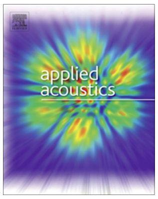
Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust



Effect of sound sequence on soundscape emotions

Zhihui Han [a], Jian Kang [b,*], Qi Meng [a,*]

[a] 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, NO. 66 Xi Da Zhi Street, Harbin, China
[b] UCL Institute for Environmental Design and Engineering, The Bartlett, University College London (UCL), London WC1H 0NN, United Kingdom

ARTICLE INFO

Article history:
Received 19 October 2022
Received in revised form 8 March 2023
Accepted 1 April 2023

Keywords:
Sound sequence
Emotion
Sound source
Soundscape design

ABSTRACT

This study analysed the effect of sound sequence on soundscape emotions with respect to three aspects of sound sources: the number of sound source/s, changing trends in the number of sound source/s (increment/decrement) and category of sound source/s. A laboratory listening test was conducted with 31 participants and data on emotions evoked by the sound source per second, with its different characteristics, were obtained. A linear regression model was established between the three aspects of sound sources mentioned above and emotions. The results reveal the following: *first*, the number of sound source/s is negatively and positively correlated to the pleasing and arousing dimensions of emotion, respectively; *second*, changing trends in the number of sound source/s (increment/decrement) has a significant effect on variations in emotions ($p < 0.05$), but not on emotions itself ($p > 0.05$); and *third*, although the category of the sound source/s has a decisive effect on the coordinate range of emotions in a two-dimensional emotion space, the number and changing trends of sound source/s have a limited effect on it. *Finally*, the linear regression model, composed of the three aspects of sound sources, could explain the values of 33.2% and 28.8% for the pleasantness and arousal dimensions of emotion, respectively.

1. Introduction

The perception of soundscapes is multi-faceted, however, perceived affective quality and emotional perception are the two major aspects [1–4]. Existing literature on the perceived affective quality of soundscape mostly focused on establishing perceptive dimensions [5]. According to the International Organization for Standardization (ISO), perceived affective quality can be measured by a two-dimensional model, composed of the pleasantness and eventfulness dimensions [6]. However, Axelsson et al. proposed three dimensions: pleasantness, eventfulness and familiarity, they argued that the first two are independent dimensions of the perceived affective quality of soundscape [7]. Similarly, Cain et al. added calmness and vibrancy as principal dimensions [8]. Several studies have demonstrated that investigating the emotional perception of soundscapes is important for future research [5]. They revealed that environmental sounds trigger extensive emotional responses. For example, the list of International Affective Digitized Sounds, which consists of 167 everyday-life natural sounds lasting for six seconds, triggers different emotional responses like pleasantness, arousal and dominance [9]. Masullo et al. examined environmental sounds of longer durations and found that the evoked emotions changed with modifications in the acoustic features like spectrum, intensity and frequency [10]. In general, the perceived affective quality of soundscape is derived from the positive or negative factors of the perception. However, the emotional perception of soundscape centres on psychology. The latter emphasises on primitive feelings about the environment, which represent the survival response to adapt to the environment [10]. Therefore, emotional perception is a more direct feeling about the auditory environment and is also the focus of this study.

Urban soundscapes can be divided into positive and negative soundscapes, based on perception. These refer to soundscapes that are perceived positively, such as through feelings of pleasure or comfort, and negatively, such as through feelings of displeasure or discomfort, respectively [11,12]. Utilising the effects of positive soundscapes to build an enhanced sound environment in cities has become an important method in soundscape designing [13]. By combing a sound source that evokes positive feelings with a piece of soundscape recording from another source, the negative soundscape's effect is reduced with the help of the different sounds' masking effects [14]. Moreover, employing natural sounds decreases the negative effects of traffic sounds and increases positive feelings about the urban soundscape [15,16]. Hong et al. combined the sounds of birds and water with real soundscape

* Corresponding authors.
E-mail addresses: J.kang@ucl.ac.uk (J. Kang), mengq@hit.edu.cn (Q. Meng).

https://doi.org/10.1016/j.apacoust.2023.109371
0003-682X/© 2023 The Authors. Published by Elsevier Ltd.

recordings from urban spaces, to assess whether the effects of natural sounds reduce the annoyance evoked by urban soundscapes. Their results showed that by incorporating natural sounds into the soundscape, the perceived loudness of traffic noises could be effectively reduced and additionally, the overall soundscape quality could be improved. They further pointed out that the signal-to-noise ratio and time features were the key factors in soundscape designs [17]. Jeon et al. conducted a laboratory listening test to investigate which kind of water sound was appropriate for masking urban noises. The subjective responses to the stimuli were rated by preference scores and 15 chosen adjectives. The findings revealed that preference scores for urban soundscapes were affected by the acoustic characteristics of water sounds [18]. Similarly, Hao et al. discussed how the masking effect of the sound of birds could improve the soundscape quality of traffic. They combined voice, wind and water sounds with that of traffic to render the experimental material more realistic. The results indicated that the sound of birds had a significant effect on improving the soundscape quality, especially when the traffic noises were low or distant [19]. It was observed that the negative soundscape that required improvement either had a single sound source or a combination of multiple sound sources. The question therefore arises: does a positive sound source affect a single negative sound source and a combination of multiple negative sound sources differently? If so, is this discrepancy related to the number of sound sources in the sound sequence? This question is important because the ultimate number of sound sources changes when one sound source is combined with another. Therefore, this study investigates the effect of the number of sound sources and changing trends (increment/decrement) in the number of sound sources, on emotion.

Another important element is capturing the emotional changes in the process of listening to a sound sequence with various sound sources. Emotion is a common psychological phenomenon that includes subjective experience, physiological arousal and behaviour [20]. There are many methods of measuring it, including subjective reports, physiological monitoring, behaviour recognition, etc. [21]. Research on soundscapes employs the method of subjective reports most widely and uses scales to evaluate emotions on the basis of emotional dimension theory [22]. This theory disassembles an emotion into several dimensions for measurement, allowing for its comprehensive assessment [23]. For example, the PAD (Pleasure–Arousal–Dominance) scale, based on Wundt's three-dimensional theory, is most commonly employed [24]. Participants have to fill in the scale after they have been exposed to all the stimuli; however, this scale does not consider the emotional changes during the process. Nonetheless, obtaining emotional data per second is particularly important for this research. The method of continuous emotion measurement, which is based on the two-dimensional theory of emotion, solves this problem; it has been utilised extensively in the emotional research of music [25,26]. For example, Schubert used the two-dimensional emotion-space software (2DES) to capture emotional changes caused by four pieces of music. He pointed out the use of 2DES as a means of collecting emotional responses to music within an ecologically valid framework and examined its validity and reliability [27]. Sharma et al. also developed a continuous, real-time, joystick-based emotion annotation framework. A laboratory test was conducted with 30 participants, who were asked to watch eight emotion-inducing videos and indicate their instantaneous emotional state in a valence-arousal (V-A) space, using a joystick. The results confirmed the framework's value, validity and usability [28]. Thus, the method of continuous emotion measurement was found to be an effective way of capturing the emotional changes caused by stimuli, and has also been used in this study.

Therefore, this study employs the method of continuous emotion measurement to capture emotional changes, and investigates three aspects of sound sources: the number of sound source/s, changing trends in the number of sound source/s and category of sound source/s, on emotions in a sound sequence. It further establishes a linear regression model to assess the relationship between these three aspects and emotion. The results are expected to be obtained according to the following aspects: first, does the number of sound sources in a sound sequence affect emotion? If yes, how? Second, do changing trends (increment/decrement) in the number of sound sources in a sound sequence affect emotion? If yes, how? Third, to what extent does the category of the sound source affect emotion, if we control for the other two aspects of sound sources in a sound sequence? Finally, to what extent can a linear regression model based on these three aspects explain emotion?

2. Method

2.1. Production of the sound sequences

According to the ISO (International Organization for Standardization) [29], urban soundscapes include sounds generated by human and non-human activities. The former mainly includes sounds of vehicles, footsteps, machines, voices, instruments etc. The latter mainly consist of natural and animal sounds. Studies have shown that there are differences between perceptions of non-natural and natural sounds [30,31]. Therefore, this study created three types of sound sequence, namely, natural, non-natural and mixed (non-natural and natural combined). The non-natural and natural sound sequences were consisted of sounds related to human and non-human activities, respectively. Of the mixed sound sequence, one half consists of sounds generated by human activities, while the other half consisted of those generated by non-human activities. Subsequently, each kind of sound sequence was further divided into an increment and decrement sequence. In the former, the number of sound sources gradually increased from one to four and, in the latter, it gradually decreased from four to one. The increment or decrement were two different experimental conditions, and there was no order between them. Therefore, there were six kinds of sound sequence in this study: non-natural increment, non-natural decrement, natural increment, natural decrement, mixed increment and mixed decrement (Table 1).

With regard to selecting the sound sources in a sound sequence, referring to previous studies, and higher frequency sounds that were part of real urban soundscapes were subsequently chosen as being representative [32,33]. Accordingly, the non-natural sound sequence comprised of four sound sources, namely, traffic sounds, voices, footsteps and advertisement sounds; the natural sound sequence included bird calls, water sounds, wind sounds and wind-rustled leaves; and the mixed sound sequence included bird calls, water sounds, voices and traffic noises, which were selected partly from the non-natural sound sequence and partly from the natural sound sequence.

With regard to creating the sound sequences (Fig. 1), the audio files were first downloaded from an international database (via the website https://www.ear0.com/), having of durations ranging from one to two minutes, waveform audio file format and 44000 Hz sampling frequency. Fig. 1 illustrates the steps for creating a sound sequence. Consider the process of creating the non-natural sound sequence as an example: First, in the non-natural increment sound sequence, there were one, two, three and four sound source/s appearing in the time frames of 0 s–30 s, 30 s–60 s, 60 s–90 s and 90 s–120 s, respectively (Table 1). Participants heard the sound sources in the following specific order: traffic noises, a combination of traffic noises and voices, a combination of traffic noises, voices and footsteps, and a combination of traffic noises, voices, footsteps and advertisement sounds. Then, the clips of 120 s,

Table 1
Composition of Sound Sources in the Sound Sequences.

	Increment sequence				Decrement sequence			
Number of sound sequences	1 (0 s–30 s)	2 (30 s–60 s)	3 (60 s–90 s)	4 (90 s–120 s)	4 (0 s–30 s)	3 (30 s–60 s)	2 (60 s–90 s)	1 (90 s–120 s)
Non-natural sound sequence	Traffic	Traffic Voice	Traffic Voice Footsteps	Traffic Voice Footsteps Advertisement	Traffic Voice Footsteps Advertisement	Traffic Voice Footsteps	Traffic Voice	Traffic
Natural sound sequence	Bird	Bird Water	Bird Water Wind	Bird Water Wind Wind blows leaves	Bird Water Wind Wind blows leaves	Bird Water Wind	Bird Water	Bird
Mixed sound sequence	Bird	Bird Water	Bird Water Voice	Bird Water Voice Traffic	Bird Water Voice Traffic	Bird Water Voice	Bird Water	Bird

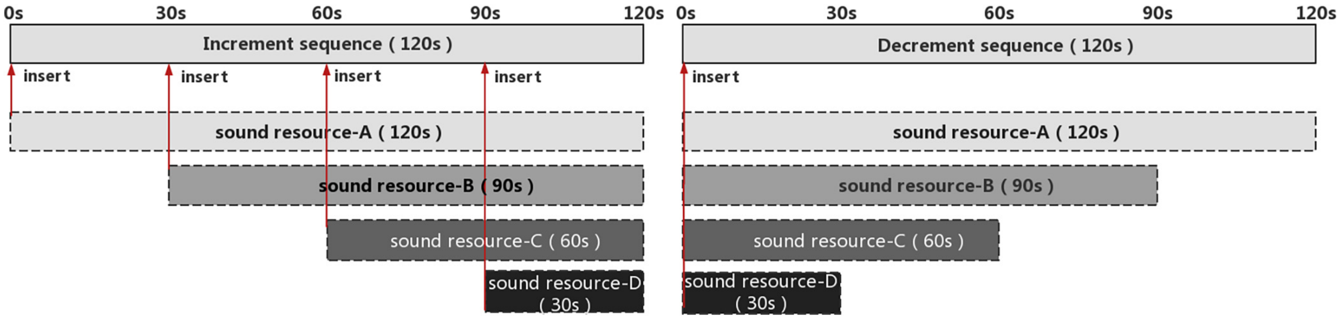


Fig. 1. Method of Creating a Sound Sequence.

90 s, 60 s and 30 s were intercepted from the traffic noises, voices, footsteps and advertisement sounds audio, respectively; subsequently, these clips were inserted at the time points of 0 s, 30 s, 60 s and 90 s, respectively. They were combined into one track to arrive at the non-natural increment sound sequence. Second, for the non-natural decrement sound sequence, there were four, three, two and one sound source/s in the time frames of 0 s–30 s, 30 s–60 s, 60 s–90 s and 90 s–120 s, respectively (Table 1). The clips of 120 s, 90 s, 60 s and 30 s were intercepted from the traffic noises, voices, footsteps and advertisement sounds audio, respectively, and then inserted at the time point of 0 s in the sound sequence; subsequently, these were combined into one track for the non-natural decrement sound sequence. The other sound sequences were made in a similar manner, using Cooledite software was used for editing. The sound pressure levels of each sequence changed with the number of sources. The following procedure was followed when setting the sound pressure levels of the sequence: when the number of the sound sources was one, the sound pressure level was set to 75 dB(A); when the number of sound sources was two, the sound pressure level was set to 78 dB(A); when the number of sound sources was three, the sound pressure level was set to 81 dB(A); and when the number of sound sources was four, the sound pressure level was set to 84 dB(A) [34].

Table 1. presents the order of sound sources in each sound sequence, the order of emerging sound sources resemble the occurrence of sound sources in an environment over the day. For example, the order of sound sources in the non-natural increment sound sequence refers to the characteristics of sound sources in the urban street sound sequence. The traffic sound source is the first one, because it appears first and most frequently throughout a day in the street. When the street is crowded with people, the human-related sound sources appear, such as voices and footsteps. When the mall opens, the sounds of advertisements emerge. The order of sound sources in the non-natural decrement sound sequence is just the opposite; from afternoon to night, the human-related sound sources in the urban street disappear first, and the traffic noises remain until late at night. The order of sound sources in the natural and mixed sound sequences mainly refer to the characteristics of sound sources in a city park and other public spaces with greenery and water sounds, respectively.

2.2. Software

This study used the EMuJoy software for continuous emotion measurement [35]. It allowed recording the participants' emotional changes when the audio files were played, in the form of points that the participants positioned on the computer screen, using a mouse. The software interface comprised of two vertically intersecting coordinate axes, as seen in Fig. 2. The X-axis represented the 'pleasantness' dimension of emotion and the left and right sides of the coordinate axis represent negative and positive, respectively, with values ranging from -1 to 1. The Y-axis represented the 'arousal' dimension of emotion, with 'calming' and 'arousing' at the bottom and top, respectively, with values ranging from -1 to 1 (Fig. 2). Upon the audio being played, participants were required to click on the screen with the mouse, to confirm the point that corresponded to their felt emotion in the two-dimensional space at that given time. Furthermore, the software recorded the values of the X- and Y-axes of the point, as well as their corresponding timings. This was how data of the pleasantness and arousal dimensions of emotional changes were obtained, while the audio was being played. The software's sampling frequency was 50 ms, and its effectiveness was proven [35–37].

Theoretically, the emotions reported by the participants were evoked only by the sounds they heard and not any other factor. In fact, this experiment was not a natural task. Although some

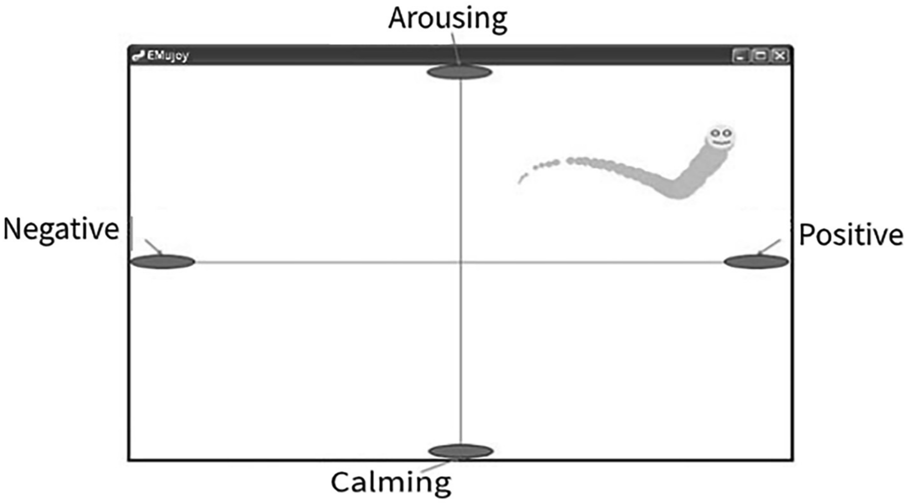


Fig. 2. Software Interface for Participants.

scholars have employed physiological methods for recording emotions indirectly [36], subjective reports are considered the most appropriate way of directly measuring emotions evoked by sound [5].

2.3. Procedure

The experiment was divided into three stages (Fig. 3). First, the entire experimental procedure was explained to the participants and their informed consent was obtained. Second, the EMuJoy software's interface and method of operating it, were explained to them. Previous research has shown that pictures are usually the ideal stimuli for allowing participants to adapt to using a particular software [35,38]. Therefore, five pictures were selected from the International Affective Picture System, corresponding to the emotions represented by the four quadrants in the two-dimensional emotion space and the emotion at the coordinate origin (0,0) [39]. The pictures were randomly displayed to the participants. After confirming that they had understood and could operate the software correctly, the formal experiment was initiated. In the formal experimental stage, the increment or decrement sequences were played to the participants randomly in order to eliminate the effect of order between the sound sequence on results. Since the focus of this study was not on acoustic parameters, the audio files were played back to the participants at the same maximum volume of 81 dB(A). That is to say, the volume of the clips where there were four sound sources exiting at the same time in the sequence, was 81 dB(A). The playback volume of the headphones was calibrated by connecting the dummy heads (Head acoustics HMS III) before the experiment. Each sound sequence was two minutes long, with a 30 s interval between each of them, to eliminate the potential influence of the previous sequence on the following one [40]. The experiment's total duration was 15 min. Participants were requested to listen to the audio carefully, identify whether their emotions were affected by the stimuli, and indicate the position of their emotions on the screen, using the mouse, when their emotions changed with the sounds. They could report their emotion at any time and without any limitation on the number of times they could do so. The whole experiment was conducted in a listening room with a background noise of 25 dB(A). As shown in Fig. 4, the laboratory measured 8.3 m × 6.4 m. During the experiment, the participant sat in front of the computer at the centre of the laboratory, and the audio was played through Sennheiser RS170 headphones, as shown in Fig. 5. The starting and ending of the audio were remotely controlled by the monitoring computer, and the experimental data were transmitted online; these ensured that the participants were not disturbed during the experiment.

2.4. Participants

Thirty-one participants were selected for this experiment, including 16 males and 15 females. Since the experiment would not focus on differences according to age group, participants aged between 20 and 30 years were finally chosen for the convenience of recruiting, with a mean age of 25 years and standard deviation of 3.8 years. Their occupations included student, office worker and freelancer. All the participants freely volunteered for participation.

2.5. Data analysis

This study established a linear regression model with three aspects of sound sources, namely, the number of sound source/s, changing trends in the number of sound source/s and category of sound source/s, and emotion/s in a sound sequence. Therefore,

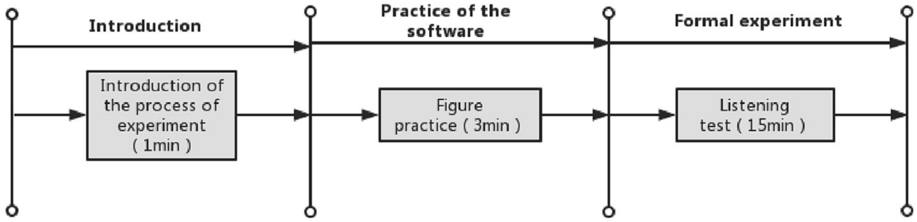


Fig. 3. The Process of the Experiment.

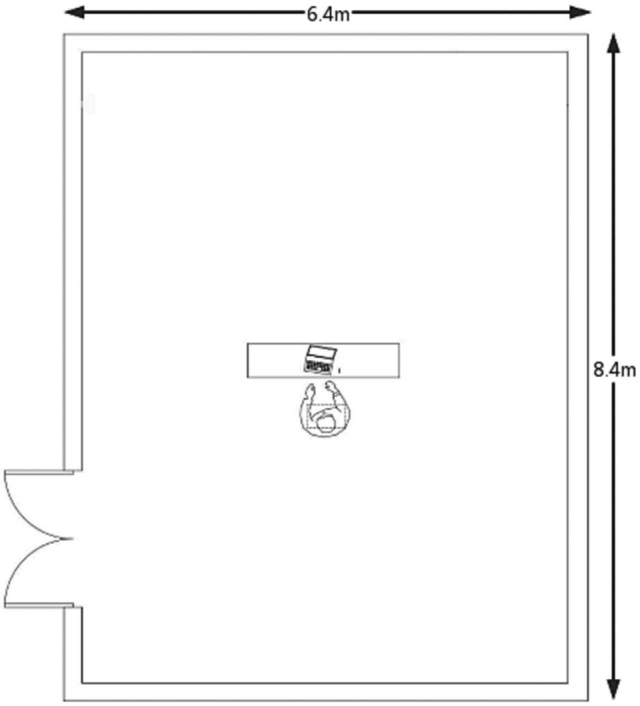


Fig. 4. The Layout of the Laboratory.

the dependent and independent variables were continuous and categorical variables, respectively. A linear regression model cannot directly establish quantitative relationships between continuous and categorical variables. Therefore, the latter should be encoded as dummy variables into the model, for regression. They usually belong to several categories, thus, a benchmark category should be selected before encoding. Changing trends in the number of sound sources is a binary variable (increment/decrement), while the sound source category is a tertiary variable (natural/non-natural/mixed sound sequences). Therefore, 'increment' and 'natural sound sequence' were selected as the benchmark categories, and coded as 00 and 000, respectively. The settings of the other categories were required to refer to those of the benchmark category. However, while the benchmark category did not appear in the final model, it aided in interpreting the meaning of the correlation coefficient between the categorical and dependent variables; for example, 'how intense (positive correlation) or apathetic (negative correlation) were the emotions evoked by a particular category, as compared to the benchmark category?'

3. Results

3.1. Overall emotional evaluation of the sound sequences

This section provides an overall description of the emotions triggered by the experiment's sound sequences. All subsequent results were based on the following data. Fig. 6 presents the data on emotions evoked by the sound sequences per second. Fig. 6 (a) depicts the actual value of the data, along with the timings, and Fig. 6 (b) is the variation value, that is, the difference in emotions between the subsequent and previous seconds. Fig. 6 (a) reveals that the emotions evoked by urban soundscapes change from second to second, and that the trends and inflection point of emotions triggered by different sound sequences, were different. It will be observed that the variation value ranges from −0.08 to 0.06, which is negligible, compared to the actual value of the fluctuation of emotions. The variation of emotions in each sound sequence is generally stable, with the larger variations occurring at 0 s–30 s and 60 s–90 s, as seen in Fig. 6. (b).

Fig. 5. The Sennheiser RS170 headphones.

Although the data on emotions per second could provide more comprehensive information, this data form presents a limitation for further analysis. Table 2 provides the mean value and standard deviation of the emotions evoked by the different sound sequences. If the non-natural increment sound sequence with the table's X_1 is taken as an example, its value represents the mean value of the emotion evoked by the non-natural increment sound sequence with one sound source, by averaging the 30 values from the 0 s–30 s sequence. The standard deviation is the dispersion of the 31 participants' emotional data within 0 s–30 s.

3.2. Effect of the number of sound source/s on emotion

Fig. 7 presents the standard deviation of emotional evaluation in different sound sequences. The standard deviation was used to compare the deviation of emotions values evoked by different sound sequences. First, the fluctuation of standard deviation in the arousal dimension in different sound sequences is greater than that in the pleasantness dimension. It ranged from 0.13 to 0.33 in the latter and 0.15 to 0.41 in the former. Specifically, the consistency of emotional evaluation in the different sound sequences was poor, especially with regard to the arousal dimension. Second, by comparing Fig. 7 (a), (c) and (e), it was found that the standard deviation of the non-natural sound sequence is greater than that of the natural and mixed sound sequences, in general. Particularly, the non-natural sound sequence's emotional evaluation had the worst consistency, while the mixed sound sequence had the best consistency. Third, the trend line in Fig. 7 shows the trend of standard deviation according to the number of sound source/s. It can be seen that, in the increment sequence, an increase in the number of sound source/s caused the standard deviation of the pleasantness and arousal dimensions to increase in the non-natural and mixed sound sequences. Specifically, as the number of sound sources

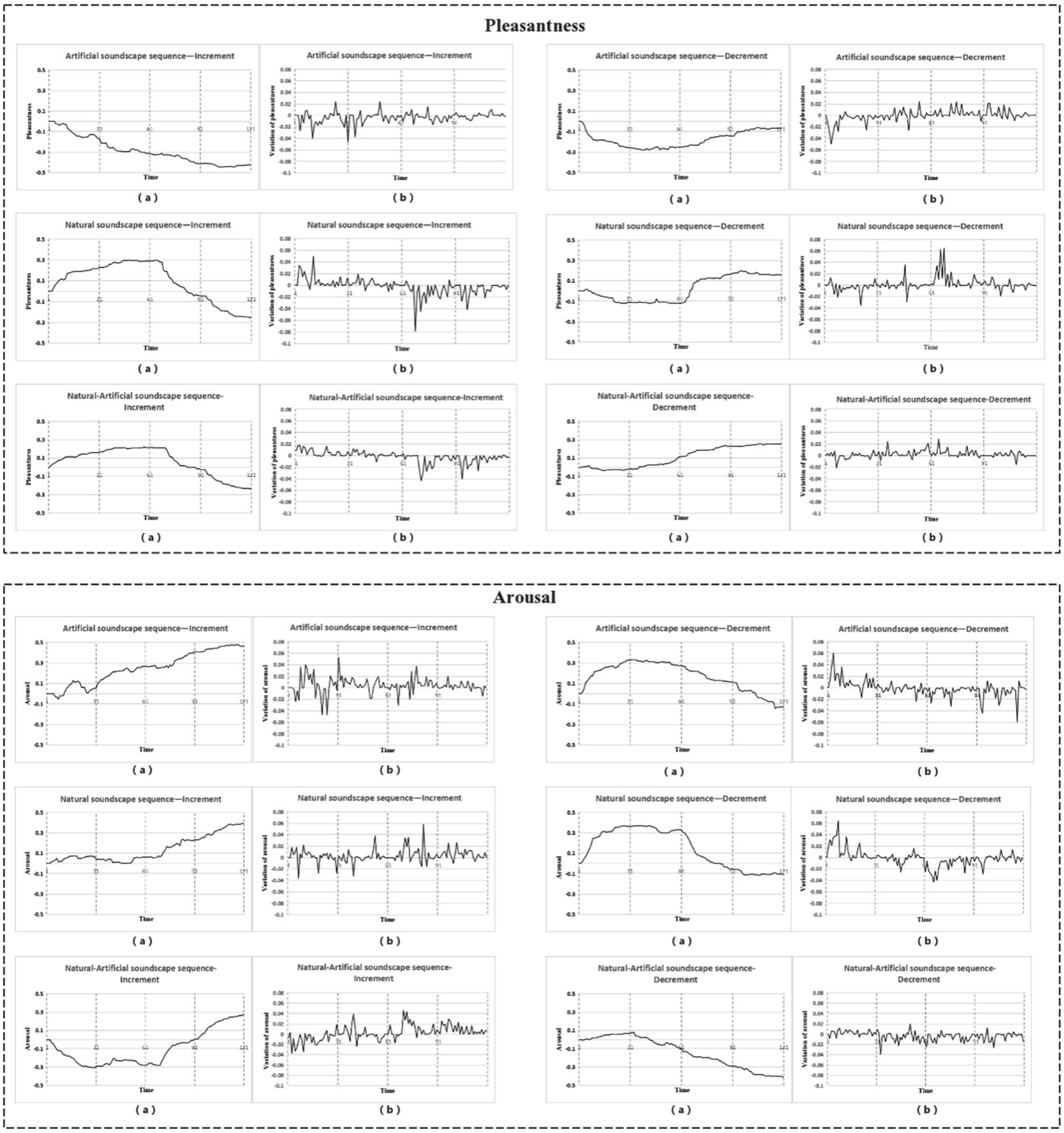


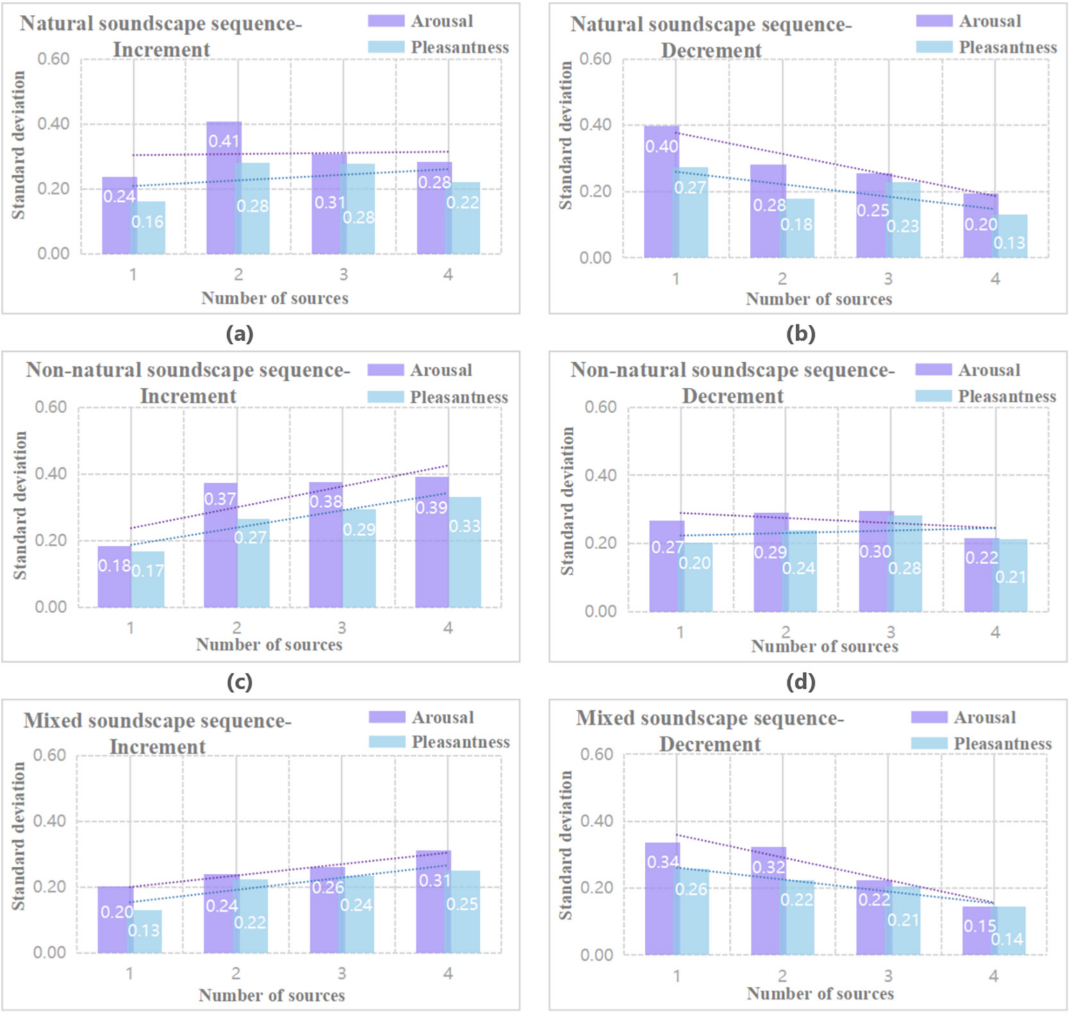
Fig. 6. Data of Emotions, Along with Time, in Different Sound Sequences. Notes: (a) is the actual value of the emotion, with time; (b) is the difference value of the emotion, with time.

Table 2
Mean Values and Standard Deviation of Emotions in Different Sound Sequences.

Type of sound sequence		Emotional dimensions							
		Pleasantness dimension(M ± SD)				Arousal dimension(M ± SD)			
		X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4
Non-natural sound sequence	Increment	−0.09 ± 0.17	−0.27 ± 0.27	−0.35 ± 0.29	−0.43 ± 0.33	0.04 ± 0.18	0.20 ± 0.37	0.31 ± 0.38	0.45 ± 0.39
	Decrement	−0.09 ± 0.20	−0.19 ± 0.24	−0.26 ± 0.28	−0.17 ± 0.21	−0.03 ± 0.27	0.17 ± 0.29	0.31 ± 0.30	0.22 ± 0.22
Natural sound sequence	Increment	0.15 ± 0.16	0.28 ± 0.28	0.11 ± 0.28	−0.18 ± 0.22	0.05 ± 0.24	0.03 ± 0.41	0.15 ± 0.31	0.32 ± 0.28
	Decrement	0.17 ± 0.27	0.08 ± 0.18	−0.11 ± 0.23	−0.06 ± 0.13	−0.10 ± 0.40	0.08 ± 0.28	0.34 ± 0.25	0.26 ± 0.20
Mixed sound sequence	Increment	0.11 ± 0.13	0.20 ± 0.22	0.09 ± 0.24	−0.16 ± 0.25	−0.19 ± 0.20	−0.25 ± 0.24	−0.14 ± 0.26	0.17 ± 0.31
	Decrement	0.24 ± 0.26	0.18 ± 0.22	0.03 ± 0.21	−0.02 ± 0.14	−0.36 ± 0.34	−0.20 ± 0.32	−0.02 ± 0.22	0.04 ± 0.15

Notes: ‘X’ and ‘Y’ represent the pleasantness and arousal dimensions, respectively; ‘1′ represents the number of sound source/s, ‘M’ represents the mean value and ‘SD’ represents the standard deviation.

increased, the standard deviation also increased. However, in the decrement sequence, a decrease in the number of sound source/s caused the standard deviation of the pleasantness and arousal dimensions to increase in the natural and mixed sound sequences. Specifically, as the standard deviation increased, the number of sound sources decreased. Thus, the aforementioned results contrast each other. However, a change in the number of sound sources is also related to the sound sequence’s duration. An increased number of sound sources in the increment sequence corresponds to an increase in the sound sequence’s duration, but nonetheless, a decreased number of sound sources in the decrement sequence also corresponds to an increase in the sound



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 7. Standard Deviation of Emotional Evaluation in Different Sound Sequences. Notes: The trend lines related to the standard deviation values gained over the number of sources. The blue and purple dotted lines represent the trend lines of the pleasantness and arousal dimensions, respectively.

sequence's duration. It can therefore be inferred that consistency of the sound sequence's emotional evaluation is related to its duration. The longer the sequence, the poorer the consistency of emotional evaluation. Evidently, this conclusion only explains the changing number of sound sources in the sound sequences. Determining whether sound sequences with a stable number of sound source/s would be suitable, requires further research.

To analyse the relationship between the number of sound sources in a sound sequence, and the emotions evoked, a correlation analysis was conducted. The test confirmed that the data distribution was non-normal, thus, the Spearman correlation analysis was employed to assess whether there was a correlation between the emotional dimension and number of sound sources. The results show that the number of sound sources had a significantly negative correlation with the pleasantness dimension, with a correlation coefficient of -0.336 ($p < 0.05$), and a significantly positive correlation with the arousal dimension, with a correlation coefficient of 0.382 ($p < 0.05$). Specifically, when the number of sound sources in the soundscape sequence ranges from one to four, it transpired that the more the number of sound sources in a sound sequence, the lower the pleasantness, and the higher the arousal. In the two-dimensional emotion space, this is indicated as an increased number of sound sources causing the emotion to move from the lower right to the upper left of the space.

3.3. Effect of changing trends in the number of sound source/s on emotion

To analyse whether changing trends in the number of sound sources in a sound sequence, that is, an increment or decrement in the sequence, affects emotion, a difference analysis was performed on the actual value of the emotions triggered by the increment and decrement sequences. The emotional data per second, yielded by each sound sequence, were employed. The test demonstrated that the data were non-normal, thus, the rank sum test was selected for performing the difference analysis. The findings revealed no significant difference between the actual value of the emotions elicited by the increment and decrement sequences, both in the pleasantness ($p = 0.694 > 0.05$) and arousal dimensions ($p = 0.190 > 0.05$). A further difference analysis was conducted for variation in emotions (that is, the difference in the value of an emotion between the subsequent and previous seconds) caused by the increment and decrement sequences. The results show a significant difference in the pleasantness and arousal dimensions, at 0.05 ($p = 0.046 < 0.05$) and 0.1 ($p = 0.070 < 0.1$), respectively.

The detailed differences are further demonstrated in Fig. 8. The X-axis and Y-axis represent the number of sound sources and absolute value of emotional changes, respectively. This means that the

Y-axis value is not obtained by directly averaging the emotional data from three sound sequences with one sound source, but rather by averaging the three values after taking the absolute. The reason for this is that the emotions evoked by the different sound sequences had different positive or negative values. If these were directly averaged, the mean value would be extremely small, making it inconducive for further analysis. In addition, a positive or negative value is merely a description of the emotional tendency, that is, they represent pleasantness and unpleasantness, respectively. The absolute mean value consequently reflects emotional changes more effectively. Section 3.4 examines whether the emotions moved towards the positive or negative direction in the emotion space.

The results show that, first, changing trends in the number of sound sources had a greater effect on the pleasantness dimension, than the arousal dimension. Specifically, the values of the pleasantness dimension, caused by the increment and decrement sequences, were different for the same number of sound sources, with a discrepancy of about 0.05 to 0.18. This difference was significantly larger for two and four sound sources, than for one and three. The difference value in the arousal dimension, caused by the increment and decrement sequences, ranged from 0.01 to 0.14, and was significantly greater for four sound sources, than for one, two or three. Second, the increment sequence seemed to cause greater emotional changes in the pleasantness dimension, as this dimension's values in the increment sequence were constantly higher than those in the decrement sequence. Third, the trend line shows that, as the number of sound sources in the decrement sequence increased, the value of the pleasantness dimension decreased; however, as the number of sound sources in the increment sequence increased, the value of the arousal dimension increased.

In a general 2-minute soundscape sequence, changing trends in the number of sound sources had a greater effect on the pleasantness dimension, than the arousal dimension. Compared to the decrement sequence, the increment sequence triggered greater changes in the pleasantness dimension.

3.4. Effect of the category of sound Source/s on emotion

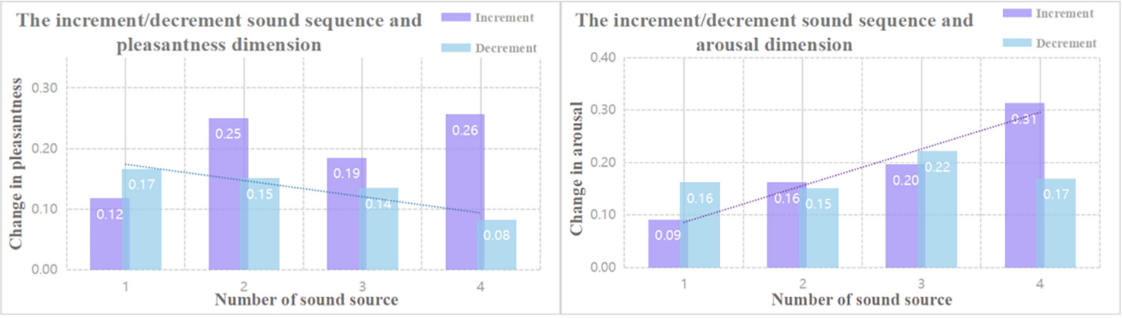
A difference analysis was conducted to analyse whether there was a statistical difference between the emotional data yielded by sound sequences composed of sound sources belonging to different categories. Data of emotions per second were employed for this analysis. The test indicated that the emotional data were non-normal, therefore, the rank sum test was selected for the difference analysis. The results demonstrate significant differences in emotions triggered by different categories of sound sources (p = 0.000 < 0.05).

The detailed differences are presented in Fig. 9. It shows the different positions of emotions in the two-dimensional emotion space, caused by different categories and number of sound sources. The blue, yellow, and green colours represent the non-natural, natural and mixed sound sequences, respectively. The solid and dashed lines denote the increment and decrement sequences, respectively. The numbers stand for the number of sound sources at each point. All the data in this figure comprises the mean value.

The results show that, first, the category of sound sources had a greater effect on emotions, compared with the other two aspects of sound sources. The category of a sound source determines the coordinate range of emotions in the emotion space, despite the different numbers and changing trends in the number of sound sources. Second, emotions evoked by the urban soundscape were basically in the range of -0.5×0.5, which was near the origin of the coordinates in the emotion space. This means that the urban soundscape did not evoke emotions of extreme value either in the pleasantness or arousal dimensions, as compared with the emotion space's entire range (-1 × 1). Third, different categories of sound sources evoke relatively different emotions. Emotions evoked by the natural sound sequence were mainly located in the first and second quadrants, positioned most closely to the origin of the coordinates, and the best of the three sound sequences. The value of pleasantness and arousal dimensions ranged from -0.2 to 0.3 and 0 to 0.4 (except for a special point), respectively. Emotions evoked by the non-natural sound sequence were mainly located in the second quadrant, positioned farthest from the origin of the coordinates, and the worst of the three sound sequences. The value of the pleasantness and arousal dimensions ranged from -0.5 to 0 and -0.1 to 0.5, respectively. Emotions evoked by the mixed sound sequence were located in the second and fourth quadrants, with the pleasantness and arousal dimensions ranging from -0.2 to 0.25 and -0.4 to 0.2, respectively. It is noteworthy that the emotions evoked by the mixed sound sequence did not simply correspond to a combination of emotions evoked by the natural and non-natural sound sequences, wherein the pleasantness dimension's value was the same as that of the natural sound sequence and the arousal dimension's value was lower than that of the other two sound sequences.

3.5. Model of Relation between emotions and the Number, changing trends in number and category of sound Source/s

Table 3 presents a linear regression relationship between emotions and the number, with changing trends in the number and category of sound source/s. The results show that, first, the model has



(a)

(b)

Fig. 8. Increment or Decrement in the Number of Sound Sequences and Emotional Changes. Notes: The values shown in the figure are the absolute values of emotional changes, and the trend lines related to the values gained over the number of sources. The trend lines in Fig. 6 (a) and (b) represent the trend of the value in the decrement and increment sequences, respectively.

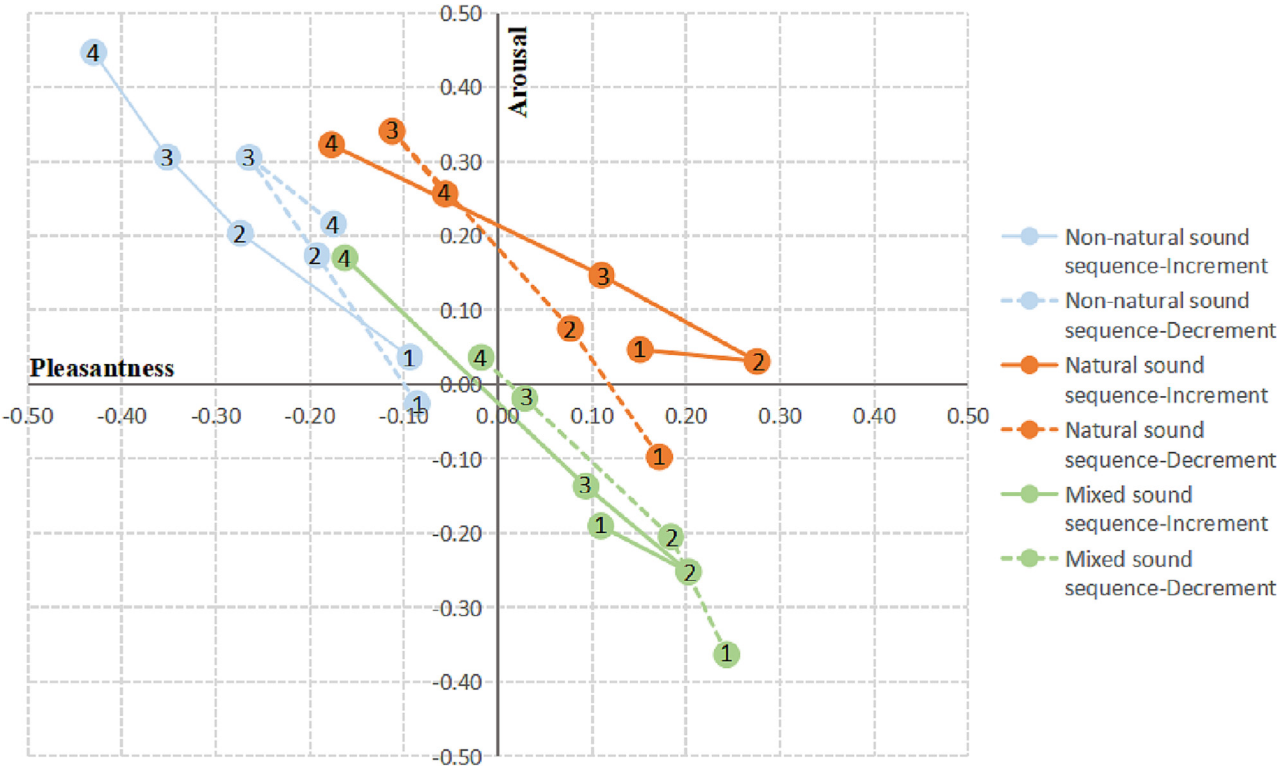


Fig. 9. Emotions Evoked by Different Categories of Sound Sources. Notes: The values shown in the figure is the average values, and “1, 2, 3,4” represent the number of sound sources at the corresponding points.

Table 3
Linear Regression Model of the Relation Between Emotions and the Three Aspects of Sound Sources.

		Unstandardised coefficient		Standardised coefficient	t	Sig.	R^2
		B	Standard deviation				
Pleasantness dimension	C	0.260	0.028		9.455	0.000	0.332
	Non-natural	−0.288	0.022	−0.451	−12.807	0.000	
	Mixed	0.030	0.022	0.048	1.352	0.177	
	Decrement	0.029	0.018	0.048	1.563	0.119	
	Number of sound source	−0.088	0.008	−0.327	−10.724	0.000	
Arousal dimension	C	−0.135	0.034		−4.000	0.000	0.288
	Non-natural	0.070	0.027	0.092	2.546	0.011	
	Mixed	−0.258	0.027	−0.340	−9.373	0.000	
	Decrement	−0.038	0.022	−0.053	−1.698	0.090	
	Number of sound source	0.117	0.010	0.365	11.616	0.000	

an explanatory ability of 33.2% and 28.8% for the pleasantness and arousal dimensions, respectively. Second, the category of the sound source contributed the most to predicting emotion, followed by the number of sound sources. Changing trends in the number of sound sources did not contribute to predicting emotion, which corresponds to the aforementioned results.

There was a negative correlation between the number of sound sources and pleasantness dimension, with a correlation coefficient of −0.327. Particularly, as the number of sound sources increases by 1, the value of the pleasantness dimension decreases by 0.327. Nevertheless, there was a positive correlation between the number of sound sources and arousal dimension, with a correlation coefficient of 0.365. Specifically, as the number of sound sources increased by 1, the value of the arousal dimension increased by 0.365. Second, interpretation of the category of sound sources’ correlation coefficient corresponds to what has been stated in section 2.5. In the pleasantness dimension model, the correlation coefficient between the non-natural sound sequence and the pleasantness dimension is −0.451 (p < 0.05), which indicates that the pleasantness evoked by the non-natural sound sequence was significantly lower than that evoked by the natural sound sequence (p < 0.05). Furthermore, the value of the pleasantness dimension for the non-natural sound sequence is 0.451 lower than that for the natural sound sequence. However, there is no significant difference in the pleasantness dimension for the mixed and natural sound sequences (p > 0.05). The same relationship exists between the non-natural sound sequence and arousal dimension. In the arousal dimension model, the arousal evoked by the mixed sound sequence was significantly lower than that evoked by the natural sound sequence. Moreover, the value of the arousal dimension for the non-natural sound sequence is 0.340 lower than that for the natural sound sequence. Finally, there is no significant difference in the values of pleasantness (p > 0.05) and arousal (p > 0.05) evoked by the increment and decrement sound sequences.

4. Discussions

4.1. Potential relationship between the number of sound Source/s and emotions

This study shows that the emotions evoked by an urban soundscape could be completely covered by a -0.5×0.5 area, centred on the coordinate origin, in the experiment's two-dimensional emotion space. In this space, the origin of the coordinate represented a neutral emotion, with the values of both pleasantness and arousal dimensions at 0. The farther away from the origin in the X-axis, the more active was the pleasantness, that is, it was either more pleasant or unpleasant. At the same time, the results also show a linear correlation between the number of sound sources and emotional dimension; the number of sound sources is negatively and positively correlated with the pleasantness and arousal dimensions, respectively. Emotions either increased or decreased with an increment in the number of sound sources. However, this may evidently not be the case, as there could be a limitation value of the two dimensions at 0.5/-0.5, with the limitation on the number of sound sources at four. Thus, it can be inferred that when the number of sound sources keeps increasing, the value of the pleasantness dimension does not decrease; however, the value of the arousal dimension would reach a peak, and eventually begin to decline. Therefore, the critical value of the number of sound sources needs further investigation.

4.2. Recommendations for soundscape design

An important approach in soundscape design has been the use of the masking effect of a sound source with positive perception to reduce the negative effect of the soundscape. This study indicates that the number of sound sources in a sound sequence plays an essential role in the emotional perception of soundscape. Therefore, it is necessary to consider both these aspects, that is, the category and number of sound sources, in soundscape design. From the perspective of the category of sound sources, natural and non-natural sound sequences trigger positive and negative emotions, respectively. When they are combined into one sequence, the non-natural sound's negative effect on emotion is greatly reduced. From the perspective of the number of sound sources, a better emotion is evoked when the number of sound sources in a sound sequence is reduced. Therefore, the best way to improve the perception of a sound sequence is not only adding positive sound sources, such as natural sounds, but also reducing the number of sound sources. Only by combining both these factors in soundscape design, can an enhanced soundscape perception be achieved.

4.3. Limitations

This research has certain limitations. First, it uses only four sound sources, thus, further studies need to be conducted on the relationship between a larger number of sound sources and emotion. Second, there are several possible orders of the appearance or disappearance of the sound sources in a sound sequence. Since this study does not focus on the order of sound sources in a sequence, only one possible situation has been employed; accordingly, the results are representative of only this situation. Third, the findings only pertain to participants aged 20–30 years; thus, more studies are required for other age groups. Finally, the results are only valid in the setting of this study's experimental conditions, therefore more studies need to be done in a larger degree of immersion and realism conditions.

5. Conclusions

This research systematically discussed the relationship between the number, changing trends in number, and category of sound source/s in a sound sequence, and emotion. Further, it established a linear regression model. The results reveal the following:

- *First*, the number of sound sources is negatively and positively correlated with the pleasantness (-0.336) and arousal dimensions (0.382), respectively. The consistency of emotional evaluation is not related to the number of sound sources, but to the duration of the sound sequence. With an increase in the number of sound sources, the position of emotions in the two-dimensional emotion space moves from the lower right to the upper left, thereby indicating that the emotions have improved.
- *Second*, the factor of changing trends in the number of sound sources (increment/decrement) does not cause a statistical difference in the actual value of emotions ($p > 0.05$); however, it does cause a significant difference in the variation value of emotions ($p < 0.05$). This factor has a greater effect on pleasantness than arousal, and the increment sequence always causes a greater change that the decrement sequence, in the pleasantness dimension.
- *Third*, the emotions triggered by an urban soundscape are positioned in the -0.5×0.5 area, centred on the coordinate origin, in the two-dimensional emotion space. Specifically, the emotions evoked by the natural sound sequence are the best, while those evoked by the non-natural sound sequence are the worst. The value range of the pleasantness dimension, caused by the mixed sound sequence, is similar to that caused by the natural sound sequence; furthermore, the value range for the arousal dimension is significantly lower than that of the other two sound sequences. The category of the sound source determines the coordinate range of emotions in the emotion space, despite the different numbers and changing trends in number of sound sources.
- *Finally*, the linear regression model of the number, changing trends in number and category of sound sources could explain 33.2% and 28.8% of the pleasantness and arousal dimensions, respectively. The number and category of sound sources contributed more to the model's explanatory power, while changing trends in the number of sound sources did not contribute to the model at all.

With regard to designing an urban soundscape, the sound source category should be the main factor. In fact, an urban soundscape is a continuously changing auditory environment; thus, more strategies that focus on the sound sequence must be proposed. This study proposed that using sound sources with a positive effect to mask the negative perception of soundscape, is insufficient. Instead, controlling for or reducing the number of sound sources in a sound sequence, while adding positive sound sources, should be considered. Nevertheless, the potential relationship between the number of sound sources and emotions, with a larger number of sound sources, is worthy of further study.

CRediT authorship contribution statement

Zhihui Han: Methodology, Formal analysis, Writing – original draft. **Jian Kang:** Conceptualization, Funding acquisition. **Qi Meng:** Funding acquisition, Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express sincere gratitude to all the participants of the experiment.

Funding

This work was supported by National Natural Science Foundation of China (NSFC) Grant Nos. 52178070 and 51778169, the Open Fund Project of Anhui Provincial Key Laboratory of Building Acoustic Environment [grant number AAE2021ZD03], Shanghai Key Laboratory of Urban Design and Urban Science, NYU Shanghai, Open Topic Grants (grant number 2022QMeng_LOUD), and European Research Council (ERC) Advanced Grant No. 740696 on "Soundscape Indices" (SSID).

References

[1] Kang J, Aletta F, Gjestland T, Brown LA, Botteldooren D, Schulte-Fortkamp B, et al. Ten questions on the soundscapes of the built environment. Build Environ 2016;108:284–94. https://doi.org/10.1016/j.buildenv.2016.08.011.

[2] Botteldooren D, Boes M, Oldoni D, De Coensel B. The role of paying attention to sounds in soundscape perception. J Acoust Soc Am 2012;131:3382. https://doi.org/10.1121/1.4708755.

[3] Kidd GR, Watson CS. The perceptual dimensionality of environmental sounds. Noise Control Eng J 2003;51:216. https://doi.org/10.3397/1.2839717.

[4] Aumond P, Can A, De Coensel B, Ribeiro C, Botteldooren D, Lavandier C. Global and continuous pleasantness estimation of the soundscape perceived during walking trips through urban environments. Appl Sci 2017;7:144. https://doi.org/10.3390/app7020144.

[5] Fiebig A, Jordan P, Moshona CC. Assessments of acoustic environments by emotions – the application of emotion theory in soundscape. Front Psychol 2020;11:. https://doi.org/10.3389/fpsyg.2020.573041573041.

[6] International Organization for Standardization. ISO/TS 12913–3:2019 Acoustics-Soundscape-Part 3: Data Analysis. Geneva: International Organization for Standardization; 2019.

[7] Axelsson Ö, Nilsson ME, Berglund B. A principal components model of soundscape perception. J Acoust Soc Am 2010;128:2836–46. https://doi.org/10.1121/1.3493436.

[8] Cain R, Jennings P, Poxon J. The development and application of the emotional dimensions of a soundscape. Appl Acoust 2013;74:232–9. https://doi.org/10.1016/j.apacoust.2011.11.006.

[9] Bradley MM, Lang PJ. The International Affective Digitized Sounds Affective Ratings of Sounds and Instruction Manual. Technical Report B-3. University of Florida, Gainesville,Fl, 2007;29–46, Retrieved from .

[10] Masullo M, Maffei L, Iachini T, Cioffi F, Ruotolo F. A questionnaire investigating the emotional salience of sounds. Appl Acoust 2021:182. https://doi.org/10.1016/j.apacoust.2021.108281.

[11] Kang J, Yang W. Sound preferences in urban open public spaces. J Acoust Soc Am 2003;114:2352.

[12] Zhang X, Ba M, Kang J, Meng Q. Effect of soundscape dimensions on acoustic comfort in urban open public spaces. Appl Acoust 2018;133:73–81. https://doi.org/10.1016/j.apacoust.2017.11.024.

[13] Nilsson M E, Botteldooren D, Coensel BD. Acoustic indicators of soundscape quality and noise annoyance in outdoor urban areas (invited paper). Proceedings of International Congress on Acoustics 2007.

[14] Jeon JY, Lee PJ, You J. Urban space design based on the perceptual assessment of soundscape. J Acoust Soc Am 2010;128:2369. https://doi.org/10.1121/1.3508414.

[15] van den Bosch K, Andringa T. The effect of sound sources on soundscape appraisal. In Proceedings of ICBEN 2014. Nara, Japan, 2014.

[16] You J, Lee PJ, Jeon JY. Evaluating water sounds to improve the soundscape of urban areas affected by traffic noise. Noise Control Eng J 2010;58:477–83. https://doi.org/10.3397/1.3484183.

[17] Hong J, Ong ZT, Lam B, Ooi K, Tan ST. Effects of adding natural sounds to urban noises on the perceived loudness of noise and soundscape quality. Sci Total Environ 2020:711134571. https://doi.org/10.1016/j.scitotenv.2019.134571.

[18] Jeon JY, Lee PJ, You J, Kang J. Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces. J Acoust Soc Am 2012;131:2101–9. https://doi.org/10.1121/1.3681938. PMID: 22423706.

[19] Hao Y, Kang J, Wörtche H. Assessment of the masking effects of birdsong on the road traffic noise environment. J Acoust Soc Am 2016;140:978–87. https://doi.org/10.1121/1.4960570.

[20] Ledoux JE. Emotion circuits in the brain. Annu Rev Neurosci 2000;23:155–84. https://doi.org/10.1146/annurev.neuro.23.1.155.

[21] Mauss IB, Robinson MD. Measures of emotion: A review. Cognition Emotion 2009;23:209–37. https://doi.org/10.1080/02699930802204677.

[22] Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. J Behav Ther Exp Psy 1994;25:49–59. https://doi.org/10.1016/0005-7916(94)90063-9.

[23] Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res in Pers 1977;11:273–94. https://doi.org/10.1016/0092-6566(77)90037-X.

[24] Li X, Zhou H, Song S, Ran T, Fu X. The Reliability and Validity of the Chinese Version of Abbreviated PAD Emotion Scales. In: Tao J, Tan T, Picard RW (eds) Affective Computing and Intelligent Interaction. ACII 2005. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2005. https://doi.org/10.1007/11573548_66.

[25] Egermann H, Nagel F, Altenmüller E, Kopiez R. Continuous measurement of musically-induced emotion: a web experiment. Int J Internet Sci 2009;4:4–20.

[26] Schubert E. Continuous self-report methods. In: Juslin PN, Sloboda JA, editors. Handbook of music and emotion: Theory, research, applications. Oxford: Oxford University Press; 2010. p. 223–53. https://doi.org/10.1093/acprof:oso/9780199230143.003.0009.

[27] Schubert E. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. Aust J Psychol 1999;51:154–65. https://doi.org/10.1080/00049539908255353.

[28] Sharma K, Castellini C, Stulp F, van den Broek EL. Continuous, real-time emotion annotation: A novel joystick-based analysis framework. IEEE T Affect Comput 2020;11:78–84. https://doi.org/10.1109/TAFFC.2017.2772882.

[29] International Organization for Standardization. ISO/TS 12913–2:2018 Acoustics-Soundscape-Part 2: Data Collection and Reporting Requirements. Geneva: International Organization for Standardization; 2018.

[30] Tan JK, Hasegawa Y, Lau SK, Tang SK. The effects of visual landscape and traffic type on soundscape perception in high-rise residential estates of an urban city. Appl Acoust 2022:189. https://doi.org/10.1016/j.apacoust.2021.108580.

[31] Liu J, Kang J, Behm H, LuoT. Effects of landscape on soundscape perception: Soundwalks in city parks. Landsc Urban Plan 2013;123:30–40. https://doi.org/10.1016/j.landurbplan.2013.12.003.

[32] Sudarsono AS, Nitidara NP, Sarwono J. The relationship between sound source and urban soundscape. J Phys Conf Ser 2018:1075. https://doi.org/10.1088/1742-6596/1075/1/012033.

[33] Jo HI, JinYJ. Urban soundscape categorization based on individual recognition, perception, and assessment of sound environments. Landsc Urban Plan 2021:216. https://doi.org/10.1016/j.landurbplan.2021.104241.

[34] Zwicker E, Fastl H. Psychoacoustics—Facts and Models. Berlin: Springer; 1999. p. 203–64.

[35] Nagel F, Kopiez R, Grewe O, Altenmüller E. Emujoy : software for continuous measurement. Behav Res Methods 2007;39:283-290. https://doi.org/10.3758/bf03193159.

[36] Nagel F, Grewe O, Kopiez R, Altenmuller E. The relationship of psychophysiological responses and self-reported emotions while listening to music. Göttingen NWG Conference, 2005.

[37] Coutinho E, Dibben N. Emotions perceived in music and speech: relationships between psychoacoustic features, second-by-second subjective feelings of emotion and physiological responses. International Conference on Music & Emotion 2013. , https://livrepository.liverpool.ac.uk/id/eprint/3002876.

[38] Schubert E. Modeling perceived emotion with continuous musical features. Music Percep 2004;21:561–85. https://doi.org/10.1525/mp.2004.21.4.561.

[39] Fernández-Abascal E, Guerra P, Martínez-Sánchez F, Domínguez J, Muñoz M, Egea-Caparrós D, et al. The International Affective Digitized Sounds (IADS): Spanish norms. Psicothema 2008;20:104–13.

[40] Wang B, Kang J, Zhao W. Noise acceptance of acoustic sequences for indoor soundscape in transport hubs. J Acoust Soc Am 2020;147:206. https://doi.org/10.1121/10.0000567.