Development of an audio-visual interaction questionnaire

Timothy Van Renterghem[1]

[1]*Institute for Biomedical Engineering (iBME/Noise & Health), Department of Information Technology, Ghent University, Belgium*

Abstract

Previous research has shown that audio-visual interactions in environmental perception—and in environmental noise perception in particular—can be effectively leveraged in urban sound planning. However, the extent to which audio-visual cues are integrated varies substantially across individuals. Accounting for such individual differences may be essential for more accurately explaining phenomena such as the benefits of green window views in epidemiological studies. In this study, results from two computer-based tests—a congruent object recognition task (CORT) and the sound-induced flash illusion (SIFI)—were used as objective measures to guide the development of a questionnaire assessing individual audio-visual integration capacity. From this pool, a subset of questions was identified that demonstrated predictive power for performance on the CORT and SIFI tasks, as well as good test–retest reliability. Although the present findings should be regarded as preliminary, they provide an initial indication that individual differences in objective audiovisual processing are reflected in self-reported experiences of everyday multisensory functioning. The observed associations suggest that the CORT and the SIFI may capture partly distinct underlying constructs.

Keywords: *environmental noise, noise perception, audio-visual interactions, personal factors.*

1. Introduction

Audio-visual interactions play a key role in environmental perception and have emerged as a promising pathway for improving the experience of environmental noise. One of the most consistent findings is that visual vegetation in a window view reduces noise annoyance, as demonstrated in multiple cross-sectional survey studies around the world [1], [2], [3], [4], [5].

Corresponding author: *timothy.vanrenterghem@ugent.be*

This beneficial effect is commonly interpreted in the context of theories describing the psychological impact of natural environments, such as Attention Restoration Theory (ART) and Stress Reduction Theory (SRT) [6], [7]. A complementary explanation is provided by multisensory perception. The concept of inattentional deafness [8] demonstrates that demanding visual processing can suppress awareness of task-irrelevant auditory information. Extrapolated to everyday environments, this suggests that visually engaging scenes may reduce attention allocated to environmental noise, thereby influencing noise perception and annoyance.

However, audio-visual interactions are known to vary substantially between individuals. Although the Colavita effect [9] indicates that vision generally dominates during simultaneous audio-visual stimulation, this dominance is far from universal. A well-established example of cross-modal integration is the Sound-Induced Flash Illusion (SIFI), in which a single flash accompanied by two auditory beeps is frequently perceived as two flashes [10]. The SIFI demonstrates that auditory information can alter visual perception and has become a widely used paradigm for quantifying individual differences in multisensory integration.

To complement the SIFI, this study also employs the Congruent Object Recognition Test (CORT) [11]. Although less frequently used, the CORT provides a broader assessment of multisensory processing by including both unimodal and bimodal stimulus presentations. Accuracy and reaction time measures jointly characterize modality dominance and multisensory integration, offering complementary information to the SIFI. Together, these two computer-based paradigms provide objective measures of individual audio-visual processing styles.

The final goal of the current study is to use such information in future surveys to help explaining the variability observed in the beneficial effect of green window view on noise annoyance, or more generally, the influence of positive visual environments on noise perception. Laboratory-based audiovisual tests such as SIFI and CORT are impractical for large-scale population studies, so the aim is to develop and validate a self-report audio-visual questionnaire.

2. Methodology

2.1 CORT

The Congruent Object Recognition Test (CORT) used in this study is based on the audiovisual object recognition paradigm [11] as implemented in Ref. [12]. Participants viewed stimuli on a computer monitor while wearing headphones and identified one of two objects (A or B) as quickly and accurately as possible.

Objects were presented in auditory-only, visual-only, or audiovisual form. Visual stimuli consisted of a circle deforming into either a horizontally (object A) or vertically (object B) oriented ellipse, while auditory stimuli were pure tones of 540 Hz (object A) or 560 Hz (object B). Response accuracy and reaction times (for correct responses only) were recorded, yielding six outcome measures characterizing individual audiovisual processing.

2.2 SIFI

The Sound-Induced Flash Illusion (SIFI) task consists of white circular flashes presented centrally on a black background, with a fixation cross continuously visible. Each flash was displayed for one frame on a 60 Hz monitor (approximately 16.7 ms). Auditory stimuli consisted of 3500 Hz beeps with a duration of 7 ms.

Illusion trials followed the classical one-flash/two-beep (1F2B) stimuli, in which the visual flash was presented simultaneously with the first beep, while the second beep followed after a variable stimulus-onset asynchrony (SOA). SOAs ranged from ±46 to ±156 ms. Positive and negative inter-stimuli intervals (ISI) corresponded to vision-lead and auditory-lead conditions, respectively. Baseline trials (1F1B and 2F2B) and unimodal control trials (2F0B, 0F2B, 1F0B, and 0F1B) were included to assess task performance and response validity.

2.3 Questionnaire

An initial pool of 40 items was intentionally constructed to broadly sample the constructs of interest in an exploratory manner. Questions addressed common situations involving audiovisual processing, including e.g. the use of subtitles despite clearly audible speech while watching television, preference for telephone versus video calls, susceptibility to auditory or visual distraction, reliance on spoken navigation instructions, and whether other people are more readily recognized by their voice or face. The questionnaire included conceptually related as well as reverse-worded items to evaluate internal consistency and reduce acquiescence bias.

To assess test-retest reliability, participants were invited to complete the same questionnaire a second time approximately 14 days after the initial assessment. This allowed evaluation of the temporal stability of the responses and identification of items showing poor repeatability or ambiguity. The initial large item pool will subsequently be refined to a smaller set of robust items based on their psychometric properties and their relationship with the objective audiovisual measures. All questions used a 1 to 5 Likert scale, indicating the degree to which the respondent agrees with the statement (from 1="totally disagree" to 5="totally agree").

3. Results

Logistic curves were fitted on the repeated raw illusion probabilities of each respondent (N=87) in the SIFI test and presented in Fig. 1. At very small ISI, the chance that a person experiences the illusion is high, while at a larger ISI, only few people will see the second flash. In a next step, a binary SIFI susceptibility indicator was calculated, demanding that for an ISI of 50 ms, the probability of experiencing the illusion should be at least 40%, while at an ISI of 100 ms, this should be 25 %. These same criteria were used for both positive and negative ISI.

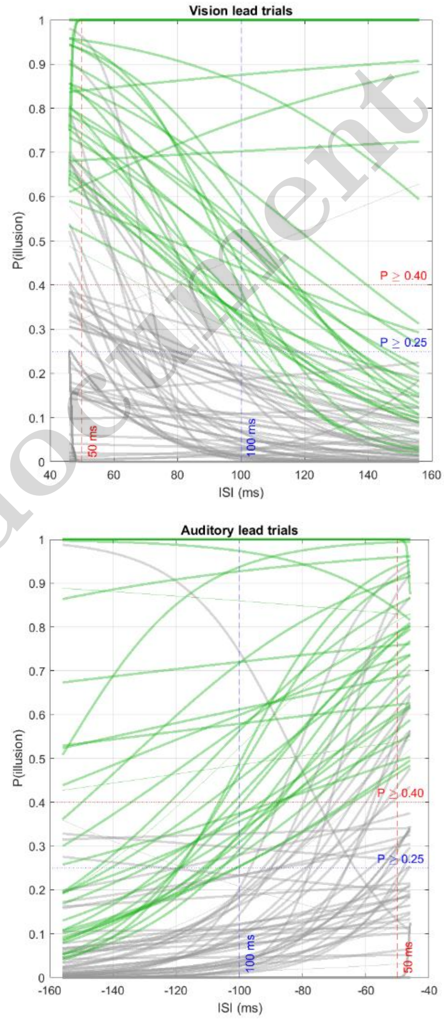

Figure 1. Logistic regression fits for the individual respondents in case of vision-lead and auditory-lead trials, with indication of the binary SIFI susceptibility criterion (green curves indicate that the criterion is fulfilled). Thin lines denote participants not passing the control trials with at least 90% accuracy.

The difference in reaction time to audio only and visual only trials (on condition of correct answering) in the CORT allowed discrimination between major clusters as previously identified [13]. Visually dominated persons show a high value on this relative indicator, while more auditory apt persons a low or negative one. The distribution of RT_A-RT_V is shown in Fig. 2.

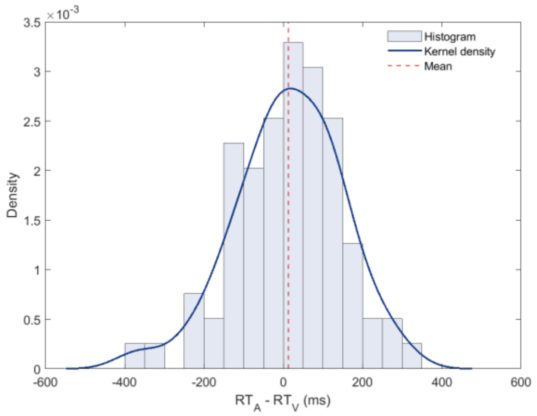

Figure 2. Histogram (all respondents) of the relative difference in reaction times between audio only and visual only trials in the CORT test.

In this preliminary analysis, correlations between each item of the questionnaire, and the above-mentioned key results of the CORT and SIFI are performed. For CORT, simple ordinary least-squares (OLS) linear regression is performed (see Table 1), and for SIFI, logistic regression given the binary outcome (Table 2). This analysis allows indicating key questions with potentially relevant predictive power. Only the items with a p-value lower than 0.1 are shown. The English translations of the questions (originally in Dutch) identified in Tables 1 and 2 are summarized in Table 3, together with the test-retest statistics.

There is some variation in the correlations between first and second answering (ranging from 0.37 to 0.78), although p-values are all small and strongly significant. Note that the link between question and objective indicators is performed on the first answering (N=87) of the survey questions only.

Table 1. Correlation statistics for the questionnaire items with the lowest p-values (<0.1).

Item	Beta	Pvalue	Adj R2
AV5_1	-0.038	0.002	0.107
AV6_1	-0.033	0.013	0.065
AV26_1	-0.038	0.034	0.045
AV13_1	-0.033	0.043	0.040
AV27_1	-0.030	0.046	0.038
AV2_1	-0.044	0.065	0.031

Table 2. Logistic regression statistics (positive ISI) for the questionnaire items with the lowest p-values (<0.1).

Item	Beta	Pvalue	OddsRatio
AV28_1	0.779	0.030	2.18
AV21_1	0.620	0.033	1.86
AV13_1	0.491	0.098	1.63

Table 3. Identified relevant questions with indication of test-retest statistics.

Question	English translation of Dutch question	test-retest r	test-retest p
AV2	I can recognize emotions in someone's voice, even without seeing the person.	0.481	1.16E-04
AV5	When driving in heavy traffic, I turn off the radio to concentrate better.	0.675	4.67E-09
AV6	When I want to focus on music, I often close my eyes.	0.704	4.90E-10
AV13	When trying to recall a previous conversation, I often visualize the situation or the person's appearance.	0.557	4.51E-06
AV21	I feel that one sense is dominant in how I process information.	0.488	8.97E-05
AV26	When using my smartphone or tablet, I enable keypress sounds.	0.781	2.96E-13
AV27	I usually lower the volume when television subtitles are on.	0.616	2.03E-07
AV28	Sometimes I pay more attention to visual details than to what people are saying.	0.366	4.38E-03

4. Discussion

The present study investigated whether individual differences in audiovisual processing measured using two objective laboratory paradigms are reflected in self-reported audiovisual experiences in everyday life. Although the observed associations between questionnaire items and objective performance were modest, a consistent pattern emerged that provides insight into the ecological relevance of both paradigms. Notably, the questionnaire items associated with CORT performance differed substantially from those associated with SIFI susceptibility, supporting the notion that these paradigms capture complementary rather than identical aspects of audiovisual functioning.

The strongest associations with CORT performance were found for questionnaire items describing deliberate modulation of sensory input in everyday situations, such as turning off the radio while driving, closing one's eyes to better appreciate music, lowering the television volume when subtitles are available, enabling auditory feedback during smartphone use, and mentally reconstructing the visual context of previous conversations. Although these behaviours appear diverse, they all involve active regulation of the relative contribution of audition and vision according to task demands. This suggests that the CORT reflects not only audiovisual processing speed but also an individual's ability to flexibly allocate attention across sensory modalities and suppress temporarily irrelevant information. Consistent with this interpretation, several associated behaviours can be viewed as strategies to minimise cross-modal distraction, indicating that the CORT may be influenced by executive mechanisms governing sensory selection and attentional reweighting.

In contrast, the questionnaire items associated with SIFI susceptibility describe relatively stable perceptual tendencies rather than deliberate behavioural strategies. Participants who reported paying greater attention to visual details than to spoken information, or who indicated that one sensory modality generally dominates their perception, were more likely to be susceptible to the audiovisual illusion. This distinction is theoretically meaningful because the SIFI is generally considered to reflect an automatic integration process that occurs largely outside conscious control. The association with visual attentional preference is particularly intriguing. Rather than indicating stronger visual dominance, it may reflect broader allocation of attention across modalities or reduced visual reliability, thereby increasing the perceptual weight assigned to auditory input in accordance with Bayesian models of multisensory integration.

One questionnaire item emerged in both analyses: participants who reported mentally visualising the context of previous conversations tended to show both altered CORT performance and, to a lesser extent, increased SIFI susceptibility. Although this finding should be interpreted cautiously, it may point towards a broader cognitive style characterised by richer visual contextual representations, influencing both strategic sensory selection and automatic audiovisual interactions. The fact that this was the only overlapping questionnaire item further suggests that the common variance between the two paradigms is limited.

Taken together, these findings reinforce the interpretation that audiovisual processing is not a unitary construct but comprises multiple complementary components. The CORT appears to relate primarily to adaptive and goal-directed control over sensory attention, whereas the SIFI predominantly reflects intrinsic perceptual weighting and automatic multisensory integration. The nature of the associated questionnaire items closely mirrors the characteristics of the laboratory tasks themselves: CORT correlates with behaviours involving intentional sensory regulation, whereas SIFI correlates with self-perceived perceptual preferences. Although the individual effect sizes were relatively small, as expected given the complexity of multisensory behaviour, the overall pattern of associations is theoretically coherent and provides converging evidence that individual differences in everyday audiovisual experiences are reflected in objective measures of multisensory processing.

5. Conclusions

Although the present findings should be regarded as preliminary, they provide an initial indication that individual differences in objective audiovisual processing are reflected in self-reported experiences of everyday multisensory functioning. Importantly, the observed associations suggest that the CORT and the SIFI may capture partly distinct underlying constructs, with the former relating more closely to strategic sensory allocation and attentional control, and the latter to intrinsic perceptual weighting and automatic audiovisual integration. However, these interpretations should be treated with caution. The analyses were exploratory and based on individual associations between single questionnaire items and the objective outcome measures, without accounting for potential interrelationships between questionnaire items or underlying latent constructs. Given the broad scope of the 40-item questionnaire, it is likely that several overlapping cognitive and perceptual dimensions contribute to the observed response patterns.

6. Acknowledgments

The author is grateful to Jill Lootens for helping developing the questionnaire and performing the first batch of respondents, and Emilia De Graeve for performing the second batch.

References

[1] H. Li, C. Chau, and S. Tang: "Can surrounding greenery reduce noise annoyance at home?" Science of the Total Environment, vol. 408, pp. 4376–4384, 2010.

[2] T. Van Renterghem and D. Botteldooren: "View on outdoor vegetation reduces noise annoyance for dwellers near busy roads," Landscape and Urban Planning, vol. 148, pp. 203–215, 2016.

[3] T. Leung, J. Xu, C. Chau, S. Tang, and P.-C. L. Pun-Cheng: "The effects of neighborhood views containing multiple environmental features on road traffic noise perception at dwellings," Journal of the Acoustical Society of America, vol. 141, pp. 2399–2407, 2017.

[4] B. Schäffer, M. Brink, F. Schlatter, D. Vienneau, and J.-M. Wunderli: "Residential green is associated with reduced annoyance to road traffic and railway noise but increased annoyance to aircraft noise exposure," Environment International, vol. 143, 105885, 2020.

[5] T. Pasanen, T. Yli-Tuomi, P. Tiittanen, and T. Lanki: "More green, less annoying? The moderating effects of greenery near home, noise sensitivity, and nature relatedness on road traffic noise annoyance," Urban Forestry & Urban Greening, vol. 104, 128625, 2025.

[6] R. Kaplan and S. Kaplan: "The experience of nature: A psychological perspective," Cambridge University Press, New York, US, 1989.

[7] R. Ulrich, R. Simons, B. Losito, E. Fiorito, M. Miles, and M. Zelson: "Stress recovery during exposure to natural and urban environments," Journal of Environmental Psychology, vol. 11, pp. 201-230, 1991.

[8] J. Macdonald and N. Lavie: "Visual perceptual load induces inattentional deafness," Attention, Perception, & Psychophysics, vol. 73, pp. 1780–1789, 2011.

[9] F. Colavita: "Human sensory dominance," Perception and Psychophysics, vol. 16, pp. 409–412, 1974.

[10] L. Shams, Y. Kamitani, and S. Shimojo: "What you see is what you hear," Nature, vol. 408, 788, 2000.

[11] M. Giard and F. Peronnet: "Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study," Journal of Cognitive Neuroscience, vol. 11, pp. 473–490, 1999.

[12] J. De Winne, P. Devos, M. Leman, and D. Botteldooren: "With no attention specifically directed to it, rhythmic sound does not automatically facilitate visual task performance," Frontiers in Psychology, vol. 13, 894366, 2022.

[13] T. Van Renterghem: "Clustering individuals based on performance in audio-visual testing," Proceedings of Forum Acusticum/Euronoise 2025, Malaga, Spain.