forum acusticum 2026
8-12 September · Graz, Austria

TRAFFIC NOISE PERCEPTION ASSESSMENT FOR TRAFFIC MANAGEMENT: AN EMPIRICAL STUDY BASED ON FACIAL ACTION UNITS

Xuejun Hu[1], Qi Meng[1], Da Yang[1], Mengmeng Li[1], Wanshi Jiang[2]

[1] *School of Architecture and Design, 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, China*

[2] *Shenhe District Market Supervision Administration, Shenyang, China*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

As traffic noise has become a major source of noise in urban public spaces, key traffic factors such as vehicle speed and honking patterns have been shown to have a significant impact on people's perception of their environment; consequently, more refined measurement and research methods are required to support relevant traffic management. This paper proposes a method for measuring the perception of traffic noise using Action Units (AUs) based on facial expression recognition (FER), and conducts perception experiments to investigate the effects of vehicle speed and honking patterns. The results show that medium- to high-speed traffic noise significantly reduces valence (by 0.04–0.10), whilst low-speed traffic noise has no significant effect; when foreground horn sounds were introduced, the initial decrease in valence was similar across conditions, whilst the subsequent recovery following the cessation of the horn varied between 0.06 and 0.14 across different patterns. This methodology and its findings help to elucidate the non-verbal emotional responses triggered by traffic noise and provide a basis for refined traffic management measures, such as speed control and horn usage regulation.

Keywords: *urban soundscape, traffic noise, facial expression recognition*

1. Introduction

Traffic noise has now become the primary source of noise in urban public spaces[1], and affects people's physical and mental health as well as their productivity[2], [3]. With the ever-increasing number of vehicles on the roads in urban open spaces, addressing traffic noise has become an effective measure for improving the acoustic environment of urban public spaces. Vehicle speed on the road is one of the key factors; as traffic speeds decrease, the level of disturbance caused by traffic noise to pedestrians also decreases. When the speed limit is reduced to 30 km/h, noise levels (Leq) can be reduced by approximately 1 dB(A) to 4.5 dB(A)[4]. Furthermore, the combined effect of vehicle speed and traffic volume further shapes pedestrians' perception of their environment; high-speed, high-traffic scenarios often result in greater stress and reduced comfort[5].

At the same time, sudden foreground sounds such as car horns, due to their high sound pressure levels and strong modulating components, are more likely to elicit negative reactions such as annoyance and startle; existing research has also indicated that foreground events are one of the most critical factors influencing the evaluation of soundscapes[6], [7].

In current research on traffic noise, the collection of human perceptions relies primarily on questionnaires and semantic scales; however, subjective measurements have limitations such as the need for large sample sizes, a strong experimenter effect, and an inability to reflect temporal dynamics. Although physiological indicators (such as EDA, HR and EEG) are objective, they do not readily allow for the direct analysis of subtle changes in emotional experience[8]. Recent research has therefore begun to explore measurement methods that are more emotionally sensitive and capable of revealing temporal trends.

Facial expression recognition (FER) offers a more promising approach in this regard. Psychological research indicates that facial expressions account for more than half of non-verbal communication[9]. Micro-expressions last for only a fraction of a second and are difficult for participants to control, whereas al-

Corresponding author: *meng.q@hit.edu.cn.*

gorithms are often more consistent and reliable than humans when it comes to recognition[10]. Compared with traditional classification methods based on 'basic emotions', this paper employs a fine-grained facial expression coding system based on Action Units (AUs). Not only can AUs distinguish subtle differences in the same emotion across varying intensities, but combinations of AUs can also describe complex experiences that are difficult to categorise under the six basic emotions; their value has already been validated in the detection of psychological states such as stress and discomfort[11].

Against this backdrop, this paper aims to address the gap in the application of FER within the field of traffic noise research, focusing on: (1) how different vehicle speeds drive the response patterns of AUs; and (2) the differing effects of foreground noise and background noise.

2. Methodology

This study was conducted in Shenyang, Liaoning Province, China, with data collected from four different categories of urban roads in the city centre, including urban expressways, main urban roads, secondary urban roads and urban side streets, see Fig. 1. All vehicles recorded during the study were light vehicles[12]. The microphone is positioned between 1.5 metres above the road surface[13]. Recorded traffic noise, including engine noise, tyre friction and vehicle honkings, provides an accurate reflection of the environment in the traffic area. Measurements were primarily taken at times when traffic or pedestrian flow was heavy on roads surrounding urban squares, mainly between 1.00 pm and 3.00 pm, with sound sources recorded in one-hour intervals. The speed limits are 60–80 km/h on urban expressways, 40–80 km/h on urban arterial roads, 40–60 km/h on urban collector roads, and 30 km/h on urban local roads.

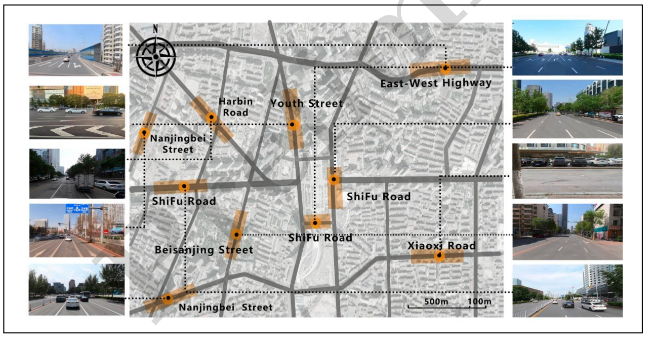

Figure 1: Recording collection locations

The recording utilised a HEAD multi-channel acoustic data acquisition front-end to capture binaural audio, whilst a zoomable handheld gimbal was used to capture corresponding video footage of the scene. To ensure the authenticity and validity of the data, the experimental procedure adhered to consistent conditions, including uniform timing of image capture, selection of sunny weather, and the use of a wide-angle lens with a roughly consistent viewing height.

For traffic noise at different speeds, audio recordings were selected from urban expressways, main urban roads and secondary urban roads. For traffic noise with different horn patterns, a 60 dB(A) traffic noise source was used as the baseline. The 60 dB(A) source was played for 10 seconds; once the subjects had adapted, a 60 dB(A) honking sound source was added. The honking sounds lasted for 10 seconds and included long-duration honking sounds (5 seconds or more), medium-duration honking sounds and short, abrupt honking sounds. The spectrogram of the honking sounds is shown in Fig. 2.

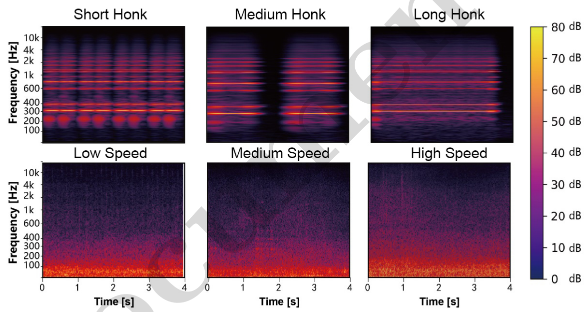

Figure 2: Spectrum plot of vehicle speed and honking patterns

The FaceReader software used in this study, developed by Noldus[14], is a programme designed for facial analysis. It can output 20 data sets per second, detect changes in 20 facial action units, and establish a more accurate facial model based on facial features specific to East Asians. The study involved a total of 33 participants, comprising 17 men and 16 women, with a mean age of 23.6 years and a standard deviation of 1.325. The layout of the screens and equipment in the laboratory is shown in Fig. 3. The experiment was conducted in the listening room, participants seated in a fixed position in front of a television screen. The experimenter gave them the instruction to watch the clip without moving their heads vigorously during the experiment because their facial expressions would be recorded. A video recording device is positioned above the television, capable of clearly capturing the subject's facial expressions. The experimenter, situated in the control room at the rear, manages the playback of the experimental video and guides the subject through the experiment. The recorded images are transmitted in real time to the FaceReader 9.0 software, which calculates facial expression data for each frame. Participants first view the visual environment of the scene on a computer screen, accompanied by background sound. After the background sound has played for 60 seconds, the main audio track begins following a 2-second transition; the audio lasts for 90 seconds.

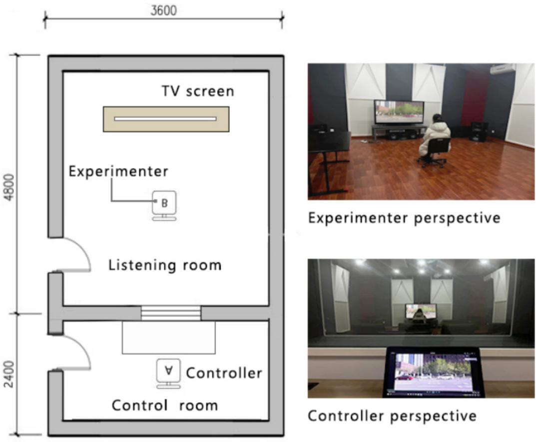

Figure 3: Laboratory layout

3. Results

3.1. Weighted Valence

According to previous research[15], the primary active units in a traffic noise environment are AU01, AU05, AU12, AU15, AU17 and AU25. The weights for each active unit were calculated, yielding the weighted summation (valence formula) as, as in Eq. (1).

$$V_a = \sum_{i=1}^{n} w_j \cdot AU_{ij}^{\text{norm}} \tag{1}$$

3.2. Traffic Speed

The valence of each second was averaged in this investigation. Determine the shift from the baseline at the present moment using the average valence when the background sound is playing as the baseline. The variation in value between traffic noise recordings at different speeds is illustrated using a violin plot, as shown in Fig. 4. A one-way analysis of variance was employed to test the significance of traffic speed on valence, with η^2 used as a measure of effect size.

The trends in changes in the valence of the participants' facial expressions under the influence of traffic noise at different speeds with a linear or quadratic regressions are shown in Fig. 5. The impact of low-speed traffic noise on the subjects was minimal, with facial expressions showing only slight negative changes that gradually stabilised over time. In contrast, medium-speed traffic noise induced a more pronounced negative trend in facial expressions during the first half-minute, followed by a gradual recovery, with the initial negative impact largely disappearing after a short period. High-speed traffic noise also caused an early negative reaction, but the recovery was slower and the negative effect persisted longer than under medium-speed conditions. These results indicate that both the magnitude and duration of facial expression responses increase with traffic speed, with medium-speed noise eliciting transient but recoverable negative reactions, while high-speed noise produces stronger and more sustained effects.

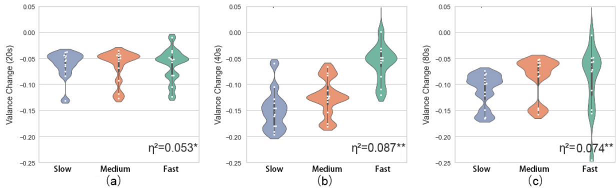

Figure 4: Violin plot: the difference effect on facial valence between traffic noise recordings of different speed after 20s(a), 40s(b) and 80s(c) of recording. η^2 is the indicator of effect size in one-way ANOVA, where *** means large effect, $\eta^2 > 0.14$, ** means medium effect,$\eta^2 > 0.06$, * means small effect, $\eta^2 > 0.01$

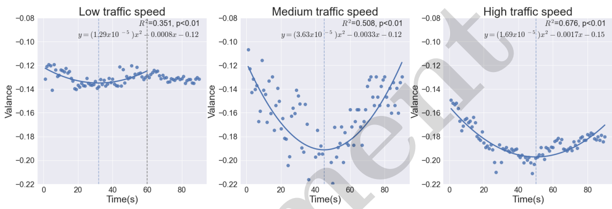

Figure 5: The trend of facial valance in 80s when listening to traffic noise recordings of different speed

3.3. Honking Pattern

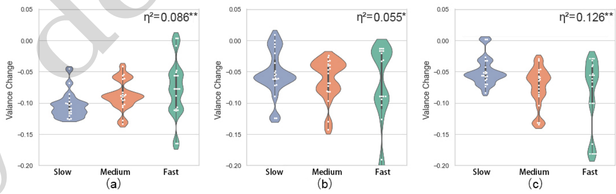

Figure 6: Violin plot: the difference effect on facial valance between traffic noise recordings of different honking patterns after 20s(a), 40s(b) and 80s(c) of recording. η^2 is the indicator of effect size in one-way ANOVA, where *** means large effect, $\eta^2 > 0.14$, ** means medium effect,$\eta^2 > 0.06$, * means small effect, $\eta^2 > 0.01$

Three types of horn sounds—a short honk, a medium-length honk and a long honk—were introduced 10 seconds after the onset of traffic noise at 60 dB(A). Independent-samples t-tests were conducted to examine changes in facial expression valence before and after the whistling started and ended. The results, as shown in Table 1, indicate that the onset of whistling led to a significant decrease in valence across all whistling modes, with the largest negative change observed for short whistling, followed by long and medium modes. After the whistling ended, valence scores generally rebounded, with short and medium whistling showing significant recovery, while long whistling showed a minimal and non-significant increase. Overall, these findings suggest that short and medium whistling have more pronounced and recoverable effects on emotional responses, whereas long whistling elicits weaker changes.

The variation in value between traffic noise recordings

Table 1: Difference in facial expression valence between traffic sound before and after the whistling starts, and between traffic sound before and after the whistling ends (independent samples T-test, reporting the level of Cohen's effect size d).

Honking Pattern	Before and after the whistling starts		Before and after the whistling ends	
	Mean Difference	T value	Mean Difference	T value
Short	-0.051	-4.71**	0.037	3.259**
Medium	-0.021	-2.01*	0.031	2.450*
Long	-0.023	-2.87*	0.007	1.065

Note: ** means medium effect, $d > 0.4$; * means small effect, $d > 0.2$

at different honking patterns is illustrated using a violin plot, as shown in Fig. 6. A one-way analysis of variance was employed to test the significance of honking patterns on valence, with η^2 used as a measure of effect size. Linear regressions were performed before the introduction of the horn sounds, during their introduction, and after they ceased; the results are shown in Fig. 7. The regression results show that prior to the horn, facial expressions were gradually declining, consistent with early trends in other experimental groups. The introduction of the horn caused an immediate drop in valence across all groups, with short and medium horns inducing stronger negative responses than the long horn. After the horn ceased, all groups exhibited a gradual recovery, though the long horn group recovered more slowly. Overall, these results suggest that shorter and medium-duration horns produce more pronounced but recoverable negative effects, whereas prolonged horns have a milder immediate impact but lead to slower recovery.

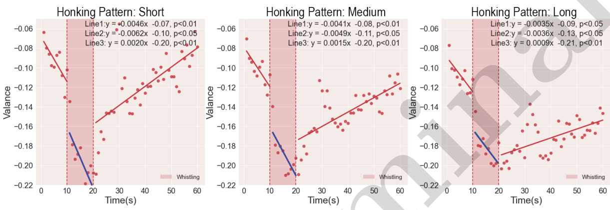

Figure 7: The trend of facial valance in 80s when listening to traffic noise recordings of different honking patterns

4. Discussion

By incorporating facial expression recognition technology into research on the perception of traffic noise, this study provides an objective, dynamic and quantifiable monitoring method for assessing the acoustic environment of urban public spaces. However, the current study has several limitations: firstly, as the facial expression analysis system used in this study can only process clear, front-facing facial images, the research was confined to a laboratory setting to examine changes in participants' perceptions under auditory stimuli, thereby introducing potential interference from the laboratory environment; Secondly, this study only considered the influence of the auditory environment, whereas in real-world settings, multiple sensory factors such as visual and olfactory cues also interact to shape an individual's perception of traffic noise; finally, in the investigation of foreground noise, only the single noise event of vehicle honking was examined, whilst other noise events—such as the passage of heavy vehicles[16]—also significantly affect human environmental perception.

Future research could further incorporate multisensory stimuli, real-world testing, and a wider range of noise events to enhance the external validity of facial expression analysis in complex acoustic environments. Concurrently, by integrating mobile imaging with multimodal models, the scope of FER applications could be expanded to achieve dynamic soundscape assessments with higher ecological validity.

5. Conclusion

This study explored the feasibility of facial expression analysis technology in assessing the perception of traffic noise in public spaces, and revealed the patterns by which urban traffic noise affects pedestrians' perceived quality of life. Through controlled experiments, the study systematically analysed the effects of different vehicle speeds and horn-honking patterns on pedestrians' facial expressions. The results indicate that traffic noise at moderate speeds causes noticeable negative changes in participants' facial expressions within a short period; however, as the stimulus persists, the negative impact gradually diminishes and stabilises. As a typical form of foreground noise, horn-blowing events instantly exacerbate negative emotional responses, with short, medium-intensity horn blasts having a more pronounced effect on facial expressions; following the cessation of the horn, the subjects' expressions showed a gradual recovery trend, indicating a certain capacity to recover from sudden acoustic events.

From the perspective of traffic management, this study suggests that unnecessary honking in public spaces—particularly brief honks—should be minimised, and that measures such as speed limits and traffic diversion should be implemented to reduce the frequency of sudden noise incidents. Furthermore, incorporating facial expression analysis into the acoustic environment assessment system is expected to provide a more responsive and people-centred monitoring tool for traffic noise management, whilst offering guidance for the formulation of more refined acoustic environment management strategies.

6. Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) [grant numbers 52478083, 52208101, 52308089]; and Shanghai Key Laboratory of Urban Design and Urban Science, NYU Shanghai Open Topic Grants [Grant No. 2022QMeng_LOUD].

References

[1] M. Z. Salim et al., “A comprehensive review of navigating urbanization induced climate change complexities for sustainable groundwater resources management in the indian subcontinent,” *Groundwater for Sustainable Development*, vol. 25, p. 101 115, 2024.

[2] V. Nedic, D. Despotovic, S. Cvetanovic, M. Despotovic, and S. Babic, “Comparison of classical statistical methods and artificial neural network in traffic noise prediction,” *Environmental Impact Assessment Review*, vol. 49, pp. 24–30, 2014.

[3] I. van Kamp, S. Simon, H. Notley, C. Baliatsas, and E. van Kempen, “Evidence relating to environmental noise exposure and annoyance, sleep disturbance, cardio-vascular and metabolic health outcomes in the context of igcb (n): A scoping review of new evidence,” *International journal of environmental research and public health*, vol. 17, no. 9, p. 3016, 2020.

[4] R. Kim et al., “Burden of disease from environmental noise,” in *WHO International Workshop on “Combined Environmental Exposure: Noise, Air Pollutants and Chemicals” Ispra*, 2007.

[5] D. Andersson, L. Wahlgren, K. S. E. Olsson, and P. Schantz, “Pedestrians’ perceptions of motorized traffic variables in relation to appraisals of urban route environments,” *International journal of environmental research and public health*, vol. 20, no. 4, p. 3743, 2023.

[6] F. Aletta, J. Kang, and Ö. Axelsson, “Soundscape descriptors and a conceptual framework for developing predictive soundscape models,” *Landscape and Urban Planning*, vol. 149, pp. 65–74, 2016.

[7] J. Foellmer and N. Tyler, “Multisensory approaches to urban pollution: A controlled study of sound-odour interactions and their impact on physiology and perception,” *Building and Environment*, p. 113 506, 2025.

[8] X. Lu, J. Tang, P. Zhu, F. Guo, J. Cai, and H. Zhang, “Spatial variations in pedestrian soundscape evaluation of traffic noise,” *Environmental Impact Assessment Review*, vol. 83, p. 106 399, 2020.

[9] A. Mehrabian, “Communication without words,” in *Communication theory*, Routledge, 2017, pp. 193–200.

[10] P. Ekman, *Telling lies: Clues to deceit in the marketplace politics and marriage*. WW Norton & Company, 2009.

[11] D. Kollias, “Abaw: Valence-arousal estimation, expression recognition, action unit detection & multi-task learning challenges,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2328–2336.

[12] A. Barros et al., “Beyond noise levels: Vehicle classification using psychoacoustic indicators from pass-by road traffic noise and their correlations with speed and temperature,” *Applied Acoustics*, vol. 214, p. 109 716, 2023.

[13] *Acoustics – measurement of the influence of road surfaces on traffic noise – part 1: Statistical pass-by method*, ISO Standard, 2023. [Online]. Available: `https://www.iso.org/standard/83223.html`

[14] A. Gudi, H. E. Tasli, T. M. Den Uyl, and A. Maroulis, “Deep learning based facs action unit occurrence and intensity estimation,” in *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, IEEE, vol. 6, 2015, pp. 1–5.

[15] X. Hu, Q. Meng, D. Yang, M. Li, and J. Kang, “Characterizing urban soundscapes via facial action coding system,” *Sustainable Cities and Society*, p. 106 918, 2025.

[16] Y. Hao, J. Kang, and H. Wörtche, “Assessment of the masking effects of birdsong on the road traffic noise environment,” *The journal of the Acoustical Society of America*, vol. 140, no. 2, pp. 978–987, 2016.