

Significance of Binaural Synthesis on the Spatial Acoustic Impression of Enclosures

Alicia ALONSO⁽¹⁾, Diego Mauricio MURILLO-GÓMEZ⁽²⁾

⁽¹⁾ *Instituto Universitario de Arquitectura y Ciencias de la Construcción*
Av. Reina Mercedes, 2, 41012 Seville, Spain; e-mail: aliciaalonso@us.es

⁽²⁾ *Institute of Sound and Vibration Research, University of Southampton*
University Road, Southampton SO17 1BJ, United Kingdom

(received October 26, 2017; accepted June 7, 2018)

Listening tests have been carried out to quantify the significance of binaural auralization over monaural auralization in accordance with the acoustic properties of the enclosure. To this end, acoustic rendering of three different rooms were generated based on synthesized monaural (two channels with the same audio material) and binaural room impulse responses. The auralizations were evaluated by means of subjective tests using headphones with non-individualized equalization. Parameters, such as localization, spatial impression and realism, were taken into consideration to determine the relevance of providing binaural information for the auralization of a given room. The analysis of the data has been conducted following a statistical approach based on ANOVA and Pearson correlation. The results indicate that spatial perception is strongly dependent on the acoustic characteristics of the rooms and on the listening condition of the audio material. Furthermore, as expected, advantages of binaural rendering in terms of source localization was also confirmed.

Keywords: architectural acoustics; acoustic simulation; auralization; listening test; spatial impression.

1. Introduction

Human hearing can be regarded as binaural since the acoustic pressure on the eardrums contains all the auditory information that is required for the brain to generate an acoustic image of the environment (MOLLER, 1992). For auralization applications, various studies have shown the advantages of providing binaural information as compared to monaural information (ROWAN *et al.*, 2015). These benefits are mainly related to the ability to recreate a three-dimensional acoustic rendering of the space, which leads to a more realistic synthesis of the sound field of the room (LEHNERT, BLAUERT, 1992).

Furthermore, the auditory spatial impression has been studied for several decades (BARRON, MARSHALL, 1981). Certain authors have demonstrated that this impression can be divided into several perceptual properties, such as spaciousness, size impression, and reverberation (BECKER, SAPP, 2001; BLAUERT, LINDEMANN, 1986). The subjective parameters related to spatial impression include the Apparent Source Width

(ASW) and the Listener Envelopment (LEV). The former is defined as the perceived spatial extent of the sound source which determines its left and right virtual boundaries (GRIESINGER, 1997). The latter refers to the listener's sense of being surrounded or enveloped by sound (SOULODRE *et al.*, 2003). These parameters enable the description of the sound field based on subjective terms.

ASW is mainly associated to the level of the early lateral reflections measured at the listener's ears within the first 80 ms after the arrival of the direct sound (BRADLEY, SOULODRE, 1995a). In contrast, LEV refers to the late lateral sound level (reverberant field), which depends on the type of space and it is generally assumed to start 80 ms after arrival of the direct sound (for reverberant spaces) (BRADLEY, SOULODRE, 1995a; OKANO *et al.*, 1998). These attributes can be correlated with other objective parameters derived from the early/late part of the energy decay curve.

Objective acoustic parameters are also considered in this analysis so that relations between subjective qualities and physical objective parameters can be es-

tablished. In this regard, CERDÁ *et al.* (2013) exposed the possibility of calculating objective measures from the musical characteristics of auralized audio signals by carrying out statistical correlations with audio features.

The aim of this work is to evaluate the relevance of providing binaural information over monaural information for the auralization of a room, by taking into consideration the influence of the stimuli type (speech vs. music) and the size of the room (small, medium, and large). This is carried out by investigating various perceptual aspects such as source localization, spatial impression, and sense of realism.

The paper is organized as follows: Sec. 2 provides a brief state of the art description or the evolution of the research of spatial impression attributes; Sec. 3 describes the methodology used in this research, including measurements, modeling, and listening tests. The statistical significance between main and interaction effects, such as the correlation between perceptual and objective parameters, is considered in Sec. 4. Finally, in Sec. 5, the conclusions are presented.

2. Evolution of spatial impression attributes

The attribute of spatial impression has been exhaustively studied for almost 50 years. Researchers have improved the knowledge of certain subjective characteristics of sound perception by conducting experiments that showed multidimensional relationships between subjective characteristics and acoustical criteria (FISCHETTI *et al.*, 1992). One example is the reverberant sound, which is commonly associated to one of these attributes, namely the LEV.

In general, acoustic measurements or simulations are carried out to assess the objective properties of a given sound field. Room acoustic modeling techniques have been successfully applied for more than 20 years for the prediction of measurable quantities associated to subjective parameters, such as listening preference or spaciousness (VAN DORP *et al.*, 2013). However, although physical quantities are able to describe the rooms and their acoustic fields, no consensus has been reached on physical measures that can accurately predict the perception of spatial aspects of sound fields. Barron and Marshall confirmed the existence of a measurable parameter called the Lateral Fraction (LF), which was shown to be highly correlated with source broadening (BARRON, MARSHALL, 1981). OKANO *et al.* (1998) presented musical signals using a multichannel system in an anechoic environment. The outcomes indicate that ASW was not only related to the interaural cross-correlation (IACC_e) and lateral fraction, but also to the absolute sound pressure level at low frequencies. Other work developed by NOWAK (2010) determines the correlation between objective measures, such as inter-aural level difference

(ILD), inter-aural time difference (ITD), IACC and subjective attributes, in order to establish an objective quality measure for sound fields.

Furthermore, a strong influence of the late lateral energy and the level of lateral reflections on the sense of envelopment was found by BRADLEY and SOULODRE (1995b). Firstly, it was proven that late arriving sound from directions corresponding to the front of the azimuthal plane has an effect on perceived listener envelopment. It was also observed that, in practice, perceived properties were influenced by the size of the musical sources (ASW), and that both ASW and LEV increase with the SPL (BARRON, 2000; KUTTRUFF, 2000). Other studies have analysed the influence of height-channel contents on enhancing spatial impressions through virtually elevated signals of sound sources (KIM, 2013).

Psychoacoustic approaches have been also considered. BLAUERT and VORLÄNDER carried out research into auditory perceptual attributes by carrying out some psychoacoustic analysis (BLAUERT, LINDERMANN, 1986; VORLÄNDER, 2008). On the other hand, GRIESINGER (1997) analysed the psychoacoustics of ASW and LEV in performance halls by presenting a hypothesis for the origins of the perception of spatial impression. He investigated how the hearing experience of the listener can extract spatial information from the type of stimuli, whether it be music or speech, based on the main characteristics of the motif (level, rise times, fall times, spaces between notes, etc.).

Other researchers resorted to testing the subjective listening experience with the aim of showing preferred conditions or to quantify ASW and LEV. Experiments were conducted in order to reveal whether objective parameters were correlated with the perceptual attributes (VIGEANT *et al.*, 2011), or to propose a framework that extends an open-source listening test software by reporting methods for spatial attributes of sound sources (WESTPHAL *et al.*, 2015). In terms of enclosures, most of the existing research has focused on concert and opera halls, but for churches, the amount of literature is more restricted. MARTELOTTA (2008) described the results of an AB test aimed to investigate the preferred subjective listening conditions inside churches by assessing diverse musical motifs. One of the outcomes is the relationship between subjective ratings and objective acoustical parameters (LF and IACC).

Alternatively, the quality of synthesized sound fields have been widely enhanced in the last decades trying to resemble as close as possible the real ones. In this sense, a new study based on listening tests to compare different type of spaces were investigated. It was proved that those tested qualities were not significantly influenced by the application of a perception-based simplification algorithm for the reduction in the number of early reflections (HACIHABIBOGLU, MURTAGH,

2008). However, there is a lack of studies that jointly address the acoustic perception in relation to factors that link the listening condition with the stimuli and the enclosure.

3. Methodology

A description of the procedure to evaluate the relevance of providing binaural information for room acoustic auralizations is shown as follows: Three spaces of different size were selected to analyze perceptual attributes by means of listening tests. Firstly, acoustic measurements were carried out based on the guidelines given by the standard ISO 3382-1 (2010). Synthesized room impulse responses (RIRs) were then predicted using the commercial Catt-Acoustics V9 package (DALENBÄCK, 2011). The acoustic models were calibrated in terms of reverberation time (T_{30}), Early Decay Time (EDT), and clarity index (C_{80}) using the data obtained from the measurements.

Subsequently, auralizations were generated by involving anechoic audio material with the synthesized monaural and binaural room impulse responses. Finally, subjective evaluations were carried out by performing listening tests to determine the statistical significance between the main and interaction effects of the evaluated factors. In the case of the monaural material, the auralization corresponds to a signal of two channels with identical audio content.

3.1. Enclosures considered

Figure 1 illustrates the spaces taken into consideration for the experiment. The enclosures were selected such that the perception of a listener would be revealed when a given attribute is rated in a small room (MR – meeting room), a medium-sized hall (IM – Ightham Mote), or a large reverberant cathedral (CA – cathedral).

To this end, the spatial impression was assessed to determine whether its correspondence to objective parameters, such as apparent source width (ASW) and listener envelopment (LEV), is the same for spaces whose dimensions and styles substantially differ. Furthermore, the relation between the geometrical char-

acteristics of the spaces and the sense of spatial impression was also considered.

The first room corresponds to a regular meeting room located in the Institute of Sound and Vibration Research of the University of Southampton, United Kingdom. It has a floor surface of approximately 30 m^2 and a volume of 90 m^3 . It is composed of two large areas of glazing, with walls covered with plaster, and a carpeted floor. The second space selected is the Great Hall of the Ightham Mote, a moated medieval manor house built in the 14th century, located in Kent, United Kingdom. The interior consists mainly of rough finished block and highly engraved wood including a large decorated fireplace and two large glass windows. The volume of the hall is about 430 m^3 , and the floor surface is 50 m^2 . The third space is one of the most important buildings of the Spanish Renaissance: the Cathedral of Granada. The Cathedral was built in the sixteenth century and it is mainly composed of an ambulatory and five naves. The naves are divided by stone-clad columns, which support 37 stone vaults. It has a floor surface of 6678 m^2 (63 m wide and 106 m long) and the volume is approximately $160\,500\text{ m}^3$.

3.2. Acoustic measurements

Acoustic measurements were independently conducted by two research groups following the guidelines given by the ISO 3382-1 (2010): the meeting room and the Ightham Mote were measured by researchers from the Institute of Sound and Vibration Research, whereas the Cathedral of Granada was characterized by researchers of the University of Seville. The acoustic measurements were conducted to support the modeling and calibration stage.

In the case of spaces (MR) and (IG), the procedure was carried out using an omnidirectional microphone (Brüel & Kjær Type 4189-L001) and a studio loudspeaker (Mackie 824 MK2). The excitation signal was an exponential sine-sweep from 20 Hz to 20 kHz. The directivity of the loudspeaker was measured in an anechoic chamber in order to include its directional pattern in the acoustic software. In the case of the Cathedral (CA) a sine-sweep signal was emitted from an omnidirectional dodecahedral sound



Fig. 1. Spaces under study: meeting room, Ightham Mote, and cathedral, respectively.

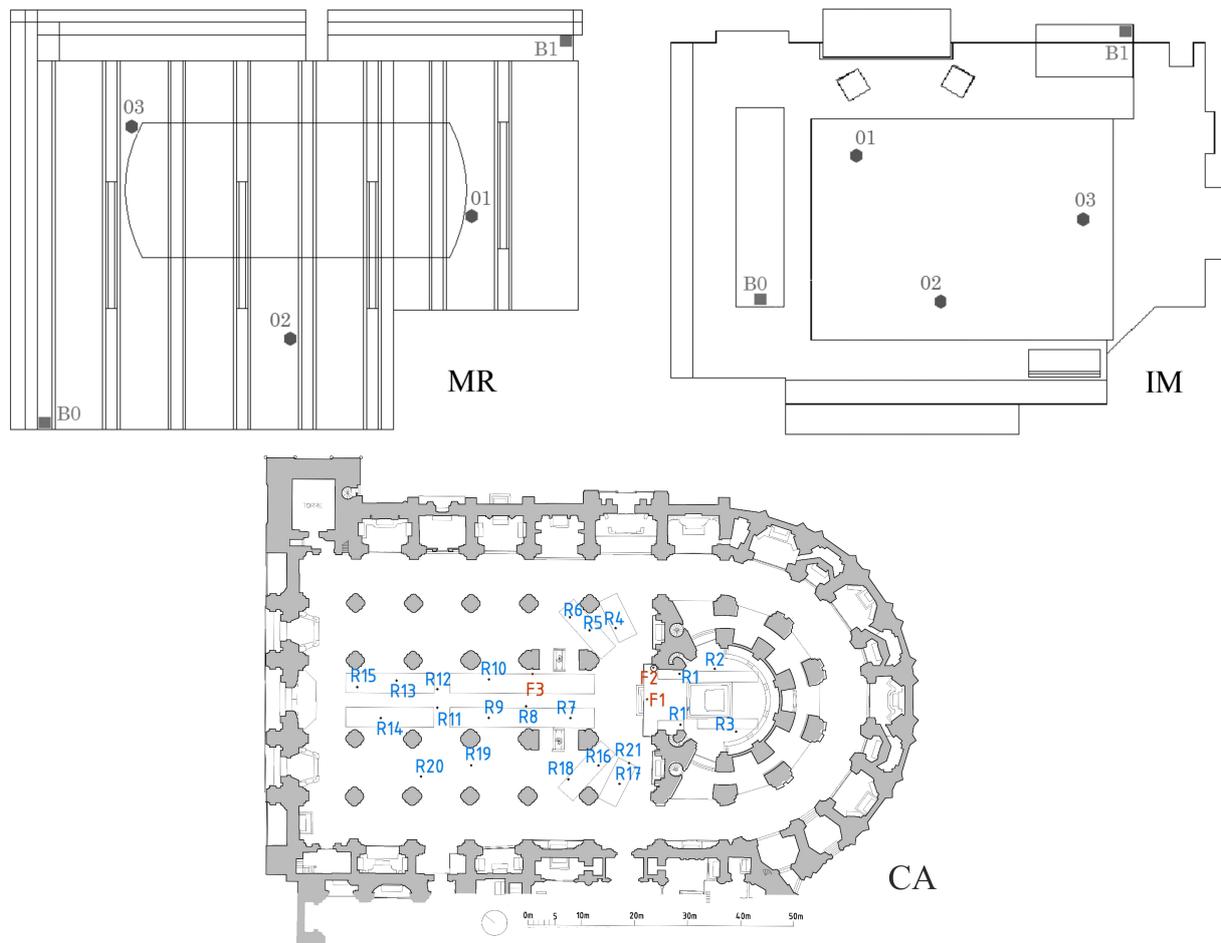


Fig. 2. Source-receiver paths for each space.

source (AVM DO-12) and the impulse responses were recorded using a multi-pattern microphone (Audio-Technica AT4050/CM5). The duration of the sweep was set to 20 s and covered the octave bands from 63 to 16 000 Hz. Figure 2 illustrates the positions of the source and receivers for each space.

Due to the differences in the setups used for the two research groups, auralizations from measured impulse responses were not considered. Instead, the acoustic measurements were selected as references to calibrate numerical models of each room, which were generated based on the same specifications. This leads to a controllable setting where synthesized room impulse responses can be used for auralization and comparison purposes. Furthermore, the use of acoustic models allows additional source-receiver paths to be generated thereby increasing the flexibility of the experiment. The simulations were performed using the commercial package CATT-Acoustics v9 (DALENBÄCK, 2011).

3.3. Acoustic simulations

Regarding the boundary conditions, absorption and scattering coefficients were extracted from CATT-

Acoustics library data and the scientific literature (VORLÄNDER, 2008; PTB Database; COX, D'ANTONIO, 2004). The selection of these materials was made according to a visual inspection of the surfaces of the enclosures.

The calibration process corresponded to the implementation of an iterative algorithm that compares the difference between the measured and predicted reverberation time by changing the absorption coefficients of the materials. Furthermore, other acoustic parameters, such as EDT and C_{80} , were compared using the just noticeable difference values specified in ISO 3382-1 (2010) and those determined for more reverberant spaces, such as churches, where the subjective threshold perception can differ (MARTELLOTTA, 2010). The use of monaural parameters as a reference for a calibration process is a well-established approach in room acoustics for the evaluation of the reliability of simulations (BORK, 2005; SILTANEN *et al.*, 2008; FOTEINOU *et al.*, 2010; RINDEL, CHRISTENSEN, 2004). Amongst the three models, the most complex calibration process was developed for the model of the Cathedral (CA) due to the great dimensions and the particular characteristics of the geometry the materials. This

process is detailed in ALONSO *et al.* (2017). The reader is referred to (MURILLO *et al.*, 2014; MURILLO, 2016) for a more complete explanation of the calibration process of the Ighthan Mote (IM) and the meeting room (MR), respectively.

Table 1. Acoustic parameters (spatially averaged). MR – meeting room; IM – Ighthan Mote; CA – Cathedral; m – measured; p – predicted.

		T_{20} [s]	EDT [s]	C_{80} [dB]
MR	m	1.09	1.03	3.3
	p	1.08	1.06	2.7
IM	m	0.92	0.87	5.9
	p	0.95	0.89	5.6
CA	m	8.97	8.4	-5.8
	p	9.16	8.96	-6.9

Furthermore, the mean square errors (MSE) of the frequency response in 1/3 octave band resolution have been calculated for the “best” and “worst” source-receiver paths in terms of the agreement with the acoustic parameters. The mean square error is defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - Y_i)^2,$$

in which n is the number of predictions, and \tilde{Y}_i and Y_i are the predicted and measured data in the i -th 1/3 octave band, respectively. Table 2 illustrates the MSE for the enclosures considered.

Table 2. Mean Squared Error of the 1/3 octave frequency band response.

Enclosure	Source-receiver path	MSE [dB]
MR	B0-R1	2.4
	B0-R3	2.6
IM	B1-R1	2.8
	B1-R2	2.4
CA	B0-R5	3.3
	B0-B8	3.5

3.4. Auralization setup

The auralizations were presented to the listeners using a pair of AKG K-702 headphones. A non-individual headphone equalization was applied at the end of the signal processing chain to reduce the frequency coloration produced by the transducers. The headphone equalization was carried out based on the methodology proposed by MASIERO and FELS (2011).

To this end, measurements of Headphone Impulse Responses (HPIRs) were conducted using a Neumann dummy head KU100 in an anechoic chamber. Figure 3 shows the measured headphone frequency responses for

the two channels (left and right). 0 dB corresponds to the magnitude measured at 1 kHz.

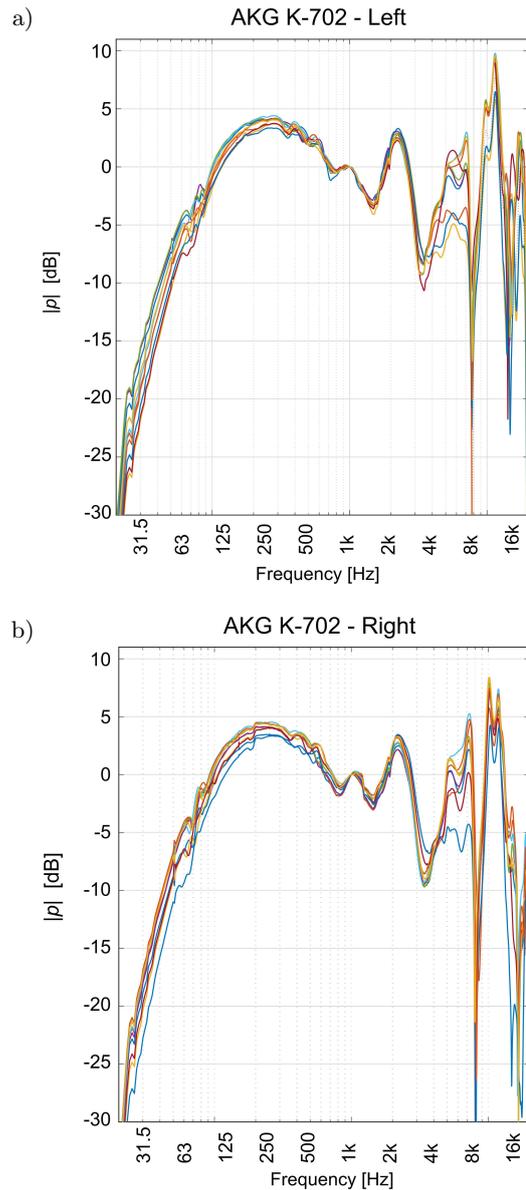


Fig. 3. Measured headphone and equalized frequency responses, AKG K-702, for the two channels: a) left and b) right. The variety of color-lines corresponds to different trials of the headphone measurements.

The assumption made in Masiero and Fels’s approach is that peaks in the inverted frequency response are the most subjectively disturbing component. Therefore, the aim is to reduce the notches in the frequency response to be inverted and thus minimize the resulting peaks of the inverted response. Assuming that the measured headphone impulse responses are normally distributed, the curve obtained from the mean μ plus 2 times the standard deviation σ of the magnitude of the measured frequency responses rises above the magnitude of all the measured frequency responses with over 95% chance. This

approach yields a relatively smooth curve and is more robust to outliers and the yielded by selecting the maximum of the magnitude of all the measurements. Finally, the phase is synthesized using minimum phase filters. The inverted frequency response is calculated as

$$H(\omega)^\dagger = \frac{H(\omega)^*}{H(\omega)^* H(\omega) + \beta},$$

in which $H(\omega)^\dagger$ is the inverted frequency response, $(\cdot)^*$ indicates the complex conjugate operation, and β is a regularization parameter implemented to reduce the boost produced by the inversion at low frequencies. Figure 4 shows the frequency responses and their inverted versions.

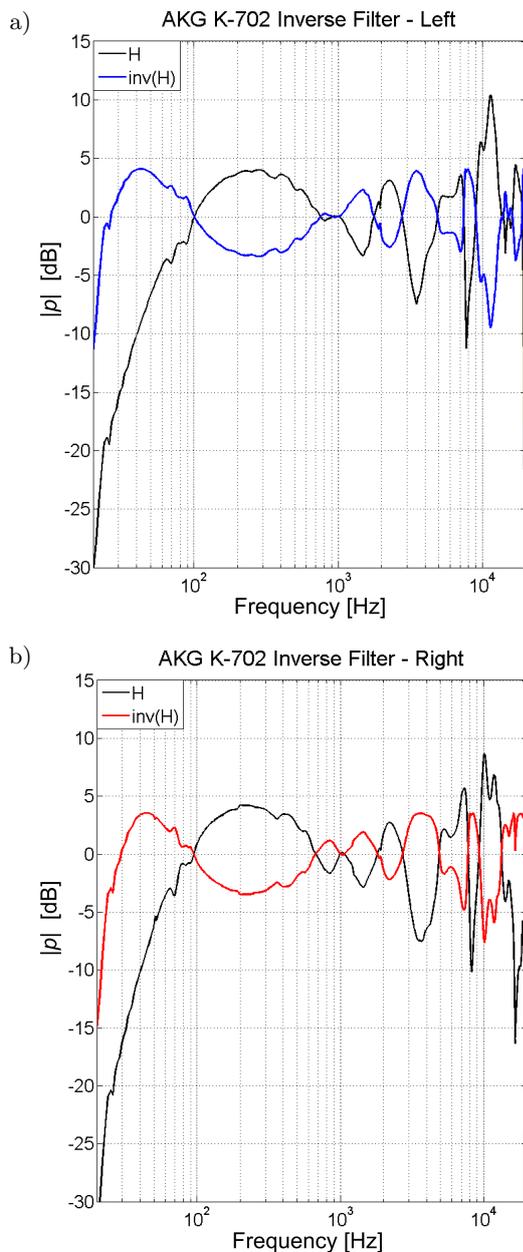


Fig. 4. Inverted headphone frequency responses, $\beta = 10 \cdot 10^6$ AKG K-702.

3.5. Subjective experiments

A set of listening tests was performed to analyse whether the typology of the space and the type of stimuli exert a noticeable effect on hearing perception when a binaural signal is reproduced as compared to a monaural signal in which the two channels reproduce identical audio content. The difference was investigated using parameters that includes localization, spatial impression and realism.

3.5.1. Factors to be evaluated

The evaluated factors are classified into three main groups: the size of the space, the stimuli, and the listening condition. The space is composed of three enclosures of different sizes: a meeting room (MR, small), Ightham Mote (IG, medium-sized); and Cathedral (CA, large). Two types of anechoic audios were used as stimuli: a female speech and a music signal. Finally, the listening condition corresponds to monaural or binaural signals. The combination of these groups yielded a total of twelve factors to be evaluated.

3.5.2. Listening room, anechoic material, and participants

The listening tests were conducted in the AudioLab of the ISVR, an acoustically isolated room used for sound-field synthesis applications. The anechoic material that was selected for the study corresponds to: i) a brief speech that was previously recorded in the anechoic chamber by a female member of the University of Southampton; ii) a music piece played by a solo of *celloist*, which corresponds to an excerpt of the piece *Old Gavotte*, composed by Martini. The anechoic audio was obtained from the library of CATT Acoustic software that contains pieces from B&O Music for Archimedes CD, recorded to support the experimental work of the Archimedes Project, funded under the European EUREKA scheme. Related to the selection of the musical sample, it should be noted that the choice was not an easy task, due to the wide range and variety of musical pieces. However, it was decided to choose a single instrument in order to adapt, in the most appropriate way possible, the musical piece to the spaces studied. In this regard, the performance of an orchestra or a choir in a small meeting room was not realistic, whereas instrumental solos are currently performed in large spaces such as cathedrals.

The length of each stimulus was 10 s and the test was fully randomized, including by changing the playback sequence. The number of repetitions of each audio was a choice determined by the participants, who were allowed to listen to the stimulus as many times as they considered suitable in order to decide on a robust judgment for each question of the test. The test samples of the stimuli presented were normalized in order to have approximately the same loudness by evaluating the

RMS level, established at -35.4 dB_{FS} (Full Scale). This process could influence the perception of the source-receiver distance, however, it should be noted that, in this work, only subjective quality measures in terms of localization and spatial impression are considered, which decreases the importance of whether the source is closer or farther away, as long as the levels are adjusted. A sample of 15 subjects, aged between 23 and 60 years old with normal hearing, was used for the listening tests. An equitably distributed sample, was sought, and was therefore composed of 46% professionals with experience in audio who worked in the ISVR itself, and of 54% of inexperienced subjects with mixed musical background and references.

3.5.3. Listening test procedure

The procedure of the subjective experiments was divided into four main parts: the first part investigated the localization performance. The second and third parts were related to the listeners' spatial impression, specifically to ASW and LEV. Finally, the last part was associated to the perception of realism. Regarding localization, participants were required to determine the position in the azimuthal plane of a virtual source based on a given auralization setup. A semicircular arc from -45° to 45° with a resolution of 5° was used as reference for the localization of the source. Figure 5 illustrates the setup built for the test. Three virtual positions of the stimuli sound source were located at 0° , -20° and 20° on the azimuth plane, respectively.

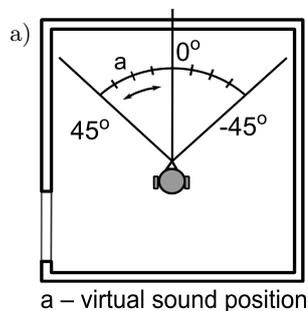


Fig. 5. Plan of the setup for the virtual source localization experiment.

The second part consisted of a subjective scaling question about ASW, defined for the participants as the width of the virtual sound source. The task for the listeners was to specify the virtual left and right edges of the sound source, that is to say, to determine how wide the source was perceived in the given auralization scenario. The subjects had to rate the attribute ASW on a scale from 1 (narrow) to 5 (broad). Subsequently, in the third part, participants were asked about LEV, defined for them as the sense of being surrounded or enveloped by sound, namely the envelopment sensation that was evoked by the sound field. In this case, listeners had to evaluate how surrounded they felt when they heard the auralization. The subjects had to rate the attribute LEV on a scale from 1 (not enveloping) to 5 (fully enveloping). Both concepts were explained by providing an example in order to ensure that all listeners understood their meanings in a similar way.

Finally, an audio-visual test was conducted to evaluate the sense of realism between six pairs of auralizations. This was performed by an AB comparison where each signal corresponded to auralizations respectively convolved with monaural or binaural RIRs of the same space and the same anechoic recording. A picture of the space where the virtual sound of the auralization is performed was also provided for the participants. Listeners were asked to rate which of the two audio signals generates a more realistic rendering of the acoustics of the space represented in the picture. The rated scale ranges from “much more realistic” to “no difference”.

4. Results

4.1. Method of analysis

The room acoustic perception has been evaluated by means of univariate and multivariate statistical techniques using the commercial package SPSS v.23. The analysis of variance of repeated measures is frequently considered in subjective preference studies. This type of analysis allows the statistical differences between different main factors to be explained when various rated attributes are evaluated (RUTHERFORD, 2001). The design of this method is characterized by including a number of observations per subject, whereby each observation is obtained under different experimental conditions. In the current research, each observation corresponds to a specific auralization condition.

A general linear model of repeated ANOVA measures was performed by assuming that the subject's factors are randomly selected, while the other within-subjects factors are evaluated by a fixed-effects model. The main factors determined were: the listening condition (binaural vs. monaural), the stimulus type (speech vs. music), and the size of the room (small, medium-sized, and large). All main effects and their interactions with each other were tested. Significant levels were con-

siderer for a p -value lower than 5%. This methodology enables the variance to be analysed when the experiment involves more than one within-subject factor and interaction effect. The analysis of variance aims to determine the statistical significance of binaural rendering by using the so-called Snedecor F distribution.

4.2. Localization

Border angles ($\pm 45^\circ$) were not selected because the difference between the binaural and monaural signals would be too obvious for the listener. Figure 6 shows the determination of the case of -20° position of the virtual sources in accordance with various levels of main factors.

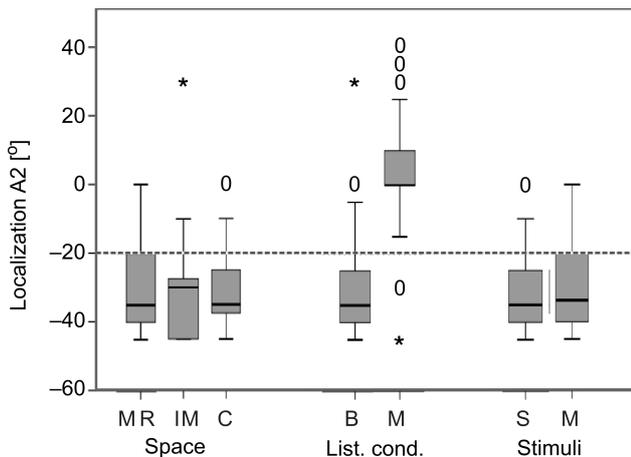


Fig. 6. Responses of the subjects for localization of -20° based on main factors. Box-whisker plots show the first and third quartiles, median of the data, the minimum and maximum values, and the outliers. Black line indicates the real position of the virtual source. Binaural results are presented for the space and stimulus factor. MR – meeting room, IM – Ightham Mote, CA – Cathedral; B – binaural, M – monaural; S – speech, M – music.

As expected, the results confirm that the localization is totally lost when the subjects listen to monaural signals. In general, participants were able to equally identify the incoming direction of the sound even in highly reverberant spaces such as the Cathedral. However, listeners felt a more lateralized source than the synthesized one, that is to say, they perceived approximately a -35° position in almost all levels of the main factor, instead of -20° , an angular range determined mainly for the space of greater volume, where the great reverberation makes the exact determination of the sound-source position more difficult. The reason may be attributed to the lack of a reference and the use of a generic HRTF, which may induce an increased number of errors in the median plane (MOLLER *et al.*, 1996). Nevertheless, it could be said that neither the stimuli nor the space have a significant effect on the localization. In addition, the objective of this work is

not focused on assessing the accuracy of positioning the sound source, but on the ability to determine the angular range of origin of the signal when binaural and monaural signals are compared.

4.3. Apparent source width

Firstly, the analysis of multivariate within-subject tests obtained from SPSS software reveals that the main effects of space and listening conditions on the ASW parameter was found to be statistically significant between their levels ($F(1.29, 18.08) = 13.78$, $p = 0.001$) and ($F(1, 14) = 26.16$, $p = 0.00$), respectively, but not in the case of Stimuli ($F(1, 14) = 2.89$, $p = 0.11$). Secondly, results of the subjective tests also reveal that the two-way interaction of factors between space-stimuli, space-listening, and stimuli-listening conditions are not significant since the null hypothesis regarding the similarity of variances is not violated ($p > 0.05$). However, a significant tendency was found between the three-way interaction of factors (space-stimuli-listening condition) at ASW ratings ($F(2, 28) = 5.01$, $p = 0.014$). This fact means that, on considering the space as the fixed factor, ASW perception varies if the levels of stimuli and listening condition are different, and it is not the same when it is convolved using a music or speech signal. Neither is it perceived in the same way for binaural nor monaural signals. Thus, it is possible to determine a significant difference between the mean values of the three factors. However, pairwise comparisons are statistically analysed to obtain more information on the experiment.

On the one hand, statistical results of the interaction between space-stimuli show that, when speech signal is considered, the difference between the variances of the most reverberant space (Cathedral) and the either of other two rooms (meeting room or Ightham Mote), were significant ($p < 0.05$). In addition, the relationship between Cathedral and Ightham Mote is no longer significant when a music signal is considered, that is to say, a Music signal is perceived as broader than a speech signal (GRIESINGER, 1997).

On the other hand, one of the highlights of this comparative analysis is the listeners' perception of the Cathedral space, regardless of the Stimuli. In this regard, as shown in Figs. 7a and 7b, negligible differences between the smallest space "room" and the most voluminous "cathedral" are found when the signal is binaural ($p = 0.928$). However, differences are statistically significant if a monaural signal is considered due to the high subjective ratings of the Cathedral for monaural signals. Thus, the results indicate that, except for huge reverberant spaces, ASW ratings are higher when the listeners perceive a binaural signal compared to a monaural signal.

In fact, when analysing the difference of means between the two levels of the listening condition factor

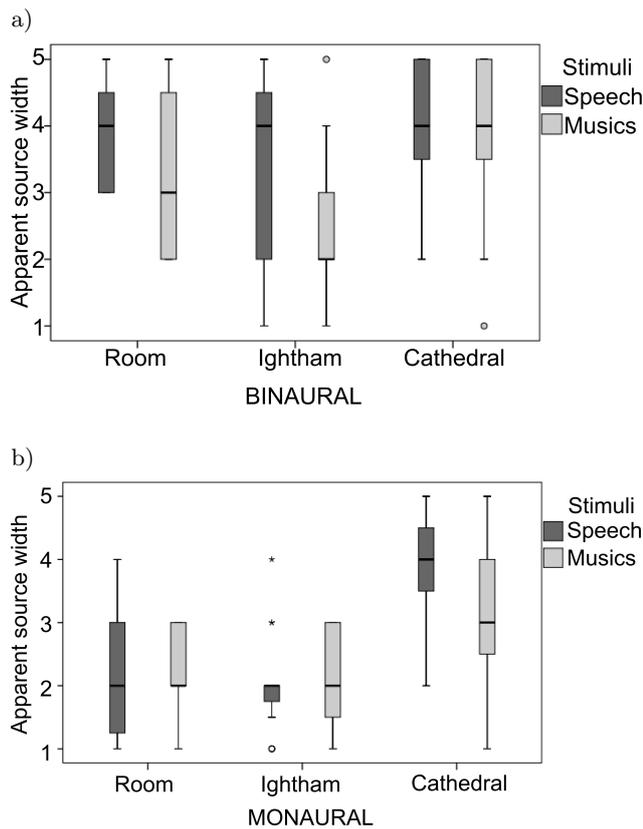


Fig. 7. Responses of the subjects for ASW: box-whisker plots show the first and third quartiles, the median of the data, the minimum and maximum values, and the outliers: a) binaural, b) monaural.

(binaural and monaural), mean values were similar in the Cathedral ($p = 0.531$), while for other combinations, there is a clear difference of mean values (mean diff. (I-J) = 1.7, $p = 0.000$).

All these statements confirm the significance of binaural synthesis since the perception of apparent source width is lost in the case of listening to monaural signals.

4.4. Listener envelopment

In the case of listener envelopment (LEV) analysis, the values of the three main factors show significant differences in the multivariate model (3-way interaction): space ($F(2, 26) = 15.54$, $p = 0.000$), stimuli ($F(1, 14) = 6.48$, $p = 0.023$), and listening condition ($F(1, 14) = 26.16$, $p = 0.000$). This tendency also remains for the 2-way interactions. In particular, the interaction between stimuli-listening condition is not significant since the evolution is similar for each space as independent factors ($p = 0.069$, see Figs. 8a and 8b).

Consequently, post-hoc multiple comparisons revealed that values between the Room and Cathedral were statistically significant, both for speech and music ($p = 0.000$ and $p = 0.003$). It is also shown that LEV ratings increase while the reverberation grows, and

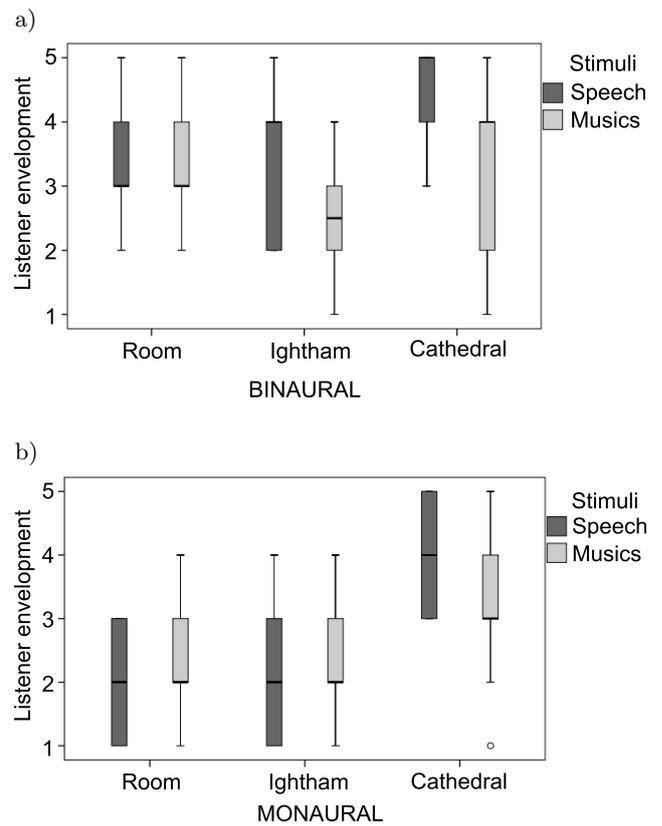


Fig. 8. Responses of the subjects for LEV: box-whisker plots show the first and third quartiles, the median of the data, the minimum and maximum values, and the outliers: a) binaural, b) monaural.

hence variances emerged when the Cathedral responses were considered, although they are on a smaller scale if the signal is convolved with music signal.

On the other hand, differences of means varied from 1.5 in the case of speech, to 0.5 in the case of music. In other words, a music signal has a slight tendency to mask the perception in the surround sense.

Multiple comparisons revealed that, except for music in the Ightham Mote, differences were noticeable both for the meeting room and Ightham Mote, but not for the Cathedral, where the binaural listening condition had not an influence in the perception. This is not the case when analyzing the stimuli factor, where speech and music have differences between their mean values, since music usually evokes less envelopment. However, a general minor dispersion of the data is observed in the case of music signals, and there is therefore an increase in the difference between mean values of both binaural and monaural signals.

Interestingly, Figs. 8a and 8b also show that there is a relevant increase in the LEV ratings between binaural and monaural signal for room and Cathedral. The results indicate that for more reverberation, substantial gains in LEV can be achieved when a sound signal is reproduced. This was also verified by the post-hoc multiple comparisons since no statistically significant

differences existed between the different groups of subjects (experts and non-experts). This fact confirms the former LEV hypothesis, which determined that subjects would perceive signals that were auralized with voluminous spaces as being more enveloping.

4.5. Realism

The evaluation of this question has been converted on a scale from 1 to 7 (see Fig. 9). The smaller the value is on the scale, the more preference is given to a Monaural signal, that is: 1. The effect of the Stimuli factor was relevant for this subjective quality, which evokes presence and realism. There were significant differences between ratings of auralizations convolved with various anechoic signals (speech or music) ($F(1, 14) = 3.04, p = 0.103$). In terms of speech, the overall mean values of room, Ightham Mote and Cathedral were 3.87, 5.40 and 5.17, respectively, while in case of music, these were 4, 4.47 and 4.20, respectively. It seems that the listener is not clearly decided by listening to a binaural signal, and therefore downplays the role of the listening condition. This statement correlates with ratings of other perceptual attributes (ASW and LEV) when the Cathedral space is assessed, that is to say, spatial information inside a large reverberant space is no longer one of the most important qualities when listening to a sound.

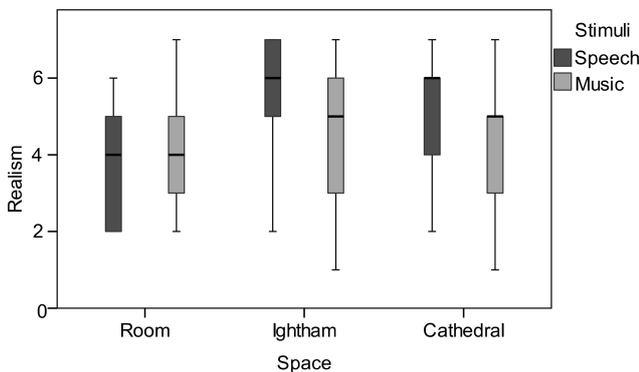


Fig. 9. Responses of the subjects for the sense of realism. The evaluation corresponds to a scale from 1 to 7: the smaller the value, the more preference is given to a monaural signal. Box-whisker plots show the first and third quartiles, the median of the data, the minimum and maximum values, and the outliers.

4.6. Relations between perceptual variables

The results were analysed using a two-way multivariate analysis of the variance model with the same three factors (space-stimuli-listening condition). In general, it could be stated that spatial impression changes noticeably both in terms of ASW and LEV, and obtains a greater difference for the former between room and Cathedral. Participants were initially

instructed to rate each attribute independently, however, the Pearson correlation coefficient between the two qualities related to spatial impression was $r = 0.745$ (ASW-LEV) are showed a statistical significance at the $p = 0.03$ level. Although these two parameters shared a relatively high correlation, differences in mean values of binaural-speech signals differed between the two attributes (ASW and LEV), thereby confirming a stronger connection between T30 and the sense of envelopment. Besides, ASW ratings between levels of stimuli (speech and music) were comparable, while in terms of envelopment, the values revealed little significant variation. In the case of the Cathedral, monaural signals showed significant differences with those of the other two spaces, a greater extent regarding LEV, whose ratings lost more surrounding perception than did those of ASW.

In the case of the relationship between localization and sense of realism, Fig. 10 shows the ability of virtual-source localization. The figure caption indicates the preference of the audio signal that inspires greater sense of realism: “BIN” means that > 3 answers were selected for binaural signal; “MONO” indicates the same but with the monaural signal; and “EQ” means an equal preference of both types thus. It could be stated that a suitable determination of the virtual-source position is found regardless of the sense of realism (Fig. 10).

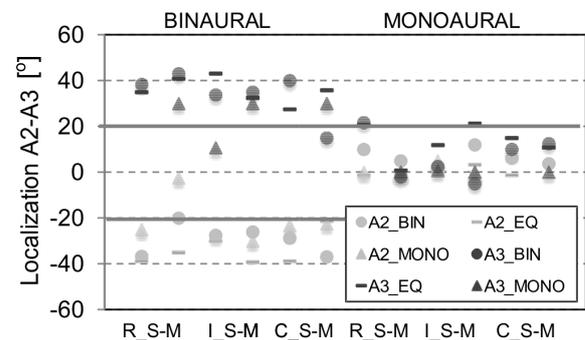


Fig. 10. Relationship between localization and sense of realism. A2 and A3 denote the acoustic sources: A2 is located at 20° and A3 is located at -20° .

5. Conclusions

Listening tests based on paired comparisons were carried out to evaluate the significance of providing binaural information when an auralization of an enclosure is intended. The analysis was conducted by means of statistical techniques using a structural model of analysis of variance (ANOVA). Four parameter associated to spatial impression were selected for the test, namely, localization, apparent source width, level of envelopment, and realism.

The results indicate that the acoustics of the enclosure plays a relevant role in the perception of spatial

information. In highly reverberant spaces, differences between binaural and monaural signals were not statistically significant in terms of ASW and LEV. This fact can be explained because the direct sound is masked in huge reverberant cathedrals and the background reflected energy is spatially diffuse. In addition, despite the fact that ASW and LEV are objectively related to different parts of the energy sound level, a moderate correlation was found when an auralization of the most highly reverberant space was considered. Regarding the stimuli factor (music or speech), the outcomes show that, in general, music signals broaden the sound source and mask the perception of spaciousness.

The analysis of the relationship between subjective qualities and physical objective parameters shows that lateral fraction is not related to ASW in the most highly reverberant space. In this case, the acoustics of the enclosure exert a significant degrading effect on the perceptual attribute.

Differences between the means of binaural signals confirmed a marked relationship between T30 and surround sensation (LEV). However, due to the small size of the sample, relations between subjective qualities and physical objective parameters cannot be considered as conclusive until further investigations have been carried out. Finally, in terms of localization, the total absence of perception of the position of the virtual source can be confirmed when anechoic monaural signals are rated. It is interesting to note that in the case of binaural signals, despite presenting significant improvements in the results with respect to monaural signals, it is still sometimes difficult for the listener to accurately determine the origin of the source, especially in reverberant spaces.

Acknowledgments

This work has been financially supported by the Spanish Government, with reference BIA2014-56755-P and under Grant FPU12/04949.

References

- ALONSO A., SUÁREZ R., SENDRA J.J. (2017), *Virtual reconstruction of indoor acoustics in cathedrals: the case of the Cathedral of Granada*, Building Simulation, **10**, 4, 431–446.
- BARRON M. (2000), *Measured early lateral energy fractions in concert halls and opera houses*, Journal of Sound and Vibrations, **232**, 1, 79–100.
- BARRON M., MARSHALL A. (1981), *Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure*, Journal of Sound and Vibrations, **77**, 2, 211–232.
- BECKER J., SAPP M. (2001), *Synthetic soundfields for the rating of spatial perceptions*, Applied Acoustics, **62**, 2, 217–228.
- BLAUERT J., LINDEMANN W. (1986), *Auditory spaciousness: some further psychoacoustic analyses*, The Journal of the Acoustical Society of America, **80**, 2, 533–542.
- BORK I. (2005), *Report on the 3rd round robin on room acoustical computer simulation. Part II: Calculations*, Acta Acustica united with Acustica, **91**, 4, 753–763.
- BRADLEY J., SOULODRE G. (1995a), *Objective measures of listener envelopment*, The Journal of the Acoustical Society of America, **98**, 5, 2590–2607.
- BRADLEY J., SOULODRE G. (1995b), *The influence of late arriving energy on spatial impression*, The Journal of the Acoustical Society of America, **97**, 4, 2262–2271.
- CERDÁ S. *et al.* (2013), *On the relations between audio features and room acoustic parameters of auralizations*, Journal of Vibration and Acoustics, **135**, 6, e-064501, 1–6, doi: 10.1115/1.4023835.
- COX T., D'ANTONIO P. (2004), *Acoustic absorbers and diffusers*, Spon Press, UK.
- DALENBÄCK B.I. (2011), *CATT-Acoustic v9 powered by TUCT, User's manual*, Gothenburg, Sweden.
- FISCHETTI A., JOUHANEAU J., HEMIM Y. (1992), *Relations between subjective spatialisation, geometrical parameters and acoustical criteria in concert halls*, Applied Acoustics, **37**, 3, 233–247.
- FOTEINOU A., MURPHY D., MASINTON A. (2010), *Investigation of factors influencing acoustic characteristics in geometric acoustics based auralization*, [in:] 13th Int. Conference on Digital Audio Effects, Graz, September 1–4, 2010.
- GRIESINGER D. (1997), *The psychoacoustics of ASW, spaciousness and LEV*, Acta Acustica united with Acustica, **83**, 4, 721–731.
- HACIHABIBOGLU H., MURTAGH F. (2008), *Perceptual simplification for model-based binaural room auralisation*, Applied Acoustics, **69**, 715–727.
- International Organization for Standardization (2010), *ISO 3382 Part 1. Acoustics – measurement of room acoustic parameters, performance spaces*.
- KIM S. (2013), *Influence of height-channel contents on enhancing spatial impressions through virtually elevated loudspeaker*, Proceedings of Meeting on Acoustics ICA2013, 19, 1, 055053, doi: 10.1121/1.4800821.
- KUTTRUFF H. (2000), *Room acoustics*, 4th ed., Spon Press, London.
- LEHNERT H., BLAUERT J. (1992), *Principles of binaural room simulation*, Applied Acoustics, 36, 3–4, 259–291.
- MARTELLOTTA F. (2008), *Subjective study of preferred listening conditions in Italian Catholic churches*, Journal of Sound and Vibrations, **317**, 378–399.
- MARTELLOTTA F. (2010), *The just noticeable difference of center time and clarity index in large reverberant spaces*, The Journal of the Acoustical Society of America, **128**, 2, 654–663.

22. MASIERO B., FELS J. (2011), *Perceptually robust headphone equalization for binaural reproduction*, Proceedings of 13th Convention Audio Engineering Society, May 13–16, London, UK.
23. MOLLER H. (1992), *Fundamentals of binaural technology*, Applied Acoustics, **36**, 171–218.
24. MOLLER H., SORENSEN M., JENSEN F., HAMMERSHOI D. (1996), *Binaural technique: do we need individual recordings?*, Journal of Audio Engineering Society, **44**, 6, 451–469.
25. MURILLO D. (2016), *Interactive auralization based on hybrid simulation methods and plane wave expansion*, Ph.D. Dissertation, Southampton University.
26. MURILLO D., COOPER C., FAZI F. (2014), *Acoustic survey of a late medieval building based on geometrical acoustics methods*, [in:] Forum Acusticum 2014, Kraków, pp. 43–55.
27. NOWAK J. (2010), *On the influence of sampling error on the perception of spatial sound fields using spherical microphone arrays for auralization*, The Journal of the Acoustical Society of America, **131**, 4, 3208–3218.
28. OKANO T., BERANEK L., HIDAKA T. (1998), *Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls*, The Journal of the Acoustical Society of America, **104**, 1, 255–265.
29. *Physikalisch-Technische Bundesanstalt (PTB) database* (n.d.), Braunschweig, Germany.
30. RINDEL M., CHRISTENSEN C. (2004), *Predicting the acoustics of ancient open-air theatres: the importance of calculation methods and geometrical details*, [in:] Baltic-Nordic Acoustics Meeting, Mariehamn, June 1–8, 2004.
31. ROWAN D., PAPADOPOULOS T., EDWARDS D., ALLEN R. (2015), *Use of binaural and monaural cues to identify the lateral position of a virtual object using echoes*, Hearing Research, **323**, 32–39.
32. Rutherford A. (2001), *Introducing ANOVA and ANCOVA: A GLM Approach*, SAGE Publications: London.
33. SILTANEN S., LOKKI T., SAVIOJA L., CHRISTENSEN C. (2008), *Geometry reduction in room acoustics modeling*, Acta Acustica united with Acustica, **94**, 3, 410–418.
34. SOULODRE G., LAVOIE M., NORCROSS S. (2003), *Temporal aspects of listener envelopment in multichannel surround systems*, Proceedings of 114th AES Convention, paper 580, March 22–25, The Netherlands, Amsterdam.
35. VAN DORP J., DE VRIES D., LINDAU A. (2013), *Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model*, The Journal of the Acoustical Society of America, **133**, 3, 1572–1585.
36. VIGEANT M., WANG L., RINDEL J. (2011), *Objective and subjective evaluations of the multi-channel auralization technique as applied to solo instruments*, Applied Acoustics, **72**, 311–323.
37. VORLÄNDER M. (2008), *Auralization*, 1st ed. Springer, Berlin.
38. WESTPHAL S., SCHOEFFLER M., HERRE J. (2015), *A Framework for Reporting Spatial Attributes of Sound Sources*, ICMI Workshop.