

# A time series analysis model of the relationship between psychoacoustic parameters of urban soundscape spatial sequences and emotional changes

Zhihui Han, Jian Kang and Qi Meng

Citation: [The Journal of the Acoustical Society of America](#) **152**, 2022 (2022); doi: 10.1121/10.0014287

View online: <https://doi.org/10.1121/10.0014287>

View Table of Contents: <https://asa.scitation.org/toc/jas/152/4>

Published by the [Acoustical Society of America](#)

---

## ARTICLES YOU MAY BE INTERESTED IN

[Prediction model of crowd noise in large waiting halls](#)

[The Journal of the Acoustical Society of America](#) **152**, 2001 (2022); <https://doi.org/10.1121/10.0014347>

[REVIEWS OF ACOUSTICAL PATENTS](#)

[The Journal of the Acoustical Society of America](#) **152**, 1995 (2022); <https://doi.org/10.1121/10.0014344>

[Transducer design for low-frequency circular close-packed array and its mutual radiation analysis](#)

[The Journal of the Acoustical Society of America](#) **151**, 2223 (2022); <https://doi.org/10.1121/10.0009579>

[Intelligibility and detectability of speech measured diotically and dichotically in groups of listeners with, at most, "slight" hearing loss](#)

[The Journal of the Acoustical Society of America](#) **152**, 2013 (2022); <https://doi.org/10.1121/10.0014419>

[Rayleigh limit extended: Scattering from a fluid sphere](#)

[The Journal of the Acoustical Society of America](#) **152**, R7 (2022); <https://doi.org/10.1121/10.0014345>

[Acoustic scattering and the exact Green function](#)

[The Journal of the Acoustical Society of America](#) **152**, 2038 (2022); <https://doi.org/10.1121/10.0014346>

---





**Advance your science and career  
as a member of the**

**ACOUSTICAL SOCIETY OF AMERICA**

LEARN MORE



## A time series analysis model of the relationship between psychoacoustic parameters of urban soundscape spatial sequences and emotional changes

Zhihui Han,<sup>1</sup> Jian Kang,<sup>2,a)</sup>  and Qi Meng<sup>1</sup> 

<sup>1</sup>Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, School of Architecture, Harbin Institute of Technology, No. 66 Xi Da Zhi Street, Harbin, China

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

### ABSTRACT:

A listening test was conducted with 32 participants to obtain data on emotional changes in response to three types of urban soundscape spatial sequences. By establishing a time series model, the relationship between psychoacoustic parameters of the sequence and changes in the two dimensions of emotion was determined. Results showed that psychoacoustic parameters can explain 44% and 40%–49% of the changes in the pleasantness and arousal dimensions of emotion, respectively. Roughness and fluctuation have the highest correlation with emotional changes, while loudness and articulation index have the lowest correlation with emotional changes. This research verified the lags between psychoacoustic changes in the soundscape and the associated perceived emotion. First, there was a 3–4 s lag between psychoacoustic parameters and emotional changes. Second, changes in roughness and loudness could cause synchronous changes in emotions, while other parameters could cause delayed changes in emotions. Finally, the lag of emotion had a strong and stable explanatory power for emotional changes. This research proves the effectiveness of the time series analysis technology in establishing the dynamic relationship between the acoustic parameters of soundscape sequences and the second-by-second perceived emotions and provides a new data analysis method for in-depth study of soundscape sequence perception.

© 2022 Acoustical Society of America. <https://doi.org/10.1121/10.0014287>

(Received 20 May 2022; revised 7 September 2022; accepted 8 September 2022; published online 3 October 2022)

[Editor: Sanford Fidell]

Pages: 2022–2037

### I. INTRODUCTION

As the scale of cities expands, soundscapes become increasingly complex. Research into urban soundscapes has not been limited to noise control but extends to soundscape perception.<sup>1–3</sup> The perception of an urban soundscape is multifaceted, and the emotional perception of the soundscape is an important area.<sup>4,5</sup> According to the definition of soundscape in ISO12913-1, an urban soundscapes are related to persons, activities, and places in space and time.<sup>6</sup> Therefore, research on emotional perception of an urban soundscape involves the dynamic changes in a soundscape with time and space and the consequent dynamic changes of emotional perception. Thus, exploring an explainable quantitative relationship between these two changing variables is the purpose of this study.

In the research into the quantitative relationship between the soundscape and its perception, different kinds of sound sources or soundscape segments are usually used as stimuli in the listening tests, and subjects are required to answer a series of subjective questions concerning perception after exposure to each stimulus. Then the acoustic parameters of each stimulus are calculated by acoustic

software, and a quantitative relationship between acoustic parameters and subjective evaluation is established by using multiple regression analysis.<sup>3,7</sup> Hall *et al.*<sup>8</sup> investigated the quantitative relationship between the acoustical parameters of soundscape segments and the perceived affective quality dimensions: pleasantness and vibrancy. Using the semantic differential method, the subjective evaluation of the pleasantness and vitality of the soundscape was obtained. Then four acoustic parameters (roughness, sharpness, loudness, and tonality) of each segment were calculated using the ArtemiS software. Following this, a multiple linear regression mode was established between pleasure/vitality and acoustic parameters to explore the stronger predictors for pleasure/vitality. Similar to previous studies, Aletta *et al.*<sup>9</sup> also studied the relationship between the vibrancy dimension of the perceived affective quality and the soundscapes in a listening test by using a questionnaire. They selected roughness, the presence of people, fluctuation strength, loudness, and the presence of music as predictors and established a linear regression model between these factors and vibrancy. The model's explanatory ability reached 76%. However, for the purpose of the relationship between the dynamic soundscapes and its emotional perception, there are still some shortcomings in the method of previous research.

<sup>a)</sup>Electronic mail: J.kang@ucl.ac.uk

First, the measurement of soundscape perception mostly uses the semantic differential scales, which are suitable for evaluating the general emotional perception caused by a stimulus but cannot track the emotional changes during the process.<sup>10</sup> Second, multiple linear regression is a method suited for cross section data, while the data obtained in this research are time series data (both the data of the acoustic parameters of the soundscape and emotional perception); hence, this method cannot analyze the study data effectively.

An important concern is capturing the emotional changes caused by the soundscape. The continuous emotional measurement method is widely used to study relationships between the acoustical features of music and the emotional changes of the participants. This method usually uses software (such as two-dimensional emotion-space and EMuJoy) to record the emotional changes of the participants, which allows researchers to report these changes by moving the mouse on the computer screen while listening to music. The software collects emotional data at a certain frequency.<sup>11–13</sup> Schubert<sup>14</sup> developed the software of two-dimensional emotion-space to capture the continuous emotional changes of music. The software was developed based on the two-dimensional theory of emotions. The theory uses the two-dimensional emotional space composed of two perpendicularly intersecting dimensions in the plane to measure emotions, and this emotional space is presented to the subjects through the computer screen. The software allows the subjects to continuously evaluate their perceived emotion by adjusting the position of the mouse on the computer screen in the two-dimensional space while listening to music. Subsequently, the software records the changes of emotion in the dimension of time, which enables a second-by-second measurement of emotion. Schubert further verified the validity and reliability of the software through experiments.<sup>14</sup> Nagel *et al.*<sup>15</sup> further developed the open resource software EMuJoy based on previous research, and this software is more convenient for research. The software allows playing not only audio but also video; it also provides ports for connecting physiological measurement instruments. The reliability and validity of the software have been proved.<sup>16</sup> Therefore, continuous emotional measurement was used to capture the emotional changes caused by the urban soundscape.

Another problem pertains to data analysis; the data obtained in this study are time series data, which contain the correlation itself. Spurious regression can result if linear regression is used to establish a model in the time series data. Conversely, the time series analysis is a method that uses the correlation of the data to build a model.<sup>17</sup> Time series techniques include the univariate and multivariate models. The former is a model that uses the correlation of the time series data itself for prediction, which includes the autoregression model (AR) and the autoregressive and moving average model (ARMA). The latter focuses on exploring the long-term relationship between multivariate variables, which include the vector autoregression (VAR)

and the autoregressive distributed lag (ARDL). Time series analysis has been widely used in the study of emotional perception of music. A specific time series model is selected to establish a dynamic relationship between the acoustic characteristics of music and the emotional perception.<sup>14,18,19</sup> For example, when the multivariate time series model was not yet fully developed, Schubert<sup>20</sup> tried to combine the AR model with the linear regression model to study the relationship of the multivariate time series variables in the research of real-time emotional perception of music. He selected six acoustic features to describe the changing characteristics of the music over time, including loudness, rhythm, melody, contour, texture, and spectral centroid. Then, by using the method of continuous emotional measurement, the data of the second-by-second perceived emotional evaluation of the four pieces of music were obtained. Finally, a combination of the AR model and linear regression model was used to establish the relationship between musical features and perceived emotions. The results showed that the model can explain 33%–73% of the perceived emotions and that perceived emotion usually has a delayed response within 1–3 s after the change of musical features. With the development of multivariate time series models, Rogert<sup>21</sup> could directly use the VAR model to establish the relationship between the perceived emotion and acoustic features (such as intensity and spectral flatness). Although the results of this research partly verified the results of the previous studies, the application of the VAR model greatly improved the efficiency and convenience of the study. Therefore, compared with multiple linear regression, the time series analysis technology not only can establish the dynamic relationship of multiple time series variables, but also can be used to explore the synchronization/lag relationship between the variables, allowing for in-depth research. Thus, the study of the dynamic relationship between the acoustic parameters of the urban soundscape sequence and the perceived emotion can also be completed by selecting an appropriate time series model.

Therefore, this study uses the method of time series analysis to study the quantitative relationship between the changes in the acoustic parameters of the urban soundscape sequence and the changes in self-reported perceived emotions. This study further hopes to draw conclusions on the following questions. First, is it possible to establish the relationship between the changes in the acoustic parameters of urban soundscape sequence and emotional changes using time series analysis, and how does the model reflect this dynamic relationship? Second, is there a synchronous or lagged relationship between the changes in acoustic parameters and the emotional changes that follow, and can it be easily reflected in the model? Finally, how is the explanatory power of the model established by time series technology in the study of emotional perception of soundscape, and is there any possibility of improving the model? Ultimately, it is expected that time series analysis techniques can provide an effective method for researching emotional perception of dynamic soundscapes.

## II. METHODOLOGY

### A. Production of urban soundscape spatial sequences

Studies have demonstrated that the impact of soundscapes on perception differs in different urban spaces.<sup>22</sup> Building on previous research on responses to soundscapes, this research selected three typical urban spaces as soundscape recording sites—urban parks, a commercial pedestrian mall, and roads with their surrounding areas. To restore the spatial aspects of the urban soundscape, audio recordings developed on-site were used in the experiment. Before recording, several representative sound scenes were selected for each urban space, as shown in Fig. 1. For example, a selection of five typical sound scenes in the park was

considered for the urban park sequence. The first one was the amusement park area. Here, the soundscape was dominated by the mechanical sound of amusement park equipment and human voices. The second category was the dating corner area, and its soundscape was dominated by the sound of bird calls and human voices; it was the most crowded of all the areas. The third was the lotus pond area, a gathering place for music lovers. Here, the soundscape was dominated by human voices and musical instruments. The fourth was the Xishan Waterfall area, where the soundscape was dominated by the sound of the breeze and bird calls; it was the quietest of all the areas. Fifth was the Taohuayuan area, the junction area between the amusement park and Xishan Waterfall. Here, the soundscape was more complex, consisting of human voices, bird calls, mechanical sounds, and the sounds of the breeze. The selection process for the other two sequences—including the pedestrian mall and the road—involved nine and eight typical sound scenes, respectively, as shown in Fig. 1. This was followed by the audio recording of each sound scene. To ensure a sufficient density of people, a particular recording time was selected: 9–11 a.m. on Saturday and Sunday. According to the ISO-2 standard,<sup>23</sup> a 5-min recording time can fully reflect the soundscape characteristics of the area, so a continuous recording of 5 min was made for each sound scene. The recordings were stored in HDF format, with 44 100–97 Hz sampling, 16-bit quantization. For the editing of the sound sequences, first, a 1-min soundscape clip from the 5-min recording was isolated. Then the 1-min clips were connected according to the order shown in Fig. 1 to form a sound sequence. The purpose was to control the time of the whole experiment to avoid fatigue of the subjects. The 1-min soundscape clip should be able to reflect the characteristics of the sound scenes and also avoid the unusual sound source that does not belong to the sound scene. Finally, three soundscape spatial sequences, namely the park, the pedestrian mall, and the road, were obtained by using the software Cooledit. Their durations were 5, 8, and 9 min, respectively.

There are a few points to explain. First, the selection of the number of sound scenes in each sequence is based on the soundscape characteristics of different urban spaces. A more real restoration of the soundscape in the laboratory can trigger a more real emotional experience. Although the number of sound scenes of the park sequence is lower than the number for the other two sequences, this does not affect the results of the experiment. As the focus of this experiment is the variation of acoustic parameters, there is no direct relationship between the number of sound scenes and the variation of acoustic parameters. Second, although the pedestrian mall and the road sequences are sometimes similar in appearance, the content of these urban soundscapes differs; hence, a learning curve does not occur. Finally, the order of the sound scenes in each sequence is preset. If this were not the case, there would be many combinations, which would not be conducive for the experimental conditions. Therefore, the results of this experiment only allow for the preset order, and deviation is expected.



FIG. 1. (Color online) The selection of sound scenes in different urban spaces. N/A, not applicable.

## B. Selection and analysis of psychoacoustic parameters

Psychoacoustic parameters are objective physical quantities that describe people's subjective feelings of sound, which play an important role in auditory sensations. Loudness, sharpness, fluctuation strength, roughness, tonality, and articulation index are the commonly used psychoacoustic parameters.<sup>24</sup> Loudness is the most important psychoacoustic quantity, describing the intensity of volume.<sup>25</sup> Sharpness is a parameter describing the proportion of high-frequency components in the sound spectrum and reflects the harshness of the sound. Fluctuation reflects the slower changes in sound. Roughness reflects the perceptual effect of fast amplitude modulation of a sound.<sup>24</sup> Tonality is the sensation of timbre, which indicates whether sound consists mainly of tonal components or broadband noise.<sup>26</sup> Articulation index reflects the transmission efficiency of speech information in background noise.<sup>27</sup> Therefore, psychoacoustic parameters are functions of the time structure and spectral distribution and can qualitatively describe the dynamic changes of soundscape sequences from different angles.

Six psychoacoustic parameters are selected in this research, namely, loudness, sharpness, fluctuation strength, roughness, tonality, and articulation index. The reason is that, on the one hand, the greater the number of parameters, the better the multi-dimensional description that can be made of the changes of the soundscape. On the other hand, the selection of the number of parameters, as the independent variable in the model, is restricted by the data amount of the dependent variable, which is emotional changes. The data amount of the emotional changes in the park sequence is the least and is 300 (the duration of the sequence is 300 s). According to previous research, the least data amount can meet the requirement of six independent variables simultaneously.<sup>20</sup>

The parameters were calculated using ArtemiS Suite 11 (Advanced Research Technology for Measurement and Investigation of Sound and Vibration). The method embedded in ArtemiS calculates the roughness, sharpness, tonality, and articulation index.<sup>28</sup> The calculation of loudness was based on the calculation method proposed by Zwicker in ISO 532-1.<sup>29</sup> The sharpness was calculated based on DIN 45692.<sup>30</sup> The time interval of calculation for each parameter needs to be set in the software, but it needs to consider the requirements of the time series analysis. In the multivariate time series analysis, the independent variables and dependent variables should have the same time interval, which is to say, the time interval of the data of the acoustic parameters needs to be the same as that of the self-reported emotional changes (1 s; see Sec. III B). Therefore, the time interval of the calculation of the acoustic parameters is also set to 1 s.

Figure 2 shows variation of acoustic parameters with time. The  $x$  axis is time with every 60 s representing a sound scene. The  $y$  axis represents the acoustic parameter and its unit, including roughness and its unit asper, fluctuation and its unit vacil, sharpness and its unit acum, loudness and its unit sone, articulation and its unit %, and tonality and its unit Hearing Model by Sottek (HMS). In general, there is a

similar variation range of the same parameter in different sequences, but the variation characteristics of the curve, such as the position of the peak, number of peaks, shape of the variation, and frequency of the variation, differ. For roughness, the position of the peaks in the three sequences is significantly different. For fluctuation strength, the frequency of variation is obviously different, and it is the highest in the road sequence. For sharpness and loudness, the shape of the variation is different in different sequences, and it shows an obvious characteristic of double-peak in the pedestrian mall sequence compared with the other two. For tonality, the value of the peak and the frequency of the variation are significantly higher than those of the other two sequences. Therefore, it can be seen that there is a significant difference in the variation characteristics of the parameters in these three sound sequences, which can provide more diversified data for the model.

## C. Measurement of emotional changes

The EMuJoy software was selected to record the emotion in the experiment, based on the two-dimensional emotional theory. Unlike the three-dimensional theory,<sup>31</sup> the two-dimensional theory describes the emotion through the dimensions of pleasantness and arousal.<sup>32</sup> As shown in Fig. 3, the software interface shown to the participants is composed of two axes (representing two dimensions of emotion) intersecting at right angles. The  $x$  axis represents the pleasantness dimension; the left and the right sides of the  $x$  axis represent the negative emotion ("displeasure") and the positive emotion ("pleasure"), respectively, with the value changing from  $-1$  to  $1$ . The  $y$  axis represents the arousal dimension, the lower and upper sides of the  $y$  axis represent calming ("low arousal") and arousing ("high arousal"), respectively, with the value changing from  $-1$  to  $1$ . When the sound stimulus plays, the participant can report their emotional changes by clicking the mouse on the screen to select an emotional point in the two-dimensional emotional space at any time during the process. The number of possible clicking times is unlimited. The software records the data every 50 ms<sup>15</sup> and obtains the variation in pleasantness and arousal over time. Previous studies have confirmed the validity of the software.<sup>12,13,15</sup>

First, the entire procedure was explained to the participants, and their informed consent was obtained. Then the concept of the emotional dimensions and the use of the software were explained to the subjects. According to previous studies,<sup>33</sup> to familiarize the subjects with the use of the software, five pictures from the International Affective Picture System manual<sup>34</sup> were selected and displayed to the subjects randomly. Then the participants were asked to listen carefully to three audio clips played through the headphones, assess whether their emotions were affected by the audio, and use the mouse to mark the corresponding emotional point once identified. They could report their emotions at any time without any limit on the number of times. If there was no emotional change during the whole process, they

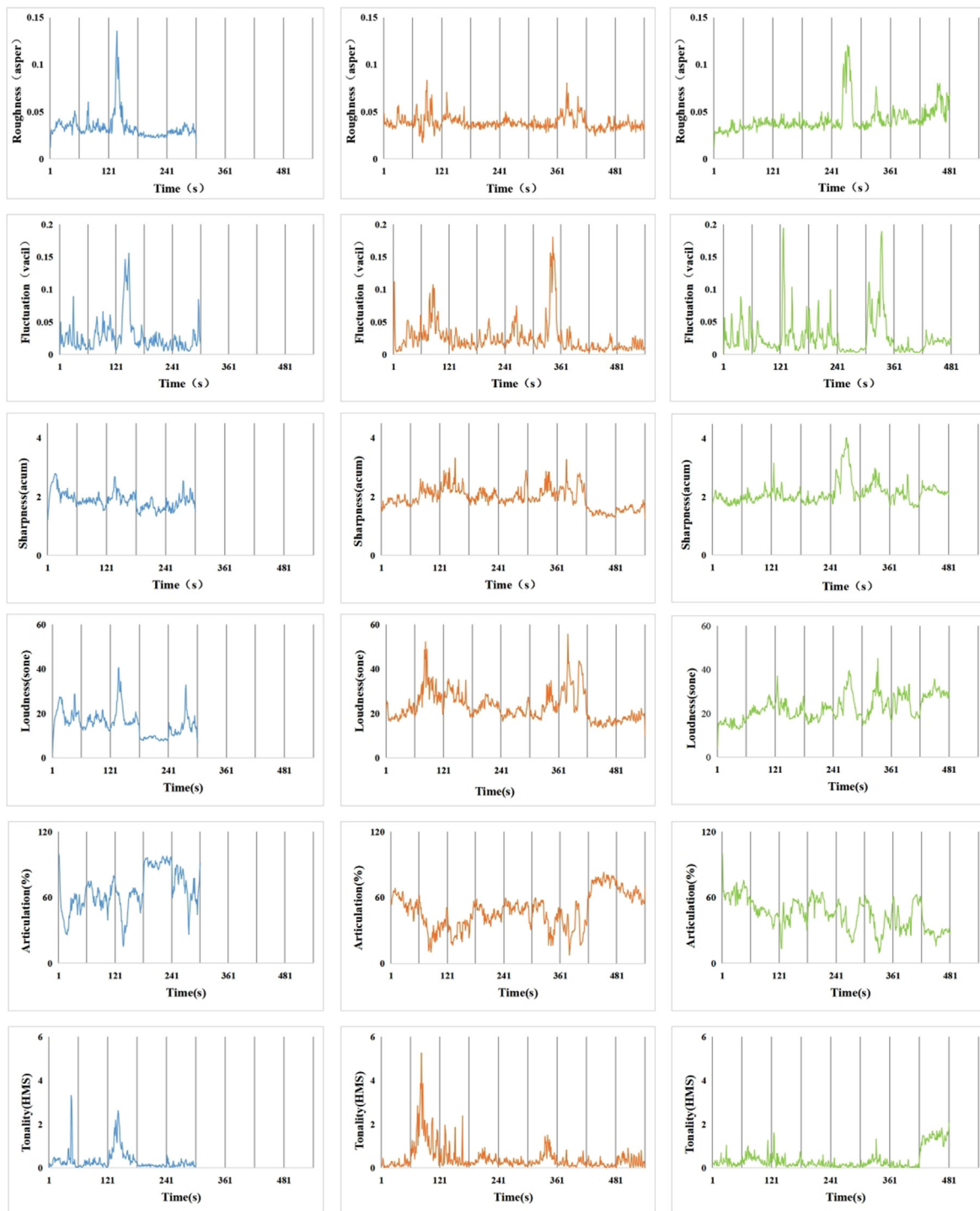


FIG. 2. (Color online) Variation of acoustic parameters with time in different sound sequences.

could elect not to click the mouse. The experiment was conducted in a listening room with a background noise of 25 dBA at a frequency of 500–1000 Hz during daytime. During the experiment, the subjects were in the listening room by themselves without any disturbance, and the audio playback was controlled by the computer in the control room. There are three sound sequences in the experiment, and each audio is played to the subjects only once randomly to eliminate the influence of the factor of the order between the

sequences on the results. The park, pedestrian mall, and road sequences lasted for 5, 9, and 8 min, respectively, and there was also a blank time of 30 s between the sequences to eliminate the potential influence in emotion of the previous audio on the next.<sup>35</sup> Therefore, the whole experiment took about 25 min. The audio was played through headphones, and the volume of the headphones was calibrated by connecting the dummy heads [Head Acoustics (Herzogenrath, Germany) HMS III] before the experiment.

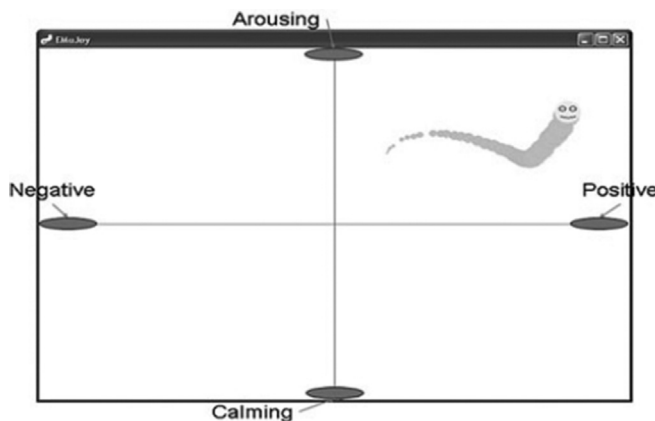


FIG. 3. The subject interface of EMuJoy software (Ref. 15). Reproduced from F. Nagel, R. Kopiez, O. Grewe, and E. Altenmüller, *Behav. Res. Methods* 39(2), 283–290 (2007). Copyright 2007 Author(s), licensed under a Creative Commons Attribution 4.0 License (Ref. 15).

#### D. Participants

The previous research on soundscapes indicates that a sample of 30 subjects can meet the needs of the listening test.<sup>24,36</sup> The participants totalled 32 (15 male and 17 female), with a normal hearing level of 20–30 dB at a frequency of 500–1000 Hz.<sup>23</sup> Since the focus of this study is not on the difference of ages, only those aged 20–30 years (mean age: 27 years, standard deviation: 4.1) were selected for convenience. The participants’ occupations included students, office workers, and freelancers. All participants were volunteers; hence, they were not compensated for participation.

#### E. Data analysis

Multiple regression analysis is usually employed to study the quantitative relationship between multiple factors in soundscape research,<sup>37,38</sup> which applies to the analysis of cross-sectional data. The data obtained in this experiment—unlike in previous studies—were time series data, which are continuous. Spurious regression can occur if the multiple regression method is used on time series data. Therefore, the time series analysis method was used to build the regression model using the correlation of the data itself. This method is widely used in emotional perception research.<sup>39</sup> The autoregressive distributed lag (ARDL) model was selected among many multivariate time series models. Compared with other models, it has the following advantages. First, it is a multivariate linear model with pre-determined causality compared with other structural models, which makes it easier to explain the practical relationship between the independent variables and dependent variables. Second, the structure includes the lag of both the independent and dependent variables, which can help quantify how the lag effect of emotion is reflected in different psychoacoustic parameters.<sup>40,41</sup>

The EViews 10.0 software was used to establish the ARDL model. The model was established in four steps—data verification, model setting, model verification, and model application—as shown in Fig. 4.

According to the requirements of the ARDL model, the time series data must be stable.<sup>40</sup> Therefore, the ADF unit root test method was used to verify the stationarity of the time series data of the psychoacoustic parameters and emotional changes. The results showed that, compared to other parameters, only the data of tonality in the road sequence were not stable in the data test of the psychoacoustic parameters. Additionally, the data for the two dimensions of emotion were not stable. The data for the non-stationary time series must be stabilized before it is entered into the model, and the difference method was used to achieve this. The first-order difference operation was used to subtract the previous value from the next value in the time series data, and  $D(X)$  was the definition assigned to the new sequence obtained, where  $X$  represented the original series. Then the first-order difference operation was performed on all non-stationary time series data, and the stability of the new series was tested again. The results showed that the new series was stable.

Another test for the data is the co-integration test, which is mainly used if there are two or more non-stationary series in one model.<sup>42</sup> Only when non-stationary variables have a co-integration relationship can they be included in the same regression model; otherwise, spurious regression will occur. Therefore, the Engle–Granger (EG) co-integration analysis was conducted on the two groups of variables in the road sequence model—pleasantness and tonality, and arousal and tonality—to test whether there was a co-integration relationship between them. The results showed no co-integration relationship between tonality and pleasantness/arousal in the road sequence. Therefore, the variable of tonality was not included in the regression of the road sequence model.

The “adjusted  $R$ -squared” was used as the selection criterion for the lag order of the variable, and the “Newey–West” estimation method was selected to avoid the autocorrelation of the model.

Finally, the Breusch–Godfrey (BG) serial correlation Lagrange multiplier (LM) test was used to assess the autocorrelation of the ARDL model. Results showed that there was no autocorrelation in the other models, except for the pleasantness model of the park sequence and the road sequence. It can be used to analyse the relationship between variables or predict data for models without autocorrelation. It can also be used for analysis—but not prediction—for models with autocorrelation, due to the “Newey–West” setting.

### III. RESULTS

#### A. The difference in correlations between different psychoacoustic parameters and emotional changes

Since the correlation in the model is complex, the absolute value of the correlation coefficient and the positive/negative of the value are discussed separately. This section is about the former, and the latter will be discussed in Sec. III B. Figure 5 illustrates the differences in correlations between different psychoacoustic parameters and emotional dimensions. Table I presents further detailed data. The results show that the correlation is the highest between

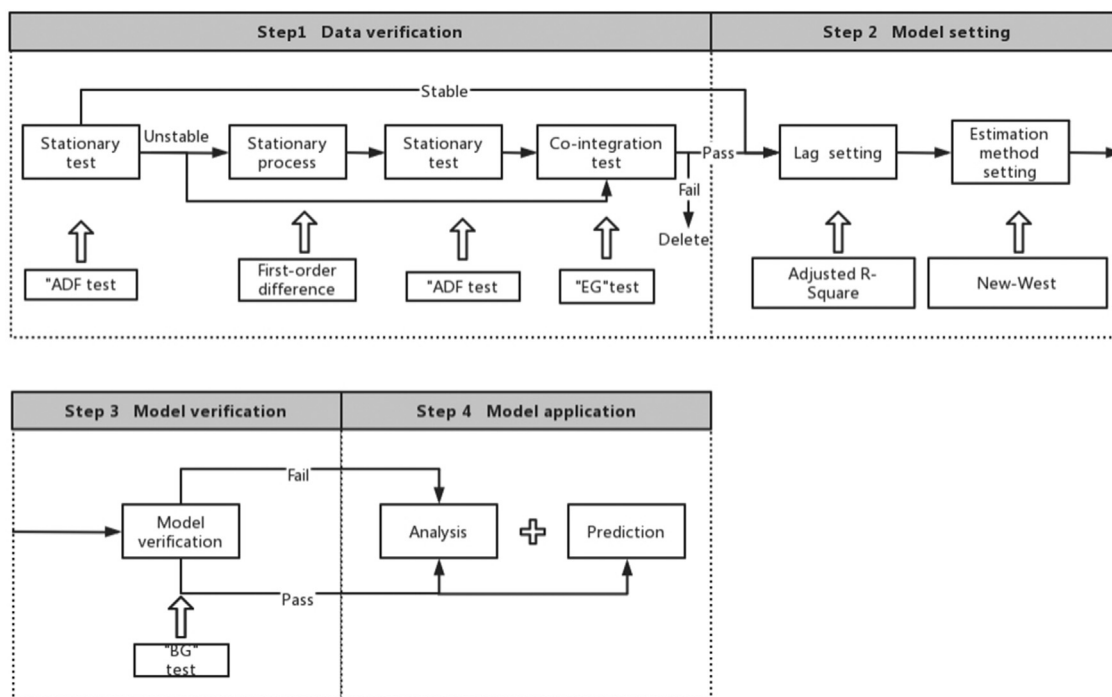


FIG. 4. The steps of the ARDL model.

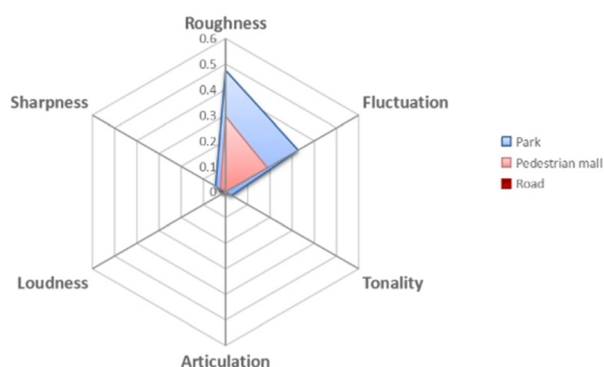
roughness/fluctuation strength/sharpness and emotional changes, while it is low between loudness/articulation index and emotional changes. There is a complicated relationship between tonality and emotional changes. The value of the coefficients between all parameters and emotion is significantly larger in the park sequence than in the road sequence.

Roughness has the highest correlation with emotional changes among all parameters and is the strongest predictor of emotional changes. However, its correlation is not significant in the traffic sequence. Its correlation coefficient can reach 0.30–0.48 in the pleasantness dimension and 0.25–0.58 in the arousal dimension. Fluctuation strength is also a parameter with a high correlation coefficient (0.19–0.59), but its correlation is also not significant in the traffic sequence. Compared with the former two parameters, the correlation coefficient of sharpness is reduced and can

increase to 0.04 in the pleasantness dimension and 0.03 in the arousal dimension. This correlation is significant in all the sound sequences, which means that sharpness is not affected by different soundscape sequences and is the most stable parameter.

Although loudness/articulation index has a weak correlation with emotional changes, they are still noteworthy. The strongest predictor parameters (such as roughness) failed to predict emotional changes in the road sequence. However, this is not the case with loudness/articulation index. Their correlation is significant in different sequences—especially the road sequence—which means that there is a weak and stable correlation between loudness/articulation index and emotional changes. Therefore, the parameters of different prediction factors can play a complementary role in predicting the emotional perception of soundscape sequences. It is

Pleasantness dimension and psychoacoustic parameters



Arousal dimension and psychoacoustic parameters

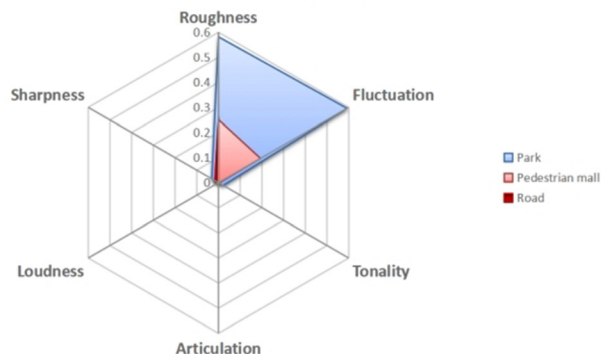


FIG. 5. (Color online) Comparison of correlation values between psychoacoustic parameters.



TABLE I. Coefficient between psychoacoustic parameters and emotional dimensions in different sequences. The correlation coefficient in the table is the largest absolute value with the highest significance, and the coefficients with significance less than 10% are blank.

Emotional dimensions in different sequences	Psychoacoustic parameters					
	Roughness	Fluctuation	Sharpness	Loudness	Articulation	Tonality
Pleasantness dimension mode						
Park	0.475 212	0.326 983	0.044 563	0.008 538		0.027 439
Pedestrian mall	0.295 424	0.188 618	0.022 613	0.002 028	0.000 827	
Road			0.020 810	0.001 792	0.000 707	
Arousal dimension mode						
Park	0.581 602	0.592 25	0.031 008		0.001 908 4	0.021 705
Pedestrian mall	0.251 091	0.193 786	0.015 25	0.001 546	0.000 324	0.002 902
Road	0.260 236		0.015 756	0.000 939	0.000 71	

important to use various parameters to predict emotion, not just the parameters with high correlation coefficients. The articulation index is a parameter to measure the transmission efficiency of speech information in background noise. Research on the emotional perception of urban soundscapes has shown that the arousal dimension is related to the informativeness of the sound.<sup>5,8</sup> Thus, the articulation index can be assumed as a parameter for describing the informativeness of the sound; however, the low correlation coefficient also demonstrates its weak ability to do this. Therefore, modifications in future studies may be required to improve its ability to predict emotional perception from the perspective of describing more sound information.

Significantly, the correlation coefficient of tonality is 0.02 in the park sequence, but it is not significant in the pedestrian mall sequence. This indicates that tonality has a complicated relationship with emotional changes and cannot be used to predict them.

**B. The lag in emotional changes in different psychoacoustic parameters**

Two parts are discussed in this section, as shown in Table II. The first part is regarding the lag in emotional changes in different psychoacoustic parameters. This means that the emotion is affected within a period after the change of parameters. The second part explores the lags in emotional changes—at particular moments—and how they are affected by previous emotional changes. The lag includes the time of the lag and the correlation (positive and negative) at different lags.

The lag time of different psychoacoustic parameters is stable at about 3–4 s. The lag time of emotion for roughness or fluctuation strength is 3 s and reaches 4 s for sharpness, loudness, articulation index, and tonality. However, the significant lag time (the lag time corresponding to the significant lags) differs in different soundscape sequences for one parameter. For example, the significant lag time for roughness and fluctuation strength in the park sequence is 3 s, which is longer than it is in the pedestrian mall sequence (1 s) or the road sequence (0 s). Similarly, the significant lag time of sharpness or loudness to emotion is longer in the pedestrian mall and road sequences (3–4 s) than it is in the

park sequence (1–2 s). The perception of the lag time varies in different soundscape sequences.

The correlation between one parameter and emotional changes is dynamic (positive or negative) at different lags. For example, the correlation between the roughness and pleasantness dimensions is first positive at lag 0, becomes negative at lag 1, and reverts to positive at lag 2/3 in the pleasantness model of the park sequence. The same situation occurs in other psychoacoustic parameters. The findings of the positive or negative correlation between a certain parameter and emotional changes are inconclusive, which is different from the results of previous studies, but this dynamic correlation may provide a new angle for future research.

In addition, emotional changes are synchronized with the change of some parameters but delayed for the change of others. Emotional changes are synchronized with the change of roughness or loudness and delayed for the changes in fluctuation strength, sharpness, and articulation index. For example, the significant lag of roughness first appears at lag 0, which shows that emotional changes are very sensitive to changes in roughness and are synchronized. Lag 3 of roughness is also significant, which also means that the impact of roughness on emotions is continuous. Therefore, the changes in roughness are synchronized with the changes in emotion, and the continuous effect of roughness can reach 3 s, which mirrors loudness. However, for sharpness and fluctuation strength, significant lags appear at the lag 2/3 or the lag 1/3, respectively, indicating that there is a delayed correlation between these parameters and emotion.

Finally, considering the emotional lag in their two dimensions, the lag time of the pleasantness dimension is 4 s, and the arousal dimension is 3–4 s. That is to say that the emotional changes at a given moment are affected by the emotional changes in the past 3–4 s. The correlation of emotional changes is significantly positive at lag 1 and then demonstrates dynamic changes at other lags.

**C. The interpretive ability of the model for emotional changes**

As shown in Fig. 6, the model explains 42%–43% of the variation in pleasantness, and 40%–49% of the variation

TABLE II. Correlation between emotional dimensions and acoustic parameters and its lags. 0, 1, 2, 3, and 4 represent the lags of one acoustic parameter. For example, “1” represents “Lag(1),” which is obtained by delaying the parameter by 1 s. “+”/“-” represent the positive/negative relationship between the acoustic parameters and emotional dimensions, and asterisks represent the significance of the correlation. \*, \*\*, and \*\*\* represent *p*-value levels of 10%, 5%, and 1%, respectively.

Lag variable	Pleasantness dimension					Arousal dimension				
	0	1	2	3	4	0	1	2	3	4
<b>Roughness</b>										
Park sequence	+**	—	+	+***		—**	+*	—	+***	
Pedestrian mall sequence	+	—**	—	+		+	—*	+	+	
Road sequence	+					+**				
<b>Fluctuation</b>										
Park sequence	+	—	—	+**		+	+	—**	+**	
Pedestrian mall sequence	+	—*	+	—		+	—***			
Road sequence	+					+	—	—	+	
<b>Sharpness</b>										
Park sequence	—	+*	—	+	—	+	+	—**		
Pedestrian mall sequence	—	—	—*	+***		+	—*	—***	+***	
Road sequence	+	+	—***	+***		—	—	—**	+***	
<b>Loudness</b>										
Park sequence	+	—***	+			+				
Pedestrian mall sequence	—*	—	—	+	+	—***	—	—*	+	+***
Road sequence	—***	—***	+***	+***		—	—	+		
<b>Articulation</b>										
Park sequence	—	—	+	—	—	+	+*	+	—*	
Pedestrian mall sequence	—	—	—	+*		—*				
Road sequence	—***	—	+***			—	+	+**	—***	
<b>Tonality</b>										
Park sequence	—**	+***	—***	+**	—*	—	+	—***	+***	
Pedestrian mall sequence	+	+	+	—		+*				
Road sequence										
<b>D(pleasantness/arousal)</b>										
Park sequence		+***	—	+	—**		+***	+	+*	
Pedestrian mall sequence		+***	—	—	+***		+***	+	+	—*
Road sequence		+***	—	+	+**		+***	—***	+	

in arousal, in terms of variables of psychoacoustic parameters (roughness, fluctuation strength, sharpness, loudness, tonality, and articulation index), as well as its lags and emotion lags. It seems that the  $R^2$  of the pleasantness dimension is more stable than that of the arousal dimension across different soundscape sequences. However, the  $R^2$  of the model is related to the interpretive ability of both the psychoacoustic parameters and the emotion. By comparing the absolute value of the correlation coefficient of the independent variables, it can be observed that the correlation between the emotional changes and the psychoacoustic parameters declines, and the correlation between the emotional changes and the lags is raised when the sequence is changed from park to pedestrian mall to road. This phenomenon exists in both the pleasantness and arousal models. However, there is a decrease in the interpretive ability of the psychoacoustic parameter and an increase in the interpretive ability of emotion itself to reach a dynamic balance in different pleasantness, but not in different arousal dimension models. Therefore, the  $R^2$  of the pleasantness dimension model is stable, while the  $R^2$  of the arousal dimension model is variable. The reason for this phenomenon may relate to the composition of the soundscape sequences. Further analysis

revealed that the composition of the park sequence was relatively simple compared with the road sequence—the emotional change relies more on itself than on acoustic features in the more complex soundscape sequences.

This section further compares the difference between the actual and predicted values of the emotional changes, by selecting the model without autocorrelation, as shown in Fig. 7. It is evident that the model has a better fit with smooth emotional changes but a poor fit for the sharp ones. As shown in Fig. 7, two sharp emotional changes appear at the beginning of minutes 2 and 7 in the pleasantness model of the pedestrian mall sequence, corresponding to the sound of musical instruments and traffic noise, respectively. A sharp emotional change also appears at the beginning of minute 5 in the arousal model of the park sequence, corresponding to the sound of mechanical noise. In the arousal model of the traffic sequence, the sharp decrease in emotion at minute 5 corresponds to the sound of construction noise. It seems that the point of poor fit in the model is related to the type of sound source. However, in the arousal model of the road sequence, the sharp emotional changes corresponding to the traffic noise demonstrate a better model fit. Further analysis shows that the soundscape of the park,

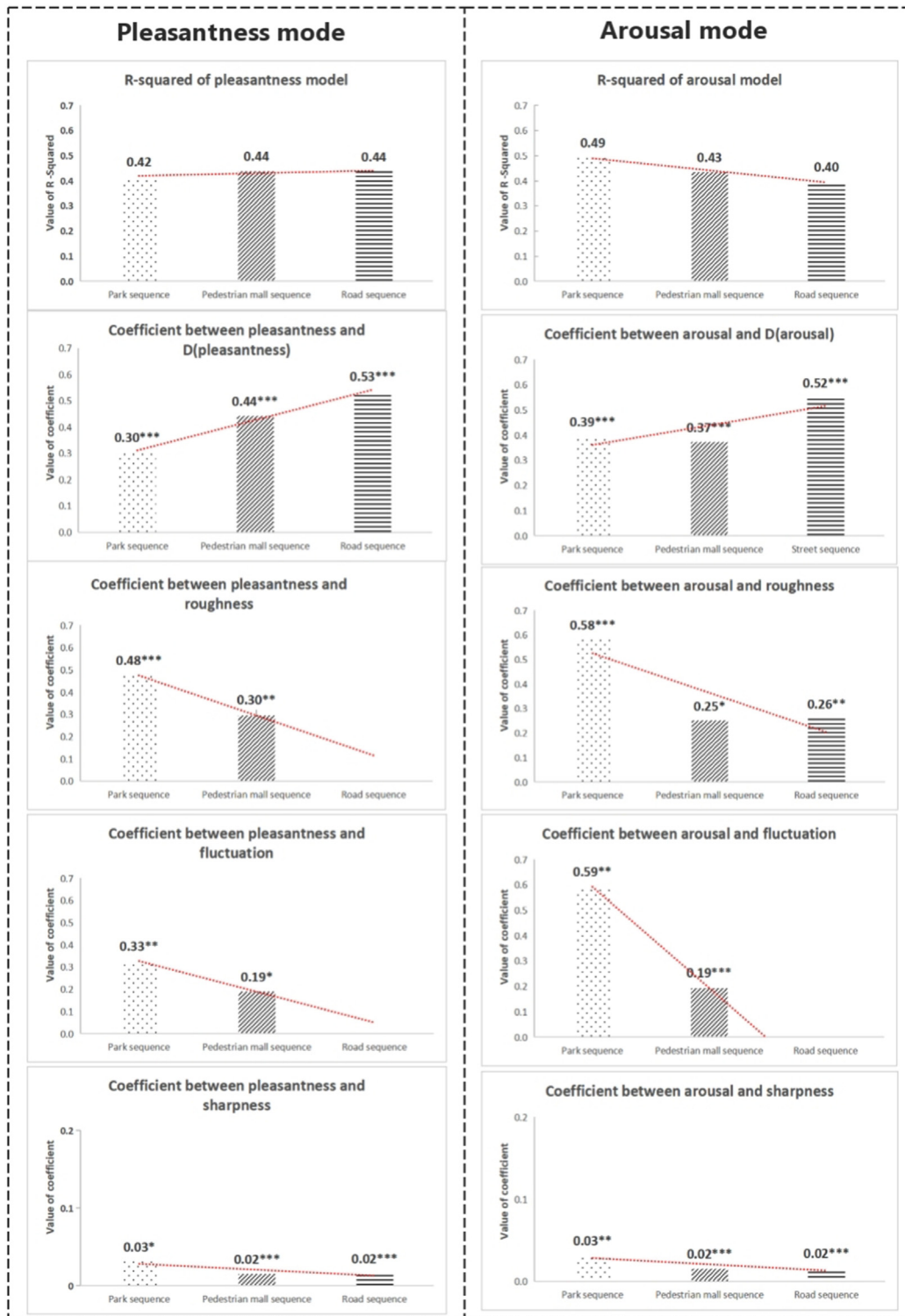


FIG. 6. (Color online) The trend of the coefficients between variables.

characterised by natural sounds, is quieter, while the soundscape of the pedestrian mall, dominated by artificial sounds, is noisy. Traffic noise—an artificial sound—is significantly different from the soundscape of the park compared to the pedestrian mall. Hence, the differences between the sound sources in the soundscapes are another important factor. Therefore, the type of sound source must be considered to

achieve a better model fit, and the contrast between the sound source and the soundscape is an important factor.

#### IV. DISCUSSION

The research shows that the time series technique provides an effective method for the study of the perception of

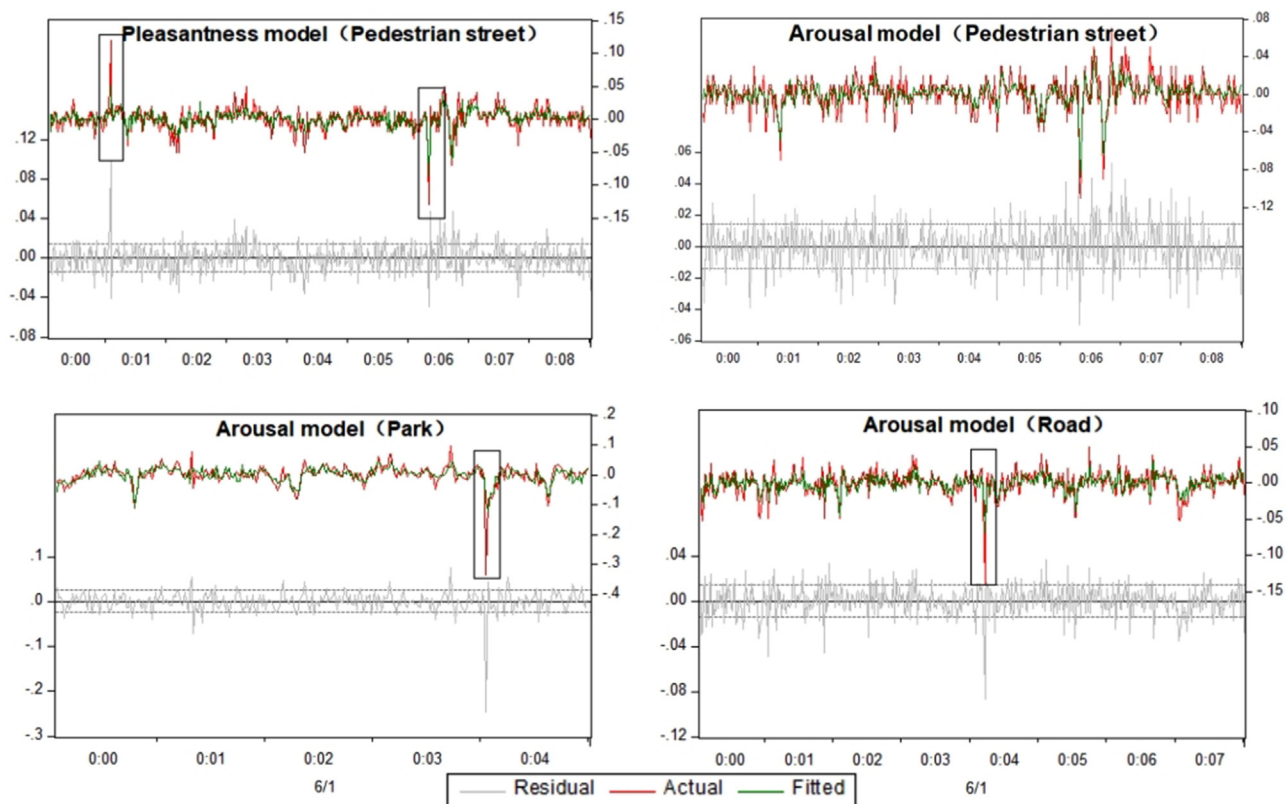


FIG. 7. (Color online) Comparison of actual value and predicted value of emotional changes.

soundscape sequences. If the research of soundscape perception focuses on soundscape sequence, rather than on the sound sources or the soundscape fragments, then analysis of the time series data becomes the primary problem. Although this study has proved that the method of time series analysis is effective for the above problems, this method is convenient only for numerical variables and ineffective for categorical variables. Therefore, the acoustic parameters, which are the numerical variables in this study, were selected as the independent variables of the model in this study. The results also showed that only using the acoustic parameters to predict emotional changes is not enough, and more potential variables may be related to the type of sound source. Previous studies have also shown that the information carried by the sound source also largely determines people's perception of the soundscape.<sup>4</sup> Therefore, the type of sound source, as a categorical variable, could be decoded in several dimensions, and its dimensions could be encoded to convert the categorical variable into a numerical variable. How to disassemble to preserve the characteristics of the sound source to the greatest extent is worth studying, as such an idea will greatly improve the convenience of using the time series model in soundscape research. However, the question of the dimensions in which the sound source could be decoded to preserve the information carried by it deserves more study. This will also improve the convenience of using the time series technique in soundscape research.

In addition, the application of time series technology can allow in-depth research on soundscape perception.

The previous research on soundscape perception is static. For example, the conclusion is usually a certain correlation coefficient between the perception and the acoustic parameter, which ignores the change of the perception in the process. However, this study shows that there is a dynamic relationship between the acoustic features and emotional perception and that emotional perception will reach a steady state within 1–4s after the change in soundscape features. However, perception is always adjusting as the soundscape is always changing. Therefore, both the magnitude and positive or negative relationship of the correlation vary at different lag times. The time series analysis technology makes it possible to investigate people's attention to the cognitive process of soundscape. However, this dynamic relationship also brings some challenges for further analysis. For example, there is no regularity in the positive/negative correlation at different lag times, leading to an inconclusive result. This is because the emotional perception of the real soundscape sequence contains complex factors. For example, the information carried by the sound resource is diverse, and it also has different meanings for different people. Simultaneously, people's previous emotional cognition of different types of soundscape also plays an important role in the process of emotional experience. Therefore, a stable conclusion for a certain parameter can be obtained only by excluding the influence of the non-acoustic factors as much as possible and designing the control experiment for a certain parameter. The significance of this study is exploratory rather than conclusive, and it hopes to provide more perspectives for further research.

**V. CONCLUSIONS**

A time series model was established in this study to explore the relationship between the changes in psychoacoustic parameters and emotional changes in three types of urban soundscape spatial sequences. The main findings are as follows.

- (1) The difference in correlation between psychoacoustic parameters and emotional changes: Roughness and fluctuation strength are strong predictors of emotional changes, and sharpness is a stable predictor across different sequences. Loudness and articulation index are weak but stable predictors of emotional changes. Tonality has a complex relationship with emotion. Finally, different parameters can play a complementary role in the prediction of emotional changes in different soundscape sequences.
- (2) The lag of emotional changes to psychoacoustic parameters: First, the lag time of emotional changes to all parameters is stable at 3–4 s, but the significant lag time differs between different soundscape sequences for the same parameters. Second, the correlation between one psychoacoustic parameter and emotional changes is dynamic (positive or negative) at different lags. Roughness and loudness have both synchronous and delayed correlations with emotional changes, while fluctuation strength, sharpness, and articulation index only have delayed correlations with emotional changes. Finally, the lag time of emotion in the two dimensions is 3–4 s, and there is a significant positive correlation with emotional changes at lag 1.
- (3) The interpretive ability of the model for emotional changes: The  $R^2$  of the pleasantness model is stable at about 43%, and the  $R^2$  of the arousal model is between 40% and 49% in a different model. There is a better model fit for smooth emotional changes and a poor model fit for sharp emotional changes.

This study verifies the role of the psychoacoustic parameters of urban spatial soundscapes in predicting the emotional changes induced by them. However, it also indicates the limitations of the predictive power of these parameters. In general, psychoacoustic parameters have a stronger ability to predict stable emotional changes but a weaker ability to predict sharp emotional changes, which may be related to the non-acoustic factors. The three indicators of roughness, fluctuation strength, and sharpness do play an important role in predicting emotional changes, but indicators with weak

predictive power, such as the articulation index, are still worthy of further research. In addition, there is a lag phenomenon between psychoacoustic parameters and emotional changes, but it does not exist in all indicators, which shows that there is a difference in emotional perception for different acoustic parameters, and the reasons for this deserve further research. Although this research provides the length of the lag of emotional changes for parameters, its composition is complicated and relates to cognition. A strong and stable “emotional inertia” is evident in the emotional changes induced by soundscapes, which cannot be ignored in future research in the study of the emotional perception of soundscapes. Finally, the perceived emotions of different age groups may be different, so this study is only valid for young people, and further research is needed for other age groups.

This study validates the usefulness of time series analysis techniques in studying emotional perception of urban soundscapes. On the one hand, this method can effectively establish the relationship in several dynamically changing factors, like the changing soundscape and its changing perception. At the same time, this method allows the lag relationship between these factors to be revealed, which provides a new angle for further exploration of cognitive-related research on soundscapes. On the other hand, how to convert categorical variables into numerical ones to incorporate them into the time series model to improve predictive ability is also a new challenge for future research. Finally, there are many options for multivariate time series models, and only the model of ARDL has been investigated in this research. Other multivariate models, such as the VAR model, which focuses on dynamic matrix system between variables, also provide great value for the study of soundscapes.

**ACKNOWLEDGMENTS**

This work is supported by National Natural Science Foundation of China (NSFC) Grant Nos. 51878210 and 51778169, Natural Science Foundation of Heilongjiang Province Grant No. YQ2019E022, and European Research Council (ERC) Advanced Grant No. 740696 on “Soundscape Indices” (SSID).

**APPENDIX**

See Tables III and IV for results of the ADF unit root test and the co-integration test and Tables V, VI, and VII for the ARDL models of the park and emotional dimensions, the pedestrian mall sequence and emotional dimensions, and the road sequence and emotional dimensions, respectively.

TABLE III. ADF unit root test. (i) The  $c$ ,  $t$ , and  $p$  in the equation form  $(c, t, p)$  represent constant, trend, and lag, respectively; (ii) the ADF unit root test uses the MacKinnon one-sided  $p$ -values; (iii)  $D(X)$  represents the first-order difference of  $X$  sequence.

Variable	Equation form $(c, t, p)$	5% level	$t$ -statistic	Probability	Result
Park sequence					
Psychoacoustic parameters					
Loudness	$(c, t, 0)$	-3.424 875	-4.234 415	0.0045	Stable
Roughness	$(c, 0, 0)$	-2.870 964	-4.263 138	0.0006	Stable

TABLE III. (Continued.)

Variable	Equation form (c, t, p)	5% level	t-statistic	Probability	Result
Sharpness	(c, t, 1)	-3.424 926	-4.656 336	0.0010	Stable
Fluctuation	(c, 0, 2)	-2.871 029	-3.590 615	0.0065	Stable
Tonality	(c, 0, 3)	-2.871 061	-3.722 487	0.0042	Stable
Articulation	(c, t, 0)	-3.424 875	-3.780 015	0.0189	Stable
Emotional dimension					
Pleasantness	(0, 0, 1)	-1.941 888	-1.696 410	0.0850	Unstable
D(pleasantness)	(0, 0, 0)	-1.941 888	-11.670 80	0.0000	Stable
Arousal	(0, 0, 1)	-1.941 888	-1.871 035	0.0586	Unstable
D(arousal)	(0, 0, 0)	-1.941 888	-9.603 602	0.0000	Stable
Pedestrian mall sequence					
Psychoacoustic parameters					
Loudness	(c, 0, 1)	-2.866 683	-3.997 975	0.0015	Stable
Roughness	(c, 0, 1)	-2.866 683	-8.189 147	0.0000	Stable
Sharpness	(c, t, 1)	-3.418 179	-4.916 618	0.0003	Stable
Fluctuation	(c, 0, 4)	-2.866 713	-4.228 677	0.0006	Stable
Tonality	(c, 0, 2)	-2.866 693	-4.886 576	0.0000	Stable
Articulation	(c, 0, 0)	-2.866 673	-3.450 376	0.0098	Stable
Emotional dimensions					
Pleasantness	(0, 0, 1)	-1.941 412	-1.726 126	0.0800	Unstable
D(pleasantness)	(0, 0, 0)	-1.941 412	-12.681 87	0.0000	Stable
Arousal	(0, 0, 1)	-1.941 412	-1.472 299	0.1318	Unstable
D(arousal)	(0, 0, 0)	-1.941 412	-12.987 53	0.0000	Stable
Road sequence					
Psychoacoustic parameters					
Loudness	(c, t, 0)	-3.419 211	-5.895 974	0.0000	Stable
Roughness	(c, t, 1)	-3.419 231	-4.512 340	0.0016	Stable
Sharpness	(c, 0, 0)	-2.867 342	-4.559 803	0.0002	Stable
Fluctuation	(c, 0, 4)	-2.867 392	-5.133 191	0.0000	Stable
Totality	(0, 0, 3)	-1.941 489	-0.399 151	0.5398	Unstable
D(totality)	(0, 0, 2)	-1.941 489	-18.774 49	0.0000	Stable
Articulation	(c, 0, 0)	-2.867 342	-4.714 486	0.0001	Stable
Emotional dimensions					
Pleasantness	(0, 0, 1)	-1.941 489	-1.434 337	0.1413	Unstable
D(pleasantness)	(0, 0, 0)	-1.941 486	-10.822 44	0.0000	Stable
Arousal	(0, 0, 2)	-1.941 487	-0.700 347	0.4132	Unstable
D(arousal)	(0, 0, 1)	-1.941 487	-12.631 53	0.0000	Stable

TABLE IV. Co-integration test—Engle–Granger.

Series	Null hypothesis	Engle–Granger tau-statistic	Probability	Co-integration
Pleasantness tonality	Series are not cointegrated	-1.857 096	0.6021	No
Arousal tonality	Series are not cointegrated	-2.557 888	0.2563	No

TABLE V. ARDL model of the park and emotional dimensions. \*, \*\*, \*\*\* represent the p-values level of 10%, 5%, and 1%, respectively.

Model 1 (pleasantness)		Model 2 (arousal)	
Variable	Coefficient	Variable	Coefficient
D[pleasantness(-1)]	0.304271***	D[arousal(-1)]	0.394204***
D[pleasantness(-2)]	-0.033520	D[arousal(-2)]	0.053083
D[pleasantness(-3)]	0.034538	D[arousal(-3)]	0.107032*
D[pleasantness(-4)]	-0.148179**	Articulation	$5.13 \times 10^{-6}$
Articulation	-0.000275	Articulation(-1)	0.000815*
Articulation(-1)	-0.000878	Articulation(-2)	0.001438
Articulation(-2)	0.002116	Articulation(-3)	-0.001908*
Articulation(-3)	-0.001531	Fluctuation	0.080814
Articulation(-4)	-0.000422	Fluctuation(-1)	0.153302

TABLE V. (Continued.)

Model 1 (pleasantness)		Model 2 (arousal)	
Variable	Coefficient	Variable	Coefficient
Fluctuation	0.235498	Fluctuation(-2)	-0.592250**
Fluctuation(-1)	-0.06634	Fluctuation(-3)	0.383462**
Fluctuation(-2)	-0.467881	Loudness	0.000110
Fluctuation(-3)	0.326983**	Roughness	-0.398345**
Loudness	0.001517	Roughness(-1)	0.657013*
Loudness(-1)	-0.008538***	Roughness(-2)	-0.416766
Loudness(-2)	0.001802	Roughness(-3)	0.581602***
Roughness	0.384042**	Sharpness	0.007781
Roughness(-1)	-0.109165	Sharpness(-1)	0.027538
Roughness(-2)	0.279385	Sharpness(-2)	-0.031008**
Roughness(-3)	0.475212***	Tonality	-0.009306
Sharpness	-0.009725	Tonality(-1)	0.005858
Sharpness(-1)	0.044563*	Tonality(-2)	-0.021705***
Sharpness(-2)	-0.000979	Tonality(-3)	0.013531***
Sharpness(-3)	0.007515		
Sharpness(-4)	-0.036818	C	-0.038111
Tonality	-0.014828**	@Trend	-3.70 × 10 <sup>-5</sup> *
Tonality(-1)	0.027439***	R-squared	0.492717
Tonality(-2)	-0.024490***	N	300
Tonality(-3)	0.014588**		
Tonality(-4)	-0.011708*		
C	0.109664		
@Trend	-2.55 × 10 <sup>-5</sup>		
R-squared	0.420824		
N	300		

TABLE VI. ARDL model of the Pedestrian mall sequence and emotional dimensions. \*, \*\*, \*\*\* represent the p-values level of 10%, 5%, and 1%, respectively.

Model 3 (Pleasantness)		Model 4 (Arousal)	
Variable	Coefficient	Variable	Coefficient
D[pleasantness(-1)]	0.440863***	D[arousal(-1)]	0.372648***
D[pleasantness(-2)]	-0.000265	D[arousal(-2)]	0.031341
D[pleasantness(-3)]	-0.006071	D[arousal(-3)]	0.030335
D[pleasantness(-4)]	0.108843***	D[arousal(-4)]	-0.070157*
Articulation	-0.000399	Articulation	-0.000324*
Articulation(-1)	-0.000411	Fluctuation	0.077681
Articulation(-2)	-2.81 × 10 <sup>-5</sup>	Fluctuation(-1)	-0.193786***
Articulation(-3)	0.000827*	Loudness	-0.001546***
Fluctuation	0.082736	Loudness(-1)	-0.000189
Fluctuation(-1)	-0.188618*	Loudness(-2)	-0.000539*
Fluctuation(-2)	0.168672	Loudness(-3)	0.000489
Fluctuation(-3)	-0.093781	Loudness(-4)	0.000794***
Loudness	-0.000923*	Roughness	0.030338
Loudness(-1)	-0.001227	Roughness(-1)	-0.251091*
Loudness(-2)	-0.000765	Roughness(-2)	0.096435
Loudness(-3)	0.002028**	Roughness(-3)	0.164536
Loudness(-4)	0.000755**	Sharpness	0.003759
Roughness	0.015578	Sharpness(-1)	-0.009497*
Roughness(-1)	-0.295424**	Sharpness(-2)	-0.012708***
Roughness(-2)	-0.035827	Sharpness(-3)	0.015257***
Roughness(-3)	0.210273	Tonality	0.002902*
Sharpness	-0.001120		
Sharpness(-1)	-0.009319		
Sharpness(-2)	-0.014883*		
Sharpness(-3)	0.022613***		
Tonality	0.001706		
Tonality(-1)	0.004329		

TABLE VI. (Continued.)

Model 3 (Pleasantness)		Model 4 (Arousal)	
Variable	Coefficient	Variable	Coefficient
Tonality(-2)	0.000710		
Tonality(-3)	-0.004719		
C	0.013560	C	0.044190*
@Trend	$-1.86 \times 10^{-6}$	@Trend	$3.57 \times 10^{-6}$
R-squared	0.439574	R-squared	0.434795

TABLE VII. ARDL model of the road sequence and emotional dimensions. \*, \*\*, \*\*\* represent the p-values level of 10%, 5%, and 1%, respectively.

Model 5 (Pleasantness)		Model 6 (Arousal)	
Variable	Coefficient	Variable	Coefficient
D[pleasantness(-1)]	0.534870***	D[arousal(-1)]	0.549060***
D[pleasantness(-2)]	-0.009834	D[arousal(-2)]	-0.187060***
D[pleasantness(-3)]	0.005411	D[arousal(-3)]	0.059509
D[pleasantness(-4)]	0.092108**	Articulation	-0.000538
Articulation	-0.000614***	Articulation(-1)	0.000105
Articulation(-1)	-0.000219	Articulation(-2)	0.000710**
Articulation(-2)	0.000707***	Articulation(-3)	-0.000547***
Fluctuation	0.034732	Fluctuation	0.012044
Loudness	-0.001792***	Fluctuation(-1)	-0.053960
Loudness(-1)	-0.001455***	Fluctuation(-2)	-0.065500
Loudness(-2)	0.001318***	Fluctuation(-3)	0.087884
Loudness(-3)	0.001035***	Loudness	-0.00135
Roughness	0.138955	Loudness(-1)	-0.000668
Sharpness	0.004766	Loudness(-2)	0.000939*
Sharpness(-1)	0.002603	Roughness	0.260236**
Sharpness(-2)	-0.020810***	Sharpness	-0.000413
Sharpness(-3)	0.014557***	Sharpness(-1)	-0.006597
		Sharpness(-2)	-0.015678**
		Sharpness(-3)	0.015756***
C	0.014579	C	0.039168
@Trend	$9.78 \times 10^{-6}$	@Trend	$7.14 \times 10^{-6}$
R-squared	0.436087	R-squared	0.396836

<sup>1</sup>F. Aletta, J. Kang, and Ö. Axelsson, "Soundscape descriptors and a conceptual framework for developing predictive soundscape models," *Landsc. Urban Plan.* **149**, 65–74 (2016).

<sup>2</sup>D. Botteldooren, M. Boes, D. Oldoni, and B. D. De Coensel, "The role of paying attention to sounds in soundscape perception," *J. Acoust. Soc. Am.* **131**(4), 3382 (2012).

<sup>3</sup>M. S. Engel, A. Fiebig, C. Pfaffenbach, and J. Fels, "A review of the use of psychoacoustic indicators on soundscape studies," *Curr. Pollut. Rep.* **7**(3), 359–378 (2021).

<sup>4</sup>A. Fiebig, P. Jordan, and C. C. Moshona, "Emotions—The Application of emotion theory in soundscape," *Front. Psychol.* **11**, 573041 (2020).

<sup>5</sup>R. Cain, P. Jennings, and J. Poxon, "The development and application of the emotional dimensions of a soundscape," *Appl. Acoust.* **74**(2), 232–239 (2013).

<sup>6</sup>ISO 12913-1:2014. "Acoustics-soundscape-part 1: Definition and conceptual framework" (International Organization for Standardization, Geneva, Switzerland, 2014).

<sup>7</sup>D. Västfjäll, "Emotional reactions to sounds without meaning," *Psychology* **3**(8), 606–609 (2012).

<sup>8</sup>D. A. Hall, A. Irwin, M. Edmondson-Jones, S. Phillip, and J. E. W. Poxon, "An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes," *Appl. Acoust.* **74**(2), 248–254 (2013).

<sup>9</sup>F. Aletta and J. Kang, "Towards an urban vibrancy model: A soundscape approach," *Int. J. Environ. Res. Public Health* **15**(8), 1712 (2018).

<sup>10</sup>M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Ther. Exp. Psychiatry* **25**(1), 49–59 (1994).

<sup>11</sup>O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller, "Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music," *Emotion* **7**(4), 774–788 (2007).

<sup>12</sup>E. Coutinho and A. Cangelosi, "Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements," *Emotion* **11**(4), 921–937 (2011).

<sup>13</sup>F. Nagel, O. Grewe, R. Kopiez, and E. Altenmüller, "The relationship of psychophysiological responses and self-reported emotions while listening to music," in *Proceedings of the Göttingen NWG Conference 2005*, Göttingen, Germany (February 17–20, 2005).

<sup>14</sup>E. Schubert, "Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space," *Aust. J. Psychol.* **51**(3), 154–165 (1999).

<sup>15</sup>F. Nagel, R. Kopiez, O. Grewe, and E. Altenmüller, "EMuJoy: Software for continuous measurement of perceived emotions in music," *Behav. Res. Methods* **39**(2), 283–290 (2007).

<sup>16</sup>K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Sci. Data* **6**, 196 (2019).



- <sup>17</sup>H. Lütkepohl, *New Introduction to Multiple Time Series Analysis* (Springer, New York, 2005).
- <sup>18</sup>K. N. Olsen, R. T. Dean, C. J. Stevens, and F. Bailes, “Both acoustic intensity and loudness contribute to time-series models of perceived affect in response to music,” *Psychomusicology* **25**(2), 124–137 (2015).
- <sup>19</sup>E. Schubert, “Correlation analysis of continuous emotional response to music: Correcting for the effects of serial correlation,” *Music. Sci.* **5**(1 Suppl.), 213–236 (2002).
- <sup>20</sup>E. Schubert, “Modeling perceived emotion with continuous musical features,” *Music Percept.* **21**(4), 561–585 (2004).
- <sup>21</sup>R. T. Dean and F. Bailes, “Time series analysis as a method to examine acoustical influences on real-time perception of music,” *Empir. Musicol. Rev.* **5**(4), 152–175 (2010).
- <sup>22</sup>D. Daniel, A. Pedrero, M. A. Navacerrada, and C. Díaz, “Relationship between the geometric profile of the city and the subjective perception of urban soundscapes,” *Appl. Acoust.* **149**, 74–84 (2019).
- <sup>23</sup>ISO/TS 12913-2:2018. “Acoustics—Soundscape—Part 2: Data collection and reporting requirements” (International Organization for Standardization, Geneva, Switzerland, 2018).
- <sup>24</sup>E. Zwicker and H. Fastl, *Psychoacoustics—Facts and Models* (Springer, Berlin), pp. 203–264 (1999).
- <sup>25</sup>H. Fletcher and W. A. Munson, “Loudness, its definition, measurement, and calculation,” *J. Acoust. Soc. Am.* **5**, 82–108 (1933).
- <sup>26</sup>W. Aures, “A model for calculating the sensory euphony of various sounds,” *Acustica* **59**, 130–141 (1985).
- <sup>27</sup>K. D. Kryter, “Methods for the calculation and use of the articulation index,” *J. Acoust. Soc. Am.* **34**(11), 1689–1697 (1962).
- <sup>28</sup>HEAD Acoustics GmbH, “Psychoacoustic analyses in ArtemiS II” (HEAD Acoustics GmbH, Herzogenrath, Germany, 2011).
- <sup>29</sup>ISO 532-1. “Acoustics—Method for calculating loudness level” (International Organization for Standardization, Geneva, Switzerland, 2017).
- <sup>30</sup>DIN 45692. “Measurement technique for the simulation of the auditory sensation of sharpness” (Deutsches Institut für Normung E.V., Berlin, 2009).
- <sup>31</sup>J. A. Russell and A. Mehrabian, “Evidence for a three-factor theory of emotions,” *J. Res. Pers.* **11**(3), 273–294 (1977).
- <sup>32</sup>J. A. Russell, “Core affect and the psychological construction of emotion,” *Psychol. Rev.* **110**(1), 145–172 (2003).
- <sup>33</sup>E. Coutinho and N. Dibben, “Emotions perceived in music and speech: Relationships between psychoacoustic features, second-by-second subjective feelings of emotion and physiological responses,” in *Proceedings of the 3rd International Conference on Music and Emotion*, Jyväskylä, Finland (June 11–15, 2013).
- <sup>34</sup>P. J. Lang, M. M. Bradley, and B. N. Cuthbert, *International Affective Picture System (IAPS): Technical Manual and Affective Ratings* (University of Florida, Gainesville, FL, 2005).
- <sup>35</sup>B. Wang, J. Kang, and W. Zhao, “Noise acceptance of acoustic sequences for indoor soundscape in transport hubs,” *J. Acoust. Soc. Am.* **147**(1), 206–217 (2020).
- <sup>36</sup>M. Rådsten-Ekman, Ö. Axelsson, and M. E. Nilsson, “Effects of sounds from water on perception of acoustic environments dominated by road-traffic noise,” *Acta Acust. united Acust.* **99**(2), 218–255 (2013).
- <sup>37</sup>J. Y. Hong and J. Y. Jeon, “Exploring spatial relationships among soundscape variables in urban areas: A spatial statistical modelling approach,” *Landsc. Urban Plan.* **157**, 352–364 (2017).
- <sup>38</sup>V. P. Romero, L. Maffei, G. Brambilla, and G. Ciaburro, “Modelling the soundscape quality of urban waterfronts by artificial neural networks,” *Appl. Acoust.* **111**, 121–128 (2016).
- <sup>39</sup>V. Kurbalija, M. Ivanović, M. Radovanović, Z. Geler, W. Dai, and W. Zhao, “Emotion perception and recognition: An exploration of cultural differences and similarities,” *Cogn. Syst. Res.* **52**, 103–116 (2018).
- <sup>40</sup>S. Johansen and K. Juselius, “Maximum likelihood estimation and inference on cointegration—With applications to the demand for money,” *Oxf. Bull. Econ. Stat.* **52**(2), 169–210 (1990).
- <sup>41</sup>M. H. Pesaran and B. Pesaran, *Working with Microfit 4.0: Interactive Econometric Analysis* (Oxford University, New York, 2001).
- <sup>42</sup>M. H. Pesaran, Y. Shin, and R. J. Smith, “Bounds testing approaches to the analysis of level relationships,” *J. Appl. Econ.* **16**(3), 289–326 (2001).