

Appendix A

Additional publications

A.1 Relevant publications

A.1.1 Paper E

Title: Does the method matter? A review of the main testing methods for the subjective evaluation of room acoustics through listening tests

Authors: D. de la Prida, A. Pedrero, L. A. Azpicueta-Ruiz, M. A. Navacerrada, C. Díaz

Conference: 23rd International Congress on Acoustics, ICA 2019

Conference location: Aachen, Germany

Conference date: 08-09-2019 to 13-09-2019

Proceedings: Proceedings of the 23rd International Congress on Acoustics, pp. 7887–7894

Publisher: Deutsche Gesellschaft für Akustik (Berlin, Germany)

Does the method matter? A review of the main testing methods for the subjective evaluation of room acoustics through listening tests

Daniel DE LA PRIDA¹; Antonio PEDRERO¹; Luis A. AZPICUETA-RUIZ²; María A. NAVACERRADA¹; César DÍAZ¹;

¹ Grupo de Investigación en Acústica Arquitectónica, Universidad Politécnica de Madrid, 28040 Madrid, Spain

² Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain

ABSTRACT

The listening tests are experimental procedures in which a group of participants are asked to express their opinion about certain questions regarding some stimuli that they have been urged to listen to. The listening tests have proved their effectiveness as methods for the subjective assessment of perception in different fields of acoustics, and particularly in room acoustics and sound insulation. The way in which the listening test is conducted, how the participant's response is expressed and how the stimuli are presented to the participants define what is known as the testing/query method. There is not yet a common methodological framework that defines which testing method is the best for each purpose and under what conditions the experiments are to be carried out. The purpose of this communication is to review the most commonly used testing methods in the subjective perception of sound and to present the advantages and disadvantages of each of them when they are used to evaluate the subjective perception of a sample of population.

Keywords: listening test, room acoustics

1. INTRODUCTION

Sensory evaluation tests allow to obtain the subjective perception of human beings regarding certain stimuli that are presented to them. This type of test has been found to be very useful in the commercial sector, when aiming to improve, consolidate or renew products, so that the population better perceives them. They are also used to assess how reducing costs on certain materials or processes affects the perceived quality of the final product. In this sense, these types of tests are used regularly in the fields of food (1), fragrances (2) and design (3), among others.

Although the use of subjective tests is not new in the field of acoustics, this kind of testing has gained great importance in recent decades. A paradigm shift in acoustics is taking place, centred on the fact that the objective descriptors used to assess acoustic quality should be more representative of the subjective perception of human beings.

In particular, in the field of environmental acoustics, concepts such as "soundscape" (4) and psychoacoustic parameters have taken on an important relevance in recent years. Listening tests are being used to evaluate how human beings perceive soundscapes and how this perception can affect their actions (5,6).

In the field of building acoustics, the most recent research related to the assessment of sound insulation is also based on listening tests. These tests are mainly focused on evaluating the degree of agreement between the SNQs (Single Number Quantities), commonly used to assess the protection

¹ arquilav.arquitectura@upm.es

² lazpicueta@tsc.uc3m.es

contributed by a sound insulation element, and the subjective perception of loudness and annoyance (7,8).

In the field of room acoustics, in recent years, listening tests have also gained great interest. They are mainly aimed to determine the just noticeable differences (JND) of the room acoustic descriptors that are frequently used to evaluate the acoustic quality of venues. Among these descriptors reverberation time, clarity or definition, as well as spatial parameters such as apparent source width (ASW) or listener envelopment (LEV) are of great interest (9,10).

Finally, since the term auralization was coined and, in particular, since this term is supported by the rise of new virtual reality applications, listening tests are of great importance in this field. Main research is related with the model calibration requirements as well as the precision of the simulation software, models and HRTFs (Head-Related Transfer Functions). The common aim is that differences between real situations and simulations are as small as possible and that auralizations are always perceived as plausible (11–13).

In other completely different fields, such as food research, great efforts are being made to develop subjective tests that are as reproducible as possible (14,15). Therefore, numerous studies have been conducted in the last years in order to determine the best testing method and design of the experiment, so that the tests are comparable and accurate. In the field of acoustics, although numerous listening tests are being performed, there are not yet clear methodologies for their design and performance. This fact can make the comparison between the results of different investigations difficult and tedious.

The purpose of this communication is to review the listening tests that have been recently undertaken in the field of room acoustics, in order to assess to what extent, the different works are following similar or different methodologies in each of the stages of the process. The main differences between researches are discussed and certain conclusions are drawn, which may be useful for improving the subjective evaluation in the field of room acoustics.

2. METHODOLOGY

The methodology followed in this review was divided into several steps: First, a detailed search for references in the fields of study was performed. Then, the references found were filtered in order to select only those that were relevant. Finally, a series of common concepts were defined and collected from each reference. Each of these steps is described individually below:

2.1 Literature search

In order to carry out this review, an extensive search has been carried out for the most recent bibliographical references (mostly between 2009 and 2017) in the field of subjective assessment of room acoustics descriptors, although some older references (between 1993 and 2005) have also been considered due to their relevance in the field under study. This search for the state of the art was done during the months of December 2018 and February 2019, although some of the references were originally collected in August 2017.

In order to perform this search, scientific databases have been used, as well as scientific social networks and specific search engines. In particular, most of the references have been found through Google Scholar, ScienceDirect and ResearchGate. Different procedures were followed for the search:

1. **Search by keywords:** This was the main search method. Scientific search engines and databases were used for keywords such as "perception", "perceptual", "subjective", "room acoustics", "reverberation time", "clarity", "JND", "just-noticeable-difference" or "virtual".
2. **Search by affinity:** Databases such as ScienceDirect suggest similar references to the last consulted by affinity between concepts.
3. **Search by references:** Consultation of the bibliography of the most recent articles allows, on some occasions, to obtain references to previous related articles.
4. **Search by author:** Some authors focus all their research activity on the same field, so that on some occasions several articles by the same researcher were found in the field of study.

2.2 Filtering of the literature

From the conducted search, 15 references were obtained (9,10,16–28), which were selected, in a preliminary way, only by reviewing the title or the abstract. After completing the search tasks, a deeper filtering of the articles was done in order to use only those that were directly related to the purpose of the review. Therefore, only journal and conference papers, in which a subjective study was performed by means of listening tests and in which the methodological design and the analysis of the listening test results had an important contribution were used. This task reduced the number of papers to 8 (9,10,16,20,22,23,25,28).

2.3 Definition of common concepts

As already mentioned in the introduction, listening tests are tools for the subjective characterization of sound in which a series of participants are presented with various stimuli and their judgement about these stimuli is collected through some system. Since a listening test is a tool that interacts with the human being, there are several sources of bias that can reduce the relevance of the experiment or even lead to erroneous conclusions. Therefore, when reviewing listening test studies, it is important to evaluate concepts such as the following:

1. **Purpose of the experiment:** What is the experiment intended to study?
2. **Testing method:** How is the consultation and collection of results addressed?
3. **Selection and preparation of stimuli:** What sound samples were used? How were they processed?
4. **Method of analysis:** Which statistical method was employed for the analysis of the results?
5. **Quality of the design:** Is the performed design robust? How many participants did the experiment have? Is it possible to identify design actions that were presumably aimed to reduce bias?

For the sake of simplicity, the particular results of each of these studies are not taken into account, as this is not within the scope of this review, which relates to the design of the listening test methods.

3. RESULTS

In this section, each of the eight articles presented in the previous section is described in detail, considering each of the five common concepts. A summary of this analysis is shown in Table 1.

In reference (22), the purpose of the research was to measure the smallest perceivable change in the objective room acoustic parameters associated with spatial impression and clarity. For that purpose, the difference limens for ELEF (early lateral energy fraction), IACC (inter-aural cross correlation), Ts (centre time) and C80(clarity) were addressed. In order to find the difference limens, the “Method of Minimal Changes” was used with 7 to 10 participants. The experiment was conducted in an anechoic environment where simulated variations of a concert hall were played through a loudspeaker system, comprising delays, effects and attenuators. Two sound samples were used for the listening test, a 8s. excerpt of “Handel’s Watermusic” and a 5s. excerpt of “Mendelsohn’s Symphony no.3 Opus 56 *The Scottish*”. The analysis of the results was mainly based on the calculation of the difference limens as the smallest perceivable changes that are perceived by at least 50% of the participants, which is the regular method of analysis for a Minimal Changes testing method. Confidence intervals at 95% were also given. Presumably, a couple of actions were aimed to the reduction of bias. First, only normal hearing participants could perform the test, even though their hearing was not tested. Fatigue effects were checked through F-test analysis.

The suitability of various objective measures as predictors of subjective ratings was examined in(23). In particular the relation between the EDT (early decay time), Ts, C80, G (loudness), early bass level and treble ratio, and the subjective perception was addressed. For that purpose, a sample of 10 subjects performed 8 different listening tests ranking several sound fields in terms of loudness, clarity, reverberance, envelopment, apparent source width and overall preference. The testing method was based on double blind paired comparisons (attribute-related 2-AFC). The sound fields were generated through near-field loudspeakers, equipped with mechanical barriers to avoid crosstalk. Just one sound sample was used during the listening test, a 15s. passage of “Mozart’s La Nozze di Figaro”. The analysis of the results was based on the regression analysis between the preference counts of the

subjective perception attributes and the room acoustic objective parameters described. Regarding bias control, it is said that being a difference test, any imperfection in the reproduction chain would be minimized. A description of the parameters to evaluate was given before the test and a brief training sequence performed.

The purpose of (9) was to find the just noticeable difference for C50. For that purpose simulated sound fields representing realistic situations were played through a 8-loudspeaker system, similar to that of (22), in an anechoic environment. 10 participants took part in a listening test, which consisted in deciding whether there was difference between pairs of sound fields, with step changes in clarity of 0.5, 1.0, 1.5, 2.0, 2.5 and 4 dB. This test was repeated for three different sound fields with different reverberation times. A sound sample was played through the loudspeakers to allow the participants to perform the test, an anechoic male modified rhyme test at 3.0 syllables per second. To determine the JND for C50, the percentage of participants that perceived a difference was accounted for each clarity step, similar to the procedure described in (22). Also, regression was calculated between the subjective results and the reverberation time, to evaluate the change in perceived clarity with reverberation time. Two actions can be found, which presumably are aimed to reduce bias. First, the pairs of sounds were randomized. Also, a training was given before the test, in which the participants tested the extreme cases.

In (25), an experiment was developed to examine the discrimination threshold for the direct-to-reverberant energy ratio (v). For that purpose, 6 participants took part in a test following an adaptive 3-down 1-up two-alternative forced choice procedure through headphones. The individualized BRIR (binaural room impulse response) was measured for 5 out of the 6 participants and the different stimuli were obtained by scaling the reverberant energy of the BRIR. Four sound samples were convolved with the BRIR: a short duration noise burst of 50 ms with 2.5 ms onset/offset times, a long duration burst of 300 ms with 150 ms onset/offset times, a male speech sample of syllable /da/ of 300 ms and an impulsive signal, which was the BRIR itself. Several trials were repeated before the procedure stopped. Runs terminated after 20 reversals and the thresholds were determined as the average of the last 5 reversals. Also, psychometric functions were fitted for each stimulus and evaluated by Chi-square test. The reduction of bias was presumably the aim of different actions, such as the performance of an audiometry for each participant, the individualized measurement of the BRIRs as well as the individualized headphone equalization (HpTF).

A research was performed in (10), with the aim of determining the difference limen for single number parameters relating the ASW and LEV. For that purpose, the BRIRs of different venues were measured with a dummy head in positions with different LF (lateral energy fraction) and LG (lateral sound level). These BRIRs were later convoluted with a sound sample and presented to the participants through a crosstalk cancellation loudspeaker system. The sound sample was a 14s. excerpt of “Haydn’s Trumpet Concerto played with an E flat Cornet”. Different participants samples were used in individual listening tests for LF changes (12 participants) and LG changes (7 participants). The testing method was similar to the others, where the participants listened to stimuli in pairs and had to make a judgment about the stimuli being the same or different. From the results of the listening test, the difference limens were calculated by means of the fitting of the psychometric function a Monte Carlo simulation for the goodness of fit. Several actions could be found, which presumably aimed to reduce bias. In particular, the participants might not repeat the hearing of the samples. That way, it was assured that identical audio presentation is given to each participant. Also, all participants received an audiometry, to check the hearing. A training and familiarization process was performed by the participants prior to the test, where extreme cases could be listened. Instructions were given by means of a written guide. The experimenter collected data about the difficulty perceived by the participants.

The approach of (16) was to investigate the possible dependence of the JND of the Ts and C80 as a function of the reverberation time in large reverberant enclosures. Impulse responses were measured in B-format and later varied in clarity parameters. These impulse responses were then convolved with two sound samples and played in a silent room through four loudspeakers. A 5s. excerpt of “Pange Lingua” and a 9s. excerpt from “Weber’s Theme for Cello” were used. The “Method of Minimal Changes” was then used, in which around 40 participants were presented with two stimuli and had to decide whether they were the same or different. If the participants identified the stimuli as different,

they had also to report which was the clearer. As the “Method of Minimal Changes” was employed, the JND was defined as the smallest perceivable change detected by 50% of the participants. Regarding bias, the participants were not test for hearing but asked about hearing issues. Sound samples were presented randomly. To avoid any learning or memory effect, listening test for each stimulus were carried in different days. Participants were asked about the design of the experiment and some reported difficulties regarding the concept of clarity, with graphical interface and with failing to find an effective listening method.

The determination of the necessary change of the primary cues, such as ITD (interaural time difference), ILD (interaural level difference) and IACC for subjects to detect them was investigated in (20) . For that purpose, the ITD and ILD of the early part of measured BRIRs, corresponding to three different environments, were manipulated. Eight participants conducted an adaptive three-interval 2-AFC test, following a 1-up 2-down staircase procedure. A 2s. excerpt of three different sound samples was randomly selected. Sound samples were recordings of a guitar, a violin and snare drum. The results were presented as mean values and standard errors of detection for the different rooms and instruments. Regarding bias control, each test is repeated four times. However, there is no mention in the article regarding training or familiarization processes, the environment, equipment or equalization of the playback chain.

Finally, in (28), the focus was on the evaluation of the JND of C80 and the assessment of whether a variation in the way the stimuli are presented to the participants gave the same or different results for the JND. For that purpose, several simulated impulses responses were used, and sound samples convoluted with these IR (impulse responses) used as stimuli. The stimuli were presented to the participants, in an anechoic environment, through a 8-channel loudspeaker system. 51 participants were presented with the first variant of the test, where the participants were presented with stimuli in pairs and had to listening to both stimuli, in order and in full, and judge whether the they were the same or not. They might listen to the stimuli as much as they wanted but they had to listen both stimuli, completely and following a fixed order, each time. 11 subjects were presented with the second variant of the test, where the participants might switch between stimuli during the reproduction of the sound sample as much as they wanted and didn't have to listen to the stimuli in full. Three different sound samples were used for the first listening test. A 10.9s. passage of the “Third movement of Bizet's L'Arlésienne Suite No. 2”, a 10.3s. excerpt of “Weber's Theme” and a 10.3s. motif of the beginning of “Handel's Water Music”. For the second test, the Bizet piece was used again. In addition, a sample which length was not reported was used of the “Debussy's Prelude to the Afternoon of the Faun”. The analysis of the results followed the same as for the method of Minimal Changes. Regarding the control of bias, the impulses responses and sound samples were randomized. Training was given prior to the listening test, with samples and impulses responses not used later during the test. Hearing level was measured for each participant.

4. DISCUSSION

Several research articles are presented, in which the main purpose is the study of the JND for several room acoustic parameters. Even though some trends can be found regarding the testing method used for the presentation of the stimuli and consultation of the participants, three main different methods are identified; the Method of Minimal Changes, the attribute-related 2-AFC and adaptive up-down approaches based on non-attributive 2-AFC procedures. In particular, the third is a more precise and efficient method of the first. Both the first and the third are based on a non-attributive 2-AFC also called same-different method. Being this method prone to tendency bias among others (29), it is considered that further research is necessary, to find newer testing methods that are more efficient and less prone to bias. Not only how the participants are consulted but also how the participants interact with the stimuli should be considered for deeper analysis. It can be found that some works (9,16,23) prefer to leave the participants freedom of choice, allowing them to control the reproduction, while others prefer to fix how the stimuli are presented (10,25). Specially interesting are the results of (28), where performing the listening test with different procedures regarding this matter reported different results.

Table 1 – Summary of common concepts for all the reviewed articles

Reference	(22)	(23)	(9)	(25)	(10)	(16)	(20)	(28)
Purpose of the study	ELEF, IACC, Ts, C80	EDT, Ts, C80, G, treble, bass	C50	ν	ASW, LEV	Ts, C80	ITD, ILD, IACC	C80
Participants	7 - 10	10	10	6	7 - 12	40	8	11 - 51
Testing method	Minimal Changes	Attribute 2-AFC	Same - different	Adaptive 2-AFC	Same - different	Minimal Changes	Adaptive 3I/2-AFC	Same-different
Stimuli and duration	S1. Water Music – 8s S2. The Scottish - 5s	S1. Nozze Figaro – 15s	S1. Male rhyme test - Loop	S1. Burst - 50 ms S2. Burst - 300 ms S3. /da/ - 300 ms S4. Impulse	S1. Trumpet concerto Cornet – 15s	S1. Pange Lingua – 5s S2. Theme for Cello – 9s	S1. Guitar – 2s S2. Violin - 2s S3. Snare drum – 2s	S1. L'Arlésienne Suite n.2 – 10.9s S2. Theme for Cello – 10.3s S3. Water Music – 10.3s S4. Prelude...
Analysis method	Diff. limen 50% F-test analysis	Regression	% difference and regression	Psychometric function. Chi-square fit test	Psychometric function. Monte Carlo fit	Diff. limen 50% Psychometric function.	Mean and standard errors of detection	Diff. limen 50% Regression
Playback system	8 Loudspeakers	2 Loudspeakers	8 Loudspeakers	Headphones	2 Loudspeakers	4 Loudspeakers	Not mentioned	8 Loudspeakers
Training	Not mentioned	Yes	Yes	Not mentioned	Yes	Yes	Not mentioned	Yes
Playback method	Not mentioned	Free switching	Free switching	Fixed	Fixed	Free switching	Not mentioned	Both (fixed/free)
Audiometry	No	No	No	Yes	Yes	No	Not mentioned	Yes
Additional bias control	-	Written instructions	Random order of stimuli	Individual BRIR and HpTF	Written instructions	Participant feedback	Random choice of sound sample	Random order of stimuli

Regarding the sound samples, each research uses different samples and different durations. Even though in some occasions the papers justify why certain sound samples are selected, no clear background is given sustaining the selection. It is not clear for the authors whether unified sound samples should be described for all listening test regarding room acoustics. However, it is considered necessary that further deeper and detailed research is performed to evaluate the importance of the sound samples and their duration in the subjective perception of room acoustics. It is also considered, that the excerpts used by published works should be available for other researchers, so that they can use them for reproducible research.

Regarding the analysis of the results, a base descriptor should be selected among existing methods of analysis or research should be done to describe a new descriptor that can be extracted, independently of the nature of the difference and the purpose of the experiment that can make the results of different researches comparable.

Bias is an important factor that must be kept under control when performing subjective tests. Some sources of bias are related with the testing method and might be reduced by using better testing methods. However, other sources of bias can be related to the design of the experiment, the interaction between the participants and the experimenter, the participants themselves and the environment in which the listening test is to be performed. In this review, several actions that are presumably aimed to control and reduce bias could be found, which was not applied by all the works. The last four rows of table 1 exemplify this fact easily. In those cases where a mention was made regarding the training and familiarization of the participants, the playback method, the hearing level audiometry and playback chain equalization, different procedures were followed. More explanations should be given in the papers about the background sustaining the taken actions and further analysis must be done in those cases where no relevant background can be given.

In summary, citing Weinzierl and Vorländer (30) “*In room acoustics, we look back on more than 50 years of research on developing psychological measuring instruments for the concept of ‘room acoustical impression’, and on more than 100 years of research on the development of physical measures which could serve as technical predictors for these perceptual qualities. On both sides, the state of the art is surprisingly unsatisfactory.*”. Therefore research in the field of subjective perception of room acoustics should not only focus on the application of listening tests to assess perceptible differences, but also on the evaluation, understanding and improvement of methods, so that clear methodologies can be described, which are followed by all researches in this field. In other words, it must be pursued that the study of the subjective perception of room acoustics is carried out through reproducible research.

REFERENCES

1. Rousseau B, Rogeaux M, O'Mahony M. Mustard discrimination by same–different and triangle tests: aspects of irritation, memory and τ criteria. *Food Qual Prefer.* 1999;10(3):173–84.
2. Ward P, Davies BJ, Kooijman D. Olfaction and the retail environment: examining the influence of ambient scent. *Serv Bus.* 2007;1(4):295–316.
3. Seckler M, Opwis K, Tuch AN. Linking objective design factors with subjective aesthetics: An experimental study on how structure and color of websites affect the facets of users' visual aesthetic perception. *Comput Human Behav.* 2015;49:375–89.
4. Kang J, Schulte-Fortkamp B. *Soundscape and the Built Environment.* CRC press; 2016.
5. de la Prida D, Pedrero A, Navacerrada MÁ, Díaz C. Relationship between the geometric profile of the city and the subjective perception of urban soundscapes. *Appl Acoust [Internet].* 2019 Jun;149:74–84. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0003682X18308168>
6. Guastavino C, Katz BFG, Polack J-D, Levitin DJ, Dubois D. Ecological validity of soundscape reproduction. *Acta Acust united with Acust.* 2005;91(2):333–41.
7. Hongisto V, Oliva D, Rekola L. Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. *J Acoust Soc Am.* 2018;144(2):1100–12.
8. Monteiro C, Machimbarrena M, de la Prida D, Rychtarikova M. Subjective and objective acoustic performance ranking of heavy and light weight walls. *Appl Acoust [Internet].* 2016 Sep [cited 2016 Apr 13];110:268–79. Available from: <http://www.sciencedirect.com/science/article/pii/S0003682X1630041X>

9. Bradley JS, Reich R, Norcross SG. A just noticeable difference in C50 for speech. *Appl Acoust.* 1999;58(2):99–108.
10. Witew IB, Behler GK, Vorländer M. About just noticeable differences for aspects of spatial impressions in concert halls. *Acoust Sci Technol.* 2005;26(2):185–92.
11. Lindau A, Weinzierl S. Assessing the plausibility of virtual acoustic environments. *Acta Acust united with Acust.* 2012;
12. Postma BNJ, Katz BFG. Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. *J Acoust Soc Am.* 2016;
13. Rychtáriková M, Chmelík V, Roozen NB, Glorieux C. Front-back localization in simulated rectangular rooms. *Appl Acoust.* 2015;
14. Stocks MA, van Hout D, Hautus MJ. Cognitive decision strategies adopted by trained judges in reminder difference tests when tasting yoghurt, mayonnaise, and ice tea. *Food Qual Prefer.* 2014;34:14–23.
15. O'Mahony M. Sensory difference and preference testing: the use of signal detection measures. In: *Applied Sensory Analy of Foods.* Routledge; 2017. p. 145–75.
16. Martellotta F. The just noticeable difference of center time and clarity index in large reverberant spaces. *J Acoust Soc Am.* 2010;128(2):654–63.
17. Bradley JS. Review of objective room acoustics measures and future needs. *Applied Acoustics.* 2011.
18. Lokki T, Pätynen J, Kuusinen A, Tervo S. Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *J Acoust Soc Am.* 2012;
19. Klockgether S, van Dorp Schuitman J, Van de Par S. Perceptual limits for detecting interaural-cue manipulations measured in reverberant settings. In: *Proceedings of Meetings on Acoustics ICA2013.* ASA; 2013. p. 15004.
20. Klockgether S, van de Par S. Just noticeable differences of spatial cues in echoic and anechoic acoustical environments. *J Acoust Soc Am.* 2016;140(4):EL352–7.
21. Klumpp RG, Eady HR. Some measurements of interaural time difference thresholds. *J Acoust Soc Am.* 1956;28(5):859–60.
22. Cox T, Davies W, Lam Y. The sensitivity of listeners to early sound field changes in auditoriums. *Acustica.* 1993;
23. Soulodre GA, Bradley JS. Subjective evaluation of new room acoustic measures. *J Acoust Soc Am.* 1995;
24. Bradley JS, Reich RD, Norcross SG. On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. *J Acoust Soc Am.* 1999;106(4):1820–8.
25. Zahorik P. Direct-to-reverberant energy ratio sensitivity. *J Acoust Soc Am.* 2002;
26. Okano T. Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections. *J Acoust Soc Am.* 2002;111(1):217–29.
27. Bradley JS. Using ISO 3382 measures, and their extensions, to evaluate acoustical conditions in concert halls. *Acoust Sci Technol.* 2005;
28. Vigeant MC, Celmer RD. Effect of Experimental Design on the Results of Clarity-Index Just-Noticeable-Difference Listening Tests. In: *Proceedings of 20th International Congress on Acoustics, ICA 2010, Paper.* 2010.
29. Prins N. *Psychophysics: a practical introduction.* Academic Press; 2016.
30. Weinzierl S, Vorländer M. Room acoustical parameters as predictors of room acoustical impression: What do we know and what would we like to know? *Acoust Aust.* 2015;43(1):41–8.

A.1.2 Paper F

Title: The effect of the rolling shutter in the perception of the sound insulation

Authors: D. de la Prida, A. Pedrero, M. A. Navacerrada, A. Díaz-Chyla

Conference: 48th International Congress and Exhibition on Noise Control Engineering, Internoise 2019

Conference location: Madrid, Spain

Conference date: 16-06-2019 to 19-06-2019

Proceedings: Proceedings of the 48th International Congress and Exhibition on Noise Control Engineering, Internoise 2019, pp. 6013–6022

Publisher: Institute of Noise Control Engineering (Reston, Virginia)

The effect of the rolling shutter in the perception of the sound insulation

De la Prida, Daniel¹; Pedrero, Antonio; Navacerrada, María Ángeles; Díaz-Chyla, Alexander

Grupo de Investigación en Acústica Arquitectónica, E.T.S de Arquitectura, Universidad Politécnica de Madrid.

Avda. Juan de Herrera 4. 28040 Madrid (España)

ABSTRACT

Rolling shutters or blinds are elements placed outside the window. They aim to protect the interior of dwellings from the direct exposure to sunlight. Even though they are not usual in some countries, they are commonly used in Mediterranean countries due to longer exposure to daylight. Acoustic insulation of window elements might be modified due to the shutter position. This research studies the subjective perception of the acoustic insulation of different window elements according to the position of the rolling shutter, by analyzing the correlation between the subjective perception and the usual SRI (Sound Reduction Index). For this purpose, a paired-comparison listening test is carried out for a normal-hearing heterogeneous panel of 120 people in which several window and shutter configurations are considered. The spectral features of each configuration are obtained by averaging several laboratory measurements for each configuration. Sound samples for the test are obtained by filtering the average spectral features of each configuration with typical urban sources of noise. Then, the responses of the participants are compared to the SRI by means of statistical methods.

Keywords: Window, shutter, insulation, perception

I-INCE Classification of Subject Number: 33

1. INTRODUCTION

Noise is one of the most harmful pollutants in urban environments and it does not only affect citizens when they are outside their homes but it can also affect them while inside dwellings. Outside noise can be transmitted to indoor environments through several propagation paths, the most common being the element that separates these indoor environments from the outside, that is, the façade.

The impact of noise on the population have been highly studied and there are numerous guidelines and regulations aimed at reducing environmental noise¹⁻³. A good protection of the façade, in addition to the application of guidelines for the reduction of environmental noise, can improve the quality of life and the rest of the citizens.

¹ d.delaprida@alumnos.upm.es

Facades are usually made up of a blind part, which generally provides a high protection against noise, and a hollow area. A window or a similar element, which commonly provides a weaker protection against noise, forms this hollow area.

It is important, therefore, to characterize accurately the protection against noise of these weaker façade elements.

In Mediterranean countries, such as Spain, Italy or Greece, the long and powerful exposure to sunlight requires the installation of rolling shutters on the outside of the windows. The position (fully extended or retracted) of these blinds, can affect the sound insulation of the window elements, as it has already been pointed out in ^{4,5}. It is, therefore, important to assess precisely the effect that the position of the blind has on the subjective perception of the sound insulation of window elements. It can also be interesting to evaluate the need of having into account the effect of the blind in the building design stage, in those cases where the window element is to be installed with integrated rolling shutter.

The aim of this work is to present the results of a subjective and objective evaluation of the effect of the rolling shutter in the sound insulation of window elements. For this purpose, a listening test comprising three different window elements and two shutter positions, as well as different sources of urban noise was performed by 120 participants.

First, the design and performance of the listening test are explained, focusing on the selection of the window-shutter configurations, the selection of the urban noise stimuli, the type of listening test and the steps of the experimental procedure followed by the participants. Then, the differences between the same window elements for both shutter positions are objectively assessed by comparing the 1/3 octave SRI of the fully retracted and fully extended shutter configurations. Finally, the results of the listening test are statistically analysed to determine whether the position of the shutter has some effect on the subjective perception of the window elements and how they relate with the objective results found.

2. METHODOLOGY

2.1 Design of the listening test

The design of the listening test is one of the most important stages of the subjective evaluation process, since different results and conclusions might be draw depending on its design.

In the design of a listening test several aspects must be addressed:

1. The quantity and type of the elements to be compared must be determined. In this case, the window elements with extended and retracted blinds must be selected.

2. The quantity and type of stimuli by which the elements to be compared are filtered. In this case, the different urban noise stimuli as well as pink noise.
3. The type of listening test to be used to achieve the proposed purpose. In this case, the listening test follows a pairwise comparison procedure.

In this section, each of these aspects is explained independently.

Selection of the window-shutter configurations

Given that the purpose of the study is the evaluation of the effect of the shutter's position on the sound insulation of window elements, it is necessary to select different window elements, with integrated shutter, and to consider their SRI both for extended and retracted shutter positions.

Based on a large database of laboratory window measurements generated over several years by our research group, which are described in greater depth in ^{5,6}, the SRI has been obtained for the elements the participants had to evaluate. To do this, three usual window types have been selected. Next, the average SRI for each of the types has been obtained, from all available measurements in database for windows of the same type, both for their extended and retracted blind configurations.

For the selection of elements, it was sought to obtain the maximum difference between the SRI, keeping, in turn, a relatively narrow SNQ (single number quantity) range. In particular, the selected windows are, in terms of R_w , in a range of 7 dB, between 32 and 39 dB.

Table 1 summarizes the main features of the selected elements:

Element	Frame Material	Opening	Drum material	Shutter position	Glazing	R_w
W1	PVC	Sliding	Empty	Retracted	4 / 12 / 8	32
W2	PVC	Sliding	Empty	Extended	4 / 12 / 8	36
W3	PVC	Openable	Polystyrene	Retracted	4 / 12 / 4	34
W4	PVC	Openable	Polystyrene	Extended	4 / 12 / 4	37
W5	PVC	Sliding	Polystyrene	Retracted	4 / 12 / 8	35
W6	PVC	Sliding	Polystyrene	Extended	4 / 12 / 8	39

Table 1: Main features of the window-shutter configurations selected for the listening test

As it can be seen, participants finally had to evaluate the differences between six different elements, although in this communication only the differences between the same window configuration and its fully extended and retracted shutter configurations are to be presented.

Selection of the stimuli: Noise fragments

An online survey was conducted to determine which types of urban noise affect the population the most when they are at home. The sample of participants in the online survey described the noise associated with vehicles (motorcycles, cars, buses, horns, sirens ...) as the most annoying type of noise in urban environments with 39% of the responses, followed by the noise caused by pedestrians (people talking, children during

school hours ...) with 12%. A lower percentage describes the noise caused by other infrastructures such as aircraft and rail traffic and leisure activities.

Finally, five stimuli were selected. Two of them related to traffic noise in the urban environment, one related to aircraft, another comprising pedestrian noise and a siren and, finally, pink noise.

The duration were set at 15 seconds for urban types of noise. A shorter duration of 5 seconds was set for pink noise, given the constant spectral and temporal features of this stimulus.

All the stimuli used, except for the pink noise, were recorded binaurally, by means of a Dummy Head B & K Type 4100, connected to a portable Pulse recording platform, operated through an iPad with the B & K Sonoscout application (figure 1). The sampling rate was 96 kHz and the bit rate was 32 bits.



Figure 1: Dummy Head during the recording of the urban sources of noise. Left, urban traffic. Middle, flyover of aircrafts. Right, pedestrian noise.

The urban stimuli selected were those that were determined to be more representative of each of the noise sources described. In particular, for traffic noise, a side investigation was carried out⁷ to characterize the noise in the centre of Madrid, which allowed the selection of two representative excerpts of different traffic configurations.

For the stimulus of aircraft noise, as well as for the stimulus of pedestrian noise, the selected 15 seconds excerpts were those that were more representative of the average of the psychoacoustic indicators of all the recorded events of each type.

Type of listening test

In the subjective characterization of sound insulation, listening tests are generally based on methods of scale^{8,9} and to a much lesser extent on methods based on forced choice procedures such as pairwise comparison^{10,11}. Scaling methods allow to place the evaluated elements orderly according to certain attribute on which the participants are consulted. However, they do not allow direct comparison between pairs of elements. Therefore, in this case, where the purpose of the study is to assess the perceptual differences between the retracted and extended shutter positions for the same window elements, the use of the pairwise comparison procedure was found to be appropriate.

A pairwise comparison approach was therefore selected, in which, the participants were presented, for each stimulus, with the six different elements in pairs, and they were asked to choose the one that was the most annoying of the two. In addition, participants could cast a blank vote, in case they did not perceive differences, although they were urged to give an answer as much as possible.

The minimum number of comparisons to be performed by participants, for a pairwise comparison procedure without repetitions, follows equation 1:

$$\# \text{ comparisons} = N * (N - 1)/2 \quad (1)$$

N determines the number of comparisons, which is the number of different elements to be assessed by each participant. Since six different window elements (W1-W6) were selected, N can be set to 6. Then, considering equation 1, the number of comparisons to be performed by each participant for each stimulus is of 15 comparisons. Given that the listening test had to be performed for five different stimuli, the total amount of comparisons to be performed by each participant was of 75. Therefore, considering that the average duration of the stimuli was of 13 seconds, it would require around 37 minutes for each participant to finish the experiment.

In order to ensure the independence of the results, which is one of the necessary statistical conditions in this type of study, the comparisons were presented in random order for each participant, to avoid, as much as possible, the bias due to learning and fatigue of the participant.

2.2 Performance of the listening test

This section describes the environment and equipment used for the performance of the listening test, as well as the different stages of the process.

Test Environment

The tests were carried out in the Laboratory for Acoustics and Vibrations of the Technical School of Architecture (Technical University of Madrid). A DEMVOX ECO100 insulated booth was used to ensure low background noise and comfort. Only a screen, a mouse, a push button and headphones were installed inside the booth. The rest of the equipment was installed outside to avoid undesired sources of noise.

The playback chain consisted of a laptop connected to an RME Fireface UFX audio interface, a RANE HC6S headphone amplifier and Sennheiser HD650 headphones. The reproduction chain was calibrated to present the sound samples at the same level at which they were recorded. In addition, the frequency response of the headphones was taken into account to counteract its effect, as much as possible.

Experimental procedure

The participants were first welcome on their arrival and invited to enter the booth. Then, they were asked to answer an anonymous demographic survey regarding age, gender, nationality and level of education. This survey allowed to gather information that

might be helpful during the analysis stage but also to help the participants to become familiar with the test environment.

Once they were finished with the survey and they were familiar with the test environment, a document was provided where a detailed explanation of the process could be found. Then, the participants were asked regarding any doubts they might have with the experimental procedure.

Following, the participants were presented, on the computer screen, with the formulary that must be answered and a short training of two comparisons was carried out, in order to assess the real understating of the test they were going to perform. After this familiarization phase were finished, the actual test began.

The listening test was performed as five individual tests, one for each stimulus, with a short break between each of them. During the break, avoiding it to look as part of the evaluation, the participant was interviewed about the previous finished test. Interesting answers to the interview were noted for further consultation, in case they were necessary.

Once the five tests were completed, all participants were given a screening audiometry to know their hearing status. Each participant received the result of the audiometry which was then explained in detail.

2.3 Analysis of the results

The results of the listening test were analysed according to the method described in ISO 5495: 2005¹². This method, though, does not consider that the participants might not make a judgment. Therefore, those cases were removed from the dataset prior to analysis.

For each pair of comparisons there were two results, describing the number of times each of the elements in the pair was considered as more annoying as the other. As explained in ISO 5495: 2005¹² any of them should rise a threshold, which is also given in the standard, so it can be stated that any of them was significantly more annoying than the other.

Depending on the total number of comparisons that were performed for a pair of elements, which can differ from case to case because of what it has been stated in the first paragraph of this section, a different threshold was established. This threshold also depend on the required significance at a certain α -level of confidence. Then, the number of comparisons was accounted, in which each of the elements of the pair was selected as more annoying. If either of the two elements exceeded the threshold, it could be determined that, for a certain level of significance a perceptible difference existed between the elements of the pair, regarding perceived annoyance.

For this research, the α -level was selected to be of 0.05.

3. RESULTS

This section shows the results of both the objective and subjective analysis, of the difference between the extended and retracted blind positions, for the same window element.

First, the results of the objective analysis are shown. Then, the subjective results are presented.

3.1 Objective comparison

For each of the three window elements, two different window configurations, regarding the position of the blind are presented. Figures 2 shows the overlapping between the 1/3 octave SRI for elements W1 and W2, W3 and W4 and W5 and W6 from left to right respectively. Below each comparison, an additional graph shows the difference between the first and the second element of each pair. For each band, positive values mean that the first element of the pair has a higher SRI. Therefore, negative values show that the second element of the pair has a higher SRI.

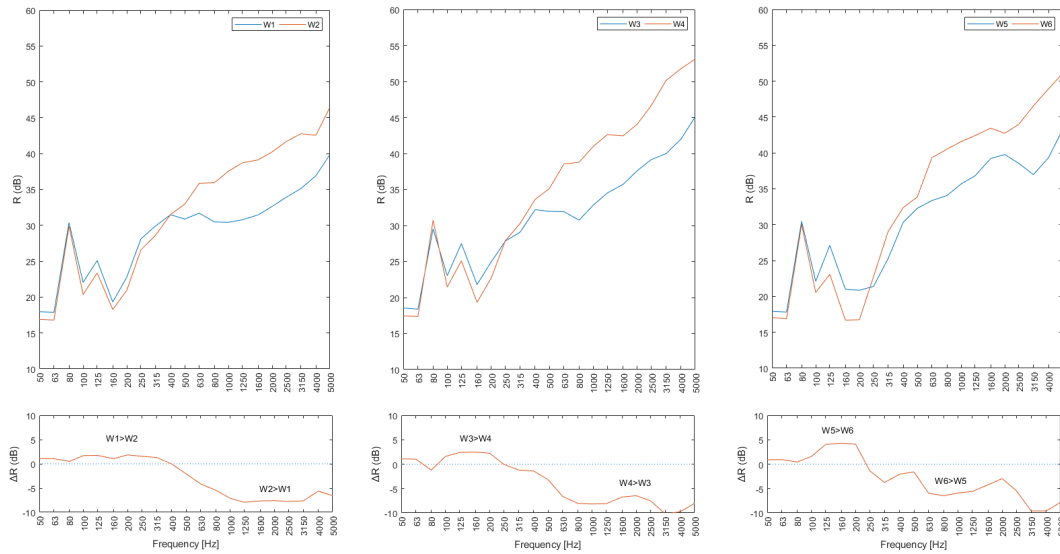


Figure 2: Overlapping of each pair of elements. Left: W1-W2, Middle: W3-W4, Right: W5-W6. Below: Difference between the first and the second element of each pair.

For all the cases it can be seen that the insulation is higher in the range of medium and high frequencies for the element in which the shutter is fully extended (W2, W4 and W6), while in the low frequencies the insulation decreases regarding to the fully retracted shutter configuration.

In particular, in the case W1-W2 (Figure 1, left), the inversion occurs at the frequency band centered at 400 Hz. The element W1 (fully retracted blind) has higher sound insulation below that band, with an average difference against W2 of 1.3 dB and a maximum difference of 1.8 dB in the band centered at 200 Hz band. Above the inversion frequency, the element W2 (fully extended blind) has higher sound insulation than W1, with an average difference against W1 of 5.8 dB and a maximum difference of 7.9 dB in the band centered at 1250 Hz.

Similarly, in the case W3-W4 (Figure 1, middle), the frequency band at which the inversion occurs is centered at 250 Hz. The element W3 (fully retracted blind) has a higher sound insulation below that band, with an average difference versus W4 of 1.8 dB and a maximum difference of 2.5 dB in the 160 Hz band. Above the inversion frequency, the element W4 (fully extended blind) has a greater sound insulation than W3, with an average difference versus W3 of 6.1 dB and a maximum difference of 10.2 dB in the 3150 Hz band.

Finally, in the case W5-W6 (Figure 1, right), the frequency band at which the inversion occurs is centered at 200 Hz. The element W5 (fully retracted blind) has a higher sound insulation below that band, with an average difference against W6 of 2.3 dB and a maximum difference of 4.3 dB in the 160 Hz band. Above the inversion frequency band, the element W6 (fully extended blind) has a higher sound insulation than W5, with an average difference against W5 of 5.2 dB and a maximum difference of 9.6 dB in the 3150 Hz band.

3.2 Listening test results

Tables 2, 3 and 4 present the decision results reported by the participants, for the cases W1-W2, W3-W4 and W5-W6 and each of the used stimuli. The participants also compared the rest of possible combinations of elements and their results will be presented in the future.

	Traffic 1	Traffic 2	Aircraft	Pedestrian	Pink
#W1	61	30	36	102	77
#W2	38	78	73	8	32
#total	99	108	109	110	109
Threshold	60	65	66	66	66
Most Annoying	W1	W2	W2	W1	W1

Table 2: Number of times W1 and W2 were chosen as more annoying. Total number of judgments. Threshold for alpha-level of 0.05. Most annoying element for that level of significance.

	Traffic 1	Traffic 2	Aircraft	Pedestrian	Pink
#W3	89	69	76	106	71
#W4	17	31	29	7	36
#total	106	100	105	113	107
Threshold	64	61	64	68	65
Most Annoying	W3	W3	W3	W3	W3

Table 3: Number of times W3 and W4 were chosen as more annoying. Total number of judgments. Threshold for alpha-level of 0.05. Most annoying element for that level of significance.

	Traffic 1	Traffic 2	Aircraft	Pedestrian	Pink
#W5	79	47	61	89	60
#W6	29	50	49	20	48
#total	108	97	110	109	108
Threshold	65	59	66	66	65
Most Annoying	W5	-	-	W5	-

Table 4: Number of times W5 and W6 were chosen as more annoying. Total number of judgments. Threshold for alpha-level of 0.05. Most annoying element for that level of significance.

It can be seen that for the case W1-W2 (table 2), depending on the type of stimulus, one element of the pair or the other was selected significantly as more annoying by the participants.

In the case W3-W4 (table 3), regardless of the stimulus that was selected, the participants perceive the element W3 (fully retracted blind) as more annoying.

Finally, for the case W5-W6 (table 4), only for two of the stimuli the participants were able to make a significant judgment. In those cases, the element W5 (fully retracted blind) was selected as more annoying. For the other cases, neither of the two elements exceeded the threshold, although in two of these cases the tendency was to select the element W5 as more annoying.

4. CONCLUSIONS

The fact of placing the shutter in different positions considerably affects the acoustic insulation of windows. Therefore, between the two extreme shutter situations (fully retracted and extended), the acoustic behavior of the element can vary drastically. A frequency can be found, in all the cases, where the difference between the elements of the pair shifts its behavior and the element with higher sound insulation becomes the element with lower sound insulation and vice versa.

Regarding the subjective perception of these objective differences, it can be stated that the participants perceived the position of the blind as changes in the sound insulation. In general, the position for which the shutter was fully extended was perceived as less annoying.

However, depending on the particular window element, the differences between the retracted and extended shutter configurations and the type of stimulus, variations in this behavior may occur.

5. ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) [BIA 2015-68914- R].

6. REFERENCES

1. Berglund, B., Thomas, L. & Dietrich, H. S. Guidelines for community noise. *WHO Expert Task Force Meet. Guidel. Community Noise* (1999). doi:10.1260/0957456001497535
2. World Health Organization. Environmental health indicators for Europe: a pilot indicator-based report. (2004).
3. Hurtley, C. *Night noise guidelines for Europe*. (WHO Regional Office Europe, 2009).
4. Díaz, C. & Pedrero, A. An experimental study on the effect of rolling shutters and shutter boxes on the airborne sound insulation of windows. *Appl. Acoust.* **70**, 369–377 (2009).
5. Díaz, C., Díaz, A. & Navacerrada, M. A. An experimental study on the effect of rolling shutters on the field measurements of airborne sound insulation of façades. *Appl. Acoust.* **74**, 134–140 (2013).

6. Díaz, C. & Pedrero, A. An experimental study on the effect of rolling shutters and shutter boxes on the airborne sound insulation of windows. *Appl. Acoust.* **70**, 369–377 (2009).
7. de la Prida, D., Pedrero, A., Navacerrada, M. Á. & Díaz, C. Relationship between the geometric profile of the city and the subjective perception of urban soundscapes. *Appl. Acoust.* **149**, 74–84 (2019).
8. Park, H. K. & Bradley, J. S. Evaluating standard airborne sound insulation measures in terms of annoyance, loudness, and audibility ratings. *J. Acoust. Soc. Am.* **126**, 208–19 (2009).
9. Hongisto, V., Oliva, D. & Rekola, L. Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. *J. Acoust. Soc. Am.* **144**, 1100–1112 (2018).
10. Monteiro, C., Machimbarrena, M., de la Prida, D. & Rychtarikova, M. Subjective and objective acoustic performance ranking of heavy and light weight walls. *Appl. Acoust.* **110**, 268–279 (2016).
11. Rychtáriková, M. *et al.* Perceived Loudness of Neighbour Sounds Heard Through Heavy and Light-Weight Walls with Equal R_w +C50–5000. *Acta Acust. united with Acust.* **102**, 58–66 (2016).
12. *ISO 5495:2005 Sensory Analysis - Methodology - Paired comparison test.* (2005).

A.1.3 Paper G

Title: Perceptual validation of virtual acoustic models

Authors: A. Pedrero, D. de la Prida, L. Marandet, J. L. Sánchez, M. A. Navacerrada, C. Díaz

Conference: 48th International Congress and Exhibition on Noise Control Engineering, Internoise 2019

Conference location: Madrid, Spain

Conference date: 16-06-2019 to 19-06-2019

Proceedings: Proceedings of the 48th International Congress and Exhibition on Noise Control Engineering, Internoise 2019, pp. 6923–6931

Publisher: Institute of Noise Control Engineering (Reston, Virginia)

Perceptual Validation of Virtual Acoustic Models

Pedrero, Antonio¹; 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 3382-1¹ 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 application². 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 impression³.

The perceptual validation of acoustic models is not a new issue. Several publications on this subject can be found in the literature⁴⁻⁷. 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 used⁸. 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 others⁹.

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 test¹⁰.

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-head¹¹, 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-AFCR¹². 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 high¹³.

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.

Listen carefully both the reference and the sound S1. Choose "Yes" if both the reference and S1 were the same and "No" otherwise. Additionally, in some of the cases, after answering to the "Yes/No" question, further questions will be presented. Then:

1. Mark, using the scale, the degree of perceived difference between the Reference and S1.
2. Also using the scales, mark whether the Reference of S1 was more in agreement with the asked questions.
3. Are the Reference and S1 real recordings or simulations? Both the Reference and S1 can be recordings or simulations and can belong to the same or different points of recording in the venue. Mark "Don't know / Can't make a choice" if you can't make a statement.

Stimuli Playback

Reference

S1

¿S1 = Reference?

☐ Yes

☒ No

How different?

Extremely different

Very different

Moderately different

Slightly different

Not at all different

Attributes of the stimuli

Which of the sound stimuli sounds clearer? Left: Reference is clearer. Right: S1 is clearer

Which of the two stimuli sounds more natural? Left: Reference is more natural. Right: S1 is more natural

For which of the two stimuli do you perceive yourself being more in the scene? Left: More present for the Reference. Right: More present for S1

More Ref.
2
3
4
More S1

More Ref.
2
3
4
More S1

More Ref.
2
3
4
More S1

Considering both stimuli... Select the statement in agreement with your perception

☐ Ref=Real, S1=Simulation

☐ Ref= Real P1, S1= Real P1

☐ Ref= Simulation P1, S1= Simulation P2

☐ Ref=Simulation, S1=Real

☐ Ref=Real P1, S1= Real P2

☐ Don't know / Can't make a choice

Next

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 America* 140.6 (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)

A.1.4 Paper H

Title: Influencia del estímulo musical sobre la percepción de la claridad C80

Authors: M. Larrosa, A. Pedrero, D. de la Prida

Conference: XI Congreso Ibérico de Acústica, Acústica 2020

Conference location: Faro, Portugal (Online)

Conference date: 21-10-2020 to 23-10-2020

Proceedings: Proceedings of Acústica 2020, pp. 395–405

Publisher: Sociedade Portuguesa de Acústica (Lisbon, Portugal)

INFLUENCIA DEL ESTÍMULO MUSICAL SOBRE LA PERCEPCIÓN DE LA CLARIDAD C_{80}

María Larrosa¹, Antonio Pedrero², Daniel de la Prida³

¹ Universidad Politécnica de Madrid
{e-mail: m.larrosa@upm.es}

² Grupo de Investigación en Acústica Arquitectónica, Escuela Técnica Superior de Arquitectura, Universidad Politécnica de Madrid, Avda. Juan de Herrera 4, 28040, Madrid, España-
{e-mail: antonio.pedrero@upm.es}

³ Universidad Politécnica de Madrid
{e-mail: danidelaprida@gmail.com}

Resumen

La Claridad C_{80} ha sido ampliamente utilizada para la determinación de la claridad musical de recintos desde su definición en 1974. A pesar de lo extendido de su uso, no se ha investigado en profundidad que este índice tenga una buena correlación con la percepción subjetiva de la claridad para cualquier entorno acústico. En esta comunicación se presentan los resultados obtenidos del análisis de la correlación existente entre el C_{80} y la percepción subjetiva para distintos estímulos y recintos. La información sobre la percepción subjetiva de la claridad se ha obtenido mediante la realización de un test psicoacústico a un conjunto de 18 participantes. Se utilizaron cinco estímulos musicales y 18 respuestas al impulso obtenidas mediante la simulación de tres recintos acústicos.

Los resultados obtenidos muestran que el estímulo musical constituye un factor significativo sobre la percepción que los oyentes tienen sobre la claridad de un recinto. De este modo se concluye que el C_{80} presenta posibilidades de mejora para mejorar su relación con la percepción subjetiva.

Palabras clave: acústica de salas, test psicoacústico, índices acústicos objetivos, claridad, C_{80}

Abstract

The acoustic parameter C_{80} has been widely used in the determination of the musical clarity of a venue since its definition in 1974. Despite its extended use, an exhaustive research has not been conducted which can conclude whether this index has a good correlation with the subjective perception of the musical clarity for certain kinds of music styles. This present communication shows the analysis of the correlation between the C_{80} and the subjective perception of the clarity for various musical stimuli and various venues. The information about the subjective perception of the clarity has been obtained from a listening test carried out by a set of 18 participants. The stimuli used were a total of five musical pieces and eighteen impulse responses obtained through the simulation of three acoustic halls.

The obtained results show that the musical stimuli were a significant factor in the perception of the clarity of the hall. Therefore it can be concluded that the C_{80} presents room for improvement in order to enhance its correlation with the subjective perception.

Keywords: room acoustics, listening test, objective acoustic indices, clarity, C_{80}

PACS nº. 43.66.Lj

1 Introducción

La investigación sobre el comportamiento acústico de las salas se remonta a principios del siglo XX [1]. El progreso en este campo se ha desarrollado principalmente en las tres últimas décadas debido al avance de otras tecnologías como la electrónica y la informática, que han permitido una mejora en los equipos de captación y en el tratamiento de las señales.

El interés que ha suscitado este campo se debe a una motivación cada vez mayor de crear espacios con propiedades acústicas mejor adaptadas al tipo de actividad que se va realizar en ellos de manera predominante (se puede hablar así de salas de conciertos, de conferencias, de cine, etc.), así como de mejorar los recintos ya existentes con el fin de obtener el mejor comportamiento acústico posible.

En el último siglo, tras la determinación del tiempo de reverberación como el primer descriptor de la acústica de salas por parte de W. C. Sabine en 1922 [2], han sido numerosas las aportaciones a la descripción de la acústica de salas mediante parámetros objetivos. Especialmente relevantes son las aportaciones en el campo de los índices acústicos de Leo L. Beranek [3]. Basándose en estudios realizados por éste en más de 100 salas se definieron un total de 25 parámetros, que él denominaba atributos, que definían todos los aspectos importantes de la música interpretada en la acústica de un recinto cerrado. Algunos de estos atributos han sido desarrollados para definir la percepción subjetiva de la acústica del recinto y han dado lugar a parámetros objetivos que se pueden definir mediante una fórmula matemática.

Los parámetros objetivos se encargan de cuantificar las cualidades acústicas de las salas mediante la realización de mediciones o estimaciones. Algunos de estos, como el tiempo de reverberación inicial, la fuerza sonora y el tiempo central, se consideran fieles descriptores de las características acústicas de los recintos, lo que ha llevado a la estandarización de sus métodos de medición en normas ISO, con el fin de contar con un marco metodológico común. El índice objetivo de interés en esta comunicación es la relación entre la energía precoz y la energía tardía, también denominado claridad C_{80} , definida por primera vez en 1974 por Reichard y Abdel Alim [4]. La norma UNE-EN ISO 3385-1:2010 [5] define la claridad como:

“... la relación entre las energías acústicas precoces y tardías. Se puede calcular para un límite precoz de 50 ms o de 80 ms, dependiendo de si los resultados se refieren a las condiciones de las salas de conferencia o de conciertos, respectivamente, utilizando la ecuación:

$$C_{te} = 10 \log \frac{\int_0^{t_e} p^2(t) dt}{\int_{t_e}^{\infty} p^2(t) dt} \text{ dB} \quad (1)$$

El índice C_{80} también se conoce como la claridad musical de un recinto. Este parámetro se calcula como el promedio del valor obtenido para las octavas de 500 Hz y 1 kHz. Se considera que cuanto mayor sea el valor de este término más sensación de claridad presentará la sala, ya que más predominará el sonido directo con respecto a la reverberación.

En 2010, Bradley [6] presentó una investigación que tuvo como fin realizar un estudio crítico de todos los errores existentes en el cálculo y la interpretación de los parámetros acústicos definidos en [5]. Considera que, en general, no se han realizado suficientes investigaciones que determinen cómo utilizar adecuadamente los parámetros para obtener la mejor correlación con la percepción subjetiva. El mayor problema encontrado por Bradley para los parámetros de claridad C_{50} y C_{80} es la inexistencia de estudios que determinen el método ideal con el que obtener un valor único para definir la claridad de una sala, ya que los valores de estos parámetros varían considerablemente en función de la frecuencia.

Actualmente sólo se consideran las octavas de 500 Hz y 1 kHz, dejando fuera mucha información con respecto al comportamiento de la sala. A pesar de ello, debido a la falta de estudios subjetivos que den más información sobre los distintos índices de claridad, Bradley considera que el C_{50} y el C_{80} son los índices más adecuados para determinarla.

Para abordar este problema a lo largo de los años se han realizado *listening test* cuyo objetivo era la evaluación subjetiva de los descriptores objetivos de la acústica de salas. Tal y como se puede observar en la revisión de *listening test* en acústica de salas hecha en [7], en la mayoría de los test realizados [8, 9, 10, 11] el estudio se centraba en la determinación del umbral diferencial (JND) de la claridad C_{50} y C_{80} y otros parámetros acústicos; mientras que en [12] el propósito era evaluar la idoneidad de diversos índices objetivos como predictores subjetivos. Los estudios [9, 10] utilizaron el método de “*Minimal Changes*”; en [8, 11] el método “*Same-different*”; y en [12] el método “*Attribute 2-AFC*”. El primer método mencionado se encuentra basado en el segundo, cuyo principal problema es que es propenso a sesgos de tendencia [13]. También resaltar la importancia de la forma en la que se pregunta a los participantes y cómo estos interaccionan con los diversos estímulos. De especial interés son los resultados de [11], donde se observa que distintos procedimientos para la realización del *listening test* dan lugar a resultados distintos.

Otro punto de gran importancia a la hora de definir un *listening test* es el número de estímulos que se van a utilizar y sus características. En los cinco estudios de interés entre los evaluados en [7] se ha podido observar un total de nueve estímulos musicales pero solo uno de ellos ha sido utilizado en dos de las pruebas. También es importante mencionar que el número de estímulos utilizados en cada *listening test* es muy bajo, siendo habitualmente uno o dos, excepto en el caso de [11] donde se han utilizado cuatro. A pesar de que en algunos casos se ha justificado la selección de los estímulos no se han encontrado antecedentes claros que sostengan estas elecciones.

El objetivo de esta investigación es determinar la correlación existente entre el parámetro C_{80} y la percepción subjetiva en función del estímulo musical. La caracterización de la percepción subjetiva de la claridad se obtuvo mediante un *listening test* realizado por un conjunto de 18 participantes, para el que se usaron 18 respuestas al impulso obtenidas mediante la simulación de tres recintos de diferentes tamaño y cinco fragmentos musicales diferentes.

Con el fin de obtener un conjunto de fragmentos musicales representativos, con anterioridad a la realización del *listening test*, se llevó a cabo un estudio de la modulación musical de un conjunto de 69 piezas grabadas en cámara anecoica. La intención de este análisis era entender cómo se ve afectada la modulación musical en función del *tempo* de la pieza, de su instrumentación y de la figuración que presenta la misma. Los resultados de este análisis, que se encuentran más allá del alcance de esta comunicación, permitieron seleccionar cinco fragmentos musicales de distintos estilos y *tempo* para comprobar si estas características son significativas a la hora de percibir la claridad de un recinto.

2 Metodología

Para caracterizar la percepción subjetiva de la claridad se realizó un *listening test* en el que los participantes tuvieron que evaluar la claridad de un total de 90 muestras sonoras. Estas muestras se obtuvieron convolucionando los estímulos musicales seleccionados con diferentes respuestas al impulso, que se calcularon mediante la modificación del nivel de claridad de los modelos acústicos de tres salas. Los participantes tuvieron que caracterizar el grado de claridad que percibieron en cada muestra mediante una escala de valores de cinco posiciones.

A continuación se describe la metodología seguida en la investigación. En primer lugar se va a explicar la implementación del *listening test*, justificando los estímulos utilizados y el tipo de test que se eligió utilizar. Seguidamente, se expone el entorno de realización del test y el procedimiento que se siguió con cada uno de los participantes. Finalmente, se explica el método de análisis empleado en la evaluación de los resultados obtenidos.

2.1 Diseño del *listening test*

Como se ha expuesto anteriormente el método utilizado para la realización del *listening test* y los estímulos utilizados en el mismo pueden tener una gran influencia sobre los resultados obtenidos. Por lo tanto, a la hora de realizar una evaluación subjetiva, el diseño del *listening test* presenta especial relevancia, a fin de reducir o controlar posibles fuentes de sesgo que puedan afectar a la calidad de los resultados. Los tres elementos más importantes a tener en cuenta en su diseño son: a) los elementos que se van a evaluar, b) los estímulos que se van a utilizar para su evaluación y c) el método de test utilizado. En la presente investigación, el elemento a evaluar fue el nivel de claridad subjetiva, los estímulos fueron cinco fragmentos musicales de distintas características y el tipo de test utilizado una escala de Likert de cinco puntos.

Determinación y cálculo de los niveles de claridad

Dado que el propósito del estudio era la evaluación de la correlación existente entre el nivel objetivo de la claridad C_{80} y la percepción subjetiva que se tiene de la claridad en función del estímulo musical, se determinó que lo ideal sería disponer de un amplio margen de variaciones de la claridad con el que poder estudiar su variación perceptual. Los recintos seleccionados para obtener las respuestas al impulso fueron una iglesia, una sala de conciertos y una sala de cámara, cuyas características iniciales se presentan en la tabla 1. La utilización de modelos acústicos de estos recintos permitieron la modificación de sus características con el fin de obtener el margen de variaciones de claridad deseado.

Tabla 1 – Características de los recintos utilizados para el cálculo de las respuestas al impulso.

Nombre del recinto	C_{80} (dB)	T_{30} (s)
Iglesia de San Cebrián de Mazote	-5,90	2,88
Sala de conciertos Jönköpin	-1,75	1,99
Auditorio del Campus Sur de la UPM	2,75	1,11

A la hora de determinar el número de respuestas al impulso a utilizar y los saltos en su nivel de claridad C_{80} se utilizó como referencia el estudio realizado por Bradley en [8]. En este trabajo, Bradley trataba determinar el JND para el parámetro acústico C_{50} . Para la realización del *listening test* se utilizó una respuesta al impulso base y, en cada prueba, los sujetos debían comparar esta señal con entornos acústicos cuyo salto de nivel del C_{50} con respecto a la señal base era de 0.0, 0.5, 1.0, 1.5, 2.5 y 4.0 dB.

A pesar de que el propósito del test fuese diferente, se utilizó el mismo método de determinación de los saltos de claridad C_{80} para la obtención de las respuestas al impulso. Las tres claridades base son las correspondientes a los recintos seleccionados. A partir de este valor, se modificaron las absorciones de los materiales de las salas para hacerlas más absorbentes, manteniendo en todo momento el balance de los elementos con el fin de que la sensación de espacialidad de la salas no se viera modificada entre señales. El valor final de la claridad C_{80} y del tiempo de reverberación T_{30} para cada una de las respuestas al impulso se presentan en la tabla 2.

Tabla 2 – Claridades C_{80} y tiempos de reverberación T_{30} para las respuestas al impulso calculadas.

Iglesia San Cebrián		Sala de conciertos Jönköping		Auditorio Campus Sur	
C_{80} (dB)	T_{30} (s)	C_{80} (dB)	T_{30} (s)	C_{80} (dB)	T_{30} (s)
-5,90	2,88	-1,75	1,99	2,75	1,11
-5,40	2,64	-1,25	1,98	3,30	1,01
-4,90	2,44	-0,70	1,74	3,80	1,02
-4,40	2,27	-0,20	1,64	4,25	0,95
-3,35	1,96	0,80	1,36	5,25	0,86
-1,90	1,61	2,30	1,14	6,75	0,73

Selección de los estímulos

Las muestras sonoras utilizadas durante la prueba fueron cinco fragmentos de piezas musicales grabadas en cámara anecoica, que posteriormente se convolucionaron con las respuestas al impulso obtenidas de las salas. Así, cada participante evaluó un total de 90 muestras sonoras de una duración media de 16,8 s.

Los estímulos se seleccionaron después de la realización de un estudio de la modulación musical de un conjunto de 69 piezas. La selección se basó en sus diferencias de instrumentación, *tempo* y estilo con el fin de poder así comprobar si el estímulo musical se trata de un factor relevante a la hora de percibir la claridad de un recinto. Las piezas finalmente utilizadas en la realización del test fueron:

1. Concierto para flauta No. 1 en Sol Mayor KV 313 de Wolfgang Amadeus Mozart: fragmento de flauta solista de *tempo* rápido. La duración del estímulo es de 24 s.
2. Elegía op. 24 para violonchelo y orquesta de Gabriel Fauré: pasaje de chelo solista de *tempo lento*. La duración del fragmento es de 18 s.
3. Novena Sinfonía en Re Menor op. 125 de Ludwig van Beethoven: fragmento formado por los siete primeros compases del cuarto movimiento, consistente en un fragmento orquestal de *tempo* rápido. La duración del estímulo es de 8 s.
4. Preludio de *La Traviata* de Giuseppe Verdi: pasaje de *tutti* orquestal de *tempo* lento y línea de bajo muy marcada. La duración es de 20 s.
5. Canto litúrgico de duración de 14 s. Se trata de una pieza musical del repertorio de canto hispánico donde un cantante solista interpreta la melodía y es contestado por un coro.

La distinta duración de los estímulos se debe a que se seleccionaron pasajes musicales que terminasen en una cadencia conclusiva. Esta decisión se tomó debido a que el corte del fragmento a mitad de una frase musical hubiera producido una tensión involuntaria en los participantes, influenciando así su concentración y la correcta realización del test.

Método de *listening test* utilizado

El objetivo del *listening test* fue obtener información sobre la claridad percibida por los oyentes para los distintos niveles de claridad C_{80} y para todas las piezas seleccionadas. Para ello se tomó la decisión de dividir el test en tres partes, una para cada uno de los recintos evaluados, y que en cada parte se evaluaran los estímulos musicales de forma independiente para todos los niveles de claridad C_{80} .

Se consideró que el método más adecuado para la realización del test, dado que la evaluación era individual y no mediante métodos de diferencia, era mediante una escala de Likert de cinco puntos [14]. Se trata de un método basado en una escala de categoría donde se fijan cinco posiciones con una etiqueta que define la respuesta para cada posición. El número de posiciones se seleccionó, basándose en la

investigación realizada en [15] donde se comprueba que a partir de seis puntos la precisión de la prueba no aumenta y, en contraposición, genera mayor fatiga y dudas al participante en la selección de la respuesta. En cuanto a la elección de las etiquetas, se consultó la bibliografía existente y se encontró que durante mucho tiempo se consideró que la posición en la que dichas etiquetas se colocaran tenía una gran importancia. A lo largo de los años se realizaron diversos estudios en la materia [16, 17, 18] y las investigaciones de Weng [17] y Subedi [18] ponen de relieve que el orden de presentación de las etiquetas no afecta a los resultados. El vocabulario utilizado para la pregunta y las explicaciones se basaron en el glosario definido por Alexander Lindau [19] y el orden de las etiquetas se presentó según la norma UNE-ISO 4121 [20].

La pregunta a contestar consiste en indicar cómo de alta se considera la claridad de cada una de las muestras sonoras presentadas mediante la escala de cinco puntos; siendo las etiquetas, de izquierda a derecha: “nada alta”, “algo alta”, “moderadamente alta”, “muy alta” y “extremadamente alta”.

A la hora de seleccionar los participantes que iban a realizar la prueba se tuvo en cuenta que el método de *listening test* utilizado se trataba de un método basado en atributos. En [21] se describe que el uso de pruebas atributivas sobre participantes sin experiencia puede suscitar la aparición de sesgos adicionales, siendo, por lo tanto, necesaria la utilización de participantes entrenados. En esta investigación se seleccionó que los participantes tuvieran como mínimo un título de Grado Profesional de Música. De los participantes seis contaban con un Título Superior de Música en la especialidad de Interpretación; uno con un Título Superior de Música en la especialidad de Dirección de Orquesta; y los restantes con un Título de Grado Profesional de Música.

2.2 Realización del *listening test*

En esta sección se va a describir el entorno en el que se llevó a cabo el test, el procedimiento que se siguió para su realización y la duración del mismo.

Entorno de realización

La realización del *listening test* a cada uno de los participantes tuvo lugar en una sala vacía, de pequeño tamaño y con un nivel de ruido de fondo muy bajo.

El equipo utilizado fue un ordenador portátil modelo HP-BS512NS y unos auriculares circumaurales de diadema abiertos Sennheiser HD-650. El nivel de volumen utilizado en los test fue siempre el mismo y se especificó a los participantes que este no se podía modificar en ningún momento.

Procedimiento

Tras el recibimiento de los participantes se les entregó una encuesta de datos demográficos, en la que debían rellenar algunos datos relativos a la edad, sexo, nivel educativo, etc... Esta encuesta tenía un doble propósito: por un lado, servía de método de aclimatación al entorno de la prueba, con el fin de que el participante se fuera acomodando al ambiente en el que iba a realizar el test y, por otro, con el fin de disponer de información adicional para las posteriores etapas de análisis. Una vez finalizada la encuesta, se le entregaba una hoja de instrucciones donde se detallaba el procedimiento del test y a qué tipo de pregunta se iba a enfrentar cada participante, con el fin de reducir el sesgo debido a la interacción entre el participante y el experimentador.

Con anterioridad al comienzo del test, el participante realizó una prueba de entrenamiento, con el fin de que se familiarizara con la interfaz utilizada y las preguntas a contestar, con el propósito de reducir un posible sesgo de interacción hombre-máquina. Las muestras sonoras utilizadas en este test de entrenamiento fueron una señal de palmadas convolucionada con distintas respuestas al impulso, que en ningún caso se utilizaron durante el desarrollo de la prueba principal.

La prueba podía, entonces, dar comienzo. Ésta se encontraba dividida en tres partes diferentes, cada una dividida, a su vez, en cinco pestañas, una para cada uno de los estímulos musicales bajo evaluación. Cada una de las tres partes se espaciaron en el tiempo, con un periodo de descanso de aproximadamente cinco minutos entre cada una de ellas, para minimizar el efecto que la fatiga y la memoria auditiva pudiesen tener sobre las caracterizaciones posteriores. El orden de presentación de las salas, los estímulos y las claridades C_{80} era aleatorio para cada participante, asegurando así una independencia estadística en las respuestas.

Una vez se iniciaba el test el participante, en primer lugar, estaba obligado a escuchar las seis muestras presentes en cada pantalla (un estímulo musical convolucionado con todas las respuestas al impulso de la sala bajo evaluación) con el fin de tener una referencia de la sala y así evitar la aparición de sesgo debido a la no selección de las posiciones extremas de la escala. Posteriormente, el participante podía escuchar las muestras tantas veces como deseara y se le instaba a dar una respuesta individual para cada una de ellas.

Durante la realización de la prueba se cronometraron las duraciones de cada parte del test y de los descansos y, una vez terminados todos, se obtuvo que el promedio de la duración del test fue de una hora y seis minutos.

2.3 Análisis de los resultados

Después de la realización del *listening test*, se procedió al análisis estadístico de los juicios perceptuales. Este análisis tuvo como propósito comprobar la hipótesis de que la claridad percibida por los participantes para un mismo recinto depende significativamente de la pieza musical interpretada.

Para el análisis se decidió utilizar la moda de las respuestas dadas. Se tomó esta decisión ya que los participantes únicamente pudieron contestar en posiciones fijas que representan números enteros y el empleo de la media implicaría la utilización de decimales que introducirían un sesgo en el estudio. Al utilizarse una escala de cinco puntos, para cada combinación de sala, estímulo y claridad C_{80} se obtendrá un resultado de la claridad percibida por cada participante en una escala entre 1 y 5.

Para comprobar la validez de la hipótesis se ha realizado un análisis de regresión, cuyo objetivo es explicar el comportamiento de una variable Y (de la cual no se conoce su comportamiento) a partir de otra variable X (controlada por el investigador). Una forma de representar gráficamente la regresión es mediante la definición de una nube de puntos que muestra el valor que toma Y para cada X y, posteriormente, mediante el cálculo de la ecuación de la recta que mejor se ajuste a dicha nube de puntos; esta recta se denomina recta de regresión.

La medida más importante para determinar la bondad de ajuste de una recta de regresión es el coeficiente de determinación R^2 que indica el grado de ajuste entre la nube de puntos y la recta de regresión calculada; toma valores entre $[0, 1]$. Cuando éste está próximo a cero indica que la capacidad de predicción de la recta de regresión es baja; mientras que si su valor está próximo a uno la fuerza de asociación entre ambas variables es muy alta.

El análisis de regresión se ha realizado para cada conjunto de estímulo y sala de forma individual con el fin de obtener una mejor representación del comportamiento de la percepción subjetiva.

3 Resultados

En este apartado se van a presentar los resultados obtenidos para la comparación realizada entre los valores objetivos de la claridad C_{80} y la claridad subjetiva percibida por los participantes del *listening test*.

En el análisis de regresión se consideró que la claridad C_{80} es la variable independiente y la claridad percibida por los sujetos es la variable dependiente. Este análisis se realizó para cada una de las salas y cada uno de los estímulos por separado ya que la evaluación subjetiva a través del *listening test* se realizó de este modo. Para cada una de las salas se tiene una escala de 1 a 5, independientemente de la claridad C_{80} de referencia de la sala, por lo que el estudio conjunto de las salas no tendría sentido puesto que se atribuirían valores de claridad percibida muy similares a juicios con diferentes niveles base de claridad.

Tabla 3 – Coeficientes de determinación para cada combinación de sala y estímulo en porcentaje. () $p > 0.15$; (*) $0.15 > p > 0.05$; (**) $0.05 > p > 0.005$; (***) $p < 0.005$.

	Sala de conciertos	Sala de cámara	Iglesia
Mozart	66,13**	78,68**	88,43**
Fauré	97,07***	29,13	49,81*
Verdi	51,04*	12,96	49,81*
Pieza litúrgica	83,76**	64,73*	91,16***
Beethoven	19,14	89,38***	79,68**

Observando los coeficientes de determinación representados en la tabla 3 se puede comprobar que la relación de dependencia existente entre la claridad C_{80} y la percepción subjetiva no se mantiene constante para todos los estímulos e, incluso, varía considerablemente dentro del propio estímulo para distintas salas. Esta variación en la capacidad de determinar la claridad de un recinto en función del estímulo puede depender de las características musicales de la pieza, como por ejemplo el *tempo* o la figuración.

En el caso de la pieza de Mozart se puede observar que los valores del coeficiente de determinación son elevados para todas las salas, siendo el más alto el obtenido para la iglesia. Esta buena correlación se puede deber a que el fragmento está interpretado por una flauta solista con un *tempo* rápido y una figuración y articulación muy marcada, siendo así posible que se perciba fácilmente el efecto que la reverberación del recinto presenta sobre la pieza.

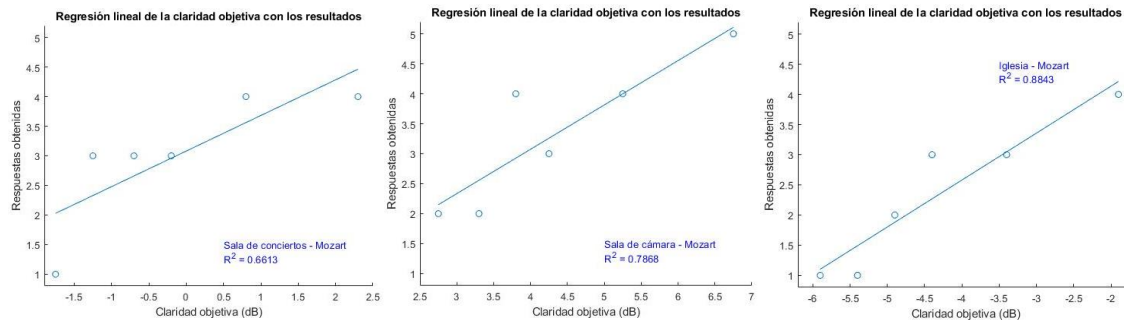


Figura 1 – Nubes de puntos y rectas de regresión para el estímulo de Mozart en la sala de conciertos (izquierda), sala de cámara (centro) e iglesia (derecha).

La pieza de Fauré se compone de un violonchelo solista interpretando una pieza de figuración larga y *tempo* lento, siendo así la pieza más influenciada por la reverberación y permitiendo también que sea sencillo diferenciar la claridad del recinto donde se interpreta. En el caso de la iglesia el nivel de claridad es tan bajo que puede dificultar la percepción de las variaciones entre niveles de claridad próximos.

Las características musicales que presenta el fragmento de Verdi son una melodía fácilmente entendible, de figuración relativamente larga y un acompañamiento a negras muy marcado, todo en un *tempo* lento. La facilidad de entendimiento de las líneas musicales de la pieza pudo derivar en una peor percepción de la claridad del recinto donde se interpreta la pieza, razón por la que los niveles de claridad seleccionados por los participantes fueron siempre los mismos (como se puede ver en la figura 2), provocando así un bajo coeficiente de determinación para todos los recintos evaluados.

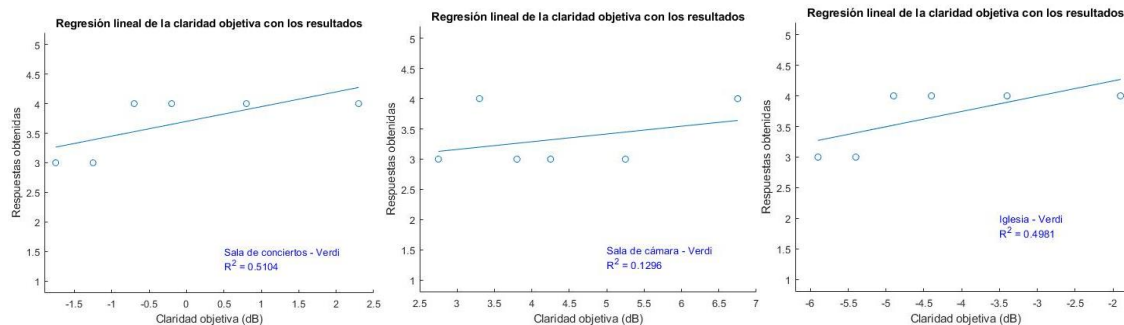


Figura 2 – Nubes de puntos y rectas de regresión para el estímulo de Verdi en la sala de conciertos (izquierda), sala de cámara (centro) e iglesia (derecha).

La pieza litúrgica se trata de un canto hispánico donde, en primer lugar, un cantante solista canta una línea melódica y, a continuación, se une un coro que contesta a dicha frase. En todos los casos el coeficiente de determinación es alto, indicando que se trata de una pieza que permite percibir la reverberación existente en las salas.

Por último, se va a evaluar el caso del estímulo con el fragmento musical de Beethoven. En este caso se puede observar un coeficiente de determinación alto tanto para la sala de cámara como para la iglesia; mientras que en la sala de conciertos el coeficiente presenta un nivel muy bajo (19,14 %). Este fragmento musical es considerablemente más complejo, con un gran número de instrumentos tocando simultáneamente con figuras musicales de corta duración y en *tempo* rápido. El valor tan bajo del coeficiente de correlación para la sala de conciertos se puede deber a una mala diferenciación de la claridad en un entorno en el que la reverberación de la sala no se percibe como excesivamente alta ni

excesivamente baja, si no que se mantiene en unos márgenes adecuados para la interpretación de este estilo musical.

4 Conclusiones

Tras la realización del *listening test* al grupo de participantes seleccionado y de la evaluación de los resultados mediante un análisis de correlación, se ha visto que la claridad percibida por dichos participantes para un mismo recinto depende significativamente del estímulo musical utilizado. Las características de la pieza musical que resultan de mayor interés y pueden presentar un mayor efecto sobre la percepción son el número de instrumentos y su tesitura, el *tempo* y el estilo de la pieza.

El hecho de que el estímulo sea un factor significativo en la percepción de la claridad implica que el estímulo con el que se determina la claridad debería ser tenido en cuenta a la hora de calcular su valor objetivo. Observando los p-valores expuestos en la tabla 3 se puede comprobar que las piezas de Verdi y Fauré no son favorables para la realización de *listening test*, ya que sus p-valores son siempre superiores a 0,05; mientras que los valores obtenidos para la pieza de Mozart muestran que es favorable para la determinación de características acústicas de recintos.

Finalmente, se puede concluir que el parámetro objetivo C_{80} , a pesar de tener posibilidades de mejora mediante la determinación de ponderaciones más adecuadas, que tengan en cuenta todo el rango frecuencial o mediante la inclusión del estímulo en su cálculo, presenta una relación aceptable con la claridad percibida y que, como ya dijo Bradley en [6], se trata del mejor índice actual para determinar la claridad de un recinto.

Referencias

- [1] Lacatis, R. et al. Historical and chronological evolution of the concert hall acoustics parameters. *Acoustics '08*, París (Francia), 29 de Junio al 4 de Julio.
- [2] Sabine, W. *Collected Papers on Acoustics*, Harvard University Press, Cambridge (Estados Unidos), 1922.
- [3] Beranek, Leo L. *Concert Halls and Opera Houses: Music, Acoustics and Architecture*, Springer, 1962.
- [4] Abdel Alim, O. "Dependence of time and register definition of room acoustical parameters with music performances", Disertación, Universidad Técnica de Dresde (Alemania), 1973.
- [5] AENOR. UNE-EN ISO 3382-1:2010: Acústica. Medición de parámetros acústicos en recintos. Parte 1: Salas de espectáculos, AENOR, Madrid (España), Junio 2010.
- [6] Bradley, J. S. Review of objective room acoustics measures and future needs. *Applied Acoustics*, Vol. 72 (10), 2011, pp. 713-720.
- [7] De la Prida, D. et al. Does the method matter? A review of the main testing methods for the subjective evaluation of room acoustics through listening tests. *Proceedings of the 23rd International Congress on Acoustics, ICA 2019*, Aachen (Alemania), 9-13 de Septiembre.
- [8] Bradley, J. S; Reich, R.; Norcross, S. G. A just noticeable difference in C_{50} for speech. *Applied Acoustics*, Vol. 58 (2), 1999, pp. 99-108.

- [9] Martellotta, F. The just noticeable difference of center time and clarity index in large reverberant spaces. *The Journal of the Acoustical Society of America* Vol. 128 (2), 2010, pp. 654–63.
- [10] Cox, T.; Davies, W.; Lam, Y. The sensitivity of listeners to early sound field changes in auditoriums. *Acta Acustica united with Acustica*, Vol. 79 (1), 1993, pp. 27-41.
- [11] Vigeant, M.C.; Celmer, R.D. Effect of Experimental Design on the Results of Clarity-Index Just-Noticeable-Difference Listening Tests. *Proceedings of 20th International Congress on Acoustics, ICA 2010*, Sydney (Australia), 23-27 de agosto, Vol. 425.
- [12] Soulodre, G. A; Bradley, J. S. Subjective evaluation of new room acoustic measures. *The Journal of the Acoustical Society of America*, Vol. 98 (1), 1995, pp. 294-301.
- [13] Prins, N. *Psychophysics: a practical introduction*, Academic Press, 2016.
- [14] Likert, R. A technique for the measurement of attitude. *Archives of Psychology*, Vol. 22, 1932, pp. 5 – 55.
- [15] Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological assessment*, Vol. 31 (4), 2019, pág. 557.
- [16] Chan, J. S. Response-order effects in Likert-type scales. *Educational and Psychological Measurements*, Vol. 51, 1991, pp. 531 – 540.
- [17] Weng, L. J.; Cheng, C. P. Effects of Response Order on Likert-type scales. *Educational and Psychological Measurements*, Vol. 60, 2000, pp. 908 – 924.
- [18] Subedi, B. P. Using Likert type data in social science research: Confusion, issues and challenges. *International journal of contemporary applied sciences*, Vol. 3 (2), 2016, pp. 36-49.
- [19] LINDAU, A. Spatial Audio Quality Inventory. Test Manual. V1.2. Marzo 2015 [Consulta: 3 de mayo de 2018]. Disponible en: https://www.researchgate.net/publication/275032444_Spatial_Audio_Quality_Inventory_SA_QI_Test_Manual_v12.
- [20] AENOR. UNE-ISO 4121: Análisis sensorial. Directrices para la utilización de escalas de respuestas cuantitativas. AENOR, Madrid (España), Octubre 2006.
- [21] Van Hout, D. Measuring meaningful differences: Sensory testing based decision making in an industrial context; applications of Signal detection theory and Thurstonian modelling. 2014.
- [22] Howell, D. C. *Statistical methods for psychology*. Cengage Learning, Wadsworth, Ohio (Estados Unidos), 2007.

A.1.5 Paper I

Title: La investigación colaborativa, clave en la obtención de magnitudes globales del aislamiento acústico representativas de la percepción

Authors: D. de la Prida, A. Pedrero, M. A. Navacerrada

Conference: XI Congreso Ibérico de Acústica, Acústica 2020

Conference location: Faro, Portugal (Online)

Conference date: 21-10-2020 to 23-10-2020

Proceedings: Proceedings of Acústica 2020, pp. 153–162

Publisher: Sociedade Portuguesa de Acústica (Lisbon, Portugal)

LA INVESTIGACIÓN COLABORATIVA, CLAVE EN LA OBTENCIÓN DE MAGNITUDES GLOBALES DEL AISLAMIENTO ACÚSTICO REPRESENTATIVAS DE LA PERCEPCIÓN.

Daniel de la Prida¹, Antonio Pedrero², María Ángeles Navacerrada³

¹ Universidad Politécnica de Madrid, Madrid, España
{e-mail: d.delaprida@alumnos.upm.es}

² Grupo de Investigación en Acústica Arquitectónica, Universidad Politécnica de Madrid, Madrid, España
{e-mail: antonio.pedrero@upm.es}

³ Grupo de Investigación en Acústica Arquitectónica, Universidad Politécnica de Madrid, Madrid, España
{e-mail: mdelosangeles.navacerrada@upm.es}

Resumen

En las últimas décadas se han realizado grandes esfuerzos para encontrar magnitudes globales del aislamiento acústico (SNQs) que muestren una alta correlación con la sensación subjetiva. Pese a estos esfuerzos, no existe consenso aún hoy en día. Para dar respuesta a este hecho se han realizado numerosos *listening tests*, en los que se ha investigado si el rango de frecuencias y los espectros de referencia utilizados en el cálculo de los SNQs representan la percepción subjetiva. Dos cuestiones principales conducen a que estas investigaciones no hayan permitido obtener aún un SNQ armonizado: (i) el número relativamente bajo de participantes que cada estudio es capaz de conseguir individualmente y (ii) la gran variedad de metodologías utilizadas para el diseño, desempeño y análisis de los juicios subjetivos.

Esta comunicación pone de manifiesto que la colaboración entre investigadores, así como el uso de una metodología armonizada y precisa para el diseño y análisis de *listening tests*, son fundamentales a fin de generar resultados concluyentes y significativos.

Palabras clave: SNQ, percepción, aislamiento acústico, listening test

Abstract

In recent decades, great efforts have been made to find single number quantities (SNQs) of the sound insulation that show a high correlation with the subjective perception. Despite these efforts, there is currently no consensus. To shed light on this subject, many listening tests have been carried out to determine whether the frequency range and reference spectra used in the calculation of the SNQs represent subjective perception. Two main issues lead to the fact that these investigations have not yet resulted in a harmonized SNQ: (i) the relatively low number of participants that each study can achieve individually and (ii) the wide variety of methodologies used for the design, development and analysis of subjective judgements.

This communication shows that using collaborative research and a harmonized and precise methodology for the design and analysis of listening tests are essential to generate conclusive and significant results.

Keywords: SNQ, perception, sound insulation, listening test

PACS nº. 43.55.Rg

1 Introducción

El aislamiento acústico es una cualidad de los elementos constructivos, cuyo conocimiento es fundamental en el diseño de nuevas construcciones, así como en las rehabilitaciones de las edificaciones existentes. Por ese motivo, el aislamiento acústico que proporcionan los elementos constructivos es una característica que debe ser conocida e interpretada no sólo por especialistas en acústica, sino por diversos actores que intervienen durante todo el proceso constructivo tales como arquitectos, aparejadores, ingenieros civiles o consultores, que no necesariamente tiene por qué ser expertos en la materia. Sin embargo, el aislamiento acústico es una característica compleja que varía en función de la frecuencia. Este hecho puede complicar su interpretación y aplicación. Con el fin de facilitar su interpretación, se han desarrollado, a lo largo de los años, diversos indicadores que se utilizan para describir el comportamiento acústico de un elemento constructivo de manera global. Estos indicadores, denominados magnitudes globales o SNQs (del inglés, *Single-Number Quantities*), expresan, a través de un único valor, el rendimiento acústico de los elementos constructivos.

De manera general, los métodos de cálculo utilizados para la obtención de estas magnitudes globales no son terreno ignoto, dado que se desarrollaron hace varias décadas y se encuentran estandarizados [1,2]. Sin embargo, aún a día de hoy, no existe una magnitud global que sea utilizada de manera armonizada, tal y como ponen de manifiesto diversas publicaciones [3–6]. Aunque en términos generales, los métodos de cálculo son muy parecidos, el fundamento en que se basan y el rango de frecuencias que se considera difiere ligeramente entre países. Estas divergencias son notables en el caso de los métodos de cálculo en Europa y Estados Unidos, donde se siguen normativas diferentes [1,2], aunque similares en su fundamento, pero también existen entre los propios países de la Unión Europea.

Asimismo, esta variabilidad de magnitudes se materializa en el establecimiento de diferentes requisitos de aislamiento, en términos de estas diferentes magnitudes globales, para los diferentes países [3,4,6–8].

Esta situación evidencia una problemática más profunda relativa al propio fundamento de las magnitudes globales: existen aún hoy en día cuestiones no resueltas con respecto a la adecuación de cada una de estas magnitudes globales para valorar el aislamiento acústico. Dicho de otro modo, es aún materia de estudio cómo han de ser estas magnitudes globales para ser fielmente representativas de la percepción subjetiva de los residentes y usuarios de los edificios.

Algunos esfuerzos por mejorar la representatividad de estas magnitudes del aislamiento acústico han llevado a la inclusión de coeficientes de adaptación espectral, tales como los índices C y C_{tr} , cuyo cálculo se describe en [1], con el fin de tener en cuenta la influencia del espectro predominante del ruido sobre el aislamiento acústico. Sin embargo, la inclusión de estos coeficientes ha llevado, en algunas ocasiones, a dificultar aún más los esfuerzos de armonización, puesto que algunos países han decidido incluirlos en sus requerimientos legales mientras otros no los contemplan.

Así, la problemática principal de la investigación en este ámbito sigue centrada en un hecho principal: la representatividad subjetiva de las magnitudes globales del aislamiento acústico no es tan alta como sería deseable. Así lo ponen de manifiesto diversas investigaciones con respecto a este hecho, que coinciden en que la representatividad de las magnitudes globales del aislamiento, con respecto a la percepción de molestia percibida, aún dispone de margen de mejora [9–22]. Por ello, la representatividad de los SNQs existentes, así como la mejora de estos indicadores, es una materia de estudio viva, tal y como se puede observar a partir de las publicaciones que se llevan a cabo cada año tratando de evaluar la percepción del aislamiento acústico.

Estas investigaciones se basan, de manera predominante, en el desempeño de pruebas perceptuales (*listening tests*). En el ámbito del aislamiento acústico a ruido aéreo, en el que se centra la presente

comunicación, estas pruebas suelen consistir en presentar a los participantes con varias muestras sonoras filtradas por el índice de reducción sonora de diversos elementos de aislamiento y consultarles acerca de un determinado atributo (generalmente sonoridad y molestia) que les evocan estos sonidos.

Si bien estas investigaciones permiten extraer conclusiones ciertamente interesantes con respecto a la percepción subjetiva del aislamiento acústico, existen ciertos inconvenientes que, en ocasiones, tienden a restar representatividad a los mismos y dificultar un avance eficiente en la armonización de las magnitudes globales.

2 Experimentos subjetivos: Situación actual

2.1 Los retos de la actual investigación

Como ya se ha descrito, numerosas son las publicaciones recientes que abordan la representatividad de los SNQs existentes, tratando, a su vez, de proponer mejoras que permitan aumentar esta representatividad [9–22].

Si bien todas estas investigaciones tienen una gran importancia, dado que mediante sus conclusiones contribuyen a incrementar el conocimiento en este ámbito, existe un inconveniente generalizado que tiende a restar potencial a los esfuerzos investigadores: la heterogeneidad de metodologías en la realización de los experimentos. Tal y como se desarrolla en detalle en [23], cuyo análisis se complementa en esta comunicación añadiendo publicaciones más recientes, las últimas investigaciones en este ámbito tienden a utilizar metodologías muy dispares para alcanzar un mismo fin, como es evaluar y mejorar la representatividad de las magnitudes globales actuales.

La Tabla 1 presenta las características y procedimientos seguidos por las investigaciones más recientes relacionadas con los *listening tests* en el ámbito del aislamiento acústico a ruido aéreo. En esta tabla se exponen, de manera resumida, el ámbito de cada estudio, el método utilizado para la consulta a los participantes, las muestras sonoras, los espectros de aislamiento y el rango frecuencial seleccionados, así como los métodos de análisis empleados. También se incluyen diversas cuestiones relacionadas con el desempeño de la prueba y que pueden tener una relevancia significativa sobre el sesgo, como son el número de participantes, las condiciones espaciales y tecnológicas de la prueba y la interacción entre los diversos entes que se interrelacionan durante los *listening tests*.

Tal y como se puede apreciar en la Tabla 1, los métodos de consulta, una de las componentes más importantes en el desarrollo de estas pruebas, tiene un carácter heterogéneo: mientras algunas investigaciones utilizan métodos de escala, otras utilizan métodos de discriminación basados en atributos; mientras algunas permiten una escucha libre de los estímulos al participante, otras permiten una única escucha; mientras algunas fuerzan al usuario a seleccionar una respuesta, otras permiten al usuario exponer que no fueron capaces de percibir diferencias. El hecho de utilizar métodos de consulta con poderes operacionales muy dispares puede acarrear consigo la deducción de conclusiones erróneas o parcialmente incorrectas si diversos estudios se comparan entre sí.

Lo mismo ocurren con los métodos de análisis, en parte motivado, pero no sólo, por la diferencia en los métodos de consulta: mientras las investigaciones basadas en escalas tienden a evaluar los juicios de escala en promedio a partir de los datos combinados de todos los sujetos [11,16,17,19], los métodos basados en discriminación tienden a evaluar los resultados mediante el conteo de preferencias

[13,20,21], aunque en algunas ocasiones se utilizan, bien de manera predominante o accesorio, métricas más complejas [12,20].

Con respecto a los atributos sonoros acerca de los que consultar a los participantes, tampoco existe un claro consenso. Si bien hay investigaciones que encuentran una alta correlación entre los juicios aportados por los participantes para los atributos “sonoridad” y “molestia”, otros encuentran que esta correlación no es lo suficientemente elevada como para que uno de ellos pueda establecerse como criterio de consulta único.

Lo mismo ocurre con respecto al entorno y las cuestiones tecnológicas del ensayo: mientras que algunos investigadores exponen que los entornos de ensayo deben ser lo más parecidos a estancias reales, introduciendo mobiliario en el entorno de ensayo y reproduciendo los estímulos por altavoces no visibles, otros opinan que es fundamental el control de cualquier otro tipo de estímulo que pueda sesgar las respuestas perceptuales puramente acústicas.

Otra de las cuestiones que permanecen sin resolver es la relativa al espectro. Si bien, las actuales corrientes paradigmáticas tienden a reforzar la inclusión del rango entre 50 y 100 [Hz] en las magnitudes globales utilizadas para caracterizar el aislamiento acústico a ruido aéreo, la mayor parte de las investigaciones destaca que estos índices no obtienen, en términos generales, una mejor descripción de la percepción subjetiva de los participantes. Sin embargo, las conclusiones individuales de cada experimento no son lo suficientemente fuertes por sí solas como para sentar un precedente en este sentido.

Este último hecho se debe, en parte, a las dos características que restan por comentar: el número de participantes y la casuística de fundamento de cada estudio, plasmada en las pruebas a través de las muestras sonoras y elementos de aislamiento evaluados.

Dado que cada experimento está supeditado a unas condiciones de estudio particulares, que circunscriben las muestras sonoras y elementos de aislamiento de cada estudio, las conclusiones de cada experimento por sí solas tienen una significancia limitada.

Adicionalmente, se ha de tener en cuenta que todas estas investigaciones se encuentran con un gran reto relativo a los participantes: contar con una muestra amplia de participantes es una tarea ardua y que puede poner en riesgo la viabilidad temporal de las investigaciones. Tal y como se puede observar en la Tabla 1, el número de participantes que cada uno de los experimentos fue capaz de abordar de manera individual es relativamente pequeño, salvo en [12,19]. Este hecho puede complicar la capacidad de extrapolación de las conclusiones a otros grupos poblacionales. Además, teniendo en cuenta que estas investigaciones se desarrollan generalmente en el ámbito académico, es común que los participantes sean principalmente estudiantes, lo que puede dificultar aún más la extrapolación de los resultados a otros grupos de edad.

Tabla 1 - Principales características comunes de las investigaciones en el ámbito de los *listening tests* para el aislamiento acústico a ruido aéreo.

Artículo	[16]	[17]	[19]	[20]	[21]	[11]	[12]	[13]
Tipo aislamiento	Paredes	Paredes ligeras y pesadas	Paredes ligeras y pesadas	Paredes ligeras y pesadas	Pared ligera y pared pesada	Elementos de fachada	Elementos de ventana	Paredes ligeras y pesadas
Atributo	Sonoridad y molestia	Molestia	Sonoridad, molestia y aceptabilidad	Sonoridad	Sonoridad	Sonoridad y molestia	Molestia	Sonoridad
Frecuencias [Hz]	50 – 5000	100 – 3150 (se usa R _w)	50 – 5000	50 – 5000	50 – 5000	50 – 5000	100 – 5000	50 – 5000
Nº participantes	30	22	59	33	39	43	119	40
Método de consulta	Escala unipolar 7 puntos	Escala unipolar continúa [0-10]	Escala unipolar 10 puntos	2-AC	2-AFC	Escala unipolar 10 puntos	2-AC	3-AFC ¹
Muestras sonoras y duración	S1 – S3: Frases (Harvard) S4: I Remain Call S5. Stand Up Tall S6. She Bop	S1. Música S2. Charla S3. Fiesta S4. Descarga cisterna 20s.	S1. Guitarra S2. Together again S3. Feuer frei S4. Llanto bebé S5. Charla S6. Ladrado 15 - 20s.	S1. Ruido rosa – 5s.	64 muestras sonoras de convivencia vecinal – 5s.	Ruido tráfico ² : S1. 50km/h S2. 80km/h S3. 100km/h S4. 60km/h S5. 60km/h 20s.	S1. Tráfico ² S2. Tráfico ² S3. Sirena S4. Aeronave S5. Ruido rosa 5-15s.	S1. Ruido Rosa - 2s. S2. Highway to Hell - 5s.
Método de análisis principal	Medias	Medias	Medias	Conteo de preferencias	Conteo de preferencias	Medias y desviaciones	Modelos thurstonianos	Conteo de preferencias
Sistema de reproducción	Altavoces	Auriculares	Altavoces (Genelec 8020A) + graves	Auriculares (HM-1 & ATH-AD900x)	Auriculares (HM-1)	Altavoces (Genelec 8020A) + graves	Auriculares (Sennheiser HD650)	Auriculares (Sennheiser HD650)
Aprendizaje	Sí	No	Sí	Sí	No	Sí	Sí	Sí
Presentación de estímulos	Escucha guiada	Escucha libre	Escucha guiada	Escucha guiada	Escucha libre	Repetición automática	Escucha guiada	Escucha libre
Audiometría	Sí	No	Sí	No	No	Sí	Sí	No

¹ Lo que los autores definen como 3-AFC es procedimentalmente equivalente a 2-AC² Las diferentes muestras sonoras de tráfico no sólo contienen tráfico a diferentes velocidades sino también con diferente comportamiento (ligero, pesado y mixto)

3 Ventajas de una investigación colaborativa

La investigación colaborativa, considerada como una investigación en la que diversos grupos de investigación interdisciplinares compartan métodos, resultados y experiencia, puede ser una herramienta muy útil a la hora de reducir los retos a los que se enfrentan actualmente las investigaciones relativas a *listening tests* en el ámbito del aislamiento acústico.

En primer lugar, el desarrollo conjunto de una metodología consensuada para el desempeño de estos experimentos resolvería los problemas derivados de la difícil comparación de resultados que, hoy en día, afecta a estas publicaciones.

Si se sigue una misma metodología, aunque cada estudio utilice unos elementos de aislamiento y muestras sonoras particulares, los resultados de cada prueba podrán utilizarse como parte de un conjunto de datos mayor, donde se aglomeren los resultados de diversos *listening tests*. Esto es, la utilización de una metodología común facilitaría la generación de grandes conjuntos de datos con características heterogéneas, más representativos de la percepción general del aislamiento acústico. También facilitaría la replicación de estudios en muestras heterogéneas de participantes, como pueden ser las de diferentes países o grupos de edad, lo que permitiría evaluar la consistencia de los resultados en términos generales.

Esta investigación colaborativa no debe ser sólo relativa al uso de una metodología común, sino también al uso de grupos de trabajo interdisciplinares tanto para el desarrollo de la metodología como para la realización de los experimentos. Es complicado, y puede llegar a ser ineficiente, que un único investigador trate de ser experto en todas las competencias necesarias para el desempeño de este tipo de pruebas. En particular, en el ámbito de la percepción subjetiva del aislamiento acústico, los expertos en psicología y estadística pueden ser de gran relevancia a la hora de determinar los métodos de consulta y análisis de mayor poder operacional y los expertos en paisaje sonoro (tanto interior como exterior) pueden ser determinantes a la hora de seleccionar muestras sonoras que representen fielmente los diferentes entornos acústicos. Por último, las empresas del sector de la construcción y los materiales acústicos pueden ser una pieza clave para la medida y obtención de abundantes datos relativos a los índices de reducción sonora de los elementos constructivos.

Para mostrar el gran potencial que podría tener la investigación colaborativa, la Figura 1 muestra dos recientes propuestas de mejora de los espectros de referencia descritos en la ISO 717-1:2013 [1], obtenidas mediante un proceso de optimización a partir de los resultados de dos *listening tests* diferentes.

La primera de las investigaciones [22] propone una optimización del espectro de referencia para ruidos de vecindad. Esta optimización está basada en los resultados del *listening test* [19] y consigue mejorar la representatividad de los SNQs con respecto al atributo subjetivo “molestia” de manera significativa. Así, el nuevo espectro de referencia propuesto, L_{opt} (mostrado en la Figura 1 como $L_{opt-living}$), mejora la correlación con la molestia para cinco de las seis muestras sonoras utilizadas durante el experimento, alcanzado valores superiores al 90% de correlación en tres de los seis casos.

La segunda de las investigaciones [14], basada en los resultados del *listening test* presentado en [12], propone una optimización del espectro de referencia para ruidos urbanos. Para ello plantea una ligera variación del algoritmo de optimización propuesto en [22] con el fin de adaptar el método de optimización a la distinta naturaleza de los resultados, dado que ambos experimentos [12,19] siguieron metodologías ciertamente dispares, como puede observarse en la Tabla 1. El nuevo espectro de referencia para ruidos urbanos $L_{opt-urban}$ consigue mejorar significativamente la representatividad del

atributo subjetivo “molestia” para las cinco muestras sonoras utilizadas en el estudio, alcanzando correlaciones superiores al 70% para cuatro de las seis muestras sonoras y superiores al 85% en dos de las seis muestras sonoras.

Es llamativo cómo ambas investigaciones han llevado a obtener curvas tan similares utilizando métodos de consulta y de análisis diferentes, para poblaciones de diferentes países y con muestras sonoras y elementos de aislamiento diferentes. Sin embargo, el hecho de haber utilizado enfoques tan diferentes no permite la generación de un conjunto de datos común que permita obtener una curva que sea realmente representativa de los resultados de ambos experimentos.

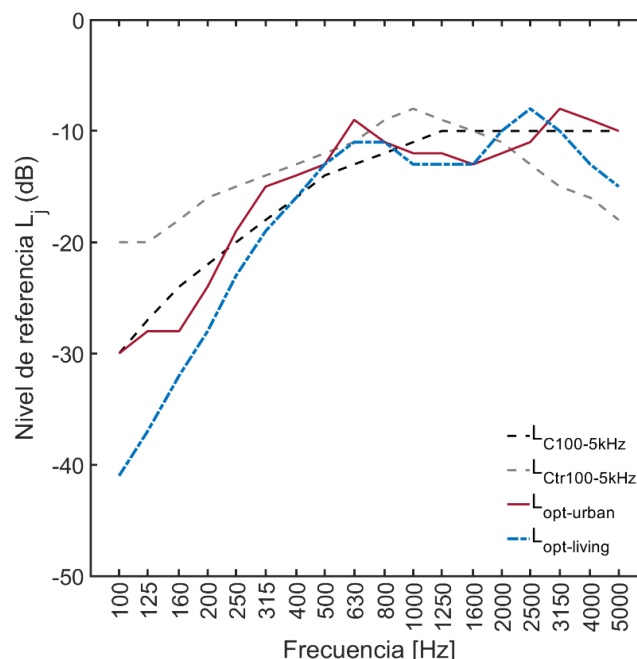


Figura 1 - Espectros de referencia de la ISO 717-1:2013 para los términos de adaptación espectral C (línea negra discontinua), C_{tr} (línea gris discontinua), así como las propuestas de [14] para ruidos urbanos (línea continua marrón) y de [22] para ruidos de vecindad (línea punteada y discontinua azul).

4 ¿Por dónde empezar?

Una gran ventaja de la investigación colaborativa en el ámbito de la armonización de magnitudes globales del aislamiento acústico es que no es necesario, hoy en día, comenzar de cero. Acciones investigadoras con COST TU0901 llevaron ya a cabo la ardua tarea de sentar las bases en este hecho. Sin embargo, no se puede considerar que el trabajo de generación de un marco común quedase totalmente resuelto. Se considera necesario, a la vista de las divergencias entre metodologías aún existentes, continuar con este tipo de acciones y proyectos colaborativos con el fin de mejorar el conocimiento en este ámbito.

Más allá de la metodología, se considera importante la generación colaborativa de una herramienta software destinada al desempeño y análisis de pruebas subjetivas. Éste es, precisamente, uno de los primeros impedimentos que se encuentran los investigadores a la hora de llevar a cabo *listening tests*.

Este desarrollo conjunto, por un lado, eliminaría posibles fuentes de sesgo derivadas de la interpretación individual de la metodología común por parte de cada investigador y, por otro, permitiría un ahorro de coste temporal y humano. Actualmente, los grupos de investigación acaban, como normal general, diseñando su propia herramienta, tal y como ponen de manifiesto las diversas publicaciones en las que se presentan herramientas diseñadas ad-hoc para la consecución de estas pruebas [13,24,25].

5 Conclusiones

Con el fin de mejorar significativamente la investigación y facilitar la obtención de conclusiones relevantes en el ámbito del aislamiento acústico, que puedan llevar a una armonización clara de las magnitudes globales, parece necesario potenciar diversas acciones:

1. El desarrollo conjunto de una metodología clara, precisa, eficiente y con un alto poder operacional para el desempeño de los *listening tests* que sea utilizada de manera común.
2. La colaboración interdisciplinar de expertos en diversas materias tales como: aislamiento acústico, pruebas subjetivas, paisaje sonoro y estadística, entre otras.
3. El apoyo de las empresas del sector de la construcción con el fin de obtener datos constructivos abundantes y representativos.
4. Contribuir con los resultados independientes de cada investigación a la conformación de un conglomerado de datos colaborativo, analizable de manera conjunta, que permita incrementar la significancia global del conocimiento en esta materia.

Agradecimientos

Los autores quieren agradecer a la acción investigadora COST TU0901 su empeño en sentar las bases de una investigación más colaborativa, en la que diversas entidades puedan aportar los resultados de sus investigaciones para mejorar, de manera conjunta, el conocimiento en este ámbito.

Asimismo, los autores quieren agradecer al Ministerio de Ciencia, Innovación y Universidades la financiación del proyecto [RTI2018-094656-B-I00], dentro del que se enmarca esta comunicación.

Referencias

- [1] ISO 717-1:2013 Acoustics — Rating of sound insulation in buildings and of building elements — Part 1: Airborne sound insulation. Geneva, Switzerland: International Organization for Standardization; 2013.
- [2] E413-10: Classification for Rating Sound Insulation. West Conshohocken, PA, USA: ASTM International; 2010.
- [3] Rasmussen B, Rindel JH. Sound insulation between dwellings—Descriptors applied in building regulations in Europe. *Appl Acoust* 2010;71:171–80.
- [4] Rasmussen B, Machimbarrena M, Fausti P, TU0901 C. Building acoustics throughout Europe. Volume 1: Towards a common framework in building acoustics throughout Europe. 2014.
- [5] Monteiro C, Machimbarrena M, Tarrero AI, Smith RS. Translation between existing and proposed harmonized airborne sound insulation descriptors: A statistical approach based on in-situ measurements. *Appl Acoust* 2017;116:94–106. <https://doi.org/10.1016/j.apacoust.2016.09.017>.

- [6] Alonso A, Patricio J, Suárez R, Escandón R. Acoustical retrofit of existing residential buildings: Requirements and recommendations for sound insulation between dwellings in Europe and other countries worldwide. *Build Environ* 2020;174:106771. <https://doi.org/https://doi.org/10.1016/j.buildenv.2020.106771>.
- [7] Rasmussen B, Machimbarrena M. Developing an international acoustic classification scheme for dwellings-From chaos & challenges to compromises & consensus? INTER-NOISE NOISE-CON Congr. Conf. Proc., vol. 259, Institute of Noise Control Engineering; 2019, p. 7751–62.
- [8] Rasmussen B. Encouraging acoustic renovation of housing in Denmark by extending acoustic classification with two lower classes E and F for old housing. 48th Int. Congr. Exhib. Noise Control Eng., INTERNATIONAL INSTITUTE OF NOISE CONTROL ENGINEERING; 2020.
- [9] Tachibana H, Hamada Y, Sato F. Loudness evaluation of sounds transmitted through walls—Basic experiment with artificial sounds. *J Sound Vib* 1988;127:499–506.
- [10] Vian J, Danner WF, Bauer JW. Assessment of significant acoustical parameters for rating sound insulation of party walls. *J Acoust Soc Am* 1983;73:1236–43.
- [11] Hongisto V, Oliva D, Rekola L. Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. *J Acoust Soc Am* 2018;144:1100–12.
- [12] de la Prida D, Pedrero A, Navacerrada MÁ, Díaz-Chyla A. Methodology for the subjective evaluation of airborne sound insulation through 2-AC and Thurstonian models. *Appl Acoust* 2020;157:107011.
- [13] Chmelík V, Rychtáriková M, Müllner H, Jambrošić K, Zelem L, Benklewski J, et al. Methodology for development of airborne sound insulation descriptor valid for light-weight and masonry walls. *Appl Acoust* 2020;160:107144. <https://doi.org/https://doi.org/10.1016/j.apacoust.2019.107144>.
- [14] de la Prida D, Pedrero A, Navacerrada MÁ, Díaz-Chyla A. An annoyance-related SNQ for the assessment of airborne sound insulation for urban-type sounds. *Appl Acoust* 2020;168:107432. <https://doi.org/10.1016/j.apacoust.2020.107432>.
- [15] Park HK, Bradley JS, Gover BN. Evaluating airborne sound insulation in terms of speech intelligibility. *J Acoust Soc Am* 2008;123:1458–71.
- [16] Park HK, Bradley JS. Evaluating standard airborne sound insulation measures in terms of annoyance, loudness, and audibility ratings. *J Acoust Soc Am* 2009;126:208–19. <https://doi.org/10.1121/1.3147499>.
- [17] Pedersen TH, Antunes S, Rasmussen B. Online listening tests on sound insulation of walls – A feasibility study. *Euronoise 2012*, European Acoustics Association; 2012.
- [18] Rychtáriková M, Roozen B, Müllner H, Stani M, Chmelík V, Glorieux C. Listening test experiments for comparisons of sound transmitted through light weight and heavy weight walls. *Akustika* 2013;19:1–10.
- [19] Hongisto V, Oliva D, Keränen J. Subjective and objective rating of airborne sound insulation—living sounds. *Acta Acust United with Acust* 2014;100:848–63.
- [20] Monteiro C, Machimbarrena M, de la Prida D, Rychtarikova M. Subjective and objective acoustic performance ranking of heavy and light weight walls. *Appl Acoust* 2016;110:268–79. <https://doi.org/10.1016/j.apacoust.2016.03.008>.
- [21] Rychtáriková M, Muellner H, Chmelík V, Roozen NB, Urbán D, Garcia DP, et al. Perceived Loudness of Neighbour Sounds Heard Through Heavy and Light-Weight Walls with Equal R

- w+ C 50–5000. *Acta Acust United with Acust* 2016;102:58–66.
- [22] Virjonen P, Hongisto V, Oliva D. Optimized single-number quantity for rating the airborne sound insulation of constructions: Living sounds. *J Acoust Soc Am* 2016;140:4428–36.
- [23] Vardaxis N-G, Bard D. Review of acoustic comfort evaluation in dwellings: Part III—airborne sound data associated with subjective responses in laboratory tests. *Build Acoust* 2018;25:289–305.
- [24] Rychtáriková M, Horvat M. Developing a Methodology for Performing Listening Tests Related to Building Acoustics. *Build. Acoust. throughout Eur. Vol. 1 Towar. a common Framew. Build. Acoust. throughout Eur.*, 2014.
- [25] de la Prida Caballero D, Rodrigues A, Monteiro C, Machimbarrena M. Los listening test como método de caracterización de la molestia e implementación de software de evaluación. *Tec.* 2014, Murcia: Sociedad Española de Acústica; 2014.

A.2 Chronological list of publications

This section presents an enumeration of all the author's contributions during the PhD period. Publications are divided in journal papers, conference papers and book chapters.

A.2.1 Journal papers

- [1] D. de la Prida, A. Pedrero, M. A. Navacerrada, C. Díaz, Relationship between the geometric profile of the city and the subjective perception of urban soundscapes, *Applied Acoustics* 149 (2019) 74-84. doi:<https://doi.org/10.1016/j.apacoust.2019.01.025> (JCR 2018 Q2)
- [2] D. de la Prida, A. Pedrero, M. A. Navacerrada, A. Díaz-Chyla, Methodology for the subjective evaluation of airborne sound insulation through 2-AC and Thurstonian Models, *Applied Acoustics* 157 (2020) 107011. doi:<https://doi.org/10.1016/j.apacoust.2019.107011> (JCR 2019 Q2)
- [3] A. Pedrero, M. A. Navacerrada, D. de la Prida, L. Iglesias, A. Díaz-Chyla, On the accuracy of the sound absorption measurement with an impedance gun, *Applied Acoustics* 158 (2020) 107039. doi:<https://doi.org/10.1016/j.apacoust.2019.107039> (JCR 2019 Q2)
- [4] D. de la Prida, A. Pedrero, M. A. Navacerrada, A. Díaz-Chyla, An annoyance-related SNQ for the assessment of airborne sound insulation for urban-type sounds, *Applied Acoustics* 168 (2020) 107432. doi:<https://doi.org/10.1016/j.apacoust.2020.107432> (JCR 2019 Q2)
- [5] T. S. Gómez, S. Zuluaga, M. Jimenez, M. A. Navacerrada, M. M. Barbero-Barrera, D. de la Prida, A. Restrepo-Osorio, P. Fernández-Morales, Evaluation of Colombian crops fibrous byproducts for the potential applications in sustainable building acoustics, *Polymers* 13 (2021) 101. doi:<https://doi.org/10.3390/polym13010101> (JCR 2019 Q1)
- [6] M. A. Navacerrada, D. de la Prida, A. Sesmero, A. Pedrero, T. S. Gómez, P. Fernández-Morales, Comportamiento acústico y térmico de materiales basados en fibras naturales para la eficiencia energética en edificación, *Informes de la Construcción* 73 (561) (2021) e373. doi:<https://doi.org/10.3989/ic.73876> (JCR 2019 Q4)
- [7] D. de la Prida, A. Pedrero, L. A. Azpicueta-Ruiz, M. A. Navacerrada, Listening tests in room acoustics: Comparison of overall difference protocols regarding operational power, *Applied Acoustics* (2021) 108186. doi:<https://doi.org/10.1016/j.apacoust.2021.108186> (JCR 2019 Q2)

A.2.2 Conference papers

- [1] P. Oliveira, D. de la Prida, A. Pedrero, C. Díaz, Caracterización subjetiva de paisajes sonoros de Madrid, in: A. Calvo-Manzano, A. Pérez-López (Eds.), *Proceedings of Tecniacústica 2017*,

- Sociedad Española de Acústica, Madrid, Spain, 2017, pp. 471–478.
- [2] D. de la Prida, A. Pedrero, M. A. Navacerrada, C. Díaz, Metodología para la caracterización subjetiva del aislamiento acústico de ventanas para entornos urbanos, in: A. Calvo-Manzano, A. Pérez-López (Eds.), *Proceedings of the XI Congreso Iberoamericano de Acústica, FIA 2018*, Sociedad Española de Acústica, Madrid, Spain, 2018, pp. 655–663.
 - [3] A. Díaz-Chyla, D. de la Prida, A. Pedrero, M. A. Navacerrada, El aumento del aislamiento a ruido aéreo de la fachada de las aulas y su efecto en la inteligibilidad del mensaje oral, in: A. Calvo-Manzano, A. Pérez-López (Eds.), *Proceedings of the XI Congreso Iberoamericano de Acústica, FIA 2018*, Sociedad Española de Acústica, Madrid, Spain, 2018, pp. 688–694.
 - [4] M. Gjura, D. de la Prida, A. Pedrero, C. Díaz, Evaluación a largo plazo de la variabilidad de los parámetros psicoacústicos en diferentes calles de Madrid, in: A. Calvo-Manzano, A. Pérez-López (Eds.), *Proceedings of the XI Congreso Iberoamericano de Acústica, FIA 2018*, Sociedad Española de Acústica, Madrid, Spain, 2018, pp. pp. 479–487.
 - [5] P. Ruiz, M. Machimbarrena, D. de la Prida, A. I. Tarrero, M. A. Martín, Estudio de la correlación entre características viarias y diversos parámetros acústicos subjetivos en la ciudad de Valladolid, in: A. Calvo-Manzano, A. Pérez-López (Eds.), *Proceedings of the XI Congreso Iberoamericano de Acústica, FIA 2018*, Sociedad Española de Acústica, Madrid, Spain, 2018, pp. 463–470.
 - [6] J. Gonzalo, D. de la Prida, A. Pedrero, C. Díaz, Auralización en entornos urbanos: caso de estudio Plaza de Jacinto Benavente, in: A. Calvo-Manzano, A. Pérez-López (Eds.), *Proceedings of the XI Congreso Iberoamericano de Acústica, FIA 2018*, Sociedad Española de Acústica, Madrid, Spain, 2018, pp. 471–478.
 - [7] D. de la Prida, A. Pedrero, M. A. Navacerrada, A. Díaz-Chyla, The effect of the rolling shutter in the perception of the sound insulation, in: *Proceedings of the 48th International Congress and Exhibition on Noise Control Engineering, Internoise 2019*, Institute of Noise Control Engineering, Reston, Virginia, 2019, pp. 6013–6022.
 - [8] A. Pedrero, D. de la Prida, L. Marandet, J. L. Sánchez, M. A. Navacerrada, Perceptual validation of virtual acoustic models, in: *Proceedings of the 48th International Congress and Exhibition on Noise Control Engineering, Internoise 2019*, Institute of Noise Control Engineering, Reston, Virginia, 2019, pp. 6923–6931.
 - [9] D. de la Prida, A. Pedrero, L. A. Azpicueta-Ruiz, M. A. Navacerrada, C. Díaz, Does the method matter? A review of the main testing methods for the subjective evaluation of room acoustics through listening tests, in: M. Ochmann, M. Vorländer, J. Fels (Eds.), *Proceedings*

of the 23rd International Congress on Acoustics, Deutsche Gesellschaft für Akustik, Berlin, Germany, 2019, pp. 7887–7894.

- [10] M. Larrosa, A. Pedrero, D. de la Prida, Influencia del estímulo musical sobre la percepción de la claridad C80, in: Proceedings of Acústica 2020, Sociedade Portuguesa de Acústica, Lisbon, Portugal, 2020, pp. 395–405.
- [11] D. de la Prida, A. Pedrero, M. A. Navacerrada, La investigación colaborativa, clave en la obtención de magnitudes globales del aislamiento acústico representativas de la percepción, in: Proceedings of Acústica 2020, Sociedade Portuguesa de Acústica, Lisbon, Portugal, 2020, pp. 153–162.
- [12] A. Díaz-Chyla, D. de la Prida, M. A. Navacerrada, A. Pedrero, C. Díaz, El efecto de las transmisiones por flancos en la mejora del aislamiento acústico entre recintos mediante capas adicionales a un elemento estructural soporte, in: Proceedings of Acústica 2020, Sociedade Portuguesa de Acústica, Lisbon, Portugal, 2020, pp. 163–173.

A.2.3 Book chapters

- [1] A. Pedrero, L. Marandet, L. Iglesias, D. de la Prida, La reconstrucción virtual del sonido primitivo del Monasterio de Piedra: Análisis y modelización del estado actual, in: H. González, D. Prieto (Eds.), Monasterio de Piedra, un legado de 800 años. Historia, arte, naturaleza y jardín, Institución Fernando el Católico, Zaragoza, Spain, 2019, pp. 316–333