Sound Perception in 3D Virtual Environments: Application to a Roman Theatre of Hispania

Javier Alayón, José A. Romero-Odero(✉), Miguel Galindo, Francisco J. Nieves, and Sara Girón

Instituto Universitario de Arquitectura y Ciencias de la Construcción (IUACC), Departamento de Física Aplicada II, Escuela Técnica Superior de Arquitectura, Universidad de Sevilla, Avenida Reina Mercedes 2, 41012 Seville, Spain
{jap01, jrodero, mgalindo, nieves, sgiron}@us.es

Abstract. The aim of this work is to present the methodology implemented for the assessment of the human perception of sound and of the degree of acoustic comfort of occupants in an ancient Roman theatre. The evaluation is carried out through a visual and acoustic experience in a virtual environment. The textured 3D visual model of the space, and the binaural auralisations based on either on-site empirical measurements or on acoustic simulations, are displayed in a listening room designed with a very short reverberation time and low background noise. By means of sophisticated equipment for 3D virtual environment reproduction to groups of people, this listening room enables the physical ambience to be recreated of the Roman theatre of Cartagena, which is located in the southeast of Hispania (Spain). Groups of people can therefore subjectively assess the intelligibility of speech and the clarity for music of this open-air performance venue. The results accentuate the strong correlation between audio and visual perceptual aspects and contribute towards a more comprehensive understanding of the architectural aural experience.

Keywords: Human acoustic perception · 3D visual texturized model · Auralisation · Acoustics of Roman theatres

1 Introduction

The psychology of perception has empirically shown that the human being, from a very early age, has a preference for stimuli that have concordant and equivalent temporal patterns when these stimuli are multimodal, that is to say simultaneously perceived by different sensory channels [1]. Research in the field of neuroscience suggests that this similarity of temporary factors of experiences from different sensory channels that occur at the same time allows us to conceive the combination of these perceptual experiences as a whole. Likewise, these correspondences usually provide us with redundant perceptual information [2], thereby making it easier at the cognitive level to highlight properties of the stimuli, such as their duration, proportion, and intensity. This experience is called transmodal perception. Several authors in the field of cinematographic language have studied the perceptual unit that causes the coincidence of optical and acoustic information of any event [3].

T. Ahram et al. (Eds.): IHIET 2020, AISC 1253, pp. 216–222, 2021.
https://doi.org/10.1007/978-3-030-55307-4_33

According to Vorländer [4], auralisation is the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data. Auralisations are used in a variety of applications from virtual concert reproduction and archaeological acoustics to architectural design [5–7]. With regard to multimodal representations, numerous studies on the perception of a sonic environment have revealed the interaction between visual and auditory modalities [8]. For instance, the effect of vision on the perception of some spatial acoustic parameters, such as localization or distance, can be found in Calcagno *et al*. [9], while Postma and Katz [10] study the impact of the presence of visual feedback on the subjective assessment of auralisations.

Digital tools provide visualization and auralisation with unique opportunities in Architecture to support the reconstruction of damaged, non-existent, or destroyed spaces in order to optimize both the materials used in its construction and the configuration of the environment to obtain the best possible acoustic response. This visualization and auralisation is also of major interest for the objective and sensory evaluation of monuments of great heritage in which a strong compromise between preservation, social satisfaction, and tourist requirements arises.

This work combines auralisation from measured room impulse responses and anechoic materials and visualization of a textured 3D digital environment in order to appreciate the sound quality for speech and music of a theatre of the Roman Hispania (Spain): The Roman theatre of Cartagena. Listening tests are conducted with the subjects immersed in a room visualizing the texturized 3D model of the theatre and binaural reproduction through loudspeakers with crosstalk cancellation to evaluate the admissibility of the model and acoustic attributes of the environment such as distance, loudness, reverberation, clarity, and speech intelligibility.

2 Case Study

In Spain, there are 22 structures documented of classical outdoor theatres from Roman Hispania. The city of Cartagena, located in the southeast territory of the Iberian Peninsula, was founded as *Qart Hadasht* by the Carthaginian Asdrubal the Beautiful, in the year 227 BC atop a previous Iberian or Tartessian settlement. Following Carthaginian domination, it was taken by the Roman general Scipio the African in the year 209 BC who renamed it *Carthago Nova*. This territory had great strategic military and commercial value.

The Roman theatre, built in the time of Emperor Augustus, between the years 5–1 BC, remained in use until the third century. With an original capacity of 7,000 spectators, it remains one of the largest theatres in Roman Hispania. The theatre typology confirms the model proposed by Marco Vitruvius, as in most of the Hispanic theatres, and consists of: *Scaenae frons* (stage front) with double columns; *Orchestra*, formed by a semicircle in front of the stage in which the authorities sat; *Cavea*, in which the spectators stood according to their social rank of *ima*, *media*, or *summa*; *Proscaenium*, which consists of space in front of the stage; and *Porticus post scaenam*, made up of a porticoed courtyard behind the stage.

The stands are projected by taking advantage of the natural slope of one of the highest hills in the city. The central part of the *cavea* is excavated into the rock of the hillside itself, while the lateral flanks include vaulted galleries. The theatre is incorporated within the historical centre of the city near emblematic buildings and port facilities. The partial overlapping of the Old Cathedral on top of the Roman theatre, is valued as one of the historical singularities of this archaeological site (Fig. 1a).

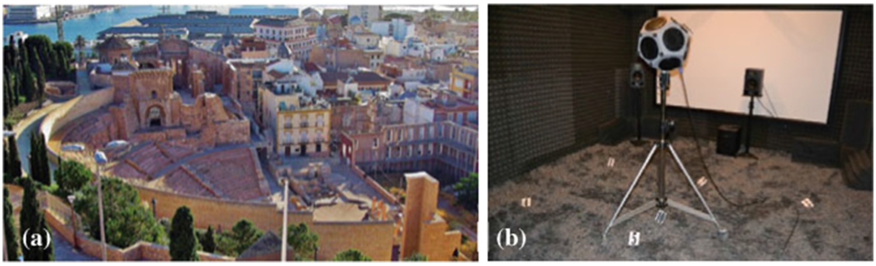

Fig. 1. (a) Aerial view of the Roman theatre of Cartagena. (b) Interior view of the listening room (Credits: (a) J. L. Sarralde, June 2019. (b) J. A. Romero-Odero, November 2019).

In order to register the binaural room impulse responses (BRIR), the theatre was excited via sine-swept signals in which the scanning frequency increases exponentially with time, produced by means of the EASERA v1.2 programme, through an AUBION x8 multichannel sound card. The generated signal was emitted through an AVM DO-12 dodecahedral sound source with a B&K 2734 power amplifier. At the reception point, the binaural RIRs were captured with a Head Acoustics HMS III torso simulator (Code 1323) and the B&K-2829 microphone polarization source [11, 12].

3 Listening Room

The acoustic perception tests were carried out in the listening room of the Acoustics Laboratory of the Applied Physics II Department of the School of Architecture of the University of Seville (Fig. 1b). The room, rectangular in shape (and of dimensions 5.1 × 7.5 × 3.0 m) in accordance with the ITU [13] requirements, is semi-anechoic, has a low background noise (L_{Aeq} = 24.6 dB), with an average mid-reverberation time of 0.2 s. The walls of the room have been treated with a foam pyramidal absorbent material and with bass traps in the corners. The ceramic floor is covered with a heavy carpet. Impulse response measurements were carried out by using three source positions and ten receivers (Fig. 1b). Table 1 shows the averaged values of the reverberation time for all source-receiver combinations at the six octave-band frequencies.

Table 1. Reverberation time (T_{30}) spatially averaged in octave band frequencies.

Frequency (Hz)	125	250	500	1000	2000	4000
T_{30}	0.34	0.24	0.22	0.18	0.10	0.07

The 3D visualization equipment of the room consists of the EH320UST OPTOMA projector, the ActivHub RF50 VOLFONI glass emitter (up to 10 glasses), and Contour Space static screen (2.6 × 1.46 m); the audio equipment consists of 2 ADAM A7X monitors on stands, and an ADAM UB8 Subwoofer. In order to improve the acoustic perception of listeners in the room, a crosstalk filter in Python has been applied to the reproduced audios. This filter enables the sound played by the right-hand speaker to only reach the listener's right-hand ear and vice versa. From the existing methods, a recursive crosstalk filter has been chosen in this work, which is easily implementable and reliable, since it takes into account the distance between sources, between ears, between source and listener, the shadow of the head, and the echo produced by the shoulders.

4 Visual Model of the Roman Theatre of Cartagena

In order to achieve a real closeness between the room user and the architectural space that makes up the case study, a 3D model of the theatre has been carried out by using the 2D digital planimetry provided by the Foundation of the Roman theatre of Cartagena and the geometrical measurements carried out during the on-site session. The 3D geometrical model is created by means of AutoCAD and Google SketchUp drawing tools based on 57,549 simple polygons, and by carrying out a material characterization of the model with the same or similar textures to those existing in the real building. In addition to the theatre, the model includes its immediate surroundings since it is incorporated inside the city (Fig. 2a).

The model is then imported into the Vizard 5-WorldViz programme for 3D visualization on the screen via the glass emitters. In this programme, the filtered audio and the various lights employed are introduced in order to attain good visibility after immersion into the model. With the system mounted and executed, the immersion of the user in the case study emulates the real experience without being physically in situ.

5 Auralisations and Survey

Auralisations have been performed by convolving with Matlab the measured BRIRs corresponding to three source-receiver combinations (the source in the centre of the *scaena*, and three receivers at the *scaena*, the *ima*, and *media cavea*, respectively) with excerpts of 25 s from anechoic recordings of female and male speech and from a Baroque music piece from Joan Baptista and Josep Pla [14], played by the Baroque Orchestra of Seville.

The perception questionnaire consists of an introductory section on personal data (gender, age, profession, and whether he/she has a musical and an acoustic background), and another two sections of nine questions each. In each question, the respondent hears a pair of stimuli of the same type in two different positions of the theatre, and has to decide whether he/she hears clearer (or more intelligibly) stimulus 1, stimulus 2, the two stimuli equally, or whether he/she does not know or fails to answer (DK/FA).

The last section of the survey is carried out in the same way as the second except that it incorporates the visualization of the 3D model in the various listening positions.

Due to the characteristics of the listening room and the equipment used, the survey could be successfully carried out on groups of three people (Fig. 2b). The total sample was made up of 18 students (9 women and 9 men), between 19 and 38 years old (14 without music studies, 3 with music studies, and 1 with acoustic studies). The approximate duration of the survey is 30 min.

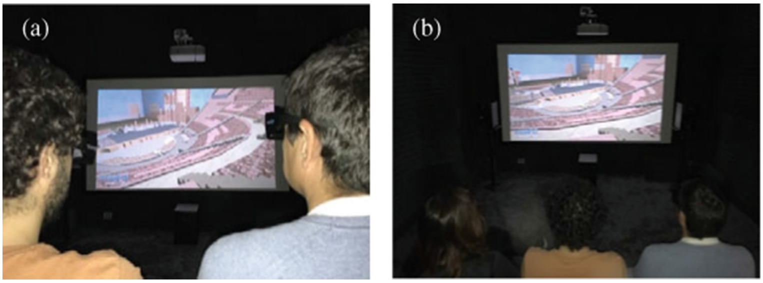

Fig. 2. (a) Texturized model of the Roman Theatre of Cartagena. (b) Participants during the survey (Credits: (a), (b) J. A. Romero-Odero, March 2019).

6 Results and Conclusions

Comparison tests of two signals belonging to various theatre locations have been established. These locations correspond to three points whose distances from the source are 9.88, 20.29, and 37.08 m respectively.

In the first part of the survey, the participants receive only the pairs of musical or oral sound signals without the support of the location visualization in the theatre. When the pairs of musical signals are heard, most respondents perceive the clarity for music as better for the closest signal than the farthest; however, when listening to the two closest positions or the two farthest positions for comparison, the number of people who choose one option or the other is almost equivalent (Fig. 3a). When the same signals are supported by 3D visualization, there is a greater preference for positions closest to the source (see Fig. 3b). For sound perception, it is therefore crucial to visualize the space and the location within this space.

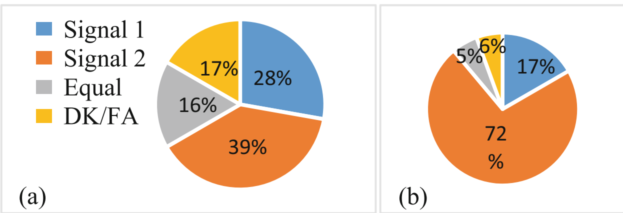

Fig. 3. Answers of the respondents when pairs of musical signals are compared for the two positions furthest from the source. (a) Only audio signals; (b) Audio signals and visualization. Signal 2 is that from the receiver closest to the source.

In order to determine the subjective limit, which is the just noticeable difference (JND) related to the speech intelligibility that a listener distinguishes in this type of venue, the same test is carried out with signals corresponding to a man and a woman individually delivering the same speech. In these cases, regardless of whether the same speech is spoken either by a man or a woman, and whether or not the theatre localization is viewed in 3D, most respondents prefer the locations closest to the source to the most distant.

This is a preliminary study in order to determine the JND values corresponding to the main acoustic parameters in these semi-open venues. To this end, it will be necessary to expand the number of respondents, to consider their possible differences in terms of gender, age, and musical education, and to increase the various acoustic parameters involved and their ranges of values. In addition, the study will be carried out using virtual reality immersion.

Acknowledgments. This work is funded by ERDF funds and the Spanish Ministry of Economy, Industry, and Competitiveness (MINECO) with ref. BIA2017-85301-P.

References

1. Bahrick, L.E.: Infants' intermodal perception of two levels of temporal structure in natural events. Infant Behav. Dev. **10**, 387–416 (1987)
2. Bahrick, L.E., Lickliter, R.: Intersensory redundance guides attentional selectivity and perceptual learning in infancy. Dev. Psychol. **2**(36), 190–201 (2000)
3. McGurk, H., McDonald, J.: Hearing lips and seeing voices: a new illusion. Nature **264**, 746–748 (1976)
4. Vorländer, M.: Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. Springer-Verlag, Berlin, Germany (2008)
5. Postma, B., Katz, B.: Creation and calibration method of virtual acoustic models for historic auralizations. J. Virtual Real. **19**, 161–180 (2015)
6. Sender, M., Planells, A., Perelló, R., Segura, J., Giménez, A.: Virtual acoustic reconstruction of a lost church: application to an order of Saint Jerome monastery in Alzira. Spain. J. Build. Perform. Simul. **11**(3), 1–22 (2017)
7. Álvarez-Morales, L., Molina-Rozalem, J.F., Girón, S., Alonso, A., Bustamante, P., Álvarez-Corbacho, A.: Virtual reality in church acoustics: visual and acoustic experience in the cathedral of Seville, Spain. In: International Congress on Sound and Vibration, pp. 1–8 (2017)
8. Stein, B.E., Meredith, M.A., Honeycutt, W.S., McDade, L.: Behavioral indices of multisensory integration: orientation of visual cues is affected by auditory stimuli. J. Cogn. Neurosci. **1**, 12–24 (1989)
9. Calcagno, E.R., Abregú, E.L., Eguía, M.C., Vergara, R.: The role of vision in auditory distance perception. Perception **41**(17), 175–192 (2012)
10. Postma, B.N., Katz, B.F.: The influence of visual distance on the room-acoustic experience of auralizations. J. Acoust. Soc. Am. **142**(5), 3035–3046 (2017)
11. ISO 3382-1:2009, Acoustics - Measurement of room acoustic parameters -I: Performance spaces. ISO Standard: International Standards Organization, (2009)

12. Álvarez-Corbacho, A., Bustamante, P., Galindo, M., Girón, S., Zamarreño, T.: Measurement and analysis of the acoustics of the Roman theatre of Segobriga (Spain). In: Internoise 2019, paper 1833, 11 p. (2019)
13. ITU-R BS.1116-1. Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, Geneva, Switzerland (1994)
14. Girón, S., Galindo, M., Gómez-Gómez, T.: Assessment of the subjective perception of reverberation in Spanish cathedrals. Build. Environ. 171, paper 106656 (2020)