MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Psychological assessment of an urban soundscape using facial expression analysis

Hu, Xuejun[1]
School of Architecture, Harbin Institute of Technology
Harbin, 150001, China
Qi Meng[2]
School of Architecture, Harbin Institute of Technology; Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology
Harbin, 150001, China
Jian Kang[3]
UCL Institute for Environmental Design and Engineering, the Bartlett, University College London, WC1H 0NN, London, UK
Yan-jun Han[4]
School of Architecture, Harbin Institute of Technology; Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology
Harbin, 150001, China

ABSTRACT

In recent years, questionnaire investigations have typically been used in soundscape studies. However, the subjective results of this method are subject to many weaknesses. Therefore, it is necessary to discover an objective method for evaluating urban soundscapes. This study explored a new method for psychological assessment based on systems of facial expression analysis. The present study used FaceReader 7.1 to measure the facial expressions of 20 participants in a laboratory setting. The recordings and pictures used were taken from typical public urban environments in Harbin, China. The results showed that this objective method performed well in the evaluation of sound sources. Based on this method, some expression details, such as the change from happiness to sadness over time, can be assessed more accurately than with questionnaires. Furthermore, these results demonstrated that facial expression analysis can be used to assess soundscapes, and that urban planners should introduce habitats for birds and areas where there is water to improve comfort in public spaces.

Keywords: urban soundscape, facial expression analysis, sound source, valence
I-INCE Classification of Subject Number: 63

[1] 523762533@qq.com
[2] mengq@hit.edu.cn (corresponding author)
[3] j.kang@ucl.ac.uk
[4] hanyj@hit.edu.cn

1. INTRODUCTION

In recent years, studies of acoustic landscapes have usually assessed people's subjective feelings about soundscapes through questionnaire surveys. Brown et al. explored the definition of 'soundscape' and emphasized the importance of subjective evaluations for researching acoustic landscapes; questionnaires are one such means of evaluation [1]. The questionnaire survey method for soundscape classification entails enquiring about people's subjective feelings after they listen to recordings then classifying the recorded clips into categories specified by the subjects themselves [2]. Studies by Cain et al. and Aletta et al. used questionnaires which asked participants to select the appropriate words for describing their feelings to find the dimension describing the subjective feelings of sound [3, 4]. Liu et al. combined field research with ratings to assign the influence of some physical and landscape indicators to the soundscape on participants' feelings [5]. Questionnaires are also sometimes used to explore the influence of a particular sound (e.g., bird song, traffic noise, water sounds, etc.) on human perception [6, 7, 8]. In addition, method of instructing participants to walk around a soundscape and then complete a questionnaire is also widely used in the study of soundscapes [9]. To reach conclusions with greater degrees of accuracy, however, a large number of participants are usually required. Furthermore, the survey method has additional limitations beyond that of sample size.

The present study applied a deep learning approach to facial expression analysis, which has been demonstrated to accurately detect occurrences of facial action units (AUs) as well as estimate their intensity [10]. While this system has been applied to a variety of research fields, including psychology [11], tourism [12], and the food industry [13], it is not used frequently in the study of soundscapes. Accordingly, the main aims of this study were as follows: (1) Compare the results of the facial expression analysis system with questionnaire results to explore whether the facial expression analysis system can be applied to soundscape research; and (2) explore the effects of different sounds on facial expressions.

2. INTRODUCTION

2.1. Survey sites

The Shengwang binaural recording system was used to document sounds for this study, and the recordings were made along the banks of the Majiagou River in Harbin (Figure 1). This waterfront is representative of the city's public spaces because it is not only a location for different kinds of human activities, but the elements of its landscape are also diverse. The sound compositions recorded varied depending on the environment of the precise recording locations. For example, the sound composition of the waterfront's 'activity area' consisted mainly of water sounds, bird sounds, human voices, and other behavioural sounds, while the sound composition of the area near the main traffic artery consisted mainly of traffic noise. A total of 8 recordings were made, each 40 seconds long. A panoramic photo of the area was taken while recording, and the sound level data was recorded with a BSWA 801 sound level meter (Figure 2).

2.2. Laboratory analysis

Twenty participants were recruited for participation in the study, 12 women and 8 men, all of whom were students at the Harbin Institute of Technology and aged 19 to 24 (mean = 21.75, variance = 2.197). The analysis was conducted in a laboratory with a panoramic screen, and participants were informed about the precautions and scheduled a time to enter the laboratory the day before the analysis. On the day of the analysis, after a subject entered the laboratory, he or she was given a minute or so to adapt to the laboratory environment.

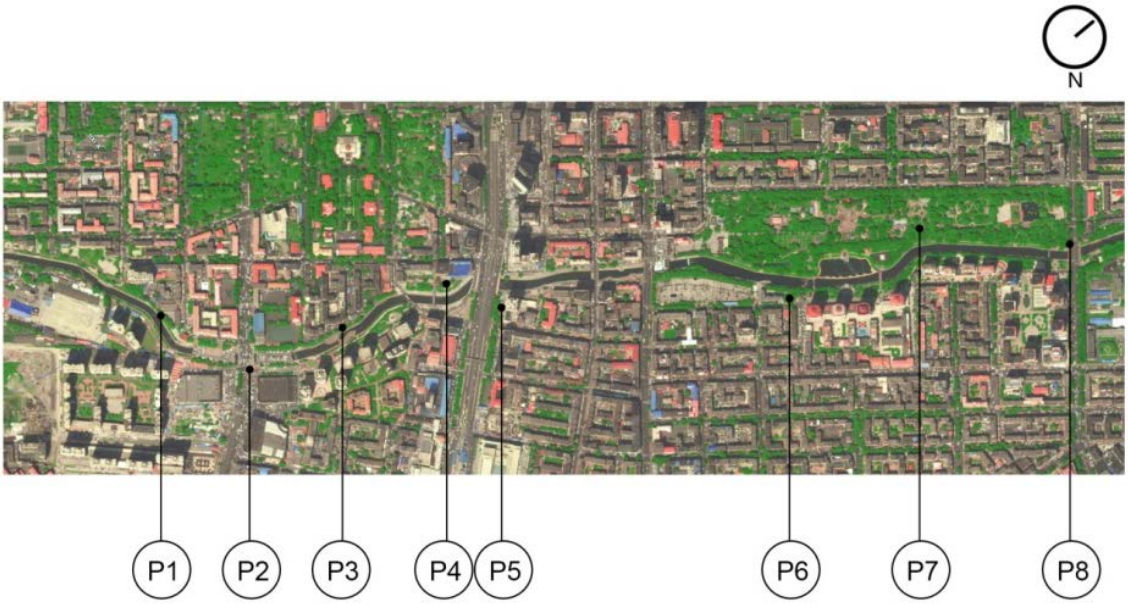


Figure 1 Recording locations

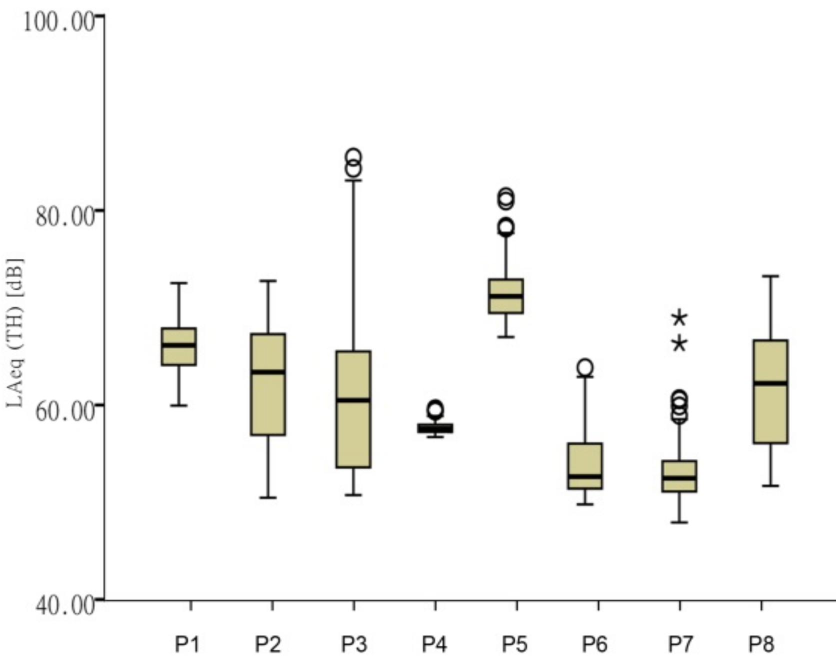


Figure 2 Sound levels of the 8 sites

A panoramic screen was placed in the laboratory, on which the panoramic photo taken of the recording area was displayed. The participant sat in a fixed seat in front of the screen, and a camera that recorded facial expressions was placed in front of the participant (Figure 3).The experimenter started playing the recordings 15 seconds after the facial analysis system started recording the expressions. Each participant listened to 5 recordings and was presented with a souvenir at the end of the session.

Table1 Locations and main function of the survey sites

	latitude	longitude	Site Function
P1	45° 44′ 47.88″	126° 38′ 27.05″	Waterfront public event space
P2	45° 44′ 53.07″	126° 38′ 30.06″	Parking lot near the main road
P3	45° 45′ 02.83″	126° 38′ 34.29″	Public activity space under the bridge
P4	45° 45′ 05.19″	126° 38′ 35.38″	Historic building area near the main road
P5	45° 45′ 07.51″	126° 38′ 40.52″	Urban main road
P6	45° 45′ 24.08″	126° 39′ 01.44″	Waterfront Plaza
P7	45° 45′ 35.28″	126° 39′ 05.17″	Acrivity areas in the park
P8	45° 45′ 41.05″	126° 39′ 13.36″	Near the bus stop

Table2 Questions in the questionnaire survey

Parts	Aims	Questions
1	Basic information	Name, gender, age
2	Acoustic comfort	Evaluate the acoustic comfort of the environment (0-10 scale)
3	Significance of sounds	Rate the significance of the sounds on a scale of 0 to 10 from ‘none’ to ‘significant’ (including vehicle, machinery, birdsong, water, music, wind, human activity, and talking)

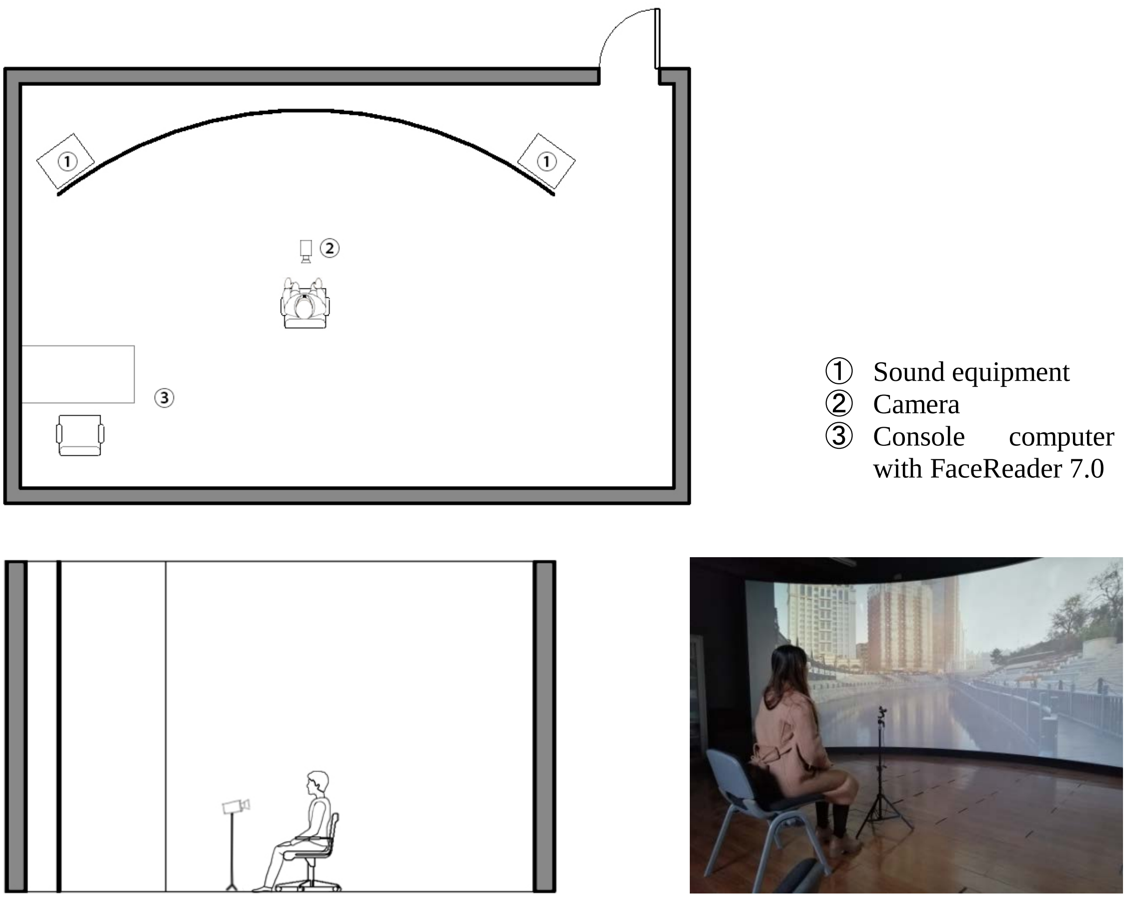


Figure.3 The laboratory setting

2.3. Questionnaire survey

To compare the results of the traditional questionnaire survey and facial expression analysis system, after listening to each recording, participants also completed a questionnaire about it. The questionnaire consisted of three parts (Table 2).

3. RESULTS

3.1. Comparison

Figure 5 shows the evaluation of participants' levels of acoustic comfort, and the results of the saliency evaluation of the different sounds are shown in Table 3. The facial expression analysis system relays a person's facial expression as a percentage of six emotions: nature, happiness, sadness, surprise, fear, and disgust. It generates 15 data per second and calculates corresponding valence and arousal. When discussing emotions in psychology, 'valence' refers to the intrinsic attractiveness/'goodness' (positive valence) or averseness/'badness' (negative valence) of an event, object, or situation [14]. Among the six emotional elements listed above, changes in happiness and sadness are more clearly represented in human facial expressions, so this study focused on the proportion of the changes in the participants' facial expressions for these two emotions. The data of every five seconds of each recording is averaged, and the trend of the subject changing with time is shown in Table 6.

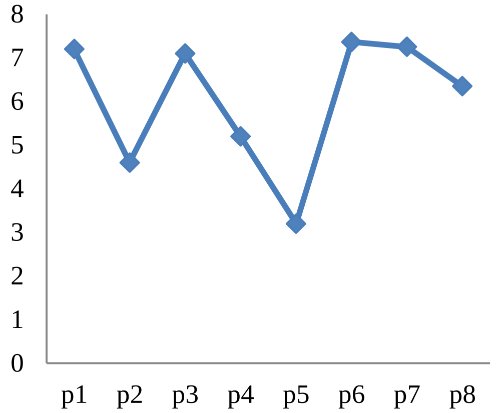
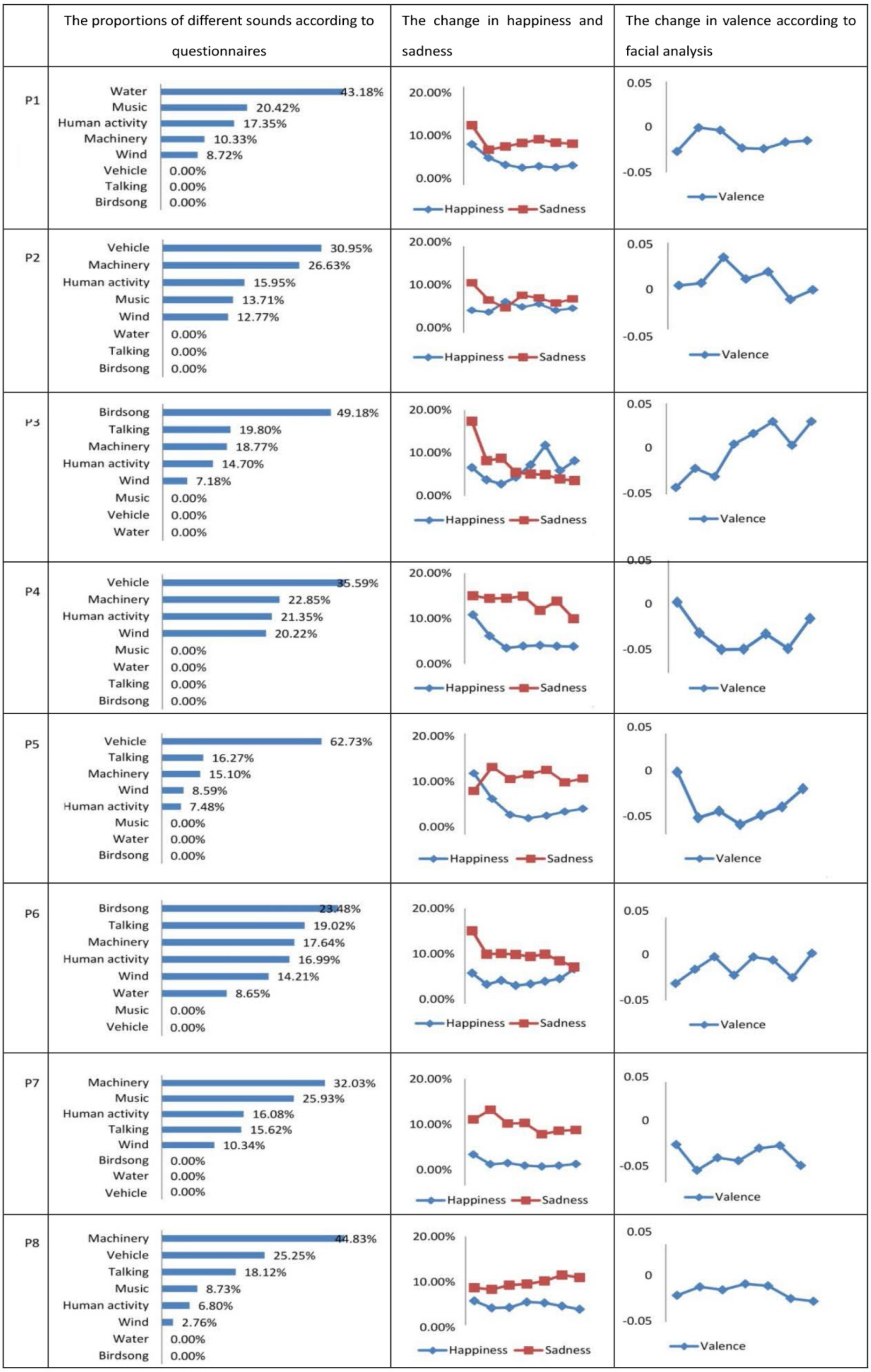


Figure. 5 the average evaluation of the recordings

Among the eight recordings, P2, P4, and P5 mainly consisted of traffic noise; P7 and P8 were composed mainly of artificial sounds other than traffic noise; and P1, P4, and P6 consisted mainly of nature sounds. Participants' evaluations of acoustic comfort levels showed that comfort was higher for P1, P4, P6, and P7; lower for P2, P4; and in the middle for P8. Meanwhile, in the facial analysis results, for P1, P3, P6, and P8, the participants' emotions improved. The recordings that had a negative impact on the participants' emotions were P2, P4, P5, and P7, largely consistent with the questionnaire results. In the questionnaire, P5 was found to have the lowest comfort levels, because P5 was located close to the main road, where the traffic was loud and continuous. In contrast, P2 was an open space facing the street, where the sound level was lower than that of P5. The location of P4 was near the same street in P2 and adjacent to P5, but the distance between P4 and the main road was farther than P5, so the amount of traffic noise was relatively small. The location of P8 was a relatively quiet street with intermittent traffic noise. At the same time, there were also music sounds, human behaviour sounds, and conversation sounds; since P8's sound level was

Table 3 Results from the questionnaire and facial analysis



relatively low, its comfort evaluation level was relatively high. The acoustic comfort of the four locations where natural sounds were prevalent was significantly higher than those where artificial sounds dominated. But, in the facial analysis, although the overall emotional level was slightly decreased, there was a slight increase at the beginning of the recording, indicating that the moderately loud artificial sounds did not have a completely negative effect on the participants' emotions.

3.2. Advantages of the facial analysis system

The results showed that the participants' emotions improved when listening to P1, P3, P6, and P8. During the process of listening to P1, P6, and P8, participants' moods improved, but then, when the recording ended, returned to levels that were close to where they were when they first started listening to the recording. Meanwhile, P3 showed a continuous increase in emotion. As P1 was recorded at the waterfront (so the sounds consisted mainly of water accompanied by music and human behaviour), participants' emotions first improved, then they tended to stabilize, indicating that water sounds have a positive effect on emotions. The improvement in mood that comes from listening to the sound of water is not sustained over time. The recording locations of P3 and P6 were community activity spaces with obvious bird sounds, human activity sounds, and conversations. In P6, the sounds of human activities were more obvious, so, when compared with P3, the degree of emotional improvement was relatively flat or even partially declined. In terms of natural sounds, different from water sounds, the degree of pleasure in birdsong gradually rises with time.

The recordings that had a negative impact on emotions were P2, P4, P5, and P7. In general, P2 'reduced' emotions, but there was nevertheless a partial increase in the process. The recording location of P2 did not consist of continuous traffic noise: the first half of the recording was quieter and the music more obvious, so the mood showed a certain upward trend; but in the second half, the traffic noise was significant and accompanied by some harsh machine noise, so participants' positive emotions were significantly reduced. These changes were not reflected in the questionnaire; when harsh noise was present, participants gave lower evaluations. In P4 and P5, the participants' emotions showed similar changes as with P1; that is, their emotions deteriorated during the process of listening to the noise but then returned to levels similar to those measured at the beginning of the recording. In P5, the reduction in mood level was more pronounced than in P4 because, as explained earlier, P4 was recorded at a location that was farther away from the main road. P4's noise level was lower than P5, but the trend observed in the participants' moods was the same. The recording location of P7 was a venue where people in the park were more active. The sound was mainly composed of human activity and some related mechanical noise. The overall sound level was not high, but there were some sudden high sound levels and high-pitched noises; as a result, participants' moods declined, which created a discrepancy with the results of the questionnaire. Because the environment was relatively quiet, the participants tended to give higher evaluations in the questionnaire, but through facial expression analysis, the emotional impact of this kind of soundscape was not necessarily shown to be positive.

4. CONCLUSION

This study used a questionnaire and the facial expression analysis system to examine the influence of acoustic landscapes on the human psychological experience in a typical public space on an urban waterfront. In the questionnaire survey, participants more positively evaluated the soundscapes where natural sounds prevailed. Evaluations of soundscapes dominated by artificial noise, such as traffic, were relatively low. Nevertheless, some artificial sound environments were also evaluated positively when sound levels were low.

In the facial expression analysis, the acoustic environment where natural sounds prevailed was shown to improve participants' emotions. However, in the water-based environment, participants tended to return to the emotional state in which they were when the recording began after a period of time, except for with the sounds of birds, which will keep people's emotions going well and will not drop at the end of the recording. Traffic and harsh mechanical noises also had a negative impact on mood and, at the end of the noise, returned participants to their initial emotional state from the beginning of the recording. However, traffic noise with low sound level may have certain positive effects. In most cases, the results of the questionnaire and facial expression analysis were consistent. However, the negative effects of some occasional noise in some of the recordings on participants' emotions may not have been reflected in the questionnaire; but, for the same recording, the resulting changes in participants' emotions over time which were not reflected in the questionnaire were measured by facial expression analysis.

Compared with the facial expression analysis system, the survey has less immediacy and is unable to reflect the impact of some transient changes in emotions. Furthermore, surveys are prone to a lack of objectivity. Facial expressions, on the other hand, are capable of displaying specific trends that are more conducive to soundscape research.

The limitations of this study were as follows:

Only 8 recordings were used. Further research should use a variety of recordings collected in different locations, including more kinds of sounds. Furthermore, as facial expressions can differ among ethnic groups, future studies should recruit diverse participants from different ethnicities to better understand the influence of ethnic and cultural factors on sound perception.

5. ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (Project Number: 51878210, 51678180, 51308145).

6. REFERENCES

1. A. L. Brown, J. Kang, T. Gjestland, ”*Towards standardization in soundscape preference assessment*”, Applied Acoustics (2011)
2. O. Bones, T. J. Cox, W. J. Davies, “*Sound Categories: CategoryFormation and Evidence-Based Taxonomies”*, Frontier in psychology (2018)
3. R. Cain, P.Jennings, John Poxon, “*The development and application of the emotional dimensions of a soundscape”*, Applied Acoustics (2013)
4. F. Aletta, Ö. Axelsson, J. Kang, “*Dimensions Underlying the Perceived Similarity of AcousticEnvironments”*, Frontier in psychology (2017)
5. J. Liu, J. Kang, T. Luo, H. Behm, “*Landscape effects on soundscape experience in city parks”*,Science of the Total Environment (2013)
6. F. Aletta, T. V. Renterghem, D. Botteldooren.*”Can attitude towards greenery improve road traffic noise perception? A case study of a highly-noise exposed cycling path”*, Euronoice (2018)
7. W. Yanga, H.J. Moon, M. Kim,*”Perceptual assessment of indoor water sounds over environmental noisethrough windows”*, Applied Acoustics (2018)

8. J. Liu, J. Kang, H. Behm, *"Birdsong As an Element of the Urban SoundEnvironment: A Case Study Concerning the Areaof Warnemünde in Germany"*, Acta Acustica United With Acustica (2014)
9. K. Filipan, B.D. Coensel, P. Aurnond, A. Can, C. Lavandier, D. Botteldooren "Auditory sensory saliency as a better predictor of change than sound amplitude in pleasantness assessment of reproduced urban soundscapes", Building and Environment (2019)
10. A. Gudi, H. EmrahTasli, T. M. den Uyl, A. Maroulis, *"Deep Learning based FACS Action Unit Occurrence and IntensityEstimation. 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) - Understanding Human Activities through 3D Sensors (UHA3DS)"*, Ljubljana, Slovenia (2015)
11. W. Sato, S. Hyniewska, K. Minemoto and S. Yoshikawa, *"Facial Expressions of Basic Emotions in Japanese Laypeople"*, Frontier in psychology (2019)
12. A. Hadinejad, B.D. Moyle, N. Scott, A. Kralj, *"Emotional responses to tourism advertisements: the application of FaceReader"*, Tourism Recreation Research (2017)
13. K.A. Leitch, S.E. Duncan, S. O'Keefe, R. Rudd, D.L. Gallagher, *"Characterizing consumer emotional response to sweeteners using an emotion terminology questionnaire and facial expression analysis"*, Food Research International (2015)
14. N.H. Frijda, *"The Emotions"*, Cambridge (UK): Cambridge University Press (1986)