forum acusticum 2026
8-12 September · Graz, Austria

Interacting with soundscapes in Virtual Reality: desktop vs. immersive VR

Yanaky, Richard[1,2], Sydney Toffa[1,2], Guastavino, Catherine[1,2]

[1]School of Information Studies, McGill University, Canada

[2]Centre for Interdisciplinary Research in Music, Media, and Technology, Montreal, Canada

Abstract

Interactive Virtual Reality can make urban soundscapes more tangible, but how does interaction modality shape user experience? This study compares desktop and immersive versions of City Ditty, an interactive soundscape simulator designed for soundscape learning, exploration, and rapid prototyping by non-specialists. Both versions provided equivalent functionality and identical binaural audio rendering, while differing in modality, namely display (computer screen vs. Head-Mounted Display), navigation (walking/turning vs. walking/turning/teleportation), and controllers (keyboard vs. controller).

Thirty-three participants completed a counterbalanced within-subjects evaluation comprising 13 structured tasks per modality. Post-session questionnaires measured user experience with each modality, followed by a comparative questionnaire.

Participants strongly preferred immersive VR for overall enjoyment, visuals, and sound, and reported greater focused attention, reward, overall presence, and visually supported presence. The preference for sound is particularly notable because the auditory presentation remained constant across modalities, indicating that the broader visual, spatial, and interactional context shaped the listening experience. Preferences for operational activities were less decisive, cautioning against treating either modality as universally preferable. These preliminary findings support maintaining interoperable desktop and immersive versions that can accommodate different users, activities, resources, and stages of soundscape work. This flexibility may broaden participation while complementing specialist acoustic expertise.

Keywords: *soundscape, auralization, virtual reality, human-computer interaction, urban planning*

Corresponding author: *catherine.guastavino@mcgill.ca*

1. Interactive Soundscape Simulation

Auralization in virtual reality offer new ways of experiencing urban soundscapes that go well beyond noise maps and sound level metrics. Considerable advances have been made in the physical modelling and audiovisual rendering of urban environments. For example, Llorca-Bofí, et al. [1] developed a modular framework combining physically based sound-source and propagation models, recorded and simulated sounds, spatial audio reproduction, and three-dimensional visualization. Such work addresses the technical challenges of constructing and presenting high-fidelity urban auralizations for perceptual research, planning, and public communication.

These advances have primarily emphasized the representation itself, including its physical plausibility, calibration, audiovisual fidelity, and suitability for controlled evaluation. Such environments are typically produced through specialist modelling workflows, while other users experience them as listeners or evaluators within expert-constructed scenes. Comparatively less attention has been given to the interactional form of soundscape simulation: how non-specialists might independently explore environments, manipulate sound sources, compare alternative configurations, and construct designs of their own. This is a complementary challenge rather than an alternative to technically sophisticated auralization. High-quality representation determines what can be experienced, while accessible interaction determines what different users can do with that representation.

Addressing this interactional challenge is important for broader soundscape adoption. Specialist expertise remains essential for advanced acoustic modelling and evaluation, but expert-led processes are difficult to apply routinely across the many decisions that shape urban environments. Municipalities and professional teams may lack continuous access to the expertise, resources, or specifications required for detailed auralization. More accessible tools could enable practitioners to consider sound earlier, explore possible interventions in the early stages of design, and better communicate their needs and alternatives with both specialists and stakeholders, including decision makers, practitioners from other built-environment disciplines, and residents or other users of the urban space.

City Ditty is an interactive soundscape simulator developed to investigate this interactional and uptake challenge. It enables users to learn soundscape concepts, explore audiovisual environments, manipulate individual sound sources, rapidly prototype alternatives, and construct scenarios for controlled evaluation. Rather than replacing advanced auralization or acoustic expertise, City Ditty investigates how accessible interaction can broaden participation in soundscape-related activities and help practitioners bring their own disciplinary requirements and perspectives into city-planning processes.

The field has made major advances in how urban soundscapes can be simulated. An increasingly important question is how people can interact with those simulations.

1.1 Developing interactive simulation practices

Developing City Ditty presented an unusual user-centered design problem. At the outset, there was no established professional role or common workflow in which non-acousticians independently used interactive simulation to explore and design urban soundscapes. The intended practice could therefore not simply be observed and improved through conventional requirements gathering. City Ditty was instead developed as both a functional system and a vehicle for investigating what interactions, functionalities, and forms of support might enable such practices to emerge. The project adopted Professionals of the Built Environment (PBEs), including urban planners, designers, and policy makers, as a broad prospective user group, particularly focusing on those with little or no formal training in sound or acoustics.

Initial research examined whether non-specialists could learn to use the simulator meaningfully. The first prototype used a desktop configuration consisting of a monitor, keyboard, mouse, and headphones. The soundscapes were auralized using stereo panning, distance propagation, and Doppler effects. In a first usability study, participants completed a scaffolded sound-awareness session that progressively introduced navigation, interface controls, soundscape principles, and forms of sound intervention [2]. Most participants completed the activities, culminating in the creation of their own soundscape designs in under one hour, providing initial evidence that soundscape could be made learnable for non-specialists, with proper support.

Subsequent work incorporated binaural audio rendering using an enhanced acoustic model (including occlusion, absorption and different reverberation conditions) [3]. We also examined broader barriers to adoption [3] and extended City Ditty from desktop to immersive VR [4] with headset, and identified design considerations and recommendations for interactive soundscape simulation [5]. Together, these efforts produced a mature research prototype supporting experiential learning, independent exploration, rapid prototyping, communication, and controlled perceptual experimentation.

Early applications began to demonstrate how these capabilities could support different forms of research and practice. A soundscape researcher used City Ditty to construct and operate scenarios for a controlled study examining the effects of music on urban public experience [6]. In a place-based design context, a multidisciplinary team used the simulator to support the biophilic redesign of a park in Chile, creating a localized digital environment and incorporating recordings of local sound sources [7, 8]. These cases demonstrate how direct scenario creation can support both controlled experimentation and collaborative, context-specific design.

A separate stakeholder study examined the broader professional implications of interactive soundscape simulation [9]. A diverse sample of PBEs identified opportunities for experiential learning, communication across professional roles, rapid exploration of alternatives, and the introduction of soundscape considerations into existing workflows. The findings also showed that uptake depends on more than a functional tool. Capacity building, role-appropriate support, organizational resources, and conditions for integrating new practices are also required.

Collectively, this work indicates that non-specialists can learn to manipulate soundscape simulations and that these capabilities can support research, situated design, and professional capacity building. However, the desktop and immersive versions of City Ditty offer substantially different forms of viewing, listening, spatial orientation, navigation, and forms of manipulation. Although they provide the same core functionality and identical binaural audio rendering, their different levels of immersion and interaction modality may shape how users experience both the system and the soundscape itself. The present study examines this question by comparing user experience across the two versions. This paper reports selected preliminary questionnaire findings.

2. Comparative VR Study

2.1 Participants and apparatus

33 participants (21 women, 11 men, 1 other, mean age 26.5) were recruited from the Montreal area through university emailing lists and postings. They were required to be aged 18+ with normal or corrected-to-normal vision and hearing to operate the software. The study compared desktop and immersive versions of City Ditty. The desktop version ran on a 24-inch HD monitor using a standard keyboard and mouse, while the immersive version used a Varjo XR-3 headset with HTC Vive controllers. Both versions presented audio through AKG K702 open back headphones. The two versions provided the same system functionality and can interchangeably open and use soundscape designs made on either version. Users could navigate the environment, inspect and manipulate individual sound sources, as well as modify broader contextual conditions such as time of day, weather, and season. Binaural audio was rendered in real time using the Wwise audio engine, including distance attenuation, reverberation, and occlusion. The same binaural rendering pipeline and audio assets were used in both conditions, as described in [3].

Interactional differences between the two modalities consist of 1) navigation methods: on desktop the user moves about the virtual space with arrow or WASD keys, whereas the immersive version supports natural walking and turning movement in a VR play space, supported by standard VR teleportation methods to travel over greater virtual distances. 2) The head-mounted display and VR controllers for the immersive version are each tracked with six degrees of freedom, supporting natural movement and interactions (e.g. grabbing and rotating objects in 3D). In contrast, the desktop version uses additional keyboard and mouse inputs to provide the same functionality, including changes in elevation and rotation. 3) The menu system is accessed through a standard Heads-Up Display on desktop, while the immersive version utilizes moveable, popup menus in the virtual space. The layout differs on the top menu level but is otherwise identical afterwards. Both versions used the same underlying virtual city model, centred on a small urban park (Figure 3). Each session used a different preconfigured soundscape within the same physical setting of a city park, modifiable by participants.

Figure 1. Immersive VR version of City Ditty.

Figure 2. Desktop version of City Ditty

The virtual space consisted of a small city centre park from which the city expands outwards for approximately six blocks on all sides. For this study, only the park (Figure 3) was populated and used using different soundscape designs across sessions, changing both through loading different pre-set designs, and through the user's actions. After user intervention, no soundscape was alike. To give a better feel for this space, Figure 3 is provided from Miao, et al. [7], who used City Ditty for a comparative soundscape experiment using pre-set soundscapes from this study and pre-built city landscape from Yanaky, et al. [3].

Figure 3. Overhead view of the city park space. Study tasks utilized the full space, while re-design tasks took place in sections 1 and 4. Image from [7].

2.2 Procedure

Participants completed two 30-minute sessions, one using the desktop version of City Ditty and one using the immersive version. A within-subjects design was used, with the order of the two modalities counterbalanced across participants. Both sessions took place in the CIRMMT Audiovisual Editing Lab. The immersive session used an approximately 3 m × 2 m play space.

Each session consisted of 13 structured tasks delivered through on-screen text. Participants received no prior training or exposure to City Ditty and learned to use the system by following these instructions. The tasks introduced controls, supported exploration and listening of the soundscape, required participants to find, add, and manipulate sound sources, and complete the redesign of a park space.

Two distinct but comparable task sets were created within the same virtual park, each using a different soundscape design. Each task set was adapted for both modalities so that the task objectives and system functions remained consistent, while the instructions reflected the different desktop and immersive control schemes. This produced four task-set versions in total. Both modality order and task-set assignment were counterbalanced, ensuring that each soundscape design could be experienced using either modality and that each participant encountered both task sets only once. This reduced repetition and prior-exposure effects while preserving comparable activities and a consistent virtual environment across sessions.

After each session, participants completed a short questionnaire on their experience interacting with City Ditty. After completing both sessions, they completed an additional questionnaire comparing between the two modalities. They received $15 for their time.

Task-completion data and user-action logs were also recorded for each task, but their analysis is beyond the scope of this paper. Here, we focus on the preliminary questionnaire results in terms of preference and user experience.

3. Analysis

3.1 Comparative evaluations

After completing both sessions, participants compared the desktop and immersive versions using a five-point preference scale ranging from strongly preferring immersive to strongly preferring desktop. Figure 4 presents the percentage distribution for each scale.

The clearest preferences concerned the overall experience. Enjoyment and visuals for immersive ranked very highly, while there was a very strong preference for sound when in immersive, despite no audible difference. Following this, there is a close split in preferences between immersive and desktop versions for items related to operating the system, including navigation, controllers, manipulating sound source objects, re-design of the park space, finding sound source objects, and reading. This preference appears to drop more when interactions shift away from spatial interactions to more standard operations like reading and operating menus. Likewise, a broader perspective from a monitor can be helpful for locating objects vs. in-situ/first person perspective.

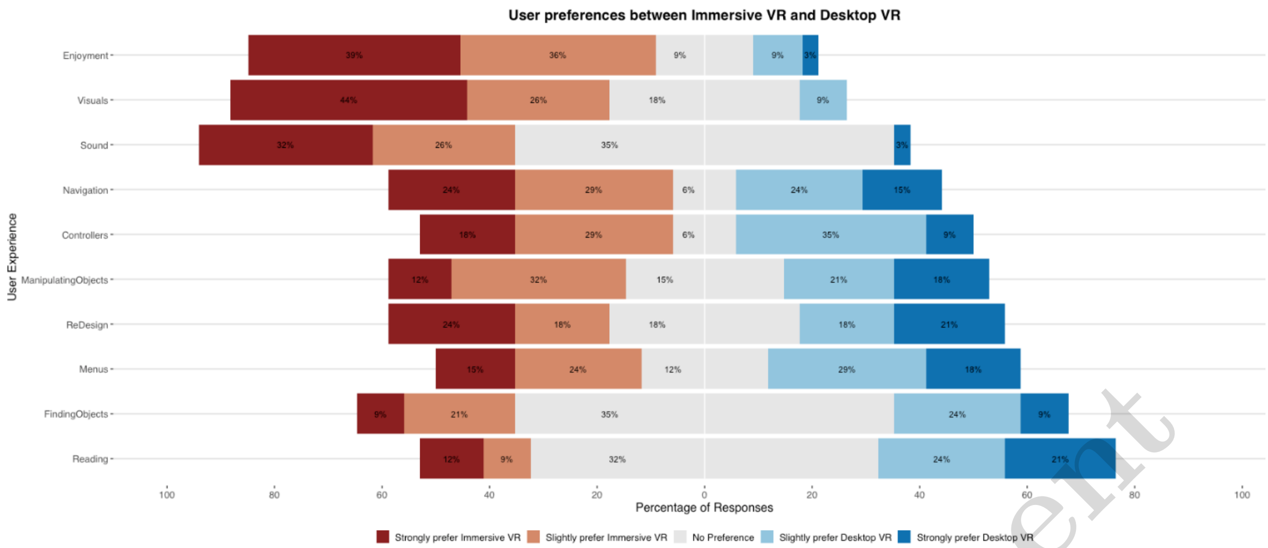

Figure 4. Comparative evaluation of immersive VR and desktop VR ordered from strongest preference for immersive VR to strongest preference for desktop VR, collapsing over all 36 participants.

3.2 User experience scales

After each session, participants evaluated their experience with a questionnaire including the User Engagement Scale–Short Form (UES-SF) [10], a scale measuring experience adapted here to VR, as well as scales on Presence, Enjoyment and Navigation with each modality. The full questionnaire is presented in Table 1. The mean participant ratings are shown in Figure 5. A one-way ANOVA revealed a significant main effect of modality (immersive vs. desktop) on user experience scale ratings ($F(1,11) = 4.13$, $p<0.002$). Post hoc tests for each scale are reported below with a p-value of 0.05.

I lost myself in this experience The time I spent using the software just slipped away I was absorbed in this experience	Focused attention (FA) *(Engagement)*
I felt frustrated while using City Ditty I found City Ditty confusing to use Using City Ditty was taxing	Perceived usability (PU) *(Engagement)*
The virtual environment was attractive The virtual environment was aesthetically appealing The virtual environment appealed to my senses	Aesthetics (AE) *(Engagement)*
Using City Ditty was worthwhile My experience was rewarding I felt interested in this experience	Reward (RW) *(Engagement)*
The sounds made me feel like I was there in the virtual environment	Presence (P1)
I felt like I was there in the virtual environment	Presence (P2)
The visuals made me feel like I was there in the virtual environment	Presence (P3)
I enjoyed the sounds in the virtual environment	Enjoyment (E1)
I enjoyed the visuals in the virtual environment	Enjoyment (E2)
I found it easy to navigate the virtual environment	Navigation (N1)
I felt disoriented or lost in the virtual environment	Navigation (N2)

Table 1. User experience scales

3.2.1 User Engagement scales

User engagement measures cognitive, affective and behavioural aspect of interaction it was evaluated for each version using the UES-SF questionnaire, which consists of 12 scales (in yellow in Table 1) grouped into 4 components for the analysis: Aesthetics (refers to the sensory appeal), Focused Attention (when users feel absorbed in the interaction and lose track of time), Perceived Usability, and Reward.

Post hoc tests revealed significant differences for Focused Attention ($p=0.002$) and Reward ($p=0.027$). Immersive was rated higher in terms of Focused Attention, which refers a flow state and losing track of time in one's activity. Similarly, using the immersive version was found to be a more rewarding experience. No significant differences were observed in terms of Perceived Usability and Aesthetics, despite a strong preference for immersive VR in terms of visuals and sound in the comparative evaluations.

3.2.2 Presence, enjoyment and navigation

Presence (3 scales in green in Table 1) refers to the illusion of being situated within the virtual environment rather than merely observing it. No significant difference was observed for sound-based presence (P1) but immersive VR was rated significantly higher than desktop for overall presence (P2, $p<0.001$) and visual presence (P3, $p = 0.001$). No differences were observed in terms of enjoyment and navigation, measured with 2 scales each (in blue and grey respectively in Table 1), despite a strong preference in terms of enjoyment for immersive VR in the comparative evaluation.

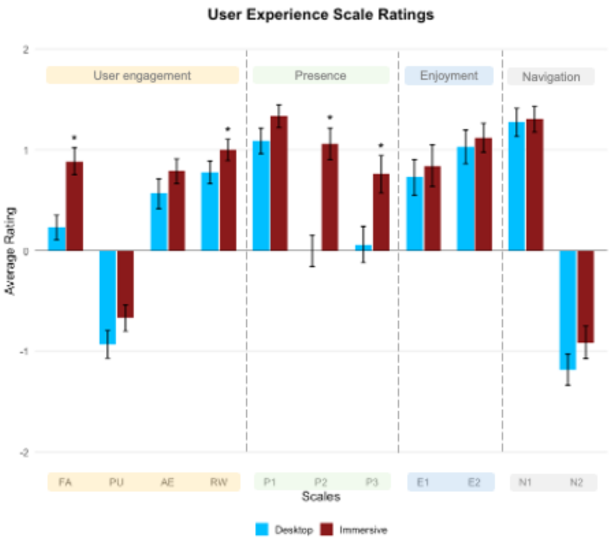

Figure 5. Mean ratings and standard error for user experience scales (User engagement, Presence, Enjoyment, Navigation, see Table 1) across all 36 participants. Significant differences ($p < 0.05$) between immersive and desktop VR indicated with *.

4. Discussion

4.1 Immersion and listening experience

Participants strongly preferred immersive VR for enjoyment, visuals, and sound (in the comparative evaluation) and reported greater focused attention, reward, overall presence, and visual presence with immersive than with desktop version (user experience ratings). The preference for sound is particularly notable because both versions used the same sound sources, binaural rendering pipeline, and headphones. Because the audio rendering remained unchanged, this preference reflects a difference in the experience of listening, rather than in acoustic fidelity. The presence results help clarify this distinction. Although immersive VR produced greater overall and visually supported presence, there was no corresponding difference in sound-based presence. Participants preferred listening within the immersive experience without necessarily perceiving sound itself as making them feel more present. Greater focused attention and reward may also have contributed to this more compelling listening experience.
Importantly, the study compared complete desktop and immersive interface configurations rather than manipulating their individual components (e.g. display, navigation, control), independently. It therefore cannot determine which components contributed most to their evaluation. What the findings do show is that changing the overall interaction modality affected participants' reported listening experience while the underlying audio rendering remained constant. Interactive auralizations should therefore be considered not only in terms of acoustic quality, but also more holistically as complete visual, spatial, and interactional experiences.

4.2 Modality flexibility and accessibility

Preferences concerning the operation of City Ditty were less decisive than those concerning enjoyment, visuals, and sound. Responses involving navigation, controls, object manipulation, menus, finding objects, and reading were more evenly distributed and did not show the same clear preference for immersive VR. Rather than identifying a universally preferred interface, this pattern raises three related considerations: particular activities may benefit from different interaction characteristics, individuals may have different modality preferences, and the two versions may be useful as complementary parts of a broader workflow. Desktop interaction provides familiar keyboard-and-mouse controls and a stable overview of the environment, while immersive VR offers embodied orientation and direct three-dimensional interaction. These characteristics suggest that users may prefer different modalities for different activities, but more research is needed to establish if particular tasks are better supported by either version. Preferences may also differ between individuals, even when completing the same activity, according to their familiarity, comfort, interaction style, or interest in immersive experiences.
Maintaining both versions is therefore important for accessibility. Although immersive VR produced a more engaging and rewarding experience, not every user or organization will be able or willing to rely on a head-mounted display and game controllers. Personal comfort and familiarity, as well as access to hardware, physical space, and technical resources may affect which version can be used. Desktop access should not be treated simply as an inferior substitute, but as another means of reaching the same core functions and soundscape designs.
City Ditty's shared functionality and interchangeable designs also allow the modalities to be used together. Users could create or organize a design on desktop, experience it in immersive VR, and return to desktop for further revision. Although these transitions were not evaluated in the present study, combining desktop and immersive interaction was repeatedly suggested by participants in a parallel stakeholder study conducted using the immersive version [9]. Because designs can be transferred between versions, City Ditty can support such mixed-modality workflows without requiring users to recreate their work. Retaining both forms of interaction is therefore an initial step toward making soundscape simulation accessible to a wider range of users, activities, and working conditions.

4.3 Limitations

This paper reports selected preliminary questionnaire findings from a broader dataset. The comparative questionnaire focuses on self-reports. Task performance (action logs, completion times) will complement this preliminary analysis.
The study compared complete desktop and immersive configurations rather than isolating immersion or individual interface features. Participants also completed distinct but functionally comparable task sets and soundscape configurations. Novelty may have influenced responses to immersive VR, while the 30-minute sessions cannot establish how preferences, engagement, comfort, or fatigue might change with repeated or sustained use.
Finally, participants were not recruited specifically from the professional groups expected to use City Ditty in urban planning and design. The findings concern general interaction and user experience rather than professional

relevance or workflow integration, which are examined more directly in the related stakeholder study with PBEs [9]. Future directions include looking into how participants' experience varies as a function of 1) their familiarity and past experience with Virtual Reality, 2) the specific task at hand, 3) the learning curve for each version, and 4) other factors contributing to their overall experience emerging from in the exit interviews.

5. Conclusions

This study compared desktop and immersive VR versions of City Ditty with equivalent functionality and identical binaural audio rendering. Participants strongly preferred HMD VR for enjoyment, visuals, and sound, while also reporting greater focused attention, reward, and overall presence. The preference for sound is especially notable because the underlying audio presentation remained constant. This indicates that listening through an interactive auralization is shaped not only by its acoustic reproduction, but also by the complete visual, spatial, and interactional context in which it is experienced.

Preferences concerning the operation of the simulator were less decisive. Desktop and immersive VR provide different interaction characteristics that may support different users, activities, and working conditions. Maintaining equivalent functionality and interchangeable designs enables both versions to operate as parts of a flexible workflow, allowing users to move seamlessly between conventional desktop interaction and more immersive exploration.

This flexibility also supports City Ditty's broader aim of widening participation in how urban soundscapes are imagined, discussed, and shaped. In this respect, it extends a foundational purpose of auralization: translating technical sound information into experiential forms that people can hear and more readily understand. By making these representations directly manipulable, City Ditty can also help people explore alternatives and establish a shared experiential reference for discussion. To that end, accessible interaction is intended to complement, rather than replace, specialist acoustic expertise by enabling PBEs and other stakeholders to consider sound earlier, explore possible interventions, and communicate their needs and perspectives more effectively. Extending this work to additional places, users, and professional contexts will require continued development, interdisciplinary partnerships, and sustained resources. Such investment could help advance more proactive soundscape practices in which a wider range of people can engage with specialists in shaping how cities sound.

6. Resources

City Ditty Project page including related works and media: https://www.sounds-in-the-city.org/en/CityDitty

7. Acknowledgments

Thank you to all the participants, as well as CIRMMT (Centre for Interdisciplinary Research on Music Media and Technology), NSERC (RGPIN 2019-06121) and the AIRS network for research support.

References

[1] J. Llorca-Bofí, C. Dreier, J. Heck, and M. Vorländer, "Urban sound auralization and visualization framework—Case study at IHTApark," *Sustainability*, vol. 14, no. 4, Art. no. 2026, 2022, doi: 10.3390/su14042026.

[2] R. Yanaky, D. Tyler, and C. Guastavino, "City Ditty: An immersive soundscape sketchpad for professionals of the built environment," *Applied Sciences*, vol. 13, no. 3, Art. no. 1611, 2023, doi: 10.3390/app13031611.

[3] R. Yanaky, G. Grazioli, Y.-Y. Zhang, and C. Guastavino, "Exploring the use of soundscape sketchpads with professionals," in *INTER-NOISE and NOISE-CON Congr. Conf. Proc.*, Chiba, Japan, Aug. 20–23, 2023, 10 pp., doi: 10.3397/IN_2023_0322.

[4] R. Yanaky and C. Guastavino, "Virtual reality for urban soundscape design: Exploring knowledge sharing, creation, and workplace integration," in *Proc. 2025 31st ACM Symp. Virtual Reality Software and Technology (VRST '25)*, New York, NY, USA: Association for Computing Machinery, 2025, Art. no. 147, pp. 1–2, doi: 10.1145/3756884.3768399.

[5] R. Yanaky and C. Guastavino, "Realistic enough? Design considerations for soundscape simulators," *Canadian Acoustics*, vol. 52, no. 1, May 2024. [Online]. Available: https://jcaa.caa-aca.ca/index.php/jcaa/article/view/4183

[6] C. Miao, R. Yanaky, and C. Guastavino, "Effect of music on the quality of urban public experience: A virtual reality experiment," presented at the 56th Annu. Conf. Environmental Design Research Assoc (EDRA), Halifax, NS, Canada, May 27–30, 2025.

[7] R. Yanaky, L. Fuentes, D. Verdugo, J. Monsalve, B. Vega, C. Villavicencio, P. Kogan, and C. Guastavino, "Application of the City Ditty soundscape tool in an interdisciplinary urban park design competition," in *INTER-NOISE and NOISE-CON Congr. Conf. Proc.*, vol. 272, no. 1, São Paulo, Brazil, Aug. 24–27, 2025, pp. 3639–3650, doi: 10.3397/IN_2025_1095787.

[8] P. Kogan *et al*., "From soundwalks to auralisation: A participatory workflow for biophilic soundscape design in urban green spaces using City Ditty," 12th Convention of the European Acoustics Association (Forum Acusticum), Graz, Austria, Sep. 8–12, 2026.

[9] R. Yanaky, S. Toffa, and C. Guastavino, "Interacting with soundscapes in virtual reality: Desktop vs. immersive VR," invited paper at the 12th Convention of the European Acoustics Association (Forum Acusticum), Graz, Austria, Sep. 8–12, 2026.

[10] H. L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form," *International Journal of Human-Computer Studies,* vol. 112, pp. 28-39, 2018/04/01/ 2018.