

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-03924495v2

HAL Id: hal-03924495 https://hal.science/hal-03924495v2

Submitted on 14 Nov 2024

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 <u>Catherine.MARQUISFAVRE@entpe.fr</u>
12	COTTET Marylise, University of Lyon, CNRS, ENS de Lyon, Lyon, France <u>marylise.cottet@ens-</u>
13	<u>lyon.fr</u>
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 <u>ellermeier@psychologie.tu-darmstadt.de</u>

- 20 Abstract
- 21

22 The presence of natural elements in the urban environment improves wellbeing. Comparisons 23 of different psychophysiological reactions to both auditory and visual perceptions of urban 24 environments are employed as a method for shedding light on the mechanisms involved in 25 audiovisual interaction, by which visual context influences auditory perception, and vice versa. 26 Combining psychophysical and physiological methodologies is likely to improve our 27 understanding of the perceptual consequences of variations in the presence of nature in the 28 city. In the present study, 61 participants, mainly students, were exposed to urban 29 soundscapes accompanied by street scenes. Audio excerpts of bird song or road traffic were 30 paired with images with or without vegetation. Half of the participants (n=30) were asked to 31 judge the auditory pleasantness after each audiovisual presentation, whereas the other half 32 (n=31) judged the visual pleasantness of the scene. Heart rate, pupil area, gaze fixations and 33 electromyography of the eyebrow muscle (corrugator supercilii) were concurrently recorded.

34 We demonstrated a strong perceptual audiovisual interaction: auditory pleasantness was 35 higher when accompanied by a natural visual context, and visual pleasantness was higher 36 when accompanied by natural sounds. Electromyography and heart rate were particularly 37 sensitive to the presence of nature in the soundscapes, whereas gaze behavior discriminated 38 only visual nature elements. The natural sounds were associated with lower heart rate and 39 less corrugator supercilii electromyographic activity, whereas natural images were associated 40 with slower gaze behavior. Pupil area was sensitive to both auditory and visual stimulations. 41 The pupil was relatively constricted by all natural stimuli and discriminated different unpleasant 42 traffic noise samples and the modulation of visual pleasantness by these sounds. This 43 suggests a distinctive role of arousal level in the modulation of pupil reactions, which appear 44 well suited to study audiovisual interactions and the perceptual effects of natural elements in 45 urban environments.

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

47 **1. Introduction**

48

The densification of urban areas, activities generated by economic growth and 49 50 transportation, are some of the main driving forces for noise pollution, which is considered as 51 a major environmental health problem. Studies have highlighted various adverse impacts of 52 noise exposure on well-being and health, such as sleep disturbance (e.g., Muzet, 2007), 53 effects on children's cognition and health (e.g., Clark et al., 2013), annoyance (e.g., Miedema 54 & Oudshoorn, 2001), myocardial infarction (e.g., Babisch et al., 2005), diastolic blood pressure 55 and hypertension (e.g., Petri et al., 2021). Noise exposure incurs a financial burden on society 56 (e.g., World Health Organization. Regional Office for Europe, 2011). Urban areas are typically 57 exposed to a multitude of noise sources (e.g., road traffic noise, railway noise, road traffic 58 noise, recreational noise). Road traffic noise is one of the prominent threats to wellbeing in 59 cities, causing emotional stress, impeding work performance, and forming a risk factor for 60 sleep disturbance, cardiovascular diseases and hypertension (e.g., Singh et al., 2018). The 61 exposure of city dwellers to noise pollution is thus an important issue. The presence of nature, 62 whether visually perceived (e.g., the presence of plants, trees, or a river) or heard (e.g., bird 63 song, water lapping, wind in trees), is known to improve wellbeing (e.g., Franco et al., 2017). 64 Soundscape studies have shown that when natural sounds like bird song or water sounds are 65 combined with road traffic noise, they improve judgments of auditory pleasantness of urban 66 environments (Lavandier & Defréville, 2006, Jeon et al., 2012). Natural sounds increase the 67 perceived tranquility of urban areas (Cassina et al., 2018), and are thought to play an important 68 role in nature-based rehabilitation (Cerwén et al., 2016). From a visual perspective, looking at 69 vegetated scenes was demonstrated to bring important benefits, such as stress reduction and 70 better psychological restoration (Kaplan, 1995). For example, it was shown that green spaces 71 are favorable for stress reduction as indicated by a reduced cortisol level (Ward Thompson et 72 al., 2012). However, there is still much that remains to be learned about the specific pathways linking green spaces to health (Markevych et al., 2017). In particular, previous studies 73 74 indicated that audiovisual interactions (here defined as the perceptual phenomenon according

75 to which visual context influences auditory perception, and vice versa) were observed in 76 pleasantness judgments, especially when nature is present. Actually, it is known that the 77 presence of nature in the visual context affects auditory judgments (e.g., Preis et al., 2016). 78 For example, the auditory unpleasantness due to road traffic noise in urban areas was judged 79 to be about 20-30 % less when auditory stimuli were combined with a visual context of natural 80 scenes compared to the context of a scene with apartment blocks only (Viollon et al., 2002). 81 Similar findings were also highlighted in studies dealing with noise annoyance (e.g., Chau et 82 al., 2018). Outdoor vegetation visible from the living room window reduces noise annoyance 83 (e.g., Van Renterghem & Botteldooren, 2016). However, not all studies confirmed such 84 findings when noise annoyance was assessed in a visual context with or without vegetation 85 (Sun et al., 2018). Thus, the mechanisms linked to the impact of natural factors and 86 audiovisual interactions on wellbeing in the context of the urban environment are therefore far 87 from clear. Several psychological and physiological human responses deserve when exploring 88 these mechanisms. In particular it could be promising to compare how the changes in 89 parameters of audiovisual excerpts of urban environments cause variations in emotional 90 valence, arousal and attention. In this manner, it may be possible to identify the key processes 91 by which the presence of nature in the urban environment can modify human behavior. 92 Pupillary dilation is known to be sensitive to rapid arousal (Bradley et al., 2008), slowing of 93 heart rate is evidence of an experienced threat, and an increase in the activity of corrugator 94 supercilii face muscle has been interpreted as evidence of discrimination between negative 95 and positive emotional valence (Bradley & Lang, 2000). Gaze behavior reveals changes in 96 visual attention. A reduced number of gaze fixations and longer fixation durations may be 97 physical manifestations of altered attentional states that have been characterized as 98 involuntary sustained attention or fascination (Berto et al., 2008). Such parameters could 99 therefore be of interest for revealing the basic mechanisms of audiovisual interactions and for 100 helping us to understand the wellbeing effects of nature in urban environments.

101 The present study aimed to explore the mechanisms involved in perceptual audiovisual 102 interactions observed during the auditory and visual appraisal of urban environments in the

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

106 (1) Both the auditory and visual pleasantness of urban environments was studied while107 varying the degree of nature elements present in the audiovisual stimulation.

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

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

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

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

117

118 2. Method

119

120 2.1 Participants

121

Sixty-one healthy adults between 19 and 57 years-of-age (mean age 27.93, SD=10.33) 122 123 were recruited through advertisements sent via the institutional mailing list of the Graduate 124 School of Civil, Environmental and Urban Engineering (ENTPE – Lyon, France), where the 125 research experiment took place. This sample size corresponded to the number of participants 126 used in previous studies with a similar design (Berto et al., 2008; Bradley et al., 2008; Bradley 127 & Lang, 2000; Viollon et al., 2002). The critical sample size calculated for our study was 26 128 participants, and obtained with G*Power 3 (Faul et al., 2007) for an effect size (f) 0.33, power 129 of 0.80, p<0.05 and four measures in repeated measures ANOVA, assuming no overall 130 correlation between the measures.

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

- 138
- 139 2.2 Auditory and visual stimuli
- 140

141 The protocol used two categories of auditory stimuli: 18 urban road traffic noise samples 142 ranging from 48 to 62 dB(A) and 18 samples of bird song ranging from 40 to 45 dB(A), each 143 sample lasting for 4 seconds. All noise samples stemmed from recordings made using the 144 ORTF technique (devised at the former Office de Radiodiffusion Télévision Française), with 145 Schoeps MSTC 64 microphones equipped with Schoeps BBG windfields, and a Tascam 146 portable recorder with 24-bit resolution and a sampling frequency of 44.1 kHz (Morel et al., 147 2012). The road traffic noise samples consisted of vehicle pass-bys (light vehicles, buses, 148 heavy vehicles or motorcycles/mopeds) recorded by Morel and colleagues (Morel et al., 2012). 149 These samples were previously used in studies dealing with noise annoyance (Gille et al., 2016). The bird song samples were recorded on the ENTPE campus site (Vaulx-en-Velin, 150 151 France). A 300-Hz high-pass filter was applied to the bird song samples to remove low 152 frequency content resulting from the background noise of suburban Lyon and to insure good 153 sound reproduction quality in the quiet experimental room. For the sound reproduction, the 154 mean sound pressure levels of all sound stimuli were set to the levels of the noise sources 155 measured on site. A 200-ms amplitude fade-in and a 500-ms fade-out with a quadratic 156 progression kernel were applied to all 36 sound stimuli.

157 The experimental protocol also used two categories of visual stimuli: 18 images 158 representing vegetated scenes and another 18 images depicting non-vegetated scenes

(examples in Fig. 1). These visual stimuli were pictures of urban scenery in and around Lyon
(France) taken specifically for this experiment during the spring and summer seasons. The
camera used to take the pictures was a SONY© DSC-HX10V, with exposure of 1/400-1/1000
s, ISO 100, and automatic white balance.

163 There was no overall significant difference in eye illuminance between the vegetated 164 and non-vegetated images when displayed on the screen during the experiment (11.0 lux, SD 165 = 1.1, n = 18 and 10.9 lux, SD = 1.2, n = 18, respectively, p=0.81, t-test). The initial resolution 166 was 4896 x 2752 pixels, while the size of the images on the screen was 45.3 x 25.5 cm with 167 a resolution of 1632 x 918 pixels. All photographs were taken in the same manner: the 168 photographer stood at the intersection of two streets with their back turned towards the 169 intersection and the lens pointing in the opposite direction. This protocol was followed to avoid 170 visual and auditory incongruities. All photographs were taken in good weather in the city of 171 Lyon and its suburbs.

172

173 2.3 Laboratory setup

174

175 Participants were seated in front of a desk (Fig. 1, A3) supporting a computer monitor 176 and eye-tracking equipment (described in 2.6) inside a dark sound-proof room with a 177 background noise level of 18.5 dB(A). A computer mouse for making responses was available on the desk. The LCD monitor used (Lenovo ThinkVision LT2423WC, 24-inch FHD LED 178 179 Backlit Monitor, TN technology) had a resolution of 1920 x 1080 with a typical contrast ratio of 180 1000:1. The viewing distance was about 50 centimeters. Given that the width/height of the 181 display was 53/30 centimeters, it subtended approximately 30 degrees of visual angle 182 vertically and 50 degrees of visual angle horizontally. Sound stimuli were reproduced using 183 two active loudspeakers (Dynaudio BM5A Actif), and one subwoofer (Dynaudio Acoustics 184 BM9S) to ensure good reproduction of low frequencies of the road vehicle pass-by noise 185 samples. The loudspeakers were located behind the seated participants at a height of 1.50 186 with the subwoofer on the floor between them. The center of the participant's interaural axis

and the loudspeakers formed an equilateral triangle with sides measuring 1.5 m. Thus, the
sounds were reproduced in such a way that the participants perceived the vehicles to pass
behind their backs, and therefore the noise sources would have been outside of their field of
vision.

191

192 2.4 Experimental paradigm

193

194 The PsychToolbox application (Brainard, 1997; Pelli, 1997) in MATLAB was used for 195 implementing the experimental paradigm based on 2 categories of sounds (18 bird song and 196 18 road traffic noise samples) and 2 categories of visual scenes (18 vegetated and 18 non-197 vegetated urban pictures). To ensure that physiological measures were reliable, stimuli were 198 presented with short duration and repeated conditions (e.g., Bradley & Lang, 2008) using 199 different samples per condition to avoid habituation (e.g., Groves & Thompson, 1970). During 200 each audiovisual presentation, the participants were simultaneously exposed to one picture 201 and one sound for a duration of 4 s. Four paradigm conditions resulted from the 2-by-2 202 categories of sounds and images: bird song and vegetated image; bird song and non-203 vegetated image; traffic noise and vegetated image; traffic noise and non-vegetated image. 204 Each of the 36 visual stimuli was randomly presented twice, once paired with a bird song 205 sample and once with a traffic noise sample. Each of the 4 conditions was repeated using 18 206 trials with different image-sound combinations, giving a total of 72 audiovisual stimulations 207 (see Fig. 1, A1 and A2). Whether participants assessed the auditory or visual pleasantness 208 of the audiovisual combination was randomly determined before the start of the experiment 209 and constituted a between-subjects factor (Fig. 1, B).

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

After each presentation, a 4-s blank screen with a fixation dot was shown. Subsequently,
participants made their response by judging the auditory or visual pleasantness of the

215 presented urban environment (depending on the task required – cf. instructions in section 2.5). 216 They gave their rating using a bipolar continuous visual-analogue scale. The scale displayed 217 numerical labels from 0 to 10, with two opposite end points "unpleasant" and "pleasant", to 218 facilitate ratings along the continuum (cf. Fig. 1, B3), as typically used in studies recording 219 affective reactions (e.g., Bradley & Lang, 2000; Frederickson & Dunsmuir, 2009). Participants 220 were instructed to move the mouse cursor from the center of the screen to the location on the 221 scale reflecting the pleasantness felt. Placing the mouse on the line made a white spot appear 222 at that position, thus providing feedback on a successfully registered judgment. The mouse 223 cursor then had to be brought back to the center of the screen so that the next judgment 224 started from about the same hand position. Participants had 4 s to make their response. 225 Subsequently, a fixation dot was displayed for a random duration between 6 and 14 s before 226 the next trial commenced.

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

232

233 2.5 Judgment task instructions

234

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

Approximately half of the participants (n=30) were randomly assigned to the "auditory task". They received the following instructions in French: *«L'expérience vous propose des extraits audiovisuels d'environnements urbains. Imaginez-vous à chaque fois dans le lieu représenté par l'image. Si vous étiez dans ce lieu, comment jugeriez-vous l'environnement* **sonore** et uniquement l'environnement **sonore** ?», which translate as "During the experiment, *you will be exposed to audiovisual excerpts of the city environment. Imagine yourself each*

time at the place shown in the image. If you were at this place how would you judge the soundenvironment and only the sound environment?".

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

250

251 2.6 Eye-tracking

252

253 The gaze position and pupil area of both eyes were recorded using an Eyelink[®] Portable 254 Duo eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada) with a sampling frequency 255 of 1 kHz. This eye-tracker operated in remote mode as a head-free system with a target 256 sticker: a standard cue placed on the forehead of the participants and automatically identified 257 by the eye-tracking system. This setup allows recording of eye position without restraining the 258 head position. The test phase was preceded by a standard nine-point calibration procedure 259 for mapping gaze position to screen position. The targets used for calibration were the default 260 targets of the Eyelink® apparatus.

261

262 2.7 Eye-tracking preprocessing

263

Pupil area and gaze position data were analyzed using MATLAB. The algorithm for the preprocessing of pupil size data and automatic artifact rejection followed recommendations proposed in the Pupil Size Preprocessing Code (Kret and Sjak-Shie, 2019). This routine sets certain criteria for the rejection of samples associated with blinks. First dilation speed outliers and edge artifacts were identified as pupil size samples that were outside of a predefined feasible range. Second, measurements with abnormally large deviations (caused by eyeblinks) from a smooth trend line were removed. To detect dilation speed outliers, the median absolute deviation was used, this is a robust and outlier resilient data dispersion metric.
Finally, short-series eye-tracking samples separated from the rest of the recording by long
temporal intervals were removed before analyzing the data.

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

281

282 2.8 Electrophysiological recording

283

284 Electromyography (EMG) and electrocardiography (ECG) were continuously recorded 285 during the experiment using AgAgCI electrodes at a 1024-Hz sampling rate (16 bits) and the 286 MicroMed SD64 electrophysiological system (Micromed, Treviso, Italy). As recommended for 287 recording of the EMG activity of the corrugator supercilii muscle (Fridlund & Cacioppo, 1986), 288 the electrodes of a bipolar channel were placed over the right eye-brow (Fig. 1, A3). For ECG 289 recording, the electrodes of the other bipolar channel were placed on the left and right wrists 290 (or on the left ankle as a function of the detectability of the R peak, the most important peak 291 of the ECG). The ground electrode was placed on the left side of the forehead. In both cases, 292 the skin was cleaned, gently abraded with gel (NUPREP; Weaver and Co.) and for better 293 contact, a conductive adhesive gel (Tensive; Parker Laboratories) was placed under each 294 electrode. The electrode impedances were kept below 5 kOhm. During the acquisition a band-295 pass filtering of 0.1-8 Hz and 0.1-1 Hz was applied to the EMG and ECG channels 296 respectively, using the BRAIN QUICK recording system. The electrophysiological signal was 297 amplified 1000 times.

298

299 2.9 EMG preprocessing

300

301 The EMG signals were preprocessed in a similar way as in other recent studies examining the activity of the corrugator supercilii muscle during emotional experiences 302 303 (Golland et al., 2018). The EMG spectrogram was estimated with a short-time Fourier 304 transform. The Fourier-transform segment length was 100 ms, with adjoining segments 305 overlapping by 50 ms. For each segment, the sum of the natural logarithms of the power 306 spectral densities was computed for the frequencies 10 – 200 Hz, excluding the range of 45-307 55 Hz to suppress power line noise. Then, the time series were interpolated to the initial 308 sampling rate (1 kHz) with shape-preserving piecewise cubic interpolation.

309

310 2.10 ECG preprocessing

311

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

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

320

321 3. Results

322

323 3.1 Behavior

324

325 The individual behavioral responses were collected for the 4 experimental conditions in 326 auditory or visual tasks. The mean pleasantness scores for all the four conditions for either

327 task are shown in Figure 2, A1-A2. These individual scores were used in 2*2 repeated-328 measures analyses of variance (RM-ANOVAs), with the within-subject factor "Sound category" 329 (2 levels: bird song, denoted later by B, and road traffic noise, denoted by T) and the within-330 subject factor "Image category" (2 levels: vegetated scene, denoted by V, and non-vegetated 331 scene, denoted by nV). RM-ANOVAs were performed separately for the two tasks (i.e., judging 332 the pleasantness of the auditory or the visual environment, hereafter referred to as the 333 "auditory task" and "visual task", respectively). To estimate the effect size, partial eta squared η_p^2 was calculated, a parameter frequently reported for RM-ANOVA designs (Bakeman, 2005). 334

As expected, the auditory or visual environments that were more natural were judged to be more pleasant than those that were less natural. An effect of the main factor "Sound category" on the pleasantness score during the auditory task was unsurprisingly observed $(F(1,29) = 241.6, p < 0.001, partial eta squared \eta_p^2 = 0.89;$ Fig. 2, B1). The effect of the main factor "Image category" on visual pleasantness was also highly significant in visual task $(F(1,30) = 34.49, p < 0.001, \eta_p^2 = 0.53;$ Fig. 2, B2).

341 Strong perceptual audiovisual interaction was also observed in both tasks: the visual 342 context affected auditory ratings and vice versa. The effect of the main factor "Image category" on auditory pleasantness was highly significant (F(1,29) = 17.56, p<0.001, η_p^2 = 0.37), 343 demonstrating that pleasantness in the auditory task was increased by the presence of 344 345 vegetation in the depicted environment (see Fig. 2, B1 and the red dotted contour to highlight 346 the perceptual audiovisual interaction). The effect of the main factor "Sound category" on visual pleasantness was highly significant (F(1,30) = 20.33, p < 0.001, $\eta_p^2 = 0.40$), showing that 347 the natural auditory context, represented by bird song, made the visual scenery more pleasant 348 349 in the visual task (see Fig. 2, B2 and the red dotted contour to highlight the perceptual 350 audiovisual interaction). Both contextual effects provided evidence of perceptual audiovisual 351 interaction in both auditory and visual pleasantness judgments.

352

354

4 **3.2** Pupil area, EMG and heart rate

355

356 The following processing and analyses were applied to pupil, EMG and heart rate 357 physiological responses. First, for each participant, seventy-two epochs were extracted, each 358 spanning from 4 s prior to the onset of the 4-s audiovisual presentation until 4 s after its end. 359 A correction was applied to the 72 epochs by subtracting a baseline. For heart rate and EMG, 360 the baseline was the average of the physiological measurements recorded during the 10 s 361 before the stimulus onset. For pupil area, the baseline was the average of the physiological 362 measurements during the 2 s before stimulus onset. The statistical noise level was estimated 363 as a minimum to maximum range during the 20-s segment around presentation onset, and 364 the 14 noisiest epochs of each participant were excluded from further analysis. Additionally, 365 those pupil size epochs, in which more than 40% of samples were rejected because of blinks 366 in the 20-s segment around the presentation onset, were rejected. There were up to three 367 such trials in the recordings of the majority of the participants, whereas 8% of participants 368 (5/61) had more than seven trials rejected, with the maximum number of rejected trials per 369 participant being 27.

370 Second, the average epoch time-series from 4 s prior to the stimulus onset until 4 s after 371 its end, was down-sampled by a factor of 100 (i.e. to 10 Hz) for each of the 4 audiovisual 372 conditions and each of the 61 participants. When compared with the pleasant counterparts, 373 the unpleasant conditions, namely traffic noise or non-vegetated scenes provoked 374 pronounced pupil dilations and smaller EMG and heart rate effects. From these individually 375 averaged temporal physiological responses, the dynamic physiological response difference 376 was considered to highlight the potential change in physiological responses due to the 377 variation in the extent of nature within the sound and image categories of the audiovisual 378 stimulations (i.e., bird song vs road traffic noise, and vegetated images vs non-vegetated). To 379 illustrate the differences in dynamic physiological responses for pupil area, EMG and heart 380 rate, they were averaged over the participants of each group (i.e., for auditory and visual tasks, 381 respectively) and are presented in Figure 3, A-C. The effect of the variation in nature presence

within the sound category (i.e., bird song vs. road traffic noise) is referred to as the auditory
effect [B-T], while the effect of the variation in nature presence within the image category (i.e.,
vegetated vs non-vegetated images) is referred to as the visual effect [V-nV].

385 For either of the two groups, a median-based 1000-iteration-bootstrap bias-corrected 386 and accelerated percentile 95%-confidence interval (DiCiccio & Romano, 1988) was 387 estimated for each timepoint of the differential responses. Then, the average differential 388 response for every participant was computed within the window of interest of 4 s-audiovisual 389 stimulation. These individual scores were used in group-level summary statistics with the 390 median-based bootstrap technique and RM-ANOVAs. Furthermore, medians and bootstrap 391 group-level 95%-confidence intervals (CI) were also considered for the differential scores of 392 pleasantness, to simultaneously illustrate the significance of the effect of audiovisual 393 stimulation on both pleasantness judgments and physiological reactions. The Figures 4 and 394 5 display the median values and 95%-CIs to highlight the auditory effect [B-T] (i.e., Bird song 395 vs. Traffic noise) and the visual effect [V-nV] (i.e., Vegetated vs. non-Vegetated) on the 396 differential pleasantness scores (Fig. 4, A) and the differential physiological responses (Fig. 397 4, B to Fig. 5, C). In these figures, a significant effect can be considered to have occurred 398 when the confidence interval does not embrace the zero line.

399 RM-ANOVAs with the factors "sound" (i.e., Bird song vs. Traffic noise, [B-T]) and "image" 400 (i.e., Vegetated vs. non-Vegetated, [V-nV]) were performed for the auditory and visual tasks, 401 respectively, to give the magnitudes of significant effects according to η_p^2 .

The median and 95% CI results show that pupil area was higher when the participant was presented with traffic noise rather than birdsong, both in the auditory task and in the visual task. This is highlighted in **Figure 5**, **A** by the significant auditory effect [B-T] on pupil area during both the auditory and visual tasks. This effect is also observed in the RM-ANOVA results (F(1,29) = 30.6, p < .001, $\eta_p^2 = 0.51$ for the auditory effect [B-T] on pupil area during the auditory task, and F(1,30) = 10.8, p = .003, $\eta_p^2 = 0.27$ during the visual task). Pupil area was also higher after observing non-vegetated scenes than after observing vegetated scenes (**Fig.**

5, **A**). The visual effect [V-nV] was significant in the RM-ANOVAs, both when the sound environment was judged (F(1,29) = 10.0, p=.004, $\eta_p^2 = 0.26$) and when the visual environment was evaluated (F(1,30) = 5.9, p=.021, $\eta_p^2 = 0.16$).

412 The bootstrap method indicated that the corrugator supercilii muscle (EMG) appeared 413 to be more strongly activated during traffic noise than during bird song, in both the auditory 414 task and the visual task (Fig. 5, B). This significant auditory effect [B-T] is also observed in the RM-ANOVA results (F(1,29) = 8.6, p=.006, $\eta_p^2 = 0.23$ during the auditory task, and F(1,30)415 = 5.1, p=.031, η_p^2 = 0.15 during the visual task). In contrast, image content had no significant 416 417 differential effect on the EMG (Fig. 5, B). Heart rate showed a slight elevation only when 418 unpleasant sound content was presented in the visual task (Fig. 5, C), with the auditory effect [B-T] showing statistical significance in the RM-ANOVA (F(1,30) = 6.2, p=.019, $\eta_p^2 = 0.17$). 419

420

421 **3.3. Relationships between pupil reactions and behavior**

422

423 The results reported so far have shown that pupil dilations were obtained in response to 424 the less pleasant sound and image categories. Beyond this finding we wanted to explore 425 whether pupil size during exposure was linked to the subsequent behavioral judgments on 426 samples within sound and image categories. Differences in pleasantness and physiological 427 reactions within the image category will be presented in a forthcoming study dedicated to 428 visual perception. Our investigation therefore concerns the potential effect of variation in 429 auditory pleasantness on the pupil size within each sound category. As there was negligible 430 variation in pleasantness judgments for the different bird song samples stemming from the 431 Lyon area, we focused on differences in reactions to the urban road traffic noise samples (light 432 vehicle, bus, heavy vehicle, and motorcycle/moped pass-bys). To investigate the potential 433 effect on pupil size, two sub-conditions of traffic noise samples were considered: traffic noise 434 samples with high unpleasantness scores versus samples with low unpleasantness scores.

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

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

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

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

A similar pattern was observed for the pupil reactions. The difference in pupil dilation was considered for low versus high unpleasantness traffic noise exposure (auditory and visual tasks pooled, n=61). The time course of differences in pupil reaction exhibited a significant relative increase in pupil size for highly-unpleasant traffic noise samples compared with lessunpleasant ones (**Fig. 6, B1**).

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

- 468
- 469 **3.4 Gaze fixations and saccade statistics**
- 470

471 Seventy-two epochs were extracted from the vertical and horizontal gaze position data 472 ranging from the onset to the end of the 4-s presentation of the audiovisual content. Epochs 473 were rejected if either less than 12% of data points per epoch were available or more than 474 40% of samples were rejected because of blinks in the 20-s segment around the presentation 475 onset. An instantaneous velocity threshold of 0.6 pixel (20°/s) was applied to identify 476 saccades. The remaining episodes were labelled as fixations. The average position and 477 duration were identified for each fixation episode. If the positions of consecutive fixations did 478 not differ by more than 25 pixels (0.8°) we considered them to belong to the same fixation. 479 The duration of such fixations was computed from the onset of the first to the end of the last 480 episode belonging to it. For each epoch, the average fixation duration and the cumulative 481 distance of the saccades, separating the fixations were computed. Then, the average scores 482 were computed for each of the 4 conditions for each participant, except for the epochs 483 excluded because of the noisiness of the pupil data. Following the same rationale as the one 484 described in section 3.2, individual response differences were considered in the group-level 485 summary statistics using the median-based bootstrap technique and RM-ANOVAs to analyze 486 the auditory effect [B-T] (i.e., Bird song vs. Traffic noise) and the visual effect [V-nV] (i.e., 487 Vegetated vs. non-Vegetated) on fixation duration and cumulative saccade distance.

488 The analysis of eve-tracking data vielded a set of significant effects that depended on 489 the audiovisual stimulation. Fixation duration was strongly affected by the image category of 490 the visual scenes, being significantly longer for the vegetated scenes in both tasks (Fig. 4, B). 491 The RM-ANOVA with the factors "sound" (i.e., Bird song vs. Traffic noise, [B-T]) and "image" 492 (i.e., Vegetated vs. non-Vegetated, [V-nV]) also revealed the significant visual effect [V-nV] on fixation duration during the auditory task (F(1,29) = 14.4, p < 0.001, $\eta_p^2 = 0.33$) and the visual 493 task (F(1,30) = 19.0, p<.001, η_p^2 = 0.39). Consistent with the longer fixation durations, the 494 495 cumulative saccade distance was shorter for vegetated scenes (Fig. 4, C), though the 496 bootstrapping technique confirmed this effect only during the visual task. This corresponded 497 to the RM-ANOVA result, with a significant visual effect [V-nV] on the cumulative saccade distance (F(1,30) = 8.8, p=0.006, $\eta_p^2 = 0.23$). A significant result was observed for the auditory 498 effect during the visual task: bird song significantly increased gaze travel distance (F(1,30) =499 4.3, p=0.047, $\eta_p^2 = 0.13$). 500

501

502 4. Discussion

503

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

510

511 4.1 Behavioral findings

512

513 Significant perceptual audiovisual interactions were demonstrated in both auditory and 514 visual judgment tasks. The presence of nature (bird song or vegetation) increased both the

auditory and visual pleasantness of the excerpts of urban environments presented in ourstudy.

517 First, we confirmed a previously reported modification of auditory pleasantness by visual 518 context (Viollon et al., 2002), with sound pleasantness being judged as greater when 519 presented in the context of vegetated scenes as opposed to non-vegetated scenes. We further 520 showed that visual pleasantness judgments were influenced by the auditory context. The 521 visual pleasantness of the depicted urban scenes was judged to be greater when they were 522 accompanied by bird song rather than road traffic noise. This result is in agreement with 523 previous findings that assessment of landscapes was negatively affected by man-made 524 sounds such as traffic noise in comparison with natural sounds such as bird song (Ren et al., 525 2018). Moreover, we found that in visual pleasantness judgments, the modulation by the 526 auditory context varied depending on the degree of unpleasantness elicited by the road traffic 527 noise samples. The depicted urban scenes were judged to be less visually pleasant when 528 accompanied by more unpleasant (compared with less unpleasant) urban road vehicle pass-529 by noise.

530 A similar phenomenon was studied in the interaction of videos and music, and was 531 termed the "affective congruency" effect. It refers to the enhancement of valence judgments 532 in combined audiovisual exposures in comparison with auditory or visual stimuli of the same 533 valence (either positive or negative) presented separately (Gao et al., 2017). This means that 534 pleasant sounds combined with pleasant images produce a stronger impression of overall 535 pleasantness. The same holds for negative valences (Gao et al., 2017). This was also shown 536 by Cassina et al. (2018) in a study dealing with urban soundscapes. They highlighted the 537 combined effect of negative visual elements and negative sound sources (e.g., road traffic) on 538 the perceived tranquility of urban environments, with tranquility being known to be closely 539 related to pleasantness (Li & Lau, 2020). However, the affective congruency effect highlighted 540 by Gao et al. (2017) was shown for positive vs. negative valence stimuli only if they were of 541 high arousal. There is evidence that the audiovisual integration of emotional stimuli yielding 542 similar results to ours is associated with specific activity in the medial prefrontal cortex (Gao

et al., 2020). This might reflect integrative top-down cognitive processes during audiovisual integration. Another indication of the largely cognitive nature of this effect is the finding that the degree of matching between visual and sound features can alter this relationship. Unlike bird song or traffic noise, pleasantness judgments of human-generated sounds like speech or footsteps do not appear to be sensitive to the visual context (Viollon et al., 2002). This reservation might be explicitly tested in further studies.

549 We also showed that a number of physiological and gaze-behavior effects accompanied 550 these modulations of pleasantness judgments: pupil size, EMG, heart rate, saccadic distance 551 and fixation duration were all affected by the studied conditions. Physiological reactions and 552 gaze behavior were typically significant in both the auditory and visual tasks, or were not 553 significant in either one. On the other hand, the sound category (bird song or traffic noise) and 554 the image category (vegetated or not) provoked significant effects in only some of the objective 555 measures. In particular, the average fixation duration and cumulative saccade length were 556 significantly affected by visual content but not significantly by the auditory content in either 557 task (Fig. 4, B-C), whereas EMG and heart rate responded significantly to the auditory 558 component but not to the visual one (Fig. 5, B-C). The pupillary response changed significantly 559 as a function of both the visual and auditory content of urban environments (Fig. 5, A) and 560 therefore constituted the most reliable physiological measure of the perceptual audiovisual 561 interaction during pleasantness assessment.

562

563 **4.2 Gaze behavior discriminated vegetated from non-vegetated urban scenes**

564

In our study, when participants observed urban scenes containing vegetation, their average fixation duration was longer and their saccadic distances were shorter in comparison with observation of urban scenes without vegetation. This is consistent with previous findings on gaze fixation during observation of natural scenes of high or low fascination (Berto et al., 2008). However, this difference should not be attributed exclusively to the presence of vegetation. It is known that even among different vegetated scenes, fixations are longest for 571 scenes evaluated as having the highest landscape quality (Cottet et al., 2018), which might 572 closely correspond to the pleasantness ratings captured in our study. Therefore, longer 573 fixations in our study may be associated with the potential of more-pleasant vegetated visual 574 environments to gently attract effortless attention. According to Attention Restoration Theory, 575 such changes improve wellbeing and facilitate recovery of attentional resources (Kaplan, 576 1995).

577 At the same time, the more-pleasant natural sounds did not affect gaze behavior in the 578 same way as the vegetated scenes. To the contrary, bird song facilitated saccades, which 579 may be evidence for a redistribution of visual attention as a function of the auditory context, 580 as has been reported by Ren & Kang (2015). Modification of the distribution of spatial visual 581 attention by the auditory context requires further exploration.

582

583 **4.3 Corrugator supercilii EMG and heart rate changes significantly discriminated** 584 **between traffic noise and bird song**

585

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

592 The corrugator supercilii EMG differed as a function of sound category, but was not 593 modified by the image category. This may be explained by the much smaller difference in 594 pleasantness across the image categories of the urban scenes presented (vegetated or not) 595 in comparison with the sound categories (bird song or traffic noise).

596 Heart rate acceleration in response to road traffic noise was observed only during the 597 visual task, thus reflecting implicit discomfort, elicited by the non-attended modality. However, 598 because of the moderate statistical significance of the heart rate data in the present study

(p<0.05), this finding requires further confirmation. In at least one study, explicit pleasantness was shown to be more strongly associated with changes in functional magnetic resonance imaging-measured cerebral activity in the amygdala than it was with heart rate responses (Irwin et al., 2011). This suggests that heart rate responses may be a less sensitive indicator of pleasantness than *corrugator supercilii* EMG.

604

4.4 The pupil reacted with greater dilation to both unpleasant auditory and visualstimulation

607

608 Pupil size increased significantly in response to both traffic noise and non-vegetated 609 images in comparison with their categorial counterparts, regardless of the task, be it visual or 610 auditory (Fig. 5, A). Compared with the image categories, the sound categories produced 611 larger pupillary response effects. However, it is conceivable that the pupillary response is 612 primarily accounted for by the difference in arousal mediated by upregulation of sympathetic 613 autonomic activity, rather than by the variation in valence (Bradley et al., 2008). Greater pupil 614 size is known to be associated with both highly negative and positive complex sounds, 615 indicating the role of arousal in regulation of this physiological variable (Oliva & Anikin, 2018). 616 Thus, the sound category effect on pupil size is very likely explained by arousal, potentiated 617 by the fact that the traffic noise samples had higher sound pressure levels than the bird song 618 samples. The finding that the arousal picked up by the pupillary dilation was more evident in 619 sound discrimination is supported by a previous study that found a relationship between 620 subjective arousal and pupil diameter in response to sound clips, but not to visual scenes 621 (Burley et al., 2017).

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

626

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

629

630 Importantly, we found that pleasantness differences were associated with pupil 631 responses not only when comparing the global categories of traffic noise and bird song, but 632 also when contrasting samples of traffic noise differing in unpleasantness. Highly-unpleasant 633 traffic noises reduced the pleasantness of the visual scenes to a greater extent than less-634 unpleasant traffic noises. In the same manner, greater pupil dilations were associated with the 635 highly-unpleasant samples rather than the less-unpleasant traffic noise samples, irrespective 636 of whether a visual or auditory scene was judged. Perceptual audiovisual interactions in 637 pleasantness ratings were thus associated with corresponding pupillary reactions.

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

643 As to the subcategories of the contrasting sounds presented in our study, the mean 644 sound pressure level of the more-unpleasant traffic noise sounds was higher than that of the 645 less-unpleasant ones (60.8 dB(A) vs. 51.7 dB(A) respectively), and the rank order correlation 646 between the sound pressure levels of the road traffic noises and their corresponding average 647 auditory unpleasantness ratings was highly significant (r=0.78, p<0.001, n=18). However, for 648 the road traffic pass-by noises used in our study the annoyance ratings previously reported 649 when sound pressure level was equalized at 60 dB(A) (Gille et al., 2016) did not correlate 650 significantly with the pleasantness ratings collected during this current study. This might be 651 explained by the important role of loudness in both (un)pleasantness ratings and the 652 physiological arousal related to the auditory salience of sounds as reflected in pupillary 653 reactions (Boswijk et al., 2020).

4.6 Integrative vision of behavioral and psychophysiological measures of audiovisual interaction

657

658 To summarize, we believe that the main driving mechanism for the lower pleasantness 659 ratings produced by the stimuli from the less-pleasant category may be an increase in arousal 660 level due to their salience, which is reflected by pupil dilation. Indeed, there is evidence that 661 both the pleasantness and restorative potential of natural sounds correlate with lower arousal 662 (Ratcliffe et al., 2020). In contrast, the emotional valence, which was reflected in the EMG 663 activity, affected only the auditory modulation of pleasantness ratings. The longer fixation 664 durations for vegetated scenes may be vision-specific manifestations of lower arousal, which is supported by an increase in parasympathetic heart rate variability in a number of studies 665 666 (Mygind et al., 2021). One critical limitation of our study is that the crucial manipulation of the 667 valence category — both in images and sounds — may have been partially confounded with 668 low-level perceptual factors like the loudness of the sounds and local variations in the 669 brightness of the images. However, since in real life these confounds appear to be tightly 670 bound to the stimuli, the current results highlight the advantages of using eye-tracking 671 technology to record pupil size and gaze behavior to better understand the mechanisms 672 involved in the audiovisual interactions observed during the appraisal of our urban 673 environment.

674

675 Conclusions

676

The current work aimed to contribute to a better understanding of the mechanisms involved in the perceptual audiovisual interaction phenomenon that is known to occur during visual and auditory perception of urban environments. In this investigation, we proposed a combined psychophysical and physiological methodology to assess both auditory and visual perceptions of urban environments, and interactions between these two modalities. In this methodology, different types of physiological measures (electromyography of the eyebrow

683 muscle, heart rate, pupil size, and gaze behavior) were considered to help us understand the 684 mechanisms. Under controlled conditions, 61 participants, mainly students, were asked to 685 assess the auditory or visual pleasantness of briefly presented excerpts from urban 686 environments that varied in the degree to which natural features (visual or sound) were 687 included. We showed that both auditory and visual pleasantness judgments are subject to 688 perceptual audiovisual interactions that strongly reflect the 'naturalness' of the exposure. The 689 current study highlighted several physiological indicators to be associated with the perceptual 690 audiovisual interactions observed during the appraisal of urban environments. While 691 electromyography and heart rate measures were primarily related to auditory pleasantness, 692 the gaze behavior measures were linked to visual pleasantness. Compared with road traffic 693 sounds, the natural sounds resulted in lower eyebrow electromyographic activity and heart 694 rate, while natural images were associated with slower gaze behavior. To the contrary, pupil 695 area reacted to both auditory and visual conditions, with pupil responses discriminating 696 between pleasant and unpleasant sound and visual environments. The pupil was relatively 697 constricted by all the natural stimuli, and was found to be sensitive to subtle variations in 698 unpleasantness between different traffic noise samples and the modulation of visual 699 pleasantness by these sounds.

700 The different physiological measures employed provide evidence that the complex 701 mechanisms involved in perceptual audiovisual interactions are largely driven by arousal level 702 due to the salience of the audiovisual stimulation. In particular, this was well reflected by pupil 703 size: pupil size measures constituted the most reliable physiological measures of the 704 perceptual audiovisual interaction during pleasantness assessment. The audiovisual 705 integration of emotional elements from urban environments might reflect integrative top-down 706 cognitive processes. Improvement in the auditory and visual presence of nature in the city can 707 have a positive impact on the wellbeing and health of city dwellers by reducing arousal and 708 stress.

Future perspectives to this work would be to consider participant groups with a higher mean age and standard deviation to have a more representative sample, and to investigate

the mechanisms involved in the audiovisual interactions observed during the appraisal of our
urban environment with the use of eye-tracking technology to record pupil size and gaze
behavior.

715 Acknowledgements

716

717 This work was funded by the LABEX CELYA (ANR-10-LABX-0060) and the LABEX IMU

718 (ANR-10-LABX-0088) of Université de Lyon, within the program "Investissements d'Avenir"

719 operated by the French National Research Agency (ANR).

721 References

- 722
- Babisch, W., Beule, B., Schust, M., Kersten, N., & Ising, H. (2005). Traffic Noise and Risk of
 Myocardial Infarction. *Epidemiology*, *16*(1), 33–40.
- 725 https://doi.org/10.1097/01.ede.0000147104.84424.24
- 726 Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs.

727 Behavior Research Methods, 37(3), 379–384. https://doi.org/10.3758/BF03192707

728 Berto, R., Massaccesi, S., & Pasini, M. (2008). Do eye movements measured across high

and low fascination photographs differ? Addressing Kaplan's fascination hypothesis.

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

- 731 https://doi.org/10.1016/j.jenvp.2007.11.004
- Boswijk, V., Loerts, H., & Hilton, N. H. (2020). Salience is in the eye of the beholder:
- Increased pupil size reflects acoustically salient variables. *Ampersand*, *7*, 100061.
 https://doi.org/10.1016/j.amper.2020.100061

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

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

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

739 https://doi.org/10.1111/j.1469-8986.2008.00654.x

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

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

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

- Burley, D. T., Gray, N. S., & Snowden, R. J. (2017). As Far as the Eye Can See:
- 745 Relationship between Psychopathic Traits and Pupil Response to Affective Stimuli.
- 746 PLoS ONE, 12(1). https://doi.org/10.1371/journal.pone.0167436

- 747 Cassina, L., Fredianelli, L., Menichini, I., Chiari, C., & Licitra, G. (2018). Audio-Visual
- Preferences and Tranquillity Ratings in Urban Areas. *Environments*, *5*(1), 1.
 https://doi.org/10.3390/environments5010001
- 750 Cerwén, G., Pedersen, E., & Pálsdóttir, A. M. (2016). The Role of Soundscape in Nature-
- Based Rehabilitation: A Patient Perspective. International Journal of Environmental
 Research and Public Health, 13(12). https://doi.org/10.3390/ijerph13121229
- 753 Clark, C., Head, J., & Stansfeld, S. A. (2013). Longitudinal effects of aircraft noise exposure
- on children's health and cognition: A six-year follow-up of the UK RANCH cohort.
- Journal of Environmental Psychology, 35, 1–9.
- 756 https://doi.org/10.1016/j.jenvp.2013.03.002
- 757 Cottet, M., Vaudor, L., Tronchère, H., Roux-Michollet, D., Augendre, M., & Brault, V. (2018).
- 758 Using gaze behavior to gain insights into the impacts of naturalness on city dwellers'
- perceptions and valuation of a landscape. *Journal of Environmental Psychology*, 60,
- 760 9–20. https://doi.org/10.1016/j.jenvp.2018.09.001
- 761 DiCiccio, T. J., & Romano, J. P. (1988). Discussion: Theoretical Comparison of Bootstrap
- 762 Confidence Intervals. *The Annals of Statistics*, *16*(3), 965–969.
- 763 https://doi.org/10.1214/aos/1176350938
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
 power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146
- Franco, L. S., Shanahan, D. F., & Fuller, R. A. (2017). A Review of the Benefits of Nature
 Experiences: More Than Meets the Eye. *International Journal of Environmental*
- 769 Research and Public Health, 14(8), 864. https://doi.org/10.3390/ijerph14080864
- 770 Frederickson, N., & Dunsmuir, S. (2009). *Measures of Children's Mental Health and*
- 771 Psychological Wellbeing: A Portfolio for Education and Health Professionals. GL
- Assessment Ltd. The Chiswick Centre. 414 Chiswick High Road. London W45TF.
- 773 UK.

- Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for Human Electromyographic
 Research. *Psychophysiology*, *23*(5), 567–589. https://doi.org/10.1111/j.14698986.1986.tb00676.x
- Gao, C., Weber, C. E., Wedell, D. H., & Shinkareva, S. V. (2020). An fMRI Study of Affective
- 778 Congruence across Visual and Auditory Modalities. Journal of Cognitive

779 *Neuroscience*, *32*(7), 1251–1262. https://doi.org/10.1162/jocn_a_01553

- Gao, C., Wedell, D. H., Kim, J., Weber, C. E., & Shinkareva, S. V. (2017). Modelling
- 781 audiovisual integration of affect from videos and music. *Cognition and Emotion*,

782 32(3), 516–529. https://doi.org/10.1080/02699931.2017.1320979

- Gille, L.-A., Marquis-Favre, C., & Klein, A. (2016). Noise Annoyance Due To Urban Road
- 784 Traffic with Powered-Two-Wheelers: Quiet Periods, Order and Number of Vehicles.

785 Acta Acustica United with Acustica, 102(3), 474–487.

- 786 https://doi.org/10.3813/AAA.918966
- Golland, Y., Hakim, A., Aloni, T., Schaefer, S., & Levit-Binnun, N. (2018). Affect dynamics of
 facial EMG during continuous emotional experiences. *Biological Psychology*, *139*,

789 47–58. https://doi.org/10.1016/j.biopsycho.2018.10.003

- Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review*, 77(5), 419–450. https://doi.org/10.1037/h0029810
- 792 Irwin, A., Hall, D. A., Peters, A., & Plack, C. J. (2011). Listening to urban soundscapes:
- Physiological validity of perceptual dimensions. *Psychophysiology*, *48*(2), 258–268.
 https://doi.org/10.1111/j.1469-8986.2010.01051.x
- Jeon, J. Y., Lee, P. J., You, J., & Kang, J. (2012). Acoustical characteristics of water sounds
- for soundscape enhancement in urban open spaces. The Journal of the Acoustical
- 797 Society of America, 131(3), 2101–2109. https://doi.org/10.1121/1.3681938
- Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework.
- 799 Journal of Environmental Psychology, 15(3), 169–182. https://doi.org/10.1016/0272-
- 800 4944(95)90001-2

- Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*(3), 1336–1342. https://doi.org/10.3758/s13428-0181075-y
- Lavandier, C., & Defréville, B. (2006). The Contribution of Sound Source Characteristics in
 the Assessment of Urban Soundscapes. *Acta Acustica United with Acustica*, *92*(6),
 912–921.
- Li, H., & Lau, S.-K. (2020). A review of audio-visual interaction on soundscape assessment
 in urban built environments. *Applied Acoustics*, *166*, 107372.

809 https://doi.org/10.1016/j.apacoust.2020.107372

- Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences
- 811 among pupillary dilation response, subjective salience of sounds, and loudness.
- 812 Psychonomic Bulletin & Review, 23(2), 412–425. https://doi.org/10.3758/s13423-015813 0898-0
- Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A. M., de

815 Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M. J., Lupp, G.,

- 816 Richardson, E. A., Astell-Burt, T., Dimitrova, D., Feng, X., Sadeh, M., Standl, M.,
- 817 Heinrich, J., & Fuertes, E. (2017). Exploring pathways linking greenspace to health:
- 818 Theoretical and methodological guidance. *Environmental Research*, *158*, 301–317.
- 819 https://doi.org/10.1016/j.envres.2017.06.028

820 Miedema, H. M., & Oudshoorn, C. G. (2001). Annoyance from transportation noise:

Relationships with exposure metrics DNL and DENL and their confidence intervals. *Environmental Health Perspectives*, *109*(4), 409–416.

Morel, J., Marquis-Favre, C., Dubois, D., & Pierrette, M. (2012). Road Traffic in Urban Areas:

- A Perceptual and Cognitive Typology of Pass-By Noises. *Acta Acustica United with*
- 825 *Acustica*, *98*(1), 166–178. https://doi.org/10.3813/AAA.918502
- 826 Muzet, A. (2007). Environmental noise, sleep and health. Sleep Medicine Reviews, 11(2),
- 827 135–142. https://doi.org/10.1016/j.smrv.2006.09.001

- 828 Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., Stevenson, M. P., Quintana, D. S., &
- 829 Bentsen, P. (2021). Effects of Public Green Space on Acute Psychophysiological
- 830 Stress Response: A Systematic Review and Meta-Analysis of the Experimental and
- 831 Quasi-Experimental Evidence. *Environment and Behavior*, 53(2), 184–226.
- 832 https://doi.org/10.1177/0013916519873376
- Oliva, M., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion recognition in
 human vocalizations. *Scientific Reports*, *8*(1), 4871. https://doi.org/10.1038/s41598018-23265-x
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming
 numbers into movies. *Spatial Vision*, *10*(4), 437–442.
- 838 Petri, D., Licitra, G., Vigotti, M. A., & Fredianelli, L. (2021). Effects of Exposure to Road,
- 839 Railway, Airport and Recreational Noise on Blood Pressure and Hypertension.
- 840 International Journal of Environmental Research and Public Health, 18(17), 9145.
 841 https://doi.org/10.3390/ijerph18179145
- Preis, A., Hafke, H., Szychowska, M., & Felcyn, J. (2016). Audio-visual interaction of
 environmental noise. *Noise Control Engineering Journal*, *64*(1), 34–43.
- 844 *R Wave Detection in the ECG MATLAB* & *Simulink*. (n.d.). Retrieved June 9, 2020, from
- 845 https://www.mathworks.com/help/wavelet/ug/r-wave-detection-in-the-ecg.html
- 846 Ratcliffe, E., Gatersleben, B., & Sowden, P. T. (2020). Predicting the Perceived Restorative
- Potential of Bird Sounds Through Acoustics and Aesthetics. *Environment and Behavior*, 52(4), 371–400. https://doi.org/10.1177/0013916518806952
- 849 Ren, X., Jian, K., PeiSheng, Z., & ShiYuan, W. (2018). Effects of soundscape on rural
- 850 landscape evaluations. *Environmental Impact Assessment Review*, 70, 45–56.
- 851 Ren, X., & Kang, J. (2015). Interactions between landscape elements and tranquility
- 852 evaluation based on eye tracking experiments (L). *The Journal of the Acoustical*
- 853 Society of America, 138(5), 3019–3022. https://doi.org/10.1121/1.4934955

- Singh, D., Kumari, N., & Sharma, P. (2018). A Review of Adverse Effects of Road Traffic
 Noise on Human Health. *Fluctuation and Noise Letters*, *17*(01), 1830001.
 https://doi.org/10.1142/S021947751830001X
- 857 Sun, K., De Coensel, B., Echevarria Sanchez, G. M., Van Renterghem, T., & Botteldooren,
- D. (2018). Effect of interaction between attention focusing capability and visual
- factors on road traffic noise annoyance. *Applied Acoustics*, *134*, 16–24.
- 860 https://doi.org/10.1016/j.apacoust.2018.01.001
- 861 Sundarasekar, R., Thanjaivadivel, M., Manogaran, G., Kumar, P. M., Varatharajan, R.,
- 862 Chilamkurti, N., & Hsu, C.-H. (2018). Internet of Things with Maximal Overlap
- B63 Discrete Wavelet Transform for Remote Health Monitoring of Abnormal ECG Signals.
- *Journal of Medical Systems*, *42*(11), 228. https://doi.org/10.1007/s10916-018-1093-4
- Van Renterghem, T., & Botteldooren, D. (2016). View on outdoor vegetation reduces noise
- 866 annoyance for dwellers near busy roads. *Landscape and Urban Planning*, 148, 203–
- 867 215. https://doi.org/10.1016/j.landurbplan.2015.12.018
- Viollon, S., Lavandier, C., & Drake, C. (2002). Influence of visual setting on sound ratings in
 an urban environment. *Applied Acoustics*, *63*(5), 493–511.
- 870 https://doi.org/10.1016/S0003-682X(01)00053-6
- Ward Thompson, C., Roe, J., Aspinall, P., Mitchell, R., Clow, A., & Miller, D. (2012). More
- green space is linked to less stress in deprived communities: Evidence from salivary
- cortisol patterns. *Landscape and Urban Planning*, *105*(3), 221–229.
- 874 https://doi.org/10.1016/j.landurbplan.2011.12.015
- 875 World Health Organization. Regional Office for Europe. (2011). Burden of disease from
- 876 *environmental noise: Quantification of healthy life years lost in Europe.* World Health
- 877 Organization. Regional Office for Europe.
- 878 https://apps.who.int/iris/handle/10665/326424

- 881 Figure captions
- 882

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

890

891 Figure 2. Mean pleasantness scores and standard errors. The mean scores are shown for the four combinations of sound (bird song (B), traffic noise (T)) and image 892 893 (vegetated scenes (V), non-vegetated scenes (nV)) categories during the task of 894 evaluation of auditory pleasantness (A1) or visual pleasantness (A2). The effects of the Sound category factor (2 levels: bird song (B) and traffic noise (T)), and the effects 895 of the Image category factor (2 levels: vegetated scenes (V) and non-vegetated 896 897 scenes (nV)) on pleasantness scores during the auditory (B1) and visual (B2) tasks. 898 *** p<0.001. Significant perceptual audiovisual interactions are highlighted with red 899 dotted contours.

900

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

908

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

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

918 919

Figure 5. Physiological measure changes due to variation in nature presence within sound or image category, i.e. bird song vs. traffic noise (Auditory effect [B-T]) or vegetated scenes vs. non-vegetated scenes (Visual effect [V-nV]), during the auditory task (Aud. task) and the visual task (Vis. task). (A) Pupil area difference. (B) Corrugator supercilii EMG difference. (C) Heart rate
difference. Individual values (ind. vals.), median values, and bootstrap 95% CIs are
shown. *** p<0.001, ** p<0.01, * p<0.05. Significant perceptual audiovisual
interactions are highlighted with red dotted contours.

928 929

Figure 6. Effects of the five most-unpleasant versus the five least-unpleasant trafficnoise samples on perceptual audiovisual interaction and pupillary reactivity.

(A1) Highly- and Less-unpleasant urban road traffic noise (HT and LT, respectively) in 932 933 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 934 previous studies employing these sounds (Gille et al., 2016; Morel et al., 2012): X 935 stands for "vehicle type" (b: bus; d: motorcycle/moped; p: heavy vehicle; v: light 936 937 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 938 of Highly-unpleasant Traffic noise (HT) and Less-unpleasant Traffic noise (LT) on 939 940 median visual pleasantness rating differences relative to bird song in the visual task 941 (n=31). (B1) Difference in the time course of pupil dilation contrasting low- and highlyunpleasant traffic noise (i.e. traffic noise effect [LT-HT]), with both visual and auditory 942 experiments pooled (n=61). (B2) Effects of variations in unpleasant traffic noise (i.e. 943 944 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, 945 and bootstrap 95% CIs are shown. Significant perceptual audiovisual interactions are 946 947 highlighted with red dotted contours.

948



















Figure 5



