Relationship between street scale and subjective assessment of audio-visual environment comfort based on 3D virtual reality and dual-channel acoustic tests

Fangfang LIU[a,b], **Jian KANG**[b,a]*

[a] *Heilongjiang Cold Region Architectural Science Key Laboratory, School of Architecture, Harbin Institute of Technology, 66 West Dazhi Street, Nan Gang District, Harbin 150006, PR China*

[b] *School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, UK*

Abstract: We examine the influence of street scales (the street width, building height, and street-width-to-building-height ratio, referred to as 'width-to-height ratio' in the paper) on visual, acoustic, and audio-visual comfort evaluation (as evaluated by a set of participants) in urban areas. In addition, we examine the relationships between the sound level and the abovementioned subjective comfort evaluation except the visual one. After measuring the street scales and recording the street visual information with a 3D camcorder, the virtual 3D models of the streets were generated. Meanwhile, dual-channel acoustic signals of the streets were collected. Subsequently, subjective tests were carried out using a 3D virtual reality with corresponding sounds using 164 participants. The analysis shows that subjective attitudes are directly related to the street scales. In particular, there is a strong positive correlation between audio-visual comfort and the street width-to-height ratio. In contrast, the three indicators (visual, acoustic, and audio-visual comfort) are strongly negatively correlated to the height, and this type of negative correlation is also observed between subjective indicators (except the visual one) and the sound level. Overall, the respondents found the audio-visual level most comfortable when the street width-to-height ratio is greater than 1, street width is within 20 m, height of street buildings is less than 26 m, and the sound level is less than 58 dBA. It is expected that these findings can aid designers in predicting the ideal audio-visual environment quality for urban streets.

Keywords: Visual comfort, Acoustic comfort, Audio-visual comfort, Street scale

2018 Building and Environment

Date Received 10 September 2017 **Date Accepted:** 27 November 2017

Publish online: 05 December 2017

* Corresponding author. School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

E-mail addresses: j.kang@sheffield.ac.uk, j.kang@hit.edu.cn (J. Kang).

1 Introduction

According to data from the United Nations (UN), 396 million people will be added to the cities of East Asia and Southeast Asia from 2015 to 2040 [1, 2]. This increase in the urban population requires a drastic expansion of the urban living and transportation space. The rapid development of large-scale buildings in new urban areas can bring in changes in urban structures which are quite different from those in the old towns because of many new functional constructions, heavy transportation, and multiply scaled streets and buildings, which can subsequently change the overall perceptive audio-visual comfort of a resident or pedestrian. Whether the newer urban space (for example, its street and square) can meet the demand of satisfactory audio-visual perception forms a key issue for exploration, and understanding the correlation between the width and height of a street and the indicators of visual comfort, acoustic comfort, and audio-visual comfort would be beneficial for designing liveable cities.

Sound propagation in a street was first measured in 1965 by Wiener et al. for detecting the noise level and reverberation [3]. Schröder et al. observed a change in reverberation along the length of the street [4]. Kang established an acoustic model by using the energy virtual source method for enclosed streets and enclosed squares with geometric mirror reflection boundary surfaces, and they proposed to increase the use of audio-visual interactions to simplify the simulation process [5].

Sounds in an environment are normally evaluated by analysing the acoustic comfort or annoyance in the subjective evaluation of a soundscape [6-8]. Annoyance is generally associated with acoustic comfort. The concept of annoyance is defined as ‘a feeling of displeasure associated with any agent or condition, known or believed by an individual or group to adversely affect them’, and it is clearly a negative environment situation [9]. In this backdrop, Guski et al. explored the concept of ‘noise annoyance’, concluding that noise annoyance can be regarded as a multifaceted psychological concept addressing the immediate behavioural (disturbance and interfering with intended activities) and evaluative aspects (nuisance, unpleasantness, and getting on one's nerves) [10]. The relationship between exposure to noise and noise annoyance has been studied *via* an integrated meta-analysis, and the relationships between noise level and noise annoyance have been further investigated [11, 12]. Background noise levels were found to be an important index in evaluating the urban soundscape in open public spaces. For example, low background levels can allow a person to feel quieter and more peaceful [6]. Furthermore, the subjective evaluation of a soundscape is not limited to the study of acoustic comfort and noise annoyance but also the pluralistic aspects of acoustic perception, acoustic memory, sound sentiment, and aesthetics [13].

The audio-visual environment has also been considered as a total environment to be studied. For example, Hong and Jeon have suggested that audio-visual interactions may affect the environmental quality, and as a result, these interactions should be considered in urban planning [14, 15]. Some researchers have determined that an individual experiences the surrounding environment as a whole, through all the sensory forms at the same time [16]. The sum of these inputs produces physiological and psychological effects which lead to the feeling of happiness or distress. The complex interplay between the auditory and visual senses has also been researched in neuroscience [17, 18]. Total environmental consciousness involves the combination of all the senses, and a reasonable assessment of an environment should be based on an overall evaluation of multiple parameters. For example, Preis et al. found that there could be many different interactions between feeling, hearing, and vision, and as a result, the subjective evaluation of the urban environment should be included [19]. Here, we note that interactions can also be multisensory. Other studies show that sensory stimulation can transfer from one individual to another [20-22].

Although the relationship between scale and sound propagation has been studied, the effect of the 3D scale of a street on visual, acoustic, and audio-visual comfort levels has not thus far been systematically explored. In addition, acoustic and visual factors are always interconnected and not independently controlled in the real world, and each street has its unique audio-visual environment. Due to inherent limitations, main effect of morphological factors of urban street on audio/visual perceptions cannot be calculated. The aim of this work is to examine the influence of the street width (*W*), building height (*H*), and *W/H* on visual perception, acoustic perception, and audio-visual comfort evaluation in urban areas. The observations and findings in this paper are based on 164 individuals who participated in audio-visual tests to evaluate the visual, acoustic, and audio-visual comfort levels using a 5-point scale, *via* the application of 3D dynamic virtual simulations and dual-channel acoustic tests for a field acoustic environment of 10 streets. Significant correlations between the scales of *W/H*, *W*, *H* and visual, acoustic, and audio-visual comfort were observed. In addition, the relationship between sound level (SL) and the abovementioned subjective indicators except the visual one is also discussed.

2. Methodology

The study methods included the selection of survey sites, 3D simulation, questionnaire survey, and analysis of acquired statistical data.

2.1 The selection of survey site

In this study, the field survey includes measuring the street and building scales and recording the sound environment of the sample. Our survey was carried out in Type 2-Environmental Noise Function Regions in Harbin, China, which refer to the areas with maximum SL of 60 dBA during the daytime (6 a.m.-10 p.m.) and 50 dBA at night (10 p.m.-6 a.m.), respectively [23]. These areas, including the Haxi new town and Daowai old town (Figure 1), are major commercial and residential areas in the city where a significant number of residents live, and they form the most common areas in the urban environment. By measuring *W* and *H* of the buildings in three dimensions with a laser range finder (Trueyard SP1500), we collected data for 10 differently scaled streets which were not influenced by noise from the railway and factories in the city. Usually, the buildings comprise two or three floors and the streets comprise two to four lanes in the old town. In comparison, in the new town, many buildings have 18 to 32 floors, while the streets follow the 8-, 10-, and 12-lane models, which is representative of a common characteristic of the streets in the new town.

Figure 1. New- and old-town locations in Harbin chosen for the study.

Next, the acoustic signal data for the 10 streets were collected over dual channels using an acoustic signal data collector (ZODIAC/DIC10). The acoustic signals were recorded at a height of 1.5 m above the ground on pedestrian roads at a distance of >50 m away from the intersection in order to avoid noise interference from pedestrian/bike crossings to obtain high-fidelity stereo recordings [5]. We recorded the acoustic signals on workdays in winter (when no leaves were on the trees) in order to avoid sound attenuation by the leaves. The traffic flow and speed are shown in Table 1.

Table 1. Traffic data.

Street No.	Traffic flow/h^{-1}	Heavy veh./%	L_{eq}/dBA	Lane	Road width/m	The speed of the car/km h^{-1}
1	102	10	56.13	2	15	40
2	91	12	52.29	2	15	40
3	99	12	54.57	2	10	40
4	101	11	55.86	2	9	40
5	262	16	58.96	2	7	40
6	308	10	58.79	10	37	50
7	404	15	63.87	6	30	50
8	379	13	63.31	6	28	50
9	397	11	63.13	10	45	50
10	401	12	64.56	10	46	50

2.2 3D simulation

Street images recorded by a 3D camera/camcorder are realistic; however, visual information such as color, architectural style, and degree of congestion in images would interfere with the responses of the subjective attitude towards street scales. In addition, certain research results show that the simulation of a 3D landscape with the coupling of the corresponding sound can deliver a more accurate and life-like experience [24]. Therefore, the visuals of the 10 streets were recorded with a 3D camcorder (Panasonic HDC-Z10000), and modelled in Unity 3D using the software CadnaA. The 3D visual roaming model was constructed by considering the combination of the height of the buildings and the width of the streets while the street styles and other visual information are omitted, as shown in Figure 2. A height of about 1.5 m from the ground is the typical 'eye height' of the landscape [25]. The camcorder of the virtual 3D model of the street sets is based on this view height. Furthermore, the integration of dynamic vision and sound provides a realistic sense of presence in the environment for the participant, and thus provokes responses and behaviours similar to those that would occur in the real environment.

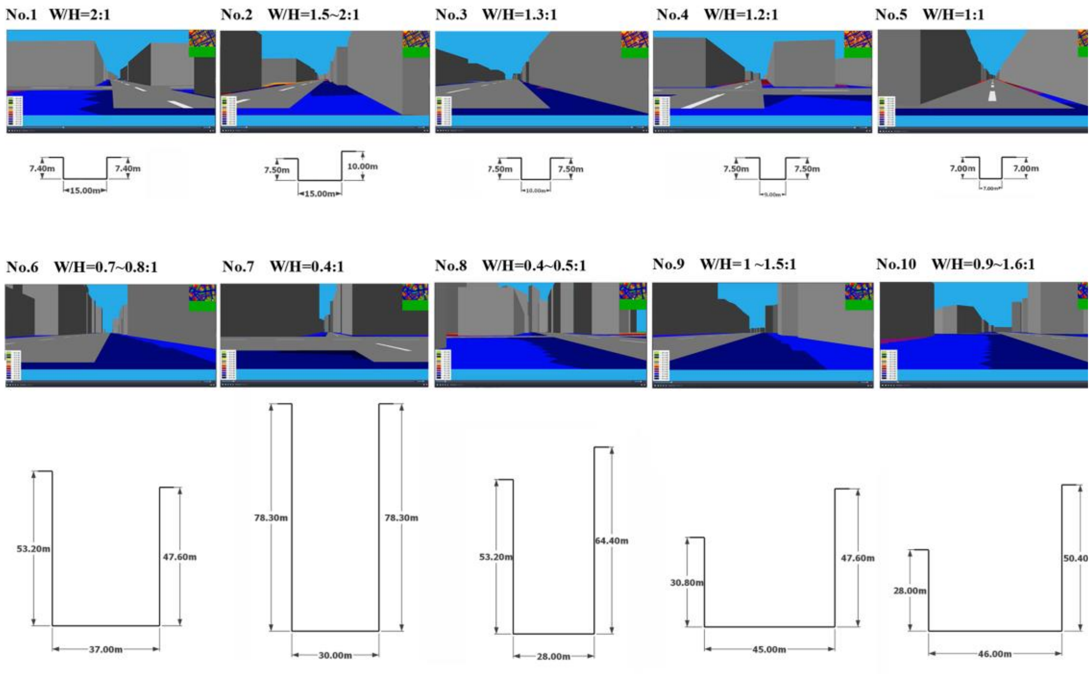


Figure 2. Scale measurements and virtual 3D models of the 10 selected streets.

2.3 Questionnaire survey

Previous studies have shown that in the study of perception on audio-visual stimuli in a controlled laboratory setting, a subject sample size of around 20 is often used [26, 27]. In this study, we increase the sample size to a total number of 164 to cover different ages while keeping the balance of the male-to-female ratio for participants. Based on the selection criteria of previous studies [28-30], young adults with normal hearing and regular or corrected-to-normal vision were selected as the study subjects. Participants were between 20 and 55 years old, and the male-to-female ratio was set to 1.05:1 (male 84, female 80) to ensure that the sample sex ratio showed balance.

This study used the questionnaire survey method to measure visual comfort, acoustic comfort, and audio-visual comfort, which are three important evaluation indicators. The attitudes of respondents were measured using Likert scale that has been widely used in survey research of environmental effects on subjective comfort, although ICBEN scale is applicable and has been utilized in previous studies [31-33]. The parameters of visual, acoustic, and audio-visual comfort were graded as per the following linear scale: 1-very uncomfortable, 2-uncomfortable, 3-neither comfortable nor uncomfortable, 4-comfortable, and 5-very comfortable. In the questionnaire, the names and locations of the 10 streets in question were not disclosed. The acoustic tests were conducted using the *in situ* sound level without any equalization, and the street soundtracks of the two SLs (The highest mean SL value is 64.6 dBA at S10, while the lowest is 52.3 dBA at S2, 'S' indicates the street.) were played to the participants through a BHS II headset connected to the acoustic signal data collector (ZODIAC/DIC10) for 30 seconds, which allowed their ears to adapt the environment to avoid misjudgement. Then, soundtracks for 10 differently scaled streets were played randomly to the participants, during which they were allowed to answer questions related to the acoustic comfort in the questionnaire. Next, the stereo playback was stopped, followed by turning on the PC monitor (Lenovo IdeaCentre B520) that repeated playing the street video corresponding to the lately played soundtrack mutely, during which the participants answered the questions related to the visual comfort in the questionnaire. Finally, the stereo playback was turned on, and the participants can watch video and at the same time hear the stereo of the same street. Meanwhile, they answered the questions related to the audio-visual comfort in the questionnaire.

2.4 Data statistics and analysis

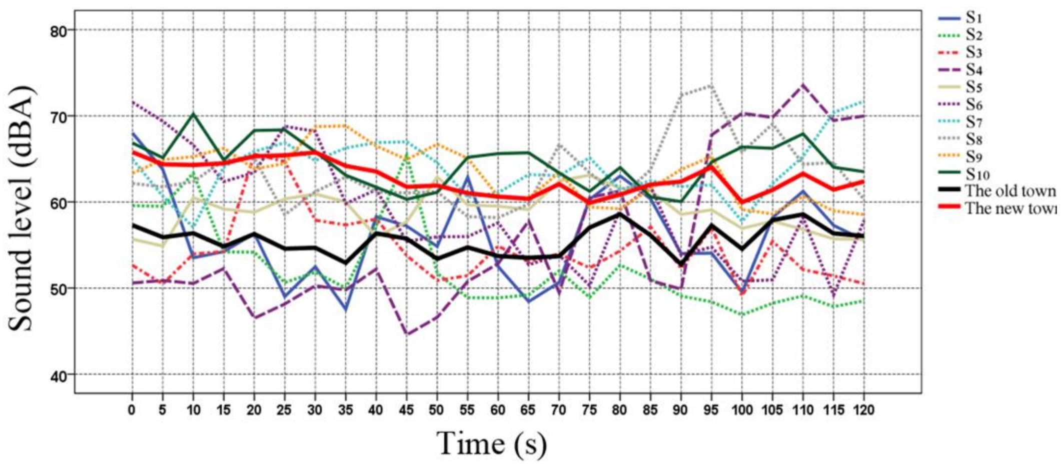


Figure 3. The sound levels (SLs) in the selected streets.

The data from the questionnaire of the 164 participants were collected. After calculation, the reliability coefficient of the questionnaire was estimated as 0.848 (Cronbach's alpha). A reliability coefficient 0.9 > $\alpha \geq 0.8$ indicates that the questionnaire satisfies the reliability requirement [34].

The analysis of A-weighted SL, which is a commonly suggested metric in the evaluation of an aural environment [35], was performed using the Artemis-software-based acoustic signal data collector to generate binaural recordings to create a realistic 3D sound. All the sounds were recorded in the WAV format with a sampling frequency of 44,100 Hz. The observation points and the SLs are illustrated in Figure 3, wherein ‘S’ indicates the street. The figure shows the SL distribution in each street, and we note that the maximum SL is about 73 dBA, while the minimum SL is 44.5 dBA. Mean SL values for old and new streets were plotted in Figure 3, and the difference is about 10 dBA.

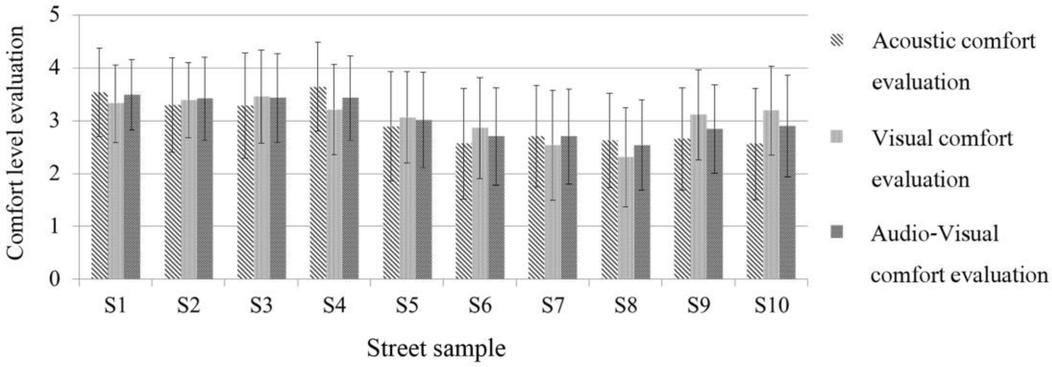


Figure 4. Mean acoustic, visual, and audio-visual comfort levels along with standard deviations.

Table 2. Street scales for the 10 street configurations.

Configu rations	**S_1**	**S_2**	**S_3**	**S_4**	**S_5**	**S_6**	**S_7**	**S_8**	**S_9**	**S_{10}**
W/H_{min}	2.03	2.00	1.33	1.20	1.00	0.78	0.38	0.53	1.46	1.64
W/H_{ave}	2.03	1.71	1.33	1.20	1.00	0.73	0.38	0.48	1.15	1.17
W/H_{max}	2.03	1.50	1.33	1.20	1.00	0.70	0.38	0.43	0.95	0.91
W (m)	15.00	15.00	10.00	9.00	7.00	37.00	30.00	28.00	45.00	46.00
H_{min} (m)	7.40	7.50	7.50	7.50	7.00	47.60	78.30	53.20	30.80	28.00
H_{ave} (m)	7.40	8.75	7.50	7.50	7.00	50.40	78.30	58.80	39.20	39.20
H_{max} (m)	7.40	10.00	7.50	7.50	7.00	53.20	78.30	64.40	47.60	50.40

Figure 4 depicts the mean of the visual comfort, acoustic comfort, and audio-visual comfort, which are the three important evaluation indicators. The means of the subjective evaluation do not oscillate significantly; and they vary between 2 and 4 with the subjective evaluation of visual comfort level, acoustic comfort level, and audio-visual comfort level. The highest and lowest mean values differ by nearly one point, as shown in the figure.

The *W*/*H* ratios of the streets are also shown in Figure 2. We note from the figure that the streets are not equal in terms of height in reality. Therefore, a given street is set to have three different *W/H* ratios *via* choosing the height from different sides of a building. In this context, we considered three situations: (1) *W/H* is calculated based on the short side (at the low end), denoted as W/H_{min}. (2) *W/H* is calculated based on the long side (at the high end), denoted as W/H_{max}. (3) *W/H* is calculated based on the average building height for both sides of the street, denoted as W/H_{ave}. The 3D scales for all 10 street configurations are listed in Table 2.

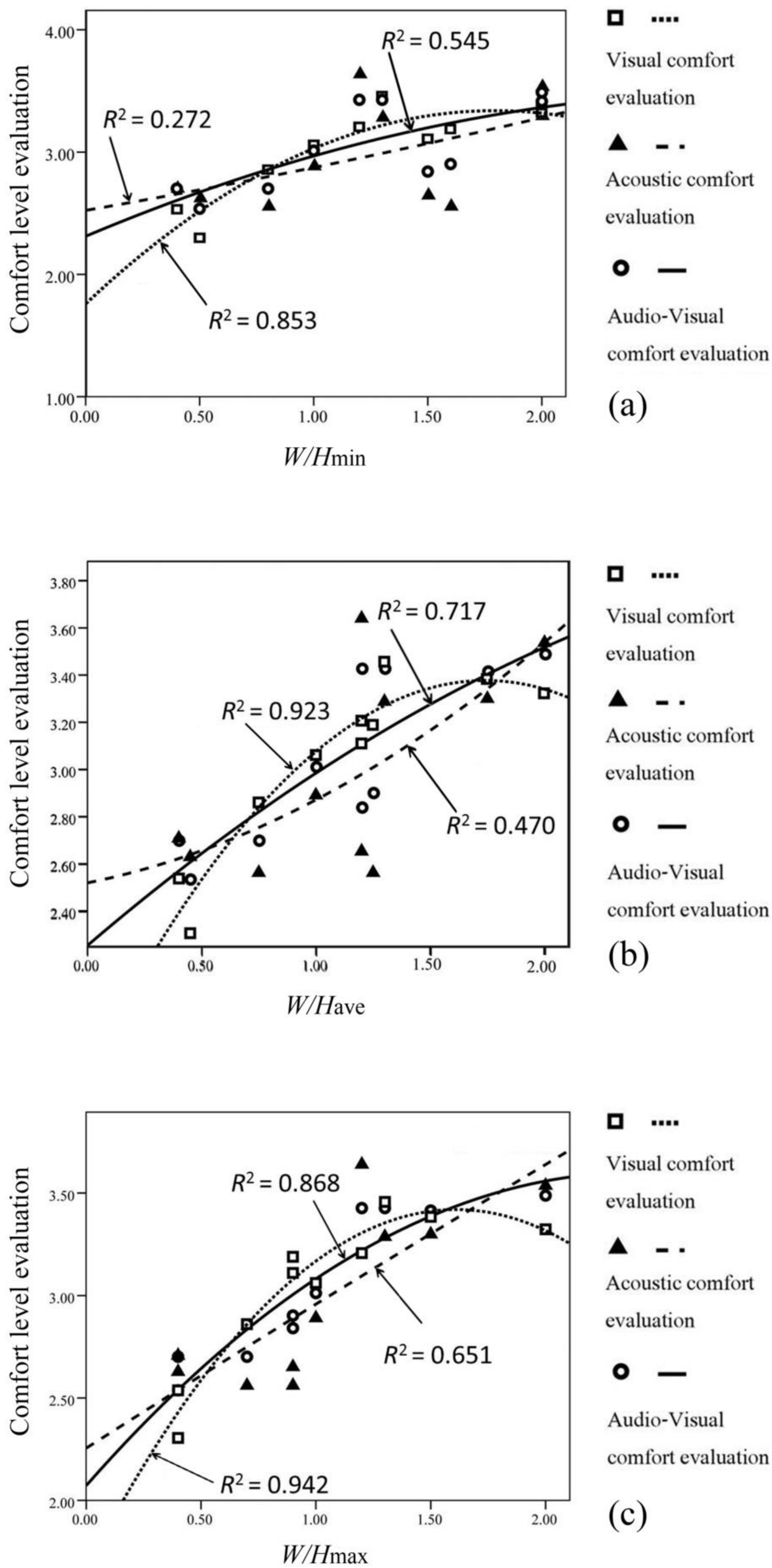
3. Results and analysis

Multivariate analysis of variance (MANOVA) was used to evaluate the effect of statistically significant mean difference in perceptual factors (visual, acoustic, and audio-visual) in terms of street scales of each configuration as well as SL. As shown in Table 3, significant mean differences are observed in visual, acoustic, and audio-visual comfort in terms of W/H_{min}, W/H_{ave}, W/H_{max}, H_{min}, H_{ave} , H_{max}, and *W*, as well as in acoustic and audio-visual comfort in term of SL (p <0.05 for all, Table 3).

Table 3. MANOVA results for street configurations and subjective comfort assessment.

Source	SS	df	MS	*F*	Sig.	η_p^2
W/H_{min}						
Visual comfort	205.737	9	22.860	30.093	**0.000**	0.142
Acoustic comfort	263.371	9	29.263	32.127	**0.000**	0.151
Audio-visual comfort	193.844	9	21.538	29.883	**0.000**	0.142
W/H_{ave}						
Visual comfort	205.737	9	22.860	30.093	**0.000**	0.142
Acoustic comfort	263.371	9	29.263	32.127	**0.000**	0.151
Audio-visual comfort	193.844	9	21.538	29.883	**0.000**	0.142
W/H_{max}						
Visual comfort	205.737	9	22.860	30.093	**0.000**	0.142
Acoustic comfort	263.371	9	29.263	32.127	**0.000**	0.151
Audio-visual comfort	193.844	9	21.538	29.883	**0.000**	0.142
H_{min}						
Visual comfort	200.318	7	28.617	37.554	**0.000**	0.139
Acoustic comfort	250.152	7	35.736	38.935	**0.000**	0.143
Audio-visual comfort	193.828	7	27.690	38.465	**0.000**	0.142
H_{ave}						
Visual comfort	200.097	7	28.585	37.506	**0.000**	0.139
Acoustic comfort	252.429	7	36.061	39.349	**0.000**	0.144
Audio-visual comfort	193.539	7	27.648	38.398	**0.000**	0.141
H_{max}						
Visual comfort	200.612	8	25.077	32.896	**0.000**	0.139
Acoustic comfort	253.115	8	31.639	34.519	**0.000**	0.145
Audio-visual comfort	193.844	8	24.230	33.639	**0.000**	0.142
W						
Visual comfort	205.432	8	25.679	33.817	**0.000**	0.142
Acoustic comfort	258.734	8	32.342	35.418	**0.000**	0.148
Audio-visual comfort	193.405	8	24.176	33.550	**0.000**	0.141
SL						
Acoustic comfort	263.371	9	29.263	32.127	**0.000**	0.151
Audio-visual comfort	193.844	9	21.538	29.883	**0.000**	0.142

SS = Type III Sum of Squares; df = degrees of freedom; MS = Mean Square; *F* = F ratio; Sig. = significance; $\eta_p{}^2$= partial eta squared (effect size). Significance (at 0.05) is in **bold**.



Comfort level evaluation
$R^2 = 0.545$
$R^2 = 0.272$
$R^2 = 0.853$
W/Hmin
(a)
$R^2 = 0.717$
$R^2 = 0.923$
$R^2 = 0.470$
W/Have
(b)
$R^2 = 0.868$
$R^2 = 0.651$
$R^2 = 0.942$
W/Hmax
(c)
Visual comfort evaluation
Acoustic comfort evaluation
Audio-Visual comfort evaluation

Figure 5.Relationship between W/H_{min} and comfort-level evaluation (a); Relationship between W/H_{ave} and comfort-level evaluation (b); Relationship between W/H_{max} and comfort-level evaluation (c).

We analysed the various relationships among the variables in question under the following classifications:

3.1 Effect of W/H on comfort evaluation

To investigate the relationship between W/H and comfort-level evaluation, the regression analyses with using mean values of subjective assessment were performed. These regression curves demonstrate similar trends to their corresponding ones analyzed based on all subjective data, and significant p values are less than 0.01 in all analyses, although R^2 values are substantially small in the latter. In addition, regression analyses based on mean values have been used in many previous studies [5, 6]. Figure 5 illustrates the relationships between W/H (W/H_{min}, W/H_{max}, W/H_{ave}) and the subjective evaluation of the visual comfort level, acoustic comfort level, and audio-visual comfort level, and each symbol represents the average of the subjective evaluation of the visual, acoustic, and audio-visual comfort levels for one specific value of W/H. Both linear and quadratic regression were performed to analyze the relationship between W/H and the subjective comfort-level, and the results indicate that coefficients of determination R^2 from quadratic regression are closer to 1 than those from linear regression. Furthermore, quadratic regression has been widely used in analyzing the relationship between environmental factors and subjective comfort assessment [5, 6]. As a result, quadratic regression was adapted in this study.

As for visual comfort, a W/H value > 0.9 corresponds to a 'positive' participant attitude (that is, the comfort evaluation is greater than 3 in the scale defined previously). This value corresponds to a critical point between positive and 'negative' evaluation. Moreover, a significant correlation was generally observed between W/H_{ave} and the subjective visual comfort level, with an R^2 value of 0.923 (Figure 5b), thereby indicating that the W/H_{ave} variation accounts for 92.3% of the variability in the subjective visual-comfort level. In addition, W/H_{min} variation has an R^2 value of 0.853 (Figure 5a), which is similar to that of W/H_{max}, with an R^2 value of 0.942 (Figure 5c). The Spearman's correlation coefficients r and p have been used to find significant correlation between the preference for integration of wind park and aural annoyance [28]. From Table 4, we note that there is a significant correlation between visual comfort and W/H_{ave} ($r = 0.918$, $p < 0.01$). Correlation is significant at the 0.01 level (two-tailed). This result is similar to the visual assessment results for W/H_{min} ($r = 0.802$, $p < 0.01$) and W/H_{max} ($r = 0.909$, $p < 0.01$), as listed in Table 4. The result is higher for this correlation, but it is not the only factor affecting the visual comfort. We observed that the living environment has a certain effect on the audio-visual evaluation of the participant: the evaluation of visual comfort for streets with high-rise buildings is high by participants coming from Shenzhen, China (a city with many high-rise buildings).

Table 4. Spearman's correlation coefficient between street scale and acoustic, visual, and audio-visual comfort levels, including the two-tailed significance levels. Significant correlations are marked with * ($p < 0.05$) and ** ($p < 0.01$).

Configurations	Visual comfort evaluation	Acoustic comfort evaluation	Audio-visual comfort evaluation
W/H_{min}	*r*: 0.802** *p*: 0.005	*r*: 0.372 *p*: 0.290	*r*: 0.691* *p*: 0.027
W/H_{ave}	*r*: 0.918** *p*: 0.000	*r*: 0.518 *p*: 0.125	*r*: 0.835** *p*: 0.003
W/H_{max}	*r*: 0.909** *p*: 0.000	*r*: 0.768** *p*: 0.010	*r*: 0.948** *p*: 0.000
W	*r:* -0.353 *p*: 0.318	*r*: -0.768** *p*: 0.009	*r*: -0.587 *p*: 0.074
H_{min}	*r*: -0.669* *p*: 0.034	*r*: -0.646* *p*: 0.044	r: -0.846** p: 0.002
H_{ave}	*r*: -0.652* *p*: 0.041	*r*: -0.664* p: 0.036	*r*: -0.853** *p*: 0.002
H_{max}	*r*: -0.909** *p*: 0.000	*r*: -0.677* *p*: 0.032	*r*: -0.844** *p*: 0.002

The acoustic comfort also depends on W/H. As can be observed from Figure 5b, participants show positive attitudes when W/H_{ave} is greater than 1.2, and a correlation was generally observed with an R^2 value of 0.470, thereby indicating that the W/H_{ave} variation accounts for 47.0% of the variability in the subjective acoustic comfort level. In addition, the corresponding R^2 values for W/H_{min}, W/H_{ave} and W/H_{max} are 0.272, 0.470 and 0.651, respectively (Figure 5). There was no significant correlation between acoustic comfort and W/H_{ave}, as can be inferred from Table 4 (r = 0. 518, $p > 0.05$). In addition, the influence of W/H_{min} on the sound comfort is not significant ($r = 0.372$, $p > 0.05$). However, a significant correlation between acoustic comfort and W/H_{max} was found to exist, as can be inferred from Table 4 ($r = 0.768$, $p = 0.01$). Correlation is significant at the 0.01 level (two-tailed). It is to be noted that the ratio W/H is calculated based on the long side (at the high end), and a strong correlation can be found between acoustic comfort and W/H_{max}.

In terms of the audio-visual comfort, a ratio of $W/H > 1$ corresponds to positive participant attitude, as can be observed in Figure 5b. A correlation was also observed between W/H_{ave} and the audio-visual comfort level, with an R^2 value of 0.717. The corresponding value R^2 values for W/H_{min} and W/H_{max} are 0.545 and 0.868 in Figures 5a and 5c, respectively. As can be inferred from Table 4 ($r = 0.835$, $p < 0.01$), similar results for W/H_{min} ($r = 0.691$, $p < 0.05$) and W/H_{max} ($r = 0.948$, $p < 0.01$) were obtained. Thus, there exists a significant positive correlation between audio-visual comfort and W/H.

It is noted that the tendencies of the three regression curves are different in Figure 5b. While the regression is nearly linear for acoustic comfort, it is parabolic for both visual and audio-visual comfort. From Figures 5a and 5c, similar trends are observed associated with W/H_{min} and W/H_{max} as well. Specifically, coefficients of determination R^2 derived from linear and quadratic regression have the same value of 0.651 for acoustic comfort in Figure 5c, suggesting that attitudes of respondents towards the acoustic comfort could be straightened with the change the street scales. Moreover, the dependence of subjective comfort on W/H_{ave} as well as W/H_{max} shows both large R^2 (Figure 5) and small Spearman's correlation coefficients *p* (Table 4), suggesting that attitudes of respondents were consistently influenced by these scales relative to W/H_{min}. As shown in Figure 5a, subjective evaluations are very scattered and fall apart from the regression line when values of W/H_{min} are between 1.33 and 2.03. One possible explanation could be that individual's attention was prone to be attracted by buildings on the high end of a wide street.

3.2 Effect of *W* on comfort evaluation

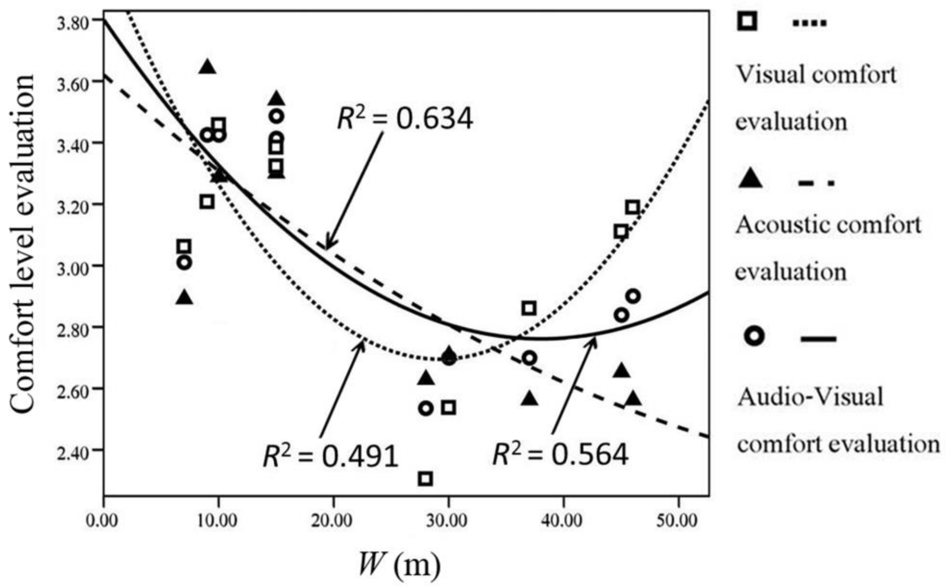


Figure 6. Relationship between *W* and comfort-level evaluation.

When *W* is very large, it can give rise to comfort change in the audio-visual environment in addition to inconveniences to pedestrians, particularly the elderly and children. Figure 6 shows the relationships between *W* and the comfort evaluation with the corresponding quadratic regressions and correlation coefficients R^2. In Figure 6, each symbol represents the average of the subjective evaluation of the visual, acoustic, and audio-visual comfort for a specific value *W*. With an increase in *W*, the mean evaluation score decreases; however, it is also interesting to note that when *W* approaches 30 m, visual comfort begins to increase. When *W* is ~38 m, the audio-visual evaluation level also begins to rise, thus suggesting that

the visual and audio-visual assessments are parabolic. However, acoustic comfort assessment is fairly linear; with an increase in *W*, the acoustic evaluation score always decreases.
As for visual comfort, a correlation was observed between *W* and the subjective visual comfort level, with an R^2 value of 0.491. This indicates that *W* variation accounts for 49.1% of the variability in the subjective visual comfort level. This value is not very large. Further, no significant correlation between visual comfort and *W* was found, as can be inferred from Table 4 (r = -0.353, $p > 0.05$). A possible explanation is that the correlation is related to the participant living environment and social background. For instance, certain participants living in rural areas said they preferred more spacious streets, but did not feel comfortable in crowded places. In addition, for participants to exhibit positive attitudes regarding visual comfort, *W* should be less than 15 m or greater than 43 m, as shown in Figure 6.

A correlation was generally observed between *W* and the acoustic comfort level with an R^2 value of 0.634 (Figure 6), indicating that the *W* variation accounts for 63.4% of the variability in the subjective acoustic comfort level. A significant negative correlation was observed between acoustic comfort and *W*, as can be inferred from Table 4 (r = -0.768, $p < 0.01$). As expected, this result is consistent with the fact that a wider road leads to the movement of more vehicles, leading in turn to more noise, which lowers the sound comfort level. Therefore, from a sound-comfort perspective, Figure 6 indicates that people prefer street widths of < 22 m.

Nevertheless, from the audio-visual comfort point of view, *W* should be limited to within 20 m. In this regard, a correlation between *W* and the audio-visual comfort was generally observed with an R^2 value of 0.564. This indicates that *W* variation accounts for 56.4% of variability in the subjective audio-visual comfort level evaluation. No significant correlation between audio-visual comfort and *W* was found, as can be inferred from Table 4 (r = -0.587, $p > 0.05$).

3.3 Effect of *H* on comfort evaluation

Figure 7 depicts the relationships between *H* and the comfort level evaluation with the corresponding quadratic regressions and correlation coefficients R^2. In Figure 7b, the comfort level of 3 forms the critical point between positive and negative participant attitudes. The participants preferred the visual, acoustic, and audio-visual comfort levels corresponding to *H* < 40 m, 20 m, and 26 m, respectively. From Figure 7, we note that when *H* is ~56 m and ~70 m, the acoustic and audio-visual evaluation levels are the lowest. However, visual evaluations always exhibit a decreasing trend. We could find a correlation between H_{ave} and the subjective audio-visual comfort level, with an R^2 value of 0.841. It is also interesting to note that all three indicators exhibit similar trends (Figures 7a and 7c).

Here, it is noteworthy that a significant negative correlation exists between visual comfort and H_{ave} (r = -0.652, $p < 0.05$), between audio-visual comfort and H_{ave} (r = -0.853, $p < 0.01$), and between acoustic comfort and H_{ave}, (r = -0.664, $p < 0.05$), as can be inferred from Table 4. Similar relationships were observed to exist between H_{min} and the three indicators and H_{max}

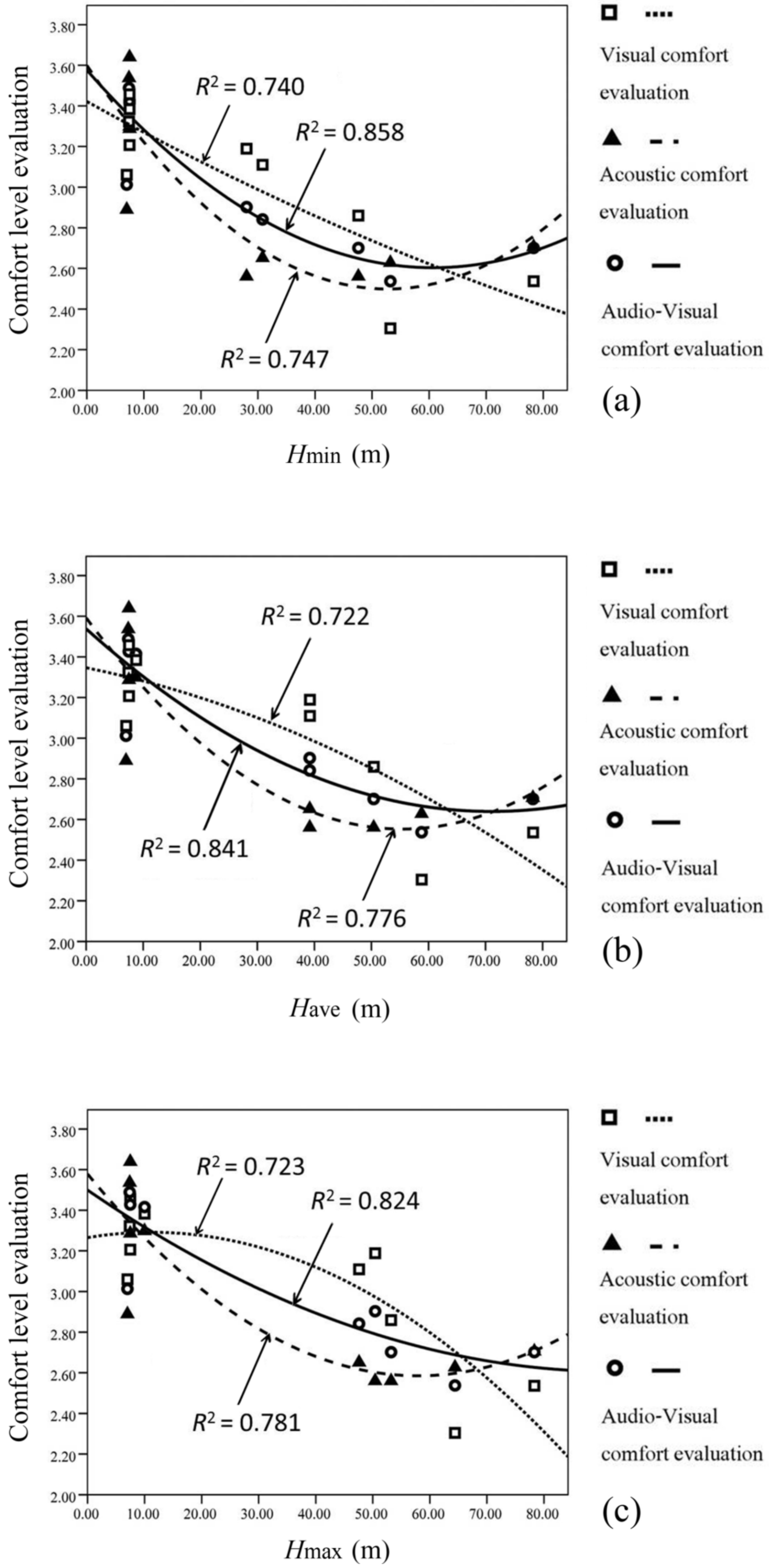

Comfort level evaluation
R^2 = 0.740
R^2 = 0.858
R^2 = 0.747
Visual comfort evaluation
Acoustic comfort evaluation
Audio-Visual comfort evaluation
(a)
Hmin (m)
R^2 = 0.722
R^2 = 0.841
R^2 = 0.776
(b)
Have (m)
R^2 = 0.723
R^2 = 0.824
R^2 = 0.781
(c)
Hmax (m)

Figure 7. Relationship between H_{min} and comfort-level evaluation (a); Relationship between H_{ave} and comfort-level evaluation (b); Relationship between H_{max} and comfort-level evaluation (c).

and the three indicators, as can be inferred from Table 4. Therefore, *H* forms an important factor in comfort evaluation.

3.4 Effect of SL

Although the results presented thus far indicate a correlation between *W*, *H*, and the indicators of acoustic comfort, and audio-visual comfort, the SL is an important factor which cannot be ignored [5, 11]. Figure 8 shows the relationships between the measured LA_{eq} values and the comfort level evaluation, with the corresponding quadratic regressions and the correlation coefficients R^2. In Figure 8, each symbol represents the average of the comfort level evaluation for a specific value of LA_{eq}. With an increase in LA_{eq}, the mean evaluation score of the two abovementioned indicators decreases.

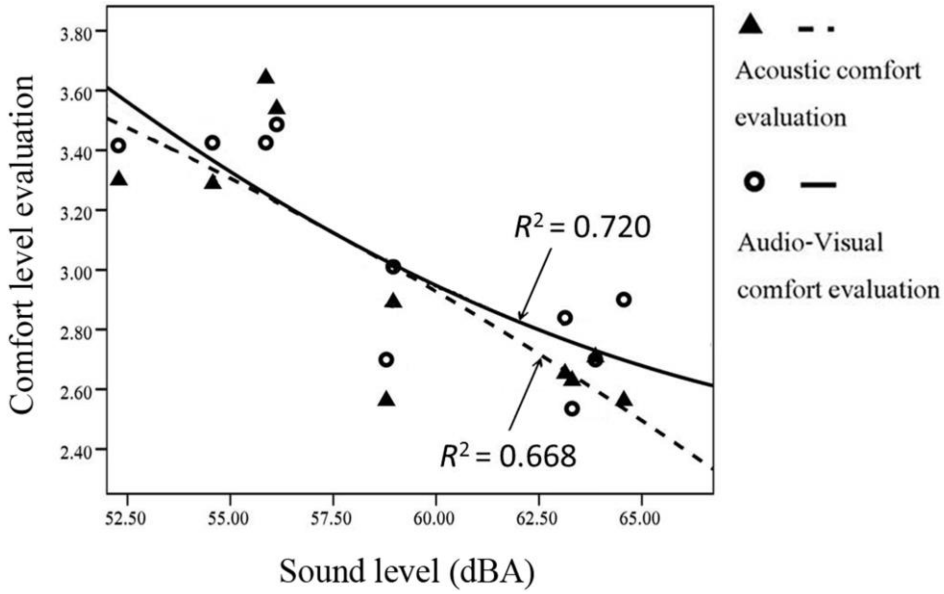


Figure 8. Relationship between SL and comfort-level evaluation.

As regards acoustic comfort, we observed a significant negative correlation between acoustic comfort and LA_{eq} (Table 5, $r = -0.717$, $p < 0.05$), which is consistent with the previous study [5]. However, LA_{eq} variation accounts for only 66.8% of variability in the subjective acoustic comfort, with an R^2 value of 0.668, thereby indicating the presence of other possible factors of influence. For example, the adaptability of the environment and lifestyle habits could possibly affect the respondents, for e.g. people with driving habits show a high tolerance for high traffic noise while people who do not drive are sensitive to traffic noise and find themselves very uncomfortable with high noise levels in the street. Living environments may have also affected the respondent responses: a respondent living near a highway is not likely very sensitive to high decibel levels, and can show more tolerance or even rate the acoustic comfort as positive.

Furthermore, SL should be limited to within 58 dBA to obtain positive attitude in terms of audio-visual comfort from participants, as shown in Figure 8. The influence of SL on audio-visual comfort for the various cases discussed in the study can be inferred from the entries in Table 5, where a significant negative correlation between audio-visual comfort and LA_{eq} can be observed (r = -0.683, $p < 0.05$). This result is consistent with that of a previous study indicating that the pleasantness of the environment can increase with reduction in the traffic noise level [7]. However, an R^2 value of 0.720 suggests that LA_{eq} may not be the only factor affecting the audio-visual comfort. Indeed, some respondents from Hong Kong and Southeast Asia said that they preferred crowds and a life of peddling and marketing, and noisy markets could be tolerated as long the environment did not feel too 'loud'. This indicates that the living environment could form another important factor which influences the evaluation of audio-visual comfort.

Table 5**.** Spearman's correlation coefficient between SL and related variables, including the two-tailed significance levels. Significant correlations are marked with * ($p < 0.05$) and ** ($p < 0.01$).

Spearman's rho		Acoustic comfort	Audio-visual comfort	H_{min}	H_{max}	H_{ave}	W	W/H_{min}	W/H_{ave}	W/H_{max}
SL	r	- 0.717*	-0.683*	0.571	0.584	0.567	0.632*	-0.426	-0.590	-0.774**
	p	0.020	0.030	0.085	0.077	0.087	0.05	0.220	0.073	0.009

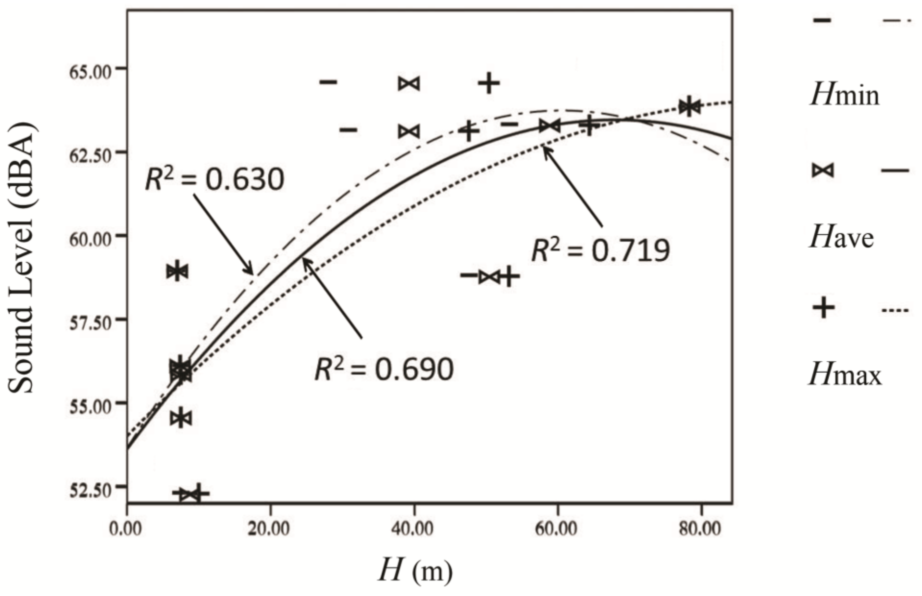


Figure 9. Relationship between SL and H.

The relationships between SL and street scale (H, W and H/W) were also examined. The dependence of LA_{eq} on H is shown in Figure 9. Although LA_{eq} increases with increasing H, and the R^2 values between LA_{eq} and H (H_{min}, H_{ave}, H_{max}) are 0.630, 0.690, and 0.719, respectively, significant correlations were not observed between LA_{eq} and H since the Spearman's correlation coefficient p is greater than 0.05 between LA_{eq} and H (Table 5).

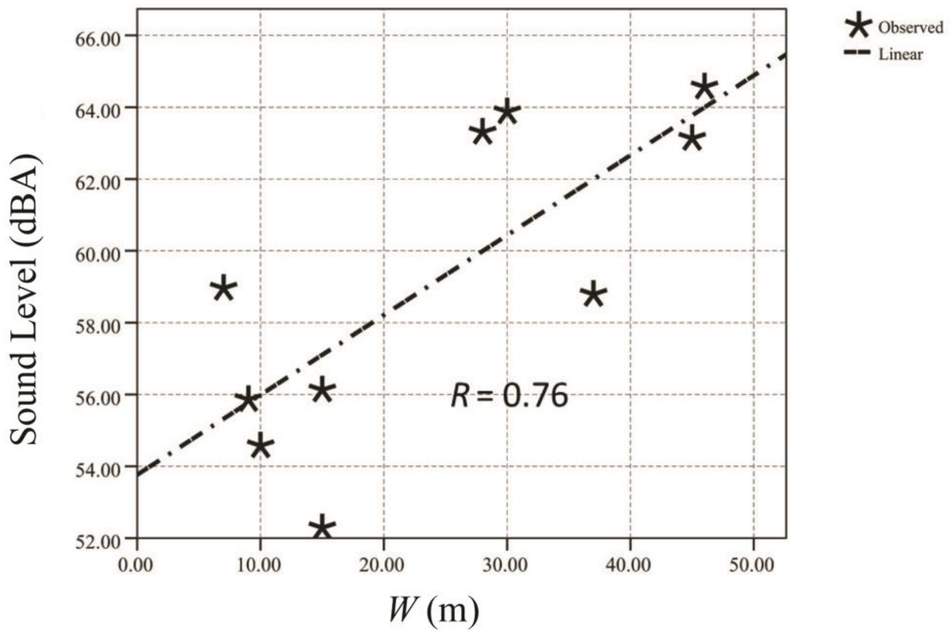


Figure 10. Relationship between W and SL.

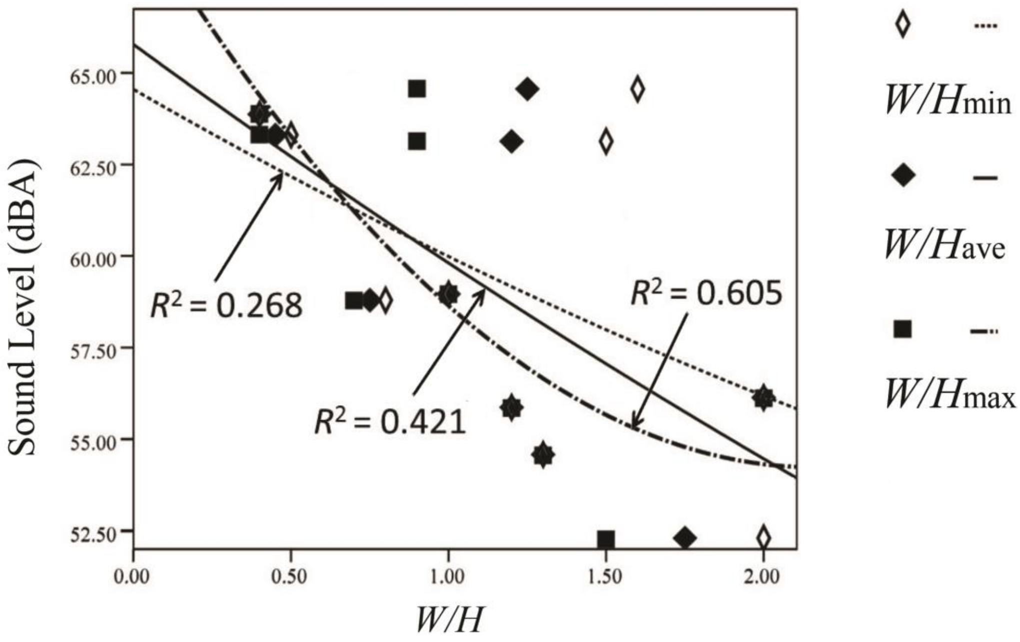


Figure 11. Relationship between SL and W/H.

Figure 10 shows the relationship between W and LA_{eq}, including the linear regressions and the correlation coefficients R. Parameter W increases with increasing LA_{eq}, with an R value of 0.76. We observed a significant correlation between LA_{eq} and W, as can be inferred from Table 5, according to Spearman's correlation (r = 0.632, p = 0.05). Therefore, limiting the width of the street can also limit LA_{eq}. In addition, we found a significant negative correlation between acoustic comfort and LA_{eq}. Therefore, it can be concluded that restricting the width of the street can improve the sound comfort level. Although our choice of streets covers a broad range of street widths from 7 to 46 m, the observed relationship between W and LA_{eq} in the present work is consistent with a previous study which compared the streets in the UK and Hong Kong (HK), where the mean values of W are 15.2 m and 26 m, respectively. The LA_{eq} values for narrow streets in the UK are slightly lower than those for the streets in the HK with a line source [5].

Figure 11 depicts the relationship between LA_{eq} and W/H with the corresponding quadratic regressions and the correlation coefficients R^2: LA_{eq} decreases with increase in W/H. In addition, the R^2 values between W/H (W/H_{min}, W/H_{ave}, W/H_{max}) and LA_{eq} are 0.268, 0.421, and 0.605, respectively, in Figure 11. Only one significant negative correlation between W/H_{max} and LA_{eq} was observed. The corresponding Spearman's correlation coefficient is listed in Table 5 (r = -0.774, $p < 0.01$). The field measurement results corresponding to the relationship between W/H_{max} and the acoustic comfort level agree with previous research based on the coupled finite-difference time-domain-parabolic equation (FDTD-PE) model which suggested that except for very narrow streets, the shielding of buildings between the streets was insensitive to the W/H value of parallel streets for sound propagation [5, 36].

4. Conclusions

Pervious research has indicated that more than 80% of the human sensory input is visual [37], and as a result, the audio-visual senses majorly contribute to obtaining information from the surrounding environment. This work demonstrates that street scales play very important roles in determining people's overall audio-visual comfort. In particular, the correlations between the three indicators (the visual, acoustic, and audio-visual comfort levels) and the street scale (*W/H*, *W*, and *H*), and those between the two indicators (acoustic, and audio-visual comfort levels) and SL, as well as the relationships between the SL and the street scale (*W/H*, *W*, and *H*) can be summarised as follows: Strong positive correlations are observed between (1) visual comfort and *W/H*, (2) audio-visual comfort and *W/H*, and (3) *W* and SL, while significant negative correlations are observed between (1) acoustic comfort and *W*, (2) visual comfort and *H*, (3) acoustic comfort and *H*, (4) audio-visual comfort and *H*, (5) acoustic comfort and SL, and (6) audio-visual comfort and SL.

Although the effect of the street scales on audio-visual perceptions cannot be directly calculated since acoustic and visual factors are not changeable independently in the real world, our survey study suggests that subjective comfort evaluations are directly related to the scales of streets. A high quality of visual, acoustic, and audio-visual comfort can be achieved by increasing *W/H* and reducing *W* and *H*, and the reduction in SL is beneficial to both acoustic and audio-visual comfort. To increase the audio-visual comfort, the following scales of streets are recommended: $W/H > 1$, $W < 20$ m, $H < 26$ m, SL < 58 dBA. However, the influence of the street scales on acoustic and visual comfort evaluation was investigated independently in the present study. Future work will be focused on revealing the effect of the street scales on audio-visual interactions in urban areas.

We believe that our study can be beneficial to urban designers and architects in reasonably predicting and controlling the street environment by varying the street scales [38–40] to design urban environments with high levels of audio-visual comfort in future urban planning and construction.

Acknowledgements
This study was supported by the National Natural Science Foundation of China (51608147, 51678180, and 51438005).

References

[1] K.W. Chan, Crossing the 50 percent population rubicon: can China urbanize to prosperity?, Eurasian Geogr. Econ. 53(1) (2012) 63-86.
[2] D. Webster, J. Cai, L. Muller, The new face of peri-urbanization in east Asia: Modern production zones, middle-class lifestyles, and rising expectations, J. Urban Aff. 36(s1) (2014) 315-333.
[3] F.M. Wiener, C.I. Malme, C.M. Gogos, Sound propagation in urban areas, J. Acoust. Soc. Am. 37(4) (1965) 738-747.
[4] E. Schröder, Nachhall in geschlossenen bebauten Straβen, Lärmbekämpfung 17 (1973) 11.
[5] J. Kang, Urban sound environment, Taylor & Francis, London and New York, 2007.
[6] W. Yang, J. Kang, Acoustic comfort evaluation in urban open public spaces, Appl. Acoust. 66(2) (2005) 211-229.
[7] G.M. Echevarria Sanchez, T. Van Renterghem, K. Sun, B. De Coensel, D. Botteldooren, Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space, Landscape Urban Plan. 167 (2017) 98-107.
[8] J. Song, H. Ma, Study on how to create a comfortable soundscape for commercial open space, J. Acoust. Soc. Am. 131(4) (2012) 3439-3439.
[9] T. Lindvall, E.P. Radford, Measurement of annoyance due to exposure to environmental factors, Environ. Res. 6(1) (1973) 1-36.
[10] R. Guski, U. Felscher-Suhr, R. Schuemer, The concept of noise annoyance: how international experts see it, J. Sound Vib. 223(4) (1999) 513-527.
[11] L. Maffei, M. Masullo, F. Aletta, M. Di Gabriele, The influence of visual characteristics of barriers on railway noise perception, Sci. Total Environ. 445 (2013) 41-47.
[12] H. Ma, S. Zhang, A case study of soundscape design based on acoustical investigation, J. Acoust. Soc. Am. 131(4) (2012) 3438-3438.
[13] F. Liu, J. Kang, A grounded theory approach to the subjective understanding of urban soundscape in Sheffield, Cities 50 (2016) 28-39.
[14] J.Y. Hong, J.Y. Jeon, Designing sound and visual components for enhancement of urban soundscapes, J. Acoust. Soc. Am. 134(3) (2013) 2026-2036.
[15] J.Y. Hong, J.Y. Jeon, The effects of audio–visual factors on perceptions of environmental noise barrier performance, Landscape Urban Plan. 125 (2014) 28-37.

[16] T. Cassidy, Environmental psychology : behaviour and experience in context, first ed., Psychology Press, New York, 1997.
[17] S. Watkins, L. Shams, O. Josephs, G. Rees, Activity in human V1 follows multisensory perception, NeuroImage 37(2) (2007) 572-578.
[18] A.L. Schormans, M. Typlt, B.L. Allman, Crossmodal plasticity in auditory, visual and multisensory cortical areas following noise-induced hearing loss in adulthood, Hear. Res. 343 (2017) 92-107.
[19] A. Preis, J. Kociński, H. Hafke-Dys, M. Wrzosek, Audio-visual interactions in environment assessment, Sci. Total Environ. 523 (2015) 191-200.
[20] V. Santangelo, S. Fagioli, E. Macaluso, The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space, NeuroImage 49(3) (2010) 2717-2727.
[21] V. Santangelo, C. Spence, Multisensory cues capture spatial attention regardless of prceptual Load, J. Exp. Psychol. Hum. Percept. Perform. 33(6) (2007) 1311-1321.
[22] D. Talsma, D. Senkowski, S. Soto-Faraco, M.G. Woldorff, The multifaceted interplay between attention and multisensory integration, Trends Cogn Sci. 14(9) (2010) 400-410.
[23] Ministry of Environmental Protection of the People's Republic of China, Environmental Quality Standard for Noise China, Environmental Science Press, Beijing, 2008.
[24] M. Lindquist, E. Lange, J. Kang, From 3D landscape visualization to environmental simulation: The contribution of sound to the perception of virtual environments, Landsc. Urban Plan. 148 (2016) 216-231.
[25] R.J. Pheasant, M.N. Fisher, G.R. Watts, D.J. Whitaker, K.V. Horoshenkov, The importance of auditory-visual interaction in the construction of 'tranquil space', J. Environ. Psychol. 30(4) (2010) 501-509.
[26] J.L.R. Joynt, J. Kang, The influence of preconceptions on perceived sound reduction by environmental noise barriers, Sci. Total Environ. 408(20) (2010) 4368-4375.
[27] X. Ren, J. Kang, Effects of the visual landscape factors of an ecological waterscape on acoustic comfort, Appl. Acoust. 96 (2015) 171-179.
[28] T. Yu, H. Behm, R. Bill, J. Kang, Audio-visual perception of new wind parks, Landsc. Urban Plan. 165 (2017) 1-10.
[29] P.J. Lee, J.Y. Hong, J.Y. Jeon, Assessment of rural soundscapes with high-speed train noise, Sci. Total Environ. 482 (2014) 432-439.
[30] F. Ruotolo, V.P. Senese, G. Ruggiero, L. Maffei, M. Masullo, T. Iachini, Individual reactions to a multisensory immersive virtual environment: the impact of a wind farm on individuals, Cogn. Process. 13(1) (2012) 319-323.
[31] L. Yu, J. Kang, Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach, J. Acoust. Soc. Am. 126(3) (2009) 1163-1174.
[32] G.N. Bratman, J.P. Hamilton, G.C. Daily, The impacts of nature experience on human cognitive function and mental health, Ann. N. Y. Acad. Sci. 1249(1) (2012) 118-136.
[33] M. Brink, D. Schreckenberg, D. Vienneau, C. Cajochen, J.-M. Wunderli, N. Probst-Hensch, M. Röösli, Effects of Scale, question Location, order of response alternatives, and season on self-reported noise annoyance using ICBEN scales: a field experiment, Int. J. Environ. Res. Public Health 13(11) (2016) 1163.

[34] D. George, P. Mallery, IBM SPSS Statistics 21 Step by Step: a Simple Guide and Reference, thirteenth ed., Pearson, Boston, 2013.
[35] L. Maffei, T. Iachini, M. Masullo, F. Aletta, F. Sorrentino, V. Senese, F. Ruotolo, The effects of vision-related aspects on noise perception of wind turbines in quiet areas, Int. J. Environ. Res. Public Health 10(5) (2013) 1681.
[36] T. Van Renterghem, E. Salomons, D. Botteldooren, Parameter study of sound propagation between city canyons with a coupled FDTD-PE model, Appl. Acoust. 67(6) (2006) 487-510.
[37] I. Rock, C.S. Harris, Vision and Touch, W.H. Freeman, San Francisco, 1967.
[38] J. Kang, Sound propagation in street canyons: comparison between diffusely and geometrically reflecting boundaries, J. Acoust. Soc. Am. 107 (3) (2000) 1394–1404.
[39] J. Kang, Sound propagation in interconnected urban streets: a parametric study, Environ. Plann. Plann. Des. 28 (2) (2001) 281–294.
[40] J. Kang, Numerical modelling of the sound fields in urban squares, J. Acoust. Soc. Am. 117 (6) (2005) 3695–3706.