Perceptual assessment of simulated aircraft cabin noise in early design stages

Knuth, D.D.; Hüpel, Y.; Pockelé, J.S.; Merino Martinez, R.; Langer, S.C.

DOI
10.1121/10.0039865

Publication date
2025

Document Version
Final published version

Published in
The Journal of the Acoustical Society of America

NOVEMBER 17 2025

Perceptual assessment of simulated aircraft cabin noise in early design stages[a)]

Daniel Knuth ; Yannik Hüpel ; Josephine S. Pockelé ; Roberto Merino-Martínez ; Sabine C. Langer

J. Acoust. Soc. Am. 158, 3870–3886 (2025)
https://doi.org/10.1121/10.0039865

JASA ARTICLE

Perceptual assessment of simulated aircraft cabin noise in early design stages[a)]

Daniel Knuth,[1,2,b)] Yannik Hüpel,[1,2] Josephine S. Pockelé,[3] Roberto Merino-Martínez,[2,3] and Sabine C. Langer[1,2]

[1]*Institute for Acoustics and Dynamics, TU Braunschweig, Langer Kamp 19, Braunschweig 38106, Germany*

[2]*Cluster of Excellence SE²A—Sustainable and Energy-Efficient Aviation, TU Braunschweig, Germany*

[3]*Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, Delft, 2629 HS, The Netherlands*

ABSTRACT:

This study explores the potential of simulation methodologies in the early stages of the acoustic design of advanced air mobility cabins. With perceptual assessment as a priority, the approach includes conducting listening experiments based on the auralization of simulation results. For this, a cabin was simulated under the stochastic load of a turbulent boundary layer and auralized as representative cabin noise. The listening experiments investigated the impact of cabin parameter variations—specifically Young's modulus, skin thickness, and fluid bulk modulus—on the participants' perception and preferences. The findings show that utilizing the presented methodology within an early design scope produced audible differences for these parameter variations. With significant changes to the signals' preference probabilities, the proposed method is able to provide a better understanding and statistical depth to the cabin acoustic design process. Loudness and A-weighted sound pressure levels reliably predicted preferences, whereas other psychoacoustic metrics were of little significance, mainly due to the stochastic, stationary, and low-frequency characteristics of the noise samples. Furthermore, the position of the passenger within the cabin model significantly affected the preferences. Adding authentic cabin sounds to the auralizations did not significantly alter the parameter variations' preference distributions.

 https://doi.org/10.1121/10.0039865

(Received 24 March 2025; revised 6 October 2025; accepted 22 October 2025; published online 17 November 2025)

[Editor: Matthew Boucher] Pages: 3870–3886

I. INTRODUCTION

The noise inside an aircraft cabin is of great importance when considering passenger comfort and overall flight experience. Excessive cabin noise can cause discomfort and fatigue, with a direct link between sound pressure levels (SPLs) inside the cabin and passenger satisfaction.[1] Pennig *et al.*[2] report that passenger noise acceptance increases with lower noise levels and with improved noise characteristics. Their findings suggest optimizing cabin noise based on individual seat positions.[2] Consequently, reducing and improving cabin noise is a crucial objective in the design of modern commercial aircraft and advanced air mobility (AAM).

The Cluster of Excellence "Sustainable and Energy-Efficient Aviation" (SE²A), coordinated by the TU Braunschweig, is investigating technologies and aircraft configurations on the path toward sustainable and carbon-neutral air travel. The research effort is focused on three aircraft designs for short-, mid-, and long-range requirements, respectively, which include state-of-the-art developments toward efficient aviation.[3] Other than sustainability and energy efficiency, a holistic aircraft design for modern technologies needs to include several defining considerations, with reliability, economic viability, or regulatory compliance being only a few examples. Product acceptance belongs to these key factors for the success of aircraft concepts. Numerous elements play a role in this, with consumer well-being and comfort being influenced by available space, seating layout, and, as noted in literature,[1,2] the impact of cabin noise.

Introducing technologies for sustainability and efficiency, as well as developing AAM concepts, cannot omit the requirement of providing a comfortable and high-quality passenger experience. This highlights the need for advanced perception-influenced methodologies to assess and optimize cabin noise and its impact on passengers in the early stages of design.

An early design of the SE²A cluster's short-range configuration, which is the object of this study, is illustrated in Fig. 1. It is a propeller-driven, battery-powered electric aircraft with top-level requirements similar to ATR-72 (see Sec. II for details).[4,5] In the context of AAM, this configuration falls into the category of regional air mobility (RAM) with conventional takeoff and landing (CTOL). AAM is most commonly associated with urban air mobility (UAM) and vertical takeoff and landing (VTOL) capabilities; see,

[a)]This paper is part of a special issue on Advanced Air Mobility Noise: Predictions, Measurements, and Perception.

[b)]Email: d.knuth@tu-braunschweig.de

https://doi.org/10.1121/10.0039865

e.g., AAM demand analysis by Goyal *et al.*[6] or NASA AAM concept vehicles,[7] both focusing on aircraft classes with under 30 passengers, except for the NASA HECTR concept, which is designed for 76 passengers. However, cabin noise assessment methodologies in the early design stages of AAM concepts should ideally be applicable regardless of concept specifics.

Traditionally, cabin noise assessment has relied on empirical measurements conducted during flight tests or via physical mock-ups. Spehr *et al.*[8] conducted flight tests with more than 250 sensors placed in the cabin cross section upstream of the wings to measure and identify the main sources of cabin noise and their transfer paths to the passenger. Hu *et al.*,[9] performed flight test measurements, which included surface pressure sensors to measure the external turbulent boundary layer (TBL) and jet noise-induced fuselage excitation, in addition to microphones inside the cabin. Of the three major sources of aircraft cabin noise—the TBL on the fuselage, jet noise, and the air conditioning system—the first two are dominant and increase towards the rear of the cabin, whereas air conditioning had a minor impact on interior noise levels. An important application for cabin noise measurements is noise exposure and occupational hazard studies. For example, Zevitas *et al.*[10] evaluated aircraft cabin sound levels for existing measurements on 200 flights. The measurements represented different flight phases and six aircraft groups. Generally, measurements are effective but resource-intensive and time-consuming, and are not available during the early design stages.

Evaluation tools need to be developed to be as efficient as possible with the limited resources available in early development processes, where geometries and parameters are uncertain, and several iterations are required. Ever since the middle of the 20th century, the simulated noise assessment of cabin noise has become increasingly important.[11] In recent years, the research has focused more on modeling the physical properties of aircraft domains and the accurate material behavior.[12,13] Furthermore, the efficient solution of the arising large-scale models has also been the subject of ongoing research.[14,15] This also includes the process of creating a fully-automated modeling and simulation chain, which allows a vibroacoustic model creation from aircraft configuration data.[12]

FIG. 1. Early visualization of the short-range aircraft considered in this study (Ref. 3).

In order to evaluate the perception of the simulated cabin noise, additional depth needs to be added with a psychoacoustic analysis and listening experiments, which introduce the requirement of representative and realistic cabin noise auralizations. Auralization, the process of rendering auditory experiences from simulation data, has emerged as a powerful tool in the scope of perception-influenced assessment.[16]

Pieren[17] and Pieren *et al.*[18,19] successfully applied auralization techniques to investigate various environmental acoustical sceneries, including road traffic, railway, and wind turbine noise, as well as aircraft flyover noise. Schade *et al.*[20] also analyzed auralized sounds to evaluate the perceptual influence of rotational speed fluctuations in an AAM vehicle with distributed ducted fans. Another recent application in environmental noise combined synthetic sound auralization and psychoacoustic listening experiments to evaluate the performance of different noise reduction measures for wind turbines.[21] Furthermore, Pockelé[22] and Bresciani[23] demonstrated fully-modeled auralization approaches for wind turbine noise, with applications in human perception.

Regarding interior noise of flight vehicles, Allen and Krishnamurthy[24] published work on modeling and auralizing a six-passenger flight vehicle. Utilizing a hybrid approach, the interior sound field was computed with finite element analysis for lower frequencies and statistical energy analysis for higher frequencies. The auralization was achieved with the NASA Auralization Framework.[25] Other examples of interior vehicle noise are comparisons of different car audio systems,[26] or auralizations of vehicle interior tire-road noise.[27]

The common aim of applying auralization in acoustic research and design is to extend or even shift classical evaluation methods based on conventional sound metrics toward perception-based considerations. By converting acoustical data into audible sound, subjective feedback can be gathered from potential passengers or stakeholders for psychoacoustic evaluations without the need for physical prototypes.

In light of the existing methodologies and advancements in simulation techniques, this study investigates the potential of combining simulation and auralization techniques in the early stages of the acoustic design of aircraft or AAM vehicle cabins. The proposed concept is proven through the use of various listening experiment results. The findings aim to highlight the efficacy of this approach in providing detailed insights into the psychoacoustic impacts of cabin noise variations, therefore contributing to an early aircraft design process.

The outline of this contribution is as follows: First, the methodology is detailed in Sec. II, consisting of the aircraft cabin modeling and simulation, the auralization framework, and the models used for calculating psychoacoustic metrics. Applying the methodology, Sec. III starts with the

auralization of the stimuli and focuses on the listening experiment, detailing the test environment, the sound reproduction setup, and the experimental design itself, with all relevant questionnaires. Next, Sec. IV describes the participant group with relevant statistics and discusses potential biases. Finally, Sec. V highlights the results of the listening experiment and discusses some of the most prominent findings with regard to cabin noise perception and correlation with psychoacoustic metrics. In the end, a conclusion and future ideas are presented.

II. METHODOLOGY

A. Cabin model and simulation setup

The starting point of this methodology, geared towards perceptually assessing an AAM vehicle concept, is building a model and computing the sound pressure in a simulation. This section describes the aircraft model and the simulation procedure used in this contribution.

As mentioned in the introduction, the object of this study is a propeller-driven, battery-powered electric aircraft with performance requirements similar to ATR-72. It falls into the RAM category of AAM. The detailed model design, including a sensitivity analysis of the aircraft characteristics to technology advancements such as battery energy density, is provided by Karpuk and co-worker.[4,5] Table I summarizes the aircraft's mission requirements.

TABLE I. SE^{2}A short-range configuration top-level requirements (from Refs. 4 and 5).

Requirement	Value	Units
Passengers	72	
Design range for max. payload	926	km
Max. payload	7500	kg
Cruise Mach number	0.42	
Service ceiling	7620	m
Take-off field length	1367	m
Landing distance	915	m

The simulation of this aircraft needs to work with varying degrees of detail and yield meaningful results. To address the latter, the wave-resolving finite element method (FEM) is utilized to compute the sound pressure inside the cabin. These simulations are computed with our in-house code *elementary Parallel Solver (elPaSo)*,[28] which is designed for high-efficiency frequency-domain simulations. This allows for the computation of the spatial SPL distribution within the whole aircraft in a frequency domain up to 1000 Hz. While higher frequencies are possible, the computational demand increases drastically.

To address the degree of detail of the model and model fidelity, limitations are set due to the available data in early design and the short times between iterations. The simulation model is based on Blech.[29] A high-fidelity approach is chosen for the material models and mesh resolution, and a low-fidelity approach is chosen for the cabin geometry. To further increase the efficiency of the simulations and enable parametric variations within a reasonable time frame, only a segment of the cabin with symmetry boundary conditions is considered in the following—the modeled cabin segment is 3 m in length, 1.96 m in height, and spans approximately 1.26 m from the symmetry boundary to the cabin's wall. The examined domain includes five seating rows. These model dimensions emphasize that the simulation methodology is applicable for both larger RAM vehicles and smaller UAM vehicles (the latter case could be efficiently evaluated as a full-scale model).

Cropping the model is expected to change the modal behavior of the cabin, and a symmetric excitation of the cabin is not realistic; however, both are considered a necessary trade-off for less computational effort. Due to the flexural bending of the airframe panels being the main excitation source, it is not expected that any unwarranted artifacts arise from the lengthwise truncation of the model. The symmetry boundary condition is implicitly fulfilled, and the model is not conditioned with any other boundary conditions. Figure 2 shows an exploded view of the cabin model and provides details on the modeling assumptions for the aircraft segment.

The prestressed airframe is made up of carbon fiber reinforced plastics (CFRP). From there, the excitation is propagated into the glass wool insulation, which is modeled as an equivalent fluid using the limp Johnson-Champoux-Allard (JCA-limp) approach. The interior lining is comprised of two surfaces, again modelled with 9-noded shell elements and made up of glass fiber reinforced plastics (GRFP), while its core is modeled as a three-dimensional

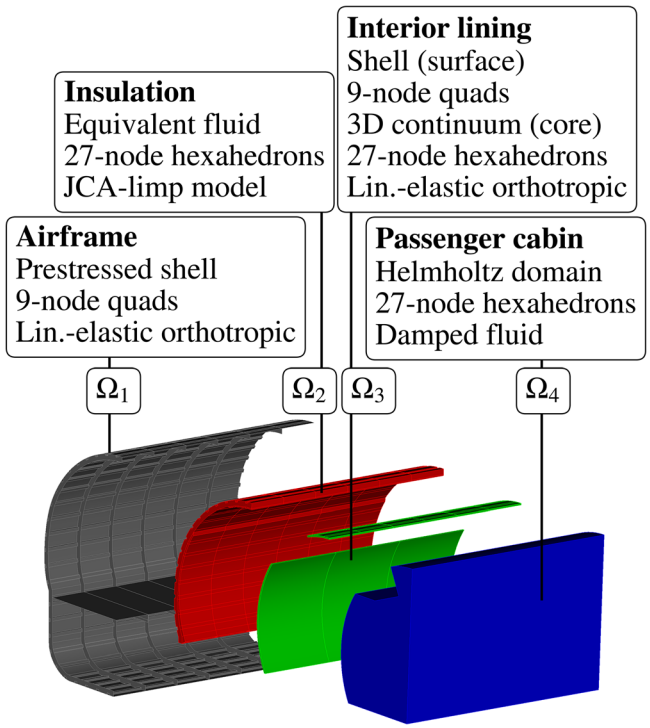

FIG. 2. Vibroacoustic FE model with detailed description of the physical subdomains (Ref. 15).

https://doi.org/10.1121/10.0039865

(3D) solid continuum. Last, the cabin is a Helmholtz fluid with a density of 0.88 kg/m^3 and wave speed of 345.6 m/s, already taking pressurization into account. As is evident, all subdomains are modeled using quadratic ansatz functions in the finite element (FE) formulation.

Table II provides relevant structural parameter values of this reference configuration, and Fig. 3 shows the frequency-dependent damping of CFRP, GFRP, cabin fluid, and the insulation. The loss factor of the cabin includes the influence of seats; for details, see Ref. 29.

As for the excitation, the outer skin is excited by the TBL in a cruise-flight scenario at Mach 0.42. A workflow to directly utilize CFD computations as the input for the excitation has been developed in Refs. 30 and 31, and further details on the pressure computation can be found in Ref. 13. In summary, RANS simulations (solving Reynolds-averaged Navier-Stokes equations) were conducted by the IFAS at TU Braunschweig and supplied the pressure fields of the TBL. Based on these pressure fields, the airframe excitation is derived using the Uncorrelated Wall Plane Wave approach. Assuming near field effects extend to three wavelengths from the source, the entire cabin should be affected by the trim's near field up to approximately 800 Hz.

It should be noted that no propeller excitation is included in this case study, as no reliable data fitting the SE^2A short-range aircraft was available at the time of the listening experiments. While it would enhance the realism of the simulated cabin scenario as a whole, this paper focuses on thoroughly investigating the TBL as a broadband excitation type. This demonstrates the capability of simulations to investigate individual noise sources in an isolated manner, which would be practically impossible to achieve with in-flight measurements.

Finally, the acoustic pressure inside the cabin can be computed according to Eq. (1),

$$\begin{bmatrix} -\omega^2\mathbf{M}_s+\mathbf{K}_s & \mathbf{C}_{FSI} \\ \mathbf{C}_{FSI}^T & \frac{1}{-\rho_f\omega^2}\left(-\omega^2\mathbf{M}_f+\mathbf{K}_f\right) \end{bmatrix} \begin{bmatrix} \mathbf{u}_s \\ \mathbf{p}_f \end{bmatrix} = \begin{bmatrix} \mathbf{f}_s \\ \mathbf{0} \end{bmatrix}. \tag{1}$$

$\mathbf{K}$ and $\mathbf{M}$ denote the stiffness and mass matrices, respectively, carrying the indices s and f, which identify them as structural or fluid components. The entries of the matrices are governed by the introduced vibroacoustic model. The angular frequency is given by ω. $\mathbf{u}_s$ is the vector of degrees of freedom (DOFs) of the structural parts. It contains displacements and rotations. Additionally, $\mathbf{p}_f$ describes the pressure in the acoustic domains. The system is strongly coupled through the coupling matrices $\mathbf{C}_{FSI}$. The four domains are restructured into two structural domains (airframe and trim) and two fluid domains (insulation and cabin), which allows for a symmetric formulation of the underlying equations, ultimately yielding Eq. (1). The right-hand side term denotes the excitation vector that includes only the structural excitation $\mathbf{f}_s$ caused by the TBL. Equation (1) can be solved for any frequency of interest. As already noted, computational effort needs to be considered, especially in an early design stage. It is impacted by resolving higher frequencies with a finer computational mesh, as well as increasing the frequency resolution of the response by requiring more steps to be computed. For this case study, a frequency domain of 16–1000 Hz with a step size of 2 Hz proved to be manageable with moderate computational effort: Solving a single configuration requires approximately 3 h on a system equipped with 24 cores (3.8 GHz) and 126 GB of RAM. This allows for computations of multiple parameter variations with a good spectral resolution to capture the broadband shape of the TBL excitation in a reasonable time frame. After solving Eq. (1) for each frequency step, a pressure distribution is obtained at any nodal point inside the aircraft cabin. This allows for an evaluation of the SPL at different positions inside the cabin, such as the passenger head position in an aisle seat, for example. An exemplary computed power spectral density (PSD) spectrum in the middle of the aisle is displayed in Fig. 4. The spectrum shows the broadband, noisy characteristic of the TBL excitation as well as a few resonant ranges that stem from the modal behavior of the radiating airframe patches between the frames and stringers, as well as the modal behavior of the cabin. These frequency responses serve as the input to the auralization framework to yield signals that can be listened to.

TABLE II. Structural parameter values of the reference FEM model.

		E_x (Pa)	E_y (Pa)	ν	ρ $\left(\frac{kg}{m^3}\right)$	t (mm)
Frame	Skin	69.2×10^9	48.6×10^9	0.166	1620	3
	Stringer	120×10^9	120×10^9	0.33	1620	3
Lining	Skin	74×10^9	74×10^9	0.129	1930.7	2
	Core	99.1×10^4	$98.7\ 74 \times 10^4$	0.978	82.7	—

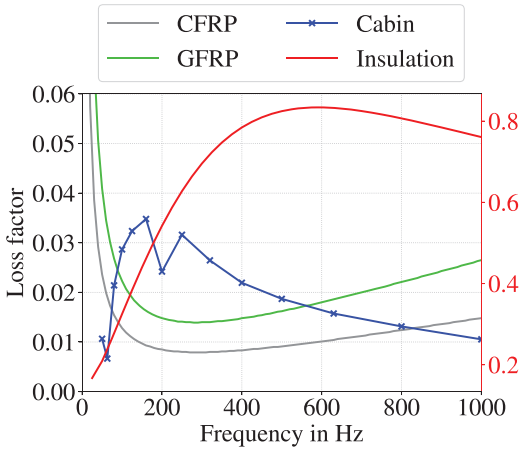

FIG. 3. Frequency dependent damping loss factors η (global) of the CFRP (airframe), GFRP (interior lining), cabin fluid and the insulation. The insulation values correspond to the axis on the right side.

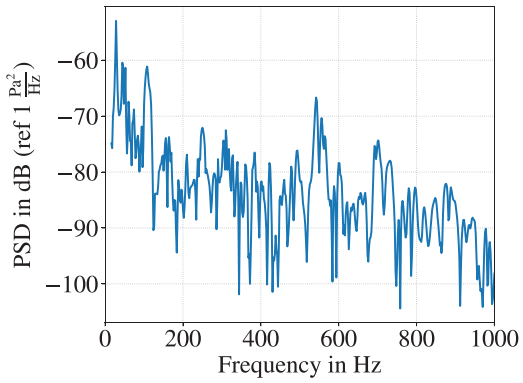

FIG. 4. PSD spectrum at a node inside the cabin at the symmetry boundary.

While the overall cabin simulation is not validated with experimental data, the included models—most importantly, the insulation's equivalent fluid approach and the cabin fluid's loss factor—are based on literature or have been validated individually. The simulation results are assumed to be a reasonably accurate representation of the simplified cabin's acoustic environment, as they are based on a wave-resolved methodology. In-depth details of the simulation model can be found in Ref. 29.

B. Auralization framework

With the cabin model established and FEM results computed, the next step in the analysis chain is auralization. The auralized signals require an appropriate length for listening tests, and they need to be representative of the simulation results; ideally, regarding a realistic application case. Here, significant challenges arise due to the limitations of the early design simulation setup considered. The frequency resolution of 2 Hz results in only 0.5 s response signal length, and the limited frequency range of 1000 Hz cannot capture the full listening experience of aircraft cabin noise.

To meet the auralization requirements while keeping the method as simple and efficient as possible for early design, subtractive synthesis[32] is chosen as the main approach. This method produces auralization by filtering a broadband signal with a transfer function. Here, the computed frequency response inside the cabin functions is used as the filter, and white noise is the chosen broadband signal. The resulting signal can be arbitrarily long for any listening purpose—the system's impulse response is simply zero-padded to the desired length of the white noise before filtering. An equivalent approach would be a circular convolution of the system's impulse response with the white noise.

By filtering white noise, the stationary stochastic behavior of it is imparted on the resulting auralized signal. Sharing the stochastic nature of a TBL and its induced fuselage excitation, this signal characteristic can be considered representative of a realistic TBL excitation.

Additionally, white noise has constant energy over all the considered frequencies in its initial state. In this case, filtering is equivalent to multiplying the frequency spectra of white noise and the cabin's response. Therefore, the spectral energy distribution of the auralized signal remains equivalent to the cabin's response, since it is ultimately a multiplication by a constant value at each frequency bin. Figure 5 shows the overlapping PSD spectra of the exemplary simulation result at an aisle position and its auralized signal with a length of 5 s, computed with Welch's method and a segment length of half a second for a matching frequency resolution. Slight stochastic deviations exist due to white noise not actually being constant at each frequency bin.

Thus, the spectral distribution and overall SPL can be considered representative of the simulation result and representative of the application case as far as the simulation model's fidelity allows.

The limited frequency range of the simulation results cannot currently be improved upon while relying solely on FEM-simulations, without leading to unreasonable computational costs. Though advances in computational efficiency and better hardware steadily increase the feasible frequency ranges, additional data, recorded or otherwise simulated, is still necessary in order to extend the auralized signals' frequency content. To investigate the impact of high frequencies on the listening experience, the option of appending frequency content above 1000 Hz with a cabin recording is added to the auralization method. The used recording includes familiar cabin noise aspects like human chatter or ventilation noise.[33] This leads to a more authentic and familiar auditory impression, therefore increasing the realism of the signal with regard to the application case. However, this approach decreases the representativeness of the simulation model, since the high-frequency content does not change with variations of the simulation's input parameters.

C. Sound Quality Analysis Toolbox (SQAT)

Sound quality metrics (SQMs) describe the subjective perception of sound by human hearing, unlike the SPL or L_p, which quantifies the purely physical magnitude of sound based on pressure fluctuations. Typically, the frequency-dependent sensitivity of the human ear is accounted for by applying A-weighting, resulting in SPL(A) or $L_{p(A)}$.

Previous studies[21,34–36] show that SQMs better capture the auditory behavior of the human ear compared to conventional

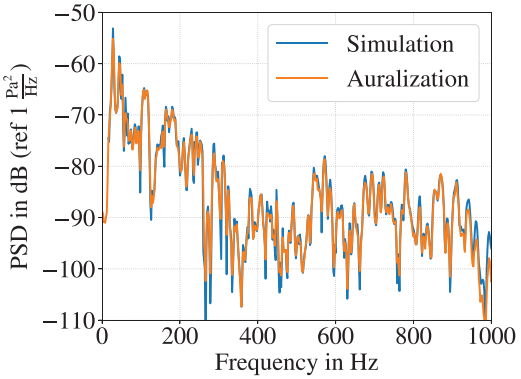

FIG. 5. PSD spectra of a simulation result at an aisle position and its auralized signal with a length of 5 s.

https://doi.org/10.1121/10.0039865

sound metrics typically employed in noise assessments. The five most commonly-used SQMs[37] are as follows:

- Loudness (N): Subjective perception of sound magnitude corresponding to the overall sound intensity.[38]
- Tonality (K): Measurement of the perceived strength of unmasked tonal energy within a complex sound.[39]
- Sharpness (S): Representation of the high-frequency sound content.[40]
- Roughness (R): Hearing sensation caused by sounds with modulation frequencies between 15 Hz and 300 Hz.[41]
- Fluctuation strength (FS): Assessment of slow fluctuations in loudness with modulation frequencies up to 20 Hz, with maximum sensitivity for modulation frequencies around 4 Hz.[42]

All SQMs and the conventional SPL(A) were computed using the open-source MATLAB toolbox SQAT v1.1.[37,43,44] Their mean values are considered, representing the stationary signals appropriately. The SQMs can be subsequently combined into a single global psychoacoustic annoyance (PA) metric. Several empirical models were designed to estimate annoyance, differing in application cases and the sub-models used for the different SQMs. Currently available in SQAT are implementations following the models outlined by Widmann,[45] Zwicker and Fastl,[46] More,[47] and Di *et al.*[48]

Several limitations stemming from the simulation setup have an effect on SQMs in this study. First, with only the TBL being included as excitation, no tonal components are present. Roughness and fluctuation strength are also not expected to be significant, due to the stationary stochastic behavior of the auralization method. Sharpness becomes significant for frequencies higher than 3000 Hz, but since the simulations of this study are limited to 1000 Hz, it is not expected to be of significance here. The signals with added authenticity include high-frequency content, but it is not varied so as not to overshadow the simulation's variation. Last, the frequency resolution of the simulations is set up to capture the broadband character of the TBL, but still does so at discrete points. There is a potential risk of not resolving important resonance interactions that could introduce amplitude modulations.

Nevertheless, SQMs provide meaningful information and are an essential part of the overall methodology. Especially tonality, as well as roughness and fluctuation strength, are indeed expected to become significant after the inclusion of propeller noise in future work. With this in mind, a short analysis in Sec. V C will address the SQMs of the stimuli in this study.

III. LISTENING EXPERIMENT

A. Sound stimuli

Applying the methodology described in Sec. II, the sound stimuli for the listening experiment were simulated and auralized for variations of the cabin parameters. This is chosen as the main point of interest of the current case study in order to prove the efficacy of the simulation and auralization approach in an early-stage scenario. Following a proof-of-concept approach regarding the methodology in an early design stage, the parameter variations were chosen in a generic, exploratory way. The focus of these variations was to produce varied and significant differences in the simulation results, as well as in the auralizations. The parameter variations include the following:

- Variation of the cabin's outer skin Young's modulus by ± 10 %, including the stringers and frames,
- variation of the cabin's outer skin thickness by ± 1 mm and
- variation of the cabin fluid's density and wave speed by ± 10 %.

These variations produce maximum differences in overall SPL of approximately 3–5 dB for this case study, depending on the position inside the cabin. The predicted SPL of the simulated cabin noise, however, was quite low on average at approximately 50 or 35 dB(A). In order to make sure the test participants would be able to focus on the spectral details of the parameter changes, a flat gain of 30 dB was added to all stimuli, elevating them to comfortable but clear 65 dB(A), on average. While the auralizations no longer represent the original excitation levels, the relative differences due to the parameter variation remain the same.

Realistic material uncertainties, alternative materials, geometric variations, or most importantly, validated excitation data will be the subject of future investigations.

The variations' most notable impacts on the physical characteristics of the vibroacoustic problem are as follows: The Young's modulus variation, labeled as "E±," affects the structure's flexural rigidity D. The thickness variation, labeled as "thick.±," has an impact on both the flexural rigidity D and the structure's mass per area ρ_A. The cabin fluid's density and wave speed variations, labeled as fluid±, affect the fluid's bulk modulus K and are referred to as a bulk modulus variation throughout the rest of this paper. Table III details the reference values and the corresponding changes of the parameter variations.

Based on these cabin parameter variations, two sets of stimuli were generated for two different positions inside the cabin model. The first position represents a passenger in a window seat (though windows are not included in the cabin model), while the other represents a passenger in the middle of the aisle, which is the farthest position spanwise from the first one. To keep it concise for the purposes of the listening

TABLE III. Impact of the parameter variations on the aircraft cabin model's vibroacoustic characteristics.

Variation	Effect	Reference value	Variation values	
			—	+
E	D (Nm)	187	168	205
thick.	D (Nm)	187	55	443
	ρ_A (kg/m^2)	4.9	3.2	6.5
fluid	K (kPa)	105	77	140

experiment, the positions are referred to as "window seat" and "aisle seat." These positions lead to different pressure spectra due to having different distances to radiating surfaces and the symmetry boundary. Also, the aisle position should be less affected by near field effects in the higher frequency range. The one-third octave-band spectra of the A-weighted SPLs of the auralizations at these two positions are displayed in Fig. 6.

It can be observed that the positional variation introduces significant differences in the spectra—especially around the 100 Hz and 500 Hz bands—with the aisle position being louder. The opposite has been observed in real aircraft [e.g., 4 dB(A) higher at window seat in Ref. 49]. This issue is discussed in Sec. V B.

A third set of stimuli was generated using the parameter variations in the aisle position. Here, the signals were enhanced with a real cabin noise recording, as described in Sec. II B to increase their overall authenticity. The authentic signal was added with a specific gain so that the overall A-weighted SPL would just barely change: the average SPL (A) increase is 0.05 dB. Figure 7 displays one-third octave-band spectra of the A-weighted SPLs of the reference auralizations with and without the authenticity addition.

With seven signals representing different combinations of cabin parameters, two positions and the authenticity addition, a total of 21 stimuli were produced to be presented to the participants in the listening experiment.

For all stimuli, a signal length of five seconds was deemed to be long enough, given the pseudostationary nature of the signals, in order to create a good impression regarding the cabin noise while keeping the listening experiment's overall duration as short as possible. These five seconds included half a second of linear fade-in and fade-out for a more pleasant listening experience.

B. Psychoacoustic Listening Laboratory (PALILA)

The listening experiments were conducted in the PALILA at the faculty of Aerospace Engineering of Delft University of Technology (TU Delft).[50] The facility consists

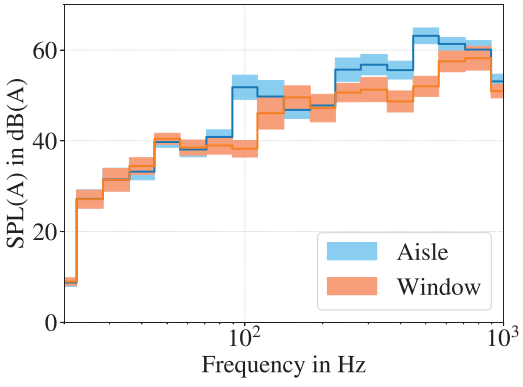

FIG. 6. One-third octave-band spectra of A-weighted SPLs of the auralizations at the two positions: window seat and mid-aisle. The bold lines represent mean values across the presented parameter variation, the colored areas span the standard deviation.

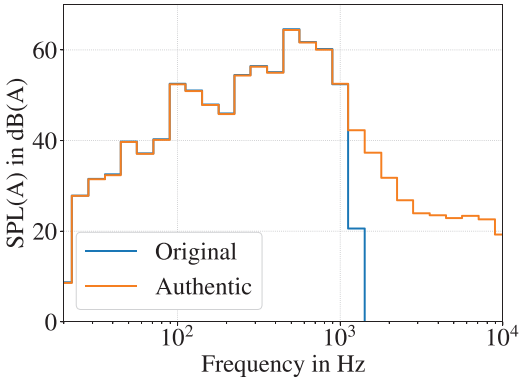

FIG. 7. One-third octave-band spectra of A-weighted SPLs of the reference auralization without (Original) and with (Authentic) the addition of a real cabin noise recording above 1000 Hz.

of a box-in-box soundproof booth with interior dimensions of 2.32 m (length) × 2.32 m (width) × 2.04 m (height). The interior walls, ceiling, and part of the floor are covered with acoustic-absorbing foam panels to prevent sound reflections. This results in free-field sound propagation conditions for frequencies higher than or equal to 1600 Hz and a reverberation time of only 0.07 s, making the facility *acoustically dead*. The walls of PALILA consist of a sandwich structure of two 9 mm-thick exterior fiberboard panels and a 52 mm-thick acoustic foam agglomerate interior, which provides a weighted average transmission loss of 45 dB. The A-weighted overall background noise level is 13.4 dBA.[51]

The listening room is equipped with a Dell Latitude 7340 touchscreen laptop (Dell, Round Rock, TX). Participants interact with the experiment through a Python-based graphical user interface for listening experiments.[52] After an initial questionnaire and an exemplary mock-up round, the interface presents participants with sound samples and questions related to those samples. For this experiment, the samples are divided into five blocks. Within these blocks, the sample order is randomized to limit learning effects due to the specific ordering.

C. Sound reproduction

In the listening experiments, the participants should perceive cabin noise audio signals that resemble the original simulation results. Therefore, the output of the sound reproduction system needs to closely match the desired input spectral characteristics. For this purpose, Sennheiser HD 660S2 (Sennheiser, Wedemark, Germany) open-back headphones were used in conjunction with the binaural headphone equalizer "labP2" from *HEAD acoustics* (Herzogenrath, Germany). The equalizer was used for its headphone-specific equalization filter, as determined by *HEAD acoustics*, and used to set the volume setting. The reproduction did not include binaural or other such filters, therefore resulting in equalized mono signals.

In order to ensure the intended output levels, an output monitoring process was conducted every day of the listening

https://doi.org/10.1121/10.0039865

experiments. First, the binaural microphone "type 4101-B" from *Brüel & Kjær* (Naerum, Denmark) was calibrated at 1 kHz with the calibrator "type 4231," also from *Brüel & Kjær*. Afterwards, two audio signals were fed to the headphones as input and measured with the calibrated microphone: a chirp and a cabin noise sample. Using the input and measured output levels, the volume setting was adjusted incrementally to minimize the level difference. The differences between the input and output levels at the left and right ears, respectively, are shown in Fig. 8, averaged over all days of the experiment. As the figure illustrates, the sound output is quite close in SPLs across all relevant one-third-octave bands up to the 1 kHz band. Here, the differences are consistently lower than ±2 dB, which is deemed an acceptable difference for evaluation purposes in this work. As displayed by the dotted lines, the day-to-day consistency between 100 and 1000 Hz is quite high, with a low standard deviation of 0.3 dB at the left ear and 0.9 dB at the right ear. Thus, the cabin simulation results are sufficiently represented by the audio reproduction.

D. Design of listening experiment

The listening experiment was designed around the cabin's parameter variation with the intention of verifying the methodology's ability to produce results that are in accordance with established sound metrics. The aim is to estimate the relative preference for each signal within the group, which can be used to rank the parameters, find trends, and compare preference sensitivities.

At its core, the listening experiment assessed the parameter variations with forced choice A/B comparisons. Comparative judgment is a widely used method to arrive at preference probability, dating back to Thurstone's description in 1927.[53] This study's method is based on a comprehensive guide provided by Pérez-Ortiz and Mantiuk:[54] Throughout the experiment, participants were presented with two sound samples at a time and had to indicate which sample they found more pleasant in an imagined cruise flight scenario. The participants were able to repeat each sound sample a second time to give their best assessment.

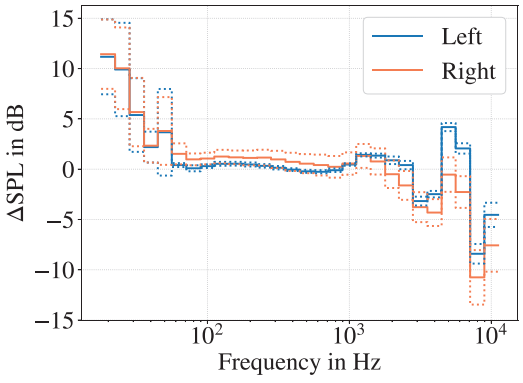

FIG. 8. Difference between input and output at the left and right ear, respectively, in one-third-octave band levels, averaged over all experiment days. Standard deviations are added as dotted lines.

Deviating from Pérez-Ortiz and Mantiuk, the pairwise comparison data is subsequently not translated into a measurement scale since a complete ranking is made with each participant. However, the preference data is post-processed a step further with perceived similarity ratings that are additionally assessed for each signal pairing. Similarity is an important quality in comparative judgments and can be assumed to have an influence on preference. A study by Berglund *et al.*,[55] for example, investigated a joint space for similarity and non-preference, and drew connections to spectral contrast based on Zwicker's specific loudness and perceived annoyance. To include similarity in this study, participants were asked to give similarity ratings on an integer scale from 0 to 6, with 0 meaning "very different" and 6 meaning "very similar." The similarity rating is used to enrich the binary preference rating with a weighting function, with the idea in mind that a preference choice is worth more when the signals are clearly distinct and is worth less when the signals are very similar. To give an example, when a signal pairing is clearly distinct and receives a 0 on the similarity scale, the participant can have a clear preference. Therefore, the preferred signal would receive a preference rating of 1 and the other signal a 0. However, when a signal pairing is very similar and receives a 6 on the similarity scale, the participant cannot give a clear preference. In that case, both the "forced" preferred signal and the other signal would receive a preference rating of 0.5. This similarity weighting function was chosen to have a linear drop-off between 1 and 0.5.

For each set of seven signals, per seating position, a 7×7 matrix is built, according to the preference choices and similarity ratings. The diagonal of the matrix is left empty because the signals were not compared to themselves. To illustrate: the matrix element p_{ij}, for the pairing of signals i and j, is set to 1 if signal i was preferred. It is then weighted according to the selected similarity rating to, for example, $p_{ij} = 0.8$. The element p_{ji} will then be set accordingly to $p_{ji} = 1 - p_{ij}$, in this case 0.2. For seven signals, 21 signal pairings need to be assessed to fill in the matrix. Once the matrix is complete, the average along the row, excluding the diagonal elements, will yield the weighted preference probability P_{wp} value of the corresponding signal.

The experiment is divided into three major parts, evaluating P_{wp} of the three sets of stimuli, which are described in the previous section: (1) for the aisle seat position, (2) for the window seat, and (3) for the aisle seat with the addition of authentic elements. There are two additional sub-parts: Part 2.2, assessing direct comparisons of window and aisle seat signals, and Part 3.2, assessing direct comparisons of signals with and without the authenticity addition.

E. Experimental protocol

Moving on from the core questions, the participants were each introduced to the experiment with information on the context and broad content of the experiment itself. The participants were informed about the data processing and

protection guidelines that were in place in this research project and asked to sign consent forms. The experiment and its corresponding documentation have been approved by the Human Research Ethics Committees of both TU Delft (application number 3599) and TU Braunschweig (identification number FV-2024-11).

The listening experiment itself started with a questionnaire in the user interface to collect demographic information for population statistics and information about the hearing health of the participants. For the population statistics, information was collected regarding age, gender identity, level of education, employment status, and the current affiliation (if any) with TU Delft (where the experiment was conducted). Hearing health was self-assessed by participants on a five-point scale from "poor" to "excellent," and further questions regarding hearing health and well-being were included. A complete overview of the question formulations and possible answers is found in the repository of the GUI.[52]

At the end of the experiment, an additional questionnaire was presented in the interface to assess possible biases. Participants were asked about flight types they had experienced: "short-range (<3 h)," "medium-range (3–6 h)," and/or "long-range (>6 h)." They also indicated their overall experience with flying, their attitude towards aviation, and whether cabin noise had previously bothered them (engine, ventilation, passenger, or other noise as possible aspects).

Since the experiment was relatively lengthy and repetitive, measures were taken to alleviate potential fatigue and strain on the participants. Breaks were included after every seven pairs of sound samples and before each new part. Each part was introduced with some broad information about the new sound situation, including light-hearted messages for context, to refocus their attention, and to remind participants to keep answering intuitively. Finally, a progress bar in the GUI was meant to provide the participants with a sense of progression.

F. Post-experiment interview

After the listening experiment, the participants were invited to a short post-experiment interview to provide feedback and insight into their experiences. The interview was guided by six main questions targeting the following aspects:

(1) well-being and fatigue,
(2) main deciding signal characteristic for preference or dislike,
(3) overall estimate of signal similarity,
(4) feedback regarding signals and procedure,
(5) attitude towards flying and aviation, and
(6) other remarks.

Since the resulting data set of the experiment was randomized with regard to the participation order to anonymize the data, the interviews were not added to the data set and were not used for quantitative statements. However, a few typical statements are discussed in Secs. IV and V to support the findings.

IV. PARTICIPANT STATISTICS

According to Beltman *et al.*,[56] about 25 to 30 participants are needed in order to have a sufficient sample size for statistical analysis. The promotion for this study's listening experiments was done in a network around the Faculty of Aerospace Engineering of the TU Delft, and 43 participants successfully registered for the listening experiments. Regarding the affiliation of the participants, 37 participants were students or employees at TU Delft, and three participants were students or employees at a partner university or company, leaving three participants without any affiliation to the university.

Starting with general information, Fig. 9 displays the demographics and hearing rating of the participant group: Most participants were young students, a male-to-female ratio of approximately 3:1 was found with two participants identifying as other, and the self-reported hearing health of all participants was at least "good." The demographics, while not representative of the overall population, are generally representative of TU Delft.[57]

The self-reported hearing health of all participants being at least "good" is considered sufficient for this type of listening experiment, since it focuses on the overall preference of broadband noise comparisons rather than just-noticeable differences in difficult frequency ranges. Regarding further hearing health questions that could potentially influence the experiment, two participants reported having to wear hearing protection at work, two (different) participants reported having tinnitus, six participants reported having a cold, two participants were feeling unwell, and one participant was feeling very tired.

No participants have been excluded from the experiment as a consequence of the health questionnaire, since it

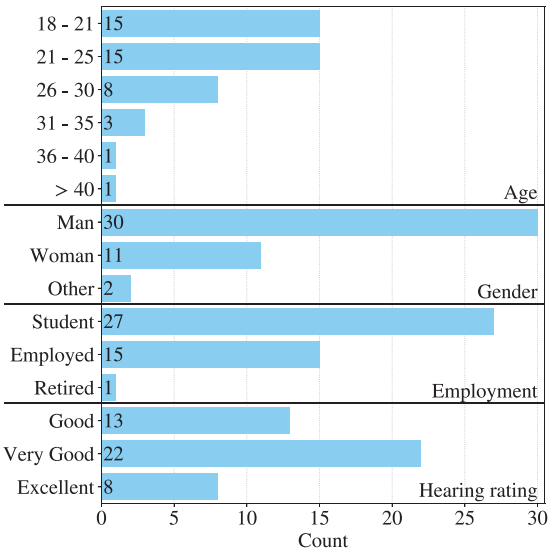

FIG. 9. Demographics and self-reported hearing rating of the participant group.

https://doi.org/10.1121/10.0039865

was taken during the experiment. A correlation analysis was applied to the participants' individual preference probabilities to assess whether their hearing health or well-being affects the results. The correlation was determined with Spearman's rank correlation coefficient[58] between the participants' individual preference probabilities. The resulting correlation matrix is presented in Fig. 10. A dendrogram is also presented, based on hierarchical clustering using the UPGMA algorithm (unweighted pair group method using arithmetic averages),[59] to highlight potential outliers.

The dendrogram highlights three participants as outliers, and indeed their correlation coefficients with other participants are noticeably low, appearing as dark streaks in the correlation matrix. The responses of these participants to the questionnaire were checked for conspicuous features in the well-being and hearing health categories. Their overall answering patterns were also checked for behaviour such as one-sided or strictly alternating answers. However, nothing of note could be found to explain the low correlation values reported. Therefore, these three outliers were not excluded from the data set.

Potential biases were assessed with the questionnaire at the end of the listening experiment, as described in Sec. III D, and are displayed in Fig. 11: All participants have flown before, with all ranges being represented well, most flight experiences have been positive, and most participants' general attitude toward aviation is positive. With the participants' aerospace affiliations and generally positive attitudes towards aviation and flying, a bias cannot be excluded. However, since the listening experiment focuses on preference choices rather than absolute estimates of comfort or other metrics, this bias is not expected to affect the results.

When asked if they had ever felt bothered by cabin noise, a significant number of participants answered "engine noise" and "passenger noise," which is further proof that these noise types need to be included in future work.

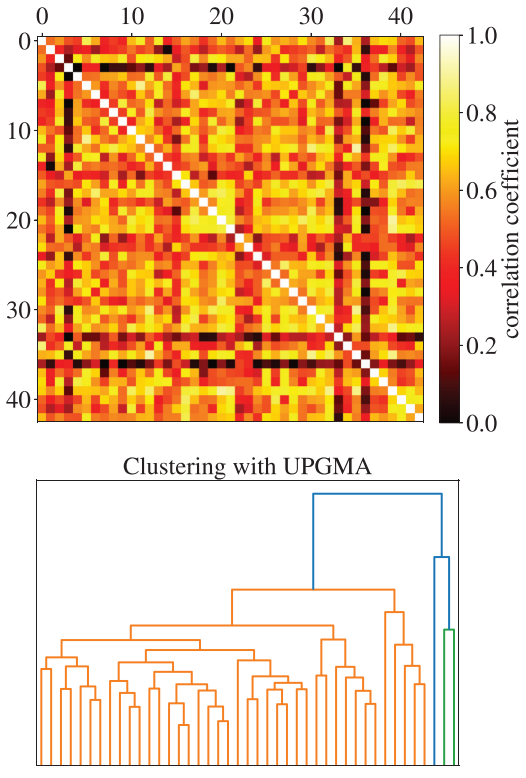

FIG. 10. Top: Correlation matrix between all participants to find outliers. Bottom: The corresponding hierarchical cluster analysis using UPGMA visualized as a dendrogram.

V. RESULTS

This section details the analysis of the listening experiment results and highlights how the simulation methodology of this project can be used in an early design process to conduct a psychoacoustic assessment.

The parameter variations, as described in Sec. III A, are referenced as specified by Table III: "E" denotes Young's modulus variation, "thick," the thickness variation, and "fluid" the bulk modulus variation via wave speed and density.

A. Analysis of cabin parameters

This first analysis section focuses on the perceptual aspects of the cabin parameter variation. The preference probabilities P_{wp} for the parameter variation at the window seat are displayed as a violin plot in Fig. 12 and they are sorted according to their respective mean values to create a preference ranking. The most notable observation for this listening experiment is that a clear trend is found, indicating that the parameter variation leads to audible differences, which have a significant impact on preference.

The least preferred signals, representing an increased cabin fluid bulk modulus and a decreased outer skin thickness, are the most distinct, with the majority of preference probabilities well below 0.5. The least preferred signal has preference values of $P_{wp} = 0.0$ for some participants, indicating that these signals were not only not preferred, but also clearly audibly different from other signals. Their

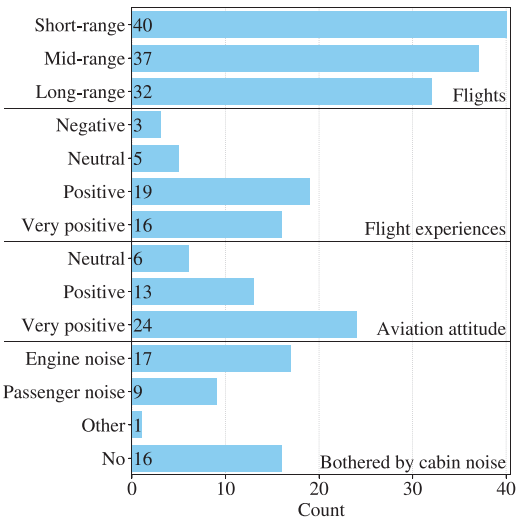

FIG. 11. Questionnaire results regarding flight experiences and attitude towards aviation.

overall P_{wp}-distributions are visibly lower than others, with the interquartile ranges (IQR) not overlapping with the other signals. To quantify the difference, a Mann–Whitney U test (MWU)[60] is conducted between all signal distributions as a non-parametric way to test the statistical significance. Generally, a difference is considered to be statistically significant with a p-value below a significance level of $p < 0.05$. For both of the least preferred signals, all p-values, especially with respect to the neighboring distributions in the ranking, are negligibly small.

The reference signal and both Young's modulus variations are found to have overlapping distributions, with close median values. Their preference values are close to $P_{wp} = 0.5$. This is caused by a balance in being preferred, but also by high similarity ratings when compared with other signals. The difference in preference distributions between the reference and the modulus increase signals is not statistically significant with a p-value of $p = 0.26$ in the MWU test. The differences between the modulus increase signal and the other two show p-values of $p = 5 \times 10^{-3}$ and $p = 6 \times 10^{-5}$, indicating that it is different with statistical significance.

The signals representing the outer skin thickness increase, and the fluid bulk modulus decrease, show the highest preference probabilities in this plot. The IQRs for both lie between $P_{wp} = 0.6$ and 0.8. Both signals exhibit a large spread over the probability range, but neither reaches the value of $P_{wp} = 1.0$ for any participant. The latter result indicates that similarity to other signals is a factor, despite these signals being preferred within the group. According to the MWU test, the two distributions are not significantly different with a p-value of $p = 0.79$.

In order to discuss these observations, the signal type and the effects of the parameter variations on the signals need to be considered. First, regarding the signal type, all auralizations in this work can be categorized as stationary stochastic noise, because the signals do not include tonal or transient aspects due to the auralization method. Therefore, audible differences between these signals stem from either frequency content or amplitude differences. This is supported by answers given in the post-experiment interview, where these parameters came up as having the biggest impact on participants' preference choices—pitch more so than amplitude.

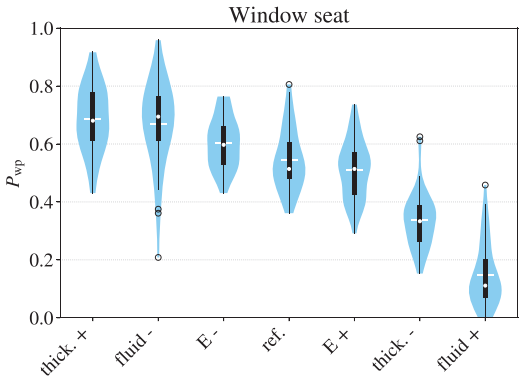

FIG. 12. Kernel density estimate (blue) and box plot of the preference probability P_{wp} of each signal based on the parameter variation at the window seat position. White circle: median, white bar: mean, black circles: outliers outside of 1.5 x interquartile range (IQR).

Therefore, regarding the effects of the parameter variations on the perceptual effect, the following statements can be made: On the structural side, the variations of the Young's modulus affect flexural rigidity. This increased stiffness results in higher resonant frequencies but has two opposing effects on the sound levels inside the cabin. The incoming sound levels reduce due to an increase in sound resistivity, but the noise levels inside the cabin are increased by the higher reflectivity. In this study's simulation, the overall result is a small SPL(A) increase in 0.4 dB(A) and its auralization is fittingly less preferred than the reference signal. The decrease in Young's modulus results in a slight decrease in SPL(A) of 0.9 dB(A) on the simulation side, and the signal is preferred over the reference.

For the thickness variation, the stiffness effect is present as described previously, but is accompanied by a mass variation. An increase in the mass of the structure results in lower resonant frequencies and lower amplitudes. The latter is caused by the increase in the required energy to excite a heavier structure. In this study, a change in skin thickness results in more significant changes in SPL (A) of −1.5 dB(A) for the increase and +2.0 dB(A) for the decrease with respect to the reference signal. This larger variation in levels is reflected in the preference probabilities.

The cabin fluid's density and wave speed variation affect the bulk modulus, as described in Sec. III A, which affects the fluid's resistance to compression. This changes how efficiently sound waves propagate through the fluid and the fluid's acoustic impedance, influencing how sound waves are transmitted and reflected at the cabin's walls. A fluid with a high bulk modulus resists compression, allowing pressure waves to travel faster and with less attenuation, generally leading to higher SPLs. Being a more extreme parameter variation in this study, the impact on the auralized signal is quite pronounced, with the SPL(A) values varying by −1.5 dB(A) for the decrease and +4.3 dB(A) for the increase in bulk modulus. This results in yielding the least preferred and the most preferred signals as far as individual ratings are concerned.

B. Comparison of experiment and metrics

As is apparent from the literature, noise level reduction increases passenger satisfaction. This is hinted at in the previous subsection's elaboration of the parameter variations' consequences: SPL(A) changes appear to correlate with the signal's preference. Therefore, in addition to displaying the experimental results as violin plots, Fig. 13 adds a direct comparison to the SPL(A)s of the signals [$L_p(A)$ on the right axis] in dB(A) and the mean loudness levels (L_N on the second right axis) in phon for the three major parameter variations of the experiment. Because satisfaction decreases with increasing noise levels, the violin plots are now inverted to

https://doi.org/10.1121/10.0039865

display an inverse preference probability $1 - P_{wp}$. To highlight the correlation between metrics and experimental results further, both metrics' axes are specifically scaled so that the first and last values overlap with the first and last distribution's median values. The reference signals have a different shade as an anchor point.

Before comparing experiment and metrics, it needs to be addressed that both $L_{p(A)}$ and L_N values are higher at the aisle position than at the window position [by approximately $3-5$ dB(A) and $3-5$ phon], yet the opposite has been observed in real aircraft [e.g., 4 dB(A) higher at window seat in Ref. 49]. Here, the positional accuracy of the cabin segment and the representativeness of the aisle position have to be critically reflected. For the auralization of the aisle position, the acoustic pressure at a node directly on the symmetry boundary was chosen. The perfect reflection at this boundary yields an overestimated sound pressure, since a realistic overlap of the sound radiated from the walls would be uncorrelated here. Additionally, the model simplifications, such as low-fidelity geometry and a global loss factor inside the cabin instead of actual seats and persons shielding sound radiation towards the aisle, will have an influence on positional sound pressure differences. These aspects, as well as the limited frequency range, are most likely the causes for the incorrect positional findings.

The significance of assessing a positional variation is obviously limited by model fidelity, but the study can still be conducted for a simplified model to highlight perceptual differences. In fact, provided a high fidelity model were available, the effects of simplifying the model could be investigated. Reduced fidelity with or without an audible impact could be distinguished.

Starting with the window seat group in (a), both $L_{p(A)}$ and L_N closely match the listening experiment result. They predict the ranking correctly and also, to some extent, predict the amount of overlap in the probability distributions. When the differences in sound metrics are large, the distributions are clearly distinct, see the difference between "E+," "thick.-," and "fluid+". With small differences in the sound metrics, the distributions are shown to overlap significantly, which is the case for "E−," "ref.," and "E+."

Next, the signal group for the aisle seat position, displayed in Fig. 13(b), is considered. Most notably, the overall trend of the signals within the ranking is not as distinct as in the window seat's case. Here, five of the signals have their mean and median values close to each other between $1 - P_{wp} = 0.4$ and 0.5 and, although the distributions still indicate a trend, their ranges overlap significantly. Applying the WMU test to these five similar probability distributions, the only significant difference is found between the reference signal and decreases in the bulk modulus and Young's modulus. The p-value is 0.04 in both cases.

High similarity is also found for the two least preferred signals, which are distinct from the other distributions but overlap more than at the window seat position. The p-value here is 0.03, indicating the difference is statistically significant.

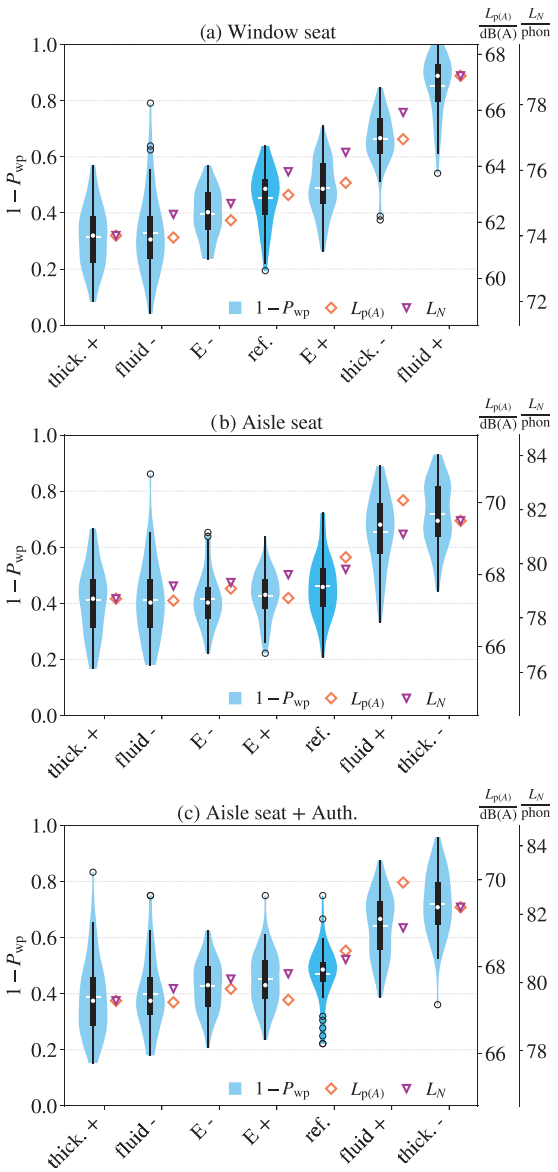

FIG. 13. Inverse preference probability $1 - P_{wp}$ of each signal based on the cases (a) window seat, (b) aisle seat, and (c) aisle seat + authenticity addition. For comparison, the signals' SPL(A) values and loudness levels are plotted next to the corresponding distributions. The respective axes are found on the right-hand side and scaled individually, so that beginning and end values match the vertical position of the median values.

A notable aspect is the change in the order of signals compared to the window seat group. First, the reference signal switched with Young's modulus increase, though the difference in preference is so small that this order is hardly conclusive. Second, the two least preferred signals switched, which represents a more significant change. The signal of the increase in bulk modulus is no longer as clearly disliked, with the main area now closer to $1 - P_{wp} = 0.7$ and some of the participants even rating it at 0.4.

Interestingly, ranking according to SPL(A) predicts the two least preferred signals incorrectly, placing the bulk modulus signals in last place with the highest value of over 70 dB(A). Loudness, however, predicted the entire trend correctly, showing similar values for the five signals that

have been rated similarly and increasing by almost 2 phon for the disliked signals.

Finally, the same overall trend is found after adding authentic cabin sounds to the aisle seat signals, displayed in Fig. 13(c). The ranking remains very similar compared to b). The comparison to $L_p(A)$ and L_N remains analogous, where SPL(A) does not predict the listening experiment ranking correctly, while loudness does. The comparisons of the authenticity change as well as the positional change are discussed more in-depth in the next two subsections.

C. Further psychoacoustic analysis

Additional SQMs—sharpness, roughness, fluctuation strength, and tonality—have been calculated to assess further perceptual aspects of the auralizations. As mentioned in Sec. II C, these have a limited significance in this study. Table IV gives an overview of the value ranges found for the signals.

It should be added that computing PA, which combines several SQMs into a single metric (see Sec. II C), is also part of the methodology. It is based on loudness and enriched by a weighting function that includes sharpness, roughness, fluctuation strength, and tonality (tonality not included in Zwicker's model). However, given the previous analysis of these metrics, the weights are found to be negligible. This results in the trends of the PA values matching loudness for all annoyance models mentioned in Sec. II C. Hence, PA is not investigated further in this work.

The correlations between all considered metrics—SPL (A), loudness, and the previously-mentioned SQMs—and the mean and median $1 - P_{\mathrm{wp}}$ values have been investigated by calculating Spearman's rank correlation coefficients[58,61] for all three major parts of the listening experiment. In accordance with the previous paragraphs, no significant correlation is found for sharpness, roughness, fluctuation strength, or tonality. Consistent with Fig. 13, however, rank correlation coefficients of 1 are found for all cases between inverse preference probability and loudness, showing a perfect monotonic relationship between both variables across parameter variations. Rank correlation coefficients of at least 0.89 or higher are found for the cases between the inverse preference probability and SPL(A). This analysis is supported by the participants' answers in the post-experiment interview regarding the signal characteristics deciding their preference, which were most commonly higher frequencies and overall loudness. Both are expected to correlate with SPL(A) and loudness (but not with sharpness because of the limited frequency range).

TABLE IV. Sharpness, roughness, fluctuation strength, and tonality value ranges of the experiment's signals. Sharpness is distinguished between the original signals without high frequency content and the authentic signals without.

SQM	Min	Max
Sharpness/acum (orig.)	0.50	0.53
Sharpness/acum (auth.)	0.65	0.70
Roughness/asper	0.07	0.35
Fluctuation Strength/vacil	0.03	0.16
Tonality/t.u.	0.05	0.10

D. Analysis of positional change

After discussing the listening experiment results regarding the simulation parameter variations, it is apparent that the position inside the cabin—even though being subject to model inaccuracies as discussed in Sec. V B—has a significant influence on human perception. To highlight this further, Fig. 14 displays a direct comparison of the violin plots resulting from the parameter variations at the two different positions. The ranking is sorted according to the mean preference probabilities of the window seat case.

The overall trends resulting from the two different positions are relatively comparable, with all IQRs overlapping, except for one. However, the resulting details of the assessments at the two positions do indeed differ. That is most notably the case for the signal of increased bulk modulus, being the only signal where the IQRs do not overlap. Additional differences can be seen for the preference ranges shifting, see, e.g., "thick. +" and "fluid−," or the distribution changing from a normally distributed shape to a bimodal one, e.g., "thick. −." When applying the MWU test between the two seating positions, per signal, five of the seven signals are statistically different, with p-values below $p = 0.02$. The two exceptions are (1) the reference signal with $p = 0.8$, and (2) the variant of decreased Young's modulus ("E−") with $p = 0.59$.

In addition to the two positions being compared by their individual parameter variations, the listening experiment included three direct comparisons of the seating positions: (1) for the reference signal, and (2) for both bulk modulus variations ("fluid+" and "fluid−"). The latter two are chosen because they represent the maximum and minimum SPL (A) values of all signals. The preference choices and average similarity given in this part of the experiment are visualized in Fig. 15

Even though the parameter variations at the two positions yield comparable trends, Fig. 15 visualizes that the

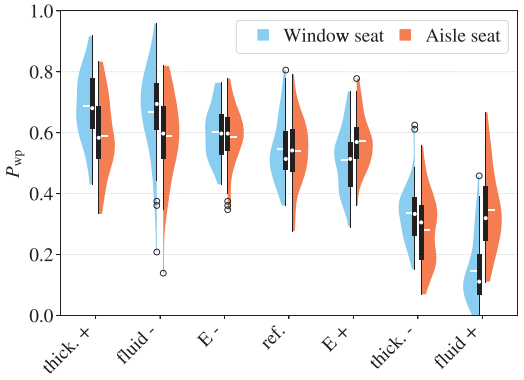

FIG. 14. Preference probability comparison between aisle seat and window seat signals.

https://doi.org/10.1121/10.0039865

individual signals are clearly perceptually different. As expected with an average SPL(A) difference of approximately 5 dB(A), most participants preferred the signals at the quieter window seat position. The accompanying average similarity ratings for these signal pairings confirm that the signals were perceptually distinct, with values well below 3.

In literature, differences in seat positions are described.[2,49] It has already been discussed that in this case study's modeling stage and with the presented node choice, the resulting physical differences are not validly representative of reality. However, the perceptual results highlight that positional preference studies could be investigated further this way, provided the model allowed for such fidelity.

E. Analysis of added authenticity

Utilizing the same comparative approach as in the previous subsection, the impact of the added authenticity, as described in Sec. III A, is investigated next. Figure 16 displays the side-by-side comparison of the preference probabilities resulting from the auralizations at the aisle seat, with and without the added authenticity from cabin recordings. The ranking is sorted according to the mean preference probabilities of the first case.

In contrast to the previous comparison, not only the trend but also the distribution ranges match closely. The IQRs are overlapping significantly for all signals, and even the whiskers span similar values. There are still differences to be found in the details of the distributions, such as the density estimate changing in some cases from being mostly normally distributed to bi-modal (see, e.g., "thick. −" or "fluid +"), and vice versa. The reference signal, in particular, has a significantly narrower IQR with added authenticity. When applying the MWU test per signal, however, none of the p-values are below $p = 0.05$, implying that the authenticity addition does not have a statistically significant influence on the preference choices made by the participants, as far as this test is concerned.

When asked about the authentic signals in the post-experiment interview, the vast majority of participants expressed that it affected their perception of the cabin noise. The opinions on the effect were divided. On one hand, participants enjoyed the authentic signals more than the original

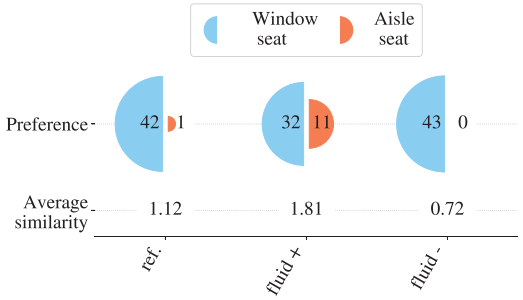

FIG. 15. Direct preference comparison between aisle seat and window seat signals with their respective average similarity rating.

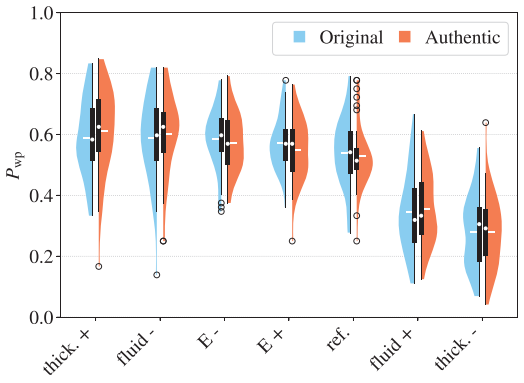

FIG. 16. Preference probability comparison between original and authentic aisle seat signals.

ones because they created a more familiar and immersive scenario. On the other hand, other participants disliked the addition, especially the background chatter, and preferred the muffled experience of the original signals, with some comparing it to using noise-canceling headphones during cruise flights. Six direct comparisons were included in the listening experiment with almost identical SPL(A) values at an average difference of 0.1 dB(A). As with the previous direct comparisons, the signal variants were: the reference signal and bulk modulus variations, but at both positions this time. The results of these direct comparisons are presented in Fig. 17.

While the participant group preferring the authentic signals is larger in five out of the six direct comparisons, the preferences are more balanced than in the positional comparison. The average similarity ratings are below the middle value of 3 for all pairings, implying that the pairings are indeed perceived as distinctly different signals.

This divided opinion indicates that the overall similarity of both variants in Fig. 16 is not because the added authenticity does not achieve the intended effect in the listening experiment. Instead, the addition of the cabin recording only improves the authenticity of the signals, without significantly influencing the parameter variation testing. This indication enables two approaches: (1) listening experiments of cabin noise can be conducted with limited bandwidth (assuming the relevant effects are represented within that bandwidth) and still yield meaningful results, or (2) listening tests can be enriched with authenticity additions to create a more familiar scenario for the participants without the results being influenced. However, more testing of different cabin models with different authenticity additions is required to confirm this hypothesis.

F. Discussion of preference probability

To conclude the results section, the processing of the responses given in the listening experiment into a preference probability is addressed. Other than the approach described in Sec. III D, which is used throughout this section, there are several additional ways to account for the similarity ratings. Some examples are as follows:

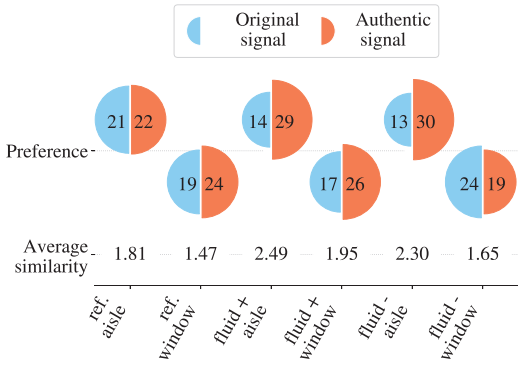

FIG. 17. Direct preference comparison between original and authentic signals at the aisle seat position with their respective average similarity rating.

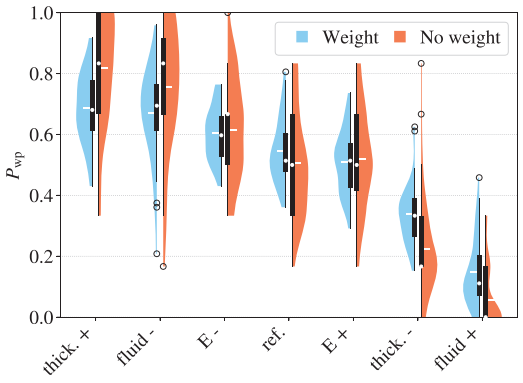

FIG. 18. Preference probability comparison between using a similarity weighting in its calculation and using no weighting.

- Strictly limiting preference probability to the preference ratings (in line with standard literature[53,54]).
- Sigmoid instead of linear weighting, rewarding a tendency to "different" and punishing a tendency to "similar."
- Adjusting the lowest weighting value, therefore still rewarding the preference choice with, e.g., 0.6/0.4 at high similarity.

These variants influence the resulting preference probability. To illustrate this, an exemplary comparison for the window seat case is drawn in Fig. 18 between this study's approach and the unweighted approach.

The distributions and median values of the unweighted approach are spread out more across the range from 0 to 1, with the preferred signals now scoring above 0.8, and the least favored signal receiving a median preference probability of 0. Ultimately, the overall results derived from the rankings, especially the trends and the overall correlation with loudness, do not change significantly. However, Fig. 18 demonstrates how applying the similarity weighting adds meaningful information regarding the comparative judgment and focuses the signals' preference distributions to provide a clearer picture.

VI. CONCLUSION AND OUTLOOK

This study was conducted as a proof of concept, highlighting the potential and importance of simulated methodologies in the early design stages of cabin acoustics, which enable psychoacoustic assessments early on. For this, a segment of a RAM aircraft was simulated under the stochastic load of a TBL, and its system response auralized to create representative stationary stochastic cabin noise. With this methodology, several samples of the cabin's parameter variations were investigated in a listening experiment to determine human preference and perceived similarity.

The findings demonstrated that the methodology's representations of the variations in Young's modulus, skin thickness, and fluid bulk modulus yield audible differences and that the listening experiments were able to capture their influence on cabin noise perception and passenger preferences.

The preference probabilities, derived from the listening experiment and visualized as violin plots, are useful in indicating perceptual similarities within the signal group, as well as identifying parameter variations that are clearly more or less preferable than others—allowing for better statistical evidence and more depth compared to what an analysis of SPLs inside the cabin would provide. In this study, specifically, the chosen variations in Young's modulus are perceived with a similar preference probability as the reference signals, while the skin thickness and the (rather extreme) bulk modulus variations result in the most and least preferred signals of the groups, respectively.

A notable aspect is the significant influence on preference probabilities of signals between seat positions of the same row, even altering the signal rankings. The reference signal and signal with increased Young's modulus switch places, although they are perceived as very similar to begin with. The least preferred signals switched as well, with a more significant change in their respective preference probabilities.

Adding authentic cabin sounds to aisle seat auralizations does not significantly alter the preference probability distributions. However, participants indicated in post-experiment feedback that it affects their perception of the presented cabin noise: some appreciate the immersive scenario, while others prefer the muffled effect of the original signals. The stable preference probabilities suggest that both limited bandwidth experiments, with only calculated results, and signals with added authenticity can yield meaningful results, though further testing is needed.

SPL(A), loudness, and several additional psychoacoustic metrics were compared with the listening experiment results. Loudness and SPL(A) were reported to be reliable predictors of preference throughout all the presented seating positions, with SPL(A) being slightly less accurate. With loudness closely aligning with the participants' assessment, it serves as a viable substitute in comparable use cases, should listening experiments not be practical or desired. The correlation was expected since the amplitude and frequency distribution in the lower frequency range are the two major distinctions between the stationary stochastic cabin noise samples. For the same reason, the other psychoacoustic

metrics, namely, sharpness, roughness, fluctuation strength, and tonality, have little variation and significance related to the participants' preferences. Once higher fidelity models and auralizations start to include higher frequency ranges, tonal noise—most notably expected due to the tonal nature of propeller noise inside aircraft and AAM cabins—and transient components, these metrics are expected to become more relevant.

Alternative approaches for processing the listening experiment's preference responses, such as different weighting functions and similarity adjustments, are explored. These approaches affect the distribution shapes and spreads, but overall trends remain consistent. Ultimately, similarity weighting and awarding value to the non-preferred signal is chosen for this study because of the added information and enhancement of the clarity of preference distributions.

Moving forward with the established methodology, there are several directions that need to be investigated further:

- For future parameter variations and the question of whether they significantly affect noise perception and user preference, realistic parameter sets of different cabin materials, variations addressing material uncertainties, and geometrical variations are of interest.
- Developing the simulation methodology further for early design assessments, integrating propeller engine noise as a major excitation type, adding refined structures to the cabin model, and expanding the auralization method to include transient noise are steps to take. This should also lead to more psychoacoustic metrics becoming relevant, therefore adding depth to the analysis.
- Employing other question types in the listening experiment, such as semantic differentials, should increase the understanding of cabin noise as a factor of passenger comfort.

ACKNOWLEDGMENTS

The authors kindly thank the 43 listening experiment participants for their time and contribution to this research. The authors acknowledge the funding provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2163/1—Sustainable and Energy Efficient Aviation—Project-ID 390881007. This publication is also part of the project *Listen to the future* (with project number 20247) of the research programme Veni 2022 (Domain Applied and Engineering Sciences) granted to Roberto Merino-Martinez, which is (partly) financed by the Dutch Research Council (NWO).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The experiment and its corresponding documentation have been approved by the Human Research Ethics Committees of both TU Delft (Application No. 3599) and TU Braunschweig (Identification No. FV-2024-11).

DATA AVAILABILITY

The data that support the findings of this study (simulation results, auralizations, and listening experiment responses) are available from the corresponding author upon reasonable request.

[1]H. Scheel, "Next generation aircraft-a challenge for interior acoustics developments," in *Proceedings of INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (2016), Vol. 253, pp. 4666–4675.
[2]S. Pennig, J. Quehl, and V. Rolny, "Effects of aircraft cabin noise on passenger comfort," Ergonomics **55**(10), 1252–1265 (2012).
[3]SE^{2}A, "Sustainable and Energy-Efficient Aviation," https://www.tu-braunschweig.de/se2a (Last viewed November 2025).
[4]S. Karpuk and A. Elham, "Influence of novel airframe technologies on the feasibility of fully-electric regional aviation," Aerospace **8**(6), 163–4310 (2021).
[5]S. Karpuk, "Influence of future airframe and propulsion technologies on energy-efficient aircraft," Ph.D. thesis, Technische Universität Braunschweig, Braunschweig, Germany (2022).
[6]R. Goyal, C. Reiche, C. Fernando, and A. Cohen, "Advanced air mobility: Demand analysis and market potential of the airport shuttle and air taxi markets," Sustainability **13**(13), 7421 (2021).
[7]W. Johnson and C. Silva, "NASA concept vehicles and the engineering of advanced air mobility aircraft," Aeronaut. J. **126**(1295), 59–91 (2022).
[8]C. Spehr, H. Hennings, H. Buchholz, M. Bouhaj, S. Haxter, and A. Hebler, "In-flight sound measurements: A first overview," in *Proceedings of the 18th AIAA/CEAS Aeroacoustics Conference (33rd AIAA Aeroacoustics Conference)* (2012), p. 2208.
[9]N. Hu, H. Buchholz, M. Herr, C. Spehr, and S. Haxter, "Contributions of different aeroacoustic sources to aircraft cabin noise," in *Proceedings of the 19th AIAA/CEAS Aeroacoustics Conference* (2013).
[10]C. D. Zevitas, J. D. Spengler, B. Jones, E. McNeely, B. Coull, X. Cao, S. M. Loo, A.-K. Hard, and J. G. Allen, "Assessment of noise in the airplane cabin environment," J. Exposure Sci. Environ. Epidemiol. **28**(6), 568–578 (2018).
[11]J. F. Wilby, "Aircraft interior noise," J. Sound Vib. **190**(3), 545–564 (1996).
[12]C. Hesse, P. Allebrodt, and A. Rege, *Multi-Physikalische Untersuchungen zum Transmissionsverhalten neuartiger Kabinenseitenwände* (*Multi-Physical Investigations on the Transmission Behavior of Novel Cabin Sidewalls*) (Deutsche Gesellschaft für Luft-und Raumfahrt-Lilienthal-Oberth eV, 2020).
[13]Y. Hüpel, S. Hoffmann, and S. C. Langer, "Influence of different turbulent boundary layer criteria on aircraft cabin noise," in *Proceedings of DAGA Hannover* (2024).
[14]C. Blech, C. K. Appel, R. Ewert, J. W. Delfs, and S. C. Langer, "Numerical prediction of passenger cabin noise due to jet noise by an ultra–high–bypass ratio engine," J. Sound Vib. **464**, 114960 (2020).
[15]C. Blech, H. K. Sreekumar, Y. Hüpel, and S. C. Langer, "Efficient solution strategies for cabin noise assessment of a wave resolving aircraft fuselage model," arXiv:2310.04734 (2023)
[16]M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (Springer International Publishing, Cham, Switzerland, 2020).
[17]R. Pieren, "Auralization of environmental acoustical sceneries: Synthesis of road traffic, railway, and wind turbine noise," Ph.D. thesis, Delft University of Technology, Delft, the Netherlands (2018); available at https://repository.tudelft.nl/islandora/object/uuid:8dbfb507-a0b0-4ccd-9772-88e213c69206?collection=research.
[18]R. Pieren, L. Bertsch, D. Lauper, and B. Schäffer, "Future low–noise aircraft technologies and procedures – Perception–based evaluation using auralised flyover," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany (September 9–13, 2019); available at http://pub.dega-akustik.de/ICA2019/data/articles/000098.pdf.
[19]R. Pieren, L. Bertsch, D. Lauper, and B. Schäffer, "Improving future low–noise aircraft technologies using experimental perception–based evaluation of synthetic flyovers," Sci. Total Environ. **692**, 68–81 (2019).

[20]S. Schade, R. Merino-Martinez, P. Ratei, S. Bartels, R. Jaron, and A. Moreau, "Initial study on the impact of speed fluctuations on the psychoacoustic characteristics of a distributed propulsion system with ducted fans," in *Proceedings of the 30th AIAA/CEAS Aeroacoustics Conference*, Rome, Italy (June 4–7, 2024).
[21]R. Merino-Martinez, R. Pieren, and B. Schäffer, "Holistic approach to wind turbine noise: From blade trailing–edge modifications to annoyance estimation," Renewable Sustain. Energy Rev. **148**, 111285 (2021).
[22]J. S. Pockelé, "Auralisation of modelled wind turbine noise for psychoacoustic listening experiments," Master's thesis, Delft University of Technology and Technical University of Denmark, Delft, the Netherlands (2023).
[23]A. P. C. Bresciani, "Physical and perceptual prediction of wind turbine noise," Ph.D. thesis, University of Twente, Twente, the Netherlands (2024); available at https://research.utwente.nl/en/publications/physical-and-perceptual-prediction-of-wind-turbine-noise.
[24]A. Allen and S. Krishnamurthy, "A hybrid transfer function procedure for broadband auralization within small flight vehicle interiors," in *Proceedings of InterNoise* (2023), pp. 927–937.
[25]A. R. Aumann, B. C. Tuttle, W. L. Chapin, and S. A. Rizzi, "The NASA auralization framework and plugin architecture," in *Proceedings of the INTER-NOISE NOISE-CON* (2015), pp. 1932–1943; available at https://ntrs.nasa.gov/api/citations/20160006430/downloads/20160006430.pdf.
[26]A. Farina and E. Ugolotti, "Subjective comparison of different car audio systems by the auralization technique," in *Proceedings of the Audio Engineering Society Convention 103* (1997).
[27]R. Sottek and B. Philippen, "Advanced methods for the auralization of vehicle interior tire-road noise," SAE Technical Paper (SAE, Warrendale, PA, 2012).
[28]H. K. Sreekumar and S. C. Langer, *elPaSo Core—Elementary Parallel Solver Core Module for High Performance Vibroacoustic Simulations* (2023).
[29]C. Blech, "Wave-resolving aircraft cabin noise prediction," Ph.D. thesis, TU Braunschweig, Düren, Germany (2022).
[30]C. Blech, C. K. Appel, R. Ewert, J. W. Delfs, and S. C. Langer, "Wave-resolving numerical prediction of passenger cabin noise under realistic loading," in *Fundamentals of High Lift for Future Civil Aircraft,* Vol. 145 of Notes on Numerical Fluid Mechanics and Multidisciplinary Design, edited by R. Radespiel and R. Semaan (Springer International Publishing, New York, 2021), pp. 231–246.
[31]Y. Hüpel, C. Blech, A. Franco, J. Delfs, B. Kirsch, and S. C. Langer, "Comparison of different noise sources for the simulative cabin noise assessment of an electrically propelled regional aircraft," in *Proceedings of Internoise in Chiba*, Tokyo, Japan (2023).
[32]E. Miranda, *Computer Sound Design: Synthesis Techniques and Programming*, 2nd ed. (Routledge, London, UK, 2002).
[33]"Note1," https://pixabay.com/sound-effects/180218-airplane-in-flight-cabin-rumble-tone-voices-loop-bahamas-23090/ (Last viewed November 2025).
[34]R. Merino-Martinez, R. Pieren, B. Schäffer, and D. G. Simons, "Psychoacoustic model for predicting wind turbine noise annoyance," in *Proceedings of the 24th International Congress on Acoustics (ICA)*, Gyeongju, South Korea (October 24–28, 2022); available at https://www.researchgate.net/publication/364996997_Psychoacoustic_model_for_predicting_wind_turbine_noise_annoyance.
[35]J. S. Pockelé and R. Merino-Martínez, "Pyschoacoustic evaluation of modelled wind turbine noise," in *Proceedings of the 30th International Congress on Sound and Vibration (ICSV)*, Amsterdam, the Netherlands (2024).
[36]F. Georgiou, B. Schäffer, A. Heusser, and R. Pieren, "Prediction of noise annoyance of air vehicle fly-overs using psychoacoustic models," in *Proceedings of the International Congress on Sound and Vibration (ICSV)*, Amsterdam, the Netherlands (July 8–11, 2024).
[37]G. F. Greco, R. Merino-Martinez, A. Osses, and S. C. Langer, "SQAT: A MATLAB-based toolbox for quantitative sound quality analysis," in *Proceedings of the 52nd International Congress and Exposition on Noise Control Engineering*, Tokyo, Japan (August 20–23, 2023); available at https://www.researchgate.net/publication/373334884_SQAT_a_MATLAB-based_toolbox_for_quantitative_sound_quality_analysis.
[38]ISO Norm 532–1: "Acoustics—Method for calculating loudness—Zwicker Method," Technical Report 1 (International Organization for Standardization, Geneva, Switzerland, 2017).
[39]W. Aures, "Berechnungsverfahren den sensorischen Wohlklang beliebiger Schallsignale (Procedure for calculating the sensory euphony of arbitrary sound signal)," Acustica **59**(2), 130–141 (1985).
[40]G. von Bismark, "Sharpness as an attribute of the timbre of steady sounds," Acta Acust. united Ac. **30**(3), 159–172 (1974).
[41]P. Daniel and R. Webber, "Psychoacoustical roughness: Implementation of an optimized model," Acta Acust. **83**, 113–123 (1997).
[42]A. Osses, R. García León, and A. Kohlrausch, "Modelling the sensation of fluctuation strength," in *Proceedings of the 22nd International Congress on Acoustics (ICA)*, Buenos Aires, Argentina (September 5–9, 2016); available at https://pure.tue.nl/ws/portalfiles/portal/52366479/Osses_Garcia_Kohlrausch_ICA2016_ID113.pdf.
[43]G. F. Greco, R. Merino-Martinez, and A. Osses, "SQAT: A sound quality analysis toolbox for MATLAB (version v1.1)," https://zenodo.org/records/10580337 (Last viewed October 2025).
[44]G. F. Greco, R. Merino-Martinez, and A. Osses, "SQAT: A sound quality analysis toolbox for MATLAB," https://github.com/ggrecow/sqat (Last viewed October 2025).
[45]U. Widmann, "Ein Modell der psychoakustischen Lästigkeit von Schallen und seine Anwendung in der Praxis der Lärmbeurteilung (A model of the psychoacoustic annoyance of sounds and its application in the practice of noise assessment)," Ph.D. thesis, Technische Universität München, München, Germany (1992); available at https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1854 (Last viewed November 2025).
[46]E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models* (Springer Science & Business Media, New York, 2013), Vol. 22.
[47]S. R. More, "Aircraft noise characteristics and metrics," Ph.D. thesis, Purdue University, West Lafayette, IN (2010).
[48]G.-Q. Di, X.-W. Chen, K. Song, B. Zhou, and C.-M. Pei, "Improvement of Zwicker's psychoacoustic annoyance model aiming at tonal noises," Appl. Acoust. **105**, 164–170 (2016).
[49]H. K. Ozcan and S. Nemlioglu, "In-cabin noise levels during commercial aircraft flights," Can. Acoust. **34**(4), 31–35 (2006); available at https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1854.
[50]R. Merino-Martinez, B. von den Hoff, and D. G. Simons, "Design and acoustic characterization of a psychoacoustic listening facility," in *Proceedings of the 29th International Congress on Sound and Vibration (ICSV)*, Prague, Czech Republic (July 9– 13, 2023).
[51]"Note2," this value corresponds to a more precise measurement featuring a B&K 4179 microphone for very low-level sound measurements, compared to the value originally reported in Ref. 50.
[52]J. S. Pockelé, "Graphical user interface for the psychoacoustic listening laboratory (PALILA GUI) (v1.1.1)" (2024). https://doi.org/10.5281/zenodo.11451003.
[53]L. L. Thurstone, "A law of comparative judgment," Psychol. Rev. **34**(4), 273–286 (1927).
[54]M. Pérez-Ortiz and R. K. Mantiuk, "A practical guide software analysing pairwise comparison experiments," arXiv:1712.03686 (2017).
[55]B. Berglund, P. Hassmén, and A. Preis, "Annoyance and spectral contrast are cues for similarity and preference of sounds," J. Sound Vib. **250**(1), 53–64 (2002).
[56]W. M. Beltman, R. A. Doherty, E. Salskov, P. J. Corriveau, D. Gabel, and E. Baugh, "Applications of psychoacoustics to information technology products," J. Acoust. Soc. Am. **123**(5), 3160 (2008).
[57]Delft University of Technology, "TU Delft facts and figures," https://www.tudelft.nl/en/about-tu-delft/organisation/facts-and-figures (Last viewed October 2025).
[58]D. Zwillinger and S. Kokoska, *CRC Standard Probability and Statistics Tables and Formulae* (CRC Press, Boca Raton, FL, 2000).
[59]D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," arXiv:1109.2378.
[60]H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," Ann. Math. Statist. **18**, 50–60 (1947).
[61]"Note3," Spearman's correlation is a "soft" correlation as it only identifies monotonic relationships compared to e.g. a Pearson's correlation, which would evaluate linear relationships.