


## Article

# Effect of Audio–Visual Factors in the Evaluation of Crowd Noise

Xiaoyin Yang <sup>1,\*</sup> and Jian Kang <sup>2,\*</sup> 

<sup>1</sup> Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, School of Architecture, Harbin Institute of Technology, Harbin 150001, China

<sup>2</sup> UCL Institute for Environmental Design and Engineering, The Bartlett, University College London (UCL), London WC1H 0NN, UK

\* Correspondence: xyin0918@126.com (X.Y.); j.kang@ucl.ac.uk (J.K.)

**Abstract:** A crowd can be both a sound source and an absorber. The sound of human voices significantly impacts evaluations of acoustic indicators in urban public spaces. This study aimed to investigate whether human sound impacts evaluations of the overall environment from both visual and auditory aspects. Primary sound sources and scenes in Harbin, China urban public space served as the research object. Four sets of sound sources and six sets of images were collected in situ in urban public spaces. A subjective evaluation of both visual and auditory aspects was then performed in the laboratory. The results showed that when different types of sounds in urban public spaces are superimposed with human sound, the volume of human sound (45.6 dBA, 55.6 dBA, and 65.6 dBA) significantly affects the acoustic evaluation of the environment. When the superimposed sounds were birds and music, the evaluation of the environment decreased with the human voice increasing in volume. Crowd density and the surrounding visual environment also influence evaluations of the overall sound environment. In this study, the sound preference and acoustic comfort of birdsong and music decreased as the human sound volume increased. The effect of human sound combined with traffic sounds significantly decreased the scores for sound preference and acoustic comfort at higher volumes. The results of the experiments on audio–visual interactions in which people evaluated visual scenes showed that the influence of the visual density of a crowd on assessments of the sound environment is negatively related to the magnitude of the sound of the crowd. When human voices are at 45.6 dBA and 65.6 dBA, there is a significant effect on the evaluation of visual scenes for high-density people. When the sound pressure level of human voices is the same, changes in the visual environment are more likely to affect people’s evaluation of the overall sound environment.



**Citation:** Yang, X.; Kang, J. Effect of Audio–Visual Factors in the Evaluation of Crowd Noise. *Appl. Sci.* **2023**, *13*, 3652. <https://doi.org/10.3390/app13063652>

Received: 3 January 2023

Revised: 26 February 2023

Accepted: 6 March 2023

Published: 13 March 2023



**Copyright:** © 2023 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/>).

**Keywords:** soundscape; human sound; audio–visual interaction; sound preference; acoustic comfort

## 1. Introduction

In many studies related to soundscapes, general evaluations of soundscapes are usually considered to be evaluations of sound levels, that is, the subjective evaluation of loudness, usually of background noise [1–6]. Soundscape evaluations are also considered to be evaluations of sound preferences, that is, evaluations of foreground sounds [7–11]. Individual sounds play an important role in the overall soundscape [12]. Therefore, evaluating sound preferences is crucial for determining the quality of a given space’s acoustic landscape. Environmental psychologists point out that the implicit attributes of social/cultural factors are interrelated with the outward attributes of the physical environment with respect to their influence on people’s perception of physical sound [13,14]. In contrast to musical preferences that focus on the sound itself, judgment of everyday sounds is based on a collection of relevant information about our surroundings [8].

In urban public spaces, natural sounds and traffic noise are particularly concerning. Natural sounds usually play an active role in urban open spaces. They can improve one’s cognitive state [15], help regulate mood, and promote feelings of pleasure. Regarding

the effect of natural sounds on sound quality, Axelsson et al. showed that water sounds can effectively improve the overall perceived quality of a soundscape by masking park noise [16]. Liu, Yang, and Xiong showed that natural sounds in historic districts positively impact the perception of quiet and harmonious sounds [17]. Ren, Kang, and Liu noted that people have a higher preference for natural sounds and melodies [18]. Conversely, traffic noise tends to play a negative role. For example, mechanical noise has been shown to affect human hearing and reduce stress recovery, leading to bad moods [19–21]. Relatively neutral sounds, such as human voices, should also be considered [22]. Human sounds are a special type of sound in which the source and receiver are both humans [23]. Hong and Jeon indicated that human sounds contributed to an increase in both pleasantness and eventfulness in commercial streets [24]. Yet, human sounds may positively or negatively affect the sound environment. Under certain specific conditions, human sounds have positive effects. For example, the sound of children playing increases the pleasantness of the sound environment [20]. However, human voices also have negative effects. For example, the sound of children crying can increase negative emotions in people [21]. Jo and Jeon demonstrated that human sound reduces the tranquility of a park but increases its vitality [25].

Eighty percent of human sensory experience relies on visual stimulation [26], so the correlation between landscapes and soundscapes has been extensively studied [16,27–29]. Most studies on the auditory response to soundscapes have focused on the auditory properties of the landscape. However, while human perception relies heavily on visual factors, most landscape studies on the auditory response to soundscapes remain limited to analyzing auditory properties. Human perception of the environment is not generated by a single, isolated sense; rather, it is an inherently multisensory experience involving visual, auditory, olfactory, and other sensory stimuli interacting [30]. Visual and auditory domains interact with each other in influencing human perception and behavior [31]. Soundscape research is closely related to ‘context,’ and visual factors are an important part of the context. Therefore, by studying audio–visual interactions from the perspective of the soundscape, we can study the whole environment in a more auditory manner. Research of audio–visual interactions in the field of soundscapes includes both field and laboratory studies. Field studies consider soundscapes based on real places. However, due to the limitations of geographical conditions, it is difficult to present the sound environment with different attributes during the same survey; furthermore, this method can downplay other factors in the natural environment, including olfactory and other sensory elements that can affect perception [32]. Hence, laboratory research involves more widespread use of methods reproducing the sound environment. Through a series of field investigations and laboratory experiments, Anderson, Mulligan, Goodman, and Regen (1983) studied the influence of sound on preferences for outdoor environments and found an interaction between sound and vision. They also compared the differences between field research and laboratory research. Laboratory research has since evolved into a developed research method for audio–visual research [33]. To study audio–visual interactions in the laboratory, researchers now mainly focus on the subjective evaluation of factors that influence the perception of the environment.

In a noisy environment, people unconsciously raise their voices [34,35]. Crowd density is an environmental factor expressed as the number of people in a space divided by the area of the space. Crowd density estimates were first used for crowd monitoring to determine when an area reaches a greater-than-expected crowd density that could endanger the safety of people [36]. Subsequently, crowd density has been used in various research areas, such as environmental pollution, building technology, and landscape design. Previous studies have shown that crowd density affects certain spatial or environmental characteristics in a given space [37]. In addition, previous studies have shown that crowd density affects behavior [38]. Regarding the masking effect of human sounds, Meng, Sun, and Kang pointed out that human sounds placed near temporary open-air markets with adjacent road traffic can effectively mask traffic noise [39]. Shu et al. also pointed out that urban

noises, such as traffic noise, construction noise, and community noise, can be effectively masked by human sounds [40]. However, studies on crowd distribution and the impact of human sound on the sound environment of urban public spaces are limited. Crowds can be regarded as sound sources and absorbers. The effect of crowd density on the sound environment is studied through subjective features such as audio–visual interaction evaluations [41].

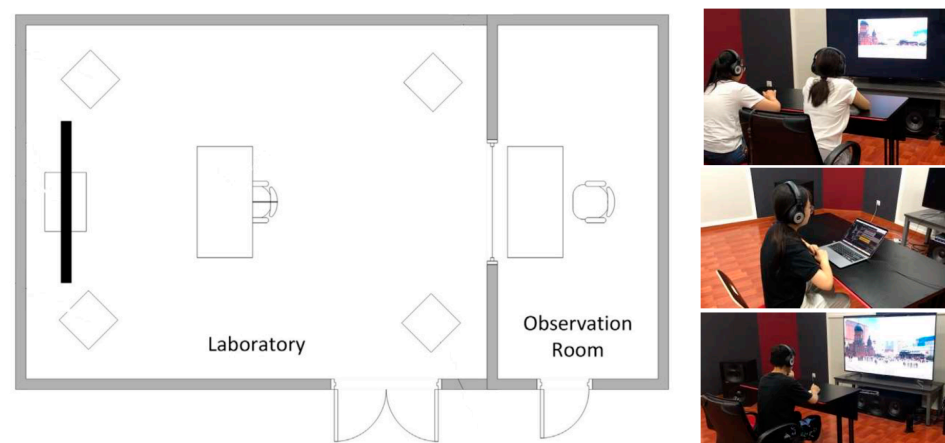
Therefore, this study aimed to research the following three questions: (1) What is the effect of human sounds at different volumes on the overall sound environment when superimposed on other sounds? (2) What is the effect of crowd density on overall evaluations of the current environment, both visual and auditory? (3) With the same human sound as the background, do changes in the visual scene affect people's evaluation of the sound environment? This study examines the audio–visual impact of human sound on evaluations of the soundscapes of urban public spaces through experiments on the audio–visual interaction of four main groups of sound sources and visual images in urban public spaces.

## 2. Materials and Methods

### 2.1. Experimental Design

Sources of sound in urban public spaces include natural, human, mechanical, and instrumental sounds [42]. Therefore, four groups of sound sources were selected in the experiment: human conversation, birdsong, traffic noise, and music. Three groups of visual images with different crowd densities in the same scene and three groups of images with different visual environments in different locations were also used.

The experiments were conducted in a laboratory at the Harbin Institute of Technology in China, shown in Figure 1. The volume of the laboratory was 186 m<sup>3</sup>, and the background noise was 11 dBA. A TV screen (Samsung H6400: 166.030 cm × 93.375 cm (75 in)), resolution: 1920\*1080 p) was used for live video playback. Participants were asked to sit in designated seats. Sennheiser rs170 headphones provided the audio. To ensure that participants could concentrate on the test, no items other than the necessary equipment were present in the laboratory.



**Figure 1.** Laboratory layout schematic.

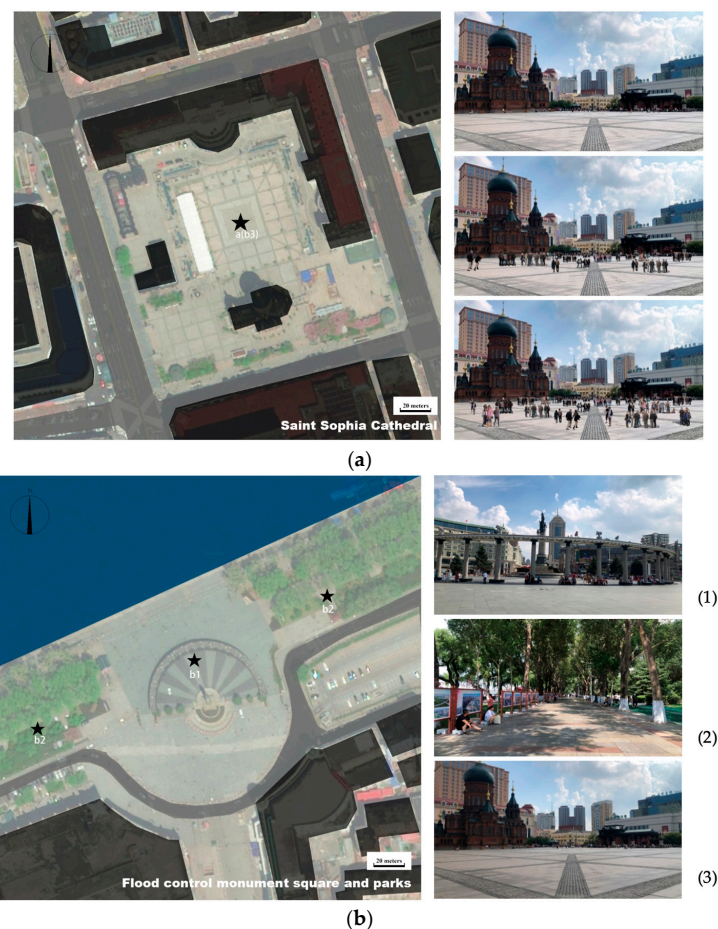
### 2.2. Sound Stimuli

For the audio stimuli, human conversations were chosen as the main object of the study. Conversational sounds were then combined with birdsong, music, and traffic noise, corresponding to natural, human, and mechanical sounds commonly found in urban environments. The recordings were made in the field using a 10-channel high-fidelity portable recorder. The recording equipment was placed at 1.5 m perpendicular to the ground, and each recording session lasted 5 min.

When the sound pressure levels simulated in the laboratory are the same as in the field, the sound may appear too loud and cause discomfort due to the different background sound pressure levels. Therefore, the actual sound pressure level was weakened in the laboratory experiment, and an artificial head (HMS IV) was used to simulate the human ear for calibration and measurement. The adjustment of sound pressure levels was performed using Adobe Audition software (Adobe Audition 2021). The sound pressure level gradient was 10 dBA, and the sound levels were 45.6 dBA, 55.6 dBA, and 65.6 dBA, representing the low-, medium-, and high-volume levels, respectively. Participants were asked to evaluate the subjective loudness based on the currently heard sound.

### 2.3. Visual Stimuli

We selected two typical squares in Harbin as sampling points: (a) the square of Saint Sophia Cathedral is semi-open, surrounded by buildings, and adjacent to the city's main traffic artery. It is mainly affected by traffic and ambient noise from the surrounding commercial area. The crowd consists mainly of foreign tourists and residents passing through; and (b) the Flood control monument square is located near Songhua River; there is flowing water on one side and a commercial area on the other side. The building complex is some distance away and the square is mainly affected by traffic noise. The crowd is mainly tourists and residents. In the square of Saint Sophia Cathedral, we selected three main sets of images with different crowd densities for the visual stimuli in the same scene (see Figure 2). We also included three scenes with different environments but the same crowd density. The scene was recorded through the photo and video function of a digital camera (canon 5d3). To record the immediate surrounding situation, the equipment was placed at a vertical height of 1.6 m from the ground, and each video recording was 5 min.

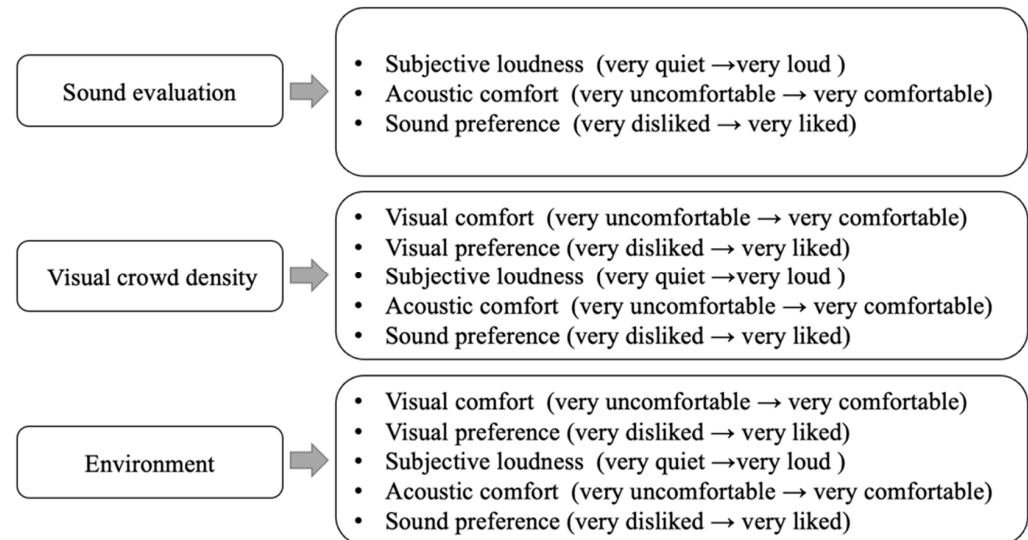


**Figure 2.** Landscape recorded: (a) crowd density; and (b) three environment scenes.



## 2.4. Questionnaires

Questionnaires were distributed in the laboratory to investigate participants' evaluations of the recorded landscapes. The questionnaire consisted of three main parts, which corresponded to our three research questions. Each part had its own evaluation indicators, as shown in Figure 3. Answers were given on a five-point Likert scale, from very dislike to very like.



**Figure 3.** Contents of questionnaire and specific evaluation indicators.

A total of 33 people participated in this study (54.5% females). To reduce intra-group differences, we chose students from our university as the participants. Participants were an average age of 26 (SD = 4.0; Min = 21; Max = 37), and all had normal hearing. Each participant was randomized to assess audio-only, visual-only, and audio-visual environments.

## 2.5. Subjective Evaluation Procedure

- (1) We first investigated the relationship between human voices and other background sounds. A total of nine audio sets were played. The participants randomly heard three groups of low, medium, and high vocals combined with traffic sounds, birdsong, and music. The questions were answered in 10 s intervals between each clip (each clip was 20 s long);
- (2) Next, we investigated the effect of crowd density on participants' evaluations of the sound environment. The three sets of images in Figure 2a were randomly combined with low, medium, and high human voices, a total of nine audio sets. Each group of materials was played for 20 s with 10 s intervals to answer the questions;
- (3) Finally, we investigated the relationship between the visual environment and sound environment evaluation. We play the three sets of videos from Figure 2b while randomly playing three sets of low, medium, and high vocals, for a total of nine combinations. Each video was played for 20 s, and the questions were answered at 10 s intervals.

## 3. Results

### 3.1. Effect of Human Voice Volume and Other Sounds on Acoustic Evaluation

IBM SPSS Statistics 21.0 was used to establish a database with all results. The acoustic evaluations of human voice volume and sound types are shown in Table 1 (SD means Standard Deviation). In terms of acoustic comfort, people had positive attitudes toward the natural sound, while the opposite was true for mechanical sound [4,43,44]. The trend in sound preferences was similar to the trend in acoustic comfort. That is, sounds with higher

acoustic comfort also had higher sound preference evaluations, consistent with previous studies [24,32].

**Table 1.** Acoustic evaluation.

Sound Types	Human Voice Volume	Acoustic Comfort		Sound Preference		Sound Loudness	
		Mean	SD	Mean	SD	Mean	SD
Traffic	Low	2.50	0.67	2.38	0.83	3.47	0.80
	Medium	2.47	0.67	2.19	0.64	4.00	0.58
	High	1.88	0.98	1.75	0.72	4.5	0.72
Bird	Low	3.66	0.87	3.53	0.92	2.81	0.86
	Medium	3.16	0.63	3.06	0.72	3.16	0.72
	High	1.47	0.57	1.38	0.55	4.41	0.80
Music	Low	3.94	0.91	4.16	0.85	2.84	0.68
	Medium	3.28	0.81	3.34	0.90	3.38	0.71
	High	2.38	0.87	2.16	0.72	4.38	0.66

Table 2 shows the significance of the indicators under the main effect. Mixed design analyses of variance (ANOVAs) were run to test the differences in mean ratings between subgroups divided by the two main variables. Sound type and volume were taken as independent variables. The results showed that in the three sound sources, the main effect of sound types and volume levels, and the interaction between them all, had significant effects on sound loudness, acoustic comfort, and sound preference ( $p < 0.05$ ).

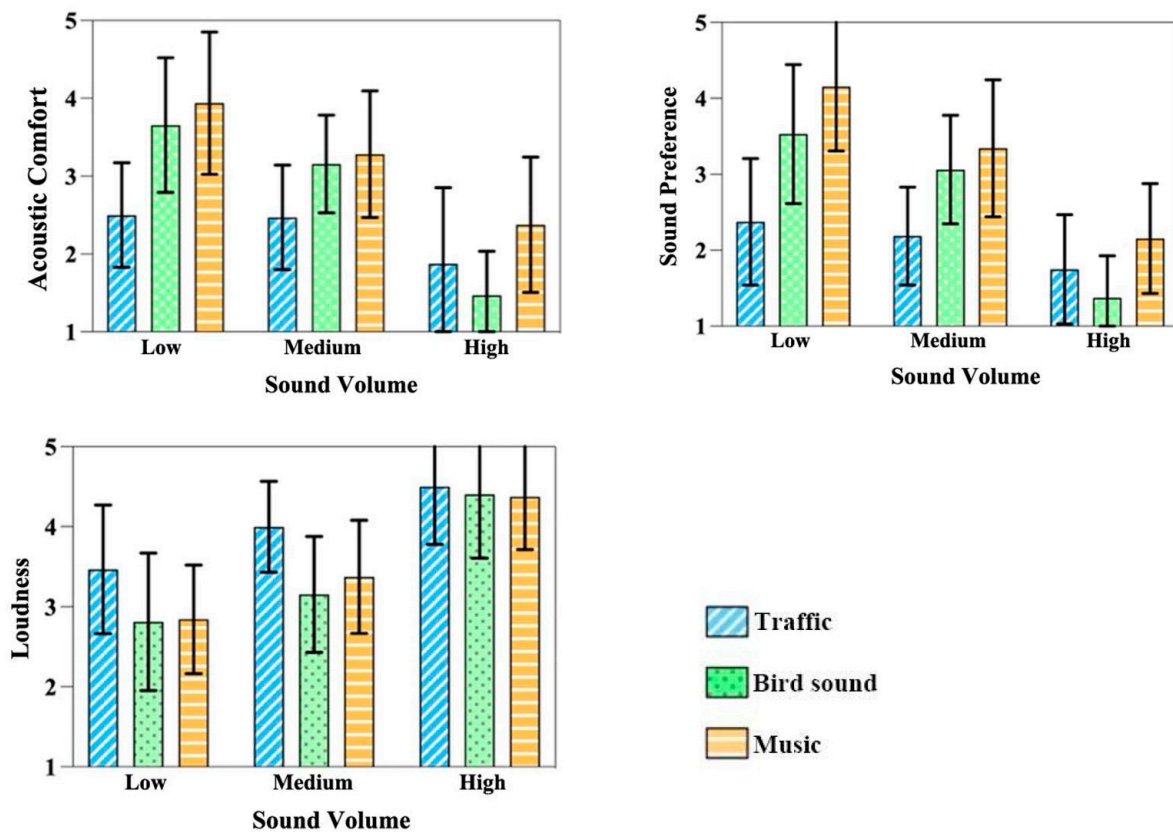
**Table 2.** Significance of the effects of sound type and volume level on acoustic evaluation.

Source	Acoustic Comfort	Sound Preference	Subjective Loudness
Traffic	<b>0.02</b>	<b>0.00</b>	<b>0.00</b>
Birds	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Music	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Low volume			
Traffic + Birds	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Traffic + Music	<b>0.00</b>	<b>0.00</b>	<b>0.01</b>
Birds + Music	0.36	<b>0.01</b>	<b>0.00</b>
Traffic + Birds + Music	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Medium volume			
Traffic + Birds	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Traffic + Music	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Birds + Music	0.76	0.31	0.40
Traffic + Birds + Music	<b>0.00</b>	<b>0.00</b>	<b>0.01</b>
High volume			
Traffic + Birds	0.12	<b>0.07</b>	0.86
Traffic + Music	<b>0.04</b>	<b>0.05</b>	0.77
Birds + Music	<b>0.00</b>	<b>0.00</b>	0.98
Traffic + Birds + Music	<b>0.02</b>	<b>0.04</b>	0.68

Note: significance ( $p < 0.05$ ) is in bold.

A one-way ANOVA was run to test the differences in mean ratings between the subgroups divided by the variables that interacted. After removing the factors without statistical significance, the model was simplified. Figure 4 shows the evaluation of the acoustic metrics after superimposing the human voice with different sound types (birdsong, traffic sounds, and music) under the auditory-only stimuli. (The error bars represent the standard deviation [S.D.] of the average values.) In terms of loudness, there was an interaction between the type of sound and the volume. With respect to sound types, regardless of the variation in volume level, the loudness of the human voice after superimposing bird sounds and musical sounds was rated lower than that of traffic noise. When the participants heard the first two sound types, they did not feel the sound loudness as much. The participants

even felt slightly quiet for the low- and medium-volume human voice superimposed with birdsong and music ( $p < 0.01$ ). With respect to volume level, sound loudness increased significantly with increasing volume level; that is, the higher the volume, the louder the participants felt ( $p < 0.01$ ).



**Figure 4.** Mean ratings of evaluation indicators under the effect of acoustic factors.

There is an interaction between acoustic comfort and sound preference in terms of sound types and volume levels. With respect to sound types, acoustic comfort and sound preference were rated significantly higher than the superimposed traffic sounds for the low- and medium-volume conversational sounds overlaid with birdsong and music sounds, but the high-volume conversational sounds overlaid with birdsong were rated lower than the superimposed music and traffic sounds. After superimposing birdsong and music sounds with human voices at low- and medium-volume levels, participants' acoustic comfort ranged between 3.0 and 4.0. Sound preference ranged between 3.0 and 4.5, and participants felt comfortable and liked it ( $p < 0.01$ ). In contrast, after superimposing birdsong and music over high-volume vocals, the acoustic comfort was between 1.0 and 2.5 and the sound preference was between 1.5 and 2.5. Participants felt uncomfortable and disliked it ( $p < 0.01$ ). When traffic sounds were superimposed over low- and medium-volume human voices, the participants felt uncomfortable and disliked it. The participants felt very uncomfortable and disliked it when traffic noise was superimposed over high-volume human voices. Their evaluation scores ranged from 2.0 to 2.5 for acoustic comfort ( $p < 0.01$ ) and 1.5 to 2.0 ( $p < 0.01$ ) for sound preference; the participants felt very uncomfortable and disliked it very much. Concerning volume, when the sound type was birdsong or music, acoustic comfort and preference decreased significantly as the volume of the human voice increased ( $p < 0.05$ ). Moreover, when the sound type was traffic, acoustic comfort and preference decreased significantly only when the human voice volume was high. Acoustic comfort and preference were higher at low volume than at high volume ( $p < 0.01$ ). It was observed that sound type and volume, as well as their interactions, affected participants'

acoustic evaluations different degrees. Regardless of voice volume, birdsong and music were rated higher than traffic, and regardless of the type of sound, acoustic evaluations were significantly higher for low-volume voices than for high.

### 3.2. Effect of Visual Crowd Density on Sound Environment Evaluation

Table 3 shows the overall evaluations under the simultaneous effect of sound and vision. When the crowd density is constant, participants' visual comfort and preference as well as acoustic comfort and sound preference decreased as the vocal volume increased. Figure 5 shows the sound preference values for different crowd densities in the same scene with different conversational sound pressures; that is, the values in the presence of both auditory and visual stimuli. (The error bars represent the S.D. of the average values.) Only when the sound volume was medium, were the loudness evaluations significantly different for the empty scene and the scene with a moderate crowd; the empty scene appeared louder ( $p < 0.05$ ).

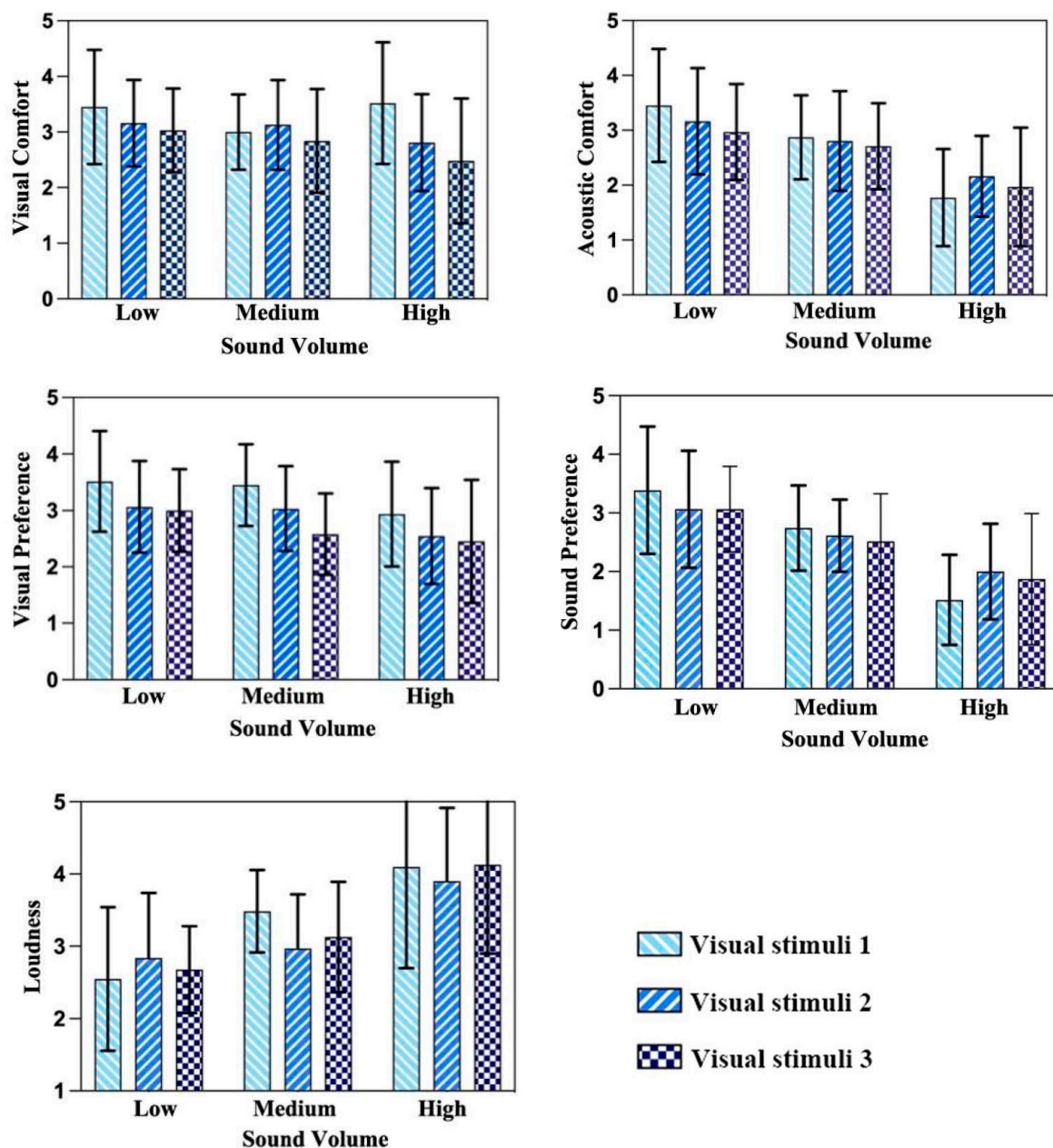


Figure 5. The mean ratings of evaluation indicators under the effect of visual crowd density.



There was an interaction between crowd density and volume level for acoustic comfort and visual comfort. In scenes with low and high crowd densities, the scores for both types of comfort were higher for moderate volume levels than for high ( $p < 0.01$ ). In scenes with medium and high volumes, visual comfort scores were higher when the crowd density was moderate than when it was high ( $p < 0.05$ ). In scenes with low volume, acoustic comfort scores were higher for scenes with low crowd density than those for with high crowd density. That indicates that even though the sound is not loud, the crowd density of the visual environment still significantly affects the overall sound preference evaluation ( $p < 0.05$ ). There is also an interactive relationship between sound preference and visual preference. There was no mutual influence between low and moderate volumes in scenes with low and moderate crowd densities. However, the relationships between high volume and low and moderate crowd densities were significant, with higher scores for visual and sound preferences when the volume level was low ( $p < 0.01$ ). Scenes with low crowd density had higher visual preference scores than the other two scenes for all three volume levels.

**Table 3.** The evaluation under the effect of sound and visual stimuli.

Subjective Evaluation	Crowd Density	Sound Volume		
		Low	Medium	High
Acoustic comfort	Low	3.45	2.87	1.77
	Medium	3.16	2.80	2.16
	High	2.97	2.71	1.97
Sound preference	Low	3.39	2.74	1.51
	Medium	3.06	2.61	2.00
	High	3.06	2.52	1.87
Sound loudness	Low	2.55	3.48	4.10
	Medium	2.84	2.97	3.90
	High	2.68	3.13	4.13
Visual comfort	Low	3.45	3.52	3.00
	Medium	3.26	3.13	2.81
	High	3.03	2.84	2.48
Visual preference	Low	3.52	3.45	2.94
	Medium	3.06	3.03	2.55
	High	3.00	2.58	2.45

### 3.3. Effect of Audio–visual Stimuli in Different Environments

Figure 6 shows the sound preferences for conversational sounds in different environments under simultaneous audio–visual stimuli (The error bars represent the S.D. of the average values). For all three environmental groups, there was no significant difference between the overall preference and visual environment preference scores for the same background sound. However, there was a significant difference in the sound preference and acoustic comfort scores for the three groups of sounds in the same set of visual environments. The change in volume also significantly affected visual preference and comfort scores. For the original three groups of voices, the lower the sound, the higher the sound preference score ( $p < 0.01$ ). The overall preference and comfort scores changed after adding visual factors. The environmental evaluation of the videos one and two changed from significant to non-significant, with a difference in scores between the low and medium groups. The difference between the scores of the medium and high groups was still significant ( $p < 0.01$ ). The louder the sound, the lower the overall preference and comfort scores. Therefore, after adding the visual factor, evaluations of the overall environment were not affected when the sound pressure level of the human voice was at 45.6 to 55.6 db. However, when the sound pressure level was 55.6 to 66.6 dB, overall environment scores were lower as the sound got louder.

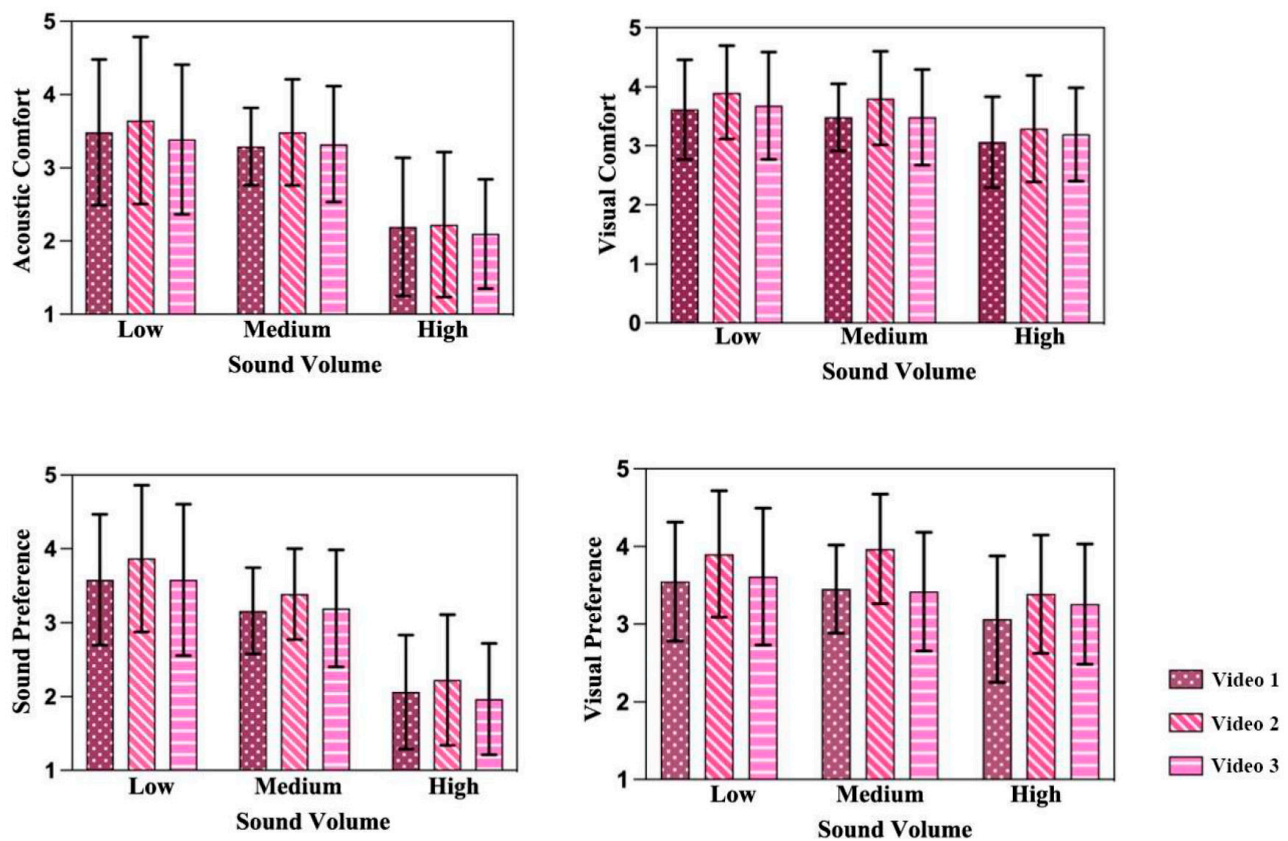


Figure 6. Mean ratings of evaluation indicators under the effect of different environments.

## 4. Discussion

### 4.1. Applications

When designing soundscapes in urban public spaces, visual factors in the surrounding environment need to be considered. In a more open environment with low crowd density, even if human voices are noisy, evaluations of the overall sound environment will not be significantly impacted. Music can be added appropriately, according to the scene, to improve the preference and comfort of the current environment. In noisy environments, people will unconsciously raise their voices [32,33]. Scenes with surrounding traffic and crowds are usually more complex, people will raise their speaking voices, which greatly reduces the comfort of the environment. Therefore, crowd density control can improve people's sound preferences and comfort.

### 4.2. Limitations and Future Study

The city of Harbin, which was selected for this study, can only represent a portion of the different forms of urban public space. Different urban forms have different sound source characteristics. The sound sources of Harbin Sofia Church, Flood Control Memorial Tower Square, and Stalin Park are a few of the study sites with large differences. However, they cannot cover all space types. In future studies, the same method of audio–visual interaction can be used to research different elements of the soundscape. Studies have currently focused on urban public spaces in high density cities. This method can be used in the future to study the sound environment of public spaces in other low-density cities. Other cities should be used as the focus of research, more types of urban public spaces should be added, and more locally appropriate sound sources should be selected for discussion to complement the existing findings.

## 5. Conclusions

This study investigated the public space of Harbin city square in the field. It assessed the influence of people's conversations, crowd density, and different environments on their evaluation of the sound environment through different combinations and comparisons of visual and auditory sensations in the laboratory. The results of the study show the following:

In urban public spaces, the magnitude of different sound types (birdsong, music, and traffic sounds), when added to human vocal sounds, significantly affect people's evaluations of the acoustic indicators of their current environment. Regardless of the superimposed sound, acoustic comfort and preference scores were higher for low- and medium-volume human sounds. The effects of birdsong and music sounds decrease as human voice levels increase. The effect of traffic sounds was not significant at low- and medium-volume voice levels, but the scores for sound preference and acoustic comfort were significantly lower at high volume ( $p < 0.05$ ).

When the crowd density is large, conversational sounds at low- and high-volume levels affect people's acoustic comfort and sound preference evaluations. When the crowd density is sparse, all three conversational sound pressure levels—low, medium, and high—will affect people's preference for the sound environment. The effect of the visual density of the plaza crowd on evaluations of the sound environment is negatively correlated with the volume of human voices.

When the sound pressure level of human voices is the same, changes in the visual environment are more likely to affect people's evaluation of the overall sound environment. When the conversational sound is between low and medium, it does not affect people's preference and comfort level with the environment. When the sound of human conversation is between medium and high, the louder the sound, the lower the people's preference and comfort.

**Author Contributions:** Conceptualization, J.K., X.Y.; Methodology, X.Y., J.K.; Validation, X.Y.; Formal analysis, X.Y.; Investigation, X.Y.; Writing—original draft, X.Y.; Writing—review & editing, J.K.; Visualization, X.Y.; Project administration, J.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China [grant number 51778169], and also, the European Research Council (ERC) Advanced Grant (no. 55740696) on “Soundscape Indices” (SSID).

**Institutional Review Board Statement:** The work was approved by the HIT committee.

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

**Data Availability Statement:** Data could be obtained by contacting the authors.

**Acknowledgments:** The authors would like to express our sincere gratitude to every participant in the experiment.

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

## References

1. Berglund, B.; Lindvall, T.; Schwela, D.H. *Guidelines for Community Noise*; World Health Organization (WHO): Geneva, Switzerland, 1999.
2. Schulte-Forkamp, B. The Meaning of Annoyance in Relation to the Quality of Acoustic Environments. *Noise Health* **2002**, *4*, 13. [[PubMed](#)]
3. Schultz, T.J. *Community Noise Rating*; Applied Science Publishers: London, UK, 1982.
4. Schultz, T.J. Synthesis of social surveys on noise annoyance. *J. Acoust. Soc. Am.* **1978**, *64*, 377–405. [[CrossRef](#)] [[PubMed](#)]
5. Botteldooren, D.; Verkeyn, A. Fuzzy models for accumulation of reported community noise annoyance from combined sources. *J. Acoust. Soc. Am.* **2002**, *112*, 1496–1508. [[CrossRef](#)] [[PubMed](#)]
6. Job, R.F.S. Community response to noise: A review of factors influencing the relationship between noise exposure and reaction. *J. Acoust. Soc. Am.* **1988**, *83*, 991–1001. [[CrossRef](#)]
7. Dubois, D. Categories as acts of meaning: The case in olfaction and audition. *Cogn. Sci. Q* **2000**, *1*, 35–68.
8. Gaver, W.W. What in the World Do We Hear? An Ecological Approach to Auditory Event Perception. *Ecol. Psychol.* **1993**, *5*, 1–29. [[CrossRef](#)]

9. Dubois, D.; Guastavino, C.; Raimbault, M. A cognitive approach to urban soundscape: Using verbal data to access everyday life auditory categories. *Acta Acust. United Acust.* **2006**, *92*, 945–951.
10. Guastavino, C. The ideal urban soundscape: Investigating the sound quality of French cities. *Acta Acust. United Acust.* **2006**, *92*, 945–951.
11. Lavandier, C. The contribution of sound source characteristics in the assessment of urban soundscapes. *Acta Acust. United Acust.* **2006**, *92*, 912–921.
12. Westerkamp, H. Sound excursion: Plano pilato, Brasilia. *Soundscape: J. Acoust. Ecol.* **2000**, *1*, 20–21.
13. Robert, G. *Environmental Psychology: Principles and Practice*, 2nd ed.; Allyn and Bacon: Boston, MA, USA, 1997.
14. Bell, A.P.; Greene, C.T.; Fisher, D.J.; Baum, A. *Environmental Psychology*, 4th ed.; Harcourt Brace College Publishers: Fort Worth, TX, USA, 1996.
15. Pheasant, R.; Horoshenkov, K.; Watts, G.; Barrett, B. Acoust. and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet place? *J. Acoust. Soc. Am.* **2008**, *123*, 1446–1457. [[CrossRef](#)]
16. Axelsson, Ö.; Nilsson, M.E.; Hellstrom, B.; Lunden, P. A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park. *Landsc. Urban. Plan.* **2014**, *123*, 49–60. [[CrossRef](#)]
17. Liu, J.; Yang, L.; Xiong, Y.; Yang, Y. Effects of soundscape perception on visiting experience in a renovated historical block. *Build. Environ.* **2019**, *165*, 106375. [[CrossRef](#)]
18. Ren, X.; Kang, J.; Liu, X. Soundscape perception of urban recreational green space. *Landsc. Arch. Front.* **2016**, *4*, 42–55.
19. Andringa, T.C.; Lancer, J.J.L. How pleasant sounds promote and annoying sounds impede health: A cognitive approach. *Env. Res.* **2013**, *10*, 1439–1461. [[CrossRef](#)]
20. Van Kamp, I.; Van Kempen, E.; Swart, W.; Hanneke, K. Perceived soundscapes, human relation and related health in green urban areas. In Proceedings of the 22nd International Congress on Acoustics, Buenos Aires, Argentina, 5–9 September 2016.
21. Bernut, S. Soundscapes and tourism-towards sustainable tourism. *Probl. Sustain. Dev.* **2014**, *12*, 107–117.
22. Yang, W.; Kang, J. Acoustic comfort evaluation in urban open public spaces. *Appl. Acoust.* **2005**, *66*, 211–229. [[CrossRef](#)]
23. Bronkhorst, A.W. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acust. United Acust.* **2000**, *86*, 117–128.
24. Hong, J.Y.; Jeon, J.Y. Comparing associations among sound sources, human behaviors, and soundscapes on central business and commercial streets in Seoul, Korea. *Build. Environ.* **2020**, *186*, 107327. [[CrossRef](#)]
25. Jo, H.I.; Jeon, J.Y. The influence of human behavioural characteristics on soundscape perception in urban parks: Subjective and observational approaches. *J. Landsc Urban. Plan.* **2020**, *203*, 103890. [[CrossRef](#)]
26. Rock, I.; Harris, C.S. Vision and touch. *Sci. Am.* **1967**, *216*, 96–104. [[CrossRef](#)] [[PubMed](#)]
27. Joynt, J.L.; Kang, J. The influence of preconceptions on perceived sound reduction by environmental noise barriers. *Sci. Total. Environ.* **2010**, *408*, 4368–4375. [[CrossRef](#)] [[PubMed](#)]
28. 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](#)]
29. Southworth, M.F. The Sonic Environment of Cities. *Environ. Behav.* **1969**, *1*, 49–70. [[CrossRef](#)]
30. Ba, M.; Kang, J.; Li, Z. The effects of sounds and food odour on crowd behaviours in urban public open spaces. *Build. Environ.* **2020**, *182*, 107104. [[CrossRef](#)]
31. Schweiker, M.; Ampatzi, E.; Andargie, M.S.; Andersen, R.K.; Azar, E.; Barthelmes, V.M.; Berger, C.; Bourikas, L.; Carlucci, S.; Chinazzo, G.; et al. Review of multi-domain approaches to indoor environmental perception and behaviour. *Build. Environ.* **2020**, *176*, 106804. [[CrossRef](#)]
32. Ba, M.; Kang, J. A laboratory study of the sound-odour interaction in urban environments. *Build. Environ.* **2019**, *147*, 314–326. [[CrossRef](#)]
33. Anderson, L.M.; Mulligan, B.E.; Goodman, L.S.; Regen, H.Z. Effects of sounds on preferences for outdoor settings. *Environ. Behavior* **2020**, *176*, 106804.
34. Pick, H.L.; Siegel, G.M.; Fox, P.W.; Garber, S.R.; Kearney, J.K. Inhibiting the Lombard effect. *J. Acoust. Soc. Am.* **1989**, *85*, 894–900. [[CrossRef](#)]
35. Cooke, M.; King, S.; Garnier, M.; Aubanel, V. The listening talker: A review of human and algorithmic context-induced modifications of speech. *Comput. Speech Lang.* **2014**, *28*, 543–571. [[CrossRef](#)]
36. Marana, A.; Velastin, S.; Costa, L.; Lotufo, R. Automatic estimation of crowd density using texture. *Saf. Sci.* **1998**, *28*, 165–175. [[CrossRef](#)]
37. Hui, M.K.; Bateson, J.E.G. Perceived control and the effects of crowding and consumer choice on the service experience. *J. Consum. Res.* **1991**, *18*, 174. [[CrossRef](#)]
38. Westover, T.N. Perceived crowding in recreational settings: An environment-behavior model. *Environ. Behav.* **1981**, *21*, 258–276. [[CrossRef](#)]
39. Meng, Q.; Sun, Y.; Kang, J. Effect of temporary open-air markets on the sound environment and acoustic perception based on the crowd density characteristics. *Sci. Total. Environ.* **2017**, *601–602*, 1488–1495. [[CrossRef](#)]
40. Shu, H.; Song, Y.; Zhou, H. Assessment of Music and Water Sounds for Urban Noise Masking. In Proceedings of the TENCON 2018–2018 IEEE Region 10 Conference, Jeju, Republic of Korea, 28–31 October 2018; pp. 1451–1455. [[CrossRef](#)]
41. Meng, Q.; Kang, J. The influence of crowd density on the sound environment of commercial pedestrian streets. *Sci. Total. Environ.* **2015**, *511*, 249–258. [[CrossRef](#)]
42. Yu, L.; Kang, J. Factors influencing the sound preference in urban open spaces. *Appl. Acoust.* **2010**, *71*, 622–633. [[CrossRef](#)]



43. Arriaza, M.; Cañas-Ortega, J.; Cañas-Madueño, J.; Ruiz-Aviles, P. Assessing the visual quality of rural landscapes. *Landsc. Urban. Plan.* **2004**, *69*, 115–125. [[CrossRef](#)]
44. Benfield, J.A.; Bell, P.A.; Troup, L.J.; Soderstrom, N.C. Aesthetic and affective effects of vocal and traffic noise on natural landscape assessment. *J. Environ. Psychol.* **2010**, *30*, 103–111. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.