



HAL
open science

Nature and the City: Audiovisual interactions in pleasantness and psychophysiological reactions

V.B. Bogdanov, Catherine Marquis-Favre, Marylise Cottet, Brice Beffara Bret, Fabien Perrin, Dominique Dumortier, Wolfgang Ellermeier

► **To cite this version:**

V.B. Bogdanov, Catherine Marquis-Favre, Marylise Cottet, Brice Beffara Bret, Fabien Perrin, et al.. Nature and the City: Audiovisual interactions in pleasantness and psychophysiological reactions. Applied Acoustics, 2022, 193, pp.108762. 10.1016/j.apacoust.2022.108762 . hal-03924495v1

HAL Id: hal-03924495

<https://hal.science/hal-03924495v1>

Submitted on 5 Jan 2023 (v1), last revised 14 Nov 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Nature and the City: audiovisual interactions in pleasantness and**
2 **psychophysiological reactions**

3

4 **BOGDANOV VB***, **MARQUIS-FAVRE C**, **COTTET M**, **BEFFARA B**, **PERRIN F**,
5 **DUMORTIER D**, **ELLERMEIER W**

6

7 * *corresponding author* **BOGDANOV Volodymyr Borysovych**, University of Lyon, Ecole Nationale
8 *des Travaux Publics de l'Etat, Laboratoire de tribologie et de dynamique des systèmes (LTDS) UMR*
9 *5513, 3 rue Maurice Audin, 69518 Vaulx-en-Velin, France* Volodymyr.BOGDANOV@entpe.fr

10 **MARQUIS-FAVRE Catherine**, University of Lyon, ENTPE, LTDS UMR 5513, Vaulx-en-Velin,
11 France Catherine.MARQUISFAVRE@entpe.fr

12 **COTTET Marylise**, University of Lyon, CNRS, ENS de Lyon, Lyon, France [marylise.cottet@ens-](mailto:marylise.cottet@ens-lyon.fr)
13 [lyon.fr](mailto:marylise.cottet@ens-lyon.fr)

14 **BEFFARA Brice**, University of Nantes, UFR Psychology Brice.Beffara@univ-nantes.fr

15 **PERRIN Fabien**, University of Lyon, CRNL, Bron, France fabien.perrin@univ-lyon1.fr

16 **DUMORTIER Dominique**, University of Lyon, ENTPE, LTDS UMR 5513, Vaulx-en-Velin, France
17 Dominique.DUMORTIER@entpe.fr

18 **ELLERMEIER Wolfgang**, Institute of Psychology, Technical University of Darmstadt, Germany
19 ellermeier@psychologie.tu-darmstadt.de

20

21 Published in [Applied Acoustics](#)
22 <https://doi.org/10.1016/j.apacoust.2022.108762>

23 Revised submitted version

24

25 **Abstract**

26

27 The presence of natural elements in the urban environment improves wellbeing. Comparisons
28 of the different psychophysiological reactions to both auditory and visual perceptions of urban
29 environments are proposed as a method for shedding light on the mechanisms involved in the
30 perceptual audiovisual interaction phenomenon, by which visual context influences auditory
31 perception, and vice versa. The results from combined psychophysical and physiological
32 methodology should improve knowledge of the perceptual consequences of variations in the
33 presence of nature in the city. In this study, 61 participants, mainly students, were exposed to
34 urban soundscapes accompanied by street scenes. Audio excerpts of bird song or road traffic
35 were paired with images with or without vegetation. Half of the participants (n=30) were asked
36 to judge the auditory pleasantness after each audiovisual presentation, whereas the other half
37 (n=31) judged the visual pleasantness of the scene. Heart rate, pupil area, gaze fixations and
38 electromyography of the eyebrow muscle (*corrugator supercillii*) were concurrently recorded.
39 We demonstrated a strong perceptual audiovisual interaction: auditory pleasantness was
40 higher when accompanied by a natural visual context, and visual pleasantness was higher
41 when accompanied by natural sounds. Electromyography and heart rate were particularly
42 sensitive to the presence of nature in the soundscapes, whereas gaze behavior discriminated
43 only visual nature elements. The natural sounds were associated with lower heart rate and
44 *corrugator supercillii* electromyographic activity, whereas natural images were associated with
45 slower gaze behavior. Pupil area was sensitive to both auditory and visual stimulations. The
46 pupil was relatively constricted by all natural stimuli and discriminated different unpleasant
47 traffic noise samples and the modulation of visual pleasantness by these sounds. This
48 suggests a distinctive role of arousal level in the modulation of pupil reactions, which appear
49 well suited to the study of audiovisual interactions and the perceptual effects of natural
50 elements in urban environments.

51 **Keywords:** audiovisual, nature, pleasantness, arousal, attention

52 1. Introduction

53

54 The densification of urban areas, activities generated by economic growth and
55 transportation, are some of the main driving forces for noise pollution, which is considered as
56 a major environmental health problem. Studies have highlighted various adverse impacts of
57 noise exposure on well-being and health, such as sleep disturbance (e.g., Muzet, 2007),
58 effects on children's cognition and health (e.g., Clark et al., 2013), annoyance (e.g., Miedema
59 & Oudshoorn, 2001), myocardial infarction (e.g., Babisch et al., 2005), diastolic blood pressure
60 and hypertension (e.g., Petri et al., 2021). Noise exposure incurs a financial burden on society
61 (e.g., World Health Organization. Regional Office for Europe, 2011). Urban areas are typically
62 exposed to a multitude of noise sources (e.g., road traffic noise, railway noise, road traffic
63 noise, recreational noise). Road traffic noise is one of the prominent threats to wellbeing in
64 cities, causing emotional stress, impeding work performance, and forming a risk factor for
65 sleep disturbance, cardiovascular diseases and hypertension (e.g., Singh et al., 2018). The
66 exposure of city dwellers to noise pollution is thus an important issue. The presence of nature,
67 whether visually perceived (e.g., the presence of plants, trees, or a river) or heard (e.g., bird
68 song, water lapping, wind in trees), is known to improve wellbeing (e.g., Franco et al., 2017).
69 Soundscape studies have shown that when natural sounds like bird song or water sounds are
70 combined with road traffic noise, they improve judgments of auditory pleasantness of urban
71 environments (Lavandier & Defréville, 2006, Jeon et al., 2012). Natural sounds increase the
72 perceived tranquility of urban areas (Cassina et al., 2018), and are thought to play an important
73 role in nature-based rehabilitation (Cerwén et al., 2016). From a visual perspective, looking at
74 vegetated scenes was demonstrated to bring important benefits, such as stress reduction and
75 better psychological restoration (Kaplan, 1995). For example, it was shown that green spaces
76 are favorable for stress reduction as indicated by a reduced cortisol level (Ward Thompson et
77 al., 2012). However, there is still much that remains to be learned about the specific pathways
78 linking green spaces to health (Markevych et al., 2017). In particular, previous studies
79 indicated that audiovisual interactions (here defined as the perceptual phenomenon according

80 to which visual context influences auditory perception, and vice versa) were observed in
81 pleasantness judgments, especially when nature is present. Actually, it is known that the
82 presence of nature in the visual context affects auditory judgments (e.g., Preis et al., 2016).
83 For example, the auditory unpleasantness due to road traffic noise in urban areas was judged
84 to be about 20-30 % less when auditory stimuli were combined with a visual context of natural
85 scenes compared to the context of a scene with apartment blocks only (Viollon et al., 2002).
86 Similar findings were also highlighted in studies dealing with noise annoyance (e.g., Chau et
87 al., 2018). Outdoor vegetation visible from the living room window reduces noise annoyance
88 (e.g., Van Renterghem & Botteldooren, 2016). However, not all studies confirmed such
89 findings when noise annoyance was assessed in a visual context with or without vegetation
90 (Sun et al., 2018). Thus, the mechanisms linked to the impact of natural factors and
91 audiovisual interactions on wellbeing in the context of the urban environment are therefore far
92 from clear. Several psychological and physiological human responses deserve when exploring
93 these mechanisms. In particular it could be promising to compare how the changes in
94 parameters of audiovisual excerpts of urban environments cause variations in emotional
95 valence, arousal and attention. In this manner, it may be possible to identify the key processes
96 by which the presence of nature in the urban environment can modify human behavior.
97 Pupillary dilation is known to be sensitive to rapid arousal (Bradley et al., 2008), slowing of
98 heart rate is evidence of an experienced threat, and an increase in the activity of *corrugator*
99 *supercilii* face muscle has been interpreted as evidence of discrimination between negative
100 and positive emotional valence (Bradley & Lang, 2000). Gaze behavior reveals changes in
101 visual attention. A reduced number of gaze fixations and longer fixation durations may be
102 physical manifestations of altered attentional states that have been characterized as
103 involuntary sustained attention or fascination (Berto et al., 2008). Such parameters could
104 therefore be of interest for revealing the basic mechanisms of audiovisual interactions and for
105 helping us to understand the wellbeing effects of nature in urban environments.

106 The present study aimed to explore the mechanisms involved in perceptual audiovisual
107 interactions observed during the auditory and visual appraisal of urban environments in the

108 presence of nature. Sixty-one participants, mainly students, were exposed in laboratory
109 conditions to audiovisual stimuli stemming from urban environments. This study presented the
110 following main features:

111 (1) Both the auditory and visual pleasantness of urban environments was studied while
112 varying the degree of nature elements present in the audiovisual stimulation.

113 (2) A combined psychophysical and physiological methodology was used to investigate
114 both the auditory and visual perceptions of urban environments.

115 (3) Audiovisual interactions in either auditory or visual perception was highlighted, i.e.
116 the influence of the visual scenery on auditory perception, and vice versa.

117 (4) Various peripheral psychophysiological measures were used to probe changes in
118 emotional valence, arousal and attention during audiovisual stimulation.

119 Our hypothesis was that both the audible and visible presence of nature in urban
120 environments would not only induce greater auditory or visual pleasantness but also affect
121 certain physiological measures.

122

123 **2. Method**

124

125 **2.1 Participants**

126

127 Sixty-one healthy adults between 19 and 57 years-of-age (mean age 27.93, SD=10.33)
128 were recruited through advertisements sent via the institutional mailing list of the Graduate
129 School of Civil, Environmental and Urban Engineering (ENTPE – Lyon, France), where the
130 research experiment took place. This sample size corresponded to the number of participants
131 used in previous studies with a similar design (Berto et al., 2008; Bradley et al., 2008; Bradley
132 & Lang, 2000; Viollon et al., 2002). The critical sample size calculated for our study was 26
133 participants, and obtained with G*Power 3 (Faul et al., 2007) for an effect size (f) 0.33, power
134 of 0.80, $p < 0.05$ and four measures in repeated measures ANOVA, assuming no overall
135 correlation between the measures.

136

137 Participants were students and staff of the aforementioned School. They declared to
138 have normal or corrected-to-normal vision, with self-reported absence of neurological,
139 psychiatric, cardiovascular, respiratory, and auditory disorders. This study followed the
140 standards of the Declaration of Helsinki and was approved by the regional ethics committee
141 (Sud-Ouest et Outre-Mer III) on 01/08/2018, Research protocol n° 2018/58, ID-RCB: n°2018-
142 A01621-54. Each participant gave written informed consent and was reimbursed for
143 participation with a gift certificate worth 10 Euros.

144

145 **2.2 Auditory and visual stimuli**

146

147 The protocol used two categories of auditory stimuli: 18 urban road traffic noise samples
148 ranging from 48 to 62 dB(A) and 18 samples of bird song ranging from 40 to 45 dB(A), each
149 sample lasting for 4 seconds. All noise samples stemmed from recordings made using the
150 ORTF technique (devised at the former Office de Radiodiffusion Télévision Française), with
151 Schoeps MSTC 64 microphones equipped with Schoeps BBG windfields, and a Tascam
152 portable recorder with 24-bit resolution and a sampling frequency of 44.1 kHz (Morel et al.,
153 2012). The road traffic noise samples consisted of vehicle pass-bys (light vehicles, buses,
154 heavy vehicles or motorcycles/mopeds) recorded by Morel and colleagues (Morel et al., 2012).
155 These samples were previously used in studies dealing with noise annoyance (Gille et al.,
156 2016). The bird song samples were recorded on the ENTPE campus site (Vaulx-en-Velin,
157 France). A 300-Hz high-pass filter was applied to the bird song samples to remove low
158 frequency content resulting from the background noise of suburban Lyon and to insure good
159 sound reproduction quality in the quiet experimental room. For the sound reproduction, the
160 mean sound pressure levels of all sound stimuli were set to the levels of the noise sources
161 measured on site. A 200-ms amplitude fade-in and a 500-ms fade-out with a quadratic
162 progression kernel were applied to all 36 sound stimuli.

163 The experimental protocol also used two categories of visual stimuli: 18 images
164 representing vegetated scenes and another 18 images depicting non-vegetated scenes
165 (examples in **Fig. 1**). These visual stimuli were pictures of urban scenery in and around Lyon
166 (France) taken specifically for this experiment during the spring and summer seasons. The
167 camera used to take the pictures was a SONY© DSC-HX10V, with exposure of 1/400-1/1000
168 s, ISO 100, and automatic white balance.

169 There was no overall significant difference in eye illuminance between the vegetated
170 and non-vegetated images when displayed on the screen during the experiment (11.0 lux, SD
171 = 1.1, n = 18 and 10.9 lux, SD = 1.2, n = 18, respectively, $p=0.81$, t-test). The initial resolution
172 was 4896 x 2752 pixels, while the size of the images on the screen was 45.3 x 25.5 cm with
173 a resolution of 1632 x 918 pixels. All photographs were taken in the same manner: the
174 photographer stood at the intersection of two streets with their back turned towards the
175 intersection and the lens pointing in the opposite direction. This protocol was followed to avoid
176 visual and auditory incongruities. All photographs were taken in good weather in the city of
177 Lyon and its suburbs.

178

179 **2.3 Laboratory setup**

180

181 Participants were seated in front of a desk (Fig. 1, A3) supporting a computer monitor
182 and eye-tracking equipment (described in 2.6) inside a dark sound-proof room with a
183 background noise level of 18.5 dB(A). A computer mouse for making responses was available
184 on the desk. The LCD monitor used (Lenovo ThinkVision LT2423WC, 24-inch FHD LED
185 Backlit Monitor, TN technology) had a resolution of 1920 x 1080 with a typical contrast ratio of
186 1000:1. The viewing distance was about 50 centimeters. Given that the width/height of the
187 display was 53/30 centimeters, it subtended approximately 30 degrees of visual angle
188 vertically and 50 degrees of visual angle horizontally. Sound stimuli were reproduced using
189 two active loudspeakers (Dynaudio BM5A Actif), and one subwoofer (Dynaudio Acoustics
190 BM9S) to ensure good reproduction of low frequencies of the road vehicle pass-by noise

191 samples. The loudspeakers were located behind the seated participants at a height of 1.50
192 with the subwoofer on the floor between them. The center of the participant's interaural axis
193 and the loudspeakers formed an equilateral triangle with sides measuring 1.5 m. Thus, the
194 sounds were reproduced in such a way that the participants perceived the vehicles to pass
195 behind their backs, and therefore the noise sources would have been outside of their field of
196 vision.

197

198 **2.4 Experimental paradigm**

199

200 The PsychToolbox application (Brainard, 1997; Pelli, 1997) in MATLAB was used for
201 implementing the experimental paradigm based on 2 categories of sounds (18 bird song and
202 18 road traffic noise samples) and 2 categories of visual scenes (18 vegetated and 18 non-
203 vegetated urban pictures). To ensure that physiological measures were reliable, stimuli were
204 presented with short duration and repeated conditions (e.g., Bradley & Lang, 2008) using
205 different samples per condition to avoid habituation (e.g., Groves & Thompson, 1970). During
206 each audiovisual presentation, the participants were simultaneously exposed to one picture
207 and one sound for a duration of 4 s. Four paradigm conditions resulted from the 2-by-2
208 categories of sounds and images: bird song and vegetated image; bird song and non-
209 vegetated image; traffic noise and vegetated image; traffic noise and non-vegetated image.
210 Each of the 36 visual stimuli was randomly presented twice, once paired with a bird song
211 sample and once with a traffic noise sample. Each of the 4 conditions was repeated using 18
212 trials with different image-sound combinations, giving a total of 72 audiovisual stimulations
213 (see **Fig. 1, A1 and A2**). Whether participants assessed the auditory or visual pleasantness
214 of the audiovisual combination was randomly determined before the start of the experiment
215 and constituted a between-subjects factor (**Fig. 1, B**).

216 During the experiment, heart rate, pupil size variations, gaze fixations, and EMG activity
217 of the *corrugator supercilii* (eyebrow muscle) were recorded. The systems used for these
218 physiological measures are described in sections 2.6 to 2.8.

219 After each presentation, a 4-s blank screen with a fixation dot was shown. Subsequently,
220 participants made their response by judging the auditory or visual pleasantness of the
221 presented urban environment (depending on the task required – cf. instructions in section 2.5).
222 They gave their rating using a bipolar continuous visual-analogue scale. The scale displayed
223 numerical labels from 0 to 10, with two opposite end points “unpleasant” and “pleasant”, to
224 facilitate ratings along the continuum (cf. Fig. 1, B3), as typically used in studies recording
225 affective reactions (e.g., Bradley & Lang, 2000; Frederickson & Dunsmuir, 2009). Participants
226 were instructed to move the mouse cursor from the center of the screen to the location on the
227 scale reflecting the pleasantness felt. Placing the mouse on the line made a white spot appear
228 at that position, thus providing feedback on a successfully registered judgment. The mouse
229 cursor then had to be brought back to the center of the screen so that the next judgment
230 started from about the same hand position. Participants had 4 s to make their response.
231 Subsequently, a fixation dot was displayed for a random duration between 6 and 14 s before
232 the next trial commenced.

233 Before the experiment, participants completed four training trials to get acquainted with
234 the task and the rating scale. The trials included all four experimental conditions and were
235 randomly composed of 4 sound excerpts (2 bird songs and 2 traffic noises) and 4 images (2
236 vegetated scenes and 2 non-vegetated scenes). These stimuli were not used in the main
237 experiment. The experimental session took about 30 minutes.

238

239 **2.5 Judgment task instructions**

240

241 The instructions given to the participants were similar to those used in a previous study
242 on audiovisual interaction (Viollon et al., 2002).

243 Approximately half of the participants (n=30) were randomly assigned to the “auditory
244 task”. They received the following instructions in French: «*L’expérience vous propose des*
245 *extraits audiovisuels d’environnements urbains. Imaginez-vous à chaque fois dans le lieu*
246 *représenté par l’image. Si vous étiez dans ce lieu, comment jugeriez-vous l’environnement*

247 **sonore** et uniquement l'environnement **sonore** ?», which translate as “During the experiment,
248 you will be exposed to audiovisual excerpts of the city environment. Imagine yourself each
249 time at the place shown in the image. If you were at this place how would you judge the sound
250 environment and only the sound environment?”.

251 The other half of the participants (n=31) were assigned to the “visual task”. They
252 received the same instructions except for the last question which was: “*Si vous étiez dans ce*
253 *lieu, comment jugeriez-vous l'environnement **visuel** et uniquement l'environnement **visuel** ?»,*
254 *which translates as “If you were at this place how would you judge the **visual** environment*
255 *and only the **visual** environment?.”*

256

257 **2.6 Eye-tracking**

258

259 The gaze position and pupil area of both eyes were recorded using an Eyelink® Portable
260 Duo eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada) with a sampling frequency
261 of 1 kHz. This eye-tracker operated in remote mode as a head-free system with a target
262 sticker: a standard cue placed on the forehead of the participants and automatically identified
263 by the eye-tracking system. This setup allows recording of eye position without restraining the
264 head position. The test phase was preceded by a standard nine-point calibration procedure
265 for mapping gaze position to screen position. The targets used for calibration were the default
266 targets of the Eyelink® apparatus.

267

268 **2.7 Eye-tracking preprocessing**

269

270 Pupil area and gaze position data were analyzed using MATLAB. The algorithm for the
271 preprocessing of pupil size data and automatic artifact rejection followed recommendations
272 proposed in the Pupil Size Preprocessing Code (Kret and Sjak-Shie, 2019). This routine sets
273 certain criteria for the rejection of samples associated with blinks. First dilation speed outliers
274 and edge artifacts were identified as pupil size samples that were outside of a predefined

275 feasible range. Second, measurements with abnormally large deviations (caused by eye-
276 blinks) from a smooth trend line were removed. To detect dilation speed outliers, the median
277 absolute deviation was used, this is a robust and outlier resilient data dispersion metric.
278 Finally, short-series eye-tracking samples separated from the rest of the recording by long
279 temporal intervals were removed before analyzing the data.

280 This preprocessing was applied to both right-eye and left-eye data. If the proportion of
281 rejected samples was more than 10% higher in one eye than in the other, the eye with the
282 most rejected samples was not used in further analysis; otherwise the valid samples of the
283 two channels were averaged for pupil size and gaze position. Subsequently, the pupil size
284 data that were rejected were replaced with cubic-interpolated samples. Depending on the
285 number and duration of blinks, the quantity of rejected data varied from a few samples to
286 several seconds. A typical blink duration was about 300-400 ms.

287

288 **2.8 Electrophysiological recording**

289

290 Electromyography (EMG) and electrocardiography (ECG) were continuously recorded
291 during the experiment using AgAgCl electrodes at a 1024-Hz sampling rate (16 bits) and the
292 MicroMed SD64 electrophysiological system (Micromed, Treviso, Italy). As recommended for
293 recording of the EMG activity of the *corrugator supercilii* muscle (Fridlund & Cacioppo, 1986),
294 the electrodes of a bipolar channel were placed over the right eye-brow (**Fig. 1, A3**). For ECG
295 recording, the electrodes of the other bipolar channel were placed on the left and right wrists
296 (or on the left ankle as a function of the detectability of the R peak, the most important peak
297 of the ECG). The ground electrode was placed on the left side of the forehead. In both cases,
298 the skin was cleaned, gently abraded with gel (NUPREP; Weaver and Co.) and for better
299 contact, a conductive adhesive gel (Tensive; Parker Laboratories) was placed under each
300 electrode. The electrode impedances were kept below 5 kOhm. During the acquisition a band-
301 pass filtering of 0.1-8 Hz and 0.1-1 Hz was applied to the EMG and ECG channels

302 respectively, using the BRAIN QUICK recording system. The electrophysiological signal was
303 amplified 1000 times.

304

305 **2.9 EMG preprocessing**

306

307 The EMG signals were preprocessed in a similar way as in other recent studies
308 examining the activity of the corrugator supercilii muscle during emotional experiences
309 (Golland et al., 2018). The EMG spectrogram was estimated with a short-time Fourier
310 transform. The Fourier-transform segment length was 100 ms, with adjoining segments
311 overlapping by 50 ms. For each segment, the sum of the natural logarithms of the power
312 spectral densities was computed for the frequencies 10 – 200 Hz, excluding the range of 45-
313 55 Hz to suppress power line noise. Then, the time series were interpolated to the initial
314 sampling rate (1 kHz) with shape-preserving piecewise cubic interpolation.

315

316 **2.10 ECG preprocessing**

317

318 The ECG signals were filtered using fast Fourier transformation, excluding the range of
319 45-55 Hz. Then, 20-ms moving-average smoothing was applied. R-peak times were detected
320 using a wavelet approach similar to the one reported in (Sundarasekar et al., 2018) and
321 following the recommendations in MATLAB guidelines (*R Wave Detection in the ECG -*
322 *MATLAB & Simulink*, n.d.).

323 The RR intervals, that is the intervals between R peaks, were then interpolated to a 1-
324 kHz sampling rate to obtain a whole-session RR-interval time series. Finally, the time series
325 were smoothed using a 1-s moving average.

326

327

328

329

330 3. Results

331

332 3.1 Behavior

333

334 The individual behavioral responses collected for the 4 experimental conditions were
335 used in 2*2 repeated-measures analyses of variance (RM-ANOVAs), with the within-subject
336 factor “Sound category” (2 levels: bird song, denoted later by B, and road traffic noise, denoted
337 by T) and the within-subject factor “Image category” (2 levels: vegetated scene, denoted by V,
338 and non-vegetated scene, denoted by nV). RM-ANOVAs were performed separately for the
339 two tasks (i.e., judging the pleasantness of the auditory or the visual environment, hereafter
340 referred to as the “auditory task” and “visual task”, respectively). To estimate the effect size,
341 partial eta squared η_p^2 was calculated, a parameter frequently reported for RM-ANOVA
342 designs (Bakeman, 2005).

343 As expected, the auditory or visual environments that were more natural were judged to
344 be more pleasant than those that were less natural. An effect of the main factor “Sound
345 category” on the pleasantness score during the auditory task was unsurprisingly observed
346 ($F(1,29) = 241.6$, $p < 0.001$, partial eta squared $\eta_p^2 = 0.89$; **Fig. 2, A1**). The effect of the main
347 factor “Image category” on visual pleasantness was also highly significant in visual task
348 ($F(1,30) = 34.49$, $p < 0.001$, $\eta_p^2 = 0.53$; **Fig. 2, A2**).

349 Strong perceptual audiovisual interaction was also observed in both tasks: the visual
350 context affected auditory ratings and vice versa. The effect of the main factor “Image category”
351 on auditory pleasantness was highly significant ($F(1,29) = 17.56$, $p < 0.001$, $\eta_p^2 = 0.37$),
352 demonstrating that pleasantness in the auditory task was increased by the presence of
353 vegetation in the depicted environment (see **Fig. 2, A1** and the red dotted contour to highlight
354 the perceptual audiovisual interaction). The effect of the main factor “Sound category” on
355 visual pleasantness was highly significant ($F(1,30) = 20.33$, $p < 0.001$, $\eta_p^2 = 0.40$), showing that
356 the natural auditory context, represented by bird song, made the visual scenery more pleasant

357 in the visual task (see **Fig. 2, A2** and the red dotted contour to highlight the perceptual
358 audiovisual interaction). Both contextual effects provided evidence of perceptual audiovisual
359 interaction in both auditory and visual pleasantness judgments.

360

361 **3.2 Pupil area, EMG and heart rate**

362

363 The following processing and analyses were applied to pupil, EMG and heart rate
364 physiological responses. First, for each participant, seventy-two epochs were extracted, each
365 spanning from 4 s prior to the onset of the 4-s audiovisual presentation until 4 s after its end.
366 A correction was applied to the 72 epochs by subtracting a baseline. For heart rate and EMG,
367 the baseline was the average of the physiological measurements recorded during the 10 s
368 before the stimulus onset. For pupil area, the baseline was the average of the physiological
369 measurements during the 2 s before stimulus onset. The statistical noise level was estimated
370 as a minimum to maximum range during the 20-s segment around presentation onset, and
371 the 14 noisiest epochs of each participant were excluded from further analysis. Additionally,
372 those pupil size epochs, in which more than 40% of samples were rejected because of blinks
373 in the 20-s segment around the presentation onset, were rejected. There were up to three
374 such trials in the recordings of the majority of the participants, whereas 8% of participants
375 (5/61) had more than seven trials rejected, with the maximum number of rejected trials per
376 participant being 27.

377 Second, the average epoch time-series from 4 s prior to the stimulus onset until 4 s after
378 its end, was down-sampled by a factor of 100 (i.e. to 10 Hz) for each of the 4 audiovisual
379 conditions and each of the 61 participants. When compared with the pleasant counterparts,
380 the unpleasant conditions, namely traffic noise or non-vegetated scenes provoked
381 pronounced pupil dilations and smaller EMG and heart rate effects. From these individually
382 averaged temporal physiological responses, the dynamic physiological response difference
383 was considered to highlight the potential change in physiological responses due to the
384 variation in the extent of nature within the sound and image categories of the audiovisual

385 stimulations (i.e., bird song vs road traffic noise, and vegetated images vs non-vegetated). To
386 illustrate the differences in dynamic physiological responses for pupil area, EMG and heart
387 rate, they were averaged over the participants of each group (i.e., for auditory and visual tasks,
388 respectively) and are presented in **Figure 3, A-C**. The effect of the variation in nature presence
389 within the sound category (i.e., bird song vs. road traffic noise) is referred to as the auditory
390 effect [B-T], while the effect of the variation in nature presence within the image category (i.e.,
391 vegetated vs non-vegetated images) is referred to as the visual effect [V-nV].

392 For either of the two groups, a median-based 1000-iteration-bootstrap bias-corrected
393 and accelerated percentile 95%-confidence interval (DiCiccio & Romano, 1988) was
394 estimated for each timepoint of the differential responses. Then, the average differential
395 response for every participant was computed within the window of interest of 4 s-audiovisual
396 stimulation. These individual scores were used in group-level summary statistics with the
397 median-based bootstrap technique and RM-ANOVAs. Furthermore, medians and bootstrap
398 group-level 95%-confidence intervals (CI) were also considered for the differential scores of
399 pleasantness, to simultaneously illustrate the significance of the effect of audiovisual
400 stimulation on both pleasantness judgments and physiological reactions. The **Figures 4 and**
401 **5** display the median values and 95%-CIs to highlight the auditory effect [B-T] (i.e., Bird song
402 vs. Traffic noise) and the visual effect [V-nV] (i.e., Vegetated vs. non-Vegetated) on the
403 differential pleasantness scores (**Fig. 4, A**) and the differential physiological responses (**Fig.**
404 **4, B to Fig. 5, C**). In these figures, a significant effect can be considered to have occurred
405 when the confidence interval does not embrace the zero line.

406 RM-ANOVAs with the factors “sound” (i.e., Bird song vs. Traffic noise, [B-T]) and “image”
407 (i.e., Vegetated vs. non-Vegetated, [V-nV]) were performed for the auditory and visual tasks,
408 respectively, to give the magnitudes of significant effects according to η_p^2 .

409 The median and 95% CI results show that pupil area was higher when the participant
410 was presented with traffic noise rather than birdsong, both in the auditory task and in the visual
411 task. This is highlighted in **Figure 5, A** by the significant auditory effect [B-T] on pupil area

412 during both the auditory and visual tasks. This effect is also observed in the RM-ANOVA
413 results ($F(1,29) = 30.6, p < .001, \eta_p^2 = 0.51$ for the auditory effect [B-T] on pupil area during the
414 auditory task, and $F(1,30) = 10.8, p = .003, \eta_p^2 = 0.27$ during the visual task). Pupil area was
415 also higher after observing non-vegetated scenes than after observing vegetated scenes (**Fig.**
416 **5, A**). The visual effect [V-nV] was significant in the RM-ANOVAs, both when the sound
417 environment was judged ($F(1,29) = 10.0, p = .004, \eta_p^2 = 0.26$) and when the visual environment
418 was evaluated ($F(1,30) = 5.9, p = .021, \eta_p^2 = 0.16$).

419 The bootstrap method indicated that the *corrugator supercilii* muscle (EMG) appeared
420 to be more strongly activated during traffic noise than during bird song, in both the auditory
421 task and the visual task (**Fig. 5, B**). This significant auditory effect [B-T] is also observed in
422 the RM-ANOVA results ($F(1,29) = 8.6, p = .006, \eta_p^2 = 0.23$ during the auditory task, and $F(1,30)$
423 $= 5.1, p = .031, \eta_p^2 = 0.15$ during the visual task). In contrast, image content had no significant
424 differential effect on the EMG (**Fig. 5, B**). Heart rate showed a slight elevation only when
425 unpleasant sound content was presented in the visual task (**Fig. 5, C**), with the auditory effect
426 [B-T] showing statistical significance in the RM-ANOVA ($F(1,30) = 6.2, p = .019, \eta_p^2 = 0.17$).

427

428 **3.3. Relationships between pupil reactions and behavior**

429

430 The results reported so far have shown that pupil dilations were obtained in response to
431 the less pleasant sound and image categories. Beyond this finding we wanted to explore
432 whether pupil size during exposure was linked to the subsequent behavioral judgments on
433 samples within sound and image categories. Differences in pleasantness and physiological
434 reactions within the image category will be presented in a forthcoming study dedicated to
435 visual perception. Our investigation therefore concerns the potential effect of variation in
436 auditory pleasantness on the pupil size within each sound category. As there was negligible
437 variation in pleasantness judgments for the different bird song samples stemming from the
438 Lyon area, we focused on differences in reactions to the urban road traffic noise samples (light

439 vehicle, bus, heavy vehicle, and motorcycle/moped pass-bys). To investigate the potential
440 effect on pupil size, two sub-conditions of traffic noise samples were considered: traffic noise
441 samples with high unpleasantness scores versus samples with low unpleasantness scores.

442 Given the variation in auditory pleasantness judgments across the 18 traffic noise
443 samples used (data from $n=30$ participants), five sound samples were selected for high
444 unpleasantness (mostly motorcycle/moped pass-by noises) and five samples for low
445 unpleasantness (mostly light vehicle pass-by noises). These subsamples are shown in **Fig. 6,**
446 **A1** in terms of their median pleasantness ratings relative to median bird song pleasantness
447 ratings during the auditory task.

448 We then examined whether these two selected sets of traffic noises differed in their
449 effects on pleasantness judgments in the visual task. As shown in **Fig. 6, A2**, we found lower
450 visual pleasantness relative to bird song during presentation of highly-unpleasant traffic noise
451 samples (HT) in comparison with presentation of less-unpleasant traffic noise samples (LT).
452 This significant perceptual audiovisual interaction is displayed in **Fig. 6, A2** in terms of median
453 ratings and 95%CIs relative to bird song.

454 Following the same rationale as the one used for the analysis described in section 3.2,
455 the response differences due to the two sets of traffic noises (Less-unpleasant Traffic noise
456 vs. Highly-unpleasant Traffic noise, i.e. [LT – HT]) are hereafter used in the group-level
457 summary statistics with the median-based bootstrap technique and RM-ANOVAs, and
458 displayed in **Fig.6, A2-B1-B2**.

459 The highly-unpleasant traffic noise samples exerted significantly more pronounced
460 perceptual audiovisual interaction effects than the less unpleasant ones (see LT-HT, rightmost
461 data plot in **Fig. 6, A2**). In a one-way RM-ANOVA with the factor LT-HT, the effect of this factor
462 on the visual pleasantness rating difference was significant ($F(1,30) = 41.8, p<0.001$) with an
463 important effect size ($\eta_p^2 = 0.58$).

464 A similar pattern was observed for the pupil reactions. The difference in pupil dilation
465 was considered for low versus high unpleasantness traffic noise exposure (auditory and visual

466 tasks pooled, $n=61$). The time course of differences in pupil reaction exhibited a significant
467 relative increase in pupil size for highly-unpleasant traffic noise samples compared with less-
468 unpleasant ones (**Fig. 6, B1**).

469 When the average pupil area change was separately computed within the auditory or
470 visual task participant groups, a significantly greater pupil size change was evident after highly
471 unpleasant traffic noise stimulation in both types of tasks (**Fig. 6, B2**). One-way RM-ANOVAs
472 with the factor LT-HT confirmed the important effect size obtained for this factor in the auditory
473 task ($F(1,29) = 25.4, p < 0.001, \eta_p^2 = 0.47$) in comparison with the effect size in the visual task
474 ($F(1,30) = 10.1, p = 0.034, \eta_p^2 = 0.25$).

475

476 **3.4 Gaze fixations and saccade statistics**

477

478 Seventy-two epochs were extracted from the vertical and horizontal gaze position data
479 ranging from the onset to the end of the 4-s presentation of the audiovisual content. Epochs
480 were rejected if either less than 12% of data points per epoch were available or more than
481 40% of samples were rejected because of blinks in the 20-s segment around the presentation
482 onset. An instantaneous velocity threshold of 0.6 pixel ($20^\circ/s$) was applied to identify
483 saccades. The remaining episodes were labelled as fixations. The average position and
484 duration were identified for each fixation episode. If the positions of consecutive fixations did
485 not differ by more than 25 pixels (0.8°) we considered them to belong to the same fixation.
486 The duration of such fixations was computed from the onset of the first to the end of the last
487 episode belonging to it. For each epoch, the average fixation duration and the cumulative
488 distance of the saccades, separating the fixations were computed. Then, the average scores
489 were computed for each of the 4 conditions for each participant, except for the epochs
490 excluded because of the noisiness of the pupil data. Following the same rationale as the one
491 described in section 3.2, individual response differences were considered in the group-level
492 summary statistics using the median-based bootstrap technique and RM-ANOVAs to analyze

493 the auditory effect [B-T] (i.e., Bird song vs. Traffic noise) and the visual effect [V-nV] (i.e.,
494 Vegetated vs. non-Vegetated) on fixation duration and cumulative saccade distance.

495 The analysis of eye-tracking data yielded a set of significant effects that depended on
496 the audiovisual stimulation. Fixation duration was strongly affected by the image category of
497 the visual scenes, being significantly longer for the vegetated scenes in both tasks (**Fig. 4, B**).
498 The RM-ANOVA with the factors “sound” (i.e., Bird song vs. Traffic noise, [B-T]) and “image”
499 (i.e., Vegetated vs. non-Vegetated, [V-nV]) also revealed the significant visual effect [V-nV] on
500 fixation duration during the auditory task ($F(1,29) = 14.4, p < 0.001, \eta_p^2 = 0.33$) and the visual
501 task ($F(1,30) = 19.0, p < 0.001, \eta_p^2 = 0.39$). Consistent with the longer fixation durations, the
502 cumulative saccade distance was shorter for vegetated scenes (**Fig. 4, C**), though the
503 bootstrapping technique confirmed this effect only during the visual task. This corresponded
504 to the RM-ANOVA result, with a significant visual effect [V-nV] on the cumulative saccade
505 distance ($F(1,30) = 8.8, p = 0.006, \eta_p^2 = 0.23$). A significant result was observed for the auditory
506 effect during the visual task: bird song significantly increased gaze travel distance ($F(1,30) =$
507 $4.3, p = 0.047, \eta_p^2 = 0.13$).

508

509 **4. Discussion**

510

511 The objective of the present laboratory study was to investigate audiovisual interactions
512 in judgments of the pleasantness of urban environments with varying levels of nature in the
513 visual landscapes or soundscapes. We hypothesized that these perceptual interactions would
514 be associated with specific objective psychophysiological responses, namely *corrugator*
515 *supercilii* electromyographic activity, heart rate, pupil reactions, and gaze behavior, with these
516 reflecting changes in valence, arousal, or attention.

517

518

519

520 **4.1 Behavioral findings**

521

522 Significant perceptual audiovisual interactions were demonstrated in both auditory and
523 visual judgment tasks. The presence of nature (bird song or vegetation) increased both the
524 auditory and visual pleasantness of the excerpts of urban environments presented in our
525 study.

526 First, we confirmed a previously reported modification of auditory pleasantness by visual
527 context (Viollon et al., 2002), with sound pleasantness being judged as greater when
528 presented in the context of vegetated scenes as opposed to non-vegetated scenes. We further
529 showed that visual pleasantness judgments were influenced by the auditory context. The
530 visual pleasantness of the depicted urban scenes was judged to be greater when they were
531 accompanied by bird song rather than road traffic noise. This result is in agreement with
532 previous findings that assessment of landscapes was negatively affected by man-made
533 sounds such as traffic noise in comparison with natural sounds such as bird song (Ren et al.,
534 2018). Moreover, we found that in visual pleasantness judgments, the modulation by the
535 auditory context varied depending on the degree of unpleasantness elicited by the road traffic
536 noise samples. The depicted urban scenes were judged to be less visually pleasant when
537 accompanied by more unpleasant (compared with less unpleasant) urban road vehicle pass-
538 by noise.

539 A similar phenomenon was studied in the interaction of videos and music, and was
540 termed the “affective congruency” effect. It refers to the enhancement of valence judgments
541 in combined audiovisual exposures in comparison with auditory or visual stimuli of the same
542 valence (either positive or negative) presented separately (Gao et al., 2017). This means that
543 pleasant sounds combined with pleasant images produce a stronger impression of overall
544 pleasantness. The same holds for negative valences (Gao et al., 2017). This was also shown
545 by Cassina et al. (2018) in a study dealing with urban soundscapes. They highlighted the
546 combined effect of negative visual elements and negative sound sources (e.g., road traffic) on
547 the perceived tranquility of urban environments, with tranquility being known to be closely

548 related to pleasantness (Li & Lau, 2020). However, the affective congruency effect highlighted
549 by Gao et al. (2017) was shown for positive vs. negative valence stimuli only if they were of
550 high arousal. There is evidence that the audiovisual integration of emotional stimuli yielding
551 similar results to ours is associated with specific activity in the medial prefrontal cortex (Gao
552 et al., 2020). This might reflect integrative top-down cognitive processes during audiovisual
553 integration. Another indication of the largely cognitive nature of this effect is the finding that
554 the degree of matching between visual and sound features can alter this relationship. Unlike
555 bird song or traffic noise, pleasantness judgments of human-generated sounds like speech or
556 footsteps do not appear to be sensitive to the visual context (Viollon et al., 2002). This
557 reservation might be explicitly tested in further studies.

558 We also showed that a number of physiological and gaze-behavior effects accompanied
559 these modulations of pleasantness judgments: pupil size, EMG, heart rate, saccadic distance
560 and fixation duration were all affected by the studied conditions. Physiological reactions and
561 gaze behavior were typically significant in both the auditory and visual tasks, or were not
562 significant in either one. On the other hand, the sound category (bird song or traffic noise) and
563 the image category (vegetated or not) provoked significant effects in only some of the objective
564 measures. In particular, the average fixation duration and cumulative saccade length were
565 significantly affected by visual content but not significantly by the auditory content in either
566 task (**Fig. 4, B-C**), whereas EMG and heart rate responded significantly to the auditory
567 component but not to the visual one (**Fig. 5, B-C**). The pupillary response changed significantly
568 as a function of both the visual and auditory content of urban environments (**Fig. 5, A**) and
569 therefore constituted the most reliable physiological measure of the perceptual audiovisual
570 interaction during pleasantness assessment.

571

572 **4.2 Gaze behavior discriminated vegetated from non-vegetated urban scenes**

573

574 In our study, when participants observed urban scenes containing vegetation, their
575 average fixation duration was longer and their saccadic distances were shorter in comparison

576 with observation of urban scenes without vegetation. This is consistent with previous findings
577 on gaze fixation during observation of natural scenes of high or low fascination (Berto et al.,
578 2008). However, this difference should not be attributed exclusively to the presence of
579 vegetation. It is known that even among different vegetated scenes, fixations are longest for
580 scenes evaluated as having the highest landscape quality (Cottet et al., 2018), which might
581 closely correspond to the pleasantness ratings captured in our study. Therefore, longer
582 fixations in our study may be associated with the potential of more-pleasant vegetated visual
583 environments to gently attract effortless attention. According to Attention Restoration Theory,
584 such changes improve wellbeing and facilitate recovery of attentional resources (Kaplan,
585 1995).

586 At the same time, the more-pleasant natural sounds did not affect gaze behavior in the
587 same way as the vegetated scenes. To the contrary, bird song facilitated saccades, which
588 may be evidence for a redistribution of visual attention as a function of the auditory context,
589 as has been reported by Ren & Kang (2015). Modification of the distribution of spatial visual
590 attention by the auditory context requires further exploration.

591

592 **4.3 Corrugator supercilii EMG and heart rate changes significantly discriminated** 593 **between traffic noise and bird song**

594

595 As expected, both the activity of the corrugator supercilii muscle and heart rate were
596 higher when participants listened to traffic noise rather than to bird song. These findings are
597 likely due to the unpleasant feelings caused by hearing motorized vehicles passing close-by.
598 This confirms earlier findings with other acoustic stimuli, in that sounds of negative valence
599 provoke unintentional activation of *corrugator supercilii* and an acceleration of heart rate
600 (Bradley & Lang, 2000; Park et al., 2018).

601 The corrugator supercilii EMG differed as a function of sound category, but was not
602 modified by the image category. This may be explained by the much smaller difference in

603 pleasantness across the image categories of the urban scenes presented (vegetated or not)
604 in comparison with the sound categories (bird song or traffic noise).

605 Heart rate acceleration in response to road traffic noise was observed only during the
606 visual task, thus reflecting implicit discomfort, elicited by the non-attended modality. However,
607 because of the moderate statistical significance of the heart rate data in the present study
608 ($p < 0.05$), this finding requires further confirmation. In at least one study, explicit pleasantness
609 was shown to be more strongly associated with changes in functional magnetic resonance
610 imaging-measured cerebral activity in the amygdala than it was with heart rate responses
611 (Irwin et al., 2011). This suggests that heart rate responses may be a less sensitive indicator
612 of pleasantness than *corrugator supercillii* EMG.

613

614 **4.4 The pupil reacted with greater dilation to both unpleasant auditory and visual** 615 **stimulation**

616

617 Pupil size increased significantly in response to both traffic noise and non-vegetated
618 images in comparison with their categorial counterparts, regardless of the task, be it visual or
619 auditory (**Fig. 5, A**). Compared with the image categories, the sound categories produced
620 larger pupillary response effects. However, it is conceivable that the pupillary response is
621 primarily accounted for by the difference in arousal mediated by upregulation of sympathetic
622 autonomic activity, rather than by the variation in valence (Bradley et al., 2008). Greater pupil
623 size is known to be associated with both highly negative and positive complex sounds,
624 indicating the role of arousal in regulation of this physiological variable (Oliva & Anikin, 2018).
625 Thus, the sound category effect on pupil size is very likely explained by arousal, potentiated
626 by the fact that the traffic noise samples had higher sound pressure levels than the bird song
627 samples. The finding that the arousal picked up by the pupillary dilation was more evident in
628 sound discrimination is supported by a previous study that found a relationship between
629 subjective arousal and pupil diameter in response to sound clips, but not to visual scenes
630 (Burley et al., 2017).

631 High-arousal images (erotic, aggression, or mutilation scenes) also provoke pupil
632 dilation, whereas natural images are associated with the strongest pupil constriction (Bradley
633 et al., 2017). Therefore, the difference in pupillary responses to non-vegetated and vegetated
634 scenes may be explained by differences in arousal level.

635

636 **4.5 Pleasantness judgments of traffic noises and audiovisual interactions were** 637 **reflected in pupillary activity**

638

639 Importantly, we found that pleasantness differences were associated with pupil
640 responses not only when comparing the global categories of traffic noise and bird song, but
641 also when contrasting samples of traffic noise differing in unpleasantness. Highly-unpleasant
642 traffic noises reduced the pleasantness of the visual scenes to a greater extent than less-
643 unpleasant traffic noises. In the same manner, greater pupil dilations were associated with the
644 highly-unpleasant samples rather than the less-unpleasant traffic noise samples, irrespective
645 of whether a visual or auditory scene was judged. Perceptual audiovisual interactions in
646 pleasantness ratings were thus associated with corresponding pupillary reactions.

647 However, this greater audiovisual modulation of pupillary reactions by the subset of
648 highly-unpleasant traffic noise samples (**Fig. 6, B1-2**) may also be explained by variations in
649 loudness, specifically the intensity of the sounds. Indeed, pupillary dilation in response to
650 auditory stimulation is positively correlated with both subjective salience and loudness (Liao
651 et al., 2016).

652 As to the subcategories of the contrasting sounds presented in our study, the mean
653 sound pressure level of the more-unpleasant traffic noise sounds was higher than that of the
654 less-unpleasant ones (60.8 dB(A) vs. 51.7 dB(A) respectively), and the rank order correlation
655 between the sound pressure levels of the road traffic noises and their corresponding average
656 auditory unpleasantness ratings was highly significant ($r=0.78$, $p<0.001$, $n=18$). However, for
657 the road traffic pass-by noises used in our study the annoyance ratings previously reported
658 when sound pressure level was equalized at 60 dB(A) (Gille et al., 2016) did not correlate

659 significantly with the pleasantness ratings collected during this current study. This might be
660 explained by the important role of loudness in both (un)pleasantness ratings and the
661 physiological arousal related to the auditory salience of sounds as reflected in pupillary
662 reactions (Boswijk et al., 2020).

663

664 **4.6 Integrative vision of behavioral and psychophysiological measures of audiovisual** 665 **interaction**

666

667 To summarize, we believe that the main driving mechanism for the lower pleasantness ratings
668 produced by the stimuli from the less-pleasant category may be an increase in arousal level
669 due to their salience, which is reflected by pupil dilation. Indeed, there is evidence that both
670 the pleasantness and restorative potential of natural sounds correlate with lower arousal
671 (Ratcliffe et al., 2020). In contrast, the emotional valence, which was reflected in the EMG
672 activity, affected only the auditory modulation of pleasantness ratings. The longer fixation
673 durations for vegetated scenes may be vision-specific manifestations of lower arousal, which
674 is supported by an increase in parasympathetic heart rate variability in a number of studies
675 (Mygind et al., 2021). One critical limitation of our study is that the crucial manipulation of the
676 valence category — both in images and sounds — may have been partially confounded with
677 low-level perceptual factors like the loudness of the sounds and local variations in the
678 brightness of the images. However, since in real life these confounds appear to be tightly
679 bound to the stimuli, the current results highlight the advantages of using eye-tracking
680 technology to record pupil size and gaze behavior to better understand the mechanisms
681 involved in the audiovisual interactions observed during the appraisal of our urban
682 environment.

683

684

685

686

687 **Conclusions**

688

689 The current work aimed to contribute to a better understanding of the mechanisms
690 involved in the perceptual audiovisual interaction phenomenon that is known to occur during
691 visual and auditory perception of urban environments. In this investigation, we proposed a
692 combined psychophysical and physiological methodology to assess both auditory and visual
693 perceptions of urban environments, and interactions between these two modalities. In this
694 methodology, different types of physiological measures (electromyography of the eyebrow
695 muscle, heart rate, pupil size, and gaze behavior) were considered to help us understand the
696 mechanisms. Under controlled conditions, 61 participants, mainly students, were asked to
697 assess the auditory or visual pleasantness of briefly presented excerpts from urban
698 environments that varied in the degree to which natural features (visual or sound) were
699 included. We showed that both auditory and visual pleasantness judgments are subject to
700 perceptual audiovisual interactions that strongly reflect the 'naturalness' of the exposure. The
701 current study highlighted several physiological indicators to be associated with the perceptual
702 audiovisual interactions observed during the appraisal of urban environments. While
703 electromyography and heart rate measures were primarily related to auditory pleasantness,
704 the gaze behavior measures were linked to visual pleasantness. Compared with road traffic
705 sounds, the natural sounds resulted in lower eyebrow electromyographic activity and heart
706 rate, while natural images were associated with slower gaze behavior. To the contrary, pupil
707 area reacted to both auditory and visual conditions, with pupil responses discriminating
708 between pleasant and unpleasant sound and visual environments. The pupil was relatively
709 constricted by all the natural stimuli, and was found to be sensitive to subtle variations in
710 unpleasantness between different traffic noise samples and the modulation of visual
711 pleasantness by these sounds.

712 The different physiological measures employed provide evidence that the complex
713 mechanisms involved in perceptual audiovisual interactions are largely driven by arousal level
714 due to the salience of the audiovisual stimulation. In particular, this was well reflected by pupil

715 size: pupil size measures constituted the most reliable physiological measures of the
716 perceptual audiovisual interaction during pleasantness assessment. The audiovisual
717 integration of emotional elements from urban environments might reflect integrative top-down
718 cognitive processes. Improvement in the auditory and visual presence of nature in the city can
719 have a positive impact on the wellbeing and health of city dwellers by reducing arousal and
720 stress.

721 Future perspectives to this work would be to consider participant groups with a higher
722 mean age and standard deviation to have a more representative sample, and to investigate
723 the mechanisms involved in the audiovisual interactions observed during the appraisal of our
724 urban environment with the use of eye-tracking technology to record pupil size and gaze
725 behavior.

726

727 **Acknowledgements**

728

729 This work was funded by the LABEX CELYA (ANR-10-LABX-0060) and the LABEX IMU
730 (ANR-10-LABX-0088) of Université de Lyon, within the program "Investissements d'Avenir"
731 operated by the French National Research Agency (ANR).

732

733 **References**

734

735 Babisch, W., Beule, B., Schust, M., Kersten, N., & Ising, H. (2005). Traffic Noise and Risk of
736 Myocardial Infarction. *Epidemiology*, *16*(1), 33–40.

737 <https://doi.org/10.1097/01.ede.0000147104.84424.24>

738 Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs.

739 *Behavior Research Methods*, *37*(3), 379–384. <https://doi.org/10.3758/BF03192707>

740 Berto, R., Massaccesi, S., & Pasini, M. (2008). Do eye movements measured across high
741 and low fascination photographs differ? Addressing Kaplan’s fascination hypothesis.

742 *Journal of Environmental Psychology*, *28*(2), 185–191.

743 <https://doi.org/10.1016/j.jenvp.2007.11.004>

744 Boswijk, V., Loerts, H., & Hilton, N. H. (2020). Saliency is in the eye of the beholder:

745 Increased pupil size reflects acoustically salient variables. *Ampersand*, *7*, 100061.

746 <https://doi.org/10.1016/j.amper.2020.100061>

747 Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli.

748 *Psychophysiology*, *37*(2), 204–215. <https://doi.org/10.1111/1469-8986.3720204>

749 Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of
750 emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

751 <https://doi.org/10.1111/j.1469-8986.2008.00654.x>

752 Bradley, M. M., Sapigao, R. G., & Lang, P. J. (2017). Sympathetic ANS modulation of pupil
753 diameter in emotional scene perception: Effects of hedonic content, brightness, and

754 contrast. *Psychophysiology*, *54*(10), 1419–1435. <https://doi.org/10.1111/psyp.12890>

755 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

756 Burley, D. T., Gray, N. S., & Snowden, R. J. (2017). As Far as the Eye Can See:

757 Relationship between Psychopathic Traits and Pupil Response to Affective Stimuli.

758 *PLoS ONE*, *12*(1). <https://doi.org/10.1371/journal.pone.0167436>

759 Cassina, L., Fredianelli, L., Menichini, I., Chiari, C., & Licitra, G. (2018). Audio-Visual
760 Preferences and Tranquillity Ratings in Urban Areas. *Environments*, 5(1), 1.
761 <https://doi.org/10.3390/environments5010001>

762 Cerwén, G., Pedersen, E., & Pálsdóttir, A. M. (2016). The Role of Soundscape in Nature-
763 Based Rehabilitation: A Patient Perspective. *International Journal of Environmental*
764 *Research and Public Health*, 13(12). <https://doi.org/10.3390/ijerph13121229>

765 Clark, C., Head, J., & Stansfeld, S. A. (2013). Longitudinal effects of aircraft noise exposure
766 on children's health and cognition: A six-year follow-up of the UK RANCH cohort.
767 *Journal of Environmental Psychology*, 35, 1–9.
768 <https://doi.org/10.1016/j.jenvp.2013.03.002>

769 Cottet, M., Vaudor, L., Tronchère, H., Roux-Michollet, D., Augendre, M., & Brault, V. (2018).
770 Using gaze behavior to gain insights into the impacts of naturalness on city dwellers'
771 perceptions and valuation of a landscape. *Journal of Environmental Psychology*, 60,
772 9–20. <https://doi.org/10.1016/j.jenvp.2018.09.001>

773 DiCiccio, T. J., & Romano, J. P. (1988). Discussion: Theoretical Comparison of Bootstrap
774 Confidence Intervals. *The Annals of Statistics*, 16(3), 965–969.
775 <https://doi.org/10.1214/aos/1176350938>

776 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
777 power analysis program for the social, behavioral, and biomedical sciences. *Behavior*
778 *Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>

779 Franco, L. S., Shanahan, D. F., & Fuller, R. A. (2017). A Review of the Benefits of Nature
780 Experiences: More Than Meets the Eye. *International Journal of Environmental*
781 *Research and Public Health*, 14(8), 864. <https://doi.org/10.3390/ijerph14080864>

782 Frederickson, N., & Dunsmuir, S. (2009). *Measures of Children's Mental Health and*
783 *Psychological Wellbeing: A Portfolio for Education and Health Professionals*. GL
784 Assessment Ltd. The Chiswick Centre. 414 Chiswick High Road. London W45TF.
785 UK.

786 Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for Human Electromyographic
787 Research. *Psychophysiology*, 23(5), 567–589. [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-8986.1986.tb00676.x)
788 [8986.1986.tb00676.x](https://doi.org/10.1111/j.1469-8986.1986.tb00676.x)

789 Gao, C., Weber, C. E., Wedell, D. H., & Shinkareva, S. V. (2020). An fMRI Study of Affective
790 Congruence across Visual and Auditory Modalities. *Journal of Cognitive*
791 *Neuroscience*, 32(7), 1251–1262. https://doi.org/10.1162/jocn_a_01553

792 Gao, C., Wedell, D. H., Kim, J., Weber, C. E., & Shinkareva, S. V. (2017). Modelling
793 audiovisual integration of affect from videos and music. *Cognition and Emotion*,
794 32(3), 516–529. <https://doi.org/10.1080/02699931.2017.1320979>

795 Gille, L.-A., Marquis-Favre, C., & Klein, A. (2016). Noise Annoyance Due To Urban Road
796 Traffic with Powered-Two-Wheelers: Quiet Periods, Order and Number of Vehicles.
797 *Acta Acustica United with Acustica*, 102(3), 474–487.
798 <https://doi.org/10.3813/AAA.918966>

799 Golland, Y., Hakim, A., Aloni, T., Schaefer, S., & Levit-Binnun, N. (2018). Affect dynamics of
800 facial EMG during continuous emotional experiences. *Biological Psychology*, 139,
801 47–58. <https://doi.org/10.1016/j.biopsycho.2018.10.003>

802 Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological*
803 *Review*, 77(5), 419–450. <https://doi.org/10.1037/h0029810>

804 Irwin, A., Hall, D. A., Peters, A., & Plack, C. J. (2011). Listening to urban soundscapes:
805 Physiological validity of perceptual dimensions. *Psychophysiology*, 48(2), 258–268.
806 <https://doi.org/10.1111/j.1469-8986.2010.01051.x>

807 Jeon, J. Y., Lee, P. J., You, J., & Kang, J. (2012). Acoustical characteristics of water sounds
808 for soundscape enhancement in urban open spaces. *The Journal of the Acoustical*
809 *Society of America*, 131(3), 2101–2109. <https://doi.org/10.1121/1.3681938>

810 Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework.
811 *Journal of Environmental Psychology*, 15(3), 169–182. [https://doi.org/10.1016/0272-](https://doi.org/10.1016/0272-4944(95)90001-2)
812 [4944\(95\)90001-2](https://doi.org/10.1016/0272-4944(95)90001-2)

813 Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code.
814 *Behavior Research Methods*, 51(3), 1336–1342. <https://doi.org/10.3758/s13428-018->
815 1075-y

816 Lavandier, C., & Defréville, B. (2006). The Contribution of Sound Source Characteristics in
817 the Assessment of Urban Soundscapes. *Acta Acustica United with Acustica*, 92(6),
818 912–921.

819 Li, H., & Lau, S.-K. (2020). A review of audio-visual interaction on soundscape assessment
820 in urban built environments. *Applied Acoustics*, 166, 107372.
821 <https://doi.org/10.1016/j.apacoust.2020.107372>

822 Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences
823 among pupillary dilation response, subjective salience of sounds, and loudness.
824 *Psychonomic Bulletin & Review*, 23(2), 412–425. <https://doi.org/10.3758/s13423-015->
825 0898-0

826 Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A. M., de
827 Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M. J., Lupp, G.,
828 Richardson, E. A., Astell-Burt, T., Dimitrova, D., Feng, X., Sadeh, M., Standl, M.,
829 Heinrich, J., & Fuertes, E. (2017). Exploring pathways linking greenspace to health:
830 Theoretical and methodological guidance. *Environmental Research*, 158, 301–317.
831 <https://doi.org/10.1016/j.envres.2017.06.028>

832 Miedema, H. M., & Oudshoorn, C. G. (2001). Annoyance from transportation noise:
833 Relationships with exposure metrics DNL and DENL and their confidence intervals.
834 *Environmental Health Perspectives*, 109(4), 409–416.

835 Morel, J., Marquis-Favre, C., Dubois, D., & Pierrette, M. (2012). Road Traffic in Urban Areas:
836 A Perceptual and Cognitive Typology of Pass-By Noises. *Acta Acustica United with*
837 *Acustica*, 98(1), 166–178. <https://doi.org/10.3813/AAA.918502>

838 Muzet, A. (2007). Environmental noise, sleep and health. *Sleep Medicine Reviews*, 11(2),
839 135–142. <https://doi.org/10.1016/j.smr.2006.09.001>

840 Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., Stevenson, M. P., Quintana, D. S., &
841 Bentsen, P. (2021). Effects of Public Green Space on Acute Psychophysiological
842 Stress Response: A Systematic Review and Meta-Analysis of the Experimental and
843 Quasi-Experimental Evidence. *Environment and Behavior*, 53(2), 184–226.
844 <https://doi.org/10.1177/0013916519873376>

845 Oliva, M., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion recognition in
846 human vocalizations. *Scientific Reports*, 8(1), 4871. [https://doi.org/10.1038/s41598-](https://doi.org/10.1038/s41598-018-23265-x)
847 [018-23265-x](https://doi.org/10.1038/s41598-018-23265-x)

848 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming
849 numbers into movies. *Spatial Vision*, 10(4), 437–442.

850 Petri, D., Licitra, G., Vigotti, M. A., & Fredianelli, L. (2021). Effects of Exposure to Road,
851 Railway, Airport and Recreational Noise on Blood Pressure and Hypertension.
852 *International Journal of Environmental Research and Public Health*, 18(17), 9145.
853 <https://doi.org/10.3390/ijerph18179145>

854 Preis, A., Hafke, H., Szychowska, M., & Felcyn, J. (2016). Audio-visual interaction of
855 environmental noise. *Noise Control Engineering Journal*, 64(1), 34–43.

856 *R Wave Detection in the ECG - MATLAB & Simulink*. (n.d.). Retrieved June 9, 2020, from
857 <https://www.mathworks.com/help/wavelet/ug/r-wave-detection-in-the-ecg.html>

858 Ratcliffe, E., Gatersleben, B., & Sowden, P. T. (2020). Predicting the Perceived Restorative
859 Potential of Bird Sounds Through Acoustics and Aesthetics. *Environment and*
860 *Behavior*, 52(4), 371–400. <https://doi.org/10.1177/0013916518806952>

861 Ren, X., Jian, K., PeiSheng, Z., & ShiYuan, W. (2018). Effects of soundscape on rural
862 landscape evaluations. *Environmental Impact Assessment Review*, 70, 45–56.

863 Ren, X., & Kang, J. (2015). Interactions between landscape elements and tranquility
864 evaluation based on eye tracking experiments (L). *The Journal of the Acoustical*
865 *Society of America*, 138(5), 3019–3022. <https://doi.org/10.1121/1.4934955>

866 Singh, D., Kumari, N., & Sharma, P. (2018). A Review of Adverse Effects of Road Traffic
867 Noise on Human Health. *Fluctuation and Noise Letters*, 17(01), 1830001.
868 <https://doi.org/10.1142/S021947751830001X>

869 Sun, K., De Coensel, B., Echevarria Sanchez, G. M., Van Renterghem, T., & Botteldooren,
870 D. (2018). Effect of interaction between attention focusing capability and visual
871 factors on road traffic noise annoyance. *Applied Acoustics*, 134, 16–24.
872 <https://doi.org/10.1016/j.apacoust.2018.01.001>

873 Sundarasekar, R., Thanjaivadivel, M., Manogaran, G., Kumar, P. M., Varatharajan, R.,
874 Chilamkurti, N., & Hsu, C.-H. (2018). Internet of Things with Maximal Overlap
875 Discrete Wavelet Transform for Remote Health Monitoring of Abnormal ECG Signals.
876 *Journal of Medical Systems*, 42(11), 228. <https://doi.org/10.1007/s10916-018-1093-4>

877 Van Renterghem, T., & Botteldooren, D. (2016). View on outdoor vegetation reduces noise
878 annoyance for dwellers near busy roads. *Landscape and Urban Planning*, 148, 203–
879 215. <https://doi.org/10.1016/j.landurbplan.2015.12.018>

880 Viollon, S., Lavandier, C., & Drake, C. (2002). Influence of visual setting on sound ratings in
881 an urban environment. *Applied Acoustics*, 63(5), 493–511.
882 [https://doi.org/10.1016/S0003-682X\(01\)00053-6](https://doi.org/10.1016/S0003-682X(01)00053-6)

883 Ward Thompson, C., Roe, J., Aspinall, P., Mitchell, R., Clow, A., & Miller, D. (2012). More
884 green space is linked to less stress in deprived communities: Evidence from salivary
885 cortisol patterns. *Landscape and Urban Planning*, 105(3), 221–229.
886 <https://doi.org/10.1016/j.landurbplan.2011.12.015>

887 World Health Organization. Regional Office for Europe. (2011). *Burden of disease from*
888 *environmental noise: Quantification of healthy life years lost in Europe*. World Health
889 Organization. Regional Office for Europe.
890 <https://apps.who.int/iris/handle/10665/326424>

891
892

893 **Figure captions**

894

895 **Figure 1.** Methods. (A1) Visual depictions of non-vegetated (top, $n=18$) and vegetated
896 (bottom, $n=18$) urban scenes. (A2) Audiovisual combinations resulting in four
897 experimental conditions. (A3) The laboratory setup showing the eye-tracking system,
898 electrophysiological system, and loudspeakers. (B) The procedure and its timeline
899 (B1-B4). Depending on the group, reminders for rating the pleasantness of either the
900 auditory environment (Auditory task) or the visual environment (Visual task) were
901 shown above the visual analogue scale (B3).

902

903 **Figure 2.** Mean pleasantness scores and standard errors. The effects of the Sound
904 category factor (2 levels: bird song (B) and traffic noise (T)), and the effects of the
905 Image category factor (2 levels: vegetated scenes (V) and non-vegetated scenes (nV))
906 on pleasantness scores during the auditory and visual tasks. *** $p<0.001$. Significant
907 perceptual audiovisual interactions are highlighted with red dotted contours.

908

909 **Figure 3.** Time course of physiological reaction changes due to variation in nature
910 presence within sound or image category, i.e. bird song vs. traffic noise (Auditory effect
911 [B-T]) or vegetated scenes vs. non-vegetated scenes (Visual effect [V-nV]). (A) Time
912 course of median pupil area changes, contrasting the more hedonic with the less
913 hedonic condition. (B) Time course of median corrugator supercillii (EMG) activity
914 changes. (C) Time course of median heart rate changes (RR intervals). 95% CI:
915 bootstrap 95%-confidence intervals.

916

917 **Figure 4.** Behavior and eye-tracking measure changes due to variation in nature
918 presence within sound or image category, i.e. bird song vs. traffic noise (Auditory effect
919 [B-T]) or vegetated scenes vs. non-vegetated scenes (Visual effect [V-nV]), during the
920 auditory task (Aud. task) and the visual task (Vis. task).

921 (A) Pleasantness score difference in favor of the a-priori more hedonic category, i.e.
922 bird song or vegetated scenes. (B) Fixation duration difference. (C) Cumulative
923 saccade distance difference. Individual values (ind. vals.), median values, and
924 bootstrap 95% CIs are shown. *** $p<0.001$, ** $p<0.01$, * $p<0.05$. Significant perceptual
925 audiovisual interactions are highlighted with red dotted contours.

926

927

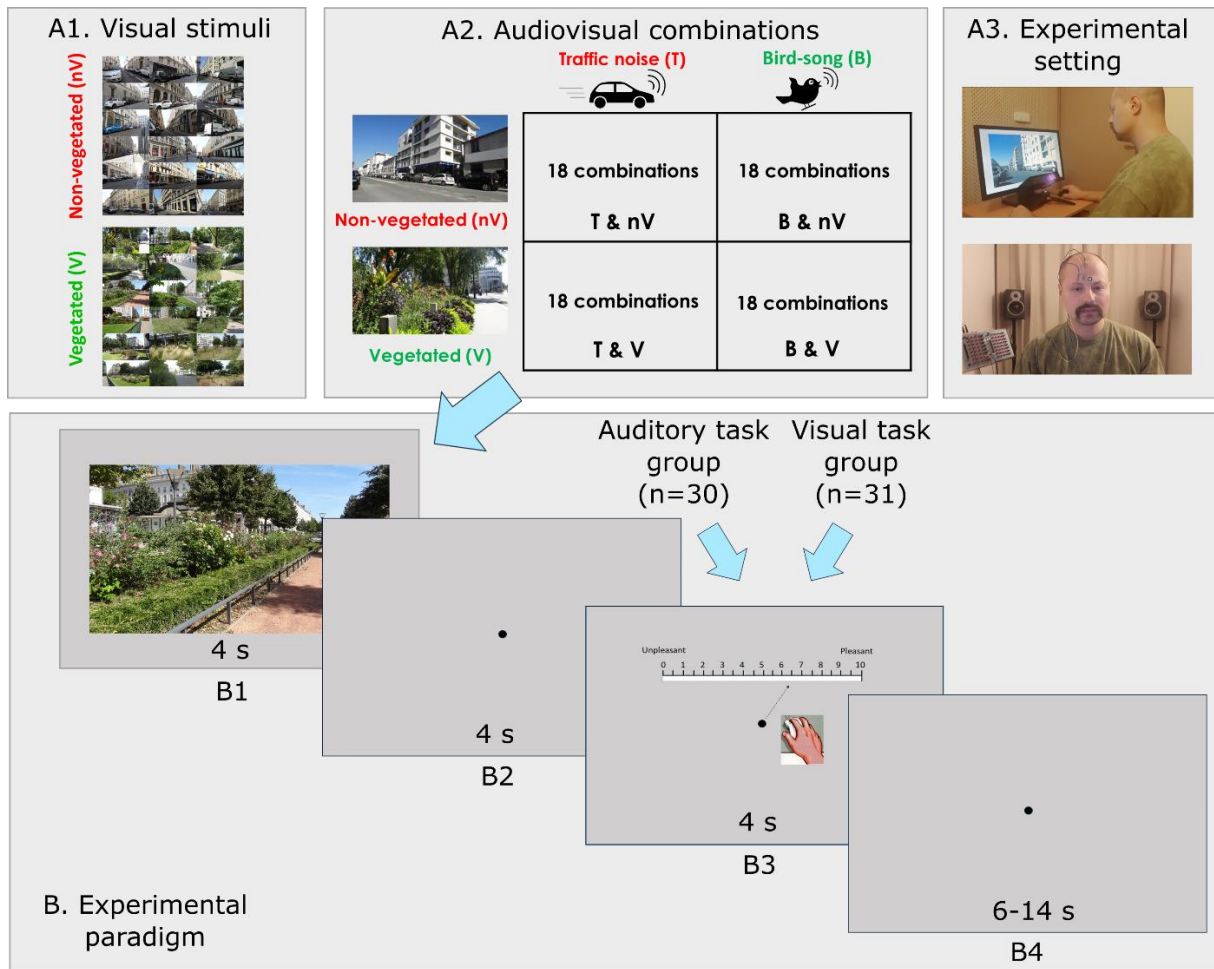
928 **Figure 5.** Physiological measure changes due to variation in nature presence within
929 sound or image category, i.e. bird song vs. traffic noise (Auditory effect [B-T]) or
930 vegetated scenes vs. non-vegetated scenes (Visual effect [V-nV]), during the auditory
931 task (Aud. task) and the visual task (Vis. task).

932 (A) Pupil area difference. (B) Corrugator supercillii EMG difference. (C) Heart rate
933 difference. Individual values (ind. vals.), median values, and bootstrap 95% CIs are
934 shown. *** $p<0.001$, ** $p<0.01$, * $p<0.05$. Significant perceptual audiovisual
935 interactions are highlighted with red dotted contours.

936

937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957

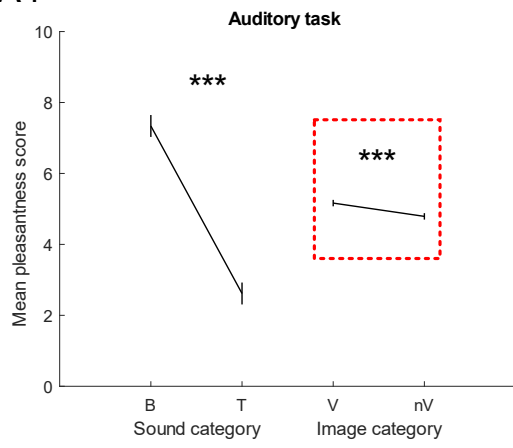
Figure 6. Effects of the five most-unpleasant versus the five least-unpleasant traffic noise samples on perceptual audiovisual interaction and pupillary reactivity. (A1) Highly- and Less-unpleasant urban road traffic noise (HT and LT, respectively) in terms of the median pleasantness rating differences relative to bird song in the auditory task (n=30). The 18 labels for traffic noise stimuli on the abscissa of A1 (XYZ) refer to previous studies employing these sounds (Gille et al., 2016; Morel et al., 2012): X stands for “vehicle type” (b: bus; d: motorcycle/moped; p: heavy vehicle; v: light vehicle), Y for “driving condition” (a = acceleration; d = deceleration; f = constant speed), Z for “road morphology” (o = open street; u = U-shaped street). (A2) Effects of Highly-unpleasant Traffic noise (HT) and Less-unpleasant Traffic noise (LT) on median visual pleasantness rating differences relative to bird song in the visual task (n=31). (B1) Difference in the time course of pupil dilation contrasting low- and highly-unpleasant traffic noise (i.e. traffic noise effect [LT-HT]), with both visual and auditory experiments pooled (n=61). (B2) Effects of variations in unpleasant traffic noise (i.e. traffic noise effect [LT-HT]) on pupil area difference during the auditory and the visual tasks. *** p<0.001, ** p<0.01, * p<0.05. Individual values (ind. vals.), median values, and bootstrap 95% CIs are shown. Significant perceptual audiovisual interactions are highlighted with red dotted contours.



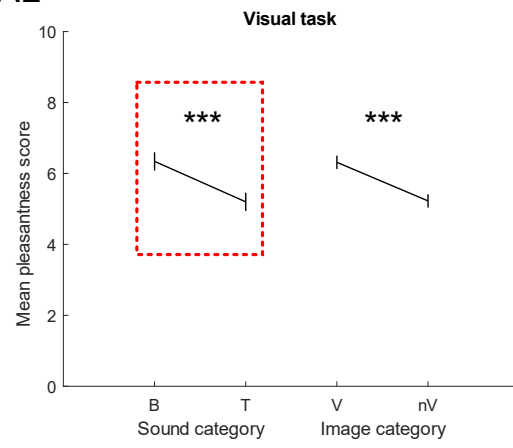
958

959 **Figure 1**

A1

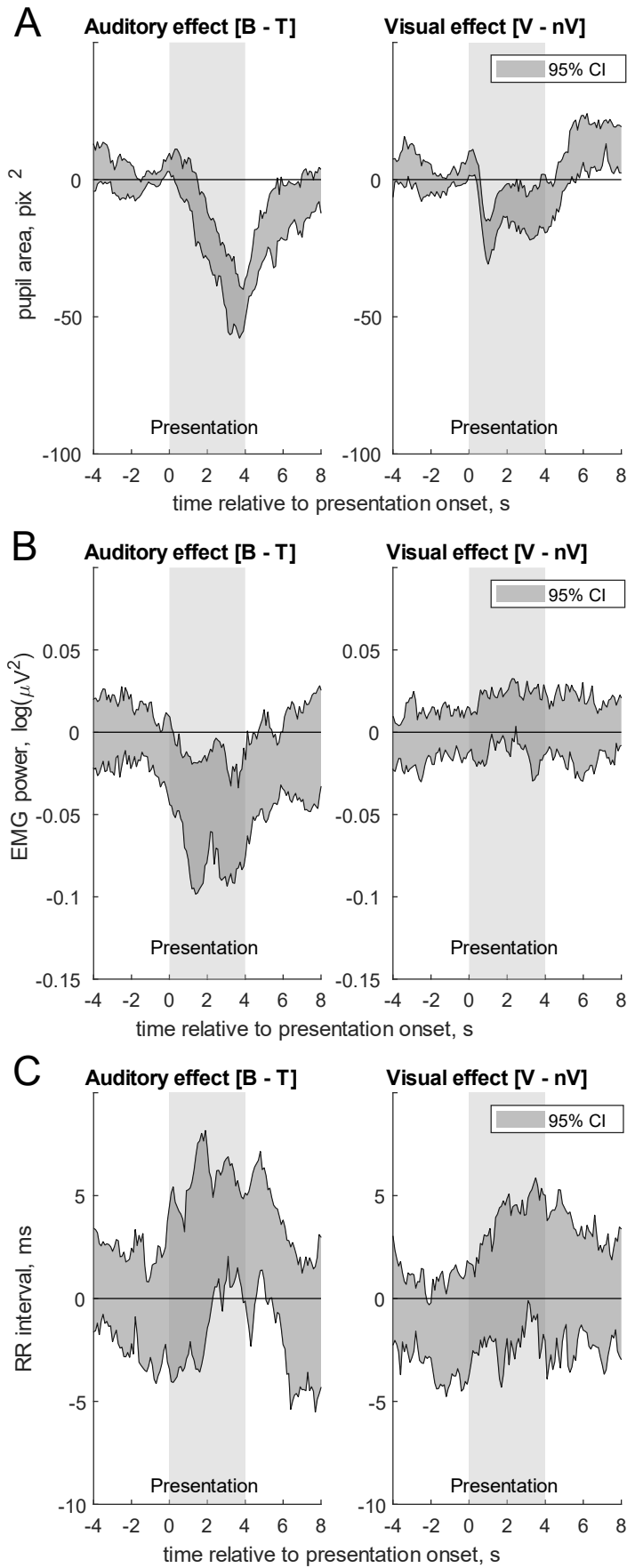


A2



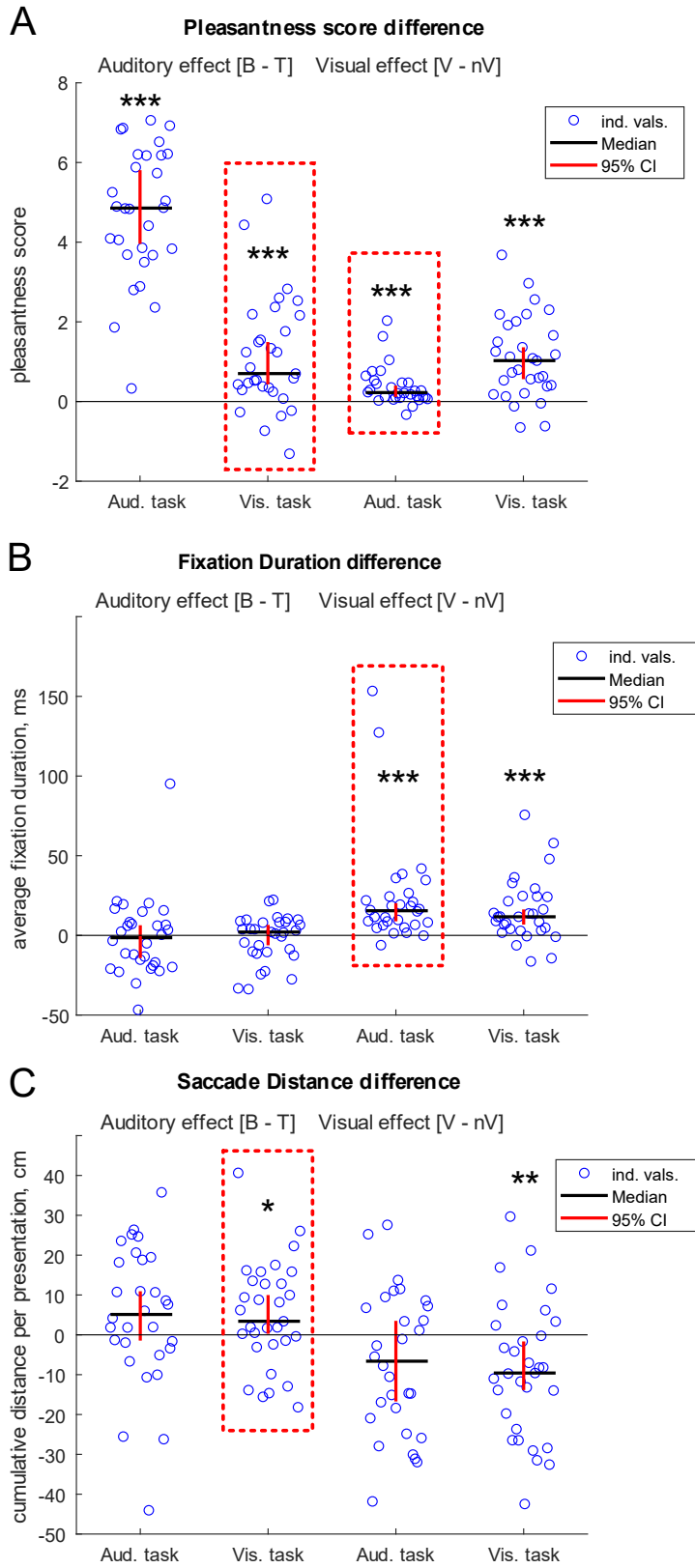
960
961

Figure 2

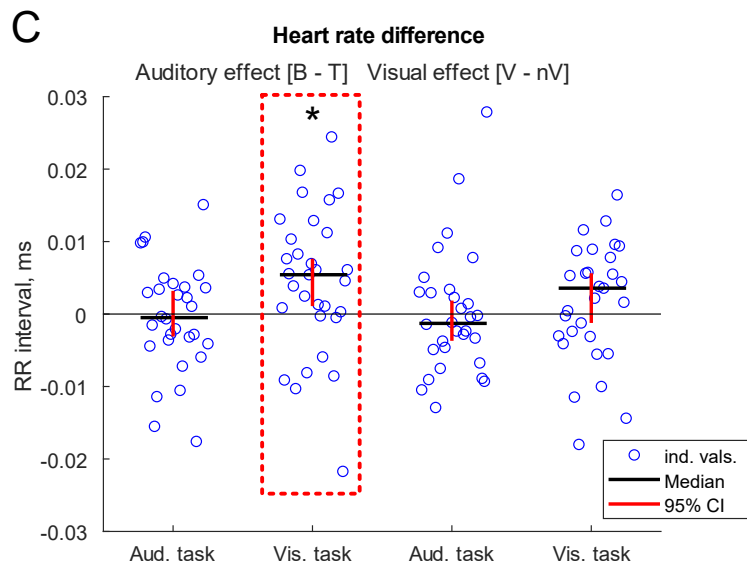
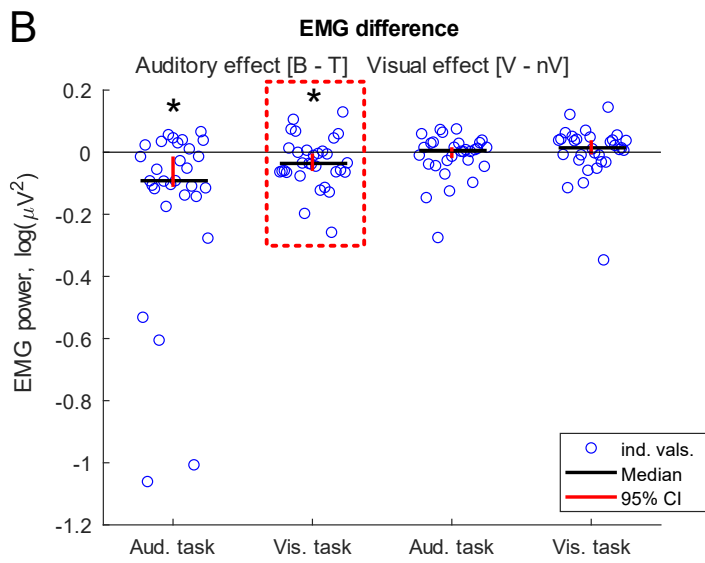
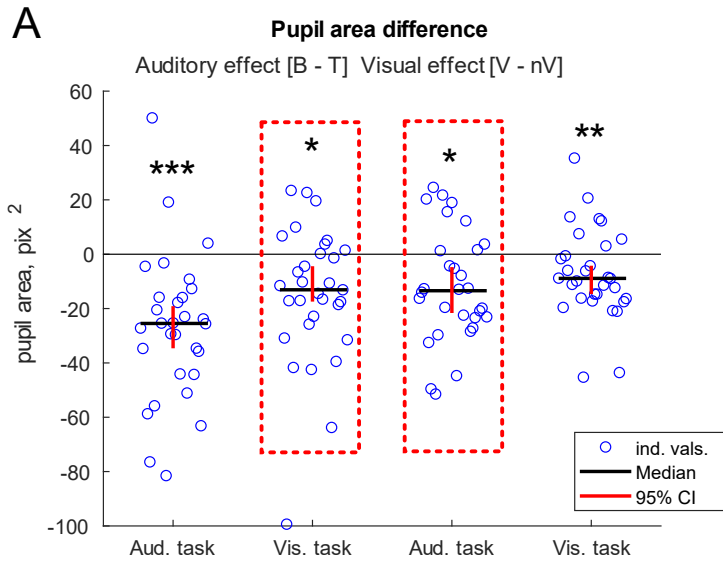


962
963

Figure 3

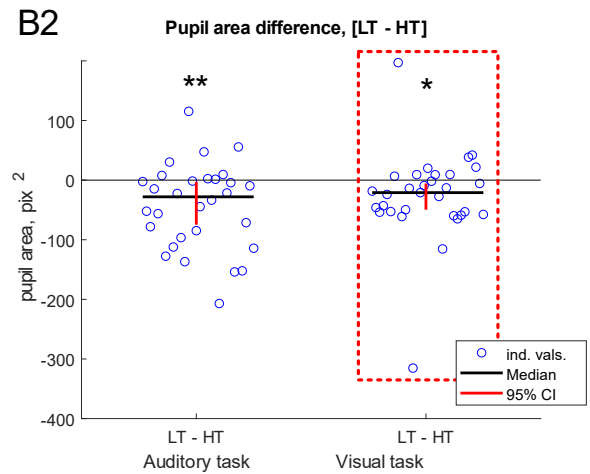
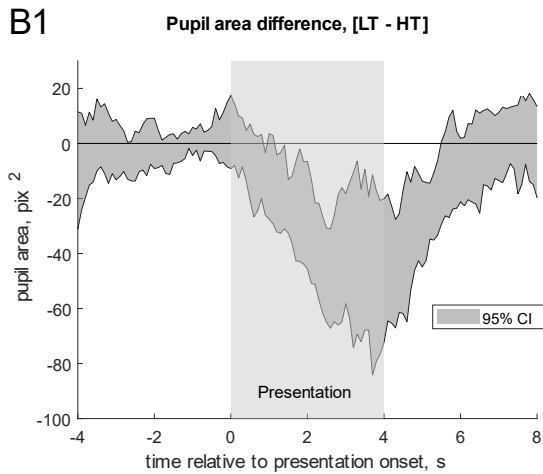
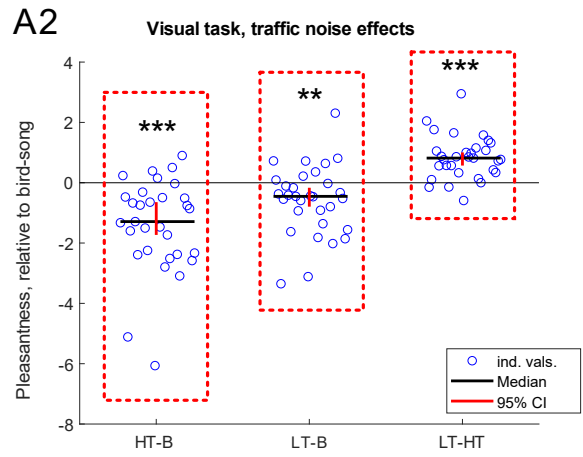
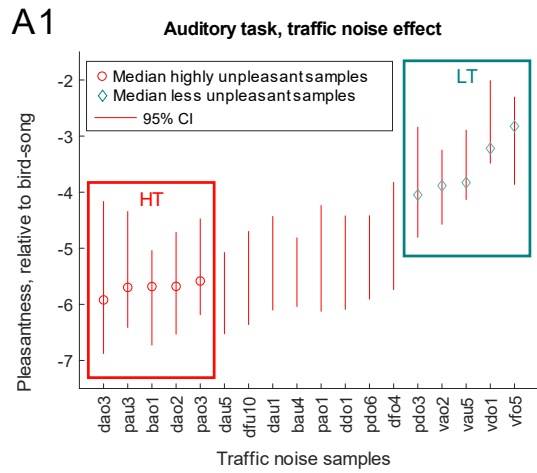


964
965 **Figure 4**



966
967

Figure 5



968
969

Figure 6