DOI: 10.61782/fa.2025.0804

FORUM ACUSTICUM EURONOISE 2025

WALKING VS. STANDING: HOW DYNAMIC VISUALS IMPACT SOUNDSCAPE PERCEPTION

Kening Guo[1*] **Jian Kang**[1]
[1] Institute for Environmental Design and Engineering,
The Bartlett, University College London (UCL), UK

ABSTRACT*

We spend about 3% to 12% of our lives walking often perceiving our surroundings in motion, especially on urban streets. However, the vast majority of soundscape assessments are focused on stationary observers, overlooking real-world dynamics. The mechanisms of soundscape perception during movement remain unclear. Therefore, this study explores whether walking visuals, as opposed to stationary ones, influence soundscape assessments. Several acoustically representative London streets were presented under three visual conditions (walking, stationary, still images) with identical audio, in a semi-anechoic laboratory. Twenty participants assessed the soundscape under all audiovisual conditions using both retrospective questionnaires and real-time PAQ (Perceived Affective Quality) feedback. The results showed that sample-level trends in overall sound quality and loudness were evident across walking, stationary, and still image conditions, though statistical significance was limited in certain streets. Real-time data revealed perceptual variation in the time domain, with early divergences and later convergence, implying adaptive perceptual processes. The findings highlight the sensitivity of human perception to subtle shifts between dynamic and static cues in complex urban scenes. Perception, it seems, can hear motion—though adaptation may soften its effects. A reminder that designing urban soundscapes means designing for both sensitivity and adaptation.

Keywords: *soundscape assessment, motion, perceptual adaptation, temporal dynamics, urban street.*

*__*Corresponding author*__: kening.guo.23@ucl.ac.uk

1. INTRODUCTION

Audio-visual interaction is a well-established concept in both urban and indoor soundscape research. Due to multisensory integration in the brain, what we see can significantly influence what we hear. For example, studies have demonstrated that urban soundscapes with water features are preferred over those without [1], and that greenery contributes to greater pleasantness in indoor environments compared to settings lacking vegetation [2]. However, unlike sitting indoors or in parks, people in urban outdoor spaces, especially on streets—are more likely to be walking than standing still. Most audio-visual soundscape studies are conducted from a static perspective, whether in laboratory experiments or in situ observations [3]. But how does soundscape perception change our experience space from a forward-moving visual perspective? Does visual motion matter? This remains an open question.

Encouragingly, neuropsychological studies suggest that physical movement can enhance auditory sensitivity [4], and that even visual motion alone can modulate perceived loudness. For example, after just a few minutes of watching a simple geometric shape (e.g., a square) move in depth, a steady auditory tone may be perceived as gradually increasing or decreasing in loudness, in the opposite direction of the visual motion [5]. Yet most of these findings are based on animal models or simplified stimuli. How such effects unfold in complex, real-world soundscapes is still unclear.

Therefore, this study explores whether dynamic visual environments influence soundscape perception. To address this, three sub-questions are considered. First, which key indicators of soundscape perception are affected, including those defined in ISO/TS 12913-3 [6]. Second, to what extent do factors beyond street identity explain differences in the impact of dynamic visuals on soundscape perception. Third, in addition to the overall impact, what temporal differences can be observed in the instantaneous impact on PAQ (Perceived Affective Quality)?

2. METHODOLOGY

This study adopts a within-subjects design with repeated measures. First, over a dozen acoustically representative urban streets in London were selected based on their sound characteristics. For each street, both stationary and walking binaural recordings were made using SQobold, accompanied by 360-degree video recordings using a GoPro Max. After careful selection, audio segments were paired with three visual conditions representing the same location: walking visuals, stationary visuals, and still images. These audio-visual stimuli were then randomly presented to participants in a semi-anechoic chamber (see Fig. 1).

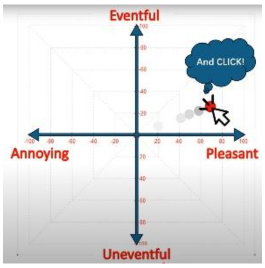

Figure 1. (Left) Audio Lab at UCL Here East.
Figure 2. (Right) Mouse-tracking interface.

A total of 20 adult participants with no reported hearing impairments were invited to take part in the listening experiment. From the moment each video began, participants were asked to continuously provide real-time feedback on their perceived sound environment using a two-dimensional Perceived Affective Quality (PAQ) scale by moving and clicking their mouse. At the end of each video, participants completed a short questionnaire assessing perceived loudness, perceived sound sources, overall PAQ, and overall sound environment quality. Each participant completed the full set of soundscape assessments across all street scenes and visual conditions. The order of audiovisual clips was fully randomized for each individual to avoid potential order effects.

A custom-designed mouse-tracking interface, developed by the author, enabled participants to continuously express changes in their emotional impressions along two perceptual axes—eventfulness and pleasantness (see Fig. 2).

Thus, continuous PAQ responses (expressed as XY coordinates), together with mouse clicks and scale ratings, were recorded for each audiovisual clip and used in subsequent analysis.

3. RESULTS

Perceptual differences between walking and stationary conditions were analyzed at two levels: (1) overall comparison based on sample-level variation and statistical testing, and (2) real-time perceptual dynamics over time.

3.1 Overall comparison: walking vs. stationary

At the sample level, the means and standard deviations of each perceptual indicator were compared across the three conditions for each street. For overall sound environment quality, more than half of the streets showed higher ratings in the moving-forward condition than in the stationary condition, while over 60% received higher ratings in the stationary condition than in the static image condition. The average difference was slightly greater between stationary and static (0.19) than between moving and stationary (0.15). In most streets, rating variance decreased from static image to stationary to moving-forward, indicating more consistent evaluations under dynamic visual input. (e.g., Malet Street, see Fig. 3).

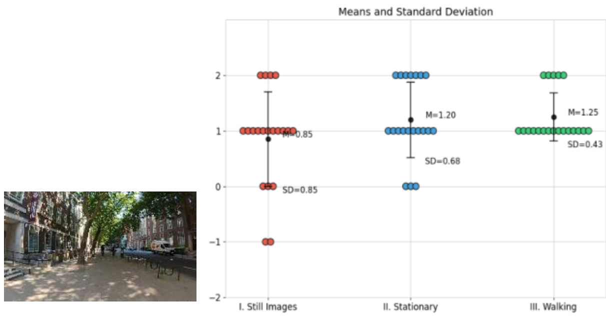

Figure 3. Ratings for overall sound environment quality (Malet St., three visual conditions).

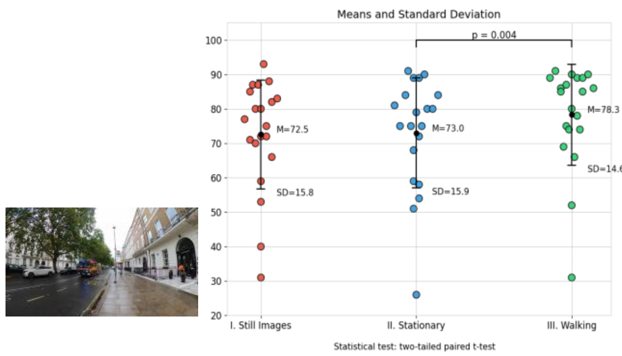

Figure 4. Ratings for perceived loudness (A4201, three visual conditions).

For perceived loudness, approximately two-thirds of the streets were associated with lower perceived loudness in the moving-forward condition than in the stationary condition. In over half of the streets, the stationary condition elicited greater perceived loudness than the static image condition. The difference was larger between moving and stationary (2.8) than between stationary and static (1.6). No consistent pattern was found in rating variance across conditions. (e.g., A4201, see Fig. 4).

Building on this, paired t-tests were used to assess overall statistical differences across all participants for each street. For perceived loudness, fewer than one-third of the streets showed significant differences between the walking and stationary conditions ($p< 0.05$). For ISO Pleasantness, less than half of the streets demonstrated significant differences between the walking condition and either the stationary or static image conditions. For ISO Eventfulness, only one street showed a significant difference between walking and stationary conditions.

These findings suggest that while walking may influence soundscape perception in certain contexts, the effect is not uniformly observed across all environments. Further examination of the streets with statistically significant differences indicates the presence of shared features—either acoustic (e.g., dominant sound source types, loudness patterns) or non-acoustic (e.g., visual complexity, spatial openness)—which may function as mediating variables beyond street identity itself.

In contrast, the majority of streets did not yield statistically significant differences across conditions. Two possible explanations are considered: first, that the sample size may have limited statistical power to detect subtle effects (with more participants currently being added to the study.); and second, that retrospective questionnaire responses may not be sensitive enough to capture dynamic or short-lived perceptual changes. This highlights the added value of real-time emotional data, which enables the identification of finer-grained perceptual patterns that might otherwise remain undetected in condition-level averages.

3.2 Variation in the time domain

To further explore subtle perceptual shifts, real-time emotional quality data from the mouse-tracking interface were analyzed. By segmenting the data into three temporal phases—F10s (first 10 seconds), M10s (middle 10 seconds), and L10s (last 10 seconds)—we examined how emotional responses evolved over time within each condition. This temporal dimension helped reveal dynamic changes that may have been obscured in average-level analyses, providing deeper insight into the role of bodily movement in soundscape perception.

Using street A4200 as an example, real-time PAQ data from one participant revealed that while ratings across the three conditions converged in the last 10 seconds, notable differences in both mean and variance were observed in the first 10 seconds (see Fig. 5). These patterns underscore the value of real-time data in revealing perceptual dynamics that static ratings may overlook. Moreover, this shift from early divergence to later similarity may reflect an adaptive process in soundscape perception over time.

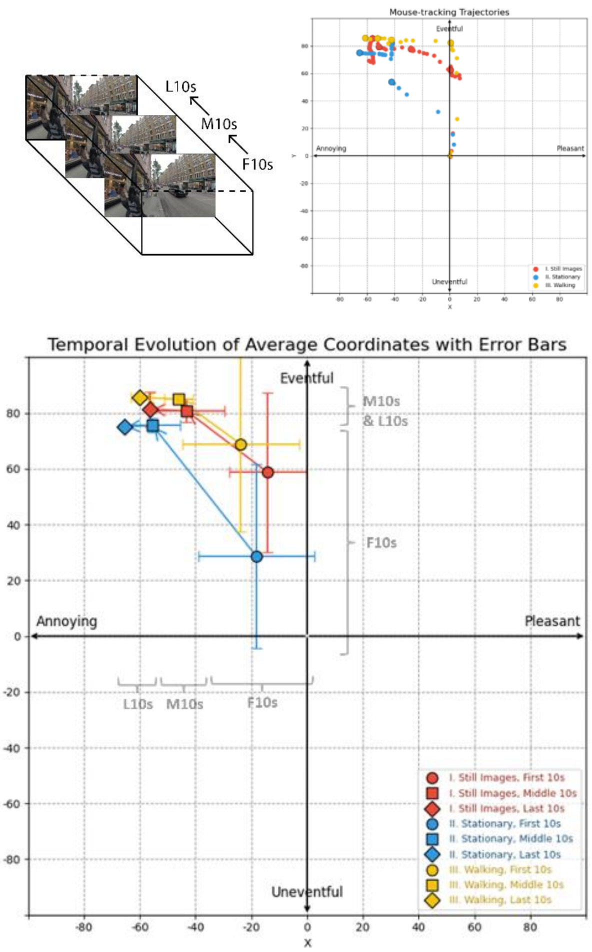

Figure 5. Dynamic PAQ ratings over time (A4200, three visual conditions, F10s/M10s/L10s).

4. DISCUSSION AND CONCLUSION

In complex real-world environments, there is no "one-size-fits-all" explanation for perceptual challenges. Human responses to soundscapes are shaped by a combination of cognitive, emotional, and situational factors that extend well beyond what physical acoustics alone can account for. Across different street scenes, the impact of dynamic versus static conditions on soundscape perception varied considerably. While some streets showed significant differences at the group level, others did not. This variation suggests that the effect cannot be fully explained by street identity alone. Therefore, further research is needed to explore which underlying factors—beyond the street label itself—contribute to these perceptual differences.

Interestingly, temporal-level analysis revealed transient emotional patterns that emerged at specific moments during the audiovisual experiences such as initial responses to a stimulus, a gradual adaptation phase in the middle, and a recalibration of perception toward the end. These fluctuations could be easily masked in aggregated scores, suggesting that perception operates in a more layered and context-sensitive manner than previously assumed. More importantly, these findings closely align with existing research highlighting the remarkable adaptability of human perception under changing conditions—an adaptability often exceeds conventional expectations. For instance, visual adaptation does not follow a linear course; its timescale depends on the accumulation of sensory evidence and the nature of stimulus transitions [7]. Similarly, sound localization in real-world settings involves flexible responses to reverberation and background noise, shaped through prolonged exposure to naturalistic acoustic environments [8].

This study contributes to a deeper understanding of how visual motion may interact with multisensory input to shape evolving impressions of soundscapes. It highlights that perception is not a passive reflection of external stimuli, but a dynamic and context-sensitive process shaped by continuous sensory integration. Human perception is inherently dynamic and becomes even more intricate when shaped by real-world conditions—continually regulated not only by environmental factors, but also by bodily and attentional mechanisms. Yet much remains to be understood about how perception adapts and recalibrates in response to the complexity, ambiguity, and multisensory nature of real-world environments.

5. ACKNOWLEDGMENTS

We are grateful to Dr. Tin Oberman, Gizem Esra Ünlü and the Acoustics Research Group at UCL for their invaluable support and insightful discussion throughout this study.

6. REFERENCES

[1] J. Y. Jeon, P. Lee, J. You, and J. Kang, 'Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces', *The Journal of the Acoustical Society of America*, vol. 131, pp. 2101–9, Mar. 2012, doi: 10.1121/1.3681938.

[2] A. Latini *et al.*, 'Virtual reality application to explore indoor soundscape and physiological responses to audio-visual biophilic design interventions: An experimental study in an office environment', *Journal of Building Engineering*, vol. 87, p. 108947, Jun. 2024, doi: 10.1016/j.jobe.2024.108947.

[3] J. Y. Hong *et al.*, 'Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications', *Building and Environment*, vol. 149, pp. 1–14, Feb. 2019, doi: 10.1016/j.buildenv.2018.12.004.

[4] D. M. Schneider, A. Nelson, and R. Mooney, 'A synaptic and circuit basis for corollary discharge in the auditory cortex', *Nature*, vol. 513, no. 7517, pp. 189–194, Sep. 2014, doi: 10.1038/nature13724.

[5] N. Kitagawa and S. Ichihara, 'Hearing visual motion in depth', *Nature*, vol. 416, no. 6877, pp. 172–174, Mar. 2002, doi: 10.1038/416172a.

[6] ISO, 'ISO/TS 12913-3:2019', ISO. Accessed: Jun. 04, 2024. [Online]. Available: https://www.iso.org/standard/69864.html

[7] B. Wark, A. Fairhall, and F. Rieke, 'Timescales of Inference in Visual Adaptation', *Neuron*, vol. 61, no. 5, pp. 750–761, Mar. 2009, doi: 10.1016/j.neuron.2009.01.019.

[8] A. Francl and J. H. McDermott, 'Deep neural network models of sound localization reveal how perception is adapted to real-world environments', *Nat Hum Behav*, vol. 6, no. 1, pp. 111–133, Jan. 2022, doi: 10.1038/s41562-021-01244-z.