\frac{\mathfrak{t}_{FE}}{\lambda_{max}}$	0.4	1
$\frac{\mathfrak{t}_{PML}}{\lambda_{max}}$	0.6	1

tion of 1 km/h/s) performed in a semi-anechoic environment. The part used for the convolution is between 51.5 and 48.5 km/h. Fig. 6 shows the amplitude of the source's power spectrum calculated with the Welch's method in Matlab [57]. The Welch's method was set on Hanning windows lasting 0.25 s with 50% overlap.

3. Qualitative analysis of numerical models

Numerical errors affect TFs, and consequently auralized sounds, due to the convolution. Therefore, it makes sense to split the analysis into two steps: first, focusing on TFs; second, taking into account the effect of the source.

The TFs depends on the quality of the numerical model. In this research, it is assumed that the quality of numerical model is sufficient to describe the TFs themselves. To assess the quality of numerical models, the response vector assurance criterion (RVAC, Sometimes called frequency response assurance criterion, FRAC, see [58,59]) is chosen.

This method consists in calculating the following correlation matrix **RVAC**:

$$\mathbf{RVAC}(f_i, f_{ref}) = \sqrt{\frac{|U_i(f_i)^H U_{ref}(f_{ref})|^2}{|U_i(f_i)^H U_i(f_i)||U_{ref}(f_{ref})^H U_{ref}(f_{ref})|}} \tag{9}$$

where U_{ref} is the vector of TFs calculated with the reference numerical model, U_i is the vector of TFs computed with the model to compare. H denotes the Hermitian transpose. Each element of the **RVAC** matrix is associated with two frequencies: one of the reference model (f_{ref}), and the other with the compared model (f_i). The magnitude of the elements represents the correlation between the two models at the given frequencies: a value equal to 0 indicates that the two vectors are orthogonal, a value of 1 indicates that the vectors are identical. Fig. 7 shows this matrix between the reference numerical model and the model with levels: $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 1$, $L\left[\frac{\mathfrak{t}_{FE}}{\lambda}\right] = 1$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda}\right] = 1$. The darkest blue areas represent points of high correlation.

Given two equal models, the diagonal of the **RVAC** matrix would be unitary; this diagonal is the red line in Fig. 7. Because

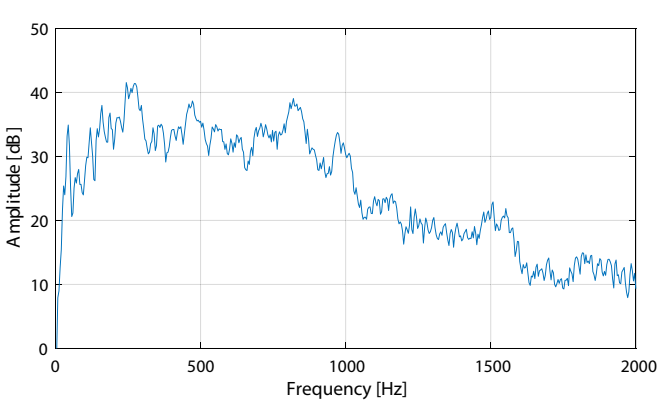

Fig. 6. Spectrum of the source signal (only the amplitude is shown).

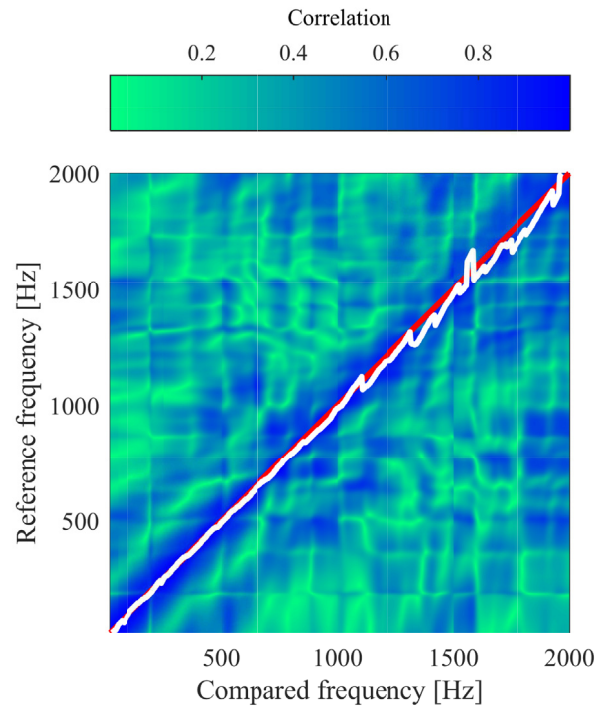

Fig. 7. **RVAC** matrix between of the model with levels: $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 1$, $L\left[\frac{\mathfrak{t}_{FE}}{\lambda}\right] = 1$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda}\right] = 1$.

of differences introduced by numerical models, the points of highest correlation (not necessarily equal to 1) do not lie always on the diagonal. The white line in Fig. 7 represents the points of highest correlation between the two models. The more this line of maximal correlation is far from the diagonal, the higher the frequency shift between the two models. Since the reference numerical model used as a reference is very accurate, a low correlation can be considered here as an index of poor numerical quality.

It is possible to group the models into four different sets according to their performance (in terms of *RVAC*) as a function of frequency, revealing also some effects of adaptivity. The four sets are:

(1) models with $L[\mathfrak{p}_{FE}] = 1$ and $L[\mathfrak{N}] = 1$ with a *RVAC* that decreases continuously, Fig. 8 (a). This set comprises models ranked 1, 2, 3 and 7 according to Table B.9;
(2) models with $L[\mathfrak{p}_{FE}] = 2$, $L[\mathfrak{N}] = 1$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda_{max}}\right] = 1$, with a good *RVAC* up to 1000 Hz, which is the maximum frequency of one of the adaptive bands, and a very poor one at higher frequencies Fig. 8 (b). This set comprises models ranked 4 and 5;
(3) models with $L[\mathfrak{p}_{FE}] = 2$, $L[\mathfrak{N}] = 1$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda_{max}}\right] = 2$, and almost constantly good, Fig. 8 (c). This set comprises models ranked 6 and 8;
(4) models $L[\mathfrak{N}] = 2$ with constantly good *RVAC*, Fig. 8 (d). This set comprises models ranked 9 and 16.

This grouping sheds light on two aspects. Firstly, the error of poor models increases together with the frequency: this effect is due to the dispersive behaviour of the FEM. Secondly, the two models with $L[\mathfrak{p}_{FE}] = 2$, $L[\mathfrak{N}] = 1$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda_{max}}\right] = 1$ behave worse than models with $L[\mathfrak{p}_{FE}] = 1$ and $L[\mathfrak{N}] = 1$, despite having higher order of polynomial shape functions (which should guarantee a higher

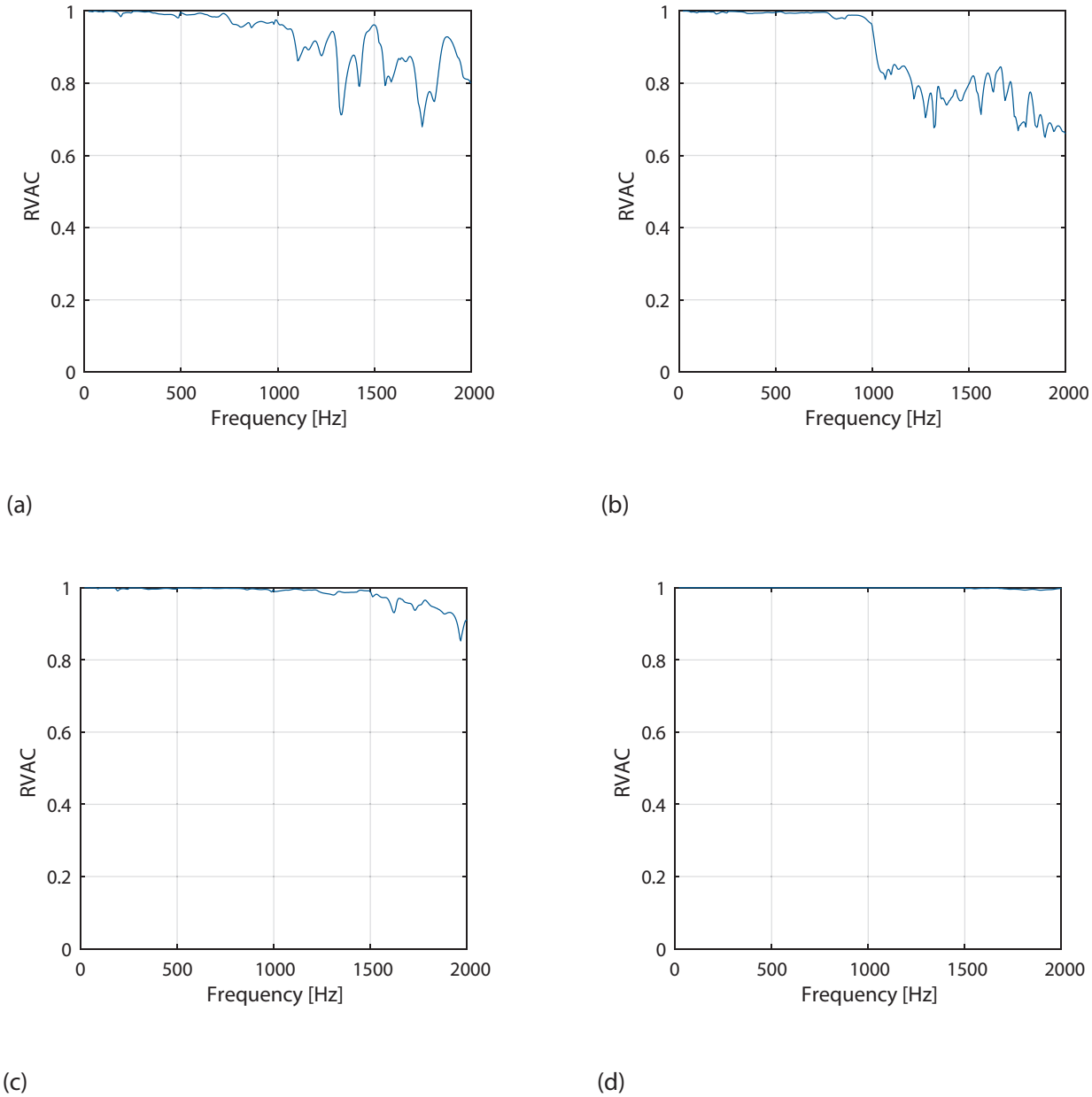

Fig. 8. Examples of *RVAC*; model factor levels: $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 1$, $L\left[\frac{t_{FE}}{\lambda}\right] = 1$ and $L\left[\frac{t_{pml}}{\lambda}\right] = 1$ (a), $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 2$, $L\left[\frac{t_{FE}}{\lambda}\right] = 1$ and $L\left[\frac{t_{pml}}{\lambda}\right] = 1$ (b), $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 2$, $L\left[\frac{t_{FE}}{\lambda}\right] = 1$ and $L\left[\frac{t_{pml}}{\lambda}\right] = 2$ (c), $L[\mathfrak{N}] = 2$, $L[\mathfrak{p}_{FE}] = 2$, $L\left[\frac{t_{FE}}{\lambda}\right] = 2$ and $L\left[\frac{t_{pml}}{\lambda}\right] = 1$ (d).

accuracy with respect to pollution and approximation errors); this lower accuracy depends on the thickness of the PML $L\left[\frac{t_{pml}}{\lambda_{max}}\right] = 1$ as explained in the following paragraphs. The exterior acoustic component of Actran does not allow setting the number of elements per wavelength in the PML; in terms of elements resolution, the user can control only the FE's number of elements per wavelength. The quadratic meshes used in this research have a number of elements per wavelength which is half of the one of the linear meshes, in order to maintain the nodal resolution $\mathfrak{N} = \frac{\lambda_{min}}{\mathfrak{h} \cdot \mathfrak{p}_{FE}}$. As a result, the quadratic meshes have a reduced number of elements in the PML. Nevertheless, the perfectly matched layer does not work with polynomial shape functions. At high nodal resolution ($L[\mathfrak{N}] = 2$), the PML mesh of quadratic FE models, despite being coarser than the linear ones, is still capable of performing correctly even with a thin layer. On the contrary, at low nodal resolution ($L[\mathfrak{N}] = 1$), to overcome the too coarse PML mesh, a thicker PML is needed.

It is possible to summarize these results by averaging the *RVAC* in the frequency domain. The resulting aritmetic average is indicated as $\overline{RVAC}$. Fig. 9 shows the $\overline{RVAC}$ in function of the size ranks of Table B.9, and highlights a certain trend between them; in the figure, the black line represents the degrees of freedom, and the dashed line with a circle represents the $\overline{RVAC}$. Not surprisingly, the eight biggest models are the ones with $L[\mathfrak{N}] = 2$ and are the most accurate ($\overline{RVAC}$ close to 1). Fig. 9 is missing one key aspect: the *RVAC* dependency on frequency as showed in Fig. 8. To overcome this limit, Fig. 10 shows four $\overline{RVAC}_{band-i}$ calculated by averaging the vector *RVAC* only in limited frequency bands i (together with the black line representing the degrees of freedom of the computational models):

(1) 20–500 Hz, blue continuous line;
(2) 504–1000 Hz, blue dashed line;

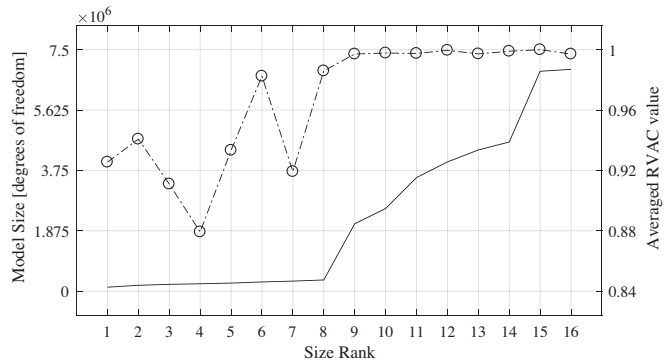

Fig. 9. DoF (black continuous line, left axis) and $\overline{RVAC}$ (right axis) in function of size rank in (a).

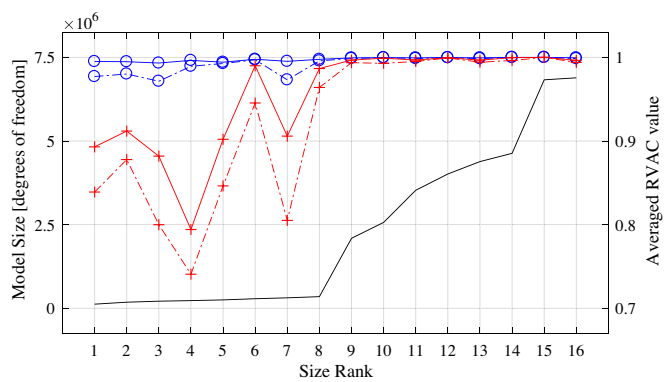

Fig. 10. DoF (black continuous line, left axis) and $\overline{RVAC}_{band-i}$ (right axis) averaged in four frequency bands (20–500 Hz: blue continuous line, circles; 504–1000 Hz: blue dashed line, circles; 1004–1500 Hz: red continuous line, crosses; 1504–2000 Hz: red dashed line, crosses) in function of size rank.

(3) 1004–1500 Hz, red continuous line;
(4) 1504–2000 Hz, red dashed line.

Fig. 9 shows how even models with poor settings perform rather good at low frequencies, but their quality quickly decreases in higher area of the spectrum, despite the use of adaptation techniques.

4. Jury tests

4.1. Test design

Engineers use different kinds of procedure for sound quality evaluation, depending on their purposes; a review is available here [60]. Here, the natural choice is a paired comparison method since it permits the evaluation of similarities between sounds in a fast but efficient way. Specifically, given a TP, all the auralized sounds can be compared with the *reference sound*, which is auralized with the *reference numerical model*.

This experiment does not require any peculiar skill. Also, the results of this research can be exploited virtually in any acoustic problem that involves simulation. Therefore, it is desirable to have untrained participants. In total, 18 people took part in this experiment, recruited among the university personnel. They reported no hearing issues. Table 5 summarizes the data and shows each participant's IDs that will be used throughout this chapter.

Sounds are presented over professional headphones for sound quality evaluation (Sennheiser HD 650). The listening environment was the sound-proof booth in Fig. 12, which has a background noise of 25 dB(A). Participants can control the graphic interface (GI) of the test on a screen with a mouse.

The sound samples are the ones described in the previous section. Each sound has a total duration of 3 s. To avoid excessive data-redundancy, only five transfer paths have been used for the jury tests. They have been selected to have a high variance of: metric and error variations, receivers, sources. The chosen TPs are: G-11, G + 10, S-21, S + 11, S + 20.

Each participant was asked to compare, for each TP, the reference sounds with a set of sixteen sounds. The latter included all the auralized sounds for that TP, including the reference sound itself. Each sound was evaluated twice to collect data for repeatability analysis. Furthermore, the test comprised five different sections, between which the participant was allowed to take a break if needed; each session corresponded to the study of sounds associated with one TP.

The test was delivered to participants via the ad hoc graphic interface in Fig. 11. By clicking on the button "(re)écouter", (french for re-listening), they could listen to a pair of sounds lasting 3 s and separated by a pause of 1 s. The pair consisted of the reference sound for the given TP and one of the other 32 sounds to evaluate. The participant could listen to a pair of sounds as many times as desired. The task was to move the horizontal scroll bar between "très similaires" (very similar) and "très différents" (very different). The resulting score was converted into a real value between 0 ("très similaires") and 10 ("très différents"). Thereafter, the participant could click on the button "suivant" (next).

The sounds were proposed in random order; this choice avoided that participants' evaluations could be influenced by sounds gradually different from the reference.

This kind of tests often requires a training phase. Namely, participants are allowed to listen to very similar sounds and very different ones, with the purpose to get an idea of what to expect during the test [60]. However, in the proposed test, there are five different sessions, and each would require a training portion; the time consumption required for these training could alter the comfort of participants, affecting the results of the test. The solution was to launch the test without any training. The researcher responsible for the test informed the participants that there was the reference itself among the sounds to compare (thus, giving an idea of when using the "very similar" rating).

Table 5
Data of the participants: identification number (ID), gender (Gen.) age in years, and nationality (Nat.).

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Gen.[1]	f	m	f	m	m	m	f	m	m	m	m	m	m	m	m	f	m	f
Age	30	27	27	28	23	26	23	34	56	23	37	32	26	26	26	25	29	25
Nat.[2]	I	I	F	S	F	F	F	Gh	F	F	Ch	I	Ge	F	I	C	C	C

[1] m = male, f = female.
[2] F = French, I = Italian, C = Chinese, Ch = Chadian, S = Senegalese, Gh = Ghanaian, Ge = German.

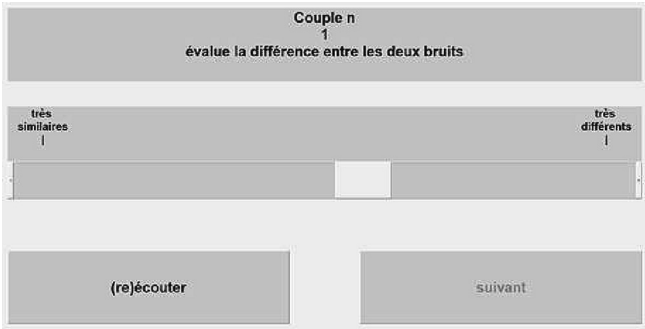

Fig. 11. Graphic interface of the jury test.

Fig. 12. Graphic interface of the jury test.

4.2. Data processing

First of all, an analysis to check repeatability is easily performed since each sound has been scored twice. For each participant p, the following repeatability index $\Im_p$ is defined:

$$\Im_p = \frac{1}{n}\sum_{i=1}^{n}\left|a_{1i,q} - a_{2i,q}\right| \tag{10}$$

in which $a_{1i,q}$ and $a_{2i,q}$ are the two similarity scores assigned to the i-th sample by the participant q, and $n = 16$ is the total number of different sounds used in each test. This index has values $0 < \Im_p < 10$, with 0 meaning that each sound has always obtained the same score, and 10 representing the maximum variation (each sound has been scored once 0 and once 10). The minimum $\Im_p$ has been 0.0446, while the maximum 4.1853, with most values between 1 and 3. Fig. 13 shows repeatability index $\Im_p$ for each participants and for each TPs; TP G-11 is in black circles, TP S-21 is in dark blue plus signes, TP G + 10 is in red asterisks, TP S + 11 is in magenta crosses and TP S + 20 is in light blue squares. Usually, in jury test analysis, repeatability is considered satisfactory if it is lower than two (on a 0–10 scale [60]), but this value needs to take into account the complexity of the test. In the performed test case, many participants exceed the threshold of 2, indicating that they faced significant difficulties. Interestingly, participants 8 and 9 consistently outperformed all the others; however, the subsequent analysis will explain that these two participants are not more precise than the others: instead, they are more conservative.

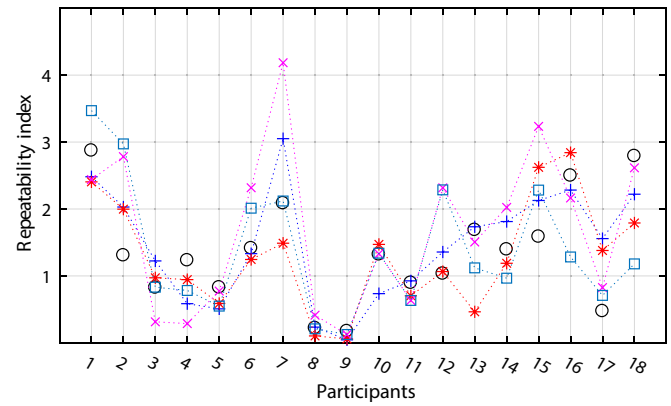

Fig. 13. Repeatability of each participant;; TP G-11 is in black circles, TP S-21 is in dark blue plus signes, TP G + 10 is in red asterisks, TP S + 11 is in magenta crosses and TP S + 20 is in light blue squares.

Fig. 14 (a) and (b) are two boxplots referring, respectively, to TPs G-11 and S + 11, showing the entire range of similarity scores used by participants. On each box, the central red mark indicates the median, and the blue box contains all values between the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, the latter being plotted individually using the red plus symbols. It is clear from these data that participants used different scales to evaluate the similarities between sounds. Surprisingly, some participants never used the minimum score, meaning that they did not recognize the equal sounds. This data reflects the difficulty of the test. It is worth noticing that participants 8 and 9 used a similar scale, much limited with respect to other people: this is the reason why their repeatability index $\Im_p$ is lower than the others in Fig. 13.

At this point, each pair of similarity scores $a_{1i,q}$ and $a_{2i,q}$ was substituted by its average $a_{i,q} = (a_{1i,q} + a_{2i,q})/2$. Then, data are standardized according to the following transformation:

$$a_{i,q} = \frac{a_{i,q} - \overline{a_q}}{\sigma_q}, \tag{11}$$

where $\overline{a_q}$ is the average of all the $a_{i,q}$ across the five TPs and σ_q is the associated standard deviation. Then, data consistency has been checked, exploiting the correlation matrice. It is found that participant 18 is always uncorrelated (or slightly negatively correlated) to the others, except in the case of TP G-11. Because of this reason, this participant has been removed from the following analysis (this subject may have misinterpreted the graphic interface of the test and assigned high scores to similar sounds). Besides participant 18, there is only one other case of negative correlation between participant 8 and 16 in the test on the G + 10 TP: they have a correlation of −6.45%.

Lastly, it is interesting to check whether a group of participants could be consistently clustered together across the different test; if yes, it would make sense to analyze it separately. Hierarchical clustering techniques have been used to answer this question. Specifically, Ward algorithm [61] based on euclidean distances has been used to cluster participants. A specific group of participants did not appear, and data can be analyzed altogether.

4.3. ANOVA

In the studied case, each transfer path represents a variable, each participant an observation. A comprehensive framework to serve these needs is the analysis of variance (ANOVA). The ANOVA requires stringent hypothesis [62]:

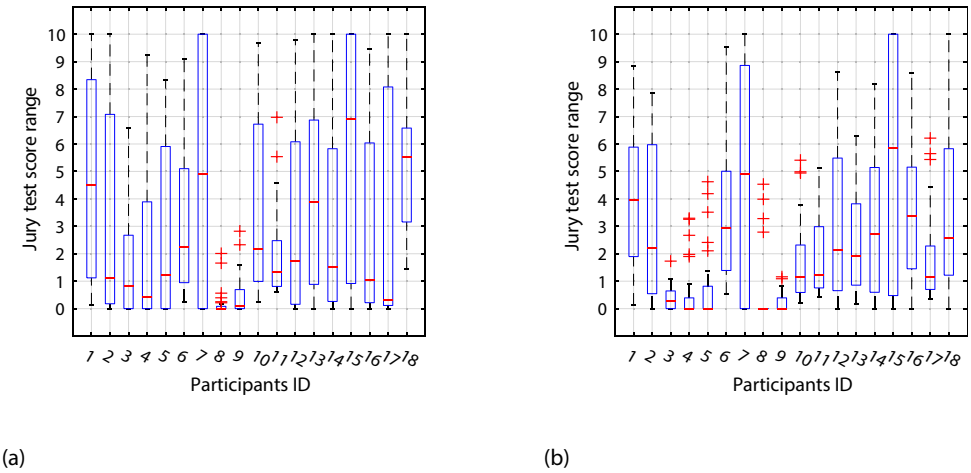

Fig. 14. Box plots showing the entire range of jury test scores used by participants in the case of TPs G-11 (a) and S + 11 (b).

1. independence of observations;
2. normal distribution of observations on the dependent variable in each group;
3. homogeneity of variance (co-variance matrix), also called homoscedasticity.

However, the robustness of ANOVA to non-normality and heteroscedasticity (namely, inhomogeneous variance) has been object of research for a long time. Recent results focused, among others, on skewness and kurtosis [63] to discuss normality, and on variance ratio (which is the ratio of the largest variance to the smallest of the groups, [64]) to discuss homoscedasticity. In general, limits for the variance ratio of 1.5, for the skewness of 2 and for the kurtosis of 6 are accepted [62]. There is not a uniform agreement on more extreme cases, but scientists regularly use these tools anyway. If these hypotheses are not satisfied, other approaches coming from non-parametric statistics are preferable, such as the Kruskal–Wallis test [65]. However, non-parametric approaches have the limits of loosing all the quantitative part of the information, being based only on ranks; for this reason, they are not considered in the following analysis.

Here, skewness and kurtosis reached levels higher than the recommended ones, depending on the considered subset. In details, if the data associated with a model and a TP had a critically high value of kurtosis (>6), the same was true for skewness (>2), and vice versa. Six datasets (each associated with a numerical model) reached such critical values: two models in the case of the TP G-11, one in the case of TP S-21, two in the case of TP G + 10, zero in the case of TP S + 11 and one in the case of TP S + 20. Even variance ratio did not satisfy the previously mentioned limits, being higher than 1.5.

Therefore, only the independence of observations is considered satisfied (although biases could be present due to the limited number of participants). However, given the small amount of subsets that do not satisfy the above mentioned criteria, the total dataset is considered sufficiently close to a normal one to perform an univariate ANOVA with multiple factors is performed. Besides the four numerical parameters used in the last chapter (that is,), also the TPs are considered as factors, bringing the total number to five. Two-factor interactions are considered as well. For each of the five factors and for each interaction, the H_1 hypothesis is that the factor or the interaction itself affects the perception of auralized sounds. Given the low number of participants, the H_0 hypothesis is rejected with a p-value of 1%.

Table 6
ANOVA test results for the five considered factors: p-values.

FACTOR	TP	$[\mathfrak{R} = \frac{\lambda}{\mathfrak{h} \cdot \mathfrak{p}_{FE}}]$	$[\mathfrak{p}_{FE}]$	$[\frac{\mathfrak{t}_{FE}}{\lambda}]$	$[\frac{\mathfrak{t}_{PML}}{\lambda}]$
p-value	<0.001	<0.001	<0.001	0.017	<0.001

4.3.1. Main effects

Table 6 shows the p-values associated to each factor. Interestingly, TPs play a critical role in this test. A quantitative analysis will provide more details later on. Regarding numerical parameters, they play a role in accordance with theory.The indices $\mathfrak{R}, \mathfrak{p}_{FE}$, and $\frac{\mathfrak{t}_{PML}}{\lambda}$ are important contributing factor since the p-value is zero. On the other hand, $\frac{\mathfrak{t}_{FE}}{\lambda}$ seems not to play any role.

In detail, the TPs with lower values are associated with acoustic samples ranked more similarly to the reference. In other words, simulation of the TPs with lower values produced more similar sounds in terms of acoustic perception, even with poor meshes; these TPs are called here *low-score* TPs: G + 10 and S + 11. On the other hand, *high-score* TPs produced sounds more different

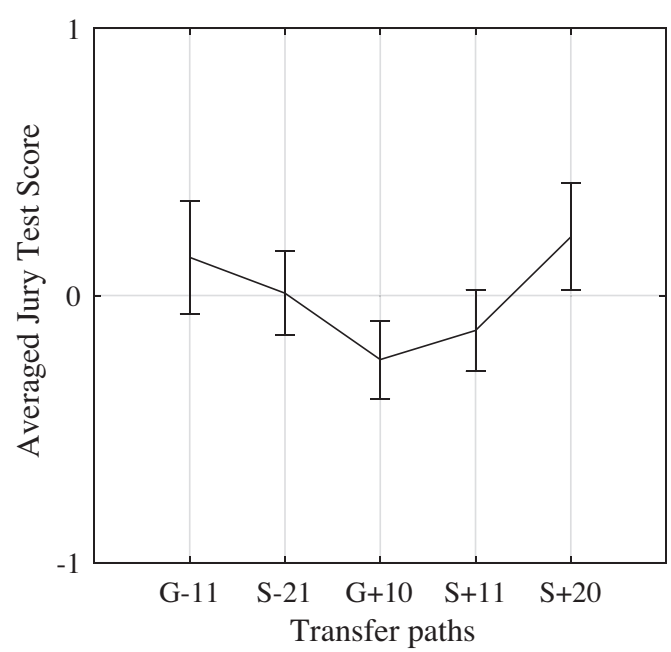

Fig. 15. Jury test score, in black (left axis), and $\overline{RVAC}$, in red (right axis), in the case of TP G-11 (a) and S + 11 (b); standard errors in the vertical bars, size ranks (see Fig. B.9) in the abscissa.

from the reference with poor meshes. The influence of the TPs reflect that the local properties of the FE errors: given a FE simulation, the local error can significantly vary when considering different points of the geometrical domain. Fig. 15 shows the ANOVA results; the error-bars represent the 99% confidence interval for a two-tailed distribution.

ANOVA results of the numerical parameters are shown in Fig. 16 (the error-bars are calculated as in Fig. 15) The parameter $\mathfrak{N}$, Fig. 16 (a), produced the biggest difference in terms of jury test score: it is by far the most critical parameter, at least in this frequency range. The parameters $\mathfrak{p}_{FE}$, Fig. 16 (b), and $\frac{\mathfrak{t}_{PML}}{\lambda}$, Fig. 16 (c), also impacted significantly the perception, but their impact is not quite as much as $\mathfrak{N}$. Lastly, the impact of $\frac{\mathfrak{t}_{FE}}{\lambda}$, Fig. 16 (d), is minimum.

4.3.2. Two-factor interactions

Table 7 shows *p*-values for two-factor cross-interactions.

It resulted that TPs interacted strongly with all other meaningful factors. This fact does not surprise: in *low-score* TPs, level 1 of numerical factors did not affect perception as much as in *high-score* TPs. In other words, this interaction reflects again the local properties of the error.

Table 7
ANOVA test results for the five considered factors: *p*-values.

INTERACTION	TP*[$\mathfrak{N}$]	TP*[$\mathfrak{p}_{FE}$]	TP*[$\frac{\mathfrak{t}_{FE}}{\lambda}$]	TP*[$\frac{\mathfrak{t}_{PML}}{\lambda}$]
p-value	<0.001	<0.001	0.560	<0.001
INTERACTION	[$\mathfrak{N}$]*[$\mathfrak{p}_{FE}$]	[$\mathfrak{N}$]*[$\frac{\mathfrak{t}_{FE}}{\lambda}$]	[$\mathfrak{N}$]*[$\frac{\mathfrak{t}_{PML}}{\lambda}$]	
p-value	<0.001	0.060	<0.001	
INTERACTION	[$\mathfrak{p}_{FE}$]*[$\frac{\mathfrak{t}_{FE}}{\lambda}$]	[$\mathfrak{p}_{FE}$]*[$\frac{\mathfrak{t}_{PML}}{\lambda}$]	[$\frac{\mathfrak{t}_{FE}}{\lambda}$]*[$\frac{\mathfrak{t}_{PML}}{\lambda}$]	
p-value	0.012	<0.001	0.012	

About numerical parameters, there have been strong interactions between all of them, with the exception of $\frac{\mathfrak{t}_{FE}}{\lambda}$. These cross-interactions are expected for two reasons. First of all, the factor $\mathfrak{p}_{FE}$ has a strong impact on pollution effect, while $\mathfrak{N}$ has a more significant impact on the approximation error; since both of these errors concur in the final quality of a simulation, it is natural to observe an interaction between these two factors. Secondly, the

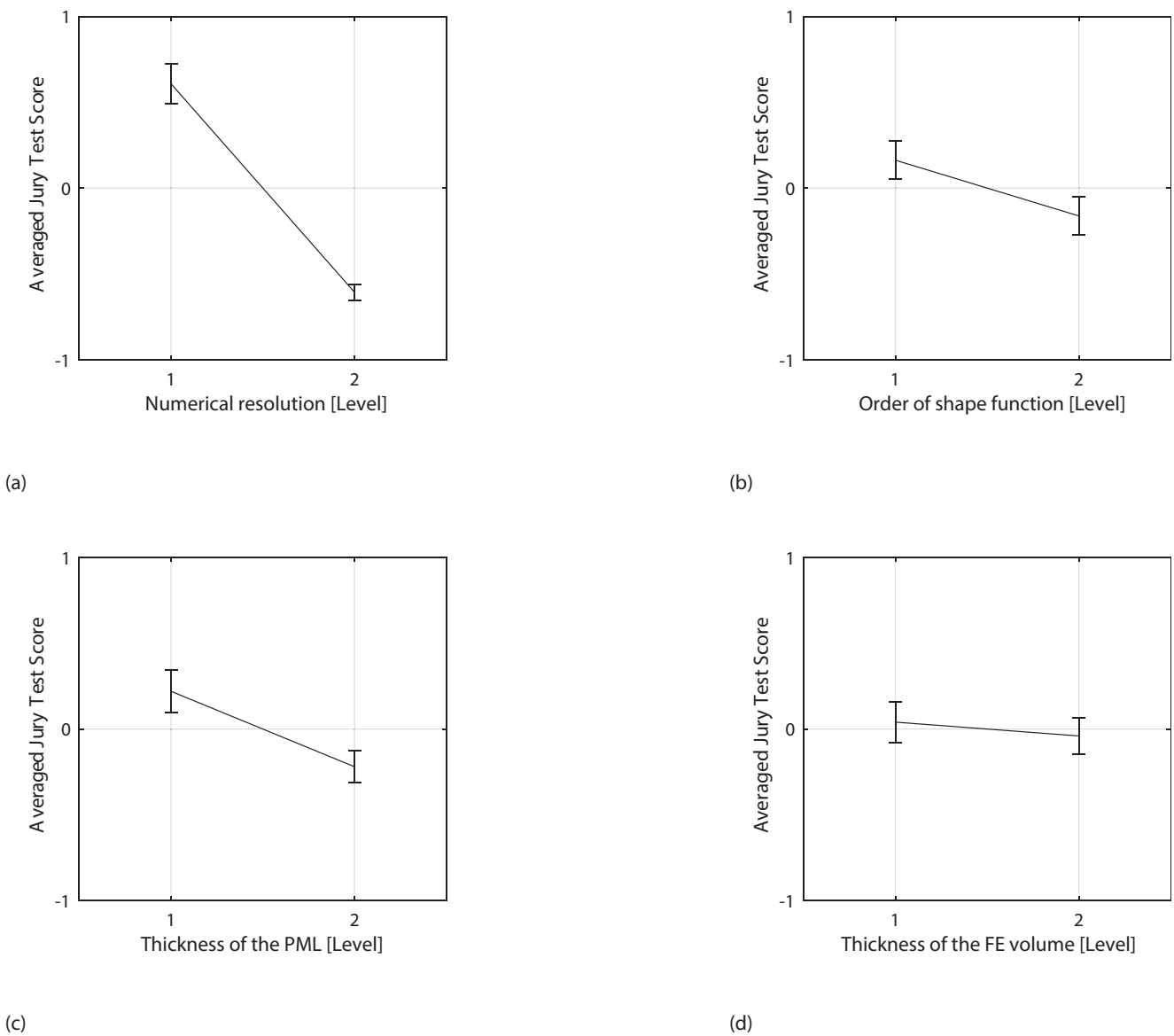

Fig. 16. Jury test score, in black (left axis), and $\overline{RVAC}$, in red (right axis), in the case of TP G-11 (a) and S + 11 (b); standard errors in the vertical bars, size ranks (see Fig. B.9) in the abscissa.

interaction between $\mathfrak{p}_{FE}$ and $\mathfrak{t}_{PML}$ is due to the PMLs of quadratic models being coarser than the ones of linear models. As explained in Section 3, this interaction takes place only at low numerical resolution, which explains the last interaction between $\mathfrak{N}$ and $\mathfrak{t}_{PML}$.

4.4. Ranking the models according to the jury-test scores

A posthoc analysis based on the Bonferroni test reveals that it is possible to define statistically separate groups in the four out of the five tested TPs with a p-value of 1% (a Tukey–Kramer test gave the same results): TPs G-11 (a), S-21 (b), G + 10 (c), and S + 20 (d). Fig. 17 shows the case of the TPs G-11 (a) and S + 11 (b); in the latter it was not possible to isolate statistically different groups; the jury test scores are in the ordinate axis, the size ranks (see Table B.9) in the abscissa. The red line is the RVAC (left y-axis), which will be discussed afterwards. Models are in ascending order based on the jury test score, and the error-bars represent the standard error. The standard error has been calculated for a two-tailed distribution with 15 degrees of freedom and a p-value of 0.01 (thus $t_{0.995}(15) = 2.947$) as:

$$err = \frac{\sigma}{\sqrt{n-1}} t_{0.995}(15),$$

where n is the sample size and σ is the standard deviation. Fig. C.23 in the appendix shows the other cases.

It appeared that the worst six models (in terms of similarity perception) are the same across the five TPs. Due to their consistently poor performances, these six acoustic samples are considered to be "different" from the reference in terms of auditory perception. Table 8 shows the models that produced such samples and their performances.

The results in Table 8 can be analyzed with the help of some graphs. Fig. 18, for instance, show the case of the transfer path G-11. The models in Table 8 (*different* models) are represented with a dotted red line, and are separated from the other models represented with a continuous black line (*similar* models). In Fig. 18 (a), the blue diamonds indicate the models with $L[\mathfrak{N}] = 2$; these models are all in the group of the *similar* ones. In other words, all the *different* models have $L[\mathfrak{N}] = 1$, which explicates why the nodal resolution is such an important factor in the ANOVA.

The other three graphs (b), (c) and (d) in Fig. 18 are built following the same principles, but refer to the other numerical factors.

In Fig. 18 (b), the blue diamonds indicate the models with $L[\mathfrak{p}_{FE}] = 2$ and represent only 1/3 of the *different* models; this explains why this parameter was important in the ANOVA, but not as much as the nodal resolution $\mathfrak{N}$. In Fig. 18 (c), the blue diamonds indicate the models with $L\left[\frac{\mathfrak{t}_{PML}}{\lambda}\right] = 2$, and the situation is exactly the same of Fig. 18 (b). Lastly, in Fig. 18 (d), the blue diamonds indicate the models with $L\left[\frac{\mathfrak{t}_{FE}}{\lambda}\right] = 2$, representing exactly 50% of the *similar* and *different* models, explaining why it does not play any role in the ANOVA.

The above analysis remains valid in the cases relative to the other TPs; indeed, although the order of the models in the abscissa can slightly vary, the groups of *similar* and *different* models are always the same.

4.5. Discussion of the jury test results

Two contrasting results can be observed: on the one hand, six sounds are consistently ranked the worst; on the other hand, they are not statistically different from the reference. However, because of the complexity of the test and the non-normal dataset, statistical results should be taken as hints. By taking into account such context, it can be suggested that models with a low $\overline{RVAC}$ (in this case, lower than 0.94) produce sounds that are perceived differently from the reference since they are consistently ranked the worst,

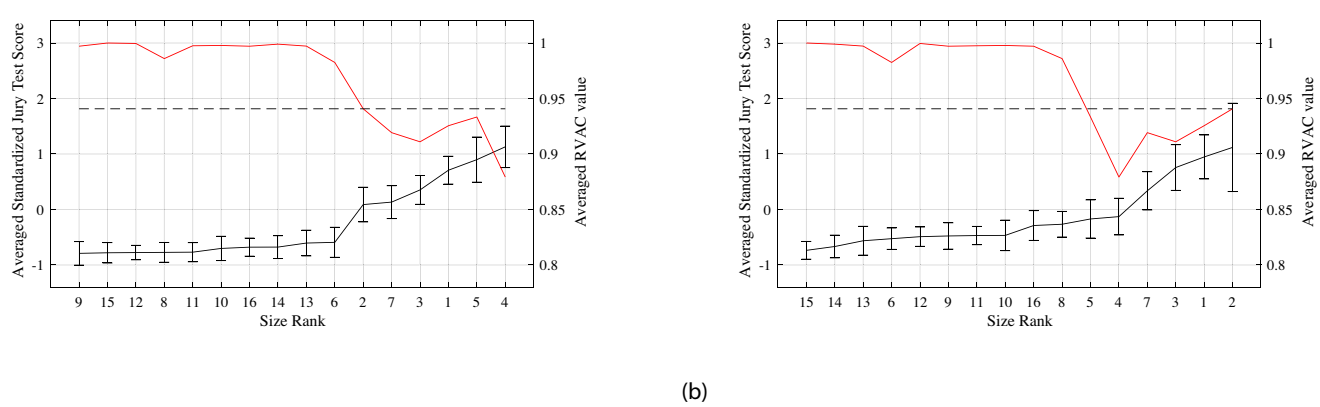

Fig. 17. Jury test score, in black (left axis), and $\overline{RVAC}$, in red (right axis), in the case of TP G-11 (a) and S + 11 (b); standard errors in the vertical bars, size ranks (see Fig. B.9) in the abscissa.

Table 8
Models producing sounds different from the reference ones.

Size rank	Model $L[\mathfrak{N}]$	$L[\mathfrak{p}_{FE}]$	$L\left[\frac{\mathfrak{t}_{FE}}{\lambda}\right]$	$L\left[\frac{\mathfrak{t}_{PML}}{\lambda}\right]$	Jury test rank G-11	S-21	G + 10	S + 11	S + 20
1	1	1	1	1	11	13	14	15	16
2	1	1	1	2	12	12	11	16	11
3	1	1	2	1	15	15	13	14	12
7	1	1	2	2	13	14	12	13	13
4	1	2	1	1	16	16	16	12	15
5	1	2	2	1	14	11	15	11	14

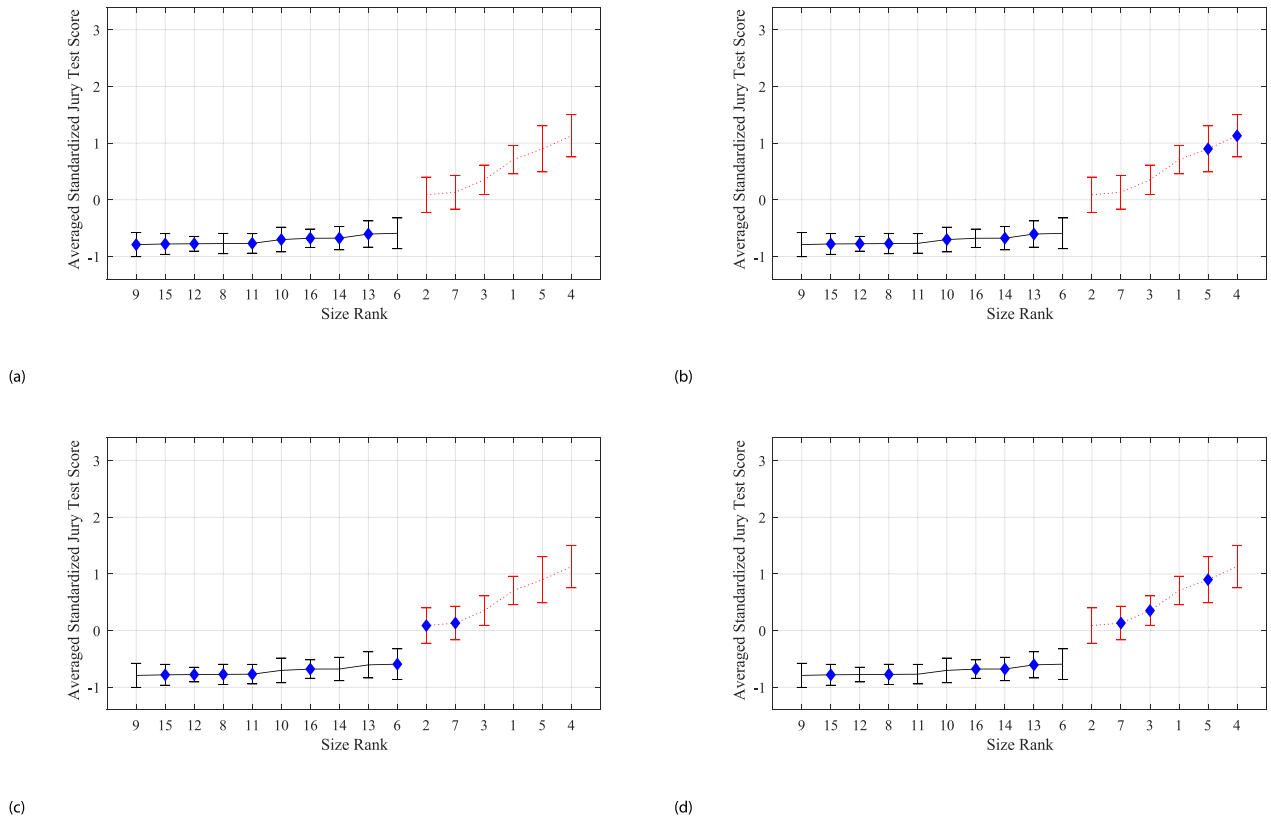

Fig. 18. Distribution of the numerical factors across the models in the case of TP G-11: the blue diamonds indicate the models with $L[\mathfrak{R}] = 2$ (a), $L[\mathfrak{p}_{FE}] = 2$ (b), $L\left[\frac{\mathfrak{p}_{PML}}{\lambda}\right] = 2$ (c), and $L\left[\frac{\mathfrak{p}_{FE}}{\lambda}\right] = 2$ (d); *similar* models in black continuous line, *different* models in red dotted line, standard errors in vertical errorbars.

even if in a different order. At the same time, models with high $\overline{RVAC}$ (in this case, higher than 0.98), are perceived very similarly since they are always ranked the best, but in a different order. If a similarity threshold for the $\overline{RVAC}$ exists, it should be between 0.98 and 0.94. Furthermore, the standard errors and the statistical distance between sounds can be interpreted as a strong signal of the difficulty faced by participants: it can therefore be hypotised that, although people could distinguish sounds auralized from models with low $\overline{RVAC}$, the differences are not so evident.

4.6. Criteria used by participants to distinguish noise samples

At this point, it is clear that participants could distinguish at least some noise samples, although with varying degrees of difficulties. A criticalstep is to understand which criteria participants used to distinguish the samples and if these criteria were different among the five TPs or were always the same. In order to do that, this section makes use of the acoustic metrics loudness N and sharpness S (calculated according to Zwicker [66].

As a first step, the following distances are defined:

$$d_N(s) = \frac{\max(N_s, N_{ref})}{\min(N_s, N_{ref})}, \quad d_S(s) = \frac{\max(S_s, S_{ref})}{\min(S_s, S_{ref})}, \quad s \tag{12}$$

where s identify the signal and the subscript *ref* identify the reference. No single metric proved to correlate well with jury test results. On the other hand, a linear combination of N and S can reproduce the perception quite satisfyingly. First, such regression has been calculated separately for each TP, using the entire set of data available, with a least square regression. The lowest value for the correlation was 93%. However, the coefficients for the regression calculated for different TPs were not similar. For this reason, a different approach has been used: a single dataset was built, comprising all the data. A least-square linear regression was then performed. In this case, correlation ρ in the case of each TP was calculated, giving the following valuess: 0.92 for the TP G-11, 0.98 for the TP S-21, 0.76 for the TP G + 10, 0.94 for the TP S + 11, and 0.92 for the TP S + 02; the correlation calculated over the entire dataset was 0.88 The coefficient of the linear regression where 24,61 (multiplicative factor of the loudness distance), 50,9 (multiplicative factor of the sharpness distance) and −74,52 (intercept term).

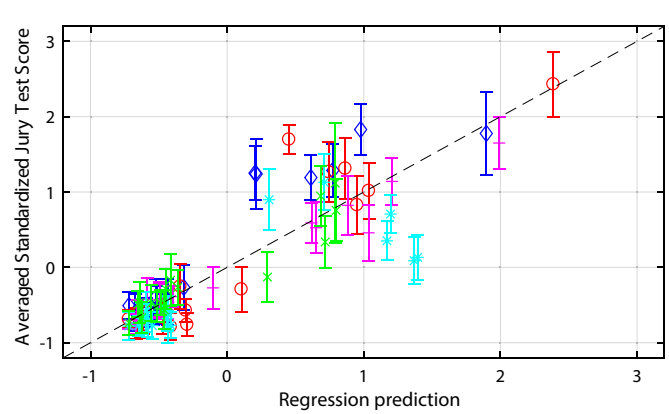

Fig. 19. Jury test scores vs. predictive model: S + 20 in blue diamonds, G-11 in red circles, S-21 in magenta plus signs, G + 10 in light blue asterisks, S + 11 in green crosses.

Fig. 19 shows the test results as a function of the predictive model. If the model was perfect, all the points would lie on the bisector (dashed black line in the graph). In the picture, colors and symbols identify the TPs: S + 20 in blue diamonds, G-11 in red circles, S-21 in magenta plus signs, G + 10 in light blue asterisks, S + 11 in green crosses. The figure shows that the points are not uniformly distributed:

(1) many points appear in the bottom-left corner, at low values of the score;
(2) a significant amount of points appear in the center of the graph, at medium values of the score;
(3) only few points appear in other areas.

Fig. 19 suggests that the three points in the top-right corner (blue diamond, magenta plus and red cross) may impact significantly the correlation of the model with the dataset. The correlation has been recalculated after rejecting these three points, resulting in 0.84. This value for the correlation is a strong index of the model's robustness.

The regressive model has a satisfying capacity to explain the jury test scores and proved to be robust. In other words, people mainly used a combination of loudness and sharpness to distinguish the noise samples. The combination of these two acoustic metrics appear to retain quite well the differences in terms of amplitude and frequency shift introduced by the numerical errors.

4.7. Linking numerical errors to acoustic metrics and to perception

The previous section showed that a regression comprising loudness and sharpness represents the perceived similarity with a correlation higher than 80%. Therefore, the link between these two acoustic metrics and the numerical errors complete this research.

In a first instance, it is worth noting that both loudness N and sharpness S depend on a quantity called specific loudness as following[66]:

$$N = \int_0^{24[Bark]} N\prime dz, \tag{13}$$

$$S = 0.11 \frac{\int_0^{24[Bark]} N\prime g(z) z dz}{\int_0^{24[Bark]} N\prime dz}, \tag{14}$$

where $N\prime$ is the specific loudness and the integral is computed across the entire audible spectrum (usually measured in Bark, see D). The specific loudness is a function of frequency that combines the spectral distribution of an acoustic input with the auditory filters of the human hearing system, and its unit of measure of $N\prime$ is sonie/bark. For more information about specific loudness and critical bands, refer to [66].

Because both loudness and sharpness depend on the integral of $N\prime$, it is crucial to evaluate the behaviour of the specific loudness of the noise samples across in the frequency domain and how it is related to the accuracy of the transfer functions, as in Fig. 20. In detail, Fig. 20 (a) shows two TFs associated to the TP S + 20, the reference one (red dashed line) and the one calculated with the following levels (blue continuous line): $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 2$, $L\left[\frac{\mathfrak{t}_{FE}}{\lambda}\right] = 2$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda}\right] = 1$. Fig. 20 (b) shows the spectra of the sound obtained via convolution with the source shown in Fig. 6. Lastly, Fig. 20 (c) shows the specific loudness of these sounds in function of the critical bands.

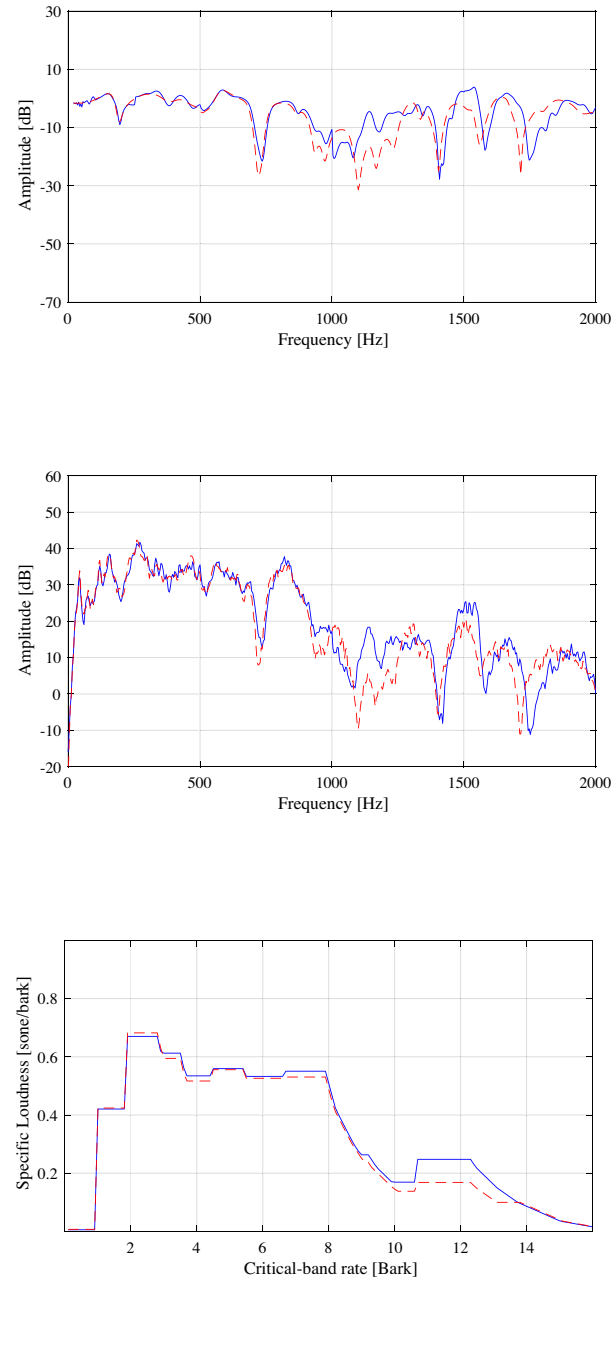

Fig. 20. Reference TF (red dashed line) and TF computed with levels (blue continuous blue line): $L[\mathfrak{N}] = 1$, $L[\mathfrak{p}_{FE}] = 2$, $L\left[\frac{\mathfrak{t}_{FE}}{\lambda}\right] = 2$ and $L\left[\frac{\mathfrak{t}_{pml}}{\lambda}\right] = 1$ in the case of TP S + 20 (a); spectra of the sounds auralized with these TFs in (b), and specific loudness $N\prime$ calculated on the spectra in (c).

As Fig. 20 shows, the changes in specific loudness are more important where there are major differences between the TFs. Since both the loudness N and the sharpness S depends on the specific loudness, the use of these metrics may help in summarizing the results. To understand the link between numerical accuracy and acoustic metrics, it is appropriate to understand how much such metrics change with respect to the reference. The differences are estimated as ratios between each i-th model and the reference:

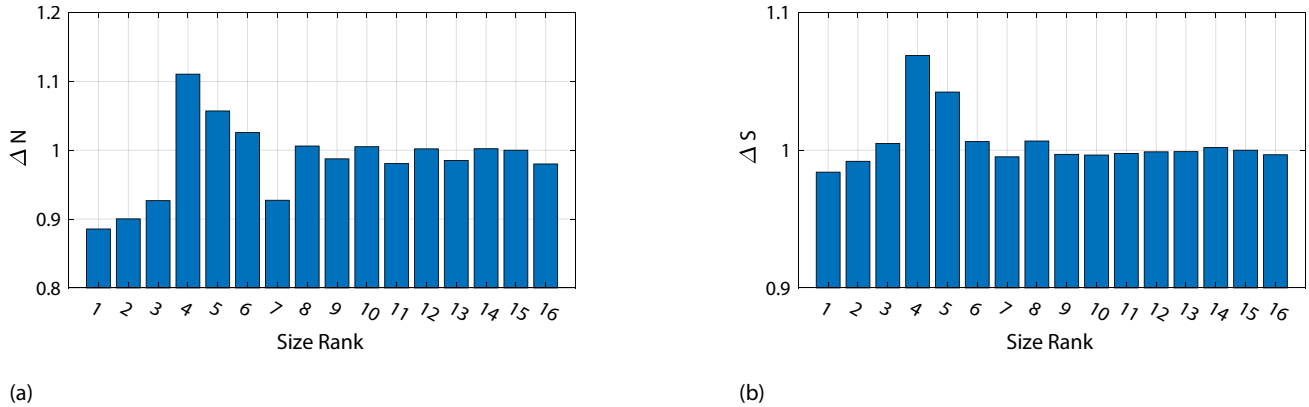

Fig. 21. ΔN_i (a) and ΔS_i (b) for the TP S + 20 as a function of the size rank (in the abscissa).

$\Delta N_i = N_i/N_{ref}$, $\Delta S_i = S_i/S_{ref}$; this approach is further justified by the fact that loudness ad sharpness define ratio scales. Moreover, it should be kept in mind that a just-noticeable-difference in loudness is around 10% [66]. Fig. 21 shows these variations in the case of TP rp02 for each model (represented by their size rank, see Table B.9), and are relatively small despite the numerical error introduced by the coarse meshes of the less accurate models, further reinforcing the concept that the sounds are not so distant in terms of perception.

It appears evident at this point why and how the perceived similarities are dependant on the accuracy and precision of the simulations.

5. Conclusions

In this research, sixteen different FE models have been used to simulate a set of sixteen TPs. One model has been used as the reference; then, other fifteen models have been built by degrading the mesh of the reference model. The quality of each model with respect to the reference has been assessed with the vector $RVAC(f)$ and the average of its values $\overline{RVAC}$. As expected, the vectors $RVAC(f)$ have lower values at higher frequencies. The analysis of the simulated acoustic signals has been carried out by means of jury tests. Participants feedback included the difficulties faced during the tests and the differences detected mainly at high frequencies. Analysis of data reflects the difficulties mentioned by participants; specifically, it was found that repeatability was higher than what usually found in industrial tests. It was not possible to find subset of participants that acted in a different way from the others with the methodologies used, except for one participants that had a negative correlation with almost all the others: for this reason, this participants has been removed from the dataset.

An univariate ANOVA shows that numerical errors are perceivable; the ANOVA also highlights the interaction between $\mathfrak{p}_{FE}$ and $\frac{t_{PML}}{\lambda}$. Results shows that $\mathfrak{N}$ and $\mathfrak{p}_{FE}$ play a critical role, while t_{PML} is especially important due to its interaction with $\mathfrak{p}_{FE}$ and t_{FE} did not play any specific role. These results are in line with the previous ones.

About similarity rankings, participants consistently evaluated six models (over sixteen) as the less similar from the reference. This consistence suggests that these acoustic signals are actually different, in perceptual terms, from the reference. In details, all the models with $\mathfrak{N} = 4$ and $\mathfrak{p}_{FE} = 1$ are part of this set of different sounds; the other two have $\mathfrak{N} = 4, \mathfrak{p}_{FE} = 2$ and $\frac{t_{PML}}{\lambda} = 0.6$. These six models are therefore considered different from the reference in perceptual terms.

It was found that a linear regression of loudness and sharpness has a high correlation (84%) with the perceived differences. Therefore, it was shown how the numerical errors affect these two metrics, thus providing a link between the accuracy of the FE models and the perception of the auralized sounds.

The most important result is probably about the difficulties that participants faced during jury tests. These difficulties indicate that acoustic signals were not so different. That is, low accuracy in the studied frequency range could be acceptable for sound quality purposes, at least for some tasks. More precisely, a nodal resolution of $\mathfrak{N} = 4$, when combined with thick PML ($\frac{t_{PML}}{\lambda} = 1$) and quadratic elements, produced noise samples very similar to the references. Only more degraded models were perceived as different from the references. Generalization of this results must be taken with cautions. As Anisworth showed [27], the accuracy of the FEM at low nodal resolution is unpredictable and erratic. In other words, similar experiments could in theory provide different results.

The results discovered in this research are auspicious and wide the door to a series of applications. To fully exploit the potentiality of this work, further research should focus on three areas: generalization of the results, practical applications, and extension of these results to different techniques.

The first way to generalize the results proposed here is to continue the same job with other examples and transfer paths. Besides using different TPs, future work could focus on some specific aspects:

(1) sources with different spectral characteristics which usually appear in engineering applications, in order to study what happens when the convolutions involve a source that introduces more high-frequency energy content and when there are tonal components in the source spectrum;
(2) studying internal problems, so to isolate the effect of pollution and approximation errors by removing the reflection errors caused by boundary conditions used to ensure the Sommerfield radiation condition; it is noteworthy that studies concerning the use of efficient methods for room acoustics exist, as previously mentioned, both for frequency and time-domain problems [12–15]: further developments of this research could concern the use of these formulations;
(3) studying external problems with a specific focus on techniques such as PML, IEM and others;

(4) extending the proposed results to higher frequencies, which are especially hard to simulate (especially the so-called mid-frequency range [11]); since this research indicates that lower accuracy might be sufficient for sound quality applications, the computational cost in the mid-frequency range may dramatically be decreased.

With respect to practical applications, an important fields of research is the study of the TFs accuracy for preference ratings problems due to their importance when comparing different TPs. Previous scientific work [18] shows that sounds auralized with heavily different TFs can preserve preference ratings. A natural extension of the research carried out during this paper moves in this direction: up to which limit does the numerical errors affect preference ratings? The difficulties that participants found during jury tests are very promising in this sense: a relatively low accuracy would likely prove to be sufficient.

Lastly, the framework proposed in this paper could be helpful for other techniques. Besides other boundary techniques to ensure the Sommerfeld radiation condition, model order reduction and substructuring techniques are crucial for practical applications. Furthermore, other deterministic methods such as the BEM and statistical techniques like ray tracing and SEA could be studied, together with all the hybrid methodologies based on them.

Data availability

The authors do not have permission to share data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was performed within the framework of the LABEX CeLyA (ANR-10-LABX-0060) of Université de Lyon, within the program "Investissementsd'Avenir" (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR). The authors gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie program through the ETN PBNv2 project (GA 721615).

Appendix A. Mock-up's multiview ortographic projections

Fig. A.22 shows the multiview ortographic projections of the virtual mockup used for this research; the light blue horizontal line is the infinite, rigid baffled plane.

Appendix B. Computational size

A high number of degrees of freedom (DoF) is related to an increased computational cost to solve the simulation. Due to the importance of the computational aspects in practical applications, it makes sense to evaluate the size of the matrices associated to the numerical models used in this research. Because of adaptivity, each

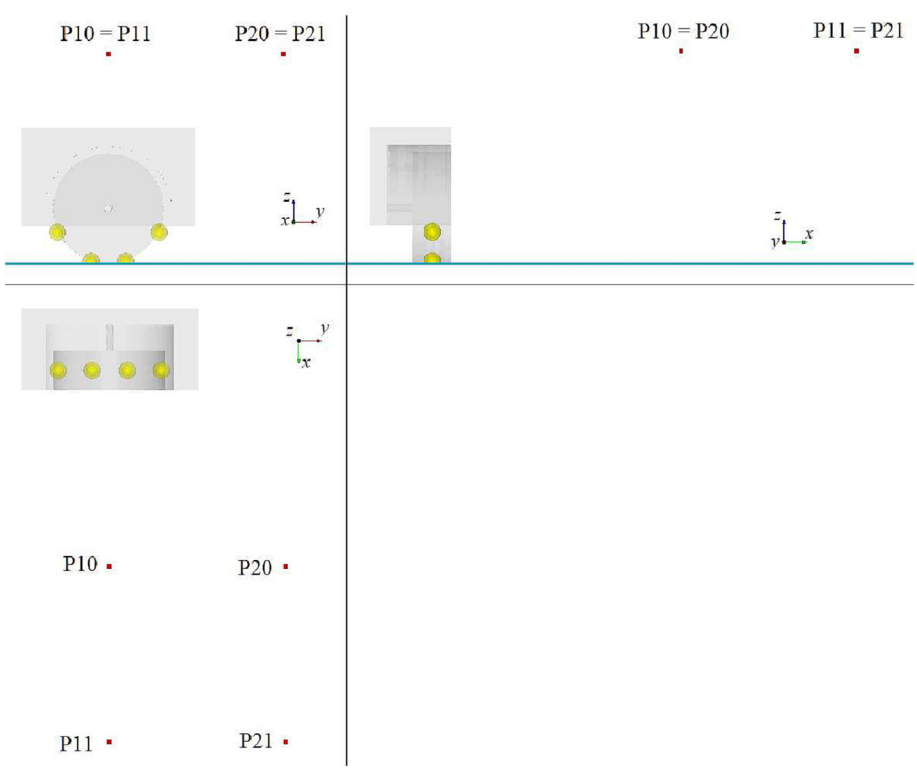

Fig. A.22. Mock-up's multiview ortographic projections; the light blue horizontal line is the infinite, rigid baffled plane.

Table B.9
Number of degrees of freedom per adaptive frequency band for each model and size rank.

$L\left[\mathfrak{R}=\frac{\lambda}{\mathfrak{h}\cdot \mathfrak{p}_{FE}}\right]$	$L[\mathfrak{p}_{FE}]$	$L\left[\frac{l_{FE}}{\lambda}\right]$	$L\left[\frac{l_{PML}}{\lambda}\right]$	N° of DoF								RANK
				20–52 [Hz]	56–100 [Hz]	104–252 [Hz]	256–500 [Hz]	504–1000 [Hz]	1004–1500 [Hz]	1504–2000 [Hz]	SUM	
1	1	1	1	12519	9649	14308	14284	21101	22174	28371	122406	1
1	1	1	2	20446	12566	22711	20940	32254	32203	40121	181241	2
1	1	2	1	26430	14842	28763	25254	39610	37746	38417	211062	3
1	1	2	2	43887	21036	37351	45984	57666	51837	54018	311779	7
1	2	1	1	16954	14900	19104	20727	29105	45224	82546	228560	4
1	2	1	2	24193	17437	26840	26916	40311	55852	93470	285019	6
1	2	2	1	31177	19764	34065	30758	43588	42603	47755	249710	5
1	2	2	2	47762	25208	50704	42766	61574	56501	62487	347002	8
2	1	1	1	192469	129341	224815	220474	381209	409307	533798	2091413	9
2	1	1	2	388728	192597	437617	382936	666792	655101	803823	3527594	11
2	1	2	1	529224	242256	564260	469812	815906	783546	980170	4385174	13
2	1	2	2	959061	389966	993873	777182	1299749	1129174	1340451	6889456	16
2	2	1	1	267314	194524	314328	314471	483768	501960	490207	2566572	10
2	2	1	2	475746	270949	537566	491095	796917	648834	792928	4014035	12
2	2	2	1	620356	320268	681518	579963	825358	753569	850889	4631921	14
2	2	2	2	957017	485461	1018774	776522	1269320	1111976	1210569	6829639	15

simulation comprises several meshes. Therefore, the computational size is calculated as the sum of DoF of all meshes of each model.

Table B.9 shows these data. The models are ranked in Table B.9 according to the sum of the numbers of degrees of freedom (last column). As a first guess, it is expected to have here a classification of the models, from the worst to the best according to the representativeness of the simulated sounds. The two metrics defined in the next section are introduced to check this first guess from an objective point of view. However, the ranking should not be taken as a strict metric of quality, but rather as an indicator. For instance, the reference numerical model is ranked 15th, while the biggest model has $L[\mathfrak{p}_{FE}] = 1$. This situation is due to the lack of central nodes in the reference model, causing a smaller number of degrees of freedom in the reference model, and due to the coarser PML mesh.

Appendix C. Jury test results - TPs: S-21, G + 10, S + 20

Fig. C.23 shows the results of the jury test in the case of TPs S-21 (a), G + 10 (b), S + 20 (c). The figure have to be read as Fig. 17:

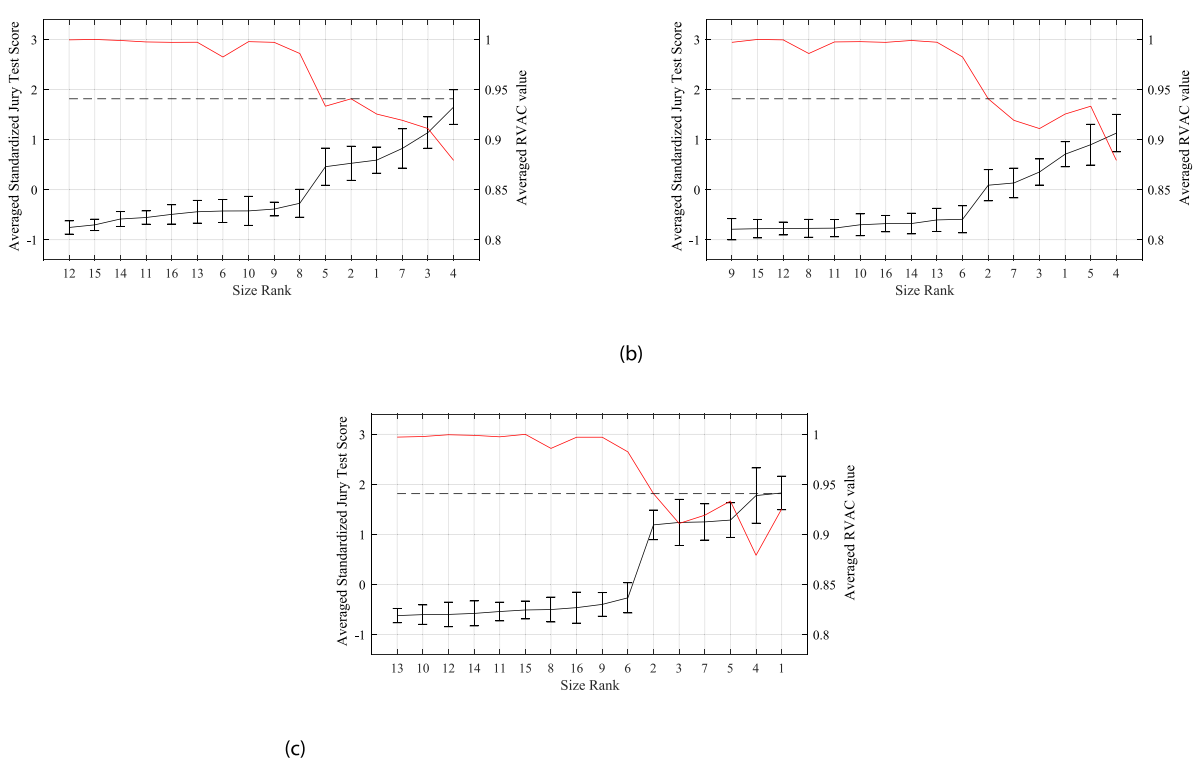

Fig. C.23. Jury test score with standard errors in the cases of S-21 (a), G + 10 (b), and S + 20 (c); size ranks in the abscissa.

Appendix D. Critical bands

Table D.10 shows the 24 critical bands measured in Bark.

Table D.10
Relation between critical-band rate z and the cut-off frequency of critical-bands f_c; more details in [66].

z [Bark]	f_c [Hz]	z [Bark]	f_c [Hz]	z [Bark]	f_c [Hz]	z [Bark]	f_c [Hz]
0	1						
1	100	7	770	13	2000	19	5300
2	200	8	920	14	2320	20	6400
3	300	9	1080	15	2700	21	7700
4	400	10	1270	16	3150	22	9500
5	510	11	1480	17	3700	23	12000
6	630	12	1720	18	4400	24	15000

References

[1] Fritschi L, Brown AL, Kim R, Schwela D, Kephalopoulo S. Burden of disease from environmental noise - quantification of healthy live years lost in europe. World Health Organization 2011:101–2.
[2] J. Blauert, U. Jekosch, Sound-quality evaluation - a multi-layered problem, Acta Acustica united with Acustica 83 (1997) 747–753(7).
[3] ISO 362–1:2015–01, Measurement of noise emitted by accelerating road vehicles – engineering method – part 1: M and n categories, 2015.
[4] Parizet E, Robart R, Pondrom P, Chamard J, Baudet G, Quinn D, et al. Additional Efficient Warning Sounds for Electric and Hybrid Vehicles. Wiley; 2016;1. p. 501–10. https://doi.org/10.1002/9781119307761.ch32.
[5] Vorländer M. Auralization: Fundamentals of Acoustics. Simulation, Algorithms and Acoustic Virtual Reality: Modelling; Springer; 2008. doi:10.1007/978-3-030-51202-6
[6] Mao J, Hao Z-Y, Zheng K, Jing G-X. Experimental validation of sound quality simulation and optimization of a four-cylinder diesel engine. J Zhejiang Univ Sci A 2013;130:341–52. https://doi.org/10.1631/jzus.A1300055.
[7] Xu Z, Xia X, Lai S, He Z. Improvement of interior sound quality for passenger car based on optimization of sound pressure distribution in low frequency. Appl Acoust 2018;130:43–51. https://doi.org/10.1016/j.apacoust.2017.08.019.
[8] Duvigneau F, Luft T, Hots J, Verhey JL, Rottengruber H, Gabbert U. Thermo-acoustic performance of full engine encapsulations - a numerical, experimental and psychoacoustic study. Appl Acoust 2016;102:79–87. https://doi.org/10.1016/j.apacoust.2015.09.012.
[9] Duvigneau F, Luft T, Hots J, Verhey JL, Rottengruber H, Gabbert U. Analysis of simulated engine sounds using a psychoacoustic model. J Sound Vib 2016;366:544–55. https://doi.org/10.1016/j.jsv.2015.11.034.
[10] E. Deckers, W. Desmet, K. Meerbergen, F. Naets, Case studies of model order reduction for acoustics and vibrations, De Gruyter, 2020, pp. 75–110. 10.1515/9783110499001-003.
[11] W. Desmet, Mid-frequency vibro-acoustic modelling: challenges and potential solutions, in: Proceedings of ISMA, volume 2, 2002, pp. 835–862.
[12] Okuzono T, Yoshida T, Sakagami K. Efficiency of room acoustic simulations with time-domain fem including frequency-dependent absorbing boundary conditions: Comparison with frequency-domain fem. Appl Acoust 2021;182. https://doi.org/10.1016/j.apacoust.2021.108212.
[13] Mukae S, Okuzono T, Tamaru K, Sakagami K. Modeling microperforated panels and permeable membranes for a room acoustic solver with plane-wave enriched fem. Appl Acoust 2023;182. https://doi.org/10.1016/j.apacoust.2021.108383.
[14] Okuzono T, Sakagami K. A frequency domain finite element solver for acoustic simulations of 3d rooms with microperforated panel absorbers. Appl Acoust 2018;129:1–12. https://doi.org/10.1016/j.apacoust.2017.07.008.
[15] Yoshida T, Okuzono T, Sakagami K. A parallel dissipation-free and dispersion-optimized explicit time-domain fem for large-scale room acoustics simulation. Buildings 2022;105. https://doi.org/10.3390/buildings12020105.
[16] Nykänen A, Johnsson R, Sirkka A, Johansson O. Assessment of the change in similarity judgements of auralized engine sounds caused by changes in frequency resolution of transfer functions. Appl Acoust 2011;72:115–23. https://doi.org/10.1016/j.apacoust.2010.10.007.
[17] Trollé A, Marquis-Favre C, Hamzaoui N. Auditory evaluation of sounds radiated from a vibrating plate inside a damped cavity: adjustment of the frequency resolution of vibro-acoustical computing. Acta Acustica united with Acustica 2012. https://doi.org/10.3813/AAA.918528.
[18] Nykänen A, Johnsson R, Sirkka A, Johansson O. Assessment of changes in preference ratings of auralized engine sounds caused by changes in frequency resolution of transfer functions. Appl Acoust 2013;74:1343–53. https://doi.org/10.1016/j.apacoust.2013.05.005.
[19] N. Aujogue, E. Parizet, Influence of the accuracy of transfer functions description on the perception of auralised sounds, in: SAE (Ed.), 10th International Styrian Noise, Vibration & Harshness Congress: The European Automotive Noise Conference, 2018–06-13. DOI: 10.4271/2018-01-1490.
[20] F.F. Technologies, Actran, 2019.https://www.fft.be/.
[21] Marburg S, Nolte B. A Unified Approach to Finite and Boundary Element Discretization in Linear Time-Harmonic Acoustics. Springer; 2008.
[22] Cohen G, Hauck A, Kaltenbacher M, Otsuru T. Different Types of Finite Elements. Springer 2008. https://doi.org/10.1007/978-3-540-77448-8.
[23] Bériot H, Prinn A, Gabard G. Efficient implementation of high-order finite elements for helmotz problems. Int J Numer Meth Eng 2016;106:213–40. https://doi.org/10.1002/nme.5172.
[24] Nettel S. Wave Phyisics. Springer 2009. https://doi.org/10.1007/978-3-540-87908-4.
[25] Harari I. Dispersion, Pollution and Resolution . Springer 2008 . doi: 10.1007/978-3-540-77448-8 .
[26] Harari I. Finite element dispersion of cylindrical and spherical acoustic waves. Comput Methods Appl Mech Eng 2001;190:2533–42. https://doi.org/10.1016/S0045-7825(00)00251-6.
[27] Ainsworth M. Discrete dispersion relation for hp-version finite element approximation at high wave number. SIAM J Numer Anal 2004;42:553–75. https://doi.org/10.1137/S0036142903423460.
[28] Ihlenburg F, Babuška I. Finite element solution of helmholtz equation with high wave number - part ii: the h-p version of the fem. J Numer Anal 1997;34:315–58. https://doi.org/10.1137/S0036142994272337.
[29] Ihlenburg F, Babuška I. Finite element solution of helmholtz equation with high wave number – part i: the h-version of the fem. Comput Math Appl 1995;30:9–37. https://doi.org/10.1016/0898-1221(95)00144-N.
[30] Harari I, Avraham D. High-order finite elements methods for acoustic problems. J Comput Acoust 1997;5:33–51.
[31] Semblat JF, Brioist JJ. Efficiency of higher order finite elements for the analysis of seismic wave propagation. J Sound Vib 2000;231:460–7. https://doi.org/10.1006/jsvi.1999.2636.
[32] Ihlenburg F, Babuška I. Dispersion analysis and error estimation of galerkin finite element methods for the helmholtz equation. Int J Numer Meth Eng 1995;38:3745–74. https://doi.org/10.1002/nme.1620382203.
[33] Babuška I, Ihlenburg F, Paik ET, Sauter SA. A generalized finite element method for solving the helmholtz equation in two dimensions with minimal pollution. Comput Methods Appl Mech Eng 1995;128:325–59. https://doi.org/10.1016/0045-7825(95)00890-X.
[34] Gerdes K, Ihlenburg F. On the pollution effect in fe solutions of the 3d-helmotz equation. Int J Numer Meth Eng 1999;170:155–72. https://doi.org/10.1016/S0045-7825(98)00239-4.
[35] Deraemaeker A, Babuška I, Bouillard P. Dispersion and pollution of the fem solution for the helmholtz equation in one, two and three dimensions. Int J Numer Methods Eng 1999;46:471–99. https://doi.org/10.1002/(SICI)1097-0207(19991010)46:4<471::AID-NME684>3.0.CO;2-6.
[36] Babuška I, Banerjee U, Osborn JE. Generalized finite element methods:main ideas, results, and perspective. Int J Comput Methods 2004;01:67–103. https://doi.org/10.1142/S0219876204000083.
[37] Schot SH. Eighty years of sommerfeld's radiation condition. Historia Math 1992;19:385–401. https://doi.org/10.1016/0315-0860(92)90004-U.
[38] D. Givoli, Numerical Methods for Problems in Infinite Domains, volume 33 of Studies in Applied Mechanics, Elsevier, 1992. First edition.
[39] Thompson LL. A review of finite-element methods for time-harmonic acoustics. J Acoust Soc Am 2006;119:1315–30.
[40] Givoli D. Computational Absorbing Boundaries. Springer 2008. https://doi.org/10.1007/978-3-540-77448-8.
[41] Bermúdez A, Hervella-Nieto L, Prieto A, Rodríguez R. Perfectly Matched Layers. Springer 2008. https://doi.org/10.1007/978-3-540-77448-8.
[42] Berenger J-P. A perfectly matched layer for the absorption of electromagnetics waves. J Comput Phys 1994;114:185–200. https://doi.org/10.1006/jcph.1994.1159.
[43] Berenger J-P. Three-dimensional perfectly matched layer for the absorption ofelectromagnetic waves. J Comput Phys 1996;127:363–79. https://doi.org/10.1006/jcph.1996.0181.
[44] Li Z, Andreas CC. A general approach for the development of unsplit-field time-domain implementations of perfectly matched layers for fdtd grid truncation. IEEE Microware Guided Wave Lett 1996;6:209–11. https://doi.org/10.1109/75.491508.
[45] S.V. Tsynkov, T.E., A Cartesian perfectly matched layer for the Helmholtz equation, Nova Science 2001, ISBN-13 978-1560729402, pp. 279–309.
[46] Turkel E, Yefet A. Absorbing pml boundary layers for wave-like equations. Appl Numer Math 1998;27:533–57. https://doi.org/10.1016/S0168-9274(98)00026-9.
[47] Kaltenbacher M, Escobar M, Becker S, Ali I. Computational Aeroacoustics based on Lighthill's Acoustic Analogy. Springer 2008. https://doi.org/10.1007/978-3-540-77448-8.
[48] Lighthill MJ. On sound generated aerodynamically. part i: General theory. London: Proceedings of the Royal Society of; 1952. p. 564–87.
[49] Curle N. The influence of solid boundaries upon aerodynamic sound. London: Proceedings of the Royal Society of; 1995. p. 505–14.
[50] J. Ffowcs Williams, D. Hawkings, Sound generation by turbulence and surfaces in arbitrary motion, in: R.S. of London (Ed.), Proceedings of the Royal Society of London, volume 264, 1969, p. 321–342.

[51] Farassat F. Acoustic radiation from rotating blades - the kirchhoff method in aeroacoustics. J. Sound Vibr. 2001;239:785–800.
[52] Brentner KS, Farassat F. Analytical comparison of the acoustic analogy and kirchhoff formulation for moving surfaces. AIAA J 1998;36. https://doi.org/10.2514/2.558.
[53] A. Hyperworks, Hypermesh, 2019.https://altairhyperworks.com/product/HyperMesh.
[54] Bohn D.A. Environmental effects on the speed of sound . J. Audio Eng. Soc. 36 : 1988 – 04 .
[55] Picard A, Davis RS, Gläiser K Fujii. Revised formula for the density of moist air (cipm-2007). Metrologia 2008;45:149–55. https://doi.org/10.1088/0026-1394/45/2/004.
[56] Montgomery DC. Design And Analysis Of Experiments. 8th ed. John Wiley & Sons Inc; 2012.
[57] Mathworks, Matlab, 2017.https://www.mathworks.com/products/matlab.html.
[58] J.F. Mercer, G.S. Aglietti, A.M. Kiley, Modal and frequency domain based techniques for finite element model correlation, in: 5th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineerig, Crete Island, Greece, 2015, pp. 191–208.
[59] C. Lein, M. Beitelschmidt, Comparative study of model correlation methods with application to model order reduction, in: Proceedings of ISMA 2014 including USD 2014, 2014.
[60] Otto N, Adams S, Eaton C, Lake S. Guidelines for jury evaluations of automotive sounds. Sound Vibr 2001:1–14.
[61] Jr JH Ward. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236–44. https://doi.org/10.1080/01621459.1963.10500845.
[62] Pituch KA, Stevenson JP. Applied Multivariate Statistics for the Social Sciences. Sixth edition. Taylor & Francis; 2016.
[63] Khan A, Rayner GD. Robustness to non-normality of common tests for the many-sample location problem. J Appl Math Decis Sci 2003;7:187–206. https://doi.org/10.1155/S1173912603000178.
[64] Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Effect of variance ratio on anova robustness: Might 1.5 be the limit? Behav Res Methods 2018;50:937–62. https://doi.org/10.3758/s13428-017-0918-2.
[65] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952;47:583–621. https://doi.org/10.1080/01621459.1952.10483441.
[66] Zwicker E, Fastl H. Psychoacoustics – Facts and Models. 2nd updated ed. Springer; 1999.