Applied Acoustics 199 (2022) 109039

Contents lists available at ScienceDirect

Applied Acoustics

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

The effect of foreground and background of soundscape sequence on emotion in urban open spaces

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 22 December 2021
Received in revised form 18 July 2022
Accepted 17 September 2022
Available online 1 October 2022

Keywords:
Urban soundscape
Acoustic sequence
Foreground sound
Background sound
Emotion

ABSTRACT

This paper discusses the influence of the soundscape sequence of different urban open spaces on emotion. Thirty participants with normal hearing were selected to listen to forty-two different acoustic sequences and report their emotional changes during the process. The data were analysed in four stages, and the results are as follows: *First*, emotional response highly correlates with background type. Only when the foreground is negative does it relate to the foreground type. *Second*, the positive foreground in the early part of a sequence, or the neutral (or negative) foreground in the later part of a sequence, induces a better emotional experience. *Third*, in an acoustic sequence, emotion changes along with a change in the foreground. The appearance of the foreground triggers emotional fluctuations, and the end of the foreground is followed by emotional recovery. *Finally*, combining foregrounds can aid in regulating negative emotions. This effect is related to the position of the positive foreground and background type. We offer suggestions on the design of urban soundscape from the perspective of emotion based on the findings.

1. Introduction

Soundscapes are a vital element of urban environment perception. As our urban environments become more complex and crowded, scholars feel the need to study soundscape perception more rigorously [1,2]. Kang pointed out that it is necessary to transform from traditional noise control to the perception of soundscape and to create a series of 'soundscape indexes' to quantify the perception of soundscape [3]. The current literature on soundscape perception mainly focuses on studying the perceived affective quality of soundscape [4]. According to iSO-12913–3, the perceived affective quality of soundscape can be measured by a two-dimensional model composed of two dimensions of pleasantness and eventfulness, and researchers have been working on building and improving this model [5]. In terms of the research on the main dimensions, Axelsson et al. proposed three main dimensions through the study of soundscape, including pleasantness and eventfulness, and established the first two independent dimensions for the perceived affective quality of soundscape, which were pleasantness and eventfulness [6]. More dimensions have also been extensively studied. For example, Cain et al. pointed out that calmness and vibrancy were also the principle dimensions [7]. Aletta et al. proposed that appropriateness is also a potential dimension that focuses on the relationship between the expectations and the real soundscape, which plays an important role in positive emotional responses [8]. Hall et al. pointed out that personal preference, history, and other social and cultural factors are also related to the perceived affective quality of soundscape [9]. In short, these dimensions of the perceived affective quality of soundscape reflect the ability to capture information related to survival through the acoustic environment, thereby choosing the environment and guiding behavior[10]. However, the perception of soundscape is multifaceted, and there is still a lack of research on the emotional perception of a soundscape.

The emotional perception of soundscape, different from the perceived affective quality of soundscape, is based on the emotion theory in psychology, focusing on the relationship between soundscape and perceived emotion. The commonly used evaluation dimensions mainly include pleasantness, arousal, and dominance [11]. It is more inclined to study positive and negative factors in soundscape perception [12]. It is shown that environmental sounds can trigger a wide range of emotional responses, and the International Affective Digitized Sounds (IADS), which consists of 167 natural sounds with a duration of 6 s, is the most typical example [13].

* Corresponding authors at: UCL Institute for Environmental Design and Engineering, The Bartlett, University College London (UCL), London WC1H 0NN, United Kingdom (Jian Kang); School of Architecture, Harbin Institute of Technology, NO. 66 Xi Da Zhi Street, Harbin, China (Qi Meng).
E-mail addresses: J.kang@ucl.ac.uk (J. Kang), mengq@hit.edu.cn (Q. Meng).

https://doi.org/10.1016/j.apacoust.2022.109039
0003-682X/

Choi pointed out that environmental sounds in the IADS database can induce an emotional response in the dimensions of pleasantness and arousal[14]. Ma and Thompson studied 24 environmental sounds and found that human emotions can track the changes in the acoustic properties of environmental sounds, just as they do for speech and music[15]. Webster studied longer acoustic stimuli, such as music, and pointed out that emotional perception of music is related to the interplay of texture and rhythm that make up the composition, rather than just acoustic properties [16]. However, environmental sounds are rarely designed to induce emotions, just like music. It usually takes the soundscape as different types of sound sources; even for the longer soundscape clip, it only considers the soundscape clip as a whole and pays little attention to its internal changes.

However, according to the definition of soundscape in ISO12913-1, a soundscape is the acoustic environment that is perceived, experienced, or understood by a person or people in context. The context includes the relationships between a person, activity, and place in space and time [17]. In other words, the perception of soundscape should not be invariable; instead, it can be modified with a change of context. Although some scholars have noticed this (e.g., Aumond et al. [18] used controlled and natural 3-min audio and audio-visual sequences to study the continuous pleasantness of the soundscape), they have primarily focused on the associated pleasure of soundscapes rather than on the emotional perception. Therefore, there is a lack of research on the relationship between the changing soundscape and the changing multi-dimension of emotional perception.

The soundscape sequence, which comprises different sound sources in different order, focuses on the variability of soundscapes. Nevertheless, the composition of the soundscape sequence in an urban open space is complicated, containing various sound sources that exist simultaneously. This complexity makes any study difficult, forcing scholars to simplify the objective.

According to the ISO/TS 12913–2, the foreground sound and background sound are the important indicators of the soundscape. *The foreground sound is the sound towards which a listener's attention is particularly directed and which can be associated with a specific source. The background sound is the sound that is heard continuously or frequently enough to form a background against which other sounds are perceived* [19]. Therefore, people can adjust their attention and direct it toward any sound they want to focus on, making it a foreground sound [20]. However, as attention is an uncontrolled element in the experiment design, the soundscape sequence design focuses on the sound source rather than on the attention. From this perspective, the background sound is a continuous listening context composed of multiple sounds. As the foreground sound is a specific sound, which may be related to a particular event, it may be more evident than the background sound at any given time for the listener. Taking the urban street space as an example, when someone walks along the street, the continuous background sound is the mixed low-frequency noise generated by the vehicles' engines. Although there are many other sound sources in the street, these may not be noticed by the listener. Furthermore, when someone passes a store along the street, their attention is easily attracted to sounds such as the advertisements from the store, which becomes the foreground sound. In other words, in an urban space, the background sound is a mixture of constant, multiple sound sources, while the foreground sound is a variable, single sound source corresponding to a specific event. With reference to previous research [21,22], we selected these two indicators to study the urban soundscape sequence. We created 'the single foreground acoustic sequence', which contains only one foreground and one background sound, to study the effect of their interaction on emotion. Further, we created 'the combined foreground acoustic sequence', which entails two different foreground sounds and one background sound, to study the effect of the interaction between the foreground sounds on emotion.

Another critical question is how to measure the changing emotion caused by sound sequence. Compared with the overall rating after the stimuli [23,24], the continuous emotion measurement based on the two-dimension theory can capture temporally-related self-reported data from participants using the software. Emotions are a common psychological phenomenon [25–27] that can be quantified through a categorical or dimensional approach. The former describes emotions with words, whereas the latter requires us to measure emotions by the linear combination of several dimensions [28]. The three-dimension and two-dimension theories are two fundamental emotional dimension models proposed by Schubert [11,29,30]. The three-dimension model evaluates the emotion through the dimensions of pleasantness, arousal, and dominance, whereas the two-dimension model only focuses on the two independent dimensions—pleasantness and arousal. The continuous emotional measurement is based on the two-dimensional model, mainly for the following two reasons: First, the two-dimension model can explain the emotional responses caused by most stimuli, such as pictorial [31,32], linguistic [33,34], and sound [35,36]. Second, the form of the two-dimensional theory is more suitable for presentation on the computer screen, then the difficulty in operating the software is reduced, and the accuracy of the data can be improved.

In the method of continuous emotion measurement based on the two-dimensional model, the pleasantness dimension is represented by the X axis and the arousal dimension is represented by the Y axis, and the two intersect perpendicularly at the origin of the coordinates; then, the numerical range is defined for the two dimensions. At each moment, the emotion corresponds to a specific X value and Y value. Subsequently, the measurement of emotion is transformed into a numerical value. Concurrently, the method also synchronises the audio and emotion measurement so that each moment of the audio corresponds well to the emotional data it stimulates. Afterwards, the continuous emotional data of the soundscape sequence can be obtained.

The method of continuous emotion measurement has been extensively used in music studies [37–39]. Schubert [40], for example, used two-dimensional emotion space to capture the continuous emotional changes caused by four musical pieces. Participants were asked to listen to four different pieces of music and to report their emotional changes while listening to them. The findings showed that the loudness, rhythm, melody contour, texture, and spectrum of music affect emotional change. Nagel, Kopiez, Grewe, and Altenmüller [41] used multiple stimuli, such as pictures and videos, to investigate the continuous emotional response using EMUJOY software.

Therefore, we used the method of continuous emotion measurement to study the influence of foreground and background sounds in a soundscape sequence of urban open space on emotion. *First*, we analysed the effect of foreground and background combinations on emotion and of different positions of the foreground in the sequence on emotion as well. *Then*, we explored whether the sudden appearance of a foreground can cause emotional fluctuations and whether the emotion recovers after the foreground is over. *Finally*, we used the combined foregrounds to regulate emotions.

2. Method

2.1. Selection and collection of the sound source

First, representative sound sources were selected for foreground and background sounds. Then, the sound source file was obtained

by recording or downloading, and finally, the foreground sound and background sound were mixed into an acoustic sequence using Cooledit software.

Regarding the selection of representative sound sources, studies have shown that different urban sound sources are significantly different in perception. For example, nature sounds usually make people feel pleased, which is positive for the soundscape. Conversely, technological sounds usually make people feel irritable, which negatively affects the soundscape [6,42]. Therefore, the representative background sounds were selected from the categories of positive, neutral, and negative, respectively. *Finally*, combining with the field survey of urban soundscapes, we selected three backgrounds, namely, 'water', 'voice', and 'traffic', representing the positive background, the neutral background, and the negative background.

In the selection of foreground sounds, the same classification criteria of positive, neutral, and negative were used. We selected the 'store', and 'mechanical' as the representative sounds of the neutral and negative foregrounds, respectively. We have further subdivided the selection of positive foreground because in the soundscape design, compared with the background, the foreground is easier to change, as shown in Fig. 1. However, due to the limitation of the time of the experiment, this study only took the positive foreground as an example, divided it into artificial and natural sounds, and chose 'music' and 'bird' as representative sounds [5].

We adopted both on-site recording and network downloading methods to collect sound sources. The on-site recording is more suitable for the background because it helps us record the spatial elements of the urban space more realistically [43]. The selection of the recording locations corresponded to the types of background, *namely*, waterfront park (water), commercial pedestrian street (voice), and the sidewalk next to the traffic arterial road (traffic noise). The 'voice' is the sound of the conversation made by the crowd, describing a noisy atmosphere in the central area of the city. The sound of the water is the regular rushing sound from the water slapping the bank, describing a tranquil atmosphere in the recreation area of the city. The recording time was from 9.00 a.m. to 11.00 a.m. on the weekend, ensuring a higher density of people. In each recording location, we selected a site where we could capture a 'pure' background. For example, if we wanted to get a 'pure' background of traffic, the recording site should only contain traffic sounds without any others. We placed the recorder at the height of 1.6 m and continuously recorded each background sound for 3–5 min to allow for editing later. Similarly, a pure foreground was downloaded from the Internet (*see* http://www.miaoyin365.com/), ensuring it was more than 1 min long; for example, the 'music' is a soothing fragment taken from a popular Chinese song.

2.2. Acoustic sequences

The acoustic sequence comprised a single foreground sequence and a combined foreground sequence. The former comprised one foreground and one background, while the latter comprised two different foregrounds ('bird' and 'mechanical' in this experiment) with different combinations and one background. The length of the sequence should stimulate an emotional response and avoid fatigue as much as possible [44,45]. Hence, we had set the background length of the single foreground sequence and the combined foreground sequence to 45 s and 60 s, respectively, while the foreground was always set to 15 s. We used Cooledit to make these sequences.

For the single foreground sequence, as shown in Table 1, we needed to make the foreground appear in different positions in the sequence. Taking the background of 'water' and the foreground of 'music' as examples, we first divided the background into three 15-second parts, the early part, the middle part, and the latter part, respectively. We inserted the foreground of 'music' separately into the early, middle, and latter parts of the background and mixed them into individual files. *Finally*, three sound sequences were obtained, all based on the background of water with the foreground of music appearing in the early part, the middle part, and the latter part of the sound sequence. In the same way, other types of foreground and background were also mixed, and we obtained 36 pieces of a single foreground sequence.

We needed to create different combinations of foreground sequences for the combined foreground sequence. We only studied the combination of 'bird' and 'mechanical' as examples. Since the foreground was two sound sources, the length was increased to 30 s, and each foreground was 15 s. As there were two different combinations of 'bird' and 'mechanical', we could put either in the front. While the length of the background increased to 60 s, the length of the early part or the latter part was still 15 s, and the middle part increased to 30 s. We put the combination of the

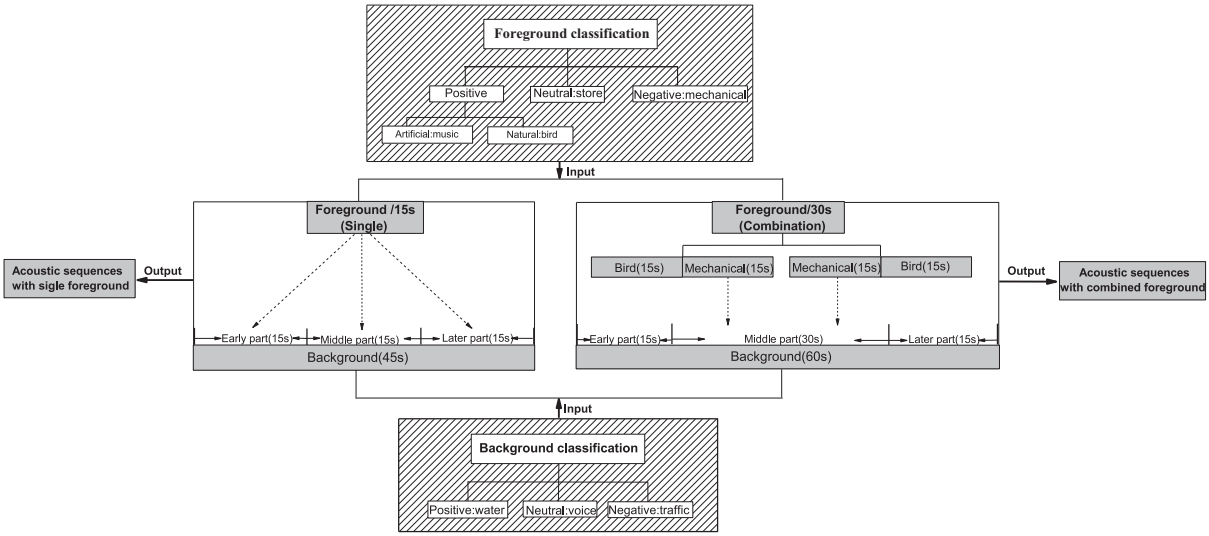

Fig. 1. Classification and creation of the acoustic sequence.

Table 1
The type of the combinations for the single foreground sequence.

Foreground type	Foreground position	Background type		
		B_{water}	B_{voice}	$B_{traffic}$
F_{music}	Early part(E)	$B_{water} + F_{music/E}$	$B_{voice} + F_{music/E}$	$B_{traffic} + F_{music/E}$
	Middle part(M)	$B_{water} + F_{music/M}$	$B_{voice} + F_{music/M}$	$B_{traffic} + F_{music/M}$
	Later part(L)	$B_{water} + F_{music/L}$	$B_{voice} + F_{music/L}$	$B_{traffic} + F_{music/L}$
F_{bird}	Early part(E)	$B_{water} + F_{bird/E}$	$B_{voice} + F_{bird/E}$	$B_{traffic} + F_{bird/E}$
	Middle part(M)	$B_{water} + F_{bird/M}$	$B_{voice} + F_{bird/M}$	$B_{traffic} + F_{bird/M}$
	Later part(L)	$B_{water} + F_{bird/L}$	$B_{voice} + F_{bird/L}$	$B_{traffic} + F_{bird/L}$
F_{store}	Early part(E)	$B_{water} + F_{store/E}$	$B_{voice} + F_{store/E}$	$B_{traffic}++F_{store/E}$
	Middle part(M)	$B_{water} + F_{store/M}$	$B_{voice} + F_{store/M}$	$B_{traffic} + F_{store/M}$
	Later part(L)	$B_{water} + F_{store/L}$	$B_{voice} + F_{store/L}$	$B_{traffic} + F_{store/L}$
$F_{mechanical}$	Early part(E)	$B_{water} + F_{mechanical/E}$	$B_{voice} + F_{mechanical/E}$	$B_{traffic}++F_{mechanical/E}$
	Middle part(M)	$B_{water} + F_{mechanical/M}$	$B_{voice} + F_{mechanical/M}$	$B_{traffic} + F_{mechanical/M}$
	Later part(L)	$B_{water} + F_{mechanical/L}$	$B_{voice} + F_{mechanical/L}$	$B_{traffic} + F_{mechanical/L}$

Note: 'B' represents the background, 'F' represents the foreground.

Table 2
Type of the combinations of two-foreground sound sequence.

The type of the background	The type of the combinations of the foreground	
	$F_{bird} + F_{mechanical}$	$F_{mechanical} + F_{bird}$
B_{water}	$B_{water} + F_{bird} + F_{mechanical}$	$B_{water} + F_{mechanical} + F_{bird}$
B_{voice}	$B_{voicer} + F_{bird} + F_{mechanical}$	$B_{voice} + F_{mechanical} + F_{bird}$
$B_{traffic}$	$B_{traffic} + F_{bird} + F_{mechanical}$	$B_{traffic} + F_{mechanical} + F_{bird}$

Note: 'B' represents the background, 'F' represents the foreground.

foreground only in the middle part of the background. *Finally*, 6 pieces of combined foreground sequence were obtained (Table 2).

2.3. Software

The EMUJOY software developed by Nagel et al. [41] was used to record the continuous emotional data in the experiment. The software can record emotional data while playing audio and is remotely controlled. The monitoring computer can choose to play or stop the audio, while the participants' computer can only be used to record emotional data. This can ensure that participants are not disturbed during the whole process.

As shown in Fig. 2, the software interface shown to the participants is composed of two axes (representing two dimensions of emotion) intersecting at right angles, where the ***X***-axis represents the pleasantness dimension, and the ***Y***-axis represents the arousal dimension. The left side of the ***X***-axis represents negative emotion ('displeasure'), and the right side represents positive emotion ('pleasure'), with the value changing from −1 to 1. The bottom of the ***Y***-axis represents calming ('low arousal'), and the top represents arousing ('high arousal'), with the value changing from −1 to 1. When the audio plays, the participant can move the mouse on the screen and confirm the point in the emotional two-dimension space during the process at any time by clicking the left button of the mouse. Subsequently, the corresponding data with pleasantness dimension value (X) and arousal dimension value (Y), as well as the time, are recorded. Participants can click the mouse at any moment when they feel emotional changes, and there is no limit to the number of times. The data is sent to the monitoring computer online as text files. The output includes values for both axes, time, and an additional trigger track. This track contains the 'start' and 'stop' of the audio. The sampling rating is 50 ms, which allowed for signal frequencies up to 10 Hz. Previous studies have shown its validity [41,46,47].

2.4. Participants

Thirty samples can meet the requirements of the acoustic sequence listening test in the laboratory [45]. In our study, we

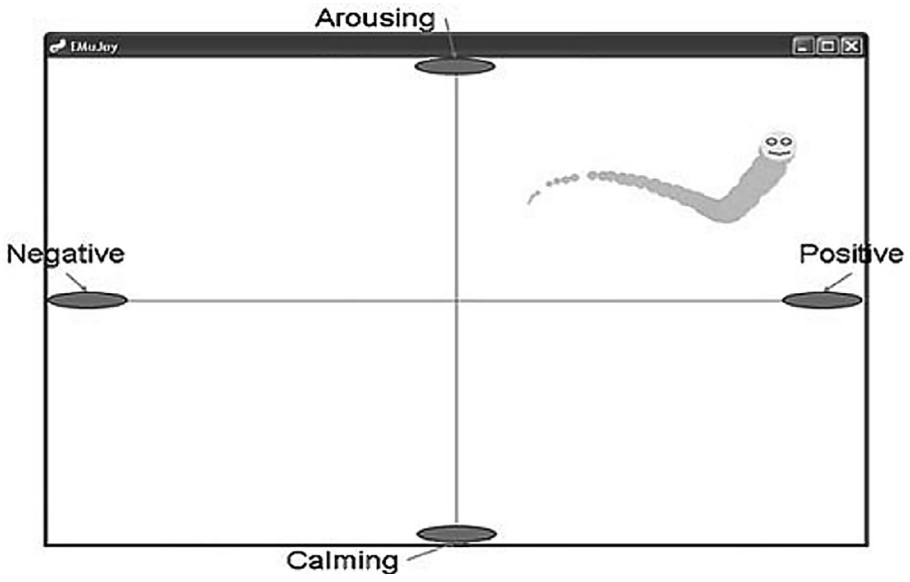

Fig. 2. Software interface of participant [41].

specifically selected 15 women and 17 men. However, two women did not complete the experiment for personal reasons, giving us an effective sample size of 30. All the participants' hearing was normal as per the hearing test results (*see* https://www.widex.com.cn/zh-cn/online-hearing-test). The details of the test are described in the section of Procedure. As emotions can vary significantly among people of different ages and younger adults tend to have a favourable perception of emotional experiences, we selected younger adults as participants (i.e., 25–35 years old) [48]. Their median age was 30 years, and the standard deviation was 4.5. Their occupations included students, office workers, and freelancers. Two of them were engaged in music-related jobs at the time, with a median working life of five years.

2.5. Procedure

As shown in Fig. 3, the experiment process consists of three stages: selecting, training, and formal experiment.

After obtaining informed consent from the participants, they were provided with an oral and written description of the experiment. After all the questions on the experiment were addressed, the participants could start the first stage. *First*, the selection stage was a hearing test to ensure the participants' normal hearing. It was a five-minute online listening test, including multiple-choice, high-frequency listening, and language listening tests (*see* https://www.widex.com.cn/zh-cn/online-hearing-test).

Second, in the training stage, we briefly explained the emotional two-dimension model to the participants and cleared any doubts. To familiarise participants with the emotional space and avoid priming effects, we used pictures from the international affective picture system for practice. The pictures were displayed in the background of the emotion space, and the participants learned to use the software by expressing their perceived emotions while looking at the pictures. After they could express themselves competently using the software, the formal experiment began.

Third, in the formal experiment stage, the participants sat in the experiment room with a length, width and height of 4 m, 5 m and 2.8 m, respectively. According to the measurement, it is a music room with a background sound level of 35dBA and a reverberation time of 0.5 s at middle frequency of 500 Hz-1000 Hz. Referring to the previous literature [8], the experimental conditions can meet the requirements of this research. Before the experiment, the participants were required to read a protocol, which indicated they were required listening each audio and report how the sounds perceived, but not how they felt due to listening of sounds. They should report the emotional changes at any time while listening to the audio, with no limit in the number of times. Then, the participants need to sit quietly in front of the computer for several minutes to adjust their emotions to a neutral state which means the emotional dimensions of pleasantness and arousal are all at the coordinate origin in the emotional space (X = 0, Y = 0). After confirming the adjustment complete, the experiment begins. To avoid fatigue, the material was divided into two sections and played to the participants at an interval of one week. The first part of the listening test included 20 pieces of audio and took 20 min; the second part played the remaining 16 pieces of audio, taking 25 min.

Finally, regarding the software, EMUJOY runs on a Lenovo computer, and the headphones used were Sennheiser RS170.

2.6. Data analysis

Software IBM SPSS 25.0 and Origin 8 were used for data analysis and figure rendering. SPSS is a general statistical analysis software; it was mainly used to analyse the mean, standard deviation, the paired-samples *t*-test, and Wilcoxon rank-sum test here, whereas Origin 8 was mainly used to draw time-varying figures of the emotional changes of the sequence.

3. Results

3.1. Effect of different foreground and background combinations on emotion

Using the data of mean and standard deviation from the single foreground sequence, we extracted the emotional data in response to when the foreground and background existed simultaneously. This helped us analyse the effect of different combinations of foreground and background on emotion (see Fig. 4). We now discuss our results.

In a sequence without the 'mechanical' foreground, the emotional response varies by background type. The values of the two dimensions of emotion are positive in the background of 'water', negative in the background of 'voice', and both positive and negative in the background of 'traffic', regardless of the foreground. However, different foregrounds can still create a significant difference in pleasantness and an insignificant difference in arousal in these three backgrounds. Moreover, the 'music' foreground can cause a relatively better emotional experience than other foregrounds, similar to the results reported in previous studies [42]. We further found that this positive effect of 'music' is the most obvious in the background of 'traffic' because the positive value in pleasantness can be achieved only in this condition (Pleasantness $_{MD}$ = 0.04, Arousal $_{MD}$ = −0.02).

However, in a sound sequence with the 'mechanical' foreground, irrespective of the background, the emotional response is always at the bottom left of the emotional space, indicating the worst emotional experience, with a slight difference in pleasantness (ranging from −0.40 to −0.32) and an obvious difference in arousal (ranging from −0.23 to −0.04). In this condition, the foreground, rather than the background, plays a decisive role in the emotional response.

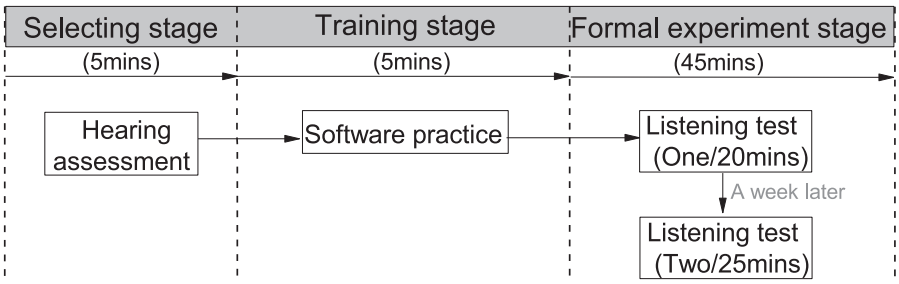

Fig. 3. Experiment process.

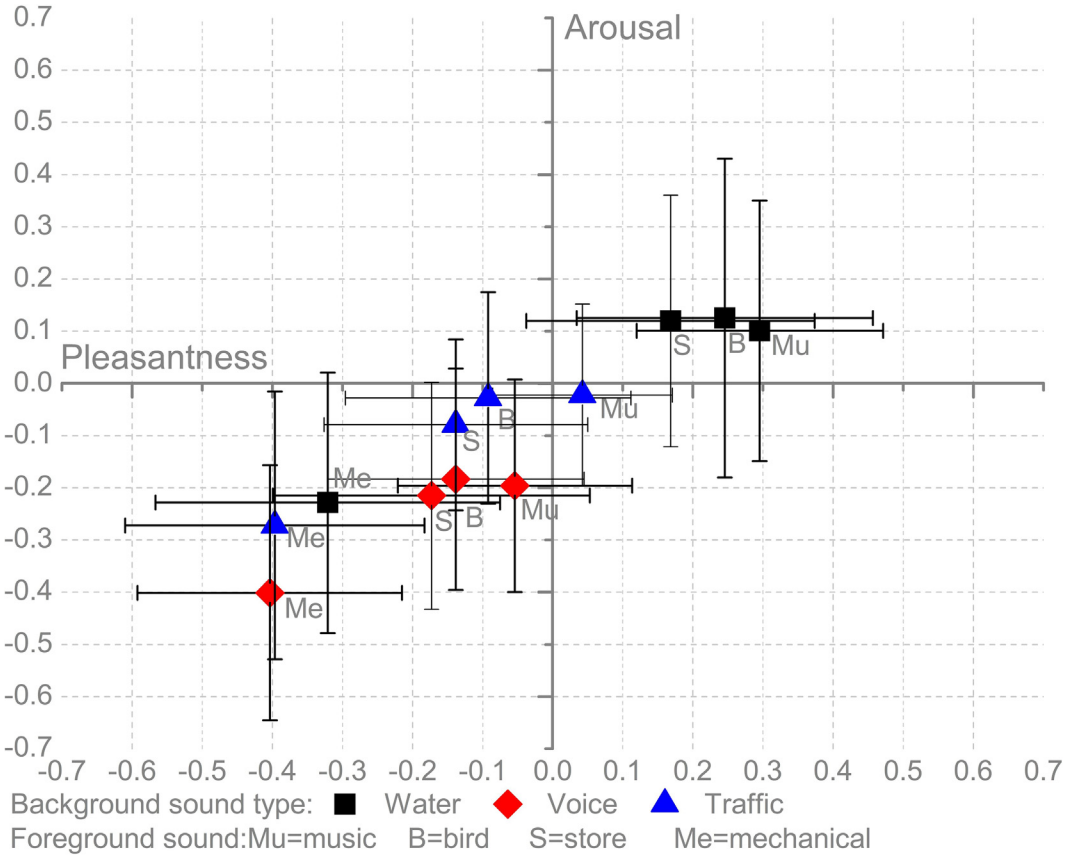

Fig. 4. Different acoustic sequences with respect to emotion (The points in the figure represent the average value, and the range represents the standard deviation of X or Y).

In summary, we find a high correlation between emotional response and background type. However, this correlation for foreground type only exists if the foreground is negative, such as a 'mechanical' sound. The emotional experience in the 'water' background is superior to that of 'traffic', which, in turn, is superior to 'voice'.

Compared with the foreground, the background has a greater effect on emotions. The background type for an already existing urban space, such as streets, cannot be changed without large-scale construction, but the designer can adjust the type of foreground. We further analyse the relationship between a controllable foreground and emotion, which may have more practical significance.

3.2. Effect of different foreground positions on emotion

Using the data of mean and standard deviation, we extract the emotional data in response to when the foreground is located at different positions (the early, middle, and later parts) in the sequence (see Fig. 5). This way, we can analyse the effect of position on emotion.

We find that, as 'music' moves ahead in a sequence, the values of pleasantness and arousal decrease, indicating that the overall emotional experience improves. The 'bird' foreground in the early part of the sequence ($Pleasantness_{SD}$ = 0.04, $Arousal_{SD}$ = 0.09) induces a significantly better emotional experience than in the middle ($Pleasantness_{SD}$ = -0.02, $Arousal_{SD}$ = -0.01) or later ($Pleasantness_{SD}$ = 0, $Arousal_{SD}$ = -0.02) parts. The 'store' foreground shows little emotional difference between the later ($Pleasantness_{SD}$ = -0.02, $Arousal_{SD}$ = 0) or early ($Pleasantness_{SD}$ = -0.03, $Arousal_{SD}$ = 0) parts, but these values are better than those for the middle part ($Pleasantness_{SD}$ = -0.05, $Arousal_{SD}$ = -0.06).

As the 'mechanical' foreground moves next in the sequence, the values of both pleasantness and arousal increase, indicating improved emotional experience. The emotional experience of this foreground in the later part ($Pleasantness_{SD}$ = -0.10, $Arousal_{SD}$ = -0.17) is significantly better than that in the early ($Pleasantness_{SD}$ = -0.25, $Arousal_{SD}$ = -0.36) or middle ($Pleasantness_{SD}$ = -0.23, $Arousal_{SD}$ = -0.31) parts.

In general, the positive foreground in the early part of a sequence, or the neutral (or negative) foreground in the later part of a sequence, induces a better emotional experience.

3.3. Single foreground acoustic sequence and the dynamic changes of emotion

We now explore if the sudden appearance of the foreground in the background will cause an emotional fluctuation and how this happens. We also determine if and how emotion recovers within a period after the foreground is over, with only the background left.

3.3.1. Acoustic sequence and emotion fluctuation

The emotional data in response to when the foreground is in the middle of the sequence are summarized to compare the emotional fluctuations from the foreground's appearance. The curve within 15–30 s in Fig. 6 is the fluctuation curve (D_{30-15} represents the difference value between the 30 s and 15 s).

A comparison of the fluctuation curves shows that the appearance of 'music' can significantly elevate emotional experience, regardless of the background. Especially when 'music' and 'water'

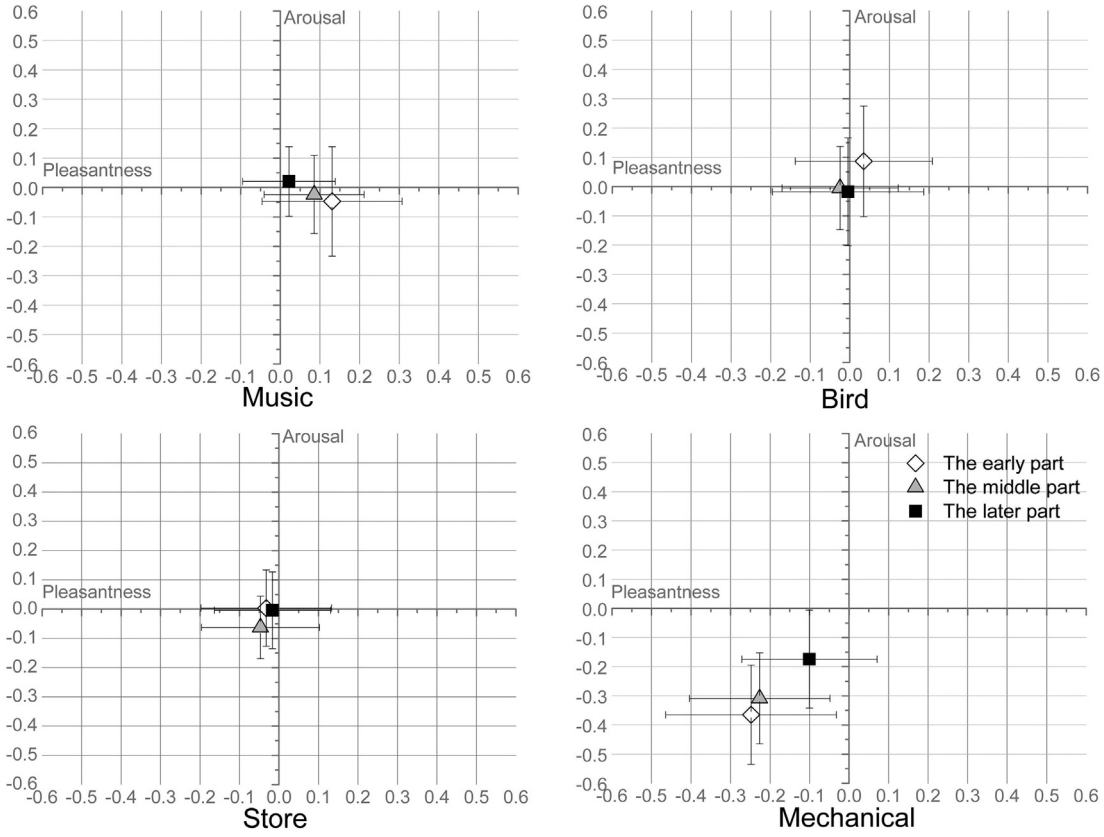

Fig. 5. Different positions of the foreground with respect to emotion.

are combined, the maximum increase in pleasantness can reach + 0.22 owing to the two positive sounds. The appearance of 'mechanical' can worsen the emotional experience, irrespective of the background. Compared with other foregrounds, this decline in pleasantness ($D_{30\text{-}15}$ = -0.76) and arousal ($D_{30\text{-}15}$ = -0.8) is the strongest. The 'store' foreground can cause a smaller decrease in both pleasantness ($D_{30\text{-}15}$ = -0.15) and arousal ($D_{30\text{-}15}$ = -0.28) compared with the 'mechanical' foreground. Moreover, the 'bird' in the background of 'voice' or 'traffic' causes weak emotional fluctuation.

Our results also show that irrespective of which foreground appears in the background of 'water', we observe a large fluctuation in emotion ($D_{30\text{-}15}$ in pleasantness ranges from –0.48 to 0.20, $D_{30\text{-}15}$ in arousal ranges from –0.38 to 0.02). In the background of 'voice', this fluctuation is the smallest ($D_{30\text{-}15}$ in pleasantness ranges from −0.48 to 0.15, $D_{30\text{-}15}$ in arousal ranges from −0.45 to –0.04).

In general, when the positive foreground appears in the positive background, the rise in emotion is highest. When the negative foreground appears in the positive background, rather than a negative one, the decline in emotion is the highest, which is an interesting finding. Finally, the foreground that appears in a positive background causes high emotional fluctuation, compared with the other types of background.

3.3.2. Acoustic sequence and emotion recovery

The emotional data in response to when the foreground is in the early sequence are summarised to obtain a longer recovery curve and, thus, compare the emotional recovery after the foreground is over. The curve within 15–45 s is the recovery curve, as shown in Fig. 7.

A comparison of the arousal recovery curve shows an upward trend in the period after the foreground is over. We find a sharp increase within 15–30 s and a steady increase within 30–45 s. A comparison with the value at 45 s shows that the value in the background of 'water' is always larger than it is in the background of 'traffic', which, in turn, is larger than in 'voice' (except for the recovery curve after the 'store' foreground). This result implies a high correlation between the recovery of arousal and the background type.

However, when we compare the pleasantness recovery curve, we see that it varies by foreground. After the 'music' finishes, the value of pleasantness in the background of 'water' is always higher than it is in the 'traffic', which, in turn, is higher than in 'voice', which is the same for arousal.

Once the 'bird' and 'store' foreground are over, the pleasantness curve increases for the background of 'water' but is stable for 'voice' and 'traffic', which is almost coincidental. Paired-samples *t*-test analysis also confirmed this (*see Pair 1* and *Pair 2* in Table 3).

After the 'mechanical' foreground is over, the pleasantness rises more rapidly for the background of 'water' compared with the others. However, the curves also rise and partially coincide for 'voice' and 'traffic'. The rank-sum test indicates that the recovery between 15 and 30 s has no statistical differences (*see Pair 4* in Table 3).

The results also indicate that, for the 'voice' and 'traffic' backgrounds, the pleasantness recovery after 'bird' is always better than that of 'store', which is, in turn, even better than 'mechanical'. Therefore, in these two backgrounds, the recovery of pleasantness correlates to the foreground type.

In general, the recovery of arousal correlates with the background type, showing that the value in the positive background is always higher than that in the negative background, which is, in turn, higher than in the neutral one. However, the recovery of pleasantness has the highest value in the positive background and is correlated with the foreground type for the neutral or negative background. The more positive the foreground is, the better pleasantness recovery will be in the neutral or negative background.

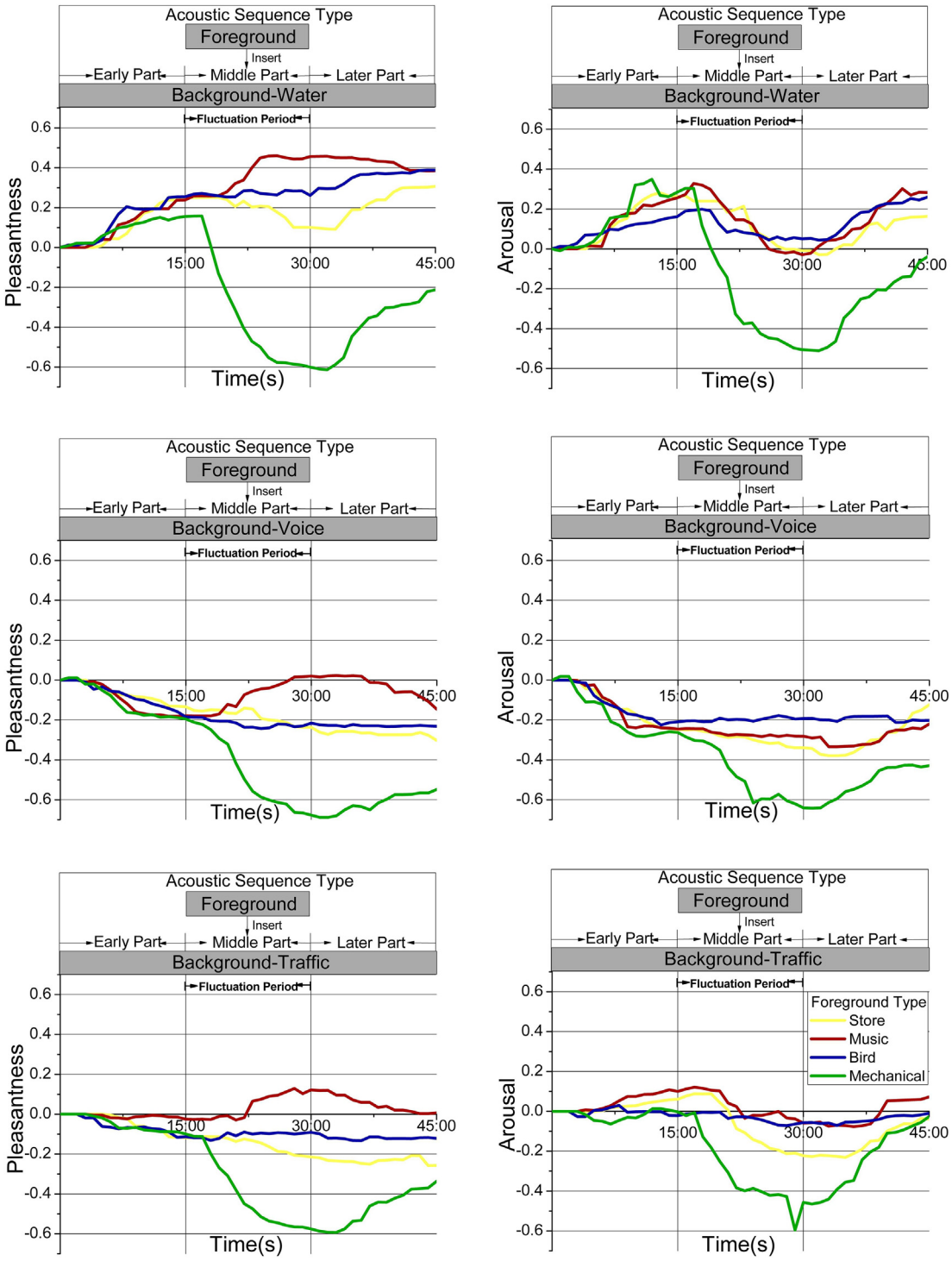

Fig. 6. Acoustic sequence and emotional fluctuation. Notes: The red/ yellow/ blue/ green line represented the foreground sound of "music"/ "store" / "bird" / "mechanical".

3.4. Effect of combined foregrounds on regulating emotion

We now take the 'mechanical' foreground, which causes negative emotions and observe its effect on regulating emotions. We do so by placing the positive foreground ('bird') before and after it. Fig. 8 shows lines of three colours; the grey line is the baseline for comparison, and we derive it from the emotional data obtained when the 'mechanical' foreground appears in the middle of the single foreground sequence. We intercept the data within 15–60 s and 0–45 s for the red line and green line, respectively, from the entire 60 s of the combined foreground sequence for mapping to obtain a more intuitive comparison.

Comparing the red line and baseline, we find that adding the 'bird' in front of the 'mechanical' foreground has a weak to no effect in regulating negative emotions in the background of 'water' within 15–30 s, and a long-lasting effect in relieving negative emotions generated by the 'mechanical' foreground in the background of 'voice' (excluding some special periods) and 'traffic' within 15–45 s. When comparing the green line and baseline within 30–45 s, we find that adding the 'bird' behind the

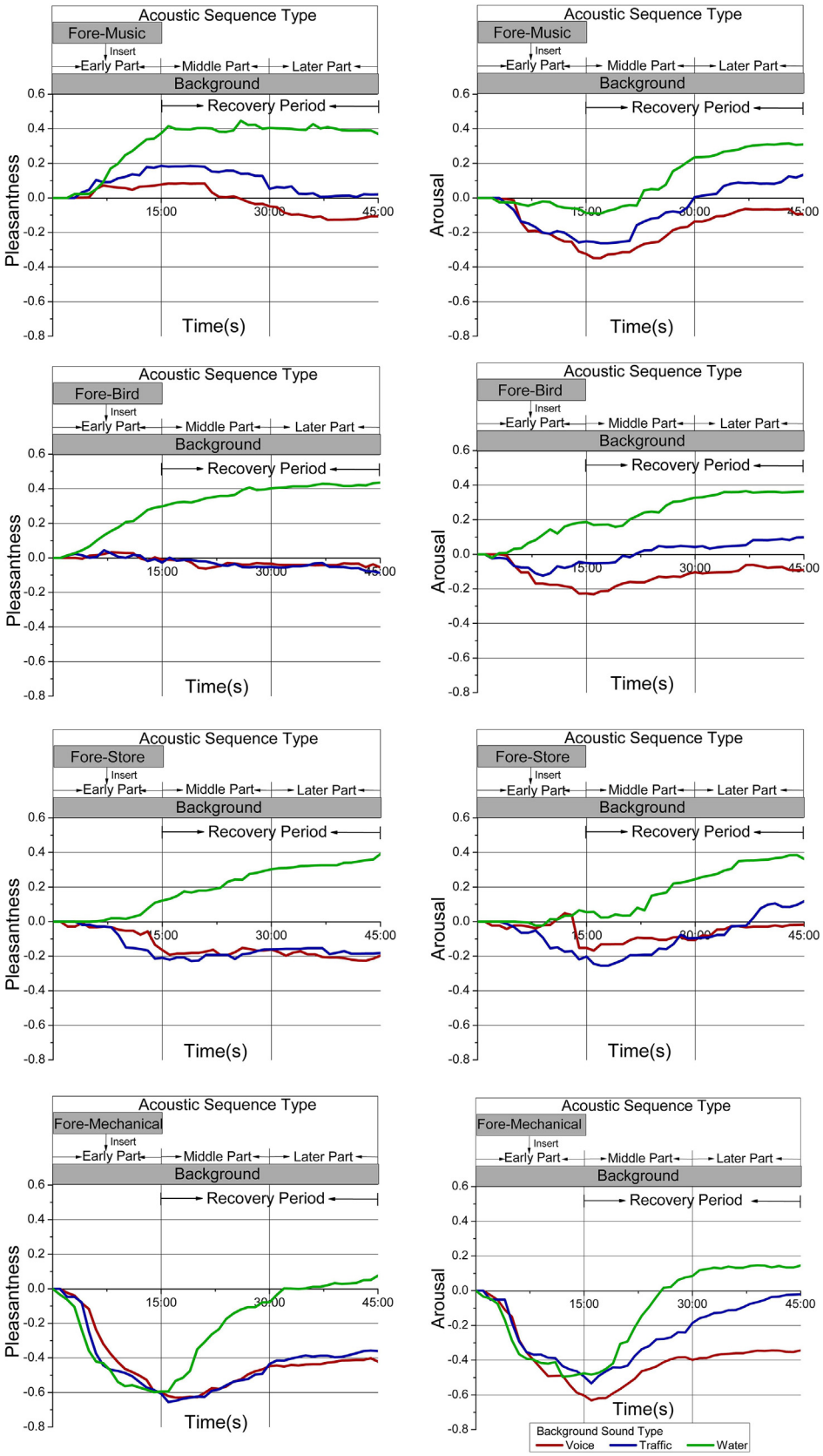

Fig. 7. Acoustic sequence and emotion recovery. Notes: "Fore" stands for foreground sound, the red / blue/ green line represented the background sound of "voice" / "traffic" / "water".

Table 3
Correlation of pleasantness recovery for paired samples.

Acoustic sequence pair	Pair 1 (Bird + voice/ bird + traffic)	Pair 2 (store + voice/ store + traffic)	Pair 3 (mechanical + voice/ mechanical + traffic)	Pair 4 (mechanical + voice/ mechanical + traffic)
Period	15–45 s	15–45 s	15–45 s	15–30 s
p	0.06	0.55	0.01	0.06

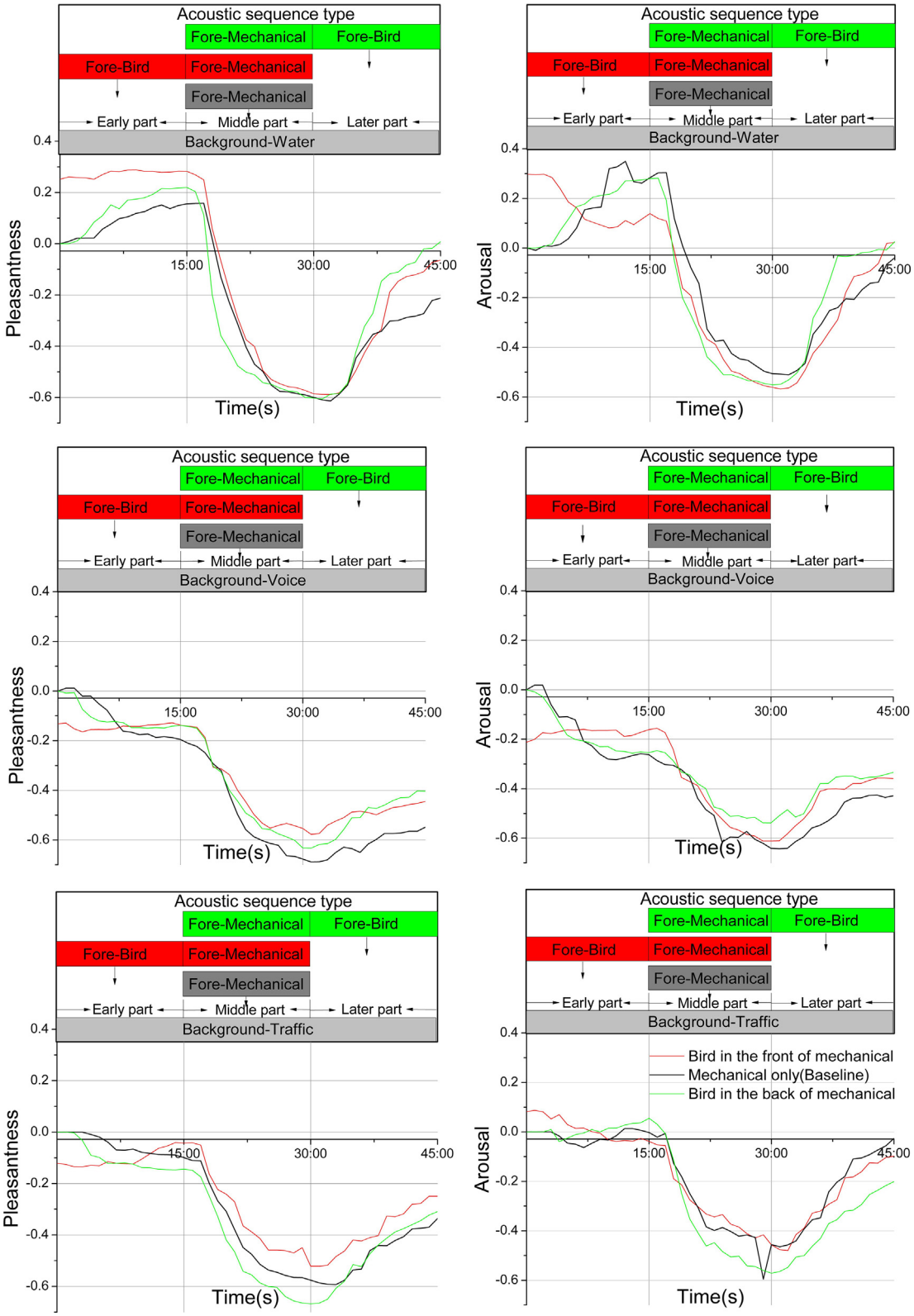

Fig. 8. Combination of foreground acoustic sequences and emotion regulation, Notes: Fore = foreground sound.

'mechanical' foreground can accelerate emotional recovery in the background of 'water' and 'voice', which is more obvious for pleasantness but is complicated in the background of 'traffic'.

We can thus conclude that a positive foreground can regulate negative emotions. This effect correlates with the position of the positive foreground and the background type. A positive foreground in front of a negative one can also relieve negative emotions with a long-lasting effect. This effect is stronger for neutral and negative backgrounds but weaker for positive backgrounds. Moreover, a positive foreground behind a negative one accelerates emotional recovery, and this effect is stronger for positive backgrounds but more complicated for negative backgrounds. More detailed research is necessary to fully uncover the relationship between acoustic sequence and emotional regulation.

4. Discussion

In general, the type of background plays a decisive role in the emotional perception of the soundscape. For example, the positive background, 'water', can cause a more positive emotional response, which coincides with the literature on the preference for the soundscape [42,49]. When we add a negative foreground to a positive background, the emotional fluctuation is the largest, and the emotional recovery is the fastest after the foreground is over. When we further add a positive foreground behind the negative one to regulate emotion, we obtain the best effect. However, the emotional response to the neutral background (voice) is lower than that to the negative one (traffic). On the contrary, the evaluation of the preference for or annoyance toward the soundscape implies that 'traffic' has a lower preference but a higher annoyance score compared with 'store' [2,49,50]. These differences also prove that the emotional perception of a soundscape sequence is quite an interesting field of research.

When we compare the emotional perception of the foreground with the mean value, we find that the more positive the foreground, the better the emotional perception. However, 'bird', as a positive foreground with a high mean value, did not cause an obvious emotional fluctuation, *that is*, the 'bird' foreground was 'swallowed' in the emotional perception of the sound sequence with the 'voice' and 'traffic' background. This swallowing phenomenon of the sound source in the sequence indicates that it is not enough to analyse the sound sequence only by the mean value. More detailed data, such as continuous emotion data, may help us better understand the complexity of the sound sequence. However, for this kind of continuous-time data, a better analysis method needs to be introduced. Finally, the forward positive foreground and the backward negative foreground allow better emotional perception, but the positive foreground placed in the front of the sequence also has a long-lasting effect on emotion regulation. The relationship between position and the effect time of the foreground should thus be further explored.

This research is crucial because it focuses on the sound sequence in urban soundscape design from the perspective of emotional perception. The emotional perception of the background in a soundscape sequence emphasises the importance of considering soundscape design at the first stage of urban space design. This is the most effective and economical approach. For example, to create a more positive urban park background, setting the park in a small area surrounded by urban arterial roads (which may also mean less greenery to shield traffic noise) may be a bad choice. For spaces with a positive background, negative stimuli can cause a strong negative effect; hence, avoiding negative stimuli, but adding more positive foregrounds, is the most effective strategy. It is also effective to regulate emotional perception for a negative foreground or background by adding more positive foregrounds. However, the selection of the foreground is critical. For example, 'music' seems to be more effective for traffic space.

5. Conclusion

This paper discusses the influence of the urban soundscape sequence on emotions and suggests some recommendations for an urban soundscape design based on the following results:

- Compared with the foreground, the background has a greater effect on emotions.The type of background sound can determine the position of emotional experience in the emotional space. The background of water can get a better emotional experience in the upper right corner in the emotional space with the value of the two dimensions of emotion are all positive(the value of pleasantness range from 0.17 to 0.30, the value of arousal range from 0.10 to 0.13),while the background of traffic or voice can get a poor emotional experience at the bottom left in the emotional space with the value of the dimensions are partial or total negative (the value of pleasantness range from −0.17 to 0.09, the value of arousal range from −0.22 to −0.02).
- The position of the foreground in the acoustic sequence triggers varied effects on emotion. A positive foreground in the early acoustic sequence, or a negative foreground in the later acoustic sequence, yields better emotional experience, that is the emotion will tend to move to the upper right corner in the emotional space with the value of the two emotional dimensions becomes higher.
- In the background, the emotion changes with the change in the foreground. The appearance of a foreground causes emotional fluctuations, while its end is followed by emotional recovery. Specifically, a positive foreground in a positive background induces positive emotions, while a negative foreground in a positive background induces negative emotions. A positive background with any foreground type causes great emotional fluctuation. Once a foreground ends, the emotional recovery of arousal depends on the background type, while the recovery of pleasantness depends on the foreground type only for a neutral or a negative background. Thus, the more positive the foreground is, the better the emotional recovery following the neutral or negative background.
- Combining foregrounds can effectively regulate emotions, but this effect is not only related to the positive foreground's position but also to the background type.

Our results showed that the sound sequence with the 'traffic' background sound had neither positive nor negative effects on the emotion. Conversely, for a space with a poor sound sequence, such as the road and its surrounding areas, it can improve the general emotional perception of the sound sequence by adding a positive foreground sound. The earlier position of foreground sound in the sequence, the better improving effect it has on emotion. However, future research is needed on the questions about the appropriately introduced sum of the positive foreground sounds and the time interval between different foreground sounds in the introduction process. In addition, if a bad sound source, such as the mechanical noise, appears in a sound sequence, it is also helpful to introduce a positive foreground sound around it to improve the overall emotional perception of the sequence if the bad sound source cannot be controlled. However, for an urban space with diverse functions, like a commercial pedestrian street, the interac-

tion between people and space may be the origin of the positive aspects of the emotion, not just the sound. Therefore, the research on the emotional perception of urban soundscape sequences in the real multisensory environment is worthy of further exploration.

There are also some limitations in this research. Theoretically, the emotional changes reported by the subjects in the software were caused only by the sounds they heard and not by other factors. There is a risk in asking people to do so because it is not a natural task. Although some scholars use the physiological indicators to evaluate emotional changes indirectly, the self-reports, as a direct way, is still considered appropriate. On the other hand, the emotions measured in this experiment are the perceived emotions, not the elicited emotions. There is a fundamental difference between them, the former emphasizes the perception of emotions represented by sounds, but the latter focuses on the emotions that the sound induces or arouses in their mind [12].This study can only explain the results in the ideal situation with only soundscape, and more research is needed in the real multisensory environment. Finally, the results of this study are only representative of young people. For people of other age groups, more studies are still needed. Simultaneously, more studies are needed for the urban sound sequence, which is longer and more complex.

CRediT authorship contribution statement

Zhihui Han: Methodology, Formal analysis, Writing – original draft. **Jian Kang:** Conceptualization, Funding acquisition, Resources. **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.

Acknowledgements

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

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) [grant numbers 51878210, 51778169], the Natural Science Foundation of Heilongjiang Province [YQ2019E022] and the European Research Council (ERC) Advanced Grant (no. 740696) on 'Soundscape Indices' (SSID).

References

[1] Botteldooren D, Boes M, Oldoni D, De Coensel BD. The role of paying attention to sounds in soundscape perception. J Acoust Soc Am 2012;131(4):3382. https://doi.org/10.1121/1.4708755.
[2] Jeon JY, Lee PJ, You J, Kang J. Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. J Acoust Soc Am 2010;127(3):1357–66. https://doi.org/10.1121/1.3298437.
[3] Brown AL, Kang J, Gjestland T. Towards standardization in soundscape preference assessment. Appl Acoust 2011;72(6):387–92. https://doi.org/10.1016/j.apacoust.2011.01.001.
[4] 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.
[5] . ISO/TS 2019;12913–3:2019.
[6] Axelsson Ö, Nilsson ME, Berglund B. A principal components model of soundscape perception. J Acoust Soc Am 2010;128(5):2836–46. https://doi.org/10.1121/1.3493436.
[7] Cain R, Jennings P, Poxon J. The development and application of the emotional dimensions of a soundscape[J]. Appl Acoust 2013;74(2):232–9. https://doi.org/10.1016/j.apacoust.2011.11.006.
[8] Aletta F, Kang J, Axelsson Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landsc Urban Plan 2016;149:65–74. https://doi.org/10.1016/j.landurbplan.2016.02.001.
[9] Hall DA, Irwin A, Edmondson-Jones M, Phillips S, Poxon JEW. An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. Appl Acoust 2013;74(2):248–54. https://doi.org/10.1016/j.apacoust.2011.03.006.
[10] van den Bosch KAM, Welch D, Andringa TC, Axelsson S. The evolution of soundscape appraisal through enactive cognition the evolution of soundscape appraisal through enactive cognition. Front Psychol 2018;9:1129. https://doi.org/10.3389/fpsyg.2018.01129.
[11] Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res Pers 1977;11(3):273–94. https://doi.org/10.1016/0092-6566(77)90037-X.
[12] Masullo M, Maffei L, Iachini T, Cioffi F, Ruotolo F. A questionnaire investigating the emotional salience of sounds. App Acoust 2021;182:3–4. https://doi.org/10.1016/j.apacoust.2021.108281.
[13] Bradley M, Lang P. The International Affective Digitized Sounds: Affective Ratings of Sounds and Instruction Manual. University of Florida; 2007.
[14] Choi Y, Lee S, Choi IM, Jung S, Park YK, Kim C. International affective digitized sounds in Korea: a cross-cultural adaptation and validation study. Acta Acust united Ac 2015;101(1):134–44. https://doi.org/10.3813/AAA.918811.
[15] Weiyi M, Thompson WF. Human emotions track changes in the acoustic environment. Proc Natl Acad Sci 2015;112(47):14563–8. https://doi.org/10.1073/pnas.1515087112.
[16] Webster GD, Weir CG. Emotional responses to music: interactive effects of mode, texture, and tempo. Motiv Emot 2005;29(1):19–39. https://doi.org/10.1007/s11031-005-4414-0.
[17] . ISO/TS 2014;12913–1:2014.
[18] 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(2):144. https://doi.org/10.3390/app7020144.
[19] . ISO/TS 2018;12913–2:2018.
[20] Truax B. Acoustic communication. Comput Music J 2000;114(5):2528–9. https://doi.org/10.1007/978-3-642-76220-8.
[21] Bolin K, Kedhammar A, Nilsson ME. The influence of background sounds on loudness and annoyance of wind turbine noise. Acta Acust united Acust 2012;98(5):741–78. https://doi.org/10.3813/AAA.918555.
[22] Lebiedowska B. Acoustic background and transport noise in urbanised areas: A note on the relative classification of the city soundscape. Transport Res Part D 2005;10(4):341–435. https://doi.org/10.1016/j.trd.2005.03.001.
[23] Kang J, Zhang M. Semantic differential analysis of the soundscape in urban open public spaces. Build Environ 2010;45(1):150–217. https://doi.org/10.1016/j.buildenv.2009.05.014.
[24] Kidd GR, Watson CS. The perceptual dimensionality of environmental sounds. Noise Control Eng J 2003;51(4):216. https://doi.org/10.3397/1.2839717.
[25] Lazarus RS. Emotion and adaptation. Oxford: Oxford University Press; 1991.
[26] Ledoux JE. Emotion circuits in the brain. Annu Rev Neurosci 2000;23:155–84. https://doi.org/10.1146/annurev.neuro.23.1.155.
[27] Ochsner KN, Gross JJ. The cognitive control of emotion. Trends Cogn Sci 2005;9(5):242–329. https://doi.org/10.1016/j.tics.2005.03.010.
[28] Mauss IB, Robinson MD. Measures of emotion: A review. Cognit Emot 2009;23(2):209–37. https://doi.org/10.1080/02699930802204677.
[29] Mehrabian A, Russell JA. An approach to environmental psychology. Cambridge, MA: MIT Press; 1974. p. 1977.
[30] Russell JA. Core affect and the psychological construction of emotion. Psychol Rev 2003;110(1):145–72. https://doi.org/10.1037/0033-295x.110.1.145.
[31] Bradley MM, Cuthbert BN, Lang PJ. Picture media and emotion: Effects of a sustained affective context. J Psychophysiol 1996;33(6):662–70. https://doi.org/10.1111/j.1469-8986.1996.tb02362.x.
[32] Pastor MC, Bradley MM, Löw A, Versace F, Moltó J, Lang PJ. Affective picture perception: Emotion, context, and the late positive potential. Brain Res 2008;1189:145–51. https://doi.org/10.1016/j.brainres.2007.10.072.
[33] Foolen A. The relevance of emotion for language and linguistics. Amsterdam: John Benjamins Publishing Company; 2012.
[34] Russell JA, Fernández-Dols JM, Manstead A. Everyday conceptions of emotion. In: Russell AJ, Fernández-Dols M, Manstead A, Wellenkamp JC, editors. An introduction to the psychology, anthropology and linguistics of emotion. Kluwer Academic; 1995. p. 17–47.
[35] Ackovska N, Kirandziska V. Finding important sound features for emotion evaluation classification. Eurocon 2013. https://doi.org/10.1109/EUROCON.2013.6625196; 2013. Zagreb Croatia: IEEE Publications [accessed July 1-4].

[36] Choi Y, Lee S, Jung S, Choi IM, Park YK, Kim C. Erratum to: Development of an auditory emotion recognition function using psychoacoustic parameters based on the international affective digitized sounds. Behav Res Methods 2016;48 (2):827–927. https://doi.org/10.3758/s13428-015-0596-x.
[37] Grewe O, Nagel F, Kopiez R, Altenmüller E. Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. Emot 2007;7(4):774–88. https://doi.org/10.1037/1528-3542.7.4.774.
[38] Schubert E. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. Aus J Psychol 1999;51(3):154–65. https://doi.org/10.1080/00049539908255353.
[39] 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.
[40] Schubert E. Modeling perceived emotion with continuous musical features. Music Percept 2004;21(4):561–85. https://doi.org/10.1525/mp.2004.21.4.561.
[41] Nagel F, Kopiez R, Grewe O, Altenmüller E. EMuJoy: Software for continuous measurement of perceived emotions in music. Behav Res Methods 2007;39 (2):283–90. https://doi.org/10.3758/bf03193159.
[42] Yang W, Kang J. Soundscape and sound preferences in urban squares: A case study in Sheffield. J Urban Des 2005;10(1):61–80. https://doi.org/10.1080/13574800500062395.
[43] Oberman T, Šćitaroci BBO, Jambrošić K, Kang J, Margaritis E. [Conference proceedings]. Potential of reverberation in squares for soundscape perception, Euronoise. Crete; 2018.
[44] Schubert E. Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music. Psychol Music 2013;41(3):350–71. https://doi.org/10.1177/0305735611430079.
[45] Wang B, Kang J, Zhao W. Noise acceptance of acoustic sequences for indoor soundscape in transport hubs. J Acoust Soc Am 2020;147(1):206. https://doi.org/10.1121/10.0000567.
[46] Coutinho E, Cangelosi A. Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. Emot 2011;11(4):921–37. https://doi.org/10.1037/a0024700.
[47] 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.
[48] Gennarina D, Santorelli E, Rebecca R, Mather MA. Perceptions of emotion and age among younger, midlife, and older adults. Aging Ment Health 2016;22(3). https://doi.org/10.1080/13607863.2016.1268092.
[49] 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(5):477–83. https://doi.org/10.3397/1.3484183.
[50] Van Gerven PW, Vos H, Van Boxtel MP, Janssen SA, Miedema HM. Annoyance from environmental noise across the lifespan. J Acoust Soc Am 2009;126 (1):87–194. https://doi.org/10.1121/1.3147510.