



HAL
open science

Effect of Environment-Related Cues on Auditory Distance Perception in the Context of Audio-Only Augmented Reality

Vincent Martin, Isabelle Viaud-Delmon, Olivier Warusfel

► **To cite this version:**

Vincent Martin, Isabelle Viaud-Delmon, Olivier Warusfel. Effect of Environment-Related Cues on Auditory Distance Perception in the Context of Audio-Only Augmented Reality. *Applied Sciences*, 2022, 12 (1), pp.348. 10.3390/app12010348 . hal-03579718

HAL Id: hal-03579718

<https://hal.science/hal-03579718>

Submitted on 18 Feb 2022

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

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

Article

Effect of Environment-Related Cues on Auditory Distance Perception in the Context of Audio-Only Augmented Reality

Vincent Martin ^{*,†}, Isabelle Viaud-Delmon and Olivier Warusfel [†] 

Sorbonne Université, Ircam, CNRS, Ministère de la Culture, Sciences et Technologies de la Musique et du Son, STMS, F-75004 Paris, France; isabelle.viaud-delmon@ircam.fr (I.V.-D.); olivier.warusfel@ircam.fr (O.W.)

* Correspondence: vmartin@ircam.fr

† These authors contributed equally to this work.

Abstract: Audio-only augmented reality consists of enhancing a real environment with virtual sound events. A seamless integration of the virtual events within the environment requires processing them with artificial spatialization and reverberation effects that simulate the acoustic properties of the room. However, in augmented reality, the visual and acoustic environment of the listener may not be fully mastered. This study aims to gain some insight into the acoustic cues (intensity and reverberation) that are used by the listener to form an auditory distance judgment, and to observe if these strategies can be influenced by the listener's environment. To do so, we present a perceptual evaluation of two distance-rendering models informed by a measured Spatial Room Impulse Response. The choice of the rendering methods was made to design stimuli categories in which the availability and reproduction quality of acoustic cues are different. The proposed models have been evaluated in an online experiment gathering 108 participants who were asked to provide judgments of auditory distance about a stationary source. To evaluate the importance of environmental cues, participants had to describe the environment in which they were running the experiment, and more specifically the volume of the room and the distance to the wall they were facing. It could be shown that these context cues had a limited, but significant, influence on the perceived auditory distance.

Keywords: audio-only augmented reality; auditory distance perception; acoustic cues weighting strategies; reverberation-related cue; visual spatial boundary



Citation: Martin, V.; Viaud-Delmon, I.; Warusfel, O. Effect of Environment-Related Cues on Auditory Distance Perception in the Context of Audio-Only Augmented Reality. *Appl. Sci.* **2021**, *12*, 348. <https://doi.org/10.3390/app12010348>

Academic Editor: Hyunkook Lee

Received: 31 October 2021

Accepted: 17 December 2021

Published: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Audio-only augmented reality (AAR) consists of using spatial audio processing to enhance the real environment of a user with virtual sound events. A key requirement of AAR is to produce the localization of virtual sound sources in a real environment as naturally as possible. Auditory distance perception is one component of the spatial localization of virtual sound sources and depends on a combination of several acoustic and nonacoustic distance cues [1]. The production of a virtual sound source in a real environment calls for the choice of a spatial audio processing method. The main purpose is to achieve a seamless integration of the virtual sound source into the real listening environment. The method has a direct influence on the reproduction of acoustic cues conveyed by the room effect, possibly affecting the auditory distance perception of a virtual sound source. In AAR scenarios, the characteristics of the environment (geometry, room acoustics) are not necessarily mastered; thus, a room divergence effect [2] may occur between the real listening environment and the acoustic cues conveyed by the virtual sound source processing.

1.1. Acoustic Cues Combination for Auditory Distance Perception

This study focuses on the auditory distance perception of virtual sound sources presented in front of the listener, in a real environment, and at distances beyond the near-field. Both auditory and nonauditory cues affecting the auditory distance perception of

frontal sources have been identified [3,4]. In the free field, the sound level at the ears decreases as the distance to the source increases. Therefore, sound level is considered as a primary monaural cue for distance perception. Sound level is often classified as a relative cue, as it needs a priori knowledge of the source sound power to be interpreted by a listener.

The direct-to-reverberant (DRR) energy ratio is often considered as the cue best defining the contribution of reverberation for distance. When the distance to a sound source increases in a reverberant environment, the DRR decreases. DRR has been proposed as an absolute distance cue, contrary to sound level, as it does not require a priori information about the source power to be interpreted as a distance cue. Nonetheless, it can be argued that familiarity with the acoustical environment is necessary to consider this cue as absolute, the reverberation decay being specific to each environment.

However, it is very unlikely that the auditory system can effectively separate the direct sound from reverberation and compute a physical DRR [5]. Different studies tried to define a more relevant reverberation-related distance cue [5–8]. Three alternative cues to DRR have already been evaluated: an estimation of the perceived distance with an early-to-late energy power ratio [6], the interaural coherence [9,10], the initial time delay gap (ITDG) depicting the time of arrival of the first reflections [11], and monaural changes in the spectral centroid or in frequency-to-frequency variability in the signal [9]. Kopčo and Shinn-Cunningham [5] have proposed that the best candidate is most likely an early-to-late power ratio. Concerning the temporal limit between what is considered as the early energy, which contains the direct sound and some early reflections, and the late reverberation, no clear conclusions have been drawn [6]. The relevance of the early-to-late energy ratio as a reverberation-related distance cue, and how this cue is combined with the sound level to process an auditory distance judgment, are examined in this study.

Moreover, the combination of the sound level and reverberation-related cues is the main factor driving the auditory distance perception of stationary sources. Zahorik [12] showed that the perceptual system uses weighting strategies of intensity and of reverberation-related cues flexibly to produce a distance percept, depending on the characteristics of the listening situation: the stimulus nature, the angular position, and the source–receiver distance. He also suggested that the weighting of cues may rely on the consistency associated with each cue [4]. Cues that are either unavailable (e.g., reverberation in a free field environment) or unreliable (e.g., a cue that stays constant with distance) are given less perceptual weight in the combination process. To conclude, he suggests that such an individual weighted sum of acoustic cues assessments may yield the final distance percept. In this study, the influence of environmental characteristics on the acoustic cues weighting strategies adopted by participants to produce an auditory distance judgment is investigated.

1.2. Vision Contribution

Another aspect that must be considered for auditory distance perception in an AAR scenario is the influence of the visual environment. Indeed, in most cases, users of an AAR scenario see their own environment while it is enhanced with virtual sound sources. The geometry of the room is not necessarily totally known and implemented in AAR applications, and several studies have suggested that the visual context influences distance perception of the sound sources through mechanisms related to multisensory integration [13,14]. Divergent visual context was also demonstrated to have an effect on the externalization of virtual sound sources [2].

It has long been argued that vision could calibrate auditory distance perception [15,16]. Calcagno et al. [13] demonstrated that a priori visual information about an environment can be beneficial to the accuracy of auditory distance perception of real sound sources.

The authors hypothesized that the representation of the visual space can serve as a spatial reference into which distance cues are integrated to create an auditory distance judgment. Consequently, the reference space conditioned by visual information scales auditory distance perception and can compress or expand the auditory distance responses, depending on the congruence of the acoustic information with visual spatial cues. This

hypothesis about calibration of space perception has been tested with stimulation of a single or multiple sensory modalities. Previous studies have notably demonstrated that distance perception responses can be calibrated by information collected by the same or a different sensory modality. For instance, Etchemendy et al. [17] demonstrated how sound reverberation conditions affect visual distance perception, with highly reverberant rooms implying an overestimation of visual distance perception.

Given these results about spatial calibration of a sensory modality, the specific impact of visual spatial boundaries on the perception of virtual sound sources is addressed.

1.3. Research Aim

The aim of this study is twofold: (1) to evaluate the use of early energy relative to late reverberation and of the global energy to judge the absolute distance to a virtual sound source; (2) to investigate the influence of two environmental characteristics, the visual spatial boundary and the volume of the environment, on the auditory distance perception of a virtual sound source.

An experiment was conducted with online listening tests in which participants had to evaluate the distance of virtual sound sources produced by different rendering methods. The choice of the rendering methods was made to design stimuli categories in which the availability and reproduction quality of acoustic cues are different. Two different distance-rendering models have been evaluated in contrast to a method based on actual measurements of spatial room impulse responses (SRIR).

A first model, labeled “envelope-based model”, was used to investigate whether energetic discrepancies between the model and measurements influence distance perception. Its evaluation enabled the investigation of whether a portion of the early energy of an impulse response can be considered as fused with the direct sound to establish an early-to-late energy ratio as a relevant distance cue.

A second model, labeled “intensity-based model”, was designed to accurately reproduce intensity while maintaining reverberation-related cues constant with distance. The perceptual evaluation of this model provided insight into participants’ acoustic cues weighting strategies, which included intensity and reverberation-related cues.

The experiment gathered two pools of 60 normal-hearing adult participants, each performing the experiment with a different inclusion criterion on the distance between the participant and the facing wall. The analysis of the differences between both groups’ results was made to investigate the influence of this visual spatial boundary condition on auditory distance perception. Participants of both groups had to report the dimensions of the room in which they were during the experiment. The volume estimated from this report was considered as a primary descriptor of the quantity of reverberation in the room. As participants were each accustomed to the specific acoustic properties of the room where they were running the experiment, it was decided to study, through each participant’s reported room volume and reported distances for each distance-rendering method, the influence of a possible acoustic familiarity on auditory distance perception.

2. Design of Distance-Rendering Methods

Three methods were designed to reproduce distance of a virtual sound source in a single acoustical environment: a classroom at IRCAM (semi-damped, dimensions: 8.7 m × 4.7 m × 3.5 m—L × W × H, T_{60} at 1 kHz of 0.55 s). Their different designs aim to partly reproduce the acoustic cues conveyed by real measurements.

2.1. Reference Method Based on Measurements

Different SRIRs were measured in the classroom with a spherical microphone array (SMA), *Eigenmike*[®] EM32 from MH Acoustics. The choice to use SRIRs was motivated by the flexibility of this solution. While a binaural impulse response (BRIR) only allows temporal modifications, an SRIR can be subject to spatial modifications and binaural decoding on any HRTF set. Nine SRIRs were measured for distances ranging from 1

to 7 m (1, 1.5, 2, 2.5, 3, 4, 5, 6, 7 m) by changing the speaker position and with a single microphone position. The different measurements were performed on an axis shifted 60 cm from the median line of the room to avoid spatial symmetry of the first reflections. This precaution was taken to favor decorrelation of the binaural stimuli, which also optimized their externalization [18]. These SRIRs were converted from a fourth-order ambisonic signal to BRIRs. Here, an “energy-preserving” decoding method was used [19]. In order to decode the SRIR to binaural format, the spherical harmonic excitation of the fourth-order ambisonic signal was sampled on virtual speakers’ positions. A virtual spherical array of 24 uniformly distributed speakers, following an equal-area grid, was used and the listener was considered at its center. To produce the final BRIR signals, each of the speakers’ signals was filtered with the head-related transfer function (HRTF) of a *Neumann KU100* dummy head, according to its direction. This set of nine resulting BRIRs will be referred to as the “reference room”. The BRIR corresponding to a distance of 1 meter was used as an initial impulse response for the two following distance-rendering models.

2.2. Envelope-Based Model

This model is based on a simplified representation of the energy envelope of the impulse response, here divided into two temporal segments: the early energy E_s , comprising both the direct sound and early reflections, and the late reverberation energy E_{rev} . The energy of these two temporal segments are modified according to the desired distance.

Different perspectives can be considered to demarcate the time limit between these two segments. This transition time can be derived from perceptual considerations, regardless of the room geometry. A time limit of 50 ms has been considered in room acoustics as criteria describing the quality of speech perception, such as *definition* ($D50$) or *clarity* ($C50$). In order to delineate the “useful sound” in opposition to the “detrimental sound”, Lochner and Burger [20] consider a weighting function equal to 1 until 35 ms and then linearly decreasing up to 95 ms.

This transition time can also refer to the physical properties of the room and of its impulse responses. After reaching a sufficiently high echo density and modal overlap, the room reverberation exhibits an exponentially decaying stochastic behavior. This lower time limit is referred to as the mixing time. Several estimators of the mixing time have been suggested in the literature. The estimation $t_m = \sqrt{V}$ (with t_m the mixing time in ms and V the volume of the room in cubic meter) was proposed in [21]. Other methods rely on the evaluation of the diffuseness of the sound field from the statistics of the echoes observed in the impulse response. This estimation may either be conducted in the time domain [22,23], or in the spatial domain, when an SRIR is available [24]. In our case, a temporal estimation of the mixing time was chosen and a value of 15 ms (after the onset delay) was provided. This time will define the temporal limit between what is considered as the early part of the BRIR and the reverberation tail.

The design of this model was based on the hypothesis that the early-to-late energy ratio is a relevant reverberation-related distance predictor [7]. Bronkhorst tested a predictive model in which he used an early-to-late energy power ratio to compute the perceived distance with success [6]. He also highlighted the role of the spatial distribution of early reflections. In [7], he showed that the early energy definition could be based on a lateralization window (i.e., aggregation of the reflections sharing an ITD close to that of the direct sound) rather than on a time window. Here, the mixing time was chosen as a first estimation of the length of the temporal integration window of the early energy.

The model applied to alter the initial BRIR to control its apparent source distance is inspired by previous work from Jot et al. described in [25]. In the proposed approach, the sound source distance is driven through the control of two temporal segments, the direct sound energy E_{dir} and the reverberated energy E_{rev} . The level of the direct sound according to the source distance is expressed as follows:

$$E_{dir}(f, d) = \frac{S_\phi(f)^2 \mu(f)^d}{4\pi c d^2} \quad (1)$$

with c the sound celerity, f the frequency, $\mu(f)$ the frequency-dependent sound absorption for a 1-m propagation in the air, and $S_\phi(f)$ the free field transfer function of the source in the direction of the receiver.

The level of the diffused part of the reverberation after the mixing time τ in the impulse response is expressed as follows:

$$E_{rev}(\tau, f) = \frac{Tr(f)S_d(f)}{13.81 * V} \exp(-13.81 * \frac{\tau + \frac{d}{c}}{Tr(f)}) \quad (2)$$

with $Tr(f)$ the reverberation time, V the volume of the room, $S_d(f)$ the diffuse-field transfer function. The dependence of the reverberation energy with the distance d agrees with Barron's revised theory on energy relations in the room response [26]. In the present study, some further simplifications are made. Air absorption is not considered. The spatial dependence of the free field transfer function is also neglected, as the source will always be heard from its frontal direction. Moreover, the attenuation law of the direct sound will be extended to the whole early energy, thus introducing discrepancies between the measured and the simulated early reflections. Under these assumptions, the modifications that should be applied to the early and late segments of the initial impulse response measured at a distance d_{ref} to derive the new impulse response at distance d can be written as follows:

$$E_s(d) = E_s(d_{ref}) * \frac{d_{ref}^2}{d^2} \quad (3)$$

$$E_{rev}(d, f) = E_{rev}(d_{ref}, f) * \exp(-13.81 * \frac{d - d_{ref}}{cTr(f)}) \quad (4)$$

This model will be referred to as the "envelope-based model". Applying Equations (3) and (4), nine BRIRs were generated upon modification of the initial impulse response measured at $d_{ref} = 1$ m to simulate the different distances selected in the classroom.

2.3. Intensity-Based Model

This simplified model extrapolates the BRIRs corresponding to different distances by applying a global gain to the initial reference BRIR measured at 1 m. The gain used to extrapolate the BRIR for a given distance was tuned so that the loudness of the resulting stimulus corresponded to the loudness of the stimulus generated with the BRIR measured at the same distance. The loudness criterion used here is EBU R128. This model is used in contrast to the envelope-based model and the rendering method based on measurements, as no modification of the reverberation-related cues is present when distance increases.

This second model gives insights into the participants' auditory cues weighting strategies, as well as a potential influence of the environment on these strategies. Participants that primarily rely on intensity to infer a distance judgment were expected to show similarities in distance reports for stimuli created with this model and for stimuli generated with the reference measurements. Participants who primarily rely on reverberation-related cues would exhibit significant differences between distance reports associated with both types of stimuli.

2.4. Energy Differences between the Methods

The differences between the BRIRs generated with the two models and the measurements taken in the reference room are analyzed under the scope of the energy contained in different temporal segments of the impulse responses. Figure 1 depicts the energy contained in the early part of the impulse response, E_s , preceding the considered mixing time, as well as in the reverberation, E_{rev} , of each impulse response. When compared to the measured impulse responses, the envelope-based model mainly underestimates the early energy E_s , a difference that increases for longer distances. This behavior comes from the attenuation law, which is proportional to the square of the distance and is applied

uniformly to the whole early energy segment and not only to the direct sound. Contrarily, the late energy tends to be slightly overestimated, as distance increases with a maximum difference of 1 dB. As a summary, the envelope-based model correctly reproduces the energy contained in the late reverberation but strongly underestimates a part of the energy of early reflections.

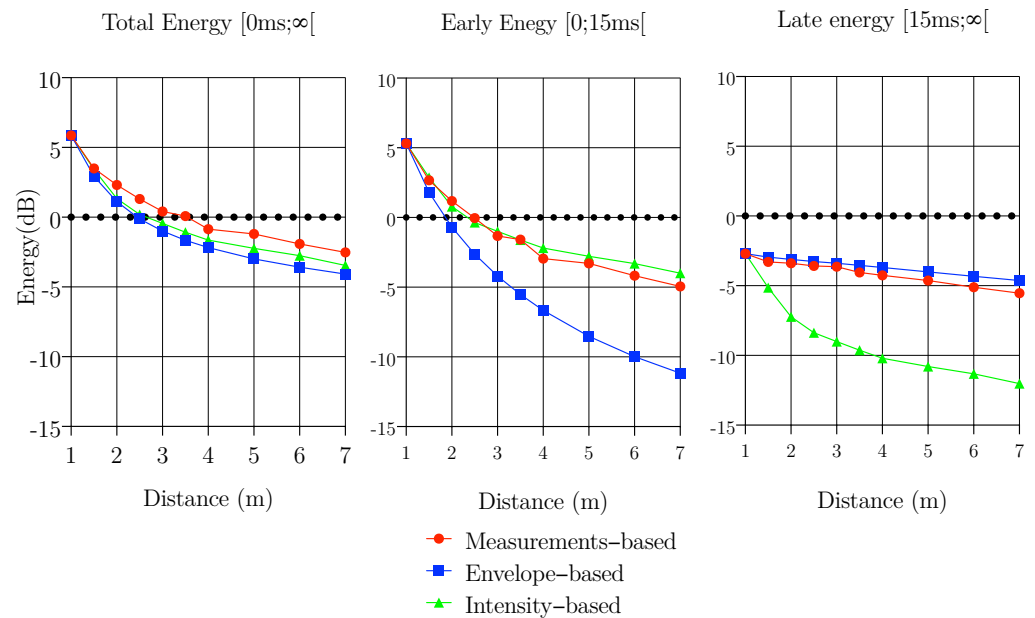


Figure 1. Evolution of the total energy E_{tot} (left), the early energy E_s (middle), and reverberation E_{rev} (right) according to the source distance for the BRIRs generated by the two models and for the measured ones.

For the intensity model, the attenuation law used to tune the global gain of the impulse responses produces opposite differences in terms of energy contained in the early and late time segments of the generated BRIRs. Early energy tends to be slightly overestimated as distance increases, with a maximum difference of 1 dB. Late energy is highly underestimated as distance increases.

Compared to the measurements, both models show a slight underestimation of the total energy as the distance increases, however not exceeding 1 dB for the intensity-based model and 2 dB for the envelope-based model, a value slightly above the just-noticeable difference threshold of intensity perception [27]. The attenuation law used for the intensity-based model is based on a loudness criterion (EBU R128) on the resulting stimuli. Hence, a difference in terms of total energy of the BRIRs is observed.

3. Material and Methods

Two different pools of online participants were recruited according to the environment in which they were running the experiment. In the Close Wall (CW) group of participants, the seating distance from the wall they were facing was required to be less than 3 m. Participants who were seated at a distance of more than 5 m from the wall were included in a Far Wall (FW) group.

3.1. Participants

Online participants were recruited via the *Prolific* recruitment platform (www.prolific.co, 15 December 2020). A total of 120 participants were recruited and paid to complete the task. An initial criterion for participation was the declared absence of hearing problems or uncorrected vision. A second criterion requested that they be equipped with headphones. A Swiss German spoken text was used for the stimuli, the reason for a final recruitment criterion being the absence of knowledge of the German language, so that participants

would not focus on its semantic content. The participants were screened on the basis of the self-reported distance from the wall they were facing. When the experiment had reached the targeted number of participants, data were manually accepted into the final results pool after inspection of the participants' individual data. After identification of statistical outliers, fifty-five out of sixty participants (age range 18 to 42; $M = 25$; $SD = 6.5$; 23 females, 32 males) were included in the CW group. Fifty-three out of the sixty participants were included in the FW group (age range 18 to 46; $M = 25$; $SD = 6$; 23 females, 30 males).

3.2. Auditory Stimuli

A 5-s anechoic recording of a sentence in Swiss German pronounced by a male speaker was preconvolved with each of the BRIRs, either measured or generated by the models. A total of 27 stimuli, simulating 9 different sound source distances for both rendering models and 9 stimuli based on the measures for the same distances, were used. Each stimulus apparent distance was evaluated 4 times throughout the test trial, for a total of 108 stimuli presentations.

3.3. Report Method

A vertical visual analogue scale (represented by a vertical continuous slider) was used to report the perceived distance of a sound stimulus. The top of the slider corresponded to a distant impression to the sound source while the bottom corresponded to an impression of proximity.

3.4. Online Experiment Design

The experiment task was designed using the software *Psychopy* [28]. This software allows running online experiments by pairing it with *Pavlovia*, a platform designed for hosting online behavioral experiments based on HTML/Javascript codes. The choice of a recruitment service such as *Prolific* allowed us to guarantee anonymity of collected data by *Pavlovia*.

3.5. Procedure

After reading the information note and approving the consent form, participants were directed to the experiment.

Participants had to fill a questionnaire in which they were invited to report the main geometric features of the room where they were running the experiment: surface (in square meters) and height of the room (in meters), as well as the distance to the wall (in meters) they were facing. They were requested to use circumaural headphones.

A screening process was launched to ensure the correct use of headphones: two audio clips corresponding to the weakest and loudest stimuli were played successively. Participants had to adjust the sound level for the quietest stimulus to be heard clearly. The loudest stimulus was then played, to ensure that it was heard at a comfortable level. Finally, broadband noise was displayed laterally to ensure that participants were correctly wearing their headphones.

Afterwards, a training session began. Participants were instructed to look at the facing wall while listening to a stimulus, and to judge the distance of the stimulus using the slider. A total of 108 stimuli (9 distances \times 3 rendering methods \times 4 presentations) were presented in a random order.

After the training session, the experiment began. The order of the stimuli within the experiment was randomized. During the trials, participants triggered the stimulus playback, but the stimulus was only played once.

At the end of the experiment, participants were invited to fill out a final questionnaire to collect feedback on the procedure. The total duration of the procedure was 15 min. Two weeks later, a second debriefing was conducted through *Prolific* electronic mail service, to ensure the validity of the answers about the room characteristics.

4. Results

4.1. General Results

In the following statistical analysis, the logarithmic value of the geometric mean of each participant's responses over a type of stimulus was considered as a dependent variable (27 different stimuli: 9 distances \times 3 rendering methods). For both groups separately, a focus was placed on ensuring the normality of the dependent variables associated with a specific stimulus. A Jarque–Bera test indicated that in all cases for both groups, the null hypothesis “the data were normally distributed” was not rejected. Homoscedasticity was checked with Levene's tests, and no large deviations of the variance were detected across the conditions. Geometric means over all participants reports for each rendering methods are displayed in Figure 2.

A repeated measures ANOVA ($\alpha = 0.05$) was conducted on geometric mean of reported distances of each subject with RENDERING (three rendering methods) and DISTANCE (nine levels ranging from 1 to 7 m) as within-subjects factors and GROUP (two groups) as intersubject factor. The analysis revealed a main effect of DISTANCE ($F(1, 8) = 580.8, p < 0.001$) and a main effect of RENDERING ($F(1, 2) = 68.0, p < 0.001$), as well as a DISTANCE \times RENDERING interaction ($F(1, 16) = 11.7; p < 0.001$) and a GROUP \times DISTANCE interaction ($F(1, 8) = 2.05; p = 0.039$). No significant main effect of GROUP was revealed ($F(1, 1) = 0.74, p > 0.05$).

4.2. Effect of Room Volume

Participants had to answer questions about the boundaries and dimensions of the room they were in. Their answers were used to estimate the volume of the room in which they performed the task (see Figure 3). In the FW group, the estimated room volumes were larger on average and more dispersed than in the CW group. A nonparametric statistical test confirmed that both distributions were significantly different (Kruskal–Wallis: $H = 16.85, p < 0.001$). Consequently, in the following analysis, the influence of volume on auditory distance perception is investigated separately in each group. For each group, separate linear regressions were applied to responses associated with a specific stimulus (27 different stimuli: 9 distance \times 3 rendering methods), with the room volume as a factor. In the CW group, significant correlations between the room volume and distances longer than 4 m generated with the intensity-based model were found ($D = 4 \text{ m } F = 4.16, p = 0.034$; $D = 5 \text{ m } F = 6.05, p = 0.017$; $D = 6 \text{ m } F = 5.06, p = 0.028$; $D = 7 \text{ m } F = 6.40, p = 0.014$). For these specific stimuli, larger volumes were significantly correlated to a larger compression of the reported distances. No significant correlations were found for shorter source distances ($p > 0.05$) nor for all distances in the FW group.

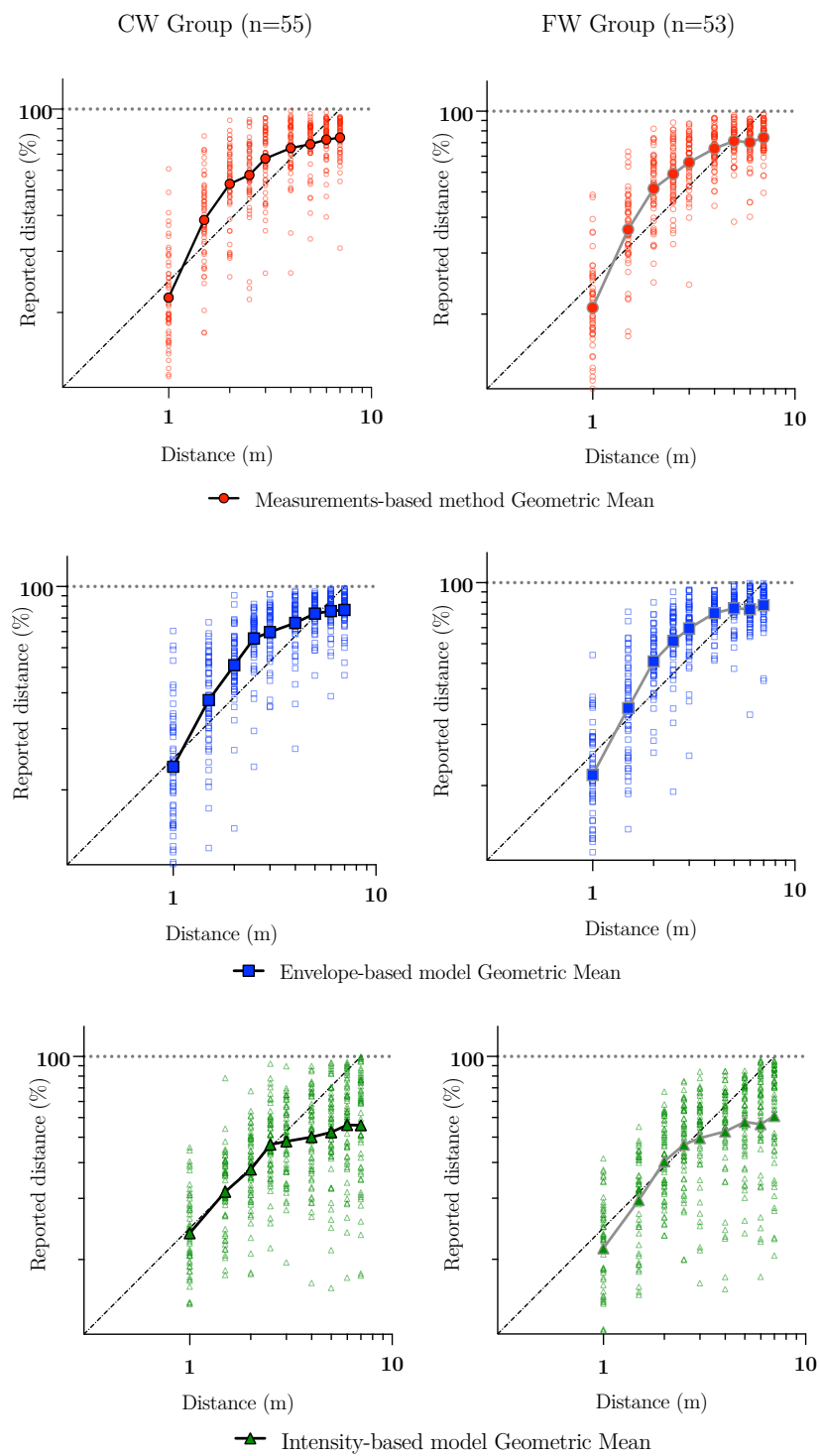


Figure 2. Geometric means of reported distance according to the rendering method used to generate the stimuli and the visual spatial boundary condition. Clear dots represent the geometric mean of a single participant for a given distance, while opaque dots represent the geometric mean over all participants for a given distance.

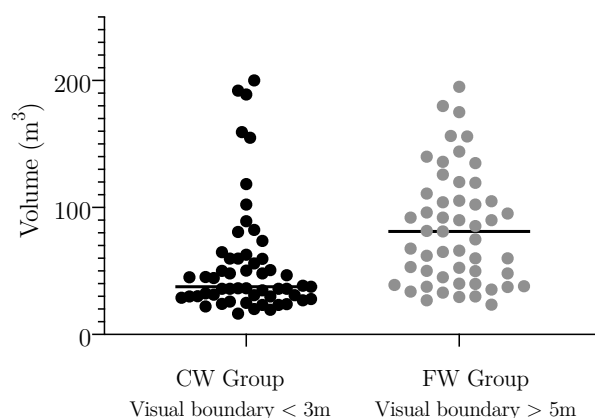


Figure 3. Estimated room volume of participants in each group (CW group: $M = 37.5 \text{ m}^3$; FW group: $M = 81.7 \text{ m}^3$).

4.3. Compression Effect Quantification across Rendering Methods

The influence of the distance-rendering method and of the environmental context (visual spatial boundary and volume) on the compression effect inherent to auditory distance perception was quantified with power functions, similarly to what is usually employed in studies implying the quantification of psychometric distance functions [13,29]. Power functions of the following form were fit (linear fit on logarithmic values) to the geometric mean reported distances with each rendering method:

$$D = k * d^a \quad (5)$$

where D is the perceived auditory distance, d is the sound–source distance, k a constant, and a a power-law exponent. For comparison purposes, the reported distances that are considered in the fitting result from a linear conversion of scaled reports made by participants on the visual analogue scale. The maximum of the visual scale (0% on the slider corresponds to the minimal distance of 0 meter and 100% to the maximum possible distance) was 7 m. The fitting parameters, k and a , were used as measures of auditory distance perception accuracy. The exponent a indicates the amount of nonlinear compression ($a < 1$) or expansion ($a > 1$) and is equivalent to the slope when representing reported distances on a logarithmic scale. The constant k indicates the amount of linear compression ($k > 1$) or expansion ($k < 1$) in the function and is equivalent to the intercept when representing reported distances on a logarithmic scale.

The fitting parameters k and a quantify the amount of compression in auditory distance perception: the closer they are to 1, the less compression there is. Hence, the fitting parameters are mentioned in the following as compression coefficients.

The cross-correlation coefficient R^2 can be used to assess the fit's quality. For each rendering method, a positive correlation was found between a fitted power function and the distances reported by each participant for the rendering method based on measurements ($\text{mean}(R^2) = 0.86$; $SD = 0.12$) and for the envelope-based model ($\text{mean}(R^2) = 0.84$; $SD = 0.11$). The intensity-based model had a moderately positive correlation ($\text{mean}(R^2) = 0.73$; $SD = 0.28$). The mean value of R^2 obtained on reported distances of stimuli generated by the intensity-based model is significantly lower than the mean obtained by the envelope-based model. In that respect, stimuli rendered with the intensity-based model tend to produce higher intrasubject variability.

To further investigate the similarity between the rendering method based on measurements and the envelope-based model at an individual level in both groups, each participant's compression coefficients were compared. In Figure 4, the individual compression coefficients a and k collected among the two groups are contrasted between those estimated for each model (ordinate) and for the rendering method based on measured

BRIRs (abscissa). For each distribution of dots, a linear regression analysis was run. In all eight analyses, the null hypothesis “the coefficient of the regression slope is equal to zero” was rejected ($p < 0.001$). The distributions evaluating individual similarity between the envelope-based model and the measurements produce regression slopes close to 1, while the distributions related to the intensity-based model produce regression slopes significantly inferior to 1.

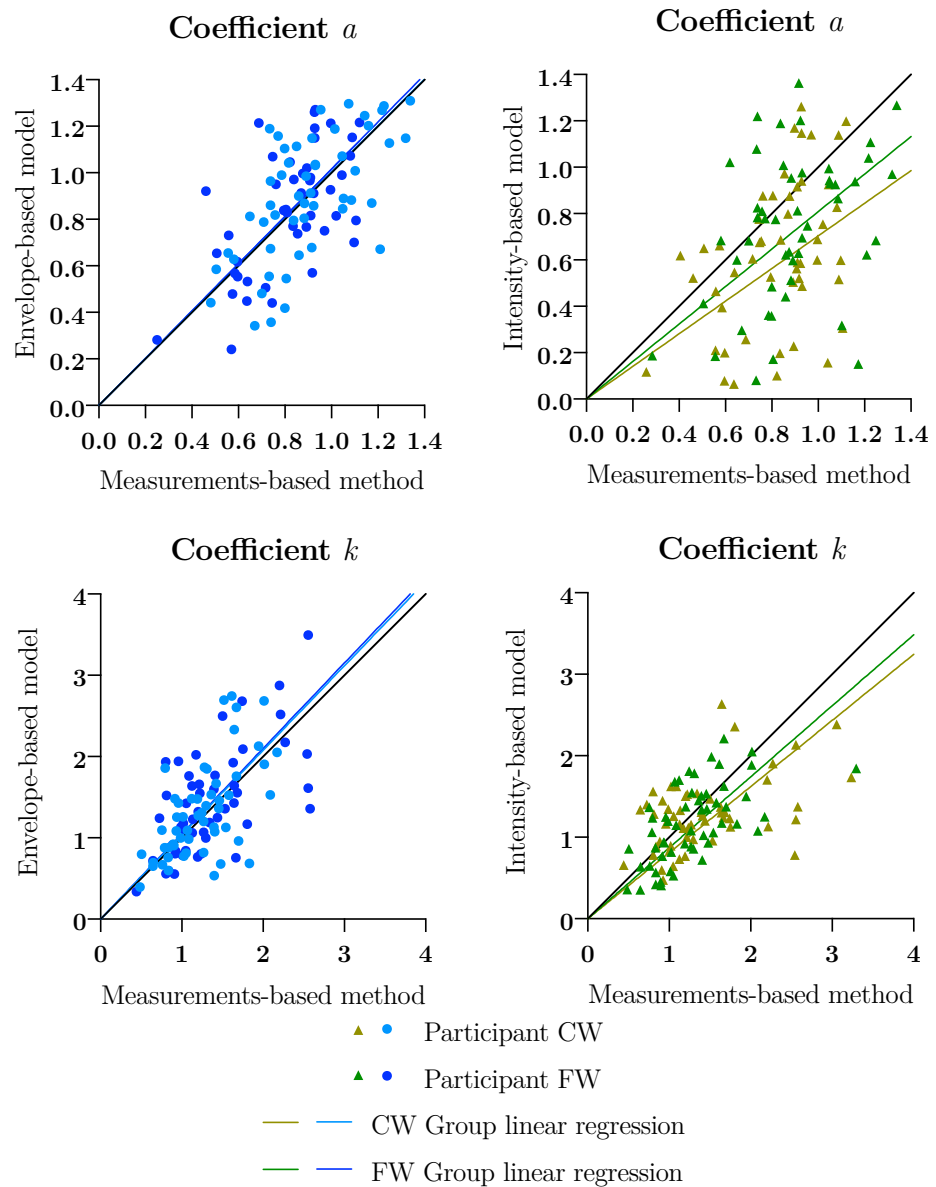


Figure 4. Comparison of the individual compression coefficients a (above) and k (below) between the rendering method based on measurements and the models. On the **left**, the envelope-based model (blue). On the **right**, the intensity-based model (green). Dark symbols correspond to parameters estimated from participants of the FW group, and light symbols are of the CW group. Dark lines correspond to the regression curves associated with the FW group, and light lines correspond to the CW group.

The distribution of compression coefficients a and k illustrates the presence of a similarity, in terms of auditory compression effect, between the envelope-based model and the rendering method based on measurements. While the distribution of coefficients for the intensity-based model is more dispersed, showing less correlation at an individual level, the distribution also highlights that most compression coefficients a estimated for the

intensity-based model are lower than those obtained for the rendering method based on the measurements.

4.4. Visual Spatial Boundary and Compression Coefficients

In order to study the influence of the environment on auditory perceived distance compression, the following analysis method, inspired by the work of Anderson and Zahorik [29] was applied. The mean values of compression coefficients for each distance-rendering method are displayed in Figure 5 for each group. The mean values of the coefficients a and k suggest that the compression effect is weaker in the FW group. For each model, independent sample t-tests were conducted between the values of coefficients a and k estimated from the data of the CW group and those estimated from the data of the FW group. The null hypothesis “the two population means are equal” is rejected for all tests, except for the comparison of the values of the coefficient a for the intensity model. According to these results, the visual spatial boundary had a moderate but significant effect on the compression effect of auditory distance perception.

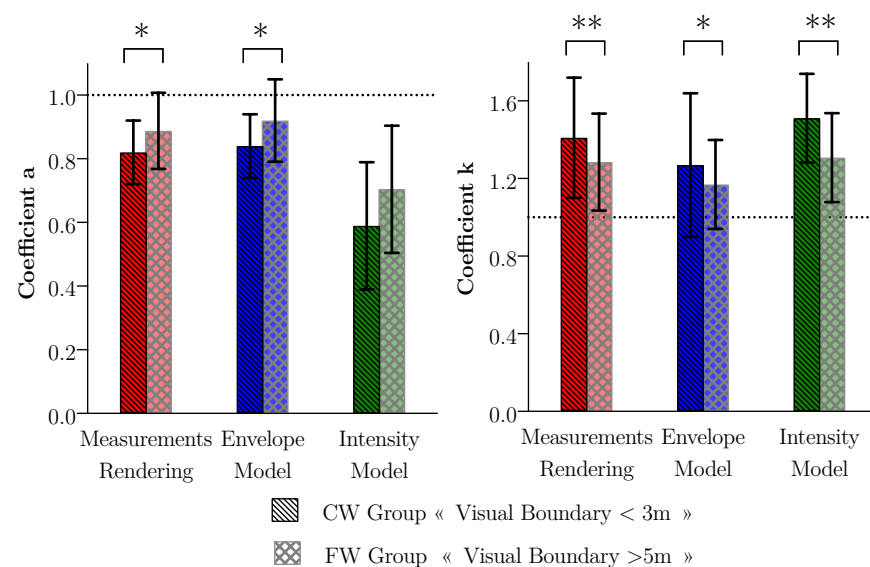


Figure 5. Mean and standard deviation of individual fitting coefficients a (left) and k (right) classed by groups (CW group (black, $n = 55$) and FW group (gray, $n = 53$)) and by rendering methods: measurements-based (red), envelope-based model (blue), and the intensity-based model (green). Coefficients for FW group are closer to 1 compared to those obtained for the CW group, demonstrating a lower compression effect of auditory distance perception. (*) t -test indicated a p -value $p < 0.05$; (**) t -test indicated a p -value $p < 0.01$.

4.5. Room Volume and Compression Coefficients

The effect of room volume on reported distances was investigated by running a linear regression analysis applied to the individual a compression coefficients estimated in the CW group for the intensity-based model (which values are displayed in Figure 6). The logarithm of each reported volume was considered as factor for this regression analysis. According to the results of this analysis, a higher volume is significantly correlated with a lower parameter a ($F = 5.60$; $p = 0.021$). This result confirms that a larger volume is linked to a stronger compression effect of reported distances of stimuli generated by the intensity-based model.

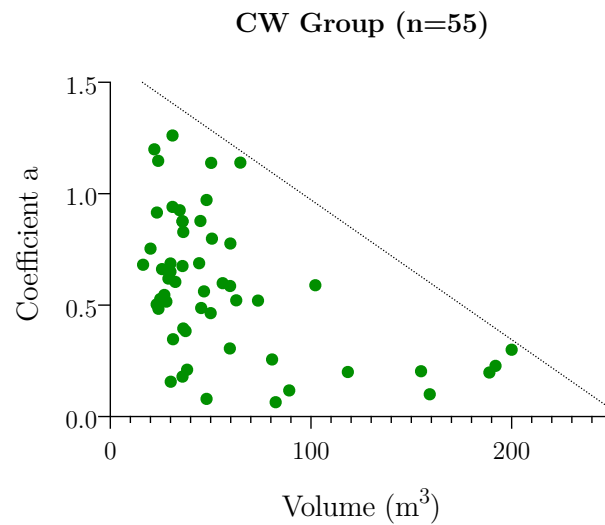


Figure 6. Value of the fitting coefficient a (green dots) estimated on the intensity-based model, for participants of the CW group as a function of the self-reported volume of the room.

5. Discussion

The current study evaluated the importance of different cues known to be important to auditory distance perception. The procedure applied here aimed to be close to a simple AAR scenario, with participants having access to the vision of the room during the experiment. The role of the global intensity and of the energy contained in various temporal segments of the room impulse response was investigated through the comparison of three different rendering methods. The results are discussed under the criteria of accuracy and variability of reported distances relative to the reference rendering method based on real measurements.

The influence of environmental context cues was also examined. The effect of the visual spatial boundary on the auditory distance perception was assessed by comparing the value of the mean compression coefficients a and k between both groups. The impact of room volume was investigated through linear regression fitting on distance reports and each rendering method associated compression coefficients.

5.1. Early-to-Late Energy Ratio as a Distance Cue

The envelope-based model and the rendering method based on measurements were shown to provide very similar auditory distance perceptions. This result was further confirmed by the post hoc analysis applied to each tested distance. However, they differ significantly by design in terms of their early-to-late energy ratio (See Figure 7). The envelope-based model strongly underestimates the early energy for offset times under 50 ms, especially with increasing distances. This perceptual similarity raises the question of the definition of a reverberation-related distance cue, and which definition could best explain the perceptual similarity between these two methods. Previously, Kopčo and Shinn-Cunningham [5] have suggested that an early-to-late power ratio is the best candidate to act as a reverberation-related distance cue. However, past studies that explored the importance of this ratio as an auditory distance cue present contrasting results on the temporal limit that should be considered between the early energy and late energy [6,7,30–32].

In our case, the differences in terms of late energy between the measurements and the extrapolated responses by the envelope-based model are not significant for any onset time considered superior to 10 ms (see Figure 7). Early energy differences tend to reach their minimum for a considered offset time of 50 ms. When considering an early-to-late ratio for a limit of 50 ms, consistent with the definition of speech clarity C50, the difference between measurements and generated impulse responses is around 1 dB. For speech signals,

the just-noticeable difference of C50 is around 1.1 dB, and a significant change value is considered to be around 3 dB [33]. This value is superior to the minimal early energy difference observed between the measured impulse responses and those generated by the envelope-based model.

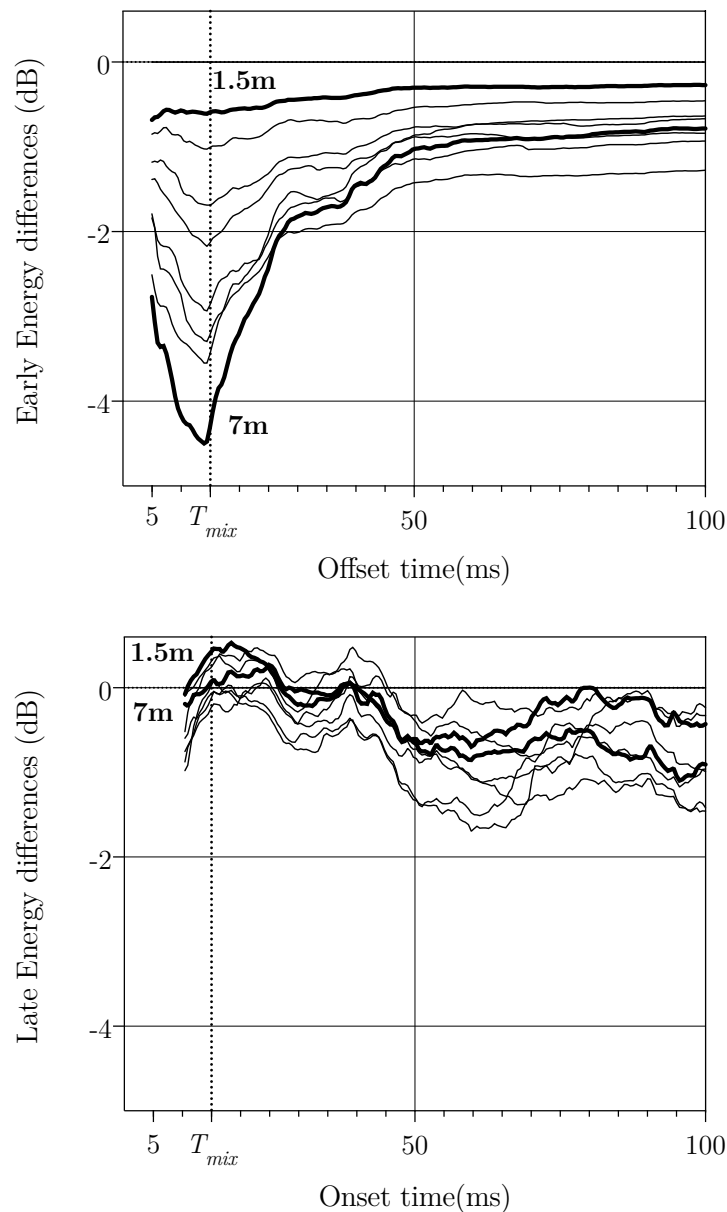


Figure 7. Early and late energy differences of the BRIRs generated with the envelope-based model compared to measured BRIRs. $T_{mix} = 15$ ms corresponds to the transition time between the early and the late reverberation segments considered in the envelope-based model to generate the BRIRs.

Thus, this similarity, in terms of early energy, could explain the perceptual similarity observed in the current experiment between the envelope-based model and the rendering method based on measurements. As the experiment results are only related to a single acoustical environment, it is not sufficient to draw a general conclusion about what appropriate limit should be considered as an early-to-late energy ratio in order to compute a distance cue. It does, however, corroborate the assumption that early reflections have little to no effect on distance perception on their own, but are perceived as part of the direct sound [7].

The design of the envelope-based model entails that numerous aspects of the generated impulse responses were inherited from the initial measurement. In this experiment, a measurement taken at a distance of 1 m was employed to extrapolate further distances. Several reverberation-related cues that were acknowledged as influential on auditory distance perception were not reproduced correctly with distance. Among these, the initial time delay gap (ITDG) [11], which denotes the arrival time of the first reflections, remains constant across all generated impulse responses. However, no significant perceptual differences were identified between the reference stimuli and those generated by the envelope-based model. Because Equation (3) was applied to the entire early part of the initial impulse response prior to T_{mix} , generated BRIRs underestimate the energy contained in the first reflections (see Figure 7). The lack of energy in the first reflections may have weakened the ITDG reliability as a reverberation-related distance cue. As a result, participants' distance judgments were driven by other cues, such as the early-to-late energy ratio. To corroborate this assumption, further studies investigating the weighting strategies of reverberation-related cues to infer a distance percept are required.

5.2. Individual Acoustic Cues Weighting Strategies

The perceived sound intensity is regarded as a primary cue for determining the distance to a sound source. It, along with the reverberation-related cue, are the two primary cues involved in auditory distance perception of frontal sources. Earlier research has produced contradictory findings regarding the relative role of intensity or a reverberation-related cue for auditory distance perception [34–36]. Kopčo and Shinn-Cunningham [5] and Zahorik [4,12] concluded that listeners weight differently the acoustic cues to produce auditory distance judgments, and that the weighting of cues might be influenced by the listening situation. The different acoustic weighting cues employed by participants are analyzed under the scope of the individual compression coefficients (see Figure 4). These coefficients quantify the compression effect for each model and each participant.

Individual participant's compression coefficients for the envelope-based model and the rendering method based on measurements are homogeneous. Each participant presents a similar compression effect for these two rendering methods. In contrast, the difference between the compression coefficients associated with the rendering method based on measurements and the intensity-based model suggests that different strategies were used to judge the distance of sound sources. Participants for which the value of the compression coefficients estimated on the intensity model is comparable to the value obtained with the rendering method based on measurements mainly base their auditory distance judgments on intensity. Some participants obtain a compression coefficient a that is nearly equal to 0 for the intensity-based model only, then do not use intensity as a distance cue, and rely primarily on a reverberation-related distance cue.

This result corroborates the findings of Zahorik [12]. It shows that the acoustic cues weighting strategies is mainly an individual characteristic. In the following section, the influence of environmental context on the weighting of acoustic cues is examined, to determine if it can be linked to some characteristics of the listening situation or if it is only an idiosyncratic characteristic.

5.3. Room Volume and Acoustic Cues Weighting Strategies

This study investigated the influence of volume on auditory distance perception. Results in Section 4.5 demonstrated a slight influence on the reported distances of stimuli generated by the intensity-based model. This specific result illustrates the influence of the reported volume on the compression effect of participants on the intensity-based model.

In the CW group, the maximal value of the nonlinear compression coefficient a tends to a smaller value when volume increases. This can be regarded as a reduction in the number of strategies observed for producing auditory distance judgments when the volume increases. In large volumes ($>100 \text{ m}^3$), participants seem to only rely on a reverberation-related cue to produce an auditory distance judgment. Under the hypothesis that a larger

volume generally leads to the perception of a more reverberant environment [37–40], this effect can be explained by the expectation induced by these environments. Participants who expect a large variation in a reverberation-related cue may be more likely to rely on it.

This effect is not observable on the compression coefficients associated with participants of the FW group who benefited from a longer visual spatial boundary. It is possible that the statistical effect observed on the a coefficient values is due solely to the seven participants with a room volume greater than to 100 m³. These participants potentially adopted the same strategies for reporting distances, that imply mainly relying on a reverberant-related cue, without being influenced by their perceived acoustic environment.

5.4. Visual Spatial Boundary and the Calibration of the Auditory Space

The influence of visual divergence was studied by investigating differences between the CW and FW groups. Results in Section 4.4 demonstrate that the FW group, whose participants were exposed to a longer spatial boundary, presents a smaller compression effect. This result can be compared to past studies concluding that the presence of congruent visual cues could enhance the accuracy of auditory distance perception [13,29,41].

In this experiment, a room divergence effect originates from the discrepancy between the reproduced room and the experimental environment specific to each participant. The geometrical characteristics of the reproduced room are, however, closer to the dimensions reported on average by the participants of the FW group than those reported in the CW group. The volume of the reproduced room is 144 m³, while the mean volume is 81.7 m³ in the FW group and 37.5 m³ in the CW group. The presented stimuli intend to reproduce auditory distances up to 7 m, while the visual spatial boundary is limited to 3 m in the CW group and ranges from 5 to 10 m for the FW group. Therefore, the presence of more congruent visual cues might have enhanced the auditory distance perception accuracy of the FW group.

It has been long argued that the visual context in which auditory judgments occur contributes to the organization of the auditory space [13,15,16]. This finding about the influence of the visual spatial boundary could be seen as an extension of the hypothesis discussed by Calcagno et al. [13]. In their experiment, participants had to evaluate the distance to a real loudspeaker, presented frontally and emitting white noise bursts. As a result, no room divergence effect was present. Participants had access to an increasing quantity of visual cues, ranging from being blindfolded to having full vision of the room. Intermediate scenarios in which only two or four LEDs were lit during the reports were also examined. The authors demonstrated in this acoustically congruent situation that the presence of congruent visual cues increases the accuracy of auditory distance perception. They hypothesized that visual range information helped calibrate the reported distances of real sound sources, visual distance perception being a sensory modality far more accurate for distance estimations [29]. In the current study, the visual spatial boundary condition could also have acted as a calibration of auditory distance perception in an acoustically divergent situation. A long visual spatial boundary condition might have influenced the representation of the auditory space and incited participants to expand the range of the reported distances. This influence on the reporting strategy explains the significantly lower compression effect of the FW group.

5.5. Study Limitations

Practical difficulties linked to online-based experiments may reduce the significance of the observed effects and limit their comparison with similar lab-based studies. The inherent lack of control on the participants' experimental conditions, such as the use of diverse headphones models and different environmental noise levels, may have increased the variability of the reported distances.

In this study, the influence of divergent environmental characteristics on auditory distance perception was examined. Participants were asked to assess the perceived distance of auditory stimuli, but no reporting about externalization was proposed. Externaliza-

tion and auditory distance perception are two related concepts; however, they primarily depend on distinct cues [18]. The presence of binaural cues in the sound signal is mandatory for externalization, which can be considered a prerequisite for an accurate distance perception for some listeners [10]. In this respect, the perception of externalization is regarded as dichotomous, whereas distance perception is continuous and mainly driven by monaural cues [3,18]. Therefore, precautions were taken to favor the externalization, such as preserving binaural cues in the auditory stimuli. Generic HRTFs were applied to generate the binaural stimuli, which is known to increase the risk of in-head localization, especially for frontal sources [42]. However, aside from practical constraints linked to the online-based experiment, this decision was motivated by previous findings showing that nonindividualized HRTFs had no impact on the auditory distance perception of frontal sound sources [36].

6. Conclusions

In this study, we presented and discussed the results of an online experiment investigating how acoustic cues are interpreted to produce an auditory distance judgment and how environmental cues influenced the auditory distance perception of a virtual sound source. Through statistical analysis of mean distance reports and individual quantification of compression effect, the effect of the self-reported visual spatial boundaries and the room volume was investigated.

First, we showed that the DRR is likely to be overtaken by an early-to-late energy ratio in defining a meaningful approach to interpret reverberation as a distance cue. This, however, requires additional testing in a variety of environments. The participants employed a broad range of acoustic cue weighting strategies, suggesting that they rely mainly on idiosyncratic choices.

Second, visual spatial boundaries influence auditory distance perception. The closer the visual boundary is, the more compressed the overall distance judgments are. The effect of this boundary on auditory distance perception is thought to be owing to a partial calibration of the auditory space by vision of the room's apparent boundaries.

Finally, volume of the room has also an effect on the acoustic cues weighting strategies. The larger the room is, the more participants rely on reverberation to judge the auditory distance. This could be explained by the greater expectation of reverberation driven by large rooms. However, the limited number of participants that led to this result casts doubt on the significance of the reported effect. Further lab-based studies with controlled conditions are necessary to confirm this finding.

Author Contributions: Conceptualization, V.M., O.W. and I.V.-D.; methodology, V.M.; software, V.M.; validation, O.W., I.V.-D. and V.M.; formal analysis, V.M.; investigation, V.M.; resources, O.W.; data curation, V.M.; writing—original draft preparation, V.M.; writing—review and editing, V.M., O.W. and I.V.-D.; visualization, V.M.; supervision, I.V.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study and methods followed the tenets of the declaration of Helsinki. Approval for the experiment design has been given by the Research Ethics Committee of Sorbonne Université (N° 2020-CER-2020-80). The online data acquisition followed the general data protection regulation.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author, V.M. (vmartin@ircam.fr).

Acknowledgments: We would like to thank all the subjects for their voluntary participation in the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*; MIT Press: Cambridge, MA, USA, 1997.
2. Werner, S.; Klein, F.; Mayenfels, T.; Brandenburg, K. A summary on acoustic room divergence and its effect on externalization of auditory events. In Proceedings of the IEEE 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016; pp. 1–6.
3. Kolarik, A.J.; Moore, B.C.; Zahorik, P.; Cirstea, S.; Pardhan, S. Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. *Atten. Percept. Psychophys.* **2016**, *78*, 373–395. [[CrossRef](#)] [[PubMed](#)]
4. Zahorik, P.; Brungart, D.S.; Bronkhorst, A.W. Auditory distance perception in humans: A summary of past and present research. *ACTA Acust. United Acust.* **2005**, *91*, 409–420.
5. Kopčo, N.; Shinn-Cunningham, B.G. Effect of stimulus spectrum on distance perception for nearby sources. *J. Acoust. Soc. Am.* **2011**, *130*, 1530–1541. [[CrossRef](#)]
6. Bronkhorst, A.W.; Houtgast, T. Auditory distance perception in rooms. *Nature* **1999**, *397*, 517–520. [[CrossRef](#)]
7. Bronkhorst, A.W. *Modeling Auditory Distance Perception in Rooms*; Forum Acusticum: Sevilla, Spain, 2002.
8. Bidart, A.; Lavandier, M. Room-induced cues for the perception of virtual auditory distance with stimuli equalized in level. *Acta Acust. United Acust.* **2016**, *102*, 159–169. [[CrossRef](#)]
9. Larsen, E.; Iyer, N.; Lansing, C.R.; Feng, A.S. On the minimum audible difference in direct-to-reverberant energy ratio. *J. Acoust. Soc. Am.* **2008**, *124*, 450–461. [[CrossRef](#)]
10. Prud'Homme, L.; Lavandier, M. Do we need two ears to perceive the distance of a virtual frontal sound source? *J. Acoust. Soc. Am.* **2020**, *148*, 1614–1623. [[CrossRef](#)]
11. Werner, S.; Füg, S. Controlled Auditory Distance Perception using Binaural Headphone Reproduction–Evaluation via Listening Tests. In Proceedings of the 27th Tonmeisterstagung, VDT International Convention, Cologne, Germany, 22–25 November 2012.
12. Zahorik, P. Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am.* **2002**, *111*, 1832–1846. [[CrossRef](#)] [[PubMed](#)]
13. Calcagno, E.R.; Abregu, E.L.; Eguía, M.C.; Vergara, R. The role of vision in auditory distance perception. *Perception* **2012**, *41*, 175–192. [[CrossRef](#)]
14. Valzolgher, C.; Alzhaler, M.; Gessa, E.; Todeschini, M.; Nieto, P.; Verdelet, G.; Salemm, R.; Gaveau, V.; Marx, M.; Truy, E.; et al. The impact of a visual spatial frame on real sound-source localization in virtual reality. *Curr. Res. Behav. Sci.* **2020**, *1*, 100003. [[CrossRef](#)]
15. Warren, D.H. Intermodality interactions in spatial localization. *Cogn. Psychol.* **1970**, *1*, 114–133. [[CrossRef](#)]
16. Warren, D.H.; Welch, R.B.; McCarthy, T.J. The role of visual-auditory “compellingness” in the ventriloquism effect: Implications for transitivity among the spatial senses. *Percept. Psychophys.* **1981**, *30*, 557–564. [[CrossRef](#)]
17. Etchemendy, P.E.; Abregú, E.; Calcagno, E.R.; Eguía, M.C.; Vechiatti, N.; Iasi, F.; Vergara, R.O. Auditory environmental context affects visual distance perception. *Sci. Rep.* **2017**, *7*, 1–10. [[CrossRef](#)]
18. Best, V.; Baumgartner, R.; Lavandier, M.; Majdak, P.; Kopčo, N. Sound externalization: A review of recent research. *Trends Hear.* **2020**, *24*, 2331216520948390. [[CrossRef](#)]
19. Carpentier, T.; Noisternig, M.; Warusfel, O. Twenty years of Ircam Spat: Looking back, looking forward. In Proceedings of the 41st International Computer Music Conference (ICMC), Denton, TX, USA, 25 September–1 October 2015; pp. 270–277.
20. Lochner, J.; Burger, J. The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech. *Acta Acust. United Acust.* **1958**, *8*, 1–10.
21. Polack, J.D. Modifying chambers to play billiards: The foundations of reverberation theory. *Acta Acust. United Acust.* **1992**, *76*, 256–272.
22. Abel, J.S.; Huang, P. A simple, robust measure of reverberation echo density. In *Audio Engineering Society Convention 121*; Audio Engineering Society: New York, NY, USA, 2006.
23. Stewart, R.; Sandler, M. Statistical measures of early reflections of room impulse responses. In Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 10–15 September 2007; pp. 59–62.
24. Massé, P.; Carpentier, T.; Warusfel, O.; Noisternig, M. Denoising directional room impulse responses with spatially anisotropic late reverberation tails. *Appl. Sci.* **2020**, *10*, 1033. [[CrossRef](#)]
25. Jot, J.M.; Cerveau, L.; Warusfel, O. Analysis and synthesis of room reverberation based on a statistical time-frequency model. In *Audio Engineering Society Convention 103*; Audio Engineering Society: New York, NY, USA, 1997.
26. Barron, A.; Rissanen, J.; Yu, B. The minimum description length principle in coding and modeling. *IEEE Trans. Inf. Theory* **1998**, *44*, 2743–2760. [[CrossRef](#)]
27. Jesteadt, W.; Wier, C.C.; Green, D.M. Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.* **1977**, *61*, 169–177. [[CrossRef](#)]
28. Peirce, J.; Gray, J.R.; Simpson, S.; MacAskill, M.; Höchenberger, R.; Sogo, H.; Kastman, E.; Lindeløv, J.K. PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **2019**, *51*, 195–203. [[CrossRef](#)]
29. Anderson, P.W.; Zahorik, P. Auditory/visual distance estimation: Accuracy and variability. *Front. Psychol.* **2014**, *5*, 1097. [[CrossRef](#)]
30. Valente, D.L.; Braasch, J. Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues. *J. Acoust. Soc. Am.* **2010**, *128*, 1952–1964. [[CrossRef](#)]

31. Moon, H.G.; Noh, J.U.; Sung, K.M.; Jang, D.Y. Reverberation cue as a control parameter of distance in virtual audio environment. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2004**, *87*, 1822–1826.
32. Messonnier, J.C.; Moraud, A. Auditory distance perception: Criteria and listening room. In *Audio Engineering Society Convention 130*; Audio Engineering Society: New York, NY, USA, 2011.
33. Bradley, J.S.; Reich, R.; Norcross, S. A just noticeable difference in C50 for speech. *Appl. Acoust.* **1999**, *58*, 99–108. [[CrossRef](#)]
34. Ashmead, D.H.; Leroy, D.; Odom, R.D. Perception of the relative distances of nearby sound sources. *Percept. Psychophys.* **1990**, *47*, 326–331. [[CrossRef](#)]
35. Brungart, D.S.; Rabinowitz, W.M. Auditory localization of nearby sources. Head-related transfer functions. *J. Acoust. Soc. Am.* **1999**, *106*, 1465–1479. [[CrossRef](#)] [[PubMed](#)]
36. Zahorik, P. Auditory display of sound source distance. In Proceedings of the International Conference on Auditory Display, Kyoto, Japan, 2–5 July 2002; pp. 326–332.
37. Mershon, D.H.; Ballenger, W.L.; Little, A.D.; McMurtry, P.L.; Buchanan, J.L. Effects of room reflectance and background noise on perceived auditory distance. *Perception* **1989**, *18*, 403–416. [[CrossRef](#)] [[PubMed](#)]
38. Sandvad, J. Auditory perception of reverberant surroundings. *J. Acoust. Soc. Am.* **1999**, *105*, 1193. [[CrossRef](#)]
39. Kolarik, A.J.; Pardhan, S.; Cirstea, S.; Moore, B.C. Using acoustic information to perceive room size: Effects of blindness, room reverberation time, and stimulus. *Perception* **2013**, *42*, 985–990. [[CrossRef](#)]
40. Cabrera, D.; Jeong, C.; Kwak, H.J.; Kim, J.Y. Auditory room size perception for modeled and measured rooms. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Dearborn, MI, USA, 12–14 July 2005; Volume 2005, pp. 2995–3004.
41. Kearney, G.; Gorzel, M.; Rice, H.; Boland, F. Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Acta Acust. United Acust.* **2012**, *98*, 61–71. [[CrossRef](#)]
42. Udesen, J.; Piechowiak, T.; Gran, F. The effect of vision on psychoacoustic testing with headphone-based virtual sound. *J. Audio Eng. Soc.* **2015**, *63*, 552–561. [[CrossRef](#)]