INTER-NOISE 2018
Impact of Noise Control Engineering
26-29 AUGUST
CHICAGO, ILLINOIS

Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation

Kang Sun[a)]
Dick Botteldooren[b)]
Bert De Coensel[c)]

Waves Research Group, Ghent University
Technologiepark-Zwijnaarde 15, 9052 Ghent, Belgium

ABSTRACT

Within the framework of the Urban Soundscapes of the World project, a comprehensive database of high quality immersive audio-visual recordings is being collected at various urban locations worldwide. The recordings combine 360-degree video, for presentation using a head-mounted display, with spatial audio, including binaural and first order ambisonics. Recording sites are being selected trough a perception-based protocol that consists of an online questionnaire conducted among panels of local experts, leading to a range of urban sites with a wide variety of soundscapes. This paper reports on the results of a two-stage immersive perception experiment, conducted using a subset of the audio-visual recordings already in the database. In the first stage, the audio-visual recordings are assessed in terms of pleasantness/eventfulness of the soundscape, and in terms of how much the soundscape might interfere with activities. In the second stage, both binaural and first order ambisonics spatial audio techniques are assessed in terms of the degree of realism and immersion they provide, for the different types of soundscapes considered. The results of this benchmark listening test will steer the design of future experiments on the effect of soundscape interventions and on the effect of inter-cultural differences on soundscape perception.

1 INTRODUCTION

The urban soundscape is an important factor in the perception of the quality of the environment of the city we live in, work in, or simply visit. Ambient sounds may evoke thoughts and emotions, may influence our mood or steer our behavior. Cities are comprised of many types of outdoor spaces, each with their distinctive soundscape. Inspired by the potential positive effects a fitting acoustic environment may have on well-being, the challenge of designing the acoustic

[a)] email: kang.sun@ugent.be
[b)] email: dick.botteldooren@ugent.be
[c)] email: bert.decoensel@ugent.be

environment of urban outdoor spaces has attracted attention since long[1,2]. During the past decade, research interest has risen considerably, partly driven by the advent of realistic and affordable immersive audio-visual reproduction systems (head-mounted displays), backed by increasingly efficient and realistic acoustic simulation and auralization models[3]. Immersive virtual reality could even become a valuable tool for interactive participatory evaluation of the soundscape in urban planning and design projects[4,5], as virtual reality reproduction systems are rapidly becoming affordable and widely available.

Physics-based methods may soon make it possible to render indoor virtual acoustic scenes that cannot be distinguished from real auditory environments[6]. However, auralization of urban outdoor spaces differs from the auralization of indoor spaces in the model scale (propagation distances and the number of surfaces and other objects), as well as in the number and complexity of sound sources that have to be considered in order to achieve a realistic acoustic environment. These issues cannot easily be resolved through software optimization and/or an increase in computing power. Therefore, high-quality immersive recordings (spatial audio combined with 360-degree spherical video) of existing spaces are highly valuable to serve as an ecologically valid baseline, on the basis of which the perceptual influence of noise control and soundscaping measures can be assessed through auralization. In order to construct such a database of immersive recordings, two questions need to be answered: how and where to record.

To date, no well-established protocols or standards exist for immersive audio-visual recording and playback of urban environments with soundscape in mind, although standardization efforts with regards to spatial audio recording have been started recently by ISO[7]. Binaural audio recordings, performed using an artificial head, are generally considered to provide the highest degree of realism. Using an artificial head, the sound is recorded as if a human listener is present in the original sound field, preserving all spatial information in the audio recording. The main disadvantage of binaural audio recordings is that the frontal direction, and as such the acoustic viewpoint of the listener, is fixed by the orientation of the artificial head during the recording. This drawback could in theory be solved using ambisonics audio recording[8], a multichannel recording technique that allows for unrestricted rotation of the listening direction after recording. In principle, this technique could therefore provide an alternative to binaural recordings in the context of soundscape studies. However, the ambisonics technique has its own disadvantages, such as the more complex process of playback level calibration and equalization as compared to the binaural technique, the necessity of head tracking and real-time HRTF updates in case of playback through headphones, and the limited spatial resolution that can be achieved with lower-order ambisonics recordings—to date, there are no truly portable higher-order ambisonics recording systems available. Nevertheless, (first-order) ambisonics has become the *de facto* standard for spatial audio in VR games and platforms providing 360 video playback such as YouTube or Facebook.

Then there remains the question of how to choose suitable recording locations among the immense number of potentially interesting spots even within a single city. Selection of urban sites for performing soundscape evaluation studies on the basis of audio recordings is most often performed in an *ad hoc* manner, e.g. based on their value for residents or visitors, the presence of particular sound sources, or on past or planned soundscape interventions at the site. When the aim is to achieve a representative coverage of urban soundscapes, for example for preservation, research or education, a more systematic site selection strategy might be more favorable.

Within the *Urban Soundscapes of the World* project[9], a comprehensive database of high quality immersive audio-visual recordings is being collected at various urban locations worldwide. Sites are hereby selected using a perception-based protocol designed to cover a range of urban sites with a wide variety of soundscapes. Subsequently, 360-degree videos are recorded at the selected locations, in combination with simultaneous binaural and first-order ambisonics spatial audio, allowing to compare both techniques. This paper reports on the results of a two-stage immersive perception experiment, conducted using a subset of the audio-visual recordings already in the database. In the first stage, the use of an alternative soundscape classification method based on activity interference is evaluated; in the second stage, binaural and first-order ambisonics techniques are assessed in terms of the degree of realism and immersion they provide. In Section 2, details on the site selection and recording techniques are provided. In Section 3, the experiment methodology is explained in detail. In Section 4, the results of the experiment are discussed.

2 URBAN SOUNDSCAPES OF THE WORLD

2.1 Site selection

The aim of the *Urban Soundscapes of the World* project[9] is to set the scope for a standard on immersive recording and reproduction of urban acoustic environments with soundscape in mind. In the process, a database of documented exemplars is created: a series of immersive audio-visual recordings of the acoustic and visual environment at a selection of locations in a range of cities worldwide. Moreover, this reference database of good (and bad) examples of urban acoustic environments may also support the further introduction of urban soundscape design in education and practice, as architects and designers commonly work by example.

The scale of the immersive audio-visual recording effort in the Soundscapes of the World project, and the way the recording locations are selected, differs from most earlier work, such as by Farina *et al*[10]. In particular, within each city, sites are selected in a systematic and perception-based way, grounded in the experience of local experts: people familiar with the sounds that can be heard in that city. An online questionnaire survey is conducted among inhabitants, in which they are asked to pinpoint outdoor public spaces within their city that they perceive in a particular way along the soundscape perception dimensions of pleasantness and eventfulness. Locations obtained from the online survey are then spatially clustered using the Google MapClusterer API, which allows to extract a shortlist of prototypical locations. This approach was designed to lead to a range of urban sites with a heterogeneous variety of soundscapes, more or less uniformly covering each of the four quadrants on the 2D core affect perceptual space[11,12]. In each city, participants are recruited among local students, and through calls for participation on relevant Facebook pages and with local guide associations. More information about the site selection protocol can be found in De Coensel *et al*[9].

2.2 Audio-visual recording

Combined and simultaneous audio and video recordings are performed at the selected locations within each city, using a portable, stationary recording setup. Photographs of this setup are shown in Fig. 1. The setup consists of the following components: binaural audio (HEAD acoustics HSU III.2 artificial head with windshield and SQobold 2-channel recording device), first-order ambisonics (Core Sound TetraMic microphone with windshield and Tascam DR-680 MkII 4-channel recording device) and 360-degree video (GoPro Omni spherical camera system, consisting of 6 synchronized GoPro HERO 4 Black cameras). The ears of the artificial head, the video camera system and the ambisonics microphone are located at heights of about 1.50m, 1.70m and 1.90m respectively. It was chosen to stack the audio and video recording devices vertically, such that no horizontal displacement between devices is introduced, which could otherwise give rise to an angular mismatch for the localization of sound sources in the horizontal plane. There needed to be a minimal distance of about 20cm between the camera and both the binaural and ambisonics microphones, such that these do not show up prominently on the recorded video, and can be masked easily using video processing software. All audio is recorded with a sample rate of 48 kHz and a bit depth of 24 bit, and are stored in uncompressed .wav format; moreover, the binaural recordings are performed according to the specifications set forth in ISO TS 12913-2[7]. Note that the recording setup is highly portable: when disassembled, all components can be carried by a single person, assembling the setup takes about 10 minutes, and batteries and memory of all recording devices allow for about a full day of recording.

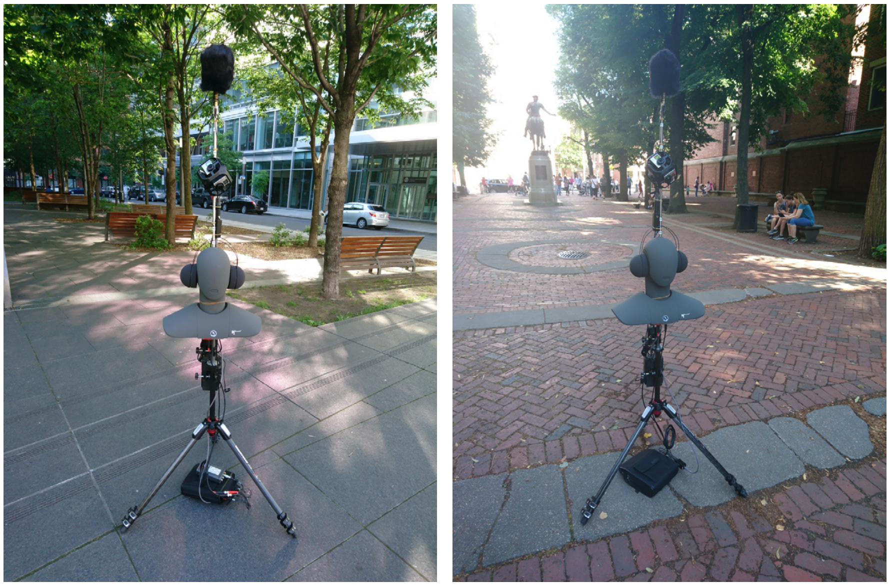


Fig. 1 – Photographs of the recording setup in Montreal (left) and Boston (right).

At each recording location, the recording system is oriented towards the most important sound source and/or the most prominent visual scene—this orientation defines the initial frontal viewing direction for the 360-degree video and ambisonics recordings, and the fixed orientation for the binaural recordings. Time synchronization is performed at the start of each recording by clapping hands directly in front of the system; this also allows to check correct 360-degree alignment of all components in post-processing. At each location, at least 10 minutes of continuous recordings are performed, such that 1-minute or 3-minute fragments containing no disturbances can be extracted easily in post-processing. During recording, the person handling the recording equipment is either hiding (in order not to show up on the 360-degree video) or, in case hiding is not possible, blending in the environment (e.g. performing the same activities as the other people around).

Post-processing of audio and video is performed using a range of software, including Kolor Autopano Video 3.0 (stitching and time synchronization of video, and masking of tripod and binaural/ambisonics microphones in the video), HEAD acoustics ArtemiS 8.3 (processing of binaural recordings and calculation of acoustical properties), VVMic 3.5 (processing of ambisonics recordings, conversion from A-format to B-format using microphone-specific calibration/equalization files), FFmpeg (synchronization of audio and video, color calibration of video, and final selection of segments and combination of media into .mov container) and Google Spatial Media Metadata Injector (for adding 360-degree video and spatial audio metadata to the videos). A software toolbox was developed to allow easy extraction of calibrated and synchronized segments, ready for playback, with any type of combination of audio and video.

3 IMMERSIVE PERCEPTION EXPERIMENT

3.1 Participants

Participants were recruited among Master and PhD students at Ghent University. To date, twenty participants took part in the perception experiment (6 female, 14 male). The mean age of the participants was 28.9 yr (standard deviation 2.8 yr, range 25-35 yr). The participants performed the perception experiment individually, and were offered a gift voucher as compensation. As of writing, the experiment is still ongoing, and a second batch of twenty participants is scheduled. Therefore, the results reported in this paper are tentative.

3.2 Stimuli

Thirty 1-minute stimuli are extracted from a subset of the recordings currently in the Soundscapes of the World database. Table 1 gives an overview of their properties (location, time, and L_{Aeq}). The L_{Aeq} of each stimulus was calculated on the basis of the binaural signal, applying an independent-of-direction (ID) equalization, and taking the energetic average between both ears. Stimuli were recorded in sunny to partly cloudy weather conditions with little to no wind.

Table 1 – Overview of the stimuli of the experiment: (upper) Stage 1, (lower) Stage 2.

Label	City	Date	Time	Location	Longitude	Latitude	$L_{\mathrm{Aeq,1min}}$
R0002	Montreal	2017-06-22	08:43	Place d'Armes	45.504683	-73.557150	66.5
R0003	Montreal	2017-06-22	09:43	Tour de l'horloge	45.511973	-73.545911	55.0
R0007	Montreal	2017-06-22	15:26	Chalet du Mont-Royal	45.503405	-73.587005	54.8
R0010	Montreal	2017-06-22	17:53	Square Phillips	45.503807	-73.568543	67.5
R0011	Montreal	2017-06-22	19:10	Place Jacques Cartier	45.507680	-73.552625	66.1
R0015	Boston	2017-06-28	12:41	Old State House	42.359039	-71.057139	69.5
R0016	Boston	2017-06-28	13:11	Quincy Market	42.359860	-71.055825	74.6
R0017	Boston	2017-06-28	13:47	Post Office Square	42.356230	-71.055600	65.8
R0018	Boston	2017-06-28	14:23	R. F. Kennedy Greenway	42.354721	-71.052073	66.1
R0020	Boston	2017-06-28	16:31	Paul Revere Mall	42.365687	-71.053446	57.4
R0022	Tianjin	2017-08-24	08:54	Peiyang Square (TJU campus)	39.107327	117.170222	62.2
R0026	Tianjin	2017-08-24	11:46	Water Park North	39.090986	117.163317	60.4
R0029	Tianjin	2017-08-24	15:29	Haihe Culture Square	39.130202	117.193256	73.5
R0031	Tianjin	2017-08-24	16:26	Tianjin Railway Station	39.133779	117.203206	65.2
R0033	Tianjin	2017-08-24	17:59	Nanjing Road	39.118566	117.185557	65.3
R0036	Hong Kong	2017-08-29	15:43	Wanchai Tower	22.279705	114.172450	68.7
R0040	Hong Kong	2017-08-30	07:44	Hong Kong Park	22.277824	114.161488	64.1
R0041	Hong Kong	2017-08-30	08:50	Wong Tai Sin Temple	22.342062	114.194042	69.7
R0047	Hong Kong	2017-08-30	13:36	Peking Road	22.296512	114.171813	77.0
R0048	Hong Kong	2017-08-30	14:30	Ap Lei Chau Waterfront	22.245093	114.155663	62.2
R0050	Berlin	2017-09-09	16:57	Breitscheidplatz	52.504926	13.336556	72.4
R0054	Berlin	2017-09-10	11:32	Gendarmenmarkt	52.513517	13.392900	60.8
R0058	Berlin	2017-09-10	14:18	Lustgarten	52.518604	13.399195	65.2
R0060	Berlin	2017-09-10	15:39	James-Simon Park	52.521787	13.399158	65.9
R0061	Berlin	2017-09-10	16:32	Pariser Platz	52.516145	13.378545	67.7
R0001	Montreal	2017-06-22	08:02	Palais des congrès	45.503457	-73.561461	65.8
R0012	Boston	2017-06-28	09:36	Boston Public Garden	42.353478	-71.070151	62.5
R0030	Tianjin	2017-08-24	16:00	Century Clock	39.132620	117.198314	63.2
R0038	Hong Kong	2017-08-29	17:07	Taikoo Shing	22.286715	114.218385	64.6
R0055	Berlin	2017-09-10	12:08	Checkpoint Charlie	52.507796	13.390011	66.5

The stimuli of the first stage of the experiment (upper 25 in Table 1) contain 360-degree video and a first-order ambisonics audio track. For reference, these stimuli have been uploaded to YouTube as 360-degree videos with spatial audio. They can be experienced using a PC (however, ambisonics playback only works in the latest versions of Firefox or Chrome), or on smartphone or tablet using the YouTube app; headphones provide the best experience. The playlist can be found at https://www.youtube.com/playlist?list=PL7YplJbeU4sKnGbO_p3EZwClZnShSkkHY. Note that the video quality of the stimuli uploads on YouTube is lower than that of the original stimuli used in the experiment due to compression performed in the upload process.

The stimuli of the second stage of the experiment (lower 5 in Table 1) contain a fixed HD video, cut out from the original video in the frontal viewing direction, and padded with black in order to obtain again a 360-degree spherical video that can be viewed through a head-mounted display. This creates a "window" or "cinema" effect, forcing the participant to watch only in the frontal direction. Fig. 2 shows screenshots of the cut-out videos. Furthermore, these stimuli are created in two flavors: with first-order ambisonics spatial audio track (allowing for head rotation) and with binaural audio track (which provides a fixed, i.e. head-locked, listening direction).

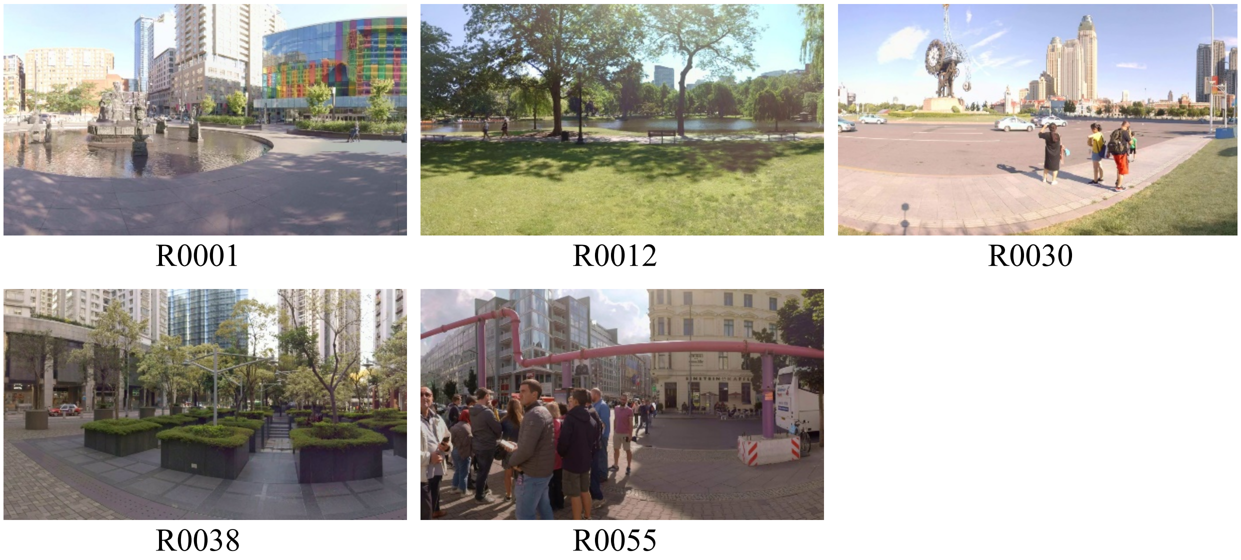

R0001 R0012 R0030
R0038 R0055

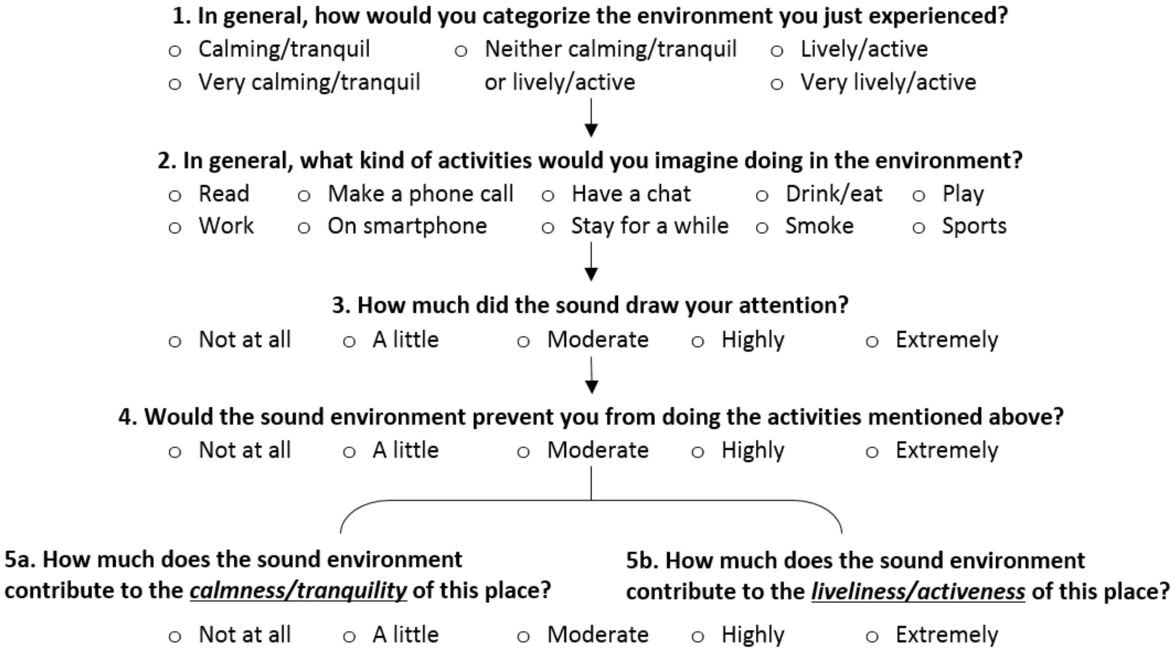
Fig. 2 – Screenshots (cut-outs) of the 5 stimuli used in the second stage of the experiment.

3.3 Audio-visual reproduction

During the experiment, participants are seated inside a soundproof booth. Recordings are played back using a PC (placed outside the booth), equipped with the GoPro VR Player 3.0 software, which allows to play back video with spatial audio. The 360-degree video is presented through an Oculus Rift head-mounted display, and the participant can freely move its head and look around in all directions. The audio is played back through Sennheiser HD 650 headphones, driven by a HEAD acoustics LabP2 calibrated headphone amplifier. Stimuli with binaural audio track (second stage of the experiment) are automatically played back at the correct level, as the headphone amplifier and headphones are calibrated and equalized for the artificial head with which the recordings were made. The gain of the ambisonics audio tracks (first and second stages of the experiment) has been adjusted such that their level is as close as possible to that of the corresponding binaural audio tracks.

3.4 Experiment outline

Before the start of the experiment, each participant was briefly informed about the experimental procedure, i.e. that he/she had to watch videos through a head-mounted display, and that he/she had to answer a small set of questions after each video. In the first stage of the experiment, the 25 spherical videos with ambisonics were presented in random order, but still grouped by city. After each video, questions were projected on screen and the participants were asked to speak out their answers, such that they did not need to take of the head-mounted display between fragments. In a first set of questions, participants were asked to rate the locations they had experienced, on an 11-point scale from 0 (not at all) to 10 (extremely), in terms of a number of adjectives (full of life and exciting, chaotic and restless, calm and tranquil, lifeless and boring) taken from the four quadrants in the principal components analysis performed by Axelsson *et al*[11]. In a second set of questions, an alternative, hierarchical soundscape classification method is tested, outlined in De Coensel *et al*[9]. This method is designed to classify locations according to how much the soundscape contributes to or interferes with the activities that could be performed at the site. Fig. 3 shows the questions asked for this alternative classification method. In particular, either question 5a or 5b is asked, depending on the answer on question 1: choosing the option "very calming/tranquil" or "calming/tranquil" leads to question 5a, all other answers lead to question 5b.



1. In general, how would you categorize the environment you just experienced?

○ Calming/tranquil ○ Neither calming/tranquil or lively/active ○ Lively/active
○ Very calming/tranquil ○ Very lively/active

2. In general, what kind of activities would you imagine doing in the environment?

○ Read ○ Make a phone call ○ Have a chat ○ Drink/eat ○ Play
○ Work ○ On smartphone ○ Stay for a while ○ Smoke ○ Sports

3. How much did the sound draw your attention?

○ Not at all ○ A little ○ Moderate ○ Highly ○ Extremely

4. Would the sound environment prevent you from doing the activities mentioned above?

○ Not at all ○ A little ○ Moderate ○ Highly ○ Extremely

5a. How much does the sound environment contribute to the *calmness/tranquility* of this place?

5b. How much does the sound environment contribute to the *liveliness/activeness* of this place?

○ Not at all ○ A little ○ Moderate ○ Highly ○ Extremely

Fig. 3 – Questions asked after each video in the first stage of the experiment.

After the first stage of the experiment, which typically lasted for about 1 hour, there was a short break, after which the second stage of the experiment was started. In this stage, the 5 pairs of recordings (fixed video with ambisonics/binaural audio track) were presented, again in random order. After each fragment, the participants were asked to rate the soundscape they experienced, on a 5-point scale, in terms of envelopment, immersion, representation, readability, realism and overall reproduction quality, based on the scales developed by Guastavino *et al*[13].

After the experiment, a small questionnaire was administered, which contained questions of demographic nature. At the end, a short hearing and vision test was performed, in order to make sure that the participant had normal hearing and was not color-blind.

4 RESULTS AND DISCUSSION

4.1 Soundscape evaluation and classification

The hierarchical soundscape classification method proposed in De Coensel *et al*[9] identifies 4 categories of soundscapes: *backgrounded*, *disruptive*, *calming* or *stimulating*. This classification can be performed based on questions 3 to 5 (as in Fig. 3). Question 3 probes for the degree to which the soundscape draws attention; if the answer is "not at all", the soundscape can be considered to be *backgrounded* (found in 18% of the cases). If not, the soundscape can be either disruptive or supportive for the activity the person might be involved in, within the environment. It is considered *disruptive* if the answer to question 4 is "highly" or "extremely" (19% of cases). If the answer to question 5a resp. 5b is "highly" or "extremely", it is considered supportive, and either *calming* (15% of cases) resp. *stimulating* (19% of cases). In all other cases (29%), the soundscape cannot be categorized in one of the four categories in a crisp way. Fig. 4 shows the distribution of soundscapes that can be categorized into one of the four categories (i.e. 71% of cases), over the overall audiovisual perception of the environment (answer to question 1).

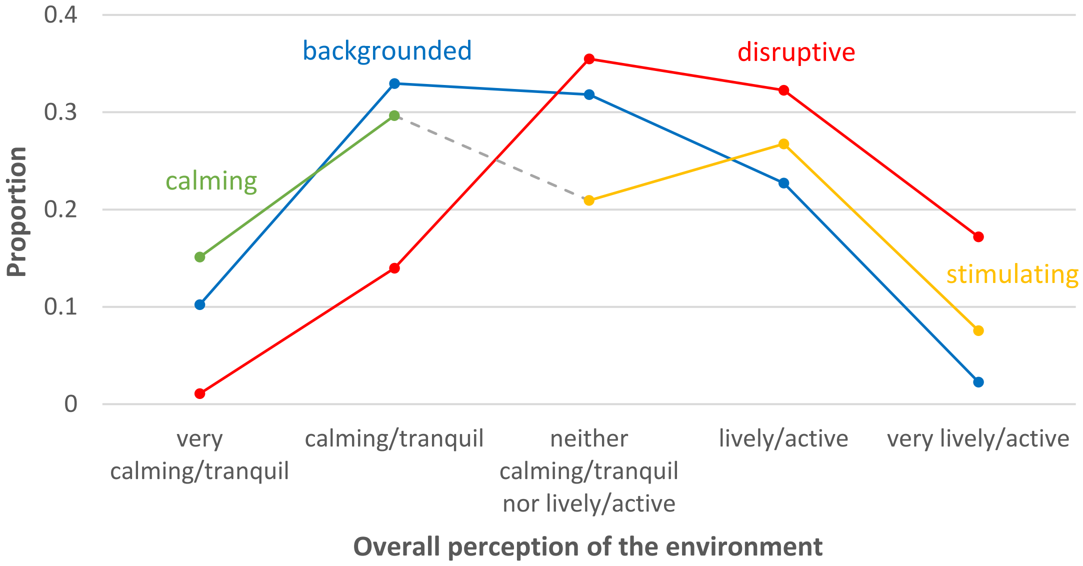


Fig. 4 – Proportion of each soundscape category as a function of overall perception.

For the *backgrounded* category, the sound at the location does not lead to awareness of the acoustical environment. The distribution shows that an overall very lively/active environment is very unlikely if the soundscape is backgrounded but also that a very calming/tranquil environment is less likely. The *disruptive* category shifts the curve towards the "lively/active" side making a very calming/tranquil overall environment very unlikely. The supportive soundscape curve is split into two parts, since people were presented with different questions (5a and 5b) based on their answer to question 1, and pushes the curve towards the extremes in overall perception. A higher proportion of *calming* resp. *stimulating* soundscapes appears in the overall perception cases of "very calming/tranquil" resp. "very lively/active". It is striking that for the option "very lively/active", the proportion of disruptive soundscapes is higher than the proportion of stimulating soundscapes, which might suggest that a relatively larger number of environments with a non-supportive soundscape was selected as stimuli for the experiment.

4.2 Realism and immersion of ambisonics/binaural reproduction

Table 2 shows the results of the comparison between ambisonics (allowing head rotation) and binaural (head-locked) audio playback. The table shows, on a scale from 1 to 5, the median scores on the questions asked (similar results are obtained with average scores). When there is a difference in median between the binaural and ambisonics playback cases, the highest value is underlined.

Table 2 – Median score of 5 pairs of soundscapes in the second stage of the experiment.

Label	Envelopment		Immersion		Representation		Readability		Realism		Overall quality	
	a	b	a	b	a	b	a	b	a	b	a	b
R0001	4.0	4.0	3.5	4.0	4.0	3.5	4.0	3.0	3.5	4.0	4.0	4.0
R0012	3.5	4.0	3.0	3.5	3.0	3.0	3.0	3.5	3.0	3.0	3.0	3.0
R0030	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
R0038	4.0	3.5	4.0	3.0	4.0	4.0	4.0	3.5	4.0	4.0	4.0	4.0
R0055	4.0	4.0	4.0	3.0	4.0	4.0	4.0	4.0	4.0	3.0	4.0	3.0

Earlier research[13] has shown that ambisonics audio results in a high degree of envelopment and immersion. Intuitively, one would expect that the possibility of rotating one's head during playback would result in a higher degree of envelopment and immersion, as compared to the case when one's listening direction is locked. On the other hand, due to the limited spatial resolution offered by first-order ambisonics, one would expect the binaural reproduction to result in a higher degree of readability and realism. The results shown in Table 2 do not allow to draw these conclusions; using a two-sample *t*-test with significance level 0.05, no significant difference is found between both sound reproduction methods, for any of the perceptual dimensions considered. Moreover, the difference between soundscapes is found to be larger than between the audio reproduction methods; some differences are significant, e.g. between R0012 and R0030 regarding representation (both ambisonics and binaural) and realism (binaural), or between R0012 and R0055 regarding immersion (ambisonics), readability (ambisonics) and representation (both ambisonics and binaural). This pilot test therefore justifies the use of ambisonics in the first stage of the experiment; either reproduction method could have been used.

5 CONCLUSIONS

In this paper, a laboratory experiment for classification of soundscapes was presented, which was conducted on the basis of a subset of the audio-visual recordings present in the database of the Urban Soundscapes of the World project. This database consists of immersive audio-visual recordings, collected at locations in several cities worldwide with a wide variety of soundscapes, and combines 360-degree video with spatial audio, including binaural and first order ambisonics.

In the first stage of the experiment, a series of 360-degree video recordings combined with first-order ambisonics spatial audio are assessed using a hierarchical soundscape classification method, based on how well the soundscape is noticed, interferes with possible activities that could be performed at the site or supports overall appreciation of the site. This method was found to be able to categorize 71% of soundscapes into one of four crisp categories, relating in a logical way to the overall perception of the environment. The proposed classification method could be an alternative to the method based on the 2D core affect model that is widely used, e.g. in the Swedish soundscape quality protocol. The advantage of this hierarchical method is that it can account for the existence of backgrounded soundscapes that do not catch attention.

In the second stage of the experiment, both binaural and first-order ambisonics spatial audio techniques were assessed in terms of the degree of envelopment, immersion, representation, readability, realism and overall reproduction quality they provide, for 5 of the different types of urban soundscapes considered. It was found that the variation between soundscapes was larger than the differences between both reproduction techniques; no significant differences were found between the ambisonics and binaural reproduction along the perceptual dimensions considered.

ACKNOWLEDGEMENTS

The research presented in this paper is supported by the HEAD Genuit Foundation through grant No. P-16/11-W, "Urban Soundscapes of the World". The support of this foundation is gratefully acknowledged.

REFERENCES

1. M. Southworth, "The sonic environment of cities", *Environment and Behavior*, **1**(1), 49-70 (1969).
2. R. M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*, Destiny Books, Rochester, Vermont (1994).
3. M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer, Berlin (2008).
4. V. Puyana-Romero, L. S. Lopez-Segura, L. Maffei, R. Hernández-Molina, and M. Masullo, "Interactive Soundscapes: 360°-Video Based Immersive Virtual Reality in a Tool for the Participatory Acoustic Environment Evaluation of Urban Areas", *Acta Acustica united with Acustica*, **103**(4), 574-588 (2017).
5. G. M. Echevarria Sanchez, T. Van Renterghem, K. Sun, B. De Coensel and D. Botteldooren, "Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space", *Landscape and Urban Planning*, **167**, 98-107 (2017).
6. M. Vorländer, "From acoustic simulation to virtual auditory displays", In *Proceedings of the 22nd International Congress on Acoustics (ICA)*, Buenos Aires, Argentina (2016).
7. ISO/PRF TS 12913-2, "Acoustics—Soundscape—Part 2: Data collection and reporting requirements", ISO Technical Specification, Geneva, Switzerland (2018).
8. M. A. Gerzon, "Ambisonics in multichannel broadcasting and video", *Journal of the Audio Engineering Society*, **33**(11), 859-871 (1985).
9. B. De Coensel, K. Sun and D. Botteldooren, "Urban Soundscapes of the World: selection and reproduction of urban acoustic environments with soundscape in mind", In *Proceedings of Internoise '17*, Hong Kong (2017).
10. A. Farina, A. Capra, A. Amendola and S. Campanini, "Recording and playback techniques employed for the Urban Sounds project", In *Proceedings of the 134th AES Convention*, Rome, Italy (2013).
11. O. Axelsson, M. E. Nilsson and B. Berglund, "A principal component model of soundscape perception", *Journal of the Acoustical Society of America*, **128**(5), 2836-2846 (2010).
12. R. Cain, P. Jennings and J. Poxon, "The development and application of the emotional dimensions of a soundscape", *Applied Acoustics*, **74**, 232-239 (2013).
13. C. Guastavino, V. Larcher, G. Catusseau and P. Boussard, "Spatial audio quality evaluation: comparing transaural, ambisonics and stereo", In *Proceedings of the 13th International Conference on Auditory Display (ICAD)*, Montréal, Canada (2007).