Applied Acoustics, Volume 193, May 2022, 108754

Effective soundscape characterisation of an acoustic metamaterial based window: A comparison between laboratory and online methods

Gioia Fusaro[1,2*], Jian Kang[3], Wen-Shao Chang[1]

[1] School of Architecture, University of Sheffield, UK
[2] Institute of High Performance Computing, A*Star, Singapore
[3] Institute for Environmental Design and Engineering (IEDE), The Bartlett, University College London, UK

Abstract

Qualitative investigation of the sound environment from human perception has lately reached wider attention through psychophysical and Soundscape descriptors; However, laboratory or in situ questionnaire can only reach a limited amount of demographic backgrounds, as, in a limited variety of cultural and geographical variation in a medium-short time, risking to bias/influence the results of the analysis. Starting from a related paper, the authors illustrated how to include human perception and psychoacoustics in evaluating an acoustic metawindow (AMW) unit effectiveness. This paper investigates the robustness of the mixed laboratory and online methodology and proves the validity of the exclusively online-based soundscape questionnaire method according to a number of headset setups. For this reason, a first *Laboratory* experiment with a fixed and controlled setup was followed by an *Online 1* experiment, where the same participants took the same questionnaire (with a three weeks' time difference). Finally, the *Online 2* experiment was used to broaden the sample background (results concerned the online method only). The setups mostly influenced the participants' responses by neutralising it (in 48% of the cases) while reporting a rate with an acceptable error (45% of *Δrate* = 0/5 and 39% of *Δrate* = 1/5). Wired headphones based rates agreed with the laboratory one consistently (36% of *Δrate* = 0 and overall *CV*=2-25% through soundscape descriptors). Wired earphones had a similar *Δrate* percentage but with an overall more neutralised soundscape perception (overall *CV*=1-40% compared with *Laboratory* rates). Wireless earphones caused the extremer *Δrate* (54% of *Δrate* = 3). Finally, natural and human noises are more detectable spontaneously, while traffic or other artificial noise are mostly recognised when directly addressed to the participants. This additional analysis gives more precise information about how efficient is the online method compared to the laboratory one. If accounted accordingly, an online-based soundscape questionnaire setup could also be viewed as a solid alternative to grasp a wider range of human perception points of view according to a broader cultural and geographical background.

1. Introduction

Quantitative methods for determining the impact of acoustic metamaterial (AMM) based building's features over the indoor environment have proved so far to be a resourceful and widespread way to test the effectiveness of these systems throughout the research community [1–4]. On the other hand, human perception has been lately included in mixed methodologies for environmental acoustics and built environment application [5–10]. These include physical and psychoacoustic parameters [11], such as insertion and transmission loss (IL and TL) and Loudness (N), Roughness (R), Sharpness (S), and Fluctuation Strength (FS), soundscape parameters [6] and extensive descriptions of sound sources and psychological effect of sound recordings over human perception [7,12]. However, these physical, psychophysical, and soundscape analyses focus on a limited amount of demographic backgrounds [13]. In most cases, laboratory or in situ tests and questionnaires may take a significant amount of time to reach a representative global cultural and geographical inclusion [14,15].

On the other hand, following an ergonomic study on the window design through a participatory approach, the authors developed an AMM-based window and tested it analytically, numerically, and experimentally from the physical point of view [4,16,17]. Afterwards, in related research, the sphere of study has been implemented from a) quantitative method approximating human perception (psychoacoustic parameters) and b) a qualitative analysis of the window effect on several environmental sound recordings according to the participant's point of view (soundscape descriptors). So, in a related paper, the authors illustrated how to include human perception and psychoacoustics in evaluating an acoustic metawindow (AMW) unit's effectiveness. The methodology presented in the related paper has been developed by first using AMMs influenced environmental sound recordings as input and then merging Soundscape based questionnaire and psychoacoustic analysis. This perception-based evaluation methodology aims to be performed in the laboratory and extended online. However, since there are no specific guidelines on conducting such an experiment on an AMM based design, a new method had to be drawn.

Therefore the methodology still needs to be further discussed to understand 1) how different listening test setups affect the online questionnaire responses compared to the controlled and standardised laboratory one, 2) if a correspondence can be drawn between the human perception based laboratory and online evaluation of an AMM based building's feature through soundscape descriptors, and finally, 3) if the AMM window influence over environmental sound recordings affects the sound sources recognition which is generally addressed in ISO standard soundscape studies [6]. For these reasons, in this technical note, firstly, the results from laboratory and online questionnaires of the same participants are compared to understand if different hearing setups may significantly affect the

AMW unit influence acoustic perception and how. Afterwards, the correlation between broader online (namely *Online 2*) and *Laboratory* results is investigated considering different headset setups and filtering capacities of the AMW unit (which affects the recording that the participants evaluate). Finally, the ability to detect specific categories of sound sources within all the recordings is tested through two different procedures (spontaneous and directed) to assess if the AMW unit affects it significantly. This would not only depend on the AMW unit filtering capacity but also on the human perception sensitivity.

This methodology could be a first step for allowing a more holistic approach over AMM based building features development defining basic guidelines to implement the already existing methods towards a more ergonomic design on a broader and more accessible user background.

2. Materials and methods

In the related paper, first, measurements and recordings of the laboratory acoustic environment were performed to generate input material to be evaluated by the users independently from the laboratory or online setup. An anechoic chamber was used in this first experimental stage (For more details, please see [4]). The AMW unit was attached on the outdoor side with a loudspeaker coupled with a power amplifier FRS 10 WP 8 OHM No. 2101 by VISATON (frequency range from 90 to 19,000 Hz and input power of 25 W) connected to the laboratory computer. The recordings were performed through a Zoom H4n Pro Handy Recorder placed at the same height of the AMW and loudspeaker centre (1.1m from the chamber floor) and 0.5 m of distance in a perpendicular direction to the AMW frontal panel (see [4]). At the same position, a sound level meter with a built-in FFT analyser (Aihua AWA6228) measured the SPL. The seven selected recordings for the study were: #1 Beach, #2 Woodlands, #3 Quiet Street, #4 Pedestrian Zone, #5 Park, #6 Shopping Mall, and #7 Busy Street. These environmental recordings were chosen to have a variety of sound sources and consequent soundscape characterisation. Since the AMW prototype could be applied to any urban environment, it was crucial to get a realistic choice of environmental sounds. Each environmental sound recording was played for 30 seconds, and it was recorded through two configurations: with and without the AMW. So, a total of fourteen recordings were presented to the participants (seven with and seven without the effect of the AMW unit). Secondly, the environmental sound recordings were analysed analytically from a psychoacoustic perspective and experimentally from a human perception point of view through soundscape descriptors. As a result, a significant noise-reducing impact was determined from experimental psychoacoustic and human perception points of view, and participants perceived an effect of neutralisation over all sound recordings with the AMW unit effect. In this paper, laboratory and online hearing setup are considered to establish the overall approximation error and allow researchers to have guidelines in analysing online Soundscape questionnaires. In the presented research, there was a first *Laboratory* experiment with a fixed and controlled setup, followed by an *Online 1* experiment, where the same participants took the same questionnaire (with a three weeks' time difference). Finally, the *Online 2* experiment broadened the sample background (results only concerned with the online method).

2.1 Soundscape questionnaire method

The soundscape questionnaire presented to the 84 participants included the following descriptors: pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful, monotonous [12,18]. The scale used to rate the 14 recordings were based on five votes: strongly agree, somewhat agree, neither, somewhat disagree, and strongly disagree. During both the laboratory and online questionnaire, the participants did not have a time limit to evaluate each listened recording and replay it as many times as possible. The questionnaire-based study was set through Gorilla, an online platform, to broaden the sample size on the second round of the online experiment. A total of 40 participants participated in the *Laboratory* experiment, while an additional 44 were included in the *Online 2* study. A comprehensive list of the participants' demographic backgrounds can be found in Table 1, and more accurate analysis is reported in the related paper. The questionnaire was developed through 15 questions. A comprehensive list of questions can be found in Appendix A.

Table 1 comprehensive list of the participants' demographic backgrounds for each experimental stage: Laboratory, Online 1, and Online 2.

	Laboratory	**Online 1**	**Online 2**
Number of participants	40	16 (among the 40 the Laboratory participants)	44
Male	21	6	22
Female	22	10	18
Non-Binary	0	0	1
Min, Max, Average Age	20-40,30	20-30, 25	23-60, 33
Nationality	Italian	Italian	Italian, French, British, Spanish, Greek, Polish, Egyptian, Libyan, Iranian, Turkish, Indian, Australian, Malaysian, Indonesian, Singaporean, Chinese, Japanese, American
Place of Experiment	Italy, Perugia	Italy, Perugia	Switzerland, U.K., Netherlands, Indiana (U.S.), Canada, Iran, Indonesia, Singapore
Time of Experiment	December 2020, January 2021	January-February 2021	Mar-21
Headset used	Lab Headphones (WH-1000XM4, by Sony)	Miscellaneous: Headphones, earphones, Wireless earphones	Miscellaneous: Headphones, earphones, Wireless earphones

2.2 Laboratory method and setup

The experimental *Laboratory* questionnaire was conducted in a controlled environment room at the University of Perugia (IT) (Italy, Cfa Köppen-Geiger climate class [19]). The test room indoor conditions are thermally controlled, and the experiment was held at stationary and thermally neutral conditions (air temperature at 21 °C and MV at fan speed level L2). The presented experimental campaign provided acoustic stimuli through wired noise-cancelling headphones (model WH-1000XM4, by Sony) to reduce the AC fan noise in the background as much as possible. More details about how the internal background noise level was mapped can be found in the related paper. Overall the sound level of internal background noise (which ranges between 22.8 and 40.4 dB) is reduced broadly by 29dB by the Active Noise Cancelling properties of the headphones used [20]; However, still, some limitations may be highlighted by sound pressure level (*SPL*) analysis in frequency, where peaks are highlighted at a low-frequency range (below 500Hz).

2.3 Online 1 method and setup

The *Laboratory* method is characterised by constant headset setup and autocorrelation noise conditions, while, when conducting the questionnaire online, these parameters may variate slightly or strongly [21,22]. This is why the online method's robustness must be assessed to understand the degree of precision of such an alternative questionnaire method. First, the laboratory and online test results related to the same participants (16 from the original 40 of the Laboratory questionnaire experiment) were compared (*Online 1*) (see Table 1 for further details on the participants *Online 1* setup). This stage of tests was run with a three weeks' time difference from the *Laboratory* one to ensure that participants would not be biased. Participants were enabled to take the soundscape questionnaire independently and in the most convenient and comfortable place for them. The only requirement was to conduct the test in a relatively quiet environment using a set of headphones they already had available. Results were compared in terms of 1) difference in the *Laboratory* and the *Online 1* answers, 2) approximation of the *Laboratory* answer, and 3) headset through which the *Online 1* test was taken. The first parameter is defined as $\Delta rate = rate_{Laboratory} - rate_{Online\ 1}$ and it goes from 0 (in case the answer is the same for both *Laboratory* and *Online 1* rate) to 4 (in case there are 4 points of rate in between the *Laboratory* rate and the *Online 1* rate). The second parameter determines how neutral the Online 1 rate is compared to the *Laboratory* one. In this case, the answer for each question could be a) neutralised (when the *Online 1* rate is closer to the neutral rate), b)

Extremer (when the *Online 1* rate is more extreme than the *Laboratory* one), and c) Opposite (when *Online 1* rate is the opposite of the *Laboratory* one). The third parameter considers wired headphones, wired earphones, and wireless earphones as the *Online 1* test headset setup. It is worth analysing the results according to the different headsets as they could lead to a different audio quality of the soundscape recordings with the AMW unit effect.

In this technical note, indeed, laboratory and online hearing setup are considered to establish the overall approximation error and allow researchers to have guidelines in analysing online Soundscape questionnaires. Even though the overall comparison of the human perception of the recordings with the AMW was run between Laboratory and Online 1 results, the sample was too small (16 participants) to comprehensively evaluate that complex research question. At the same time, 16 participants were considered representative to investigate different headsets influence over soundscape evaluation since the same *Laboratory* experiment participants were observed. On the other hand, the *Online 2* sample (44 participants) was much broader than the *Online 1* (and more comparable to the *Laboratory 1,* which was 40 participants). So, for the same representability reason, Online 2 results were considered to serve better the purpose of a more detailed discussion on the relationship between different environmental recordings perception and the AMW unit Amplitude filtering capacity.

2.4 Statistical reliability of the sample

An Intra-Rater Reliability evaluation has been performed to understand if there was any significant influence of the different headset setup on the same participants' response to the auditory perception questionnaire. Intra-rater reliability based on the five-point Likert scale questions was obtained from Cronbach's alpha calculation. Although previous research has identified appropriate and reliable measures (e.g. using psychoacoustic parameters) to assess soundscape comfort, little is known about the reliability of a given individual to determine soundscape comfort on different occasions (i.e. headset setup intra-rater reliability). Findings from Kosters et al. [23], based on an ambience scale in nursing homes, suggest differences in the reliability among individuals about the overall effect of physical and social features of the perceived environment on feelings and moods, behaviour, actions, and reactions. Although intuition would suggest that most individuals have high intra-rater reliability when the mechanical, neurophysiological, and psychological factors that may influence soundscape comfort remain constant, this has not been tested specifically. Individuals with poor headset setup intra-rater reliability would provide inconsistent soundscape assessments under consistent questionnaire judgment and, therefore, might not truly assess the acoustic perception impact of the AMW unit. Thus, soundscape assessments with poor reliability might be potentially misleading in studies investigating the effectiveness of AMMs-based building features on human indoor comfort and may also provide window manufacturers inaccurate evaluations of future products. Therefore, it is crucial to investigate intra-rater reliability of different headset setup based soundscape assessments to identify and better understand individuals with low reliability and develop strategies to account for this poor reliability.

Considering the *Laboratory* results as headset quality benchmark, intra-rater reliability for the *Online 1* experiment was determined by calculating the intra-class correlation coefficient (ICC) [24]. ICC is an attempt to overcome some of the limitations of the classic correlation coefficients. It is a single index calculated variance obtained by partitioning total variance into between and within-subject variance (known as analysis of variance or ANOVA). Thus, it reflects both degrees of consistency and agreement among rating ICC. Furthermore, it is a coefficient with no unit compared across different reliability studies. Therefore, thresholds have been established to judge the reliability quality [25,26]. By using these established thresholds, participants can be classified into groups of individuals with high or acceptable reliability (ICC ≥ 0.7) and individuals who are not able to reliably assess Soundscape with different headset setups (ICC < 0.7).

2.5 Online 2 method and setup

In the *Online 2* questionnaire, the soundscape recordings have been evaluated through the same questionnaire method but with different headsets (recorded by the system for further evaluation) and different background noise conditions. Thus, participants could do the test independently and at any time. The only requirement was to conduct the test in a relatively quiet environment using a set of headphones they already had available. So for accuracy, different headsets results have been compared with the *Laboratory* results to show how the setup in the *Online 2* method affects human perception of environmental sound recordings affected or not by the influence of the AMW unit. A ponderation was included to allow clearer visualisation of different soundscape recordings evaluation based on the 5 point Likert rates related to each *Laboratory* or *Online 2* headset. Participants' responses for each soundscape recording were multiplied by 0-4 according to the participants' rate. Respectively, 'strongly disagree' rates were multiplied times 0, 'somewhat disagree' rates were multiplied times 1, 'neither' rates were multiplied times 2, 'somewhat agree' rates were multiplied times 3, and 'strongly agree' rates were multiplied time 4. Through this ponderation process, the overall soundscape descriptors showed in Figure 2.a-c were calculated as Mean Pondered Vote (*MPV*):

$$MPV = \frac{\sum X_0 \cdot 0 + X_1 \cdot 1 + X_2 \cdot 2 + X_3 \cdot 3 + X_4 \cdot 4}{N} \quad 1$$

where

- X_0= total 'strongly disagree' votes for specific soundscape recordings
- X_1= total 'somewhat disagree' votes for specific soundscape recordings
- X_2= total 'neither' votes for specific soundscape recordings
- X_3= total 'somewhat agree' votes for specific soundscape recordings
- X_4= total 'strongly agree' votes for specific soundscape recordings
- N= sample size

Moreover, *Laboratory* results are used in this analysis to understand the validity of the online method in terms of the Likert scale and about each soundscape descriptor. The agreement between different setup rates are observed in terms of coefficient of variation (*CV*) according to each soundscape descriptor:

$$\boldsymbol{CV} = \frac{\boldsymbol{RMS_1}}{\boldsymbol{RMS_2}} \times \mathbf{100} \quad (\%) \qquad 2$$

where the Root Mean Square (RMS_2) = MPV_{Lab} while the Root Mean Square (RMS_1) expresses the measure of the magnitude of each Likert scale and is used to determine standard deviation and mean between the *Online 2* and *Laboratory* rate values:

$$\boldsymbol{RMS_1} = \sqrt{\sum_{i=1}^{N}\left(\boldsymbol{MPV_{Online2,i}} - \boldsymbol{MPV_{Lab}}\right)^2} \qquad 3$$

where:

- RMS_1= RMS to determine the standard deviation.
- $MPV_{Online2,i}$= *MPV* calculated from the *Online 2* method with a specific headset.
- MPV_{Lab}= *MPV* calculated from the *Laboratory* method.

The CV is the ratio of the standard deviation to the mean and shows the extent of variability concerning the population's mean. So, the higher the *CV*, the greater the dispersion RMS.

2.6 Sound sources detection method

When considering the analysis of questionnaire-based soundscape evaluation, it is fundamental to analyse the perception of specific sound sources, which can be done through either a spontaneous or directed detection process. In the first case, the participants are asked to report any sound sources they can detect within a specific environmental sound recording. In the second case, some sound sources options or categories are given as an example, and the participants have to rate how much they perceive a specific source category within the acoustic environmental recordings (on a five-point Likert scale). These two methods have been widely used separately [12,27]; However, in this study, both were used for each environmental sound recording with the AMW unit effect.

In this research, spontaneous sound sources detection was performed in an extended written form where a comma separated each sound source. The directed detection was based on a five-point Likert scale which included the following categories: 1) Traffic Noise (e.g. cars, bus, trains, aeroplanes), 2) Other Noise (e.g. sirens, construction, industry, loading of goods), 3) Sounds from human beings (e.g. conversation, laughter, children at play, footsteps), 4) Natural sounds (e.g. singing birds, flowing water, wind in vegetation). Due to their different report nature, a ponderation and alignment of the results scale had to be performed, followed by a percentage transformation in order to compare them. For the directed votes, the ponderation was performed as in Equation (4). The percentage expression was then calculated considering the sum of the total votes for all the sound sources categories (Traffic Noise, Other Noise, Sound from human beings, and Natural sound) as the 100% of sound sources TOT_{ss}. So each sound sources percentage (SS%) was calculated as:

$$SS\% = \frac{(SS_{votes} \cdot 100)}{TOT_{SS}} \quad 4$$

Where

SS_{votes}=overall votes for specific sound sources

TOT_{ss}= overall sum of all the votes for all the sound sources for each spontaneous or directed rate

2.7 AMW unit filtering over seven different environmental sound recordings

As introduced in the related paper, different environmental sound recordings may be affected differently by the AMW unit, so a preliminary analysis of its filtering capacity is necessary to set a term of comparison for the soundscape questionnaire results. The environmental sound recordings considered are #1 Beach, #2 Woodlands, #3 Quiet Street, #4 Pedestrian Zone, #5 Park, #6 Shopping Mall, and #7 Busy Street; However, for the sake of simplicity, the most representative sound recordings in terms of AMW unit filtering and human perception are here discussed (#1,2,7) and showed in Figure 2. In all three analysed recordings, the AMW unit Amplitude filter was calculated as:

$$AMW\ unit\ Amplitude\ Filter = \frac{SPL\ (with\ AMW\ unit)}{SPL\ (without\ AMW\ unit)} \quad 5$$

Where $SPL\ (with\ AMW\ unit)$ is the SPL measured in the first experimental conditions (Singapore) with the AMW unit effect applied, while $SPL\ (without\ AMW\ unit)$ is the original environmental sound recording measured in the same conditions but without the AMW unit effect applied. *AMW unit Amplitude Filter* values <1 highlight a filtering capacity while values ≥1 indicate no filtering effect.

3. Results

In this section, different hearing setups are firstly discussed by comparing the related results with the laboratory ones. This may clarify if such factors significantly affect the AMW unit's acoustic perception and how. Secondly, the correlation between *Online 2* and *Laboratory* results is investigated considering different headset setups and filtering capacities of the AMW unit (which affects the recording that the participants evaluate). Finally, sound sources detection is tested from a spontaneous or directed approach to assessing if the AMW unit affects it significantly. Within this results analysis, the human perception sensitivity (qualitative) is considered on top of the AMW unit filtering capacity (quantitative).

3.1 Robustness of the Online 1 method through the laboratory results

Overall (without considering the different headsets), the participants' rates of the soundscape recordings are Neutralised in 48% of the cases, Extremer in 33% of the cases and the Opposite to those they gave in the *Laboratory* test for the 19% of the cases. The difference in the participants' rate of soundscape recordings between the *Laboratory* and the *Online 1* ($\Delta rate$) is shown in percentages and according to the overall (TOT= total) and each specific headset in Figure 1. Overall there is a good correspondence of the *Online 1* rating with the *Laboratory* ones. 44.6% of the answers are the same, followed by 37.8% with 1 rating point of difference, 15.3% with 2 rating points of difference, 2.0% with 3 rating points of difference, and 0.3% with 4 rating points of difference. From the headset setup point of view, all the *Online 1* used systems result reliable in causing the same *Laboratory* response. The majority of perfect correspondence (*Δrate* = 0) is related to the wired headset. However, 29% of it is linked to wireless reproducing systems. Even more equal proportion is found in results with a *Δrate* equal to 1, meaning that the overall headset setups are also good at approximating the *Laboratory* setup. A big difference in the headset acoustic quality reproduction starts to be highlighted from *Δrate*= 2, where Wireless systems turn to cause less compatible responses towards the *Laboratory* ones (same is highlighted for *Δrate*= 3, even if this has an overall percentage of 2%). Furthermore, only wired headphones caused the most extreme and opposite response in the participants' *Online 1* test; However, the overall percentage of such low agreement is very low (0.3%), and it is negligible. As expected, the *Δrate* related to the Wireless earphones is the worst at approximating the *Laboratory* perception judgements of the participants. This will be considered in another discussion of analysis related to the *Online 2* questionnaires, including questionnaire results from the experiment's second part (exclusively online participants).

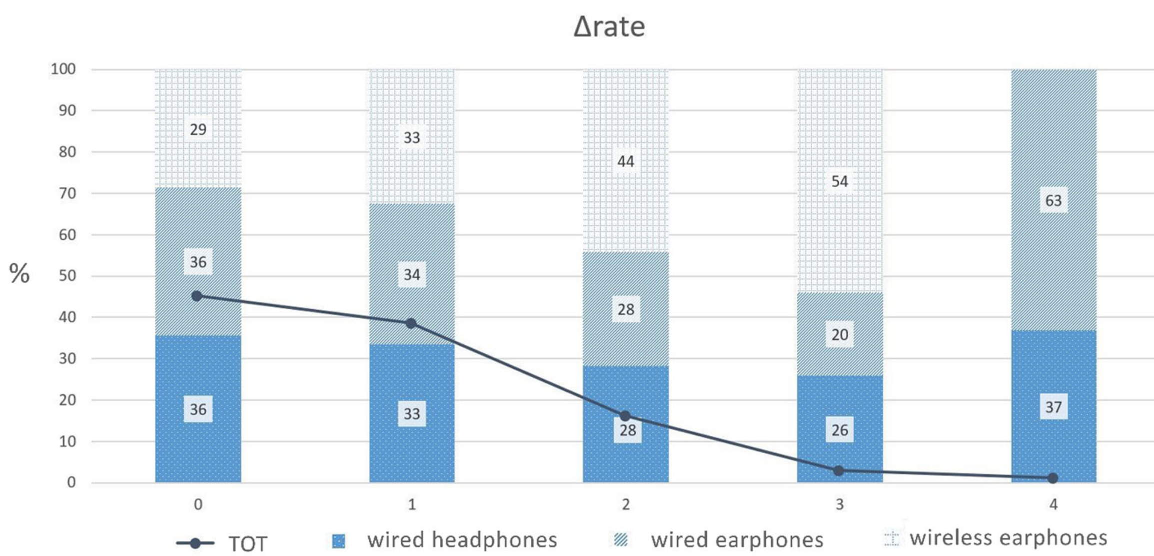

Figure 1 Difference in the participants' rate of soundscape recordings between the Laboratory and the Online 1 (Δrate) according to the overall (TOT= total) and each specific headset (expressed in percentages).

The intra-rater reliability based on the five-point Likert scale questions was assessed using Cronbach's alpha (Cortina [28]) calculated through SPSS Statistics. A Cronbach's alpha value ≥ of 0.7 was considered as acceptable reliability (Cortina [28]). As a result, all the 16 participants tested reliability with an ICC ≥ 0.7 and Cronbach's alpha value ≥ 0.7 (minimum α=0.723, maximum α=0.875), proving the robustness of the online method used for the questionnaire assessing AMW unit soundscape impact.

3.2 Robustness of Online 2 method through the laboratory results

For accuracy and due to the previous consideration in Section 3.1, different *Online 2* results were compared with the *Laboratory* results to show how the headset used in the *Online 2* method affects human perception of environmental sound recordings with the influence of the AMW unit. For simplicity, Figure 2 shows participants' perceptions regarding the environmental recordings #1 Beach, #2 Woodland, and #7 Busy Street. From the analysis of the environmental recordings #1 from a different online headsets point of view, it is highlighted that participants who used wired headphones and earphones seem to have the closest answers' tendency to those who participated in the *Laboratory* test (MPV_{Online2}=2-12%)(see Figure 2.a). Environmental sound recording #1 Beach was evaluated slightly more eventful, vibrant and annoying through the *Online 2* headset (MPV_{Online2}=+ 10-20%) and less Calm (MPV_{Online2}=- 10%) when compared to the Laboratory results; However, wired earphones response seems more neutralised for this recording as eventfulness, monotonousness, and annoying rate perfectly agrees with the Lab results. #2 Woodland soundscape rates show that similarly to #1, wired headphones and earphones highlight a generally not chaotic and not annoying component and, in turn, a calmer and more pleasant soundscape (see Figure 2.b). On the other hand, wireless earphones used on recording #2 result in a worst approximation of the Laboratory results (More annoying and chaotic + 15-18%, while less pleasant, calm, and uneventful - 7-25%) probably due to the smaller and more sensitive sound sources quality and variation and lower acoustic reproduction quality. Figure 2.c shows that for environmental recording #7, namely Busy Street, a slightly more vibrant perception is rated through all the online headsets (+ 7-20% compared to the Laboratory results) and significantly higher when compared with the other online headsets (MPV_{Online2}= + 10-15%). For the other soundscape descriptors, their tendency is well approximated within a range of MPV_{Online2}= ± 6% compared to the *Laboratory* ones. This overall good approximation from the online headsets is probabily due to the easier sound sources recognition combined with the loudness characterising the environmental recording. For example, car engines and car horns contribute to a more chaotic and annoying while less pleasant and calm soundscape. In #7, indeed, these descriptors are used from both *Laboratory* and *Online 2* participants (*CV*= 6%), meaning that this soundscape component is overall assessed from both methods. Overall, the tendency is the same (*MPV* difference of ±7%) for the other soundscape descriptors.

Considering the agreement between the *Online 2* and *Laboratory* participants' rates, in #1 Beach, an overall CV between 1-21% for headphones, 2-25% for earphones, and 4-40% for wireless earphones is highlighted, in #2 Woodland, CV ranges between 3-35% for headphones, 1-40% for earphones, and 2-53% for wireless earphones, while in #7 Busy Street, CV is between 4-47% for headphones, 8-38% for earphones, and 8-61% for wireless earphones. Therefore, there is a significant agreement between the *Online 2* and *Laboratory* participants' rates but an increasing disagreement between wired and wireless setups. Overall, as shown in Figure 2.a-c, wired headphones tend to give overall more comparable responses to the *Laboratory* ones in terms of soundscape descriptors (overall *CV*=1-47%), despite causing a more vibrant, more annoying, and less calm human perception of environmental recordings #1 and #7, and slightly less vibrant and more monotonous perception of the #2. On the other side, online wireless earphones disagree the most with the *Laboratory* headset response (MPV_{Online2} = 5-20%), according to an average of 5/8 soundscape descriptors for each environmental sound recording. Wireless earphones' quality determines a lower agreement between *Laboratory* rates than the online wired headsets (overall *CV*=2-61%). For example, for #2 Woodland, in terms of rate tendency, the wireless headset slightly disagrees with the online response tendency; however, the answers fall into an overall MPV_{Online2}= ± 20% for 7/8 soundscape descriptors compared to the *Laboratory* ones. Wired earphones have slightly the same effect (overall *CV*=1-40% compared with *Laboratory* rates) but for an average of 3/8 soundscape descriptors in each environmental recording. This is mostly related to noise heard in the background and overlapping to the heard recordings from the *Online 2* participants while taking the questionnaire. 81% of the participants who used wireless headset setup reported it, while 20% pointed this issue out as a significantly disturbing factor in the soundscape evaluation.

So, in conclusion, as highlighted by the previous comparison between *Laboratory* and *Online 1* results, the *Online 2* experiment confirms that wireless headsets reduce the overall response tendency agreement with the *Laboratory* one and are most sensitive to the background noise. Therefore, for online soundscape questionnaires, the participants should better identify disturbing background noise before the experiment, and wireless headset setup should be avoided. However, excluding wireless earphones results, differences between *Laboratory* and *Online 2* rates are always within 1-47% CV. So, these two methods are equally useful tools for studying the human perception of environmental sound recordings according to soundscape descriptors. Moreover, this method has proved to be applicable also to AMM based buildings' features analysis from the human perception point of view. Of course, as highlighted in this section, attention must be paid to which headset is used, and the rates accordingly pondered, as they might slightly change the results.

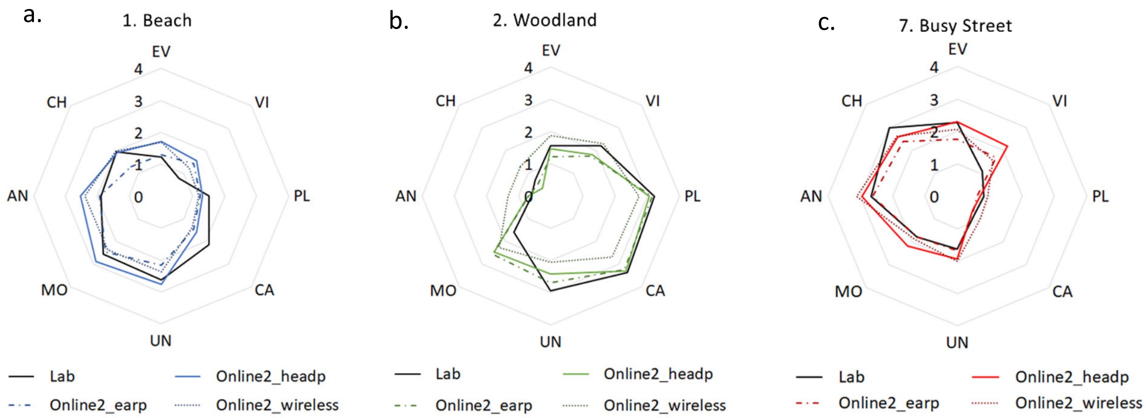

Figure 2 Schematics of the soundscape recording a) #1 Beach, b) #2 Woodland, and c) #7 Busy Street through Laboratory and Online 2 participants evaluation (with wired headphones, wired earphones, and wireless headphones and earphones), and through. Abbreviations below figure a-c respectively indicate questionnaire results from: Lab=Laboratory test, Online2_headp=Online test with headphones setup, Online2_earp=Online test with earphones setup, Online2_wireless=Online test with wireless headset setup. Soundscapes descriptors have been expressed through the following abbreviation for the sake of the readership: EV=Eventful, VI=Vibrant, PL=Pleasant, CA=Calm, UN=Uneventful, MO=Monotonous, AN=Annoying, CH=Chaotic.

It is also worth noticing from Figure 3 that there is a correlation between the filtering capacity of the AMW unit and the Soundscape perception of the participants for each presented environmental recording. Indeed, Figure 3 shows the AMW unit filtering capacity in the frequency domain (50-10kHz) and the related soundscape rating from the Laboratory to the Online 2 with and without the AMW effect within the recordings. Overall, as depicted in Figure 3.a the AMW unit filters the environmental recordings effectively from 300 to 5kHz, 7k to 8kHz, and 9k to 10kHz, while not effectively at 5k-7kHz and 8k-9kHz, and 50-300Hz. In previous research, indeed, the AMW unit noise reducing capacity demonstrated a limitation over a lower frequency range since all the studied files are affected by a magnification of the signal culminating at 280Hz [4]. This issue has been overcome and numerically demonstrated in that specific journal paper through an acoustic broadband optimisation; However, due to the current pandemic situation, the AMW unit model used for these experiments was the basic one (without broadband optimisation). From Figure 3.b-d, there is an overall good filtering capacity and significant neutralisation of the signal from 2000 to 5000Hz, resulting in a human perception response to the AMW unit effect that is, in most of the environmental sound recordings, less eventful and vibrant but calmer (less chaotic) than the original environmental recordings perception. Nevertheless, it is worth noticing that the calmness potential of the AMW unit is not as performative for the environmental recording #2 Woodland in terms of human perception (Figure 3.c). As shown in Figure 3.c (solid green line), this trend is due to the decrease of the AMW unit Amplitude Filter performance in the frequency ranges 1700-2800Hz, and due to the absence of filtering effect near 5200Hz, 5400-5700Hz, 6400-6700Hz, 7900-9500Hz. Even if this turns into a slight change in the human perception, it is worth noticing it as a limit of the AMW unit, which needs to be considered and addressed in the future implementation of the system tuning.

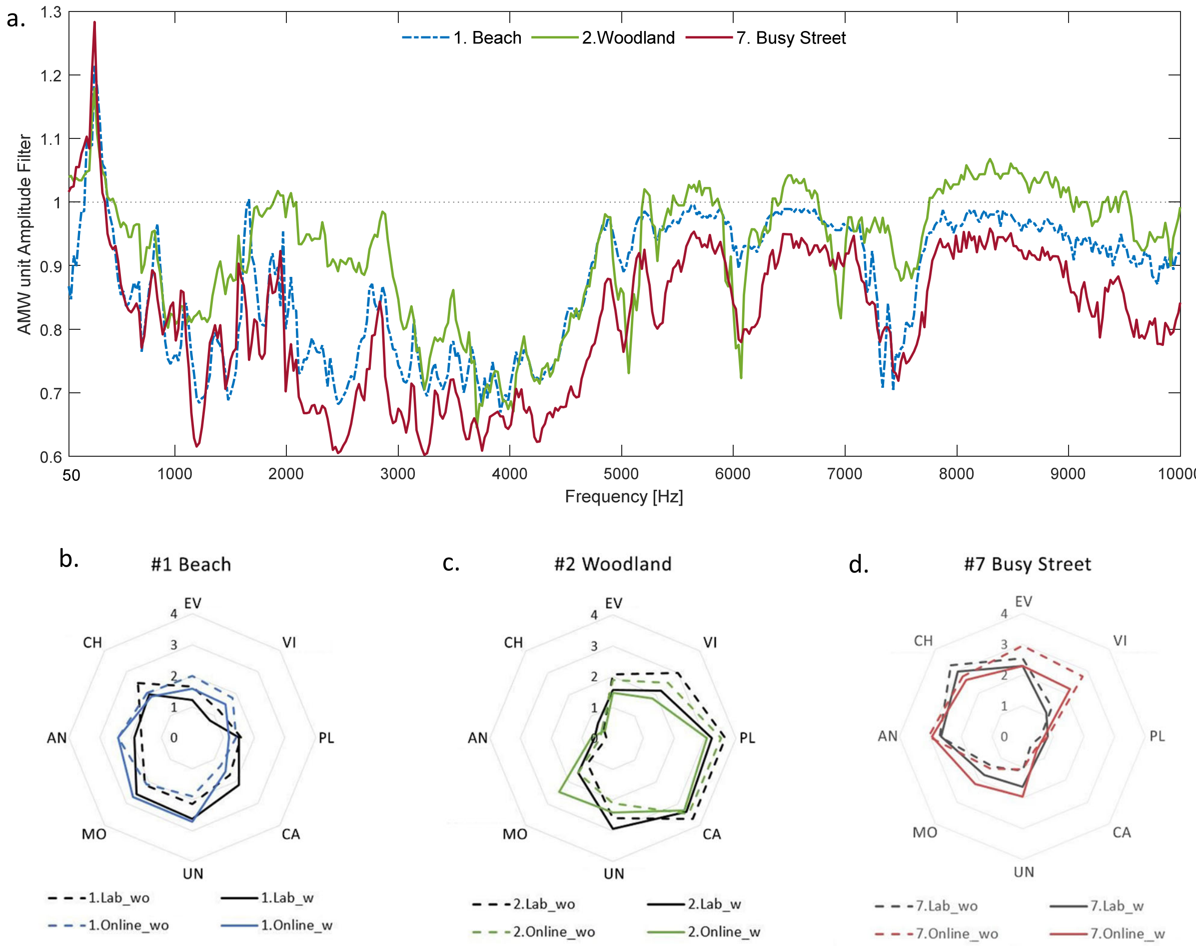

Figure 3 Schematics of the a) AMW unit Amplitude Filter analysis of the environmental sound recordings and soundscape evaluation of recording b) #1 Beach, c) #2 Woodland, and d) #7 Busy Street through Laboratory and Online 2 experiments (for the sake of simplicity, only wired headphones results are shown). Abbreviations below figure b-d respectively indicate questionnaire results from: Lab_wo=Laboratory evaluation of the recording without the AMW effect, Lab_w=Laboratory evaluation of the recording with the AMW effect, Online_wo=Online evaluation of the recording without the AMW effect, Online_w= Online evaluation of the recording with the AMW effect. Soundscapes descriptors have been expressed as for Figure 2 through the following abbreviation for the sake of the readership: EV=Eventful, VI=Vibrant, PL=Pleasant, CA=Calm, UN=Uneventful, MO=Monotonous, AN=Annoying, CH=Chaotic.

3.3 Spontaneous and directed sound sources detection

Figure 4 compares the pondered results for spontaneous and directed sound sources detection. This figure highlights that some SS are most likely to be detected by the participants through the directed rate rather than the spontaneous one: *Traffic Noise* overall +6% and *Other Noise* overall +13%. The other two SS tend to be less detected through directed rate: *Sound from human beings* overall -14% and *Natural Sounds* overall -2%. Thus, compared to other positive SS such as *Sounds from human* *beings* or *Natural sounds*, participants tend to pay less attention to negative noises spontaneously while recalling them once they are directly asked to detect them. In terms of results analysis through the effectiveness of the AMW unit, participants detected more Traffic Noise SS overall +8%, and less Natural sounds overall -5%. This is probably due to the decrease in the low-frequency range (50-350 Hz) of the noise reduction potential of the AMW unit (see Figure 3.a); However, because of the influence of the directivity sources detection, there is a range of variation (or error) in the questionnaire results. So, this should be considered for future studies and needs further investigation related to the robustness of the soundscape investigation approach to assess AMM based technologies. Another further research could be, for example, related to different environmental sound recordings.

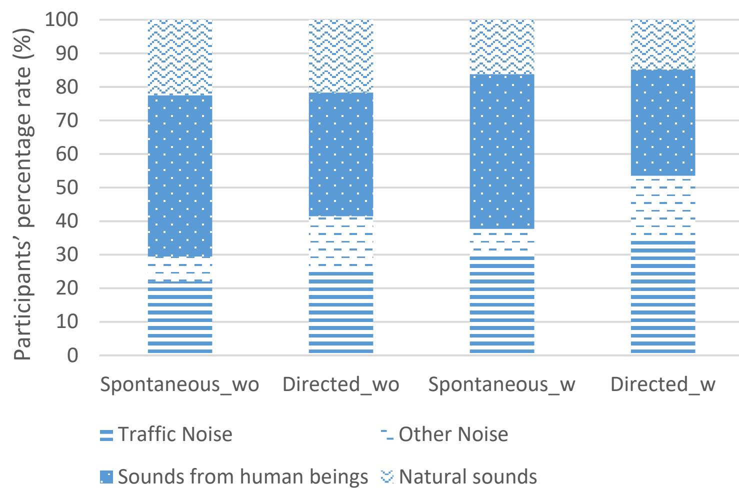

Figure 4 Comparison of the pondered results for both spontaneous and directed sound sources detection in the soundscape recordings without and with the AMW unit effect and evaluated through 4 categories of sound sources: . 1) Traffic Noise (e.g. cars, bus, trains, aeroplanes, etc.), 2) Other Noise (e.g. sirens, construction, industry, loading of goods, etc.), 3) Sounds from human beings (e.g. conversation, laughter, children at play, footsteps, etc.), 4) Natural sounds (e.g. singing birds, flowing water, wind in vegetation, etc.).

4. Conclusions

The robustness of a mixed laboratory-online soundscape questionnaire involving the acoustic effect of an AMMs based window has been assessed through the mean of statistical comparisons. A significant correspondence from the online version of the soundscape questionnaire was drawn with the laboratory setup results, highlighting that different hearing setups influence the AMW unit's effect on acoustic perception. Furthermore, answers from online participants were mostly the same as the laboratory ones or slightly neutralised, which means that these two methods are equally useful tools for studying AMM based buildings' features according to human perception. Of course, attention must be paid to which headset is used, and the results accordingly pondered, as they might slightly change the results.

The background noise component should be considered for all the online headset setups involved in this study. Especially for wireless headsets, participants reported this as a significantly disturbing factor in the soundscape evaluation. So, online wireless headsets reduce the overall response tendency agreement with the laboratory one and are most sensitive to the background noise. For this reason, this kind of headset should be avoided in future AMMs human perception online questionnaires; However, excluding wireless earphones results, differences between laboratory and online rates are always within a CV of 1-47%, which mean that these two methods are an equally useful tool for studying AMM based buildings' features according to human perception. Of course, as highlighted in this section, attention must be paid to which headset is used, and the results accordingly pondered, as they might slightly change the results. Specifically, wired headphones and earphones contribute to a slightly more vibrant, more annoying and less calm human perception (overall *CV*=1-40%); However, soundscape characterisation with wired earphones seem more neutralised (decrease in the overall soundscape descriptors rate for each environmental sound recording).

Finally, from the sound sources detection and sensitivity study from the human perception point of view (in terms of spontaneous and directed procedure), traffic and other noise (e.g. sirens, construction, industry, loading of goods) are most likely to be detected by the participants through the directed rate (overall +6% and +13%). In comparison, sounds from human beings and natural sounds tend to be less likely to be detected through directed rate (respectively -14% and -2%), showing that participants tend to pay less attention to negative noises spontaneously while recalling them once they are directly asked to detect them. So, while planning an online soundscape questionnaire, both complementary techniques (spontaneous and directed) should be used.

The customised mixed methodology has highlighted potential over an exclusive online-based questionnaire rather than the standard laboratory one. If appropriately customised, this online-based method could lead to a more holistic approach for the assessment of AMMs based building's features through human perception and a broader and more accessible sample for the investigation.

Acknowledgements

The authors would like to thank Prof Zhenbo Lu from Sun Yat-Sen University (CN) for the support in the first laboratory measurements in Singapore, and Dr Ilaria Pigliautile and Prof. Anna Laura Pisello from Perugia University (IT) for providing and helping set up the laboratory in Perugia for running the first part of the human perception experiments.

Bibliography

[1] Yu X. Design and in-situ measurement of the acoustic performance of a metasurface ventilation window. Appl Acoust 2019;152:127–32. https://doi.org/10.1016/j.apacoust.2019.04.003.

[2] Jiménez N, Groby J-P, Pagneux V, Romero-García V. Iridescent Perfect Absorption in Critically-Coupled Acoustic Metamaterials Using the Transfer Matrix Method. Appl Sci 2017;7:618. https://doi.org/10.3390/app7060618.

[3] Morandi F, Marzani A, De Cesaris S, D'Orazio D, Barbaresi L, Garai M. I Cristalli Sonici Come Barriere Antirumore Sonic Crystals As Tunable Noise Barriers. Riv Ital Di Acust 2016;40:1–19.

[4] Fusaro G, Yu X, Lu Z, Cui F, Kang J. A Metawindow with optimised acoustic and ventilation performance. Appl Sci 2021;11:1–16. https://doi.org/10.3390/ app11073168.

[5] Belyi V, Gan WS. Integrated psychoacoustic active noise control and masking. Appl Acoust 2019;145:339–48. https://doi.org/10.1016/j.apacoust.2018.10.027.

[6] UNI en ISO. ISO/TS 12913-2:2018 - Soundscape – Part 2. 2018.

[7] Kim G-W, Hwan Yun M. Understanding the impression of product sounds by integrating quantitative and qualitative findings. Int J Ind Ergon 2018;63:98–109. https://doi.org/http://dx.doi.org/10.1016/j.ergon.2017.04.002.

[8] Gontier F, Lagrange M, Aumond P, Can A, Lavandier C. An efficient audio coding scheme for quantitative and qualitative large scale acoustic monitoring using the sensor grid approach. Sensors (Switzerland) 2017;17. https://doi.org/10.3390/s17122758.

[9] Genuit K, Fiebig A. Psychoacoustics and its benefit for the soundscape approach. Acta Acust United with Acust 2006;92:952–8.

[10] Fusaro G, D'alessandro F, Baldinelli G, Kang J. Design of urban furniture to enhance the soundscape: A case study. Build Acoust 2018;25:61–75. https://doi.org/10.1177/1351010X18757413.

[11] Zwicker E, Fastl H (Hugo). Psychoacoustics : facts and models. Springer; 1999.

[12] Torresin S, Albatici R, Aletta F, Babich F, Kang J. Assessment methods and factors determining positive indoor soundscapes in residential buildings: A systematic review. Sustain 2019;11. https://doi.org/10.3390/su11195290.

[13] Dokmeci Yorukoglu PN, Kang J. Development and testing of indoor soundscape questionnaire for evaluating contextual experience in public spaces. Build Acoust 2017;24:307–24. https://doi.org/10.1177/1351010X17743642.

[14] Yang W, Kang J. Acoustic comfort evaluation in urban open public spaces. Appl Acoust 2005;66:211–29. https://doi.org/10.1016/j.apacoust.2004.07.011.

[15] Pigliautile I, Fusaro G, Kang J, Chang W-S, Pisello AL. Environmental thermal influence over soundscape perception: a test room experimental campaign involving the psychological and physiological description of the indoor environment. J Phys Conf Ser 2021;2042:012136. https://doi.org/10.1088/1742-6596/2042/1/012136.

[16] Fusaro G, Kang J. Participatory approach to draw ergonomic criteria for window design. Int J Ind Ergon 2021;82. https://doi.org/10.1016/j.ergon.2021.103098.

[17] Fusaro G, Yu X, Kang J, Cui F. Development of metacage for noise control and natural ventilation in a window system. Appl Acoust 2020;170:107510. https://doi.org/10.1016/j.apacoust.2020.107510.

[18] Aletta F, Oberman T, Axelsson Ö, Xie H. Soundscape assessment : towards a validated translation of perceptual attributes in different languages. Inter-Noise 2020.

[19] Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci 2007;11:1633–44. https://doi.org/10.1002/ppp.421.

[20] Music headphones used as hearing protection at work — The Noise Chap | Audiometry and Noise Assessments n.d. https://www.thenoisechap.com/hearing-protection-advice/headphones-as-hearing-protection (accessed September 18, 2021).

[21] Zhao S, Brown CA, Holt LL, Dick F. Robust and efficient online auditory psychophysics with the right auditory hygiene. BioRxiv 2021:2021.07.17.452796.

[22] D'Orazio D, Garai M. The autocorrelation-based analysis as a tool of sound perception in a reverberant field. Riv Estet 2017:133–47. https://doi.org/10.4000/estetica.3234.

[23] Kosters J, Kunz M, van den Bosch KA, Andringa TC, Zuidema SU, Luijendijk HJ, et al. Validation of a modified ambiance scale in nursing homes. Aging Ment Heal 2020;0:1–7.

https://doi.org/10.1080/13607863.2020.1747049.

[24] McGraw KO, Wong SP. Forming Inferences about Some Intraclass Correlation Coefficients. Psychol Methods 1996;1:30–46. https://doi.org/10.1037/1082-989X.1.1.30.

[25] Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? Physiotherapy 2000;86:94–9. https://doi.org/10.1016/S0031-9406(05)61211-4.

[26] Chinn S. Statistics in respiratory medicine. 2. Repeatability and method comparison. Thorax 1991;46:454–6. https://doi.org/10.1136/thx.46.6.454.

[27] Xiao J, Aletta F. A soundscape approach to exploring design strategies for acoustic comfort in modern public libraries: A case study of the Library of Birmingham. Noise Mapp 2016;3:264–73. https://doi.org/10.1515/noise-2016-0018.

[28] Cortina JM. What Is Coefficient Alpha? An Examination of Theory and Applications. J Appl Psychol 1993;78:98–104. https://doi.org/10.1037/0021-9010.78.1.98.

[29] Lionello M, Aletta F, Kang J. A systematic review of prediction models for the experience of urban soundscapes. Appl Acoust 2020;170. https://doi.org/10.1016/j.apacoust.2020.107479.

Appendix A – Soundscape Questionnaire

In all the experimental stages (*Laboratory*, *Online 1*, and *Online 2*), participants listened and assessed seven proposed soundscape recordings with and without the effect of the AMW unit (14 recordings in total) comprising the following categories: #1 Beach, #2 Woodlands, #3 Quiet Street, #4 Pedestrian Zone, #5 Park, #6 Shopping Mall, and #7 Busy Street. The recordings were presented randomly. The questionnaire-based study was set through an online platform (Gorilla) to broaden the sample size on the online second part of the experiment. The questionnaire was developed through 15 questions. Soundscape descriptors such as eventful, vibrant, pleasant, calm, uneventful, monotonous, annoying, and chaotic [12,18,29] helped the participants to describe each listened recording. In the comprehensive list presented below, the first question was open reply-based, while the other 14 were based on a 5 point Likert scale:

1. While listening, please write down in the following tab any sound sources you can identify in this sound environment (please separate each sound source with a comma).

2. How did you hear the following four sounds? (Not at all; A Little; Moderately; A lot; Dominates Completely):

2.A Traffic Noise (cars, bus, trains, aeroplanes, etc.)

2.B Other noise (e.g. sirens, construction, industry, loading of goods)

2.C Sounds from human beings (e.g. conversation, laughter, children at play, footsteps)

2.D Natural sounds (e.g. singing birds, flowing water, wind in vegetation)

3. For each of the eight scales below, to what extent do you agree or disagree that the outdoor public space you heard is... (Strongly Agree; Somewhat Agree; Neither; Somewhat Disagree; Strongly Disagree) - 3.A Pleasant; 3.B Chaotic; 3.C Vibrant; 3.E Uneventful; 3.F Calm; 3.G Annoying; 3.H Eventful;

4. Overall, how would you describe the outdoor public space you have just heard? (Very good; Good; Neither bad nor good; Bad; Very bad)

5. How loud would you say this environment was? (Not at all; Slightly; Moderately; Very; Extremely)