

## Article

# Sound Water Masking to Match a Waterfront Soundscape with the Users' Expectations: The Case Study of the Seafront in Naples, Italy

Virginia Puyana-Romero <sup>1,\*</sup>, Luigi Maffei <sup>2</sup>, Giovanni Brambilla <sup>3</sup> and Daniel Nuñez-Solano <sup>1</sup>

<sup>1</sup> Grupo de Investigación Entornos Acústicos, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito EC170125, Ecuador; dnunezsolano@gmail.com

<sup>2</sup> Dipartimento di Architettura e Disegno Industriale, Università degli Studi della Campania Luigi Vanvitelli, 81031 Aversa, Italy; luigi.maffei@unicampania.it

<sup>3</sup> Department of Acoustics and Sensors "O.M. Corbino", Institute of Marine Engineering (INM), National Research Council of Italy (CNR), I-00133 Rome, Italy; giovanni.brambilla@artov.inm.cnr.it

\* Correspondence: virginia.puyana@udla.edu.ec

**Abstract:** In the last decades, the soundscape approach has attracted the attention of architects and urban planners, leading them to incorporate the acoustic features into the enjoyment of their creations. One of the key aspects for an appreciated urban environment is to match the expectations of the users. In this study, the matching of the waterfront soundscape with the users' expectations is evaluated by laboratory tests using semantic differential scales applied to reproduced virtual scenarios obtained adding different water sound pressure levels (SPLs) to the original in-situ setting. The tests were carried out by an immersive virtual reality (IVR) device, using 360° videos and spatial audio recorded in two sites of the waterfront in Naples, Italy. The scenarios were presented to the participants according to three experimental protocols, namely audio-only (A), video-only (V), and simultaneous audio-video (AV) reproduction. The examined different acoustic scenarios were the original one recorded in situ and others obtained adding seawater sounds at SPL increments of 5 dB. The results show that all the scenarios with water sounds added are rated more pleasant than the original one for the audio-only scenario. When video and audio are displayed simultaneously, two scenarios are more pleasant than the original one, likely because there is a need for coherence between the water sound SPL heard and the visible noise sources. Sounds coherent with the type of shore show a higher matching with expectations and pleasantness appraisals, rather than those that are uncoherent with the layout scenario.

**Keywords:** soundscape; expectations; water masking sounds; waterfront



**Citation:** Puyana-Romero, V.; Maffei, L.; Brambilla, G.; Nuñez-Solano, D. Sound Water Masking to Match a Waterfront Soundscape with The Users' Expectations: The Case Study of the Seafront in Naples, Italy. *Sustainability* **2021**, *13*, 371. <https://doi.org/10.3390/su13010371>

Received: 24 November 2020

Accepted: 28 December 2020

Published: 3 January 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

Understanding the complex phenomena involved in environmental noise perception is becoming a priority in strategies to fight against outdoor noise, especially because noise exposure represents a significant risk to physical and mental health; among other effects, it can cause sleep disturbances, cardiovascular problems, and cognitive impairment [1–9]. Environmental perception can be defined as a multi-dimensional phenomenon catching objects and events by the senses, understanding and identifying them in the environment and, eventually, creating a reaction [10,11]. Thus, gathering sensorial information is a three-stage process: Sensation (input), organization of the perceived elements and their identification/recognition (interpretation), and the response (outcome) [12]. The context of the acoustic environment, including non-auditory factors (e.g., the location features, individual differences, socio-cultural aspects, etc.) [13–16], can also determine how the auditory sensation is interpreted. The dynamism of the places can also affect the soundscape perception, for instance preferring the acoustic environments that best match the

type of activities occurring in the place [17–19]. Dealing with people activity, Jo and Jeon have recently found that sounds made by people decreased the perceived tranquility of the soundscape, but made it more dynamic [16]. Other contextual factors, such as the visual stimuli and the landscape features, can strongly determine the perception of an acoustic environment [20–24]. For instance, Menzel et al. [25] found differences in the loudness perception of a car pass-by depending on its color. Liu et al. studied two types of landscape factors, visual and functional, evidencing that two of them are highly correlated with the overall soundscape preference [26]. Visual aspects may influence the perception of the acoustic environment [27], and vice-versa [28–31]. In that regard, Anderson et al. evidenced that the perception of an outdoor setting was conditioned by the sounds heard there [29], and Gan et al., in a study performed in natural and urban areas, found out that the acoustic preference played a much more important role in landscape evaluation than visual preferences [28]. The urban morphology (street patterns, low or high building density, etc.) is another aspect to take into account when studying the influence of the context on the acoustic perception [32–34]. A study by Liu et al. evidenced that the composition and structure of the landscape characterized through spatial patterns could influence the soundscape perception of urban areas [35]. In previous studies carried out by the research group of this paper in the waterfront of Naples, landscape factors related to the shape of the waterfront were found to be related with the acoustic perception of the environment [36,37].

The individual characteristics of respondents should also be considered when evaluating the factors influencing the soundscape perception. In a top-down processing, also called the conceptually-driven processing, the assimilation of sensory information is influenced by expectation, stored knowledge, interests, beliefs, and context [38,39]. The expectations and previous experiences that subjects have on a specific environment can influence its perception [40] through the codification of a complex cognitive structure. Several studies have evaluated the influence of the expectations on certain aspects of the perception of an acoustic environment [40–43]. Brambilla and Maffei suggested that the annoyance caused by a sound source is influenced by the expectations to hear it and its congruence within the context [41]. The expectations from context were included as one of the components in a model for the human auditory perception developed by Botteldooren and De Coensel [44]. Bruce and Davies developed a model in which the expectations of an acoustic environment depend on the expectation of the different elements composing it [40]. Bruce et al. analyzed through sense-walkings the influence of the expectations on singular factors, such as the soundscape and smellscape, on the environmental perception [45]. Some research indicates that certain types of expectations, such as those generated by negative information of certain noise sources, can increase negatively the perceived noise nuisance. This was observed in the study carried out by Crichton et al. on the perception of wind farm noise [46], in which participants receiving positive information were less annoyed by noise than those receiving negative information. They also found a relation between noise sensitivity and predicted noise annoyance in those receiving negative information about wind farms.

### *1.1. Water Masking Sounds*

There is clear evidence in the literature that in open spaces people prefer to hear natural sounds over artificial man-made sounds [47,48]. Sound masking techniques based on natural sounds are sometimes used to make the users' stay more pleasant in public spaces where technological sounds, such as road traffic noise, are predominant. One of the most used masking sounds is that of water, generally coming from water fountains in their different forms (interactive floor water fountains, wall fountains, water curtain walls, water sculptures, water games for playgrounds, etc.). The large use of this sound source is due in part to experimental studies showing the benefits of water sounds: Verena-Thoma et al. evidenced that water sounds reduce stress on individuals with somatic complaints, not observed during listening to music or in silence [49]; Jeon et al. stated that water sounds are the preferred type of natural sound masking, compared to bird twittering, wind, or

church bells [50]; likewise, Hao and Nilsson et al. suggested that the water sound reduces the loudness perception of road traffic noise [34,51], and You et al. found that the preferred water sound level for masking is 3 dB below the background noise level [52]. In that regard, the feeling of improved tranquility occurs even with low sound pressure levels (SPLs) of water sounds [53].

### *1.2. Immersive Virtual Reality as a Tool for the Appraisal of Soundscape*

The validity and reliability of immersive virtual reality (IVR) applied to the soundscape appraisal compared to on-site surveys have been proven successfully in the literature: This technique combines spatial audio recording and playback (e.g., Ambisonic coding) with visual scenarios generated entirely by computer [18,54], like 360° stereoscopic videos [55]. The results of applying these techniques evidenced that, comparing on-site surveys and IVR test results, there was no statistically significant difference in the responses given on the soundscape quality. In the last years, different studies can be found in the literature using 360° videos and spatial audio [55]; the flexibility and affordability of these audio-visual systems encouraged their application for different research purposes, such as soundscape classification [56], evaluation of the pleasantness of natural sounds in residential areas [57], or studying the effect of social interaction on the perception of the soundscape [16].

A recent study has explored the perception of loudness and annoyance of different sound sources, evaluated with the images of the anechoic chamber, and with panoramic videos of the original sound measurement locations [57]. The outcomes of the study reveal that the appraisals of the loudness and annoyance of the sound sources were significantly lower when panoramic videos of the original sound measurement locations were shown, suggesting that visual information of the real environment may influence the results with respect to those obtained for audio-only scenarios.

### *1.3. Motivation and Objectives of the Study*

The sound masking studies listed above were carried out by adding water sources visually and acoustically to the scenarios, or only adding water sound to a reproduced acoustic environment. However, masking studies on urban environments including the presence of water (seafronts, riverfronts, etc.), not heard because it is masked by the road traffic noise or people activities, have not been undertaken so far, and also the evaluation of how water sounds at different SPLs match the people expectations in the soundscape has not been analyzed yet.

Unfortunately, the important aspect of the users' expectations on typologies of outdoor spaces are not taken into account in urban design. Thus, it was deemed necessary to investigate new techniques to face this need, and also to get tools to improve the acoustic design of spaces. The present study deals with the performance of water sound masking in seafront scenarios and assessing the degree of adequacy of the seafront soundscape with users' expectations. Two areas of the seafront in Naples have been considered, and water sounds at different SPLs have been added to the original sound recordings, also taking into account the appraisals on expectations, artificiality of the water sounds added and their relevance with the context, and the sound sources that could be heard.

The objective of the study was to test the following hypotheses, named from H1 to H5:

**Hypothesis 1 (H1).** *The SPL of the seawater sounds added to the original audio-visual environment does not affect the degree of satisfaction regarding the expectations on the acoustic environment (masking water sound SPLs and expectations).*

**Hypothesis 2 (H2).** *Two different audio-visual scenarios of the seafront with the same seawater SPL added will satisfy the participants' expectations at the same degree (locations and expectations).*

**Hypothesis 3 (H3).** *Semantic differential scale (SDS) ratings on the soundscape are the same for two different sound masking conditions (SDS and water sound SPLs).*

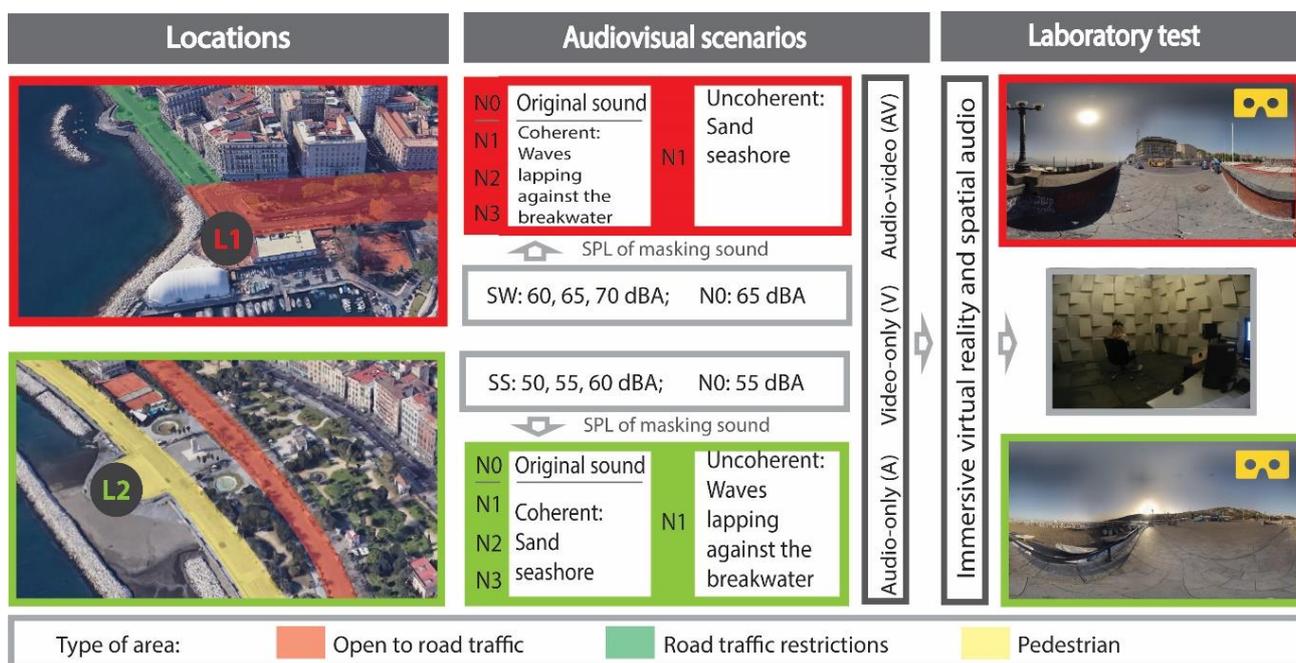
**Hypothesis 4 (H4).** Considering equal acoustic environment conditions, SDS ratings are the same for audio-visual scenarios (AV) rather than for audio-only (A) scenarios (SDS and experimental stimuli).

**Hypothesis 5 (H5).** The addition of water sounds not coherent with the layout of the promenade does not influence the degree of satisfaction of the user's expectations of the acoustic environment (coherence and expectations).

## 2. Materials and Methods

In Naples, despite the proximity of the sea, the characteristics of the places are such that the sea waves' sound could not be heard almost anywhere along the seafront. The absence of this sound is due to the sea normally being calm and, in most areas, there is a certain distance (vertical or horizontal) between the seashore or the breakwater and the seafront promenade. This setting was suggested to evaluate which water sound levels would match people's expectations to improve their ratings of the acoustic environment.

Two sites were selected as scenarios of the IVR surveys, based on the type of sounds that could be heard, and the topological relationship between the promenade and the water. Site 1 (Acton street) is a street dominated by traffic noise, and natural sound sources are seldom perceived, except for the wind. There is a considerable vertical distance between sea level at the breakwater and the promenade. Site 2 (beach) is a pedestrian zone next to an artificial seashore; a busy street is nearby, and the background traffic noise is not so high (Figure 1). During the sound recordings, the sea was very calm, and the sound of the sea waves was hardly heard.



**Figure 1.** Scheme of the locations (L1—Acton street, and L2—beach) and audiovisual scenarios (original audio recorded in situ, N0, and three sound pressure levels (SPL) of masking water sounds at 5dB steps, N1, N2, and N3, for coherent and uncoherent sounds) used in the laboratory tests.

The visual and acoustic environment were recorded simultaneously using Go-Pro Hero4 cameras and an Ambisonic microphone (Soundfield SPS 200) at 1.65 m height, according to the recommendations of the ISO 12913-2 [58]. The duration of the recordings was 120 s. A class 1 sound level meter (SOLO 01DB) was used to measure the SPLs and the spectrum of the acoustic environment. The Surround Zone VST Plugin was used to convert the registered A-format to B-format required for the reproduction in laboratory [59]. Images taken from six camera positions, arranged to capture the visual field of a sphere,

were edited and merged to create the 360° videos. The images of the microphone, the sound level meter and the tripods that held the devices were deleted by means of video editing. Audio and visual scenarios were embedded using the Vizard 5.2 software.

The listening tests were carried out in the anechoic chamber of the University of Campania. Five speakers (DynaudioBM5AMKII) and one subwoofer (DynaudioBM9S) arranged symmetrically in a 5.1 configuration were located in a 3 m diameter circle. An external sound card Motu 828 MKII drove the loudspeakers. The reproduction chain was calibrated in order that the equivalent sound levels measured at the participant's position were equal to those obtained in the field measurements, with a tolerance of 1.5 dBA. The experiment was split into a pilot and a main phase.

### 2.1. Pilot Study

This study, formed by two listening sessions, was undertaken to select the appropriate SPL steps of the seawater sounds to be added to the in situ recordings. The study was carried out in the anechoic chamber of the University of Campania. Each session was attended by a group of 10 students, different between the two sessions. In both sessions, the assessment of the stimuli was collected by four semantic differential scales (SDS), namely unpleasant–pleasant, chaotic–calm, eventful–uneventful, boring–exciting [60], shown on the display of the Oculus headset used for the stimuli reproduction. Hereinafter, each scale is named SDS followed by the left-hand side attribute of the scale itself. For the first session, 8 different audio-visual scenarios were played back, formed by the 360° video taken at site L1 and 8 different acoustic environments, namely the original audio recording in situ (N0 with LAeq set to 65 dBA) and the others obtained adding to N0 seawater SPLs, varying from 56 to 74 dBA at 3 dB steps (−9, −6, −3, 0, +3, +6, or +9 dB referred to the LAeq of N0). The criterion used by Jeon et al. and Galbrun et al. in [61,62] was applied for the selection of the SPL steps. During this session, audio-visual (AV) scenarios were randomly shown to participants, following a balanced Latin square design. Once the participants had rated all the scenarios, they were asked to evaluate again the audio-visual (AV) scenario to which they had given the highest rating on the SDS-Unpleasant, as a control question. The second session differed in the SPL increments: A 5 dB step was chosen, like in a test by Jeon et al., in which semantical differential scales were used [61]. Six acoustic settings were considered, formed by N0 again and the others obtained, adding to N0 seawater SPLs, varying from 55 to 75 dBA at 5 dB steps. A summary of the experimental design is reported in Table 1.

**Table 1.** Experimental design of the pilot study: Session number, group of participants (A or B, with 10 participants each), location (L1—Acton street), audio-visual stimuli (AV) and control question. Acoustic environments used at session 1: The original audio recording N065 (with  $L_{Aeq} = 65$  dBA) and seven acoustic stimuli obtained adding to N065 seawater SPLs, varying from 56 to 74 dBA at 3 dB steps (N0 + SW56, N0 + SW59, N0 + SW62, N0 + SW65, N0 + SW68, N0 + SW71, N0 + SW74). Acoustic environments at session 2: N065, and five acoustic stimuli obtained adding to N065 seawater SPLs, varying from 55 to 75 at 5 dB steps (N0 + SW55, N0 + SW60, N0 + SW65, N0 + SW70, N0 + SW75).

Session	Participants' Groups	Site	Stimuli (AV)
1	A (10 subjects)	L1	N065, N0 + SW56, N0 + SW59, N0 + SW62, N0 + SW65, N0 + SW68, N0 + SW71, N0 + SW74 Control question: Choice of the AV scenario with highest rating on semantic differential scale (SDS)-Unpleasant. If two scenarios have the same rating, the one with the highest SW level was chosen. For the chosen acoustic condition, the four SDS were repeated.
2	B (10 subjects)	L1	N065, N0 + SW55, N0 + SW60, N0 + SW65, N0 + SW70, N0 + SW75 Control question: Same selection criteria applied in the session 1.

## 2.2. Main Study

The main study was split into two sessions separated by 15 days and involved 30 subjects (14 women, 16 men), aged from 20 to 47. Each session lasted approximately 30 min. Participants were tested one by one in the anechoic chamber and, before starting the test, they were asked to read the instructions to fill the questionnaire in. General questions (age, gender, education level, and number of visits to the waterfront of Naples) were asked to the participants before starting the first IVR session. Afterwards, they were asked to put on the head mounted display, in order to explore the IVR scenarios. Regarding the stimuli, three experimental settings were arranged and presented to the participants in the following order: Audio-only (A), video-only (V), and audio-video (AV). For each experimental setting, the scenarios were ordered and shown to the participants based on a randomly generated Latin square. The questions were shown on the head-mounted display, and the participants had to answer them out loud so that they were aware of the noise level. A 11-item questionnaire was designed to evaluate the acoustic and visual environment, expectations of the participants regarding the perceived sound sources, SPL of water sounds (SW), qualitative features of the soundscape (through SDS), and the artificiality of the water sounds added (Table 2). Each item was rated on a 7-point Likert scale, except for the questions on the appraisal of the acoustic and visual environment, which were rated in a 4-point Likert scale.

**Table 2.** Questions asked to participants using a head mounted display.

Questions	Audio-Visual Scenarios	Measurement Scale
1. Which of the following sounds would you expect to hear in the Naples waterfront? Options: Buses, Cars, Motorcycles, Clackson, Brakes, Bikes, Background road traffic, Boats, Construction Works, Voices/footsteps, Birds, Other animals, Wind, Sea, Leaves	No video, no audio	-
2. How do you consider this acoustic environment in relation to the following attributes? <ul style="list-style-type: none"> <li>• Unpleasant–Pleasant</li> <li>• Chaotic–Calm</li> <li>• Uneventful–Eventful</li> <li>• Boring–Exciting</li> </ul>	Audio-only (A): N0, N1, N2, and N3	7-point Likert scale
3. Which of the following sounds do you expect to hear in this scenario? Options: Buses, Cars, Motorcycles, Clackson, Brakes, Bikes, Background road traffic, Boats, Construction Works, Voices/footsteps, Birds, Other animals, Wind, Sea, Leaves	Video-only (V)	-
4. How do you consider this acoustic environment in relation to the following attributes? <ul style="list-style-type: none"> <li>• Unpleasant–Pleasant</li> <li>• Chaotic–Calm</li> <li>• Uneventful–Eventful</li> <li>• Boring–Exciting</li> </ul>	Audio-video (AV): N0, N1, N2, and N3 (coherent)	7-point Likert scale
5. How do you consider this acoustic environment in relation to the following attributes? <ul style="list-style-type: none"> <li>• Unpleasant–Pleasant</li> <li>• Chaotic–Calm</li> <li>• Uneventful–Eventful</li> <li>• Boring–Exciting</li> </ul>	Audio-video (AV): N1 (uncoherent)	7-point Likert scale

Table 2. Cont.

	Questions	Audio-Visual Scenarios	Measurement Scale
6.	To what extent does this acoustic environment meet your expectations of the seafront?	Audio-video (AV): N0, N1, N2, and N3 (coherent)	7-point Likert scale
7.	To what extent does the sound of water seem artificial to you?	Audio-video (AV): N0, N1, N2, and N3 (coherent)	7-point Likert scale
8.	Considering the visual and acoustic aspects, how much do you like this scenario?	Audio-video (AV): N0 and N1 (coherent)	4-point Likert scale
9.	How do you judge the quality of the following aspects of this scenario? <ul style="list-style-type: none"> <li>• Acoustic environment</li> <li>• Visual environment</li> <li>• Overall environmental quality</li> </ul>	Audio-video (AV): N0 and N1 (coherent)	7-point Likert scale
10.	Which sounds can you hear in this scenario? Option: Buses, Cars, Motorcycles, Clackson, Brakes, Bikes, Background road traffic, Boats, Construction Works, Voices/footsteps, Birds, Other animals, Wind, Sea, Leaves	Audio-video (AV): N0 and N1 (coherent)	7-point Likert scale
11.	How does the sound of “(selected sound source)” contribute to the quality of the acoustic environment of this scenario?	Audio-video (AV): N0 and N1 (coherent)	7-point Likert scale

To shorten the duration of the test, and considering the results of the pilot study (reported in Section 2.1), it was decided to evaluate the semantic differential scales, expectations, and artificiality of the sound considering four acoustic environments (N0 + three acoustic scenarios with masking water sounds), and the questions about noise sources, which required more time to be answered, considering only two scenarios. The  $L_{Aeq}$  of N0 was set at 65 and 55 dBA for L1 and L2 sites, respectively. The  $L_{Aeq}$  of seawater sound (SW) added N1, N2, and N3 were 60, 65, and 70 dBA at L1, and 50, 55, and 60 dBA at L2, respectively. The experimental design is summarized in Table 3.

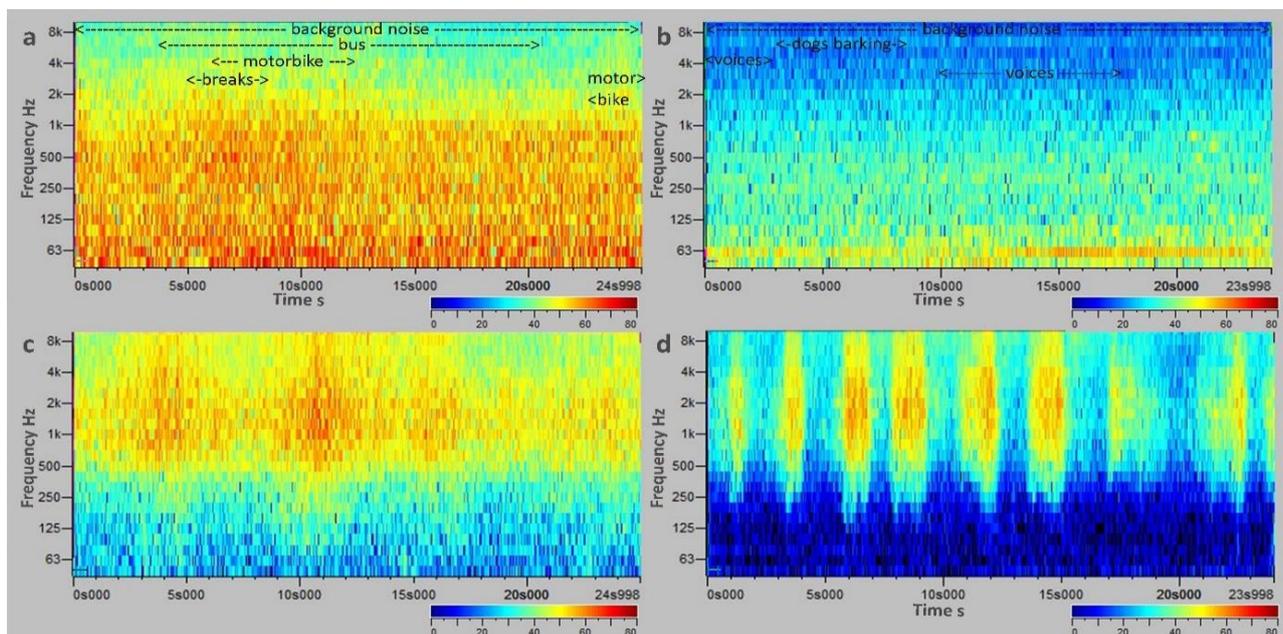
**Table 3.** Experimental design of the main study: Session number, group of participants (C, with 30 participants), location (L1—Acton street and L2—beach), type of sensorial stimuli (Audio-only, A, video-only, V, and audio-video, AV) and type of questions (semantic differential scale SDS, artificiality, expectations, soundscape quality SQ, landscape quality LQ, and overall environmental quality EQ). Acoustic environment for location L1: The original audio recording N065, and three acoustic stimuli obtained adding to N065 seawater SPLs varying from 60 to 70 at 5 dB steps (N1, N2, and N3). Acoustic environment for location L2: The original audio recording N055, and three acoustic stimuli obtained adding to N055 seawater SPLs varying from 50 to 60 at 5 dB steps (N1, N2, and N3).

Session	Participants' Groups	Site	Stimuli			
			A	V	AV	AV
			Coherent Added Sound	-	Coherent Added Sound	Uncoherent Added Sound
1	C(30)	L1	N065; N1 = N0 + SW60; N2 = N0 + SW65; N3 = N0 + SW70 Type of questions: SDS	VL1	N065; N1 = N0 + SW60; N2 = N0 + SW65; N3 = N0 + SW70 Type of questions: SDS, artificiality, expectations	-
		L2	N055; N1 = N0 + SS50; N2 = N0 + SS55 N3 = N0 + SS60 Type of questions: SDS	VL2	N055; N1 = N0 + SS50; N2 = N0 + SS55 N3 = N0 + SS60 Type of questions: SDS, artificiality, expectations	-

Table 3. Cont.

Session	Participants' Groups	Site	Stimuli			
			A	V	AV	AV
			Coherent Added Sound	-	Coherent Added Sound	Uncoherent Added Sound
15 days of delay						
2	C(30)	L1	-	-	N065; N1 = N0 + SW60 SDS questions only on N1, sound sources heard, SQ, LQ, overall EQ	N1 = N0 + SS60 SDS questions
		L2	-	-	N055; N1 = N0 + SS50 SDS questions only on N1, sound sources heard, SQ, LQ, overall EQ	N1 = N0 + SW50 SDS questions

Figure 2 shows the spectrograms (in dB) at L1 and L2 of the original audio recording N0, and of the seawater sounds (SW) added at N2 SPL. The levels were adjusted in amplitude, resulting in similar spectral characteristics to the original sample, at different SPL [61]. Typical high noise values of traffic noise can be observed at 63 Hz in the original audio recordings, even in L2 where the traffic noise was distant. Coherent and uncoherent seawater sound sources were also evaluated through SDS. In location L1, waves sound lapping against the breakwater (SW) was added as coherent sound, and seashore sound (SS) as uncoherent; contrariwise in L2.



**Figure 2.** Spectrograms of the audios used in the study; (a) original audio N0 registered at L1, (b) original audio N0 registered at L2, (c) coherent seawater sound (SW) added at L1, and (d) coherent seashore sound (SS) added at L2.

According to the assumptions of the study, the data obtained from the surveys were analyzed to assess whether there are significant differences between pairs of variables, or to obtain the degree of association between them. For this purpose, the Wilcoxon signed rank test and Spearman's correlations were performed, respectively. Cohen's classification has been used to evaluate the effect size.

### 3. Results

#### 3.1. Pilot Study

Comparing the responses given to the repeated questions (Table 4), the consistency of the results was more satisfactory for the second session, where the masking sound changed by 5 dB steps. In particular, the ratings on SDS-Unpleasant were the same for all subjects, excepting two of them showing a difference of  $\pm 1$  point; for SDS-Chaotic and SDS-Boring, the maximum differences were also  $\pm 1$  point, and in SDS-Eventful the maximum difference was  $\pm 3$ , but occurring for only one participant. For that reason, it was decided to use 5 dB steps for the acoustic conditions in the main study.

**Table 4.** The difference of ratings given to the repeated condition between session 1 (3 dB steps) and 2 (5 dB steps).

	Repeated Acoustic Condition	Semantic Differential Scale	Difference of Ratings Given to the Repeated Condition				
			0	$\pm 1$	$\pm 2$	$\pm 3$	$\pm 4$
Number of subjects 3 dB steps (session 1)	Highest rating given to Unpleasant–Pleasant	Unpleasant–Pleasant	5	3	1	1	-
		Chaotic–Calm	3	3	1	3	-
		Eventful–Uneventful	1	3	2	3	1
		Boring–Exciting	2	2	4	2	-
Number of subjects 5 dB steps (session 2)	Highest rating given to Unpleasant–Pleasant	Unpleasant–Pleasant	8	2	-	-	-
		Chaotic–Calm	6	4	-	-	-
		Eventful–Uneventful	3	4	2	1	-
		Boring–Exciting	2	3	4	1	-

For session 2, higher appraisals were observed for the 60 dBA audio-visual scenario; furthermore, most participants reported equal or similar responses to the scenarios N0 + SW65 and N0 + SW60. In particular, 31 responses over 40 (10 participants  $\times$  4 SDS) were equal for both acoustic conditions, or with a difference of  $\pm 1$ . Similarly, this also happened with the ratings given to the SDS for the scenarios N0 + SW70 and N0 + SW75 (28 responses over 40).

Considering the results of the pilot study, and the planned long duration for the main study, it was decided to not include the acoustic conditions N0 + SW55 and N0 + SW75.

#### 3.2. Main Study

The different analyses undertaken to test the hypotheses of the study, ordered from H1 to H5, are described below together with the results.

##### 3.2.1. Hypothesis 1 (H1): Masking Water Sound SPLs and Expectations

This hypothesis deals with evaluating whether adding seawater sounds to real audio-visual environments has an effect on matching the expectations placed on the soundscape. The Wilcoxon signed rank test can be used to compare two repeated measurements on a single sample, assessing whether their population mean ranks differ, and to understand whether the ranked values of one group are consistently higher or lower than the other. Actually, the test was used to compare whether the mean ranks of the expectations on two different acoustic environments—each one with a different SPL of water sound masking the real acoustic environment, but the same visual scenario—differ. The Wilcoxon signed rank test was conducted on all the possible pair-combinations of acoustic scenarios (six), at locations L1 and L2.

As reported in Table 5, at location L2, four of the six paired combinations show significant differences at 95% confidence level. Comparing the medians, 84% of participants—that gave different ratings to each compared scenario—consider that the sea sound level N1 better meets their expectations than the real acoustic environment (N0) at position L2. The same trend is observed for N2 level versus N0. Seventy-seven percent of participants also consider that N3 matches their expectations worse than N1. N2–N3 pairs show the same

trend. Sound N3 satisfies the expectations worse than N1 or N2, and it does not show a significant difference with N0. It should be also noted that, also considering not-significant differences, the masking conditions N1, N2, and N3 satisfy more widely the expectations of participants than the original acoustic environment N0.

**Table 5.** Wilcoxon signed rank test (*p*-Value and comparison of ratings between paired-variables), and effect size results for all possible pair combinations between N0, N1, N2, and N3 at locations L1 and L2. For the analysis of the participants' ratings, "a" is the first, and "b" the second variable of the pair.

Location	a	b	Wilcoxon <i>p</i> -Value	Lower (a < b)	Equal (a = b)	Higher (a > b)	Effect Size $\frac{z}{\sqrt{n}}$
L1 (Acton street)	N0	N1	0.052	8	20	2	0.354
		N2	0.192	15	5	10	0.238
		N3	0.171	15	7	8	0.250
	N1	N2	0.823	11	8	11	0.041
		N3	0.704	13	7	10	0.069
	N2	N3	0.806	7	15	8	0.045
L2 (Beach)	N0	N1	0.002	16	11	3	0.555
		N2	0.020	17	7	6	0.424
		N3	0.662	12	9	9	0.080
	N1	N2	0.721	9	8	13	0.065
		N3	0.031	5	8	17	0.395
	N2	N3	0.001	0	19	11	0.586

The effect size is normally used to report the magnitude of differences between two groups—a large effect size means that the difference is important [63]—and it is a useful tool when interpreting the effectiveness of the results [64]. The effect size was calculated using Rosenthal expression for parametric tests [65], by dividing the absolute standardized test statistic *z* by the square root of the number of subjects (last column in Table 5). According to Cohen's classification, the effect sizes at L2 are considered moderate (range 0.3–0.5) for the paired groups N0–N2 (0.424) and N1–N3 (0.395), and large (>0.5) for N0–N1 (0.555) and N2–N3 (0.586).

At location L1, although N1, N2, and N3 match the expectations of more participants better than the original acoustic environment N0, the difference between groups is not significant at 95% confidence level.

### 3.2.2. Hypothesis 2 (H2): Locations and Expectations

This hypothesis deals with evaluating whether there are statistically significant differences in matching the participant's expectations on the soundscape of two locations when adding the same SPL of seawater sound. Again, the Wilcoxon signed rank test was used to calculate significant differences between the repeated measurements of the user's expectations at locations L1 and L2 using three masking conditions (N1, N2, and N3) of the original setting N0. Table 6 shows the results for each acoustic environment and considering all environments together as one group. Participants' expectations are different at both locations for N1 and N2 with a significance level below 0.5, and for N0 with a significance level below 0.1. In particular, the acoustic environment satisfies better participants' expectations at L2 than at L1 at all levels (e.g., for the N2 level, 18 participants gave higher ratings on the expectations at L2 rather than L1, and only 3 gave higher ratings to L1).

**Table 6.** Wilcoxon signed rank test ( $p$ -Value and comparison of ratings between paired-variables), and effect size results between the expectations at L1 and L2, for each masking condition (N1, N2, and N3) and original setting N0. For the analysis of the participants' ratings, "a" is the first, and "b" the second variable of the pair.

Acoustic Condition	a	b	Wilcoxon $p$ -Value	Lower (a < b)	Equal (a = b)	Higher (a > b)	Effect Size $\frac{Z}{\sqrt{n}}$
N0	Exp_L1	Exp_L2	0.052	14	11	5	0.355
N1	Exp_L1	Exp_L2	0.001	18	9	3	0.609
N2	Exp_L1	Exp_L2	0.009	17	9	4	0.474
N3	Exp_L1	Exp_L2	0.543	10	12	8	0.111
All levels	Exp_L1	Exp_L2	0.000	59	41	20	0.398

Considering only the statistically significant differences, for all the paired variables on the expectations the effect sizes are large ( $e_{e\_N1} = 0.609$ ), moderate ( $e_{e\_N0} = 0.355$ ,  $e_{e\_N2} = 0.474$ , and  $e_{e\_all} = 0.398$ ), and low ( $e_{e\_N3} = 0.111$ ).

### 3.2.3. Hypothesis 3 (H3): SDS and Water Sounds SPLs

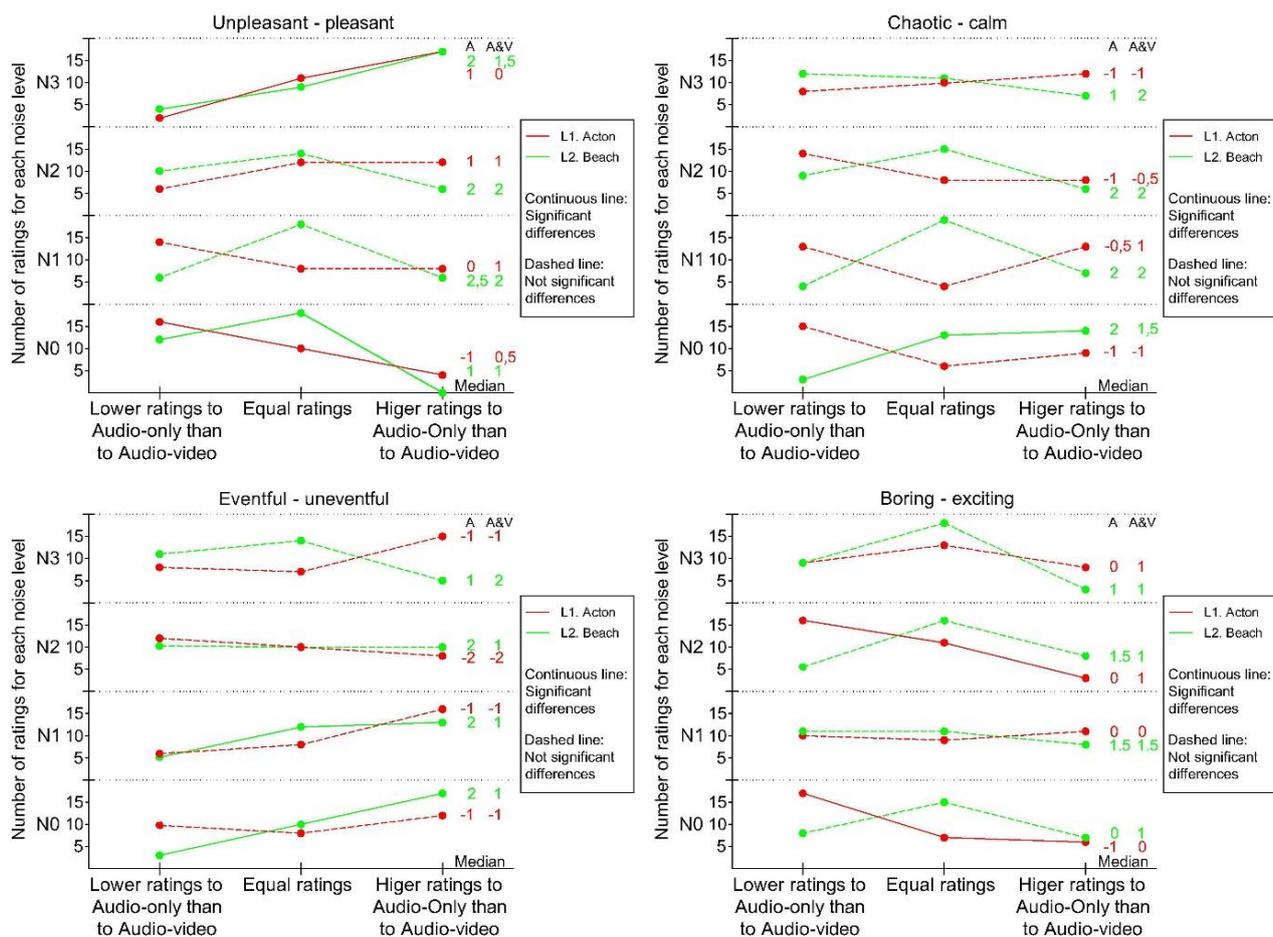
This hypothesis was formulated to understand whether there are differences in the SDS ratings for audio-only (A), and for audio-video (AV) scenarios. Figure 3 shows a comparison of the ratings given on the four SDS, that is unpleasant–pleasant (top left), chaotic–calm (top-right), eventful–uneventful (bottom-left), and boring–exciting (bottom-right) with and without video, at locations L1 (red line) and L2 (green line). The horizontal axis was divided into lower, equal, and higher ratings to A than to AV experimental settings. The vertical axis gathers the results according to the four acoustic stimuli. Significant (continuous line) and not significant differences (dashed line) between paired variables using Wilcoxon signed rank test are also shown.

For the SDS-Unpleasant, at both locations there are significant differences (at 95% confidence level) between the ratings given to the scenarios with and without video for N0 and N3 sounds; furthermore, (excluding equal ratings) for paired variables at N0, a higher number of participants gave lower ratings to audio-only (A) than to audio-video (AV) at both locations. This trend changes with increasing the masking noise level until stimulus N3, for which the opposite behavior is observed.

For the SDS-Chaotic and the N0 acoustic environment, there are only statistically significant differences at L2 ( $p\text{-value}_{L2\_N0} = 0.019$ ). For these paired variables, a higher number of participants perceived the acoustic environment as more calm with audio-only than with audio-video.

SDS-Eventful results also show significant differences only at L2, at stimuli N0 and N1 ( $p\text{-value}_{L2\_N0} = 0.001$ ;  $p\text{-value}_{L2\_N1} = 0.002$ ); in particular, more participants gave higher ratings to audio-only than to audio-video. At L2, this trend is reversed for high masking SPL.

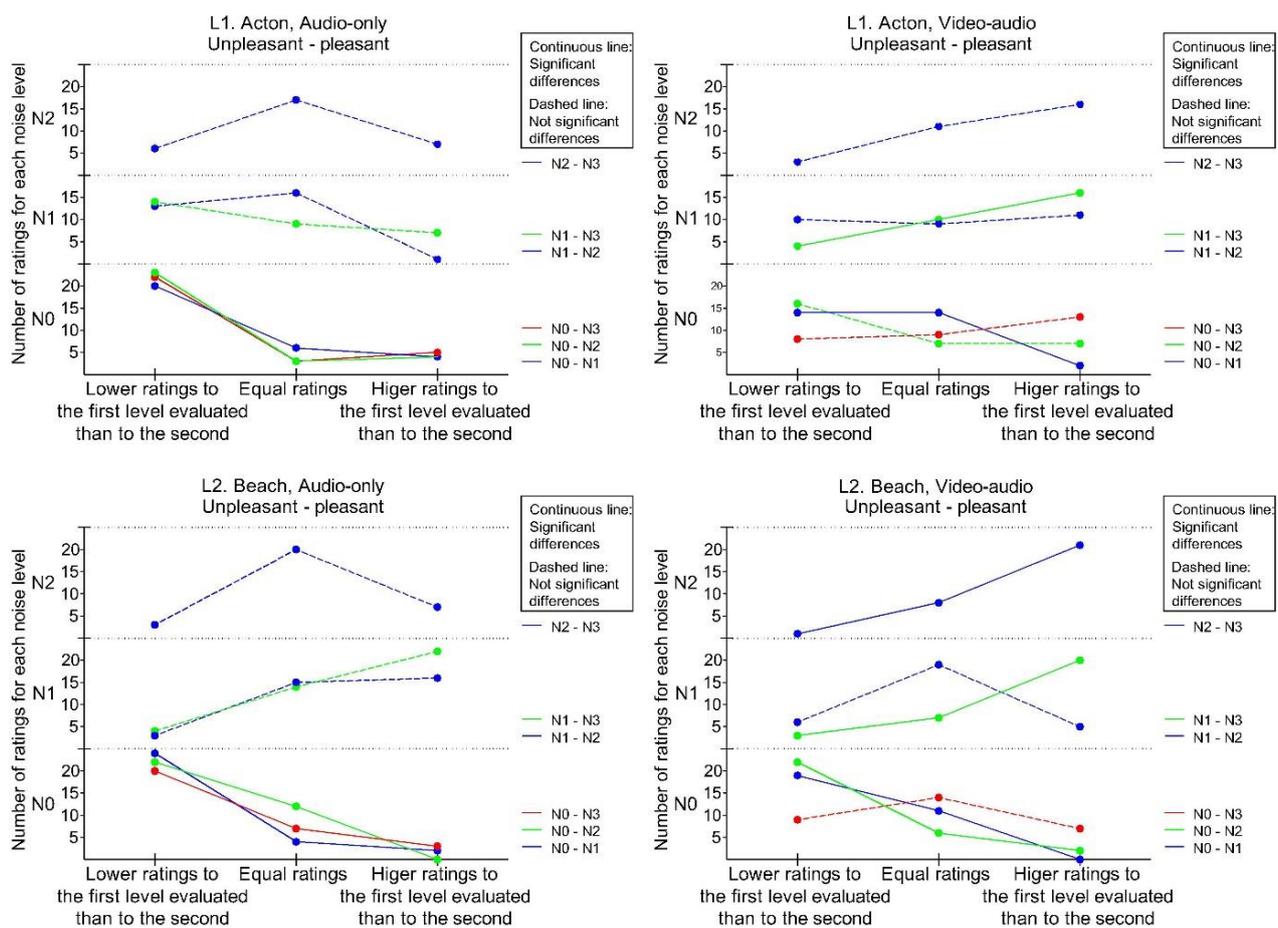
SDS-Boring paired analysis at L1 location shows that most people considered the acoustic environment more exciting when video images are added to the scenarios (in particular, 17 participants felt the soundscape was more exciting with audio-video scenarios than with audio-only, and only six respondents felt the opposite).



**Figure 3.** Wilcoxon signed rank test results for audio-only (A), and audio-video (AV) scenarios of the appraisals on the semantic differential scales (Unpleasant–pleasant, top left; chaotic–calm top right; eventful–uneventful, bottom left; and boring–exciting, bottom right). Results are shown for the acoustic conditions N0, N1, N2, and N3, at locations L1 (red line) and L2 (green line).

### 3.2.4. Hypothesis 4 (H4): SDS and Experimental Stimuli

This hypothesis was formulated to assess whether there are differences in the SDS ratings for paired acoustic conditions. Figure 4 shows a comparison of the ratings for the SDS-Unpleasant without (left) and with video images (right), evaluated at locations L1 (top) and L2 (bottom). The horizontal axis was divided into higher, equal, and lower ratings to the first variable than to the second one. The vertical axis gathers the results for N0, N1, and N2. For both locations and audio-only conditions, results show that there are statistically significant differences between the ratings given to the paired variables N0–N1 ( $p\text{-value}_{L1} = 0.001$ ;  $p\text{-value}_{L2} = 0.000$ ), N0–N2 ( $p\text{-value}_{L1} = p\text{-value}_{L2} = 0.000$ ), and N0–N3 ( $p\text{-value}_{L1} = 0.001$ ;  $p\text{-value}_{L2} = 0.000$ ); for those paired groups, participants gave lower ratings to N0 than to the rest of the sound masking levels. However, the results are not as conclusive for audio-video scenarios, and only significant differences are found between N0–N1 pair at L1 ( $p\text{-value}_{L1} = 0.010$ ), and between the N0–N1 ( $p\text{-value}_{L2} = 0.000$ ) and N0–N2 ( $p\text{-value}_{L2} = 0.000$ ) pair at L2.



**Figure 4.** Wilcoxon signed rank test results for each of the possible paired combinations between N0, N1, N2, and N3, for the semantic differential scale Unpleasant–Pleasant. The test was conducted at L1 (top) and L2 (bottom), for scenarios with audio-only (left) and with audio-video (right).

Results show statistically significant differences in the pleasantness of N1 and N3 when images are displayed simultaneously to the audio at locations L1 and L2 ( $p\text{-value}_{L1} = 0.004$ ;  $p\text{-value}_{L2} = 0.000$ ). In both cases, most people gave higher ratings to N1 than to N3. Although the scenarios with masking sound are more pleasant than the original acoustic environment, N1 is preferred over N3, and N2 over N3. Thus, it can be concluded that lower sound masking SPLs are more pleasant than the higher ones, and that N1 was rated the most pleasant acoustic environment.

### 3.2.5. Hypothesis 5 (H5): Coherence and Expectations

Wilcoxon signed rank test was performed between coherent and not coherent masking sounds (for the acoustic condition N1) added to the original acoustic environment, for each of the four semantic scales (Table 7). Results show that the soundscape is considered significantly more pleasant when sounds are coherent with the visual scenario than when they are uncoherent, for both locations ( $p\text{-value}_{L1} = 0.009$ ;  $p\text{-value}_{L2} = 0.013$ ). They are also rated more calm (less chaotic) when sounds are coherent than when they are uncoherent ( $p\text{-value}_{L1} = 0.033$ ;  $p\text{-value}_{L2} = 0.003$ ). Furthermore, at L2, when coherent sounds of water are added to the original acoustic environment, the soundscape is considered more uneventful ( $p\text{-value}_{L2} = 0.001$ ) and exciting ( $p\text{-value}_{L2} = 0.001$ ) than when the sounds added are uncoherent with the environment. It can be assumed that the differences between groups are more remarkable at L2 than at L1 (with medium and high size effect for all the SDS at L2, and only medium size effect for SDS-Unpleasant and SDS-Chaotic at L1). The correlation

coefficients between the artificiality of the sounds and the SDS are very low and considered not significant.

**Table 7.** Wilcoxon signed rank test ( $p$ -value and comparison of ratings between paired-variables), and effect size results between coherent and not coherent sounds at L1 and L2. For the analysis of the participants' ratings, "a" is the first, and "b" the second variable of the pair.

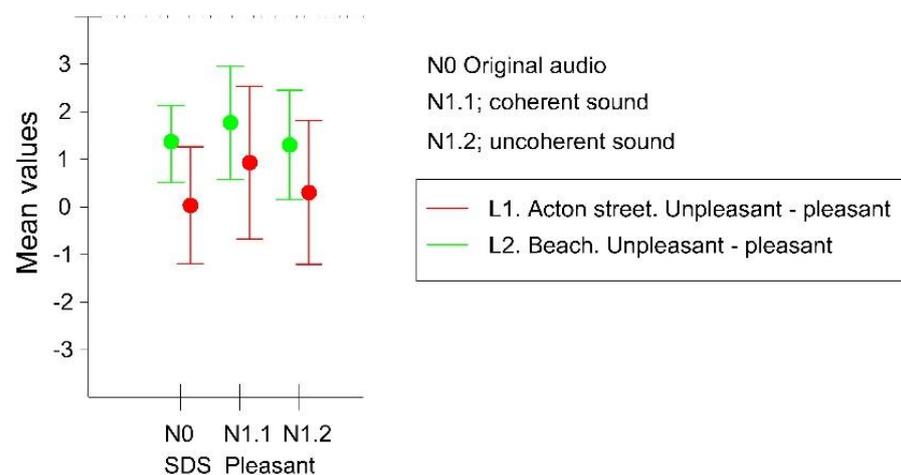
SDS	a	b	Wilcoxon $p$ -Value	Lower (a < b)	Equal (a = b)	Higher (a > b)	Effect Size $\frac{Z}{\sqrt{n}}$
L1_N1_SDS-Unpleasant	Coherent	Uncoherent	0.009	10	4	16	0.477
L1_N1_SDS-Chaotic	Coherent	Uncoherent	0.033	8	6	16	0.389
L1_N1_SDS-Eventful	Coherent	Uncoherent	0.982	12	10	8	0.004
L1_N1_SDS-Boring	Coherent	Uncoherent	0.721	17	7	6	0.065
L2_N1_SDS-Unpleasant	Coherent	Uncoherent	0.013	7	6	17	0.453
L2_N1_SDS-Chaotic	Coherent	Uncoherent	0.003	12	3	15	0.535
L2_N1_SDS-Eventful	Coherent	Uncoherent	0.001	13	2	15	0.599
L2_N1_SDS-Boring	Coherent	Uncoherent	0.001	10	2	17	0.600

The Spearman's correlation coefficient between pleasantness and expectations when added sounds are coherent with the visual environment is higher (high association:  $r_{L1} = 0.710$ ,  $p\text{-value}_{L1} = 0.000$ ;  $r_{L2} = 0.551$ ,  $p\text{-value}_{L2} = 0.002$ ) than when they are not (medium association:  $r_{L1} = 0.387$ ,  $p\text{-value}_{L1} = 0.035$ ;  $r_{L2} = 0.470$ ,  $p\text{-value}_{L2} = 0.009$ ) at both locations (Table 8). For the other SDS variables analyzed, significant correlation coefficients were obtained only for location L2.

**Table 8.** Spearman correlation coefficients between the expectations and the semantical differential scales (SDS-Unpleasant, SDS-Chaotic, SDS-Eventful, and SDS-Boring) evaluated with coherent and uncoherent water sound masking at N1 level. \* The correlation is significant at the 0.05 level. \*\* The correlation is significant at the 0.01 level.

Groups of Variables	Expectations N1		Groups of Variables	Expectations N1	
	Corr. Coef.	$p$ -Value		Corr. Coef.	$p$ -Value
L1. N1.Coherent_SDS-Unpleasant	0.710 **	0.000	L1. N1.Uncoherent_SDS-Unpleasant	0.387 *	0.035
L1. N1.Coherent_SDS-Chaotic	0.266	0.155	L1. N1.Uncoherent_SDS-Chaotic	0.230	0.221
L1. N1.Coherent_SDS-Eventful	0.191	0.312	L1. N1.Uncoherent_SDS-Eventful	0.118	0.534
L1. N1.Coherent_SDS-Boring	0.069	0.719	L1. N1.Uncoherent_SDS-Boring	0.179	0.343
L2. N1.Coherent_SDS-Unpleasant	0.551 **	0.002	L2. N1.Uncoherent_SDS-Unpleasant	0.470 **	0.009
L2. N1.Coherent_SDS-Chaotic	0.488 **	0.006	L2. N1.Uncoherent_SDS-Chaotic	0.645 **	0.000
L2. N1.Coherent_SDS-Eventful	0.298	0.110	L2. N1.Uncoherent_SDS-Eventful	0.499 **	0.005
L2. N1.Coherent_SDS-Boring	0.519 **	0.003	L2. N1.Uncoherent_SDS-Boring	0.324	0.086

Figure 5 shows the mean and standard deviation of the appraisals given on the SDS-Unpleasant for N0 and for the audio-video scenarios in which coherent and not coherent water sounds were added, at locations L1 and L2. The soundscape is more pleasant when coherent water sounds are added ( $\text{mean}_{L1} = 0.93$ ,  $\text{sd}_{L1} = 1.60$ ;  $\text{mean}_{L2} = 1.77$ ,  $\text{sd}_{L2} = 1.19$ ) than when only the real audio-video environment is displayed at both locations ( $\text{mean}_{L1} = 0.03$ ,  $\text{sd}_{L1} = 1.23$ ;  $\text{mean}_{L2} = 1.37$ ,  $\text{sd}_{L2} = 0.77$ ). This pleasantness decreases when uncoherent water sounds are added ( $\text{mean}_{L1} = 0.30$ ,  $\text{sd}_{L1} = 1.51$ ;  $\text{mean}_{L2} = 1.30$ ,  $\text{sd}_{L2} = 1.15$ ).



**Figure 5.** Mean and standard deviation of the appraisals given to the SDS Unpleasant–Pleasant, for the original (N0), coherent (N1.1), and uncoherent (N1.2) acoustic conditions, at L1 and L2.

#### 4. Discussion

The level of satisfaction with the expectations on the soundscape has been evaluated in different acoustic and visual scenarios using sound masking. The acoustic conditions of the scenarios vary both by the coherence of the sound sources added, and by the SPL of the masking water sound.

##### 4.1. Hypothesis 1 (H1)

All the masking water sound SPLs satisfy the expectations of a larger number of participants better than that observed for the original sound of the area, although significance in the differences occurred only between the pairs of variables N0–N1 and N1–N2 in L2 at 0.05%, and for N0–N1 in L1 at 0.1%. Thus, the hypothesis H1 can be rejected. The lack of statistically significant differences at L1 (for any of the six possible combinations of audio-video scenarios) may be because traffic noise levels are high and, therefore, difficult to mask. Consequently, adding seawater sound does not involve a change in meeting participants' expectations. From the comparison between the acoustic environments N1, N2, and N3, it can be inferred that the lowest masking level matches the expectations of the users better than the higher ones in the two locations, possibly because when the SPL is high, there is no coherence between the noise level and the sound sources shown in the video.

##### 4.2. Hypothesis 2 (H2)

Comparing the two locations at the same audio condition, location L2 (beach) matches the expectations of more participants than L1 (Acton street) for all the noise masking conditions, probably because the background noise at L1 has less artificial sounds, and noise levels are lower and easier to mask. Consequently, the findings do not meet the baseline assumption H2, as the acoustic environments of one location satisfy better participant's expectations than the ones of the other.

##### 4.3. Hypothesis 3 (H3)

Comparing the SDS dealing with the audio-only (A) and audio-video (AV) scenarios, for the SDS-Unpleasant a greater number of participants gave higher ratings to AV than to A for scenarios without masking at both locations, and with low masking SPLs at L1. The sounds of seawater include broadband frequency components that make the sound similar to white or pink noise [57]. When no images are displayed (A scenarios), the participants may not recognize which noise source generates the sound and may have identified the first masking condition N1 as traffic noise, and therefore, have rated the acoustic environment as unpleasant.

This trend is reversed for high levels of masking noise—being the acoustic environment more pleasant with A than with AV. This can be because for A setting, participants are only paying attention to what they are hearing: As listening to the sounds of water is a pleasant experience [50,52,53,61], even with high SPL of water, more participants report the acoustic environment as pleasant. For the original acoustic environment (N0), the addition of images seems to have a significantly positive effect on the appraisals of the participants. However, when sound levels of water are very high, there is no coherence between what people see and hear, and probably for this reason participants gave lower ratings to the acoustic environment at AV than to A scenarios. A similar effect regarding high and low levels of masking noise also occurs for the SDS-Chaotic and SDS-Uneventful, but only at L2. The results for the SDS-Chaotic at L2 may be because the video allows seeing in the distance a street with a traffic activity; people can, therefore, associate these images with the background traffic noise that is heard which, given its nature, may not have been identified in the A scenario, or may have been mistaken with electrical noise from the speakers. When considering SDS-Uneventful, similarly to what happened to chaotic–calm appraisals, visual images may allow identifying elements that generate sound events, and therefore to rate the acoustic environment as more eventful at AV than A scenarios.

Consequently, hypothesis H3 is not met, as SDS ratings on the soundscape are not the same for at least two different sound masking conditions.

#### 4.4. Hypothesis 4 (H4)

When evaluating the differences between the appraisals given to pairs of acoustic environment conditions, most participants gave higher ratings to N3, N2, and N1 than to N0 in the A scenarios; consequently, the results do not confirm hypothesis H4. Thus, all masking SPLs led to a more pleasant acoustic environment than the real one. This is because when A is reproduced, the participant's attention focuses on the acoustic environment only. When video images are simultaneously displayed with audio, the AV scenarios contain much more information that distracts the attention of participants. Regarding the AV scenarios, the appraisals are also higher for N2 and N1 than for N0, although when comparing N3 and N0, most participants give equal or worse ratings to N3; this implies that the video stimulus is capable of altering the opinion about the acoustic environment. A similar conclusion was reached by Asakura et al. when evaluating loudness and annoyance perception with images of the anechoic chamber and images of the original scenarios [57].

#### 4.5. Hypothesis 5 (H5)

The correlation coefficient between the SDS-Unpleasant and expectations is higher when the sounds heard are coherent, and therefore, the original H5 hypothesis is not met. Furthermore, a greater number of participants give higher ratings when the sounds are coherent than when they are not coherent. This happens even if both sounds correspond to seashore sounds, one of water lapping softly against the breakwater, and the other from a sandy shore. The differences between the appraisals are more remarkable at L2 than at L1, probably because the noise levels of the original scenarios are lower, and composed of more natural sounds in L2, and that allows a greater symbiosis between real and added sounds.

## 5. Conclusions

The present study evaluates the relationship of the expectations with qualitative scales of the soundscape appraisal and with different masking noise levels. In the laboratory test, 360° videos and spatial audio recorded in the waterfront of Naples were shown to the participants using the Oculus headset. The study was conducted using audio-only (A), video-only (V), and audio-video (AV) scenarios. The results show that, for A scenarios, all the added masking water SPLs make the acoustic environment more pleasant. For AV scenarios, only the first two levels of masking sounds are more pleasant. Similarly, when using two types of seawater sounds, the ones consistent with the morphology of the shore match the expectations better than non-coherent ones. The results of this study show

that both the evaluation of the degree of satisfaction of expectations can be very useful tool when designing urban soundscapes. The role of expectations and coherence with the environment has been shown to have a great influence on the soundscape perception. Thus, it would be important to consider these aspects as key factors in the analysis of soundscapes and how they are perceived, in order to improve the quality of recreational areas in cities.

**Author Contributions:** V.P.-R., L.M., G.B., and D.N.-S. conceived, wrote, and revised the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work was partially funded by the People Programme (Marie Curie Actions) of the European Union's 7th Framework Programme FP7/2007–2013 under REA grant agreement no. 290110, SONORUS "Urban Sound Planner". The article processing charge for this publication was funded by the "V:alere 2019 program" of the University of Campania Luigi Vanvitelli (Italy).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and of the Ethics Committee of the Università degli Studi della Campania Luigi Vanvitelli.

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

**Data Availability Statement:** The data presented in this study are available on justified request from the corresponding author. The data are not publicly available due to privacy reasons.

**Acknowledgments:** We would like to thank Jose Aguilar Silva for his collaboration in the data analysis, and the participants for their selfless contribution to the study.

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

## References

- Ouis, D. Annoyance from road traffic noise: A review. *J. Environ. Psychol.* **2001**, *21*, 101–120. [[CrossRef](#)]
- World Health Organization. *European Union Environmental Noise Guidelines for the European Region*; WHO Regional Office for Europe: Copenhagen, Denmark, 2018.
- Van Gerven, P.W.M.; Vos, H.; Van Boxtel, M.P.J.; Janssen, S.A.; Miedema, H.M.E. Annoyance from environmental noise across the lifespan. *J. Acoust. Soc. Am.* **2009**, *126*, 187–194. [[CrossRef](#)] [[PubMed](#)]
- Fuks, K.; Moebus, S.; Hertel, S.; Viehmann, A.; Nonnemacher, M.; Dragano, N.; Möhlenkamp, S.; Jakobs, H.; Kessler, C.; Erbel, R.; et al. Long-term urban particulate air pollution, traffic noise, and arterial blood pressure. *Environ. Health Perspect.* **2011**, *119*, 1706–1711. [[CrossRef](#)] [[PubMed](#)]
- Halperin, D. Environmental noise and sleep disturbances: A threat to health? *Sleep Sci.* **2014**, *7*, 209–212. [[CrossRef](#)] [[PubMed](#)]
- Babisch, W. The noise/stress concept, risk assessment and research needs. *Noise Health* **2002**, *4*, 1–11.
- Miedema, H.M.E.; Passchier-Vermeer, W.; Vos, H. *Elements for a Position Paper on Night-Time Transportation Noise and Sleep Disturbance*; Inro, T., Ed.; TNO Inro: Delft, The Netherlands, 2003.
- Gan, W.Q.; Davies, H.W.; Koehoorn, M.; Brauer, M. Association of long-term exposure to community noise and traffic-related air pollution with coronary heart disease mortality. *Am. J. Epidemiol.* **2012**, *175*, 898–906. [[CrossRef](#)]
- Selander, J.; Nilsson, M.E.; Bluhm, G.; Rosenlund, M.; Lindqvist, M.; Nise, G.; Pershagen, G. Long-term exposure to road traffic noise and myocardial infarction. *Epidemiology* **2009**, *20*, 272–279. [[CrossRef](#)]
- Zimbardo, P.G.; Gerrig, R.J. *Perception. Foundation of Cognitive Psychology: Core Readings*; Daniel, J.L., Ed.; The MIT Press: Cambridge, MA, USA, 2002.
- Zube, E.H. Environmental perception. In *Environmental Geology*; Springer: Dordrecht, The Netherlands, 1999; pp. 214–216. ISBN 978-1-4020-4494-6.
- Davis, S.F.; Buskist, W. *21st Century Psychology: A Reference Handbook*; Sage: Southend Oaks, CA, USA, 2008.
- International Organization for Standardization. *ISO 12913-1:2014 Acoustics—Soundscape—Part 1: Definition and Conceptual Framework*; International Organization for Standardization: Geneva, Switzerland, 2014.
- Brambilla, G.; Maffei, L.; Di Gabriele, M.; Gallo, V. Merging physical parameters and laboratory subjective ratings for the soundscape assessment of urban squares. *J. Acoust. Soc. Am.* **2013**, *134*, 782–790. [[CrossRef](#)]
- Kang, J.; Zhang, M. Semantic differential analysis of the soundscape in urban open public spaces. *Build. Environ.* **2010**, *45*, 150–157. [[CrossRef](#)]
- Jo, H.I.; Jeon, J.Y. The influence of human behavioral characteristics on soundscape perception in urban parks: Subjective and observational approaches. *Landsc. Urban Plan.* **2020**, *203*, 103890. [[CrossRef](#)]
- Brown, A.L.; Kang, J.; Gjestland, T. Towards standardization in soundscape preference assessment. *Appl. Acoust.* **2011**, *72*, 387–392. [[CrossRef](#)]

18. Maffei, L.; Masullo, M.; Pascale, A.; Ruggiero, G.; Puyana-Romero, V. Immersive Virtual Reality in community planning: Acoustic and visual congruence of simulated vs. real world. *Sustain. Cities Soc.* **2016**, *27*. [CrossRef]
19. Hong, J.Y.; Jeon, J.Y. Influence of urban contexts on soundscape perceptions: A structural equation modeling approach. *Landsc. Urban Plan.* **2015**, *141*, 78–87. [CrossRef]
20. Maffei, L.; Puyana-Romero, V.; Brambilla, G.; di Gabriele, M.; Gallo, V. Characterization of the soundscape of urban waterfronts. In Proceedings of the Forum Acusticum, European Acoustic Association, Krakow, Poland, 7–12 September 2014.
21. Carles, J.L.; Barrio, I.L.; De Lucio, J.V. Sound influence on landscape values. *Landsc. Urban Plan.* **1999**, *43*, 191–200. [CrossRef]
22. Hong, J.Y.; Jeon, J.Y. Designing sound and visual components for enhancement of urban soundscapes. *J. Acoust. Soc. Am.* **2013**, *134*, 2026–2036. [CrossRef]
23. Song, J.; Ma, H.; Yang, Q.; Zhang, S. Interaction of color perception and annoyance evaluation of road traffic noise. In *Proceedings of the Internoise*; Institute of Noise Control Engineering/Japan & Acoustical Society of Japan: Osaka, Japan, 2011; pp. 1–6.
24. Galbrun, L.; Calarco, F.M. a Audio-visual interaction and perceptual assessment of water features used over road traffic noise. *J. Acoust. Soc. Am.* **2014**, *136*, 2609. [CrossRef]
25. Menzel, D.; Haufe, N.; Fastl, H. Colour-influences on loudness judgements. In Proceedings of the 20th ICA, International Commission for Acoustics, Sydney, Australia, 23–27 August 2010; pp. 2–6.
26. Liu, J.; Kang, J.; Luo, T.; Behm, H. Landscape effects on soundscape experience in city parks. *Sci. Total Environ.* **2013**, *454–455*, 474–481. [CrossRef]
27. D’Alessandro, F.; Evangelisti, L.; Guattari, C.; Grazieschi, G.; Orsini, F. Influence of visual aspects and other features on the soundscape assessment of a university external area. *Build. Acoust.* **2018**, *25*, 199–217. [CrossRef]
28. Gan, Y.; Luo, T.; Breitung, W.; Kang, J.; Zhang, T. Multi-sensory landscape assessment: The contribution of acoustic perception to landscape evaluation. *J. Acoust. Soc. Am.* **2014**, *136*, 3200–3210. [CrossRef]
29. Anderson, L.M.; Mulligan, B.E.; Goodman, L.S.; Regen, H.Z. Effects of Sounds on Preferences for Outdoor Settings. *Environ. Behav.* **1983**, *15*, 539–566. [CrossRef]
30. López Barrio, I.; Carles, J. Acoustic Dimensions of Inhabited Areas: Quality Criteria. *Soundsc. Newsl.* **1995**, *10*, 6–8.
31. Jeon, J.Y.; Jo, H.I. Effects of audio-visual interactions on soundscape and landscape perception and their influence on satisfaction with the urban environment. *Build. Environ.* **2020**, *169*, 106544. [CrossRef]
32. Hong, J.Y.; Jeon, J.Y. Relationship between spatiotemporal variability of soundscape and urban morphology in a multifunctional urban area: A case study in Seoul, Korea. *Build. Environ.* **2017**, *126*, 382–395. [CrossRef]
33. Yu, L.; Kang, J.; Liang, H.; Yang, Y. Soundscape identification in terms of urban morphology. In Proceedings of the ICA 2016 22nd International Congress on Acoustics, International Commission for Acoustics, Buenos Aires, Argentina, 5–9 September 2016.
34. Hao, Y. Effects of Urban Morphology on Urban Sound Environment from the Perspective of Masking Effects. Ph.D. Thesis, University of Sheffield, Sheffield, UK, 2014.
35. Liu, J.; Kang, J.; Luo, T.; Behm, H.; Coppack, T. Spatiotemporal variability of soundscapes in a multiple functional urban area. *Landsc. Urban Plan.* **2013**, *115*, 1–9. [CrossRef]
36. Puyana-Romero, V.; Maffei, L.; Brambilla, G.; Ciaburro, G. Acoustic, visual and spatial indicators for the description of the soundscape of water front areas with and without road traffic flow. *Int. J. Environ. Res. Public Health* **2016**, *13*, 934. [CrossRef]
37. Puyana-Romero, V.; Maffei, L.; Brambilla, G.; Ciaburro, G. Modelling the soundscape quality of urban waterfronts by artificial neural networks. *Appl. Acoust.* **2016**, *111*, 121–128. [CrossRef]
38. University of Alberta Dictionary of Cognitive Science. Available online: [http://www.bcp.psych.ualberta.ca/~mike/Pearl\\_Street/Dictionary/dictionary.html](http://www.bcp.psych.ualberta.ca/~mike/Pearl_Street/Dictionary/dictionary.html) (accessed on 2 January 2021).
39. Williams, T. The Effects of Expectations on Perception: Experimental Design Issues and Further Evidence. *Fed. Reserve Bank Boston Work. Pap.* **2007**, 7–14. [CrossRef]
40. Bruce, N.S.; Davies, W.J. The effects of expectation on the perception of soundscapes. *Appl. Acoust.* **2014**, *85*, 1–11. [CrossRef]
41. Brambilla, G.; Maffei, L. Responses to noise in urban parks and in rural quiet areas. *Acta Acust. United Acust.* **2006**, *92*, 881–886.
42. Brambilla, G.; Gallo, V.; Asdrubali, F.; D’Alessandro, F. The perceived quality of soundscape in three urban parks in Rome. *J. Acoust. Soc. Am.* **2013**, *134*, 832–839. [CrossRef]
43. Puyana-Romero, V.; Ciaburro, G.; Maffei, L. The soundscape and the degree of match of a waterfront with the expectations placed on it. The cases study of Naples and Brighton. In Proceedings of the Internoise 2016 45th International Congress and Exposition on Noise Control Engineering: Towards a Quieter Future, Hamburg, Germany, 21–24 August 2016.
44. Botteldooren, D.; De Coensel, B. A model for long-term environmental sound detection. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 2017–2023. [CrossRef]
45. Bruce, N.; Condie, J.; Henshaw, V.; Payne, S.R. Analysing olfactory and auditory sensescapes in English cities: Sensory expectation and urban environmental perception. *Ambiances Int. J. Sens. Environ. Archit. Urban Spaces* **2015**. [CrossRef]
46. Crichton, F.; Dodd, G.; Schmid, G.; Petrie, K.J. Framing sound: Using expectations to reduce environmental noise annoyance. *Environ. Res.* **2015**, *142*, 609–614. [CrossRef] [PubMed]
47. Krzywicka, P.; Byrka, K. Restorative Qualities of and Preference for Natural and Urban Soundscapes. *Front. Pediatr.* **2017**, *8*, 1705. [CrossRef] [PubMed]

48. Liu, J.; Kang, J.; Behm, H.; Luo, T. Effects of landscape on soundscape perception: Soundwalks in city parks. *Landsc. Urban Plan.* **2014**, *123*, 30–40. [[CrossRef](#)]
49. Thoma, M.V.; Mewes, R.; Nater, U.M. Preliminary evidence: The stress-reducing effect of listening to water sounds depends on somatic complaints. *Medicine* **2018**, *97*, e9851. [[CrossRef](#)]
50. Galbrun, L.; Ali, T.T. Perceptual assessment of water sounds for road traffic noise masking. In Proceedings of the Acoustics Nantes Conference, Nantes, France, 23–27 April 2012; pp. 2147–2152.
51. Nilsson, M.E.; Alvarsson, J.; Rådsten-Ekman, M.; Bolin, K. Auditory masking of wanted and unwanted sounds in a city park. *Noise Control Eng. J.* **2010**, *58*, 524–531. [[CrossRef](#)]
52. You, J.; Lee, P.J.; Jeon, J.Y. Evaluating water sounds to improve the soundscape of urban areas affected by traffic noise. *Noise Control Eng. J.* **2010**, *58*, 477–483. [[CrossRef](#)]
53. Watts, G.R.; Pheasant, R.J.; Horoshenkov, K.V.; Ragonesi, L. Measurement and subjective assessment of water generated sounds. *Acta Acust. United Acust.* **2009**, *95*, 1032–1039. [[CrossRef](#)]
54. Maffei, L.; Masullo, M.; Pascale, A.; Ruggiero, G.; Puyana-Romero, V. On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces. *Energy Procedia* **2015**, *78*, 471–476. [[CrossRef](#)]
55. Puyana-Romero, V.; Solange-Lopez-Segura, L.; Maffei, L.; Hernández-Molina, R.; Masullo, M. Interactive Soundscapes: 360°-Video Based Immersive Virtual Reality in a Tool for the Participatory Acoustic Environment Evaluation of Urban Areas. *Acta Acust. United Acust.* **2017**, *103*, 574–588. [[CrossRef](#)]
56. Sun, K.; De Coensel, B.; Filipan, K.; Aletta, F.; Van Renterghem, T.; De Pessemer, T.; Joseph, W.; Botteldooren, D. Classification of soundscapes of urban public open spaces. *Landsc. Urban Plan.* **2019**, *189*, 139–155. [[CrossRef](#)]
57. Asakura, T.; Tsujimura, S.; Yonemura, M.; Hyojin, L.; Sakamoto, S. Effect of immersive visual stimuli on the subjective evaluation of the loudness and annoyance of sound environments in urban cities. *Appl. Acoust.* **2019**, *143*, 141–150. [[CrossRef](#)]
58. International Organization for Standardization. *ISO 12913-2 Soundscape*; International Organization for Standardization: Geneva, Switzerland, 2018; Volume 7.
59. Farina, A.; Tronchin, L. Measurements and reproduction of spatial sound characteristics of auditoria. *Acoust. Sci. Technol.* **2005**, *26*, 193–199. [[CrossRef](#)]
60. Axelsson, Ö.; Nilsson, M.E.; Berglund, B. A principal components model of soundscape perception. *J. Acoust. Soc. Am.* **2010**, *128*, 2836–2846. [[CrossRef](#)] [[PubMed](#)]
61. Jeon, J.Y.; Lee, P.J.; You, J.; Kang, J. Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. *J. Acoust. Soc. Am.* **2010**, *127*, 1357–1366. [[CrossRef](#)] [[PubMed](#)]
62. Galbrun, L.; Ali, T.T. Acoustical and perceptual assessment of water sounds and their use over road traffic noise. *J. Acoust. Soc. Am.* **2013**, *133*, 227–237. [[CrossRef](#)] [[PubMed](#)]
63. Cohen, J. Using Effect Size—Or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **2012**, *4*, 279–282.
64. Coe, R. It's the Effect Size, Stupid: What effect size is and why it is important. In Proceedings of the Conference of the British Educational Research Association, Exeter, UK, 12–14 September 2002.
65. Rosenthal, R. Parametric measures of effect size. In *The Handbook of Research Synthesis*; Russell Sage Foundation: New York, NY, USA, 1994; pp. 231–244. ISBN 0-87154-226-9.