Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2897–2906
https://doi.org/10.54941/ahfe1008513

The Immersive Power of Binaural Sound: Reproducing the Experience of a Music Festival

Coralie Bureau[1], Jeanne Nicole[1], Cindy Chamberland[1], Alysun Paradis[1], Calypso Cases[2], and Sébastien Tremblay[1]

[1]Co-DOT Laboratory, School of Psychology, Université Laval, Québec, QC G1V 0A6, Canada
[2]École Nationale de l'Aviation Civile (ENAC), 31055 Toulouse, France

ABSTRACT

Immersive technologies overwhelmingly prioritize visual fidelity, leaving the auditory channel comparatively under-exploited, even though spatial sound is a powerful driver of presence. This study examined whether binaural (3D) sound can reproduce the experience of a live music festival in a digital setting, and how auditory and visual information combine to support immersion and emotional response. Performances at the 2025 Fono Festival in Quebec City were recorded on site using a Neumann KU 100 dummy head paired with a head-mounted camera. With this setup, we could produce binaural and standard stereo versions of the same content directly at capture. Because spatial encoding occurs acoustically in the dummy head rather than being synthesized in post-production, the binaural signal preserves the complete auditory scene, stage, crowd, and surrounding ambience, with its full set of natural localization cues. Forty-eight adults, half of whom had attended the festival in person, experienced songs in binaural or stereo, with or without accompanying video, and reported on their immersion experience and emotional response, the latter assessed along the valence and arousal dimensions of Russell's (1980) circumplex model of affect. Results revealed a significant interaction between audio format and visual condition. Without video, binaural sound elicited significantly greater immersion and presence than stereo. Adding visual support raised stereo to binaural levels but produced no further gain for binaural. Binaural audio alone matched the immersive quality of audiovisual stereo. Visual information heightened arousal but did not affect valence. Attendees judged the digital version a plausible reproduction of their lived experience. Spatial auditory cues can therefore partially compensate for the absence of visual information, though the two sources appear to overlap rather than combine additively, consistent with the principle of inverse effectiveness in multisensory integration. Because binaural delivery requires only headphones, it offers an accessible route to immersive experience.

Keywords: Binaural audio, Spatial sound, Presence, Immersion, Multisensory integration, Live music, Emotional response

INTRODUCTION

Immersive media have advanced primarily along the visual axis. Headsets, resolution, and field of view have improved rapidly, while the auditory

Received September 22, 2026; Revised October 17, 2026; Accepted October 28, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

channel is often treated as an accompaniment rather than a carrier of spatial information. This is a missed opportunity. Potter et al. (2022) estimated that improving the quality of spatial audio rendering yields an immersion gain comparable to a fivefold increase in the visual resolution of a 3D virtual environment. Sound captures attention and generates sensory and emotional immersion in ways that are difficult to achieve visually (Salselas and Penha, 2019; van Laer et al., 2019), and it does so with hardware most people already own.

Two psychological states organize this literature. Presence is the subjective sense of "being there", of being physically located within a mediated space (Chen et al., 2024; Cummings and Bailenson, 2016). It supports immersion, the state in which the user feels enveloped by and involved in the environment (Witmer and Singer, 1998). Binaural sound, also called 3D audio, is a spatialization technique that reproduces the acoustic cues of human hearing: interaural time and level differences, and the spectral filtering imposed by the head and pinnae. Listeners can therefore localize sources realistically, which strengthens the feeling of being present in the recorded environment (Magal-Royo and García Laborda, 2020; Shin et al., 2019). Prior work shows that 3D audio content improves immersion, satisfaction, and social presence, and supports emotional engagement (Nagele et al., 2021; Rees-Jones and Murphy, 2018; Shin et al., 2019).

Despite these advances, few empirical studies have systematically compared the perceptual impact of binaural rendering against more conventional stereo while simultaneously manipulating the availability of visual information, and fewer still have done so for a live outdoor event. Performing arts are a demanding test case: recreating the richness and authenticity of a live performance remains a major challenge, and the value of an authentic digital reproduction is highest for those who cannot attend in person. A methodological obstacle explains part of this gap. Comparisons of spatial and stereo rendering usually rely on either separate recording sessions or on post-production mixing, which confounds audio format with performance variability or with signal processing. The present study avoids both by capturing the two formats simultaneously, from the same performance, at the same position, with two independent devices. The study addressed three objectives: (1) to compare binaural and stereo rendering in their capacity to elicit presence and immersion during a digital festival experience; (2) to examine how audio format and visual support affect emotional response, in terms of valence and arousal; and (3) to assess the perceived authenticity of the reproduction (its perceived experiential fidelity to the event as it was experienced), particularly among participants who had attended the festival in person.

METHOD

Binaural Capture at the Fono Festival

Musical performances at the 2025 Fono Festival in Québec City were recorded with a Neumann KU 100 dummy head, nicknamed "Fritz", whose two microphones reproduce human listening. A GoPro camera mounted

on the head provided audiovisual synchronization and, at the same time, a standard stereo capture. This configuration yielded two independent recordings of the same performance: a binaural version from the dummy head and a stereo version from the camera. Both versions were therefore available at capture, with no transformation of the source signal. Because spatial encoding occurs acoustically in the dummy head rather than being synthesized afterwards, the binaural signal preserves the full auditory scene (stage, crowd, and surrounding ambience) with its natural localization cues.



Figure 1: The KU 100 dummy head in position during a recorded performance at the Fono Festival, silhouetted against the stage.

The dummy head was installed in a protected area at stage right, approximately 5 m from the main stage (Figure 2). This position offered physical protection from crowd movement while remaining close enough to capture a high-quality signal. The lateral placement introduces a slight asymmetry in the binaural signal, the right ear receiving the main source with greater intensity, which is consistent with binaural capture in real settings and was taken into account when editing the recordings. Deploying a studio-grade dummy head outdoors at high sound pressure levels was itself a technical achievement, as this equipment is designed primarily for indoor and studio use.



Figure 2: Left: the Neumann KU 100 dummy head with the head-mounted GoPro camera in position at the festival. Right: schematic of the capture position, at stage right, approximately 5 m from the main stage.

Stimulus Selection

Binaural and stereo recordings were transferred and synchronized in post-production, with audio and video aligned on the GoPro signal. A preselection based on acoustic quality (background noise, level saturation, capture artifacts) reduced the corpus to eight song recordings. Six raters then evaluated each song recording on four criteria: differential quality between the binaural and stereo versions, general audio quality, video quality, and musical interest. Aggregating their rankings identified the four songs with the broadest consensus. Editing consisted only of synchronizing the two versions and standardizing volume. The source signal was not transformed. Final song files lasted between 4 and 8 minutes.

Participants and Design

Forty-eight adults took part (32 women; $M = 28.2$ years, $SD = 10.8$, range 18–70). Twenty-four had attended the 2025 Fono Festival in person and twenty-four had not. The design crossed two within-subject factors, audio format (binaural vs. stereo) and visual support (video present vs. absent), with festival attendance as a between-subjects factor. Each participant experienced all four combinations of audio format (stereo vs. binaural) and visual support (with vs. without video), one per song, using a counterbalanced assignment. Across participants, each song was presented six times in each combination, ensuring that song recording content was not confounded with audio format or visual support.

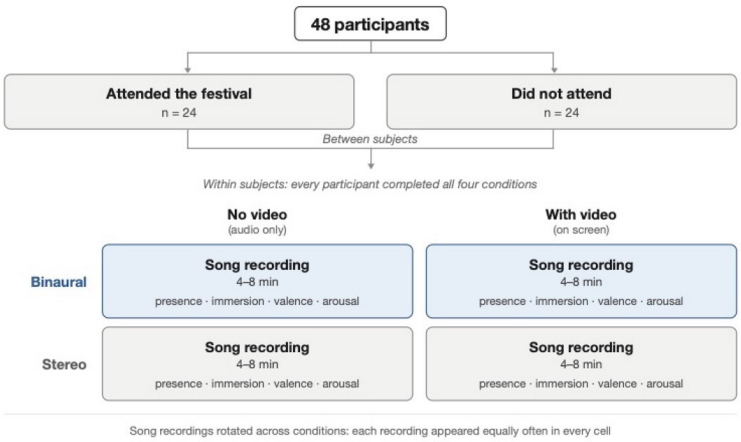


Figure 3: Experimental design. Festival attendance varied between subjects; audio format and visual support varied within subjects, with songs rotated across conditions.

Apparatus and Procedure

Participants were tested individually in a quiet laboratory room (Figure 4). Song recordings were delivered over headphones; when video was present, it was displayed on a computer screen. Questionnaires were completed on a tablet. Participants fitted a Zephyr BioHarness chest strap recording cardiac

and respiratory activity, with a three-minute resting baseline recorded while they fixated a static cross on the screen, and eye movements were calibrated on a Tobii Pro Spectrum screen-integrated eye tracker. They then listened to the four songs and completed the experience and emotion questionnaires after each one. Participants who had attended the festival answered three additional questions on the perceived authenticity of the reproduction (after all recordings). The physiological and oculometric measures will be reported elsewhere (analyses are still in progress); the present paper is focused on the self-report outcomes.

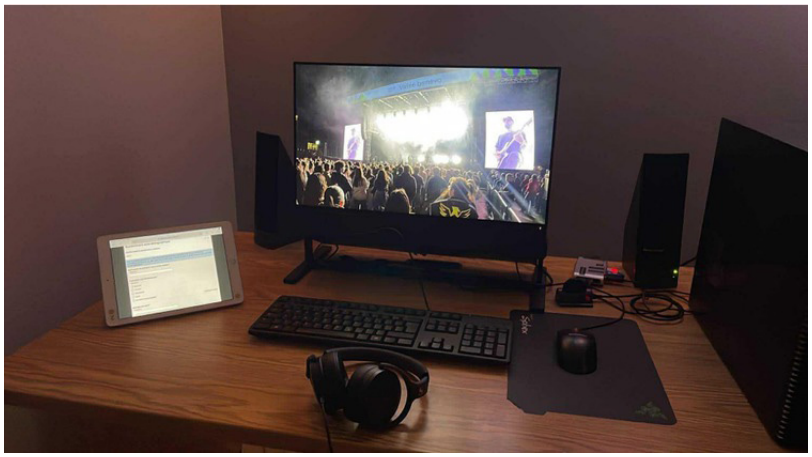


Figure 4: Laboratory setup: viewing screen with headphones, and tablet for questionnaires.

Measures

Immersive audio experience was measured with a French adaptation of the Overall Immersive Audio eXperience questionnaire (OIAX; Kailas and Tiwari, 2021). Twenty-three of the 33 items were retained. They cover four subscales: presence (8 items), immersion (7 items), clarity (4 items) and readability (4 items). Items were rated on a 10-point Likert scale (1 = *strongly agree*, 10 = *strongly disagree*). Six items were negatively worded (e.g., "Some sounds interfered with others") and were scored in the opposite direction to the others, and all responses were recoded so that higher scores indicate more of the measured attribute. Internal consistency was good for presence ($\alpha = .94$), immersion ($\alpha = .87$), readability ($\alpha = .77$), and weaker for clarity ($\alpha = .66$). Subscale scores are means of their items; the composite score is the sum of the four subscale means (range 4-40). Emotional response was measured after each song with the Affective Slider (Betella and Verschure, 2016), which provides continuous valence and arousal ratings. We assessed perceived authenticity in attendees using three purpose-built items on ambience, immersion, and emotional re-experiencing, rated on the same scale and recoded in the same way ($\alpha = .84$). We measured immersive propensity (Witmer and Singer, 1998) and attitude toward technology (Edison and Geissler, 2003) before the session.

Analysis

Effects were estimated with linear mixed models including random intercepts for participant and song, with contrasts for audio format, visual support, and attendance. Planned contrasts were tested with paired t-tests, Holm-corrected within each dependent variable, and accompanied by Bayes factors so that null results could be read as evidence rather than as absence of evidence.

RESULTS

Presence and Immersion

The mixed models revealed an interaction between audio format and visual support on presence, $z = -2.22$, $p = .026$, and on immersion, $z = -2.12$, $p = .034$, with a main effect of audio format on presence, $z = 2.00$, $p = .045$. There was no effect of festival attendance and no three-way interaction (all ps $> .25$). Descriptive statistics are presented in Table 1 and the interaction between audio format and visual support is plotted in Figure 5.

Table 1: Mean OIAX subscale ratings by audio format and visual support.

Condition	Presence	Immersion	Clarity	Readability
Stereo, no video	6.48 (2.07)	5.57 (2.00)	6.12 (1.96)	6.59 (2.26)
Stereo, with video	7.44 (1.83)	6.76 (2.03)	6.30 (1.90)	6.94 (1.76)
Binaural, no video	7.39 (2.07)	6.40 (1.76)	5.96 (1.91)	6.92 (2.19)
Binaural, with video	7.40 (2.01)	6.60 (1.91)	6.27 (2.19)	6.68 (2.10)

Note. N = 48. Values are means, with standard deviations in parentheses. All subscales are rated on a 10-point scale. Mean arousal ratings are shown in Figure 5.

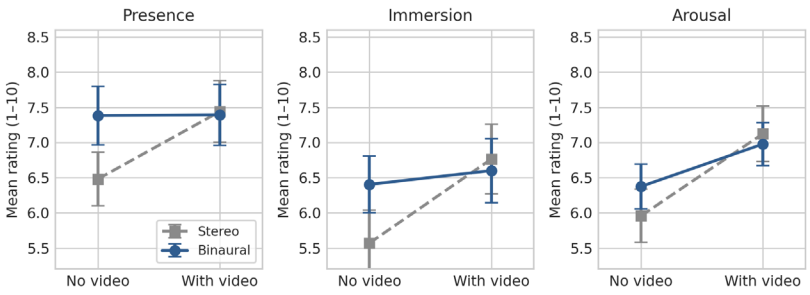


Figure 5: Presence, immersion, and arousal as a function of audio format and visual support. Error bars are 95% within-subject confidence intervals (N = 48).

Further analyses were carried out to decompose the interaction: presence was higher for binaural than for stereo when no video was present, $t(47) = 4.00$, $p = .001$ after correction, $dz = 0.58$, BF10 = 113, and the same pattern held for immersion, $t(47) = 3.04$, $p = .015$, $dz = 0.44$, BF10 = 8.74. Adding video raised stereo to binaural levels, $t(47) = 3.15$, $p = .011$, $dz = 0.49$ for presence, but produced no further gain for binaural ($dz = 0.01$, BF10 = 0.16). Binaural audio alone was not different from audiovisual stereo

($dz = 0.03$, BF10 = 0.16). Audio format had no effect on perceived clarity, $z = -0.44$, $p = .659$, or readability, $z = 0.20$, $p = .844$, and all contrasts on these two subscales yielded BF10 ≤ 0.46 (Figure 6).

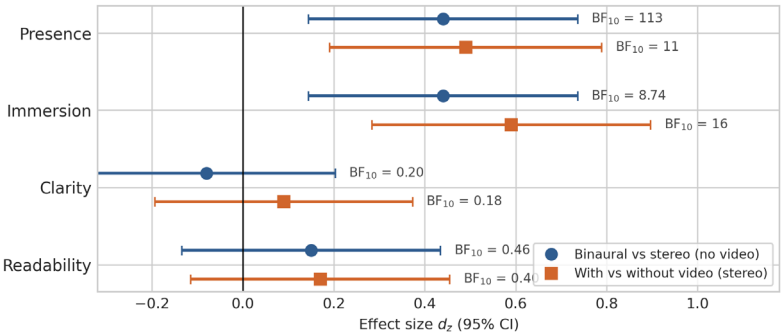


Figure 6: Effect sizes with 95% confidence intervals for the two key contrasts across the four OIAX subscales. Bayes factors below 1 indicate evidence for the absence of an effect.

Emotional Response

Arousal was higher when visual support was present than when it was absent, $z = 4.91$, $p < .001$, for both stereo ($dz = 0.54$) and binaural ($dz = 0.37$) recordings. Valence was unaffected by either factor (all BF10 ≤ 0.88). This pattern is consistent with multimodal gain: combining auditory and visual cues generally increases the intensity of emotional response relative to unimodal presentation (Lee et al., 2026). Independently of prior attendance, ratings fell in the upper-right quadrant of the valence–arousal dimensional model (Figure 7). The experience was both pleasant and moderately intense, and the condition centroids of the two attendance groups occupied the same region of the affective space, between joy and fascination. Within each group, the two conditions with video showed higher arousal than their audio-only counterparts, while the four centroids differed little in valence.

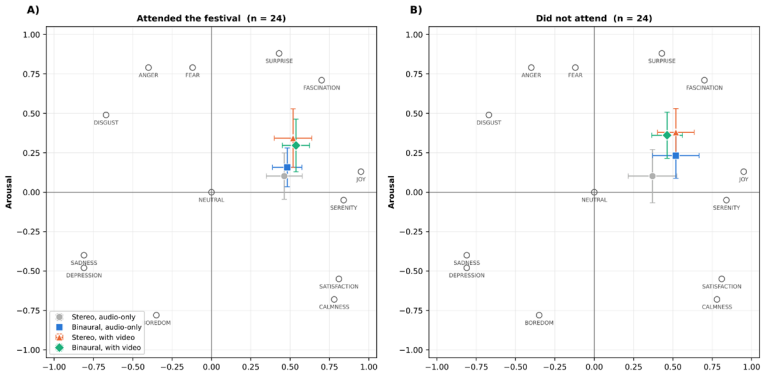


Figure 7: Condition centroids in the valence–arousal plane for participants who attended the festival (A) and those who did not (B). Open circles mark theoretical emotion coordinates (Paltoglou and Thelwall, 2013); bars are 95% confidence intervals.

Perceived Authenticity of the Reproduction

Among attendees, perceived authenticity was rated slightly above the scale midpoint ($M = 6.61, SD = 2.50$), $t(23) = 2.18, p = .040, d = 0.45$. The item-level pattern is more informative. Attendees agreed that the laboratory experience faithfully reproduced the ambience of the festival ($M = 7.04$, $p = .009$) and allowed them to re-experience the emotions they had felt there ($M = 7.25$, $p = .006$), but did not agree that they felt as immersed in the laboratory as at the festival ($M = 5.54$, $p = .948$, BF10 = 0.22). Perceived authenticity correlated strongly with immersion, $r = .75$, $p < .001$, and presence, $r = .63$, $p = .001$, but not with clarity, $r = .21$, or readability, $r = .17$, a pattern that supports the discriminant validity of both constructs. Neither immersive propensity nor attitude toward technology predicted the composite score ($R^2 = .03$), and neither moderated the benefit of binaural rendering or the gain from adding video. Participants who had attended the festival did not differ from those who had not on any outcome, though this could be due to a lack of statistical power.

DISCUSSION

Spatial auditory cues can partially compensate for the absence of visual information. When listeners had nothing to look at, they had to build a mental model of the environment from sound alone, and binaural rendering supplied the spatial information needed to do so. The auditory and visual sources of spatial information appear to overlap rather than combine additively, which is what the principle of inverse effectiveness in multisensory integration predicts: when one modality already affords a reliable estimate, adding a second yields limited benefit (Opoku-Baah et al., 2021; Spence et al., 2011).

Emotion responded to the visual manipulation but not to the audio manipulation. Visual information raised arousal for both audio formats, consistent with multimodal gain. Presence and arousal therefore respond to different aspects of the experience: spatial sound places listeners in the scene without making it more intense, and visual information intensifies the experience without adding to the sense of being there. Figure 7 situates all four conditions in a narrow region of the affective space between joy and fascination, and the two attendance groups overlap. The same region was occupied in a museum study that compared original artworks viewed in situ with digital reproductions viewed in the laboratory, where fascination was the most frequently reported emotion and neither valence nor arousal separated the two contexts (Chamberland et al., in press). The affective signature of a cultural encounter therefore appears to survive digital reproduction in the gallery as it does on the festival field.

Despite some technical challenges and sample-related limitations, the applied implication is promising. Digital reproductions of live events need not rely on costly, fully immersive visual technology to be perceived as authentic and engaging. Carefully spatialized sound delivered over ordinary headphones offers a meaningful way to feel present at an event, which matters most for people who cannot attend in person for reasons of distance, cost, health, or mobility. The same rationale extends beyond culture. Immersive technologies

are already used to support wellbeing in populations with restricted mobility, such as residents of nursing homes and assisted living facilities (Moore et al., 2023), and an augmented reality device has been shown to reduce preoperative anxiety in children without medication (Chamberland et al., 2024). Binaural audio could serve comparable purposes at a fraction of the cost and setup.

CONCLUSION

Binaural sound captured on site at a live music festival produced greater presence and immersion than stereo when no visual support was available. The advantage was specific to presence and immersion and absent from perceived sound quality. The emotional consequence of audio format seems carried entirely by presence. Attendees judged the reproduction authentic in ambience and emotion but not equivalent to being there. Future work should refine these reproduction methods (head tracking, individualized rendering, more enveloping delivery) and examine how technical and individual factors shape the perceived authenticity of virtually recreated experiences.

ACKNOWLEDGMENT

The authors thank Festival Fono for access to the site and performances. We also thank Steven Thomas for signal processing and Denis Ouellet for technical setup. This work was supported by the Vice-Rectorate for Research, Creation and Innovation (VRRCI) of Université Laval.

REFERENCES

Betella, A., Verschure, P. F. M. J. (2016). "The Affective Slider: A digital self-assessment scale for the measurement of human emotions", PLoS ONE, 11 No. 2, pp. 1–11.

Chamberland, C., Bransi, M., Boivin, A., Jacques, S., Gagnon, J., Tremblay, S. (2024). "The effect of augmented reality on preoperative anxiety in children and adolescents: A randomised controlled trial", Pediatric Anesthesia, Volume 34 No. 2, pp. 153–159.

Chamberland, C., Perichon, V., & Tremblay, S. (in press). Emotional responses to art in situ and on screen: A test of the circumplex model and the genuineness effect. Frontiers in Psychology.

Chen, C., Hu, X., Fisher, J. (2024). "What is "being there"? An ontology of the immersive experience", Annals of the International Communication Association, Volume 48 No. 4, pp. 391–414.

Cummings, J. J., Bailenson, J. N. (2016). "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence", Media Psychology, Volume 19 No. 2, pp. 272–309.

Edison, S. W., Geissler, G. L. (2003). "Measuring attitudes towards general technology: Antecedents, hypotheses and scale development", Journal of Targeting, Measurement and Analysis for Marketing, Volume 12 No. 2, pp. 137–156.

Kailas, G., Tiwari, N. (2021). "Design for immersive experience: Role of spatial audio in extended reality applications", in: Design for Tomorrow, Volume 2, Chakrabarti, A., Poovaiah, R., Bokil, P., Kant, V. (Eds.). Singapore: Springer, pp. 815–826.

Lee, Y., Lee, Y., Lee, D. (2026). "Emotion-specific modality effects in auditory and visual perception of emotion", Psychological Research, Volume 90 No. 4, p. 123.
Magal-Royo, T., García Laborda, J. (2020). "Binaural sound to enhance listening comprehension in language tests", International Journal of Interactive Mobile Technologies, Volume 14 No. 1, pp. 4–14.
Nagele, A. N., Bauer, V., Healey, P. G. T., Reiss, J. D., Cooke, H., Cowlishaw, T., Baume, C., Pike, C. (2021). "Interactive audio augmented reality in participatory performance", Frontiers in Virtual Reality, Volume 1, Article 610320.
Moore, R. C., Hancock, J. T., Bailenson, J. N. (2023). "From 65 to 103, older adults experience virtual reality differently depending on their age: Evidence from a large-scale field study in nursing homes and assisted living facilities", Cyberpsychology, Behavior, and Social Networking, Volume 26 No. 12, pp. 886–895.
Opoku-Baah, C., Schoenhaut, A. M., Vassall, S. G., Tovar, D. A., Ramachandran, R., Wallace, M. T. (2021). "Visual influences on auditory behavioral, neural, and perceptual processes: A review", Journal of the Association for Research in Otolaryngology, Volume 22 No. 4, pp. 365–386.
Paltoglou, G., Thelwall, M. (2013). "Seeing stars of valence and arousal in blog posts". IEEE Transactions on Affective Computing, Volume 4 No. 1, pp. 116–123.
Potter, T., Cvetković, Z., De Sena, E. (2022). "On the relative importance of visual and spatial audio rendering on VR immersion", Frontiers in Signal Processing, Volume 2, Article 904866.
Rees-Jones, J., Murphy, D. T. (2018). "The impact of multichannel game audio on the quality and enjoyment of player experience", in: Emotion in Video Game Soundtracking, Williams, D., Lee, N. (Eds.). Cham: Springer, pp. 143–163.
Salselas, I., Penha, R. (2019). "The role of sound in inducing storytelling in immersive environments", in: Proceedings of the 14th International Audio Mostly Conference. New York: ACM, pp. 191–198.
Shin, M., Song, S. W., Kim, S. J., Biocca, F. (2019). "The effects of 3D sound in a 360-degree live concert video on social presence, parasocial interaction, enjoyment, and intent of financial supportive action", International Journal of Human-Computer Studies, Volume 126, pp. 81–93.
Spence, C., Parise, C., Chen, Y.-C. (2011). "The Colavita visual dominance effect", in: The Neural Bases of Multisensory Processes, Murray, M. M., Wallace, M. T. (Eds.). Boca Raton: CRC Press, pp. 529–556.
van Laer, T., Feiereisen, S., Visconti, L. M. (2019). "Storytelling in the digital era: A meta-analysis of relevant moderators of the narrative transportation effect", Journal of Business Research, Volume 96, pp. 135–146.
Witmer, B. G., Singer, M. J. (1998). "Measuring presence in virtual environments: A presence questionnaire", Presence: Teleoperators and Virtual Environments, Volume 7 No. 3, pp. 225–240.