MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Perceptual Validation of Virtual Acoustic Models

Pedrero, Antonio[1]; De la Prida, Daniel; Marandet, Léo; Sánchez, José Luis; Navacerrada, María Ángeles; Díaz, César
Grupo de Investigación en Acústica Arquitectónica, E.T.S de Arquitectura, Universidad Politécnica de Madrid.
Avda. Juan de Herrera 4. 28040 Madrid (España)

ABSTRACT

The objective of a virtual acoustic model is to faithfully reproduce the sound of the acoustic environment that is represented in the model. Therefore, the model is often calibrated by in situ acoustic measurements. The usual procedure to calibrate the model is to compare the values of the room acoustical parameters obtained by measurements with those calculated in the acoustic model.
In this work a simplified perceptual validation procedure of virtual acoustic models is presented, in which, by means of a listening test, the listeners compare an anechoic signal emitted and recorded in the real environment with the same signal convolved with the impulsive response obtained from the virtual acoustic model.
The results of the application of this method to some case studies are shown and compared with those of the traditional calibration methods of virtual acoustic models.

Keywords: Validation, Acoustic Models, Listening Test
I-INCE Classification of Subject Number: 76

1. INTRODUCTION

In recent years, virtual acoustics is being used for numerous applications. Thanks to the auralization, it is now possible, by modelling, to listen to the sound that would be produced in a real built environment and even in environments that do not exist. Therefore, virtual acoustics can be used to evaluate the acoustics that a room will have that has not yet been built, or to know how a historic building that no longer exists would have sound in the past and which acoustic features have changed over time.

It is always recommendable to check to what extent the virtual acoustic model produces results which are comparable to those produced by the environment which is represented by the model. This can be done in two different ways that are not mutually exclusive: calibration and validation.

In metrology, calibration can be defined as the operation that establishes the relation between the indication of a measuring system and the quantity values provided by measurement standards. In the case of modelling, calibration involves estimating the values of various constants and parameters in the model and comparing them with those obtained experimentally in the real environment. Often model calibration is associated with a process of adjustment for maximizing the correlation between the estimates of the model and the experimental data.

---

[1] antonio.pedrero@upm.es

In contrast, validation is the confirmation that a given item fulfils the requirements for a stated intended use. The purpose of the validation is to ensure that the model adequately performs the functions for which it has been intended.

The purpose of model calibration is the authenticity, defined as the property of two entities to be indistinguishable to a human observer. It is a physics-based approach whose ideal goal would be to get a virtual environment that was an exact copy of the real environment.

For the calibration of the acoustic models, the room acoustic parameters defined by the ISO 3382-1[1] standard are usually used. These parameters are obtained both for the real environment by in situ measurements and for the model by calculation. As a quality criterion, it is generally considered that the model is correctly calibrated if the difference between the measured values and the calculated is less than the Just Noticeable Difference (JND) value established for each of these parameters described by ISO 3382-1.

The validation evaluates a very different concept, the plausibility. That is, instead of pursuing an exact copy of the real environment, it assesses to what extent the model reproduces all required quality features for a given application[2]. Therefore, it can be considered as a perception-based approach whose objective would be to obtain a model suitable enough for a given task.

When the purpose of the acoustic model is auralization, validation should be the method used to assess its quality. Calibration is not suitable for this purpose because there is no certainty that the usual room acoustical parameters are adequate predictors of room acoustical impression[3]

The perceptual validation of acoustic models is not a new issue. Several publications on this subject can be found in the literature [4-7]. However, these studies have used different methodologies and the results have not always been in accordance with expectations.

Generally speaking, the concept behind the perceptual validation of acoustic models is simple: it is to compare, by listening, a particular sound program produced in the real environment with the same sound program obtained from a virtual acoustic model by auralization. However, the variety of methodological approaches available for each of the phases of the process makes it necessary to have some guidelines to avoid mistakes that incur in a bad experimental design and, therefore, drawing wrong conclusions.

Mainly, the process consists therefore of two stages: in the first one, the pairs of sound stimuli to be compared are obtained: that is, the stimulus corresponding to the real environment and its counterpart in the virtual environment. The second stage consists in performing a listening test for comparing these stimuli and draw conclusions about the plausibility of model.

This paper describes the general aspects that must be taken into account in the perceptual validation of acoustic models. The methods and results of a pilot study conducted on room models with different acoustic characteristics are also presented.

2. OBTAINING THE SOUND STIMULUS

The aim of this task is to obtain the pairs of stimuli corresponding to the real and the virtual environments to be presented to the participants for comparison during listening test. Obtaining and preparing the stimuli requires some processes that can affect the audio signals used. For this comparison to be effective, the effects produced by these processes must be minimized or, at least, these effects must be the same in the two stimuli that are to be compared.

In order to obtain the stimulus corresponding to the real environment, an in situ recording of the sound produced by the sound emitter must be performed. The recording method must be consistent with the reproduction method that will be used in the listening test. In order to obtain the stimulus corresponding to the virtual environment, it is necessary to perform an anechoic recording of the sound emitter, and to convolve it with the impulsive response resulting from the modelling of the emitter and the sound propagation. The features of the impulsive response must also be adapted to the chosen reproduction method.

The first decision to be taken is about the sound program that will be used in the comparison. Since the interest is to evaluate the suitability of the model for a given task, the sound program should be of the same type that would be expected to be used in the usual application of the room or environment that has been modelled. The type of sound program is associated with the sound emitter, speakers if it is a vocal program or musicians if it is a musical program.

One of the issues of working with real sound emitters is that it is necessary the in situ and the anechoic chamber performances to be identical, which is not an easy task. In addition, the natural movements of sound emitters (speakers or musicians) during execution might generate instantaneous variations in directivity that are difficult to control. One possible solution is to use, as a sound source, a loudspeaker through which an anechoic signal is reproduced. This ensures that the base signal for both stimuli is identical.

2.1 Obtaining the real stimulus

As stated above, the real stimulus is obtained by an in situ recording. The most commonly used recording systems are:

- Monaural: The recording is made by an omnidirectional microphone. The main drawback is that the spatial attributes of the sound field are lost. To reproduce the sound in the listening test, a monophonic system must be used.
- Binaural: A pair of microphones located at the entrance of the auditory channel of the listener are used for the recording. Alternatively, the recording can be carried out with a dummy head, although this could alter some aspects of the spatial perception of the participants in the listening test. In both cases, sound reproduction must be done through headphones.
- Ambisonics: In its simplest form, the recording is made using a soundfield microphone that generates a B-Format signal. Sound reproduction must be done using a multichannel speaker system in a controlled acoustic environment.

In all cases, corrections must be applied to the resulting audio signal to compensate for the lack of linearity of the transducers used in the recording and reproduction processes.

2.2 Obtaining the virtual stimulus

The virtual stimulus is obtained by the convolution of the signal produced by the acoustic emitter, recorded in anechoic conditions, and the impulse response of the environment obtained from the virtual acoustic model. This model must include all the elements that affect the sound transmission from the emitter to the receiver.

For the modelling of the sound source it is necessary to determine its emission features, that is, its directivity and frequency response. When working with natural emitters, there might be a risk of adding the overall frequency response twice to auralization output; once from the directivity pattern and once more from the auralization signal which inherently includes the same source spectrum. To avoid this problem in auralization, the directivity pattern should be equalized with the inverse spectrum of that recorded at the source axis of the natural source signal.

In the modelling of the sound propagation, all the factors affecting the transfer of acoustic energy from the position of the emitter to the position of the receiver must be taken into account. In large rooms, the absorption of air can play an important role, especially at high frequencies. Since this factor depends to a great extent on the meteorological conditions, it is mandatory to replicate its effect in the same meteorological conditions as those that occurred when the recordings of the real stimuli were made. It is convenient to calibrate the propagation model by acoustic measurements in situ.

Finally, the calculated impulse response must be adapted to the method of sound reproduction that will be used in the listening test. If listening is done through headphones it is necessary to include the HRTF, which must be the same as those applied in the recording process of the real stimuli.

In addition, in order to obtain virtual stimuli as close as possible to the real stimuli recorded in situ, it is necessary to add a background noise, which is similar to that existing during the recording, and then normalize the levels of both signals so that there are no differences in loudness between the real and virtual stimuli.

3. PERFORMANCE OF THE LISTENING TEST

Many sensory discrimination techniques can be used for the assessment of sensory differences or sensory discrimination. Among these methods, the triangular test, the pairwise comparison (2-AFC), the m-AFC, the duo-trio test, the A-Not (A) and the Same-Different are the most used[8]. Each of these methods has its pros and cons. Therefore, the selection of the sensory discrimination test method and the design of the experiment should be very important steps of the preliminary stage. Although some of these methods have the same statistical power, since the same number of stimuli are used, they may not have the same sensitivity for the detection of differences between stimuli.

While some of them are easy to apply for non-expert participants and do not require knowledge of any particular attribute, others require some prior knowledge or learning by the participants. Some of them are more prone to certain types of bias, such as those related to sequencing (order in which the stimuli are presented), learning (making different judgments as the samples are better known) or memory (time between stimuli). On the other hand, the different methods and the questions asked to the participants can lead them to apply different cognitive decision strategies, some of them being more effective than others[9].

It is also important that the vocabulary used in the test properly describe the auditory qualities that are being investigated. In this regard, a vocabulary has been developed that contains all the perceptual attributes for the evaluation of all spatial audio technologies, and which should be used as a reference for the design of the test[10].

4. PILOT STUDY

A pilot study has been performed, in order to assess the applicability of a simple method for the validation of acoustic models. In this study, a perceptual assessment of the differences between real recordings and auralizations has been addressed.

For this purpose, two different acoustic environments have been used: a very reverberant room and a venue with a low reverberation time. The acoustic models of both venues have been generated through the program ODEON version 14. Both models have been calibrated with the data obtained from the acoustic measurements carried out in situ. The values of the main room acoustics parameters obtained, both for the measurements and for the models, are shown in Tables 1 and 2.

Table 1. Room 1: Values of the main room acoustics parameters measured and simulated, and differences in JND's.

Frequency (Hz)		63	125	250	500	1000	2000	4000	8000	Average JND's
EDT (s)	Simulated	5,80	5,99	6,14	6,32	5,88	4,71	2,98	1,39	1,4
	Measured	5,15	5,93	5,99	6,25	5,74	4,58	2,74	1,49	
	JND's	3,4	2,3	0,8	0,6	0,6	0,6	1,9	1,5	
T30 (s)	Simulated	5,79	5,97	6,13	6,29	5,87	4,70	2,99	1,41	0,9
	Measured	6,10	6,12	6,10	6,30	5,88	4,74	3,10	1,73	
	JND's	1,2	0,6	0,3	0,1	0,2	0,3	0,8	3,7	
Ts (ms)	Simulated	425	438	448	458	427	339	212	105	2,0
	Measured	351	427	427	441	413	317	189	107	
	JND's	4,6	2,2	1,2	1,0	1,0	1,6	2,7	1,5	
D50	Simulated	0,11	0,11	0,11	0,11	0,11	0,15	0,23	0,38	0,9
	Measured	0,18	0,11	0,10	0,13	0,14	0,18	0,27	0,37	
	JND's	1,6	1,1	0,4	0,7	0,5	0,8	1,1	0,8	
C80 (dB)	Simulated	-7,8	-8,0	-7,9	-7,9	-7,6	-6,2	-3,6	0,4	1,5
	Measured	-5,1	-7,3	-7,7	-6,8	-6,5	-5,1	-2,5	0,1	
	JND's	3,2	1,7	0,8	1,5	1,3	1,3	1,3	0,8	

Table 2. Room 2: Values of the main room acoustics parameters measured and simulated, and differences in JND's.

Frequency (Hz)		63	125	250	500	1000	2000	4000	8000	Average JND's
EDT (s)	Simulated	0,41	0,33	0,19	0,11	0,10	0,10	0,10	0,07	5,5
	Measured	0,52	0,34	0,24	0,11	0,10	0,13	0,11	0,04	
	JND's	2,9	1,8	3,9	4,1	1,7	5,1	3,4	21,4	
T30 (s)	Simulated	0,48	0,40	0,26	0,19	0,15	0,20	0,22	0,18	2,0
	Measured	0,40	0,35	0,22	0,17	0,14	0,19	0,22	0,19	
	JND's	4,1	3,3	3,8	1,6	1,7	1,7	0,3	1,3	
Ts (ms)	Simulated	25	19	11	7	5	6	6	4	6,1
	Measured	32	21	14	8	6	7	6	2	
	JND's	4,0	4,9	4,0	4,6	2,0	4,0	3,0	22,4	
D50	Simulated	0,84	0,90	0,97	0,99	1,00	0,99	0,99	1,00	0,3
	Measured	0,87	0,91	0,95	0,99	1,00	0,99	0,99	1,00	
	JND's	0,9	0,9	0,3	0,1	0,1	0,0	0,1	0,1	
C80 (dB)	Simulated	11,7	14,4	21,5	29,8	36,6	29,0	27,2	34,5	1,9
	Measured	17,3	17,8	22,5	30,9	35,3	29,6	27,7	35,6	
	JND's	5,6	3,4	1,1	1,6	1,3	0,6	0,5	1,2	

Instead of a natural sound source, a Brüel & Kjaer Echo Speech Source 4720 speaker was used as a sound emitter. By means of this sound source, two different anechoic signals have been played: a speech signal and a musical signal.

An in situ recording of the sound produced by the sound emitter at various points of each venue has been made. In order to record the background noise of each environment, sound fragments with the sound source turned off have also been recorded. The recording system has consisted of an omnidirectional microphone DPA ST 4006, and an audio interface RME Fireface UFX. The recording was made using the REAPER software at 24-bit and 44.1 KHz sampling rate.

The emission of the sound source Echo Speech Source 4720, reproducing the two sound signals, has been recorded in anechoic chamber as if it were a real interpreter. This recording has been made with the same equipment that has been used for on-site recordings. In addition, the directivity and the frequency response of the sound source have been measured.

The auralizations have been produced by the ODEON program. Since the sound recording in the real environment has been made with an omnidirectional microphone, HRTF filters have not been applied. Auralized signals have been added to the background noise recorded in real environments, trying to make the Signal to Noise Ratio the same as in the on-site recordings.

The listening test was done via headphones, so although all the signals were monophonic, a HRTF filter has been applied to improve the feeling of realism. The HRTF filter used was the corresponding to KEMAR dummy-head[11], for an incidence angle of 0°.

In the pilot study described in this communication, the auditory discrimination between stimuli was carried out by means of a balanced Same-Different test (AB, BA, AA, BB) being A and B two stimuli that can be in-situ recordings or auralizations. A and B were presented as “Reference” and stimulus “S1” during the test. The Same-Different method has proven its high sensitivity in several studies. It is also one of the recommended methods for non-expert participants. The fact that the reference is presented in each comparison avoids certain types of bias, as has been shown for other tests such as A-NOT (A) R or 2-AFCR [12]. The authors may use, for further studies on this matter, an incomplete Same-Different configuration, since it has been proven efficient in cases where there are numerous comparisons and the bias due to fatigue and sensory adaptation could be high [13].

Additionally, if the participants could perceive a difference between the stimulus S1 and the reference, the degree of difference perceived between stimuli had to be pointed out in a 5-point scale. In this sense, the test method can be considered as a version of a Same-Different with Sureness Rating test.

Apart from the overall difference, the participants were presented with a series of questions regarding three attributes of the stimulus (clarity, naturalness and presence) as described in 5. To carry out this task, participants were asked to use a 5-points bipolar intensity scale of as suggested by Lindau et al.[10]

To finish, participants were asked to label the stimuli among six options, in order to assess whether the participants were able to identify the real measurements and the auralizations or not.

Regarding the experimental procedure, after receiving the participant, the process was deeply explained with an example. Then, the participant was asked to listen to samples of dummy Reference and S1 stimuli, in order to become familiar with the types of sounds to be heard during the test. Then, the participant was asked about any possible doubts. In case no doubts were presented, the test began. Figure 1 shows the user interface of the graphic visual interface (GUI) designed for the listening test.

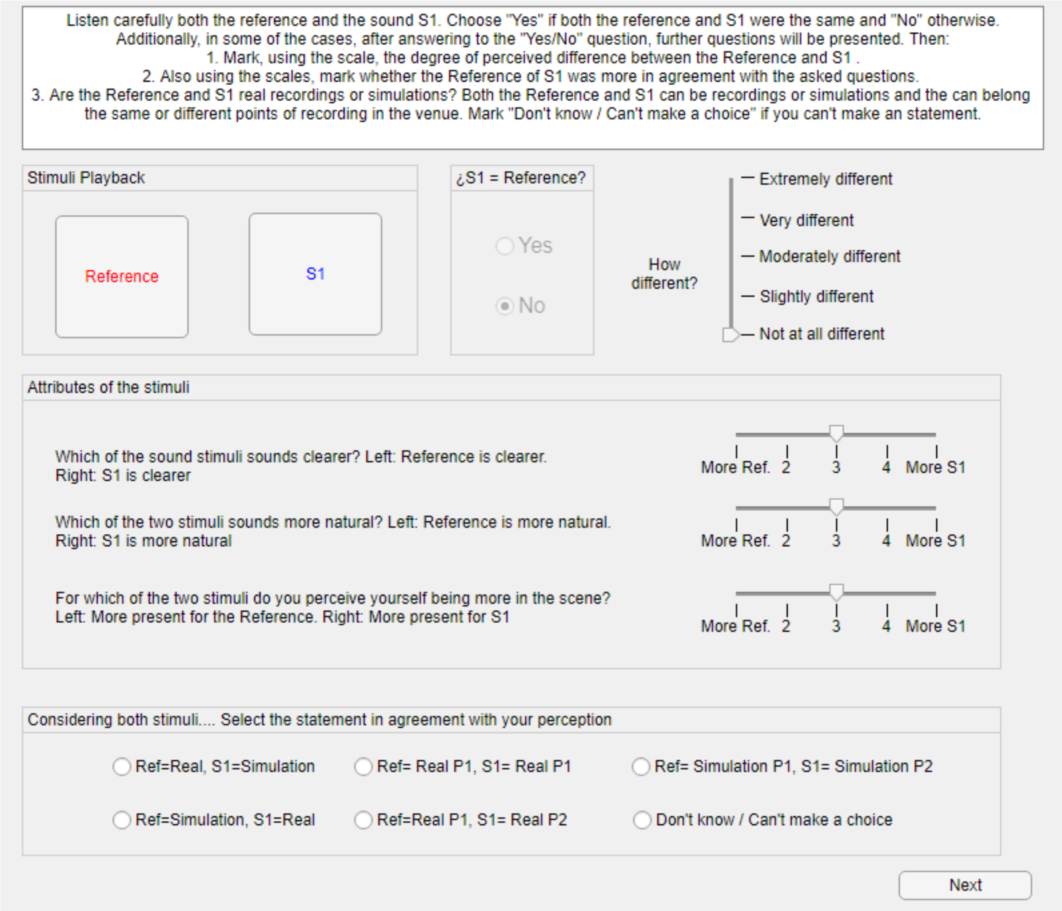

Figure 1. Graphic visual interface (GUI) designed for the listening test.

The test, in which the stimuli were presented through headphones, was performed in an insulated booth. The process was repeated four times, since two different stimuli (the speech and the music excerpt) and two different venues (the very dry room and the very reverberant room) were addressed. The duration of each of the four blocks was variable depending on the participant's skill, but it was between 6 and 9 minutes for all the participants. Therefore, the duration of the whole test ranged between 24 and 36 minutes.

4.1 Pilot study results

The results of the tests show that most of the participants find differences between the signals recorded in situ and the signals obtained by auralization. In the case of the speech signal, 100% of the participants were able to detect differences in the most reverberant room and 95% did so in the dry room. For the musical signal, the percentages were slightly lower, detecting 83% of differences for the room with high reverberation time and 75% for the room with low reverberation time.

Regarding the degree of similarity between these signals, most of the participants rated the difference between real and virtual signals as “Slightly different”. Only two out of the eight comparisons for speech signal in the most reverberant room were described as “Moderately different”.

Regarding the labelling of signals most of the participants were not able to clearly discern which signals were recorded and which were obtained by auralization.

Comparing, in qualitative terms, the results of the perceptual validation and those of the calibration by means of objective parameters, it can be seen that the perceptual validation does not show such a big difference in the quality of the models of the two rooms under test as could be deduced from the differences in JND's obtained in the calibration. It is interesting to note that the room with the highest reverberation, which showed, in terms of JNDs, values of differences significantly lower than the dry room, was the one where the participants perceived greater differences between the real signals and the signals obtained by auralization.

5. CONCLUSIONS

In this work, the most important aspects that must be taken into account for the perceptual validation of acoustic models have been discussed.

A simple methodology has been presented to carry out these validations. It uses a speaker as a emitter, which simplifies the characterization and modelling of the sound source. In addition, the same equipment is used for the recording of sound in the real environments and for the anechoic recording of the signals that will serve as the basis for the auralizations.

A listening test has been carried out which has shown that, for the two cases under study, the participants were able to detect differences between real signals and auralizations. However, the differences found were small and the participants were not able to label if the signals were recorded in a real environment or if they were obtained from a model by means of auralization.

In both cases, there was not a clear correspondence between the objective calibration and the perceptual validation when assessing the quality of the models. Therefore, it is suggested that further research on this matter is performed.

6. REFERENCES

1. ISO 3382-1, "*Acoustics. Measurement of room acoustic parameter - Part 1: Performance spaces*" International Organization for Standardization, Geneva, Switzerland (2009)

2 Renato S. Pellegrini, "*Quality assessment of auditory virtual environments*." Proceedings of the 2001 International Conference on Auditory Displays. (2001)

3. Stefan Weinzierl and Michael Vorländer, "*Room acoustical parameters as predictors of room acoustical impression: What do we know and what would we like to know*?" Acoustics Australia 43.1 (2015)

4. T. Lokki and H. Jarvelainen, "Subjective evaluation of auralization of physics-based room acoustics modeling," in International Conference on Auditory Display (2001)

5. Y.-J. Choi and F. Fricke, "*A comparison of subjective assessments of recorded music and computer simulated auralizations in two auditoria*" Acta Acust. united with Acust.92 (2006).

6. W. Yang and M. Hodgson, "*Validation of the auralization technique: Comparative speech-intelligibility tests in real and virtual classrooms*" Acta Acust. united with Acust.93 (2007)

7. Barteld N.J. Postma, and Brian F.G. Katz. "*Perceptive and objective evaluation of calibrated room acoustic simulation auralizations*" The Journal of the Acoustical Society of America140.6 (2016)

8. Prins, N. "*Psychophysics: a practical introduction*", Academic Press (2016)

9. Jeong, Y.-N. et al. "*Sensory discrimination by consumers of multiple stimuli from a reference: Stimulus configuration in A-Not AR and constant-ref. duo-trio superior to triangle and unspecified tetrad?*" Food Qual. Prefer. 47 (2016)

10. Lindau, A. et al. "*A spatial audio quality inventory (SAQI)*", Acta Acust. united with Acust. 100 (2014)

11. Bill Gardner and Keith Martin. "HRTF Measurements of a KEMAR Dummy-Head Microphone". In: MIT Media Lab Perceptual Computing - Technical Report #280 (1994). url: http://sound.media.mit.edu/resources/KEMAR.html.
12. Stocks, M. A., van Hout, D. & Hautus, M. J. "*Cognitive decision strategies adopted by trained judges in reminder difference tests when tasting yoghurt, mayonnaise, and ice tea*". Food Qual. Prefer. 34, (2014).
13. Choi, Y.-J. et al. "*Superior performance of constant-saltier-reference DTF and DTFM to same-different tests by consumers for discriminating products varying sodium contents*". Food Qual. Prefer. 37, (2014)