Bibliographical reference

Acoustic Characterisation by Using Different Room Acoustics Software Tools: A Comparative Study, L. Álvarez-Morales, M. Galindo, S. Girón, T. Zamarreño and R. M. Cibrián, Acta Acustica **vol. 102** (Number 3), 2016, pp. 578-591

DOI

https://doi.org/10.3813/AAA.918975

Acta Acustica united with Acustica
Vol. 102 (2016) 578 – 591

DOI 10.3813/AAA.918975

Acoustic Characterisation by Using Different Room Acoustics Software Tools: A Comparative Study

L. Álvarez-Morales[1)], M. Galindo[1)], S. Girón[1)], T. Zamarreño[1)], R. M. Cibrián[2)]
1) Escuela Técnica Superior de Arquitectura, Instituto Universitario de Arquitectura y Ciencias
de la Construcción (IUACC), Universidad de Sevilla, Av. Reina Mercedes 2, 41012-Seville, Spain.
calvarez6@us.es
2) Universidad de Valencia. Av. Blasco Ibáñez 15, 46010-Valencia, Spain

Summary

A room impulse response (RIR) describes the acoustic behaviour of a certain source-receiver combination within a room. In order to characterise an enclosure, several emission and reception positions are used. Current computers, equipped with suitable hardware and software, enable these RIRs to be registered, processed and analysed in an efficient way. This paper presents a study of the differences in the results obtained from the use of four commercial software tools widely used in room acoustics. Comparisons are drawn up from the RIRs measured with each tool under the same conditions. To this end, the main room acoustic parameters defined in the ISO 3382-1 are calculated. For each parameter, the differences between both the spatially averaged value in octave bands and the spectral average at each receiver point are studied, since these values are normally used to characterise the acoustics of a room. Furthermore, a more detailed study is carried out to evaluate these differences point-by-point in statistical terms, by analysing the dependence on several factors, such as the position of the source, the relative positions of the receiver points, and the frequency band. In order to assess the importance of the differences found, the Just Noticeable Difference (JND) is taken as a reference.

PACS no. 43.55.Mc, 43.58.Gn

1. Introduction

Measurement and analysis of room impulse responses (RIRs) forms the basis of the computational techniques of room acoustics for the determination of reverberation time (Schroeder's backward integration method), acoustical parameters and functions (echograms, energy-time graphs), as well as other acoustic characteristics (echoes, coupled volumes etc.). An impulse response in a real room is not free from the influence of background noise. Consequently, decay curves (backward-integrated energy decay curve (EDC)) and energy-time curves (ETC) depend both on the upper limit of integration and noise. Their corresponding peak-to-noise ratio (PNR) should be sufficient so that the measured impulse response is valid, since an insufficient PNR can significantly affect the calculation of the parameters, particularly the accuracy of the determination of reverberation time [1, 2]. Various methods have been proposed in the literature for noise compensation [3, 4, 5, 6], all of which are well-established techniques commonly implemented in commercial software in order to reduce the effect of background noise. There are also alternative ways to deal with noise that use non-linear regression methods [7], but these are currently not widely adopted. In the experimental chain, there are other sources of error which can converge on the results, producing lack of objectivity and reproducibility in measurements. Thus, Pelorson *et al.* [8] discuss, among other factors, how the use of different transducers and their placement in relation to the measurement positions can dramatically affect the spectral behaviour of the evaluated parameters. Along these lines, De Vries *et al.* [9] carry out a more in-depth study into the influence of this variation on the parameters related to the apparent source width. Likewise, the influence of characteristics of source directionality on impulse response, and therefore on the dispersion of the results of the acoustic parameters, has been studied by San Martín *et al.* [10], who quantify these differences in terms of JNDs. Furthermore, the algorithms of each separate system for the calculation of acoustic parameters introduce differences caused by the definition of time window, band pass filtering, implementation of backward integration and background noise compensation [3]. Several pieces of work have been published in this regard; for example, two international round robins were organized in order to perform a comparison of post-processing algorithms of several measurement systems by using a synthe-

---

Received 11 January 2016,
accepted 08 April 2016.

sized impulse response [11] and a real impulse response [12].

Other studies have compared the values of the acoustic parameters obtained with a variety of combinations of hardware and software tools [13] as well as using different excitation signals (interrupted stationary noise, impulsive signals, swept sine) [14] and measurement conditions (occupancy levels, source-receiver positions). Even experimental and simulated results have been compared by analysing whether the obtained differences between the values of the acoustic parameters computed from simulated RIRs (corresponding to two simulation software platforms) and experimental results (with two different sets of commercial equipment and various excitation signals) are within a reasonable range of acceptability [15].

The aim of this paper is to assess the agreement in room acoustic parameter values when using different software tools to characterise the acoustic behaviour of a room. For this purpose, a blind comparison between the room acoustic parameter values obtained by separately using four commercial software tools (ARTA, Dirac, EASERA and WinMLS), widely used in room acoustics today, has been carried out, whereby the RIRs measured under the same conditions in the same room are each taken as the starting point. An analysis of the concordance of the results of the most relevant room acoustic parameters (monaurals and binaurals) is performed, both in general and in statistical terms. These parameters are calculated from the RIRs recorded and processed separately with each software platform. A large set of reception points is characterised, whereby two positions of the source are used. Likewise, by taking advantage of the possibility of exchanging formats between the software tools, a transversal analysis is performed such that the same signal, in standardized format, is processed with each of the four software tools to assess their possible differences in processing and hence to verify the interchangeability between them. The difficulties encountered in the format exchange of signals are also discussed.

2. Methodology

In order to assess the uncertainties in room acoustic parameter values associated to the software used to measure and post-process the RIRs, a methodology based on three different types of analyses is set up: repeatability of measurements taken separately with each software; comparison of room acoustic characterisation by using each software tool independently, including both measurement and post-processing; and comparison of room acoustic characterisation by using each software tool to analyse the same set of RIRs, including only post-processing. The study has been structured into three types of analyses in order, on the one hand, to evaluate how the characterisation is affected when measurements of the RIRs are taken by using different software tools, and, on the other hand, to analyse the differences caused by the processing algorithms implemented in each of these tools and to value the possibility of using these tools interchangeably, which may lead to the possibility of using a post-processing software different from that used to measure the RIRs. Four acoustic software tools, ARTA, Dirac, EASERA, and WinMLS (randomly named A, B, C, and D to maintain anonymity in the presentation of the results), are considered. To ensure that no other source of uncertainties affects the results, exactly the same measurement chain is used to take measurements within the same room with each software tool, ensuring that, at each source-receiver combination, both source and receiver remain fixed in location and orientation.

Figure 1. Auditorium of the School of Architecture of Seville. View of the podium and the audience area.

2.1. Experimental procedure

Experimental measurements are carried out following the procedure described in ISO 3382-1 [16], so that a room is characterised by using the four software tools considered. This characterisation is necessary since the comparison is based on collating the values of acoustic parameters defined throughout the informative annexes of this standard.

The auditorium of the School of Architecture of Seville has been chosen to conducting the acoustic measurements. It is a multifunctional space which is suitable for conferences, academic and cultural activities. It is a prismatic room whose dimensions are approximately 18×24×8 m^3, with a ground surface of around 432 m^2. The slightly sloping audience area hosts 364 moderately upholstered seats, and is fragmented by two perpendicular aisles that define the four existing sectors of the audience. The stage, with 198 m^2 surface, is elevated 80 cm from the lower level of the floor. In the rear part of the stage there is a screen that can be hidden by a draped curtain, of heavy, thick, and velvety cloth. The side and back walls are finished with dark agglomerate cork panels of about 3 cm thick. The ceiling is plastered and features a metal structure, which is divided into parallel frames with lattice (see Figure 1).

The enclosure was unoccupied during the measurement session. The sound source was placed at two positions on the podium (S1/S2 in Figure 2), on the longitudinal plane of symmetry of the room and moved towards one side, 1.5 m above ground level. Fifteen receiver points were selected and located within the audience area (Figure 2), at 1.2 m above the floor. Note that in order to prevent pos-

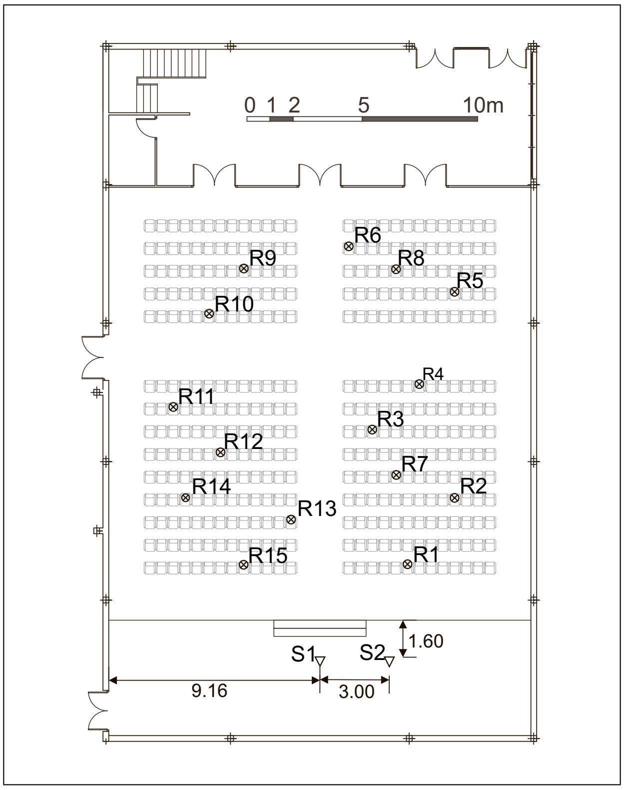

Figure 2. Ground plan of the enclosure (dimensions in m), in which the positions of the source (S1 and S2) and receptors (R1 to R15) are shown.

sible errors caused by repositioning the microphone both in its location and orientation, it remained fixed at each position while measurements with the four software tools were taken and only the connections were changed. Neither the position nor the orientation of the sound source was changed.

The environmental conditions were monitored during the measurement period. Maximum variation of temperature during a measurement session was 0.7 °C. Moreover, the RIRs were measured with the four software tools consecutively at each reception point, and therefore the variation of the environmental conditions between measurements was negligible. The background-noise level, recorded using the Svan958 analyser from Svantek, was 29.2 dBA, and the frequency spectrum of the background noise remained below the NC-25 rating curve in all octave bands.

Sine sweep signals were used as excitation signals, whose range was adjusted to cover the octave bands between 125 and 4000 Hz. Sweep time and other characteristics of the excitation signal were adjusted as recommended by the developers of each software tool, in order to achieve an PNR higher than 45 dB for all octave bands of interest. It should be noted that, to be able to obtain an appropriate PNR, the analysed systems require an appropriate choice of the running time of measurement as well as the treatment algorithm of the background noise, which can cause significant differences between measurement systems if it is not properly managed, even when the requirement of exceeding 45 dB has been accomplished for all octave bands [4].

The constituent hardware of the measurement chain, common for the four software tools, is composed of an EDIROL UA-101 USB audio device from Roland, which is controlled by each software platform. The excitation signal is emitted through the AVM DO-12 dodecahedral omnidirectional source, which has been previously amplified by the B&K-2734 power amplifier. The RIRs are captured with an Audio-Technica AT4050/CM5 multi-pattern microphone, which enables the omnidirectional or figure-of-eight pattern to be set manually, and is connected to the polarization source SoundField SMP200. In order to register binaural responses, the Head III torso simulator from Head Acoustic and the signal conditioner unit B&K–2829 were used.

Due to the difficulties of calibrating each software properly using the same procedure in order to obtain the ***G*** parameter, a reference RIR has been measured under the same conditions with systems A, B, and C (*G* parameter values calculated with D are not considered here since a manual procedure) is needed, and results can hence be affected by the particular user who analyses the data), with the aim of comparing the relative *G*. This experimental procedure involves placing the omnidirectional microphone on the axis of symmetry of the room, 10 m from the sound source located at position S1, at a height such that the straight line between them is parallel to the ground. A thick rockwool blanket ($1.30\times1.20\times0.05\,\mathrm{cm}^3$) is also placed 5 m from the source on the ground to attenuate the first reflection from the floor. The signals obtained with each measurement system have been edited to remove the reverberant field. Although it has been demonstrated that an in-situ calibration method can introduce a small calculating error in the value of *G* [17, 18, 19], the procedure described guaranteed that those discrepancies found between the values of the sound strength parameter given by the different software tools, are not attributable to the reference signal since the same measurement conditions (microphone, source and blanket positions remain fixed) are used in all cases. This work is focused on the comparison of the results rather than on the characterisation of the room, and therefore, this error can be assumed.

2.2. Processing of RIRs

In the first place, the RIRs are processed independently with each system with the aim of calculating the main acoustic parameters and comparing the results obtained (repeatability and individual comparison). This means that the same software used to register the RIRs is used to analyse them in its own format.

Secondly, a crossed analysis is set out so that a common set of RIRs is processed with all the software tools, in an attempt to separate the adquisition of RIRs from their processing variances. To this end, the RIRs previously registered are converted to common standard formats and reprocessed with all of the software tools. The choice of the format is crucial to obtain coherent results and to minimize the error. The four systems compared are able to export/import signals at least in two standard audio formats: the commonly used WAVE form audio-file format (*.wav*); and the MLSSA time-domain data files (filenames with *.tim* extension).

The problem with the *.wav* format is the scale factor, which is individually modified when exporting signals in order to offer the widest dynamic range, thus losing the ratio between correlated signals (for instance, the omnidirectional and figure-of-eight measurements at the same receiver point). Not all software platforms offer the option of exporting *.wav* signals without automatically applying the scale factor, and for this reason the calculation of certain acoustic parameters is incorrect; this option therefore remains unviable for the calculation of the lateral fractions and the Interaural Cross Correlation parameters. In view of the above, the *.tim* format, which is broadly used in room acoustics, has been chosen since it includes the information about "scaling factor" in the heading of the file, thereby maintaining the correct ratio between signals. Due to the large volume of data involved and to prevent redundancy, software A has been chosen as the arbitrary reference so that the set of RIRs measured with this software are exported in *.tim* format and analysed/processed with the other systems to study their differences.

2.3. Analysis procedure

The comparison of the results obtained with the many combinations of measurement software and analysis software is carried out following the same procedure both for the individual and crossed analyses. This comparison consists of assessing the differences of the room acoustic parameters values (T_{30}, EDT, T_S, C_{50}, C_{80}, D_{50}, G, L_J, J_LF, $\mathrm{IACC_E}$, $\mathrm{IACC_L}$, and STI) both in terms of spatial average and point-by-point. Furthermore, the spectrally averaged values recommended in the ISO 3382-1 and by other authors (Table I) are taken into account since they are normally used as acoustic valuation criteria. Additionally, a statistical study is carried out in order to evaluate these differences in statistical terms, by analysing their dependence on several factors, such as the position of the source, the location of the receiver point, the frequency band, and the software used.

All these results are analysed in absolute terms as well as in terms of the threshold of perception of each parameter (*Just Noticeable Difference*, JND, see Table I), in order to relate the absolute differences with the audible discrimination threshold, commonly used as reference in room acoustic evaluations [11, 12].

For the statistical analysis, the SPSS Statistics v19 software is used [20]. To study the general concordance between the acoustic parameters calculated with the various software tools, the Intraclass Correlation Coefficient (ICC) and the linear regression analysis with assessment of the slope, intercept and Pearson's r coefficient, are used. The ICC is extended in the context of assessing the repeatability of multiple measurements. The ICC range runs from 0 to 1, whereby the extremes indicate absence of agreement and absolute reliability of the results, respectively [21]. In order to statistically compare the parameter values following the methodology described above, repeated measures ANOVA are used taking into account the following variation factors: the measurement software, the acoustic parameter, the frequency band or the source-receiver distance. Either Scheffe's or Games-Howell tests are used for multiple comparisons, depending on the indication of the degree of equality of variances given with Levene's test. When comparing only two means the t-Student test is used. All the statistical analyses are performed with a significance level of $\alpha = 0.05$.

3. Results and discussion

3.1. Measurement repeatability

Before starting with the individual study, in which the results obtained with the software tools are separately compared, it is crucial to quantify the measurement error made when each platform is used independently.

Repeatability is understood as the concordance between the results of successive measurements taken with the same system under the same measurement conditions (repeatability conditions) which include: the same procedure in the same place by the same operator using the same equipment within short intervals of time.

In order to assess the repeatability of the results obtained with each measurement system under identical conditions, eight consecutive impulse responses are registered with the four systems at a certain receiver point of the room with the source placed at position S1, from which only monaural parameters are deduced at all octave frequency bands from 125 Hz to 4000 Hz, in order to prevent additional influences.

To analyse the variability of the data, the mean value of each parameter is calculated in each octave frequency band together with its standard deviation (SD). The standard deviations for all frequency bands are of the same order of magnitude for each parameter and software tool, therefore their mean value can be considered the standard deviation in the measure of each parameter as a single number. Table II shows these average values of the standard deviation (considering all frequency bands) calculated for each parameter using each software, also in terms of JND (see Table I). Expressing these differences in terms of JNDs allows their evaluation in terms of perception, and also makes the results intercomparable even between parameters with different units. It can be seen that the differences are negligible, for example T_{30} variations remain well below 0.01 s and for C_{80} below 0.07 dB. In general, the variability in all parameters mostly remains below 0.05 JNDs, and in no case does it exceed 0.2 JNDs.

Nevertheless, even though there is no error in perceptual terms, since a measurement method in extreme repeatability conditions considering a single point is used, then the relative error made when taking a measurement

Table I. Frequency bands used for spectrally averaged values and just noticeable differences (JND) of the objective room acoustic parameters. Avg.: Spectral averages according to Reference [16] (arithmetic average except for L_J which is energy averaged), except for IACC which is according to Okano *et al.* [24]. JNDs are referred to in Reference [15], Vorländer [25] and Bradley *et al.* [26]. STI parameter is measured through omnidirectional impulse responses rather than modulated noise. For a more robust procedure see Cabrera *et al.* [27].

	$T_{30\mathrm{m}}$ (s)	EDT$_\mathrm{m}$ (s)	$T_{S\mathrm{m}}$ (ms)	C_m (dB)	$D_{50\mathrm{m}}$	G_m (dB)	$L_{J\mathrm{m}}$ (dB)	J_{LFm}	IACC$_\mathrm{m}$	STI
Avg.	500-1k	500-1k	500-1k	500-1k	500-1k	500-1k	125-1k	125-1k	500-2k	—
JND	5%	5%	10	1	0.05	1	1	0.05	0.075	0.03

Table II. Standard deviations found at each parameter (mean value in all frequency bands) when analysing the RIR measured at one receiver point with each software in extreme repeatability conditions, both in absolute terms and in terms of JNDs.

		Software A	Software B	Software C	Software D
T_{30}	SD (s)	0.003	0.006	0.003	0.002
	SD (JND)	0.046	0.095	0.040	0.024
EDT	SD (s)	0.003	0.013	0.004	0.003
	SD (JND)	0.041	0.197	0.069	0.041
T_S	SD (ms)	0.144	0.561	0.262	0.133
	SD (JND)	0.014	0.056	0.026	0.013
C_{50}	SD (dB)	0.035	0.039	0.051	0.015
	SD (JND)	0.035	0.039	0.051	0.015
C_{80}	SD (dB)	0.031	0.064	0.031	0.014
	SD (JND)	0.031	0.064	0.031	0.014
D_{50}	SD (-)	0.001	0.001	0.003	0.001
	SD (JND)	0.029	0.025	0.063	0.015
G	SD (dB)	0.135	0.045	0.097	–
	SD (JND)	0.135	0.045	0.097	–
STI	SD (-)	0.000	0.000	0.000	0.000
	SD (JND)	0.000	0.012	0.000	0.011

needs to be evaluated. This relative error is determined as (SD/Mean)·100, expressed as a percentage. The statistically significant difference between the relative errors obtained is evaluated from the corresponding analysis of variance (ANOVA) with 3 variation factors taken into account: the software, the parameter, and the frequency band.

The four software tools present similar relative errors, below 3% ($p = 0.263$), but there are statistical differences for frequencies ($p \ll 0.001$) and parameters ($p \ll 0.001$). In general, greater dissimilitaries are found in the 2 kHz and 4 kHz bands, with a variability of 4.8% and 2.4% respectively, while for the other bands, this is less than 1.7%. These large values of relative error at high frequencies are caused by low mean values of clarity and strength parameters in these frequency bands. Specifically, a relative error of 5.3% is observed for C_{80} and of 6.4% for G, larger than for all the remaining parameters (T_{30}, EDT, T_S, and D_{50}) with relative errors lower than 0.5%. Moreover, C_{50} with a relative error of 2.4% has no statistical difference with the rest of the parameters. The relative error calculated for each parameter in each frequency band by using each software tool is shown in Figure 3. Not only may high values of relative error be caused by greater values of the standard deviation (SD), but also by measures with very small

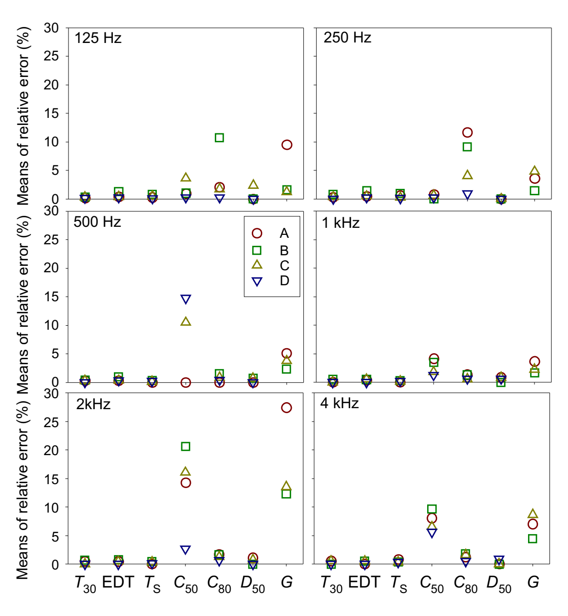

Figure 3. Means of relative error (%) estimated for each parameter with each software tool at all frequency bands for the repeatability study.

Table III. Mean values and standard deviation for those parameters whose relative error exceed the 5 percent threshold.

			125	250	500	1000	2000	4000
C_{50} (dB)	A	Mean	−3.89	−6.08	−0.20	1.13	−0.33	0.58
		SD	0.04	0.05	0.00	0.05	0.05	0.05
	B	Mean	−5.35	−6.60	0.01	1.33	−0.23	−0.54
		SD	0.05	0.00	0.04	0.05	0.05	0.05
	C	Mean	−3.74	−5.80	0.30	1.30	−0.30	−0.53
		SD	0.13	0.03	0.03	0.02	0.05	0.03
	D	Mean	−2.30	−6.37	0.11	1.34	−0.28	0.45
		SD	0.01	0.02	0.02	0.02	0.01	0.03
C_{80} (dB)	A	Mean	1.71	−0.65	3.10	3.43	2.78	3.88
		SD	0.04	0.08	0.00	0.05	0.05	0.05
	B	Mean	1.00	−0.98	3.08	3.58	2.83	2.84
		SD	0.11	0.09	0.05	0.05	0.05	0.05
	C	Mean	1.81	−0.71	3.30	3.62	2.86	2.82
		SD	0.03	0.03	0.03	0.02	0.04	0.05
	D	Mean	2.20	−0.96	3.11	3.62	2.80	3.76
		SD	0.01	0.01	0.01	0.02	0.02	0.02
G (dB)	A	Mean	−2.21	−3.49	−2.03	−2.71	0.43	1.48
		SD	0.21	0.12	0.10	0.10	0.12	0.10
	B	Mean	−2.21	−3.66	−2.26	−2.88	0.29	1.16
		SD	0.04	0.05	0.05	0.05	0.04	0.05
	C	Mean	2.12	−2.63	−2.83	−4.89	0.58	1.54
		SD	0.03	0.13	0.11	0.11	0.08	0.13
	D	Mean	–	–	–	–	–	–
		SD	–	–	–	–	–	–

average values (even if the standard deviation is small). Table III shows the mean values and their standard deviation for those parameters whose relative error exceed 5% (threshold value for considering that the measurement accuracy is adequate) in Figure 3. It can be seen that the relative errors above 5% obtained for C_{50} and G parameters in the 125 Hz frequency band are due to high values of SD. Nevertheless, in the rest of the cases in which the estimated error is certainly in excess of 5%, the average values of the parameters are very small, while the SD remains in the same order than in the other cases with negligible values of the relative error.

3.2. Individual analysis: Acoustic characterisation with different measurement software tools

After analysing the variability of the measurements for a case of extreme repeatability at a single receiver point, the correlation between the measurement systems considering several receiver points is analysed in this section, while taking into account other factors such as the position of the source, the source-receiver distance, and the influence of the frequency band.

3.2.1. Concordance between the four measurement systems. Statistical analysis

In a first general analysis, the aim is to determine whether all systems measure in the same way, by studying the intraclass correlation coefficient (ICC) of the acoustic parameters calculated with the four software tools from its own measured RIRs. The high values of ICC calculated for most of the parameters denote a *very good* strength of concordance (ICC > 0.95), except for G (ICC = 0.77), J_{LF} (ICC = 0.87) and IACC (ICC ≈ 0.89) for which the agreement is *good*, and for L_{J} which is within the *moderate* range (ICC = 0.62) [21].

Pearson's product-moment correlation coefficients (Pearson's r) between results obtained with software A (taken as the arbitrary reference) and the other platforms have been analysed. For all cases, the correlation is highly significant ($p \ll 0.001$). Most parameters have very good correlation since Pearson's r coefficient is higher than 0.91. Those with lower values represent intermediate situations with different degrees of discordance depending on the parameter–software pair analysed (Software C: G with $r = 0.747$, L_{J} with $r = 0.673$ and J_{LF} with $r = 0.820$; Software B: IACC with $r = 0.859 - 0.884$; Software D: IACC with $r = 0.853 - 0.912$).

Graphically, if the agreement is good between measurement systems, then regression lines for parameters calculated by different software tools at each receiver point for all frequencies involved must fit straight lines with slope m equal to 1 and with the y-intercept n at the origin. The linear relationship between the different acoustic software tools has been analysed through the regression lines calculated for each pair of systems, individually for each parameter. A high degree of concordance is found for T_{30}, EDT and T_S values in all cases (the coefficient of determination (r^2) is about 0.99 and a slope of regression line close to 1). Only minor discrepancies of EDT are observed for results obtained at 125 Hz with software D. In the regression analysis, involving clarity and definition parameters, results indicate that values calculated with software B slightly differ from the others (r^2 from 0.93 to 0.97). Regarding G and L_J, the agreement between software A and B ($r^2 = 0.98$ and $m = 0.94$), contrasts with the results of software C ($r^2 = 0.56$ for G and r^2=0.45 for L_J). Software C also differs from the others when analysing J_LF (r^2 from 0.65 to 0.68; and m from 0.75 to 0.87), but in this case the outliers that greatly affect r^2 and the slope of regression line correspond with receiver points located within the critical distance. Binaural parameters show a different degree of concordance depending on the pair of software tools compared; thus if software A is involved, then outliers (that correspond with values measured at R9) reduce the coefficient of determination, and if software D is involved, then the low value of the coefficient of determination (r^2 minimum of 0.73) is mainly caused by the IACC values in the 125 Hz frequency band.

In order to analyse the discrepancy between parameter values calculated with each software tool, a comparison of means has been carried out, using a paired-sample ANOVA with one variation factor, in this case the measurement system. The test indicates that, for G and L_J parameters, there are differences in variances statistically significant depending on the measurement platforms used. For all the other measured parameters, all systems behave equivalently. Through the analysis of those which statistically have the same behaviour, it can be considered that parameters with an ANOVA significance above 0.8 (T_{30}, EDT, T_S, J_LF, STI) have great equality in means, therefore all systems measure them equivalently, while other parameters, such as $\mathrm{IACC_E}$, with a significance of 0.203, have an equality which is not as complete as in the other cases. Parameters with intermediate values are not equal, but it cannot be said that significant differences exist.

Regarding parameters which show differences according to the ANOVA test, it is interesting to ascertain which systems are responsible for these differences. Thus, a study of multiple comparisons has been carried out by using Scheffe's test when variances were equal, a Games–Howell's test when variances were unequal, and a t-Student in the case of only two measurement systems. Results indicate that, in statistical terms, software A and software B estimate G equally but differently to software C, and L_J is calculated in a different way for A and C.

Although there are statistically significant differences in the values, these differences are not very large. Consequently, negligible differences in terms of JNDs are expected except in the case of G and L_J. In order to determine the specific cause of these differences, parameter values are statistically analysed depending on several factors. Firstly, work is carried out in order to determine whether source-receiver distance has a significant influence on the results. The analysis points out that at both mid and high frequencies, the differences in valuation of parameters do not depend on the position of the receiver, although they are slightly affected at low frequency. Analysing the influence of the point itself on the differences found at each parameter, only isolated cases appear in which this factor is determinant, although system B slightly underestimates C_{50}, C_{80} and D_{50} values with respect to the other systems.

Statistical analysis also reveals that the use of one or other source position (S1 and S2) is indistinguishable; therefore this factor has no influence on the differences. In the second part of this study, frequency bands are grouped according to the ISO: LF (125 and 250 Hz), MF (500 and 1 kHz), and HF (2 kHz and 4 kHz). The spectrally averaged values have also been included. Results show that considering ISO single values no disagreement was found for any parameter. Furthermore, in T_{30}, EDT, T_S, C_{80}, and J_LF, there is no significant difference in any group or at any frequency independently of the system used. In those parameters in which the test for multiple comparisons was significant, discrepancies mainly occur in the LF group (C_{50}, D_{50} and IACC). G has statistically significant differences in the LF and MF groups, while L_J differs in all groups with any of the two measurement systems which are able to calculate it.

Since parameter values are considered different when discrepancies exceed 1 JND, software A has been chosen as the arbitrary reference in order to assess the differences between the values obtained with the software tools in terms of JNDs, so that the significance of these differences can be determined. Three intervals have been set: differences below or equal to 0.5 JNDs are considered negligible, that is, both software tools measure equally; difference values between 0.5 and 1 JND represent disagreements of minor importance since they are not perceptible; and a difference in the determination of the parameter equal to or greater than 1 JND defines a noticeable discrepancy in the results. For this reason, the latter interval is considered the most critical.

Table IV shows the percentage of compared data (each source-receiver combination compared point-by-point in each frequency band) in which the discrepancy found between the value determined by using software A and the other tools remains within each interval. Note that the second interval (0.5–1.0) is not in the table since it is considered the least significant and it is implicit from the other intervals. It is observed that, for all the parameters, there is an association between the number of cases whose differences belong to each interval and the software used ($p < 0.05$). Furthermore, in those cases where there is an association, the percentage of each interval has been compared and, in each row, each letter denotes a subset of the software category whose proportions do not differ significantly at the 0.05 level. For example, in the case of T_{30}, the number of values whose differences are within the range $\leq$ 1 JND are of the same order (less than 2%) for software B and D, while C has 10.6% of data in this range, with greater statistically significant differences.

Discrepancies between the values of the parameters obtained with A and the other software tools remain below 1 JND in more than 91% of the cases when G and L_J are not taken into account. In the group of temporal parameters, it is EDT which presents greater differences, espe-

Table IV. Number of items of data (%) in which the parameter values measured with software tools B, C, and D differ by less than 0.5 JND (thereby considered equal), or differ by more than 1 JND (thereby considered different) when comparing with values measured with software A. Frequency bands from 125 Hz to 4 kHz and all receiver and source positions have been taken into account.

	Intervals (JND)	B	C	D	Chi-square Sig.
T_{30}	[0–0.5]	85.0	77.2	94.4	$p \ll 0.001$
	≤ 1	1.7	10.6	1.7	
EDT	[0–0.5]	72.8	67.2	54.4	$p<0.001$
	≤ 1	9.4	16.1	25.0	
T_S	[0–0.5]	83.9	93.3	98.9	$p \ll 0.001$
	≤ 1	2.2	1.1	0.0	
C_{50}	[0–0.5]	68.3	90.0	83.9	$p \ll 0.001$
	≤ 1	10.6	1.7	9.4	
C_{80}	[0–0.5]	80.0	92.2	87.8	$p=0.016$
	≤ 1	4.4	1.7	3.3	
D_{50}	[0–0.5]	63.3	78.9	71.1	$p=0.009$
	≤ 1	11.1	3.3	9.4	
G	[0–0.5]	95.0	21.7	–	$p \ll 0.001$
	≤ 1	3.3	55.6	–	
L_J	[0–0.5]	–	17.2	–	–
	≤ 1	–	67.2	–	
J_{LF}	[0–0.5]	71.7	85.0	73.9	$p=0.021$
	≤ 1	11.1	7.2	12.8	
$IACC_E$	[0–0.5]	77.8	80.0	72.2	$p<0.001$
	≤ 1	12.2	5.0	18.3	
$IACC_L$	[0–0.5]	75.6	92.2	83.3	$p<0.001$
	≤ 1	13.3	2.2	10.0	
STI	[0–0.5]	100.0	100.0	100.0	–
	≤ 1	0.0	0.0	0.0	

cially between D and A (25% of values differ more than 1 JND). Energy parameters show very good agreement, more significantly between C and A. Spatial parameters have perceptible differences in less than 15% of the cases with B and D, and only in 4.4% of cases when comparing results obtained with A and D. Regarding the first interval, results indicate that there are no differences in more than 80% of data.

3.2.2. General analysis of the discrepancies. Mean values

The behaviour of the measurement systems when characterising a room, in terms of discrepancies in parameter values and their relative errors has been examined above.

Despite varying with the distance, parameter values are usually spatially and spectrally averaged to generally describe the acoustic behaviour of a room, and therefore an analysis in this regard is performed. Table V shows the mean of the spatially averaged spectral values of each parameter (including 15 receiver points and 2 positions of the sound source) measured with each software tool, together with the maximum variance found between them. Note the close agreement between the four measurement systems, especially in the temporal acoustic parameters, which present insignificant differences (maximum variance for EDT of 0.138 s in the 125 Hz frequency band). Regarding energy parameters, differences appear within the range of 1 dB for C_{50} and C_{80} at low frequencies. The greatest variation between software tools is obtained for G, especially in the 125 Hz frequency band, a fact attributable to the treatment of the calibration signal (see Section 2.1). In J_{LF} parameter, there are slight differences (in the range of 0.008 to 0.023) in the mean values calculated with the four systems in each frequency band. IACC values show a maximum variance at low frequency of around 0.085.

Figure 4 shows the spatially averaged spectral value calculated with each software tool for those parameters with the highest differences. Despite the differences, it can be seen that C_{50}, C_{80}, and IACC parameters have a similar spectral tendency with all the software tools, although software B slightly underestimated clarity values in low frequency bands and also in the 4 kHz frequency band. However, as regards G, software C clearly differs from the others, especially in the 125 Hz frequency band.

Spectrally averaged values of the most relevant acoustic parameters calculated at each receiver point (in accordance with Table I, denoted by subscript m), have been compared for each pair of software tools used. The mean values of these differences, calculated point by point (in absolute values), are shown in Figure 5. Error bars correspond to standard deviations of these mean differences. According to this figure, the largest differences in tempo-

Table V. Maximum variance found between spatially averaged values calculated for each parameter considering the four software tools (15 receiver points and 2 positions of the sound source are included). Mean values obtained by averaging the spectral results calculated with the four software tools are given as a reference. Individual analysis. ^{a}: Software D is not included in the mean values of G.

		125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
T_{30} (s)	Max Variance	0.019	0.019	0.016	0.006	0.006	0.051
	Mean	2.144	1.630	1.178	1.010	1.083	0.977
EDT (s)	Max Variance	0.138	0.023	0.016	0.011	0.011	0.060
	Mean	1.889	1.418	1.155	0.978	1.101	1.008
T_S (ms)	Max Variance	4.795	0.516	0.930	0.580	0.582	6.085
	Mean	137.797	111.826	76.979	60.342	70.860	66.496
C_{50} (dB)	Max Variance	1.070	1.080	0.325	0.113	0.047	0.693
	Mean	-3.368	-2.817	0.044	1.667	0.605	0.747
C_{80} (dB)	Max Variance	0.768	0.424	0.243	0.101	0.043	0.613
	Mean	-0.179	0.567	2.496	4.317	3.285	3.615
D_{50}	Max Variance	0.052	0.052	0.023	0.017	0.002	0.037
	Mean	0.326	0.352	0.500	0.590	0.532	0.538
G (dB)^a	Max Variance	4.220	1.112	0.565	1.869	0.439	0.701
	Mean	-0.023	-1.887	-3.084	-4.295	-0.334	0.477
J_{LF}	Max Variance	0.020	0.017	0.008	0.016	0.009	0.023
	Mean	0.145	0.198	0.255	0.271	0.218	0.185
$\mathrm{IACC_E}$	Max Variance	0.082	0.086	0.068	0.055	0.047	0.052
	Mean	0.900	0.791	0.508	0.387	0.331	0.273
$\mathrm{IACC_L}$	Max Variance	0.090	0.081	0.052	0.043	0.047	0.049
	Mean	0.885	0.664	0.247	0.159	0.106	0.096

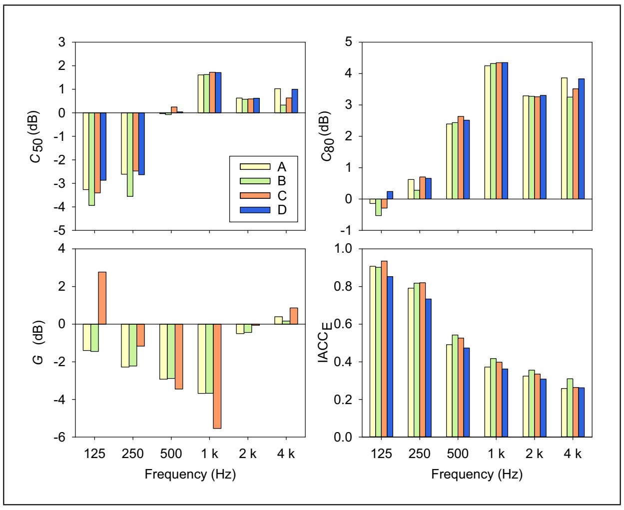

Figure 4. Spatially averaged spectral values calculated with each software tool for C_{50}, C_{80}, G, and $\mathrm{IACC_E}$.

ral parameters are found in $\mathrm{EDT_m}$ when results obtained using system D are compared with the other systems, although neither mean nor standard deviation differences exceed 1 JND. T_{Sm} point-by-point differences remain below 1 ms (0.1 JNDs). Regarding monaural energy parameters, G_{m} shows a disagreement of the order of 1 JND when compared with system C, while A and B exhibit high concordance. The discrepancies in the behaviour of this parameter can be attributed to the reference signal processing and its treatment in system C. Error bars show that directional parameters present differences of about 1 JND, or even higher for $\mathrm{IACC_{Em}}$ at certain receiver points, even though a 0.5 JND limit is not exceeded in terms of mean differences. Lastly, the STI calculation has a very low error. As a summary of Figure 6, it must be pointed out that the greatest differences, although not perceptually significant, are in the initial part of the impulse response, as evidenced by the fact that $\mathrm{EDT_m}$ and the initial energy parameters ($C_{50\mathrm{m}}$, $D_{50\mathrm{m}}$, J_{LFm}, and $\mathrm{IACC_{Em}}$) exhibit the most significant discrepancies.

3.3. Crossed analysis: Acoustic characterisation with different measurement software tools by processing the same set of RIRs

In the study of the crossed comparison, in which a certain set of RIRs is processed with all the systems, the same procedure as for the individual comparison has been followed. In particular, in this section, an analysis has been carried out on the acoustic parameter values obtained with B, C and D when processing the RIRs measured with A. Therefore, software A has been used as the arbitrary reference.

3.3.1. Concordance between the four processing systems. Statistical analysis of the discrepancies

Concordance has been studied between measurement systems from the Intraclass Correlation Coefficient (ICC) and the corresponding regression lines, always with reference to software A. This analysis evaluates the agreement of the parameter values calculated by different software tools

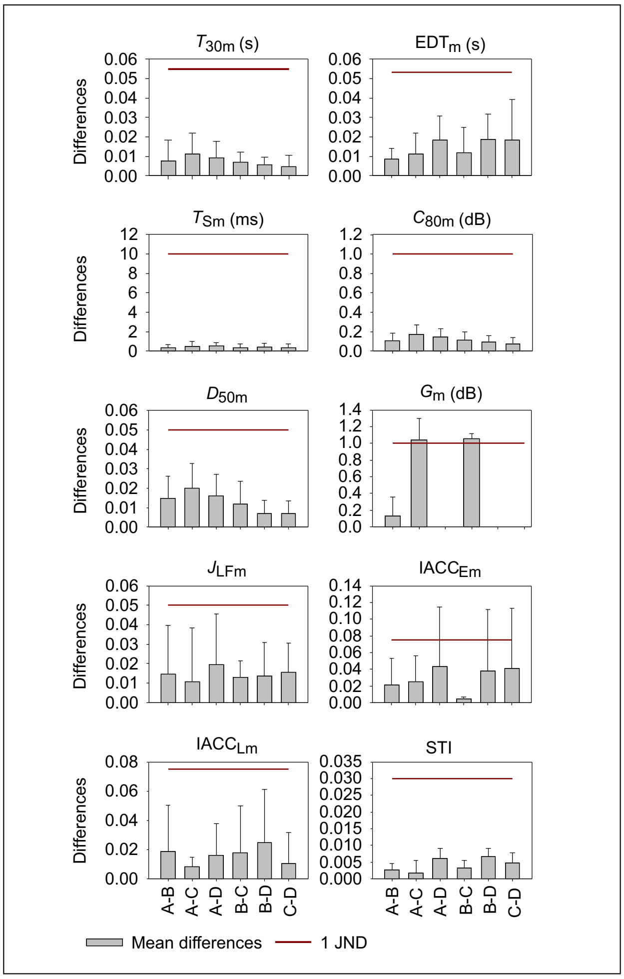

Figure 5. Mean differences found between spectrally averaged values of the acoustic parameters calculated with each software tool compared point-by-point (individual comparison). The error bars represent the standard deviation of these averages.

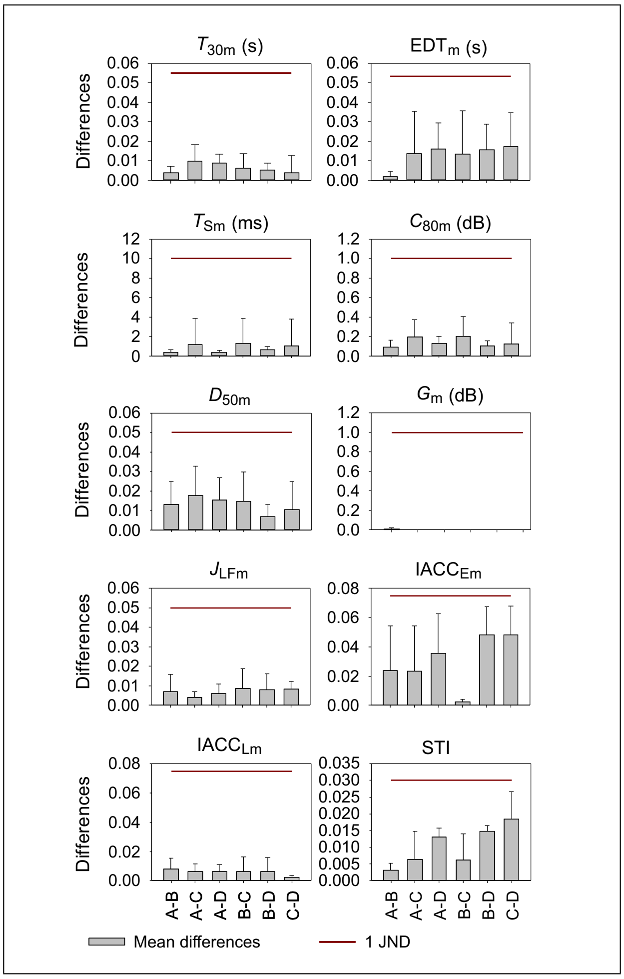

Figure 6. Mean differences found between spectrally averaged values of the acoustic parameters calculated with each software tool compared point-by-point (crossed comparison). The error bars represent the standard deviation of these averages.

at each receiver point. The intraclass correlation is even higher than that in the individual analysis, since all parameters denote a *very good* strength of concordance (ICC > 0.951). In order to ascertain the degree of correlation between software A and the others separately, Pearson's r coefficient for each parameter has been calculated. All parameters show a high degree of correlation (Pearson's r coefficient varies between 0.909 and 0.999).

Analogously to the individual study, regression lines have been generated to assess the correspondence between software A (taken as a reference) and the other systems. As in the case of individual analysis, it is observed that the group of T_{30}, EDT, and T_S shows a higher degree of similarity between the systems, since the regression lines are closer to the straight line with slope $m = 1$ and to the y-intercept n at the origin. G and J_{LF} also exhibit major agreements ($r^2 > 0.95$; $m > 0.97$; $n \sim 0$), while clarity and definition show more dispersion, especially in low frequency bands when software B or D are involved. Binaural parameters show the same discrepancies as in the individual analysis.

A stastistical analysis depending on several factors is carried out on the differences, considering all receiver and source positions. When all systems are globally considered, Pearson's chi-square test determines that there is a significant association between the differences in terms of JNDs and the frequency bands. Discrepancies in excess of 1 JND are found at LF in 11.64% of the compared data, while at MF and HF this percentage is considerably lower (less than 3%).

However, when parameter values are studied separately, the discrepancies are not always associated with the frequency band. In particular, T_{30}, T_S, G, and J_{LF} have a chi-square significance above 0.05, which means that there is no statistically significant difference associable to the grouped frequency band for these parameters. Table VI shows the percentage of cases calculated for each parameter, whose discrepancies with the values calculated with the reference software are greater than 1 JND depending on the grouped frequency band. The high degree of agreement at MF and HF both for temporal and energy parameters should be noted.

To complete this analysis, the influence of the location of the receiver with respect to the source position has been analysed as a separate factor, by distinguishing between receivers located within or beyond the critical radius [22].

Table VI. Number of items of data (%) with discrepancies greater than 1 JND depending on the grouped frequency band considered and calculated independently for each parameter and software.

	Software B			Software C			Software D		
	LF	MF	HF	LF	MF	HF	LF	MF	HF
T_{30}	1.67	3.33	1.67	5.00	3.33	3.33	0.00	0.00	1.67
EDT	6.67	0.00	0.00	11.67	13.33	5.00	58.33	6.67	3.33
T_S	0.00	0.00	0.00	10.00	3.33	3.33	0.00	0.00	0.00
C_{50}	33.33	1.67	0.00	11.67	6.67	8.33	25.00	0.00	0.00
C_{80}	16.67	0.00	0.00	8.33	1.67	3.33	10.00	0.00	0.00
D_{50}	31.67	1.67	1.67	11.67	5.00	6.67	25.00	0.00	1.67
G	0.00	0.00	0.00						
J_{LF}	5.00	1.67	5.00	0.00	0.00	0.00	3.33	0.00	0.00
$\mathrm{IACC}_{\mathrm{E}}$	28.33	3.33	10.00	3.33	3.33	10.00	5.00	5.00	21.67
$\mathrm{IACC}_{\mathrm{L}}$	25.00	3.33	0.00	3.33	1.67	0.00	3.33	1.67	0.00

Table VII. Number of items of data (%) in which the parameter values calculated with software tools B, C, and D differ by less than 0.5 JND (thereby considered equal) or differ by more than 1 JND (thereby considered different) when comparing with values measured with software A. Frequency bands from 125 Hz to 4 kHz and all receiver and source positions have been taken into account. Crossed comparison.

	Intervals (JND)	B	C	D
T_{30}	[0–0.5]	97.22	91.11	97.78
	≤ 1	2.22	3.89	0.56
EDT	[0–0.5]	91.67	76.67	60.56
	≤ 1	2.22	10.00	22.78
T_S	[0–0.5]	98.33	90.00	100.00
	≤ 1	0.00	5.56	0.00
C_{50}	[0–0.5]	78.33	90.00	86.11
	≤ 1	11.67	8.89	8.33
C_{80}	[0–0.5]	88.33	89.44	87.78
	≤ 1	5.56	4.44	3.33
D_{50}	[0–0.5]	68.89	77.78	73.89
	≤ 1	11.67	7.78	8.89
G	[0–0.5]	100.00	–	–
	≤ 1	0.00	–	–
J_{LF}	[0–0.5]	90.56	99.44	93.33
	≤ 1	3.89	0.00	1.11
$\mathrm{IACC}_{\mathrm{E}}$	[0–0.5]	76.67	84.44	70.00
	≤ 1	13.89	5.56	10.56
$\mathrm{IACC}_{\mathrm{L}}$	[0–0.5]	86.11	95.56	95.56
	≤ 1	9.44	1.67	1.67
STI	[0–0.5]	100.00	90.00	73.33
	≤ 1	0.00	0.00	0.00

There is no association between this factor and the results, and only when comparing results of tool C with those of tool A, is the number of items of data with discrepancies above the 1 JND threshold higher within the critical radius ($r_c = 6.2\,\mathrm{m}$). In general, when analysing the RIRs with B, C and D, 5.96%, 5.91%, and 6.24% of compared data differ more than 1 JND from those obtained with A, respectively. Furthermore, for most parameters, the results obtained by processing a signal with software A can be considered equal to those obtained by processing the same signal with the other software tools, since differences below 0.5 JND are found in more than 87% of the data compared. Table VII provides information about the agreement in acoustic parameter values when comparing the values calculated with software tools B, C and D, with those obtained with A from the same set of RIRs.

3.3.2. General analysis of the discrepancies. Crossed analysis

Table VIII shows the maximum variance found between the spatially averaged parameter values calculated by the four software tools through the analysis of the set of RIRs measured with software A, together with the mean values obtained by averaging the spectral results calculated with each software tool, considering all source-receiver combinations. In this case, results obtained with the different software tools are in even better agreement than in the individual analysis, especially at high frequency bands; only certain parameters (C_{50}, C_{80}, D_{50}, and IACC) show differences at low frequency, which remain around 1 dB for clarity parameters, 0.05 for definition, and 0.1 for IACC parameters, due to an underestimation by software B. It must be borne in mind that G is only calculated by software tools A and B, which are in good agreement. In Figure 6 the mean values of the differences (in absolute values) found when comparing spectrally averaged values calculated point-by-point (according to Table I) with each software tool, are presented. Error bars correspond to standard deviations of these differences, which provide information on the dispersion of the results. It is observed that all parameters have the same spectral behaviour as in the individual comparison. In this case, G, J_{LF}, and IACC differences are significantly reduced and also show less dispersion. However, discrepancies increase for the STI parameter, but remain below the 1 JND threshold.

4. Summary and conclusions

This paper presents a blind comparison of acoustic parameters obtained by using four widely utilized commercial room acoustics software tools (ARTA, Dirac, EASERA, and WinMLS) in order to ensure that the acoustic char-

Table VIII. Maximum variance found between the spatially averaged values calculated for each parameter considering the four software tools (15 receiver points and 2 positions of the sound source are included). Mean values obtained by averaging the spectral results calculated with the four software tools are given as a reference. Crossed analysis. ^{a}: Software C and software D are not included in the mean values of G.

		125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
T_{30} (s)	Max Variance	0.017	0.015	0.028	0.032	0.008	0.003
	Mean	2.151	1.631	1.185	1.015	1.088	0.960
EDT (s)	Max Variance	0.141	0.026	0.013	0.008	0.007	0.006
	Mean	1.894	1.416	1.159	0.977	1.103	0.981
T_S (ms)	Max Variance	3.178	0.783	0.733	0.514	0.507	1.047
	Mean	138.048	111.920	77.572	60.292	71.108	64.030
C_{50} (dB)	Max Variance	1.109	1.137	0.241	0.108	0.093	0.108
	Mean	-3.324	-2.895	-0.024	1.666	0.579	1.026
C_{80} (dB)	Max Variance	0.986	0.455	0.120	0.150	0.037	0.066
	Mean	-0.195	0.531	2.442	4.326	3.269	3.869
D_{50}	Max Variance	0.053	0.055	0.014	0.016	0.003	0.011
	Mean	0.328	0.348	0.496	0.590	0.531	0.553
G (dB)^a	Max Variance	0.043	0.057	0.013	0.007	0.083	0.083
	Mean	-1.375	-2.248	-2.913	-3.680	-0.458	0.438
J_{LF}	Max Variance	0.019	0.012	0.009	0.010	0.007	0.006
	Mean	0.138	0.193	0.253	0.273	0.220	0.176
$\mathrm{IACC_E}$	Max Variance	0.102	0.024	0.045	0.055	0.048	0.043
	Mean	0.901	0.805	0.509	0.378	0.317	0.266
$\mathrm{IACC_L}$	Max Variance	0.098	0.025	0.014	0.004	0.001	0.004
	Mean	0.879	0.675	0.247	0.152	0.099	0.084

acterisation of a room is equivalent when using one or any of the other three tools. Two of these software tools, EASERA and ARTA, have never been part of the international round robin comparison. Furthermore, this is the first study in which only the influence of the software tool on the registration of the RIRs is assessed, leaving aside the implication of any other element of the measurement chain. Room acoustic parameters (T_{30}, EDT, T_S, C_{50}, C_{80}, D_{50}, G, L_{J}, J_{LF}, $\mathrm{IACC_E}$, $\mathrm{IACC_L}$, and STI) have been calculated from the RIRs measured according to the ISO 3382-1 with the four software tools in the same room under identical conditions. A number of these parameters are included here for the first time in this type of study. As the software might perform differently in different acoustic environments, it is crucial to follow the guidelines laid out by the developers for setting up the equipment and measuring the RIRs with each software tool successfully. In that sense, if the software tools are correctly used, considering the characteristics of the room under test in each case, the results obtained in this study are transferable to other source-receiver combinations as well as to other rooms.

The comparative process to determine the precision of the acoustic software tools has been structured in three parts: a *Repeatability Analysis* of each measuring system; an *Individual Analysis*, in which RIRs have been measured and processed with each software tool independently; and a *Crossed Analysis*, in which RIRs measured with software A have been processed with all the other systems (using a common standard format).

The repeatability analysis reveals negligible errors for measurements of the four software tools, and no variations in perceptual terms, since the variability in the majority of the parameters lies below 0.1 JND, and in no case does the difference exceed 0.2 JND. Statistically, the relative error between repetitions has no dependence on the system used. In general, the average relative error committed on giving a parameter value stands at about 2%. More specifically, this relative error depends both on the frequency, where it is higher at 2 and 4 kHz, and on the parameter, where it is higher for C_{50}, C_{80}, and G. It must be taken into account that the greater values of relative error (calculated as Relative error = SD/Mean·100%) are mainly due to mean values lower than unity of these parameters at certain frequency bands. These repeatability results indicate that both the accuracy and the stability of the four acoustic software tools are satisfactory, and hence the comparison established for individual and crossed analyses is valid for the identification of the differences between these four tools when characterising a room.

Ideally, differences found on the comparison of the acoustic parameter values when characterising a room by using any of the four systems should remain below 1 JND, since this is the minimum variation that a listener can detect (see Table I). In order to ensure that those differences which exceed the subjective threshold of 1 JND are minimal, as well as to detect where these differences are encountered, it is necessary to determine the equivalence between systems. *Individual Analysis* aims to determine whether all the room acoustics software tools measure and process the RIRs in the same way. Statistically, the general study of concordance between the four measurement systems shows that, according to the ICC value, the strength of concordance is *good* for G, J_{LF} and IACC, *moderate* for L_{J}, and *very good* for the rest of the parameters. Only G and L_{J} significantly differ when using one or another software tool.

Statistical studies show that these differences are not associated with the position of the source or the receiver. Furthermore, there are only significance differences, negligible in terms of JND, between clarity and definition parameters at low frequency bands. Using software A as the reference to assess the discrepancies between the values of the parameters in terms of JNDs, the agreement between software tools when measuring and processing the RIRs is revealed, since differences remain below 1 JND in more than 91% of the cases, when G and L_{J} are not taken into account. Moreover, when characterising a room with the spatially averaged value of acoustic parameters, results obtained by using A, B, C, or D are equivalent (maximal variances below 0.5 JND), except for EDT in 125 Hz band; for C_{50}, C_{80}, and D_{50} at low frequencies; and for L_{J} and G at almost every frequency band (see Table V). The discrepancies in EDT values are mainly found when comparing with software D, and they are probably caused by the background-noise detection algorithm. Figure 4 depitcs that clarity and definition parameters are underestimated by software B at low frequency bands, which is responsible for these variances. These variances might be caused by differences in the algorithm of arrival time detection of the direct sound (t_0) implemented in each software tool. Following ISO 3382-1 recommendations, t_0 should be determined through the analysis of the RIRs in broad band and subsequently the RIRs should be filtered by frequency bands to estimate parameter spectral values. The arrival time determined in broad band with the four software tools is very similar, with a maximal variance of 0.04 ms. Nevertheless, software B seems to perform filtering of the signal prior to t_0 detection, which broadens the response at low frequency bands. Therefore, for the same RIR, t_0 varies by up to 13 ms at 125 Hz with respect to the arrival time estimated in broad band at certain receiver points. This variation may especially affect the results of these parameters at low frequency and may be the reason for the underestimation when comparing with the other software tools.

It is important to point out that spectrally averaged values (see Table I), commonly used as a single value to characterise a room, have no statistically significant differences for any parameter, since they show a maximal variance lower than 1 JND. As discussed in Section 3.2.1, the high values of maximal variances in IACC parameters are affected by an isolated and marginal point, and therefore, these values must be carefully interpreted.

In the general study of concordance for the crossed analysis, where the aim is to validate the software tools in terms of processing as well as to determine whether the four software tools are interchangeable for the analysis of RIRs, reveals a very good strength of concordance for all parameters according to the ICC. In addition, differences in terms of JNDs, considering system A as a reference, are slightly reduced with respect to the individual analysis, since differences are below the 1 JND threshold in more than 94.3% of cases (Table VII). Regarding frequency bands, there are more discrepancies at LF with 11.64%, versus 3% of data at MF and HF. With respect to parameters, there is an improvement in the agreement of T_{30}, EDT, and T_S, with a significant increase in the number of values that can be considered equivalent to those obtained with software A. This improvement also appears in the J_{LF} parameter, irrespective of the software tool used. In general, the percentage of perceptible differences decreases around 5% when analysing the same signal with the 4 software tools with respect to the *Individual Analysis*. Moreover, the *Crossed Comparison* of mean values shows that there are negligible differences when characterising a room either with spatially or spectrally averaged parameter values, and that the maximal variations of mean values are of the same order as those found in the *Individual Analysis*.

This analysis also highlights that, when using these four software tools, the *.wav* format is recommendable only for the exchange of omnidirectional RIRs from which monaural parameters will be extracted, whereas if binaural or relative parameters are required, whose calculation involves two correlated signals, it is recommended that *.tim* format be used.

Both the *Individual* and the *Crossed Analysis* show major agreement between the four software tools included in the comparative study, with a similar degree of discrepancies depending on the parameter and the frequency band analysed, which indicates that the cause of these discrepancies is due to the processing algorithm and not due to the acquisition of the RIRs. Unfortunately, limitations in handling the background noise and the lack of precision in the arrival time detection (both problems previously detected by Bradley [11] and Katz [12] in other systems) remain present in a number of the software tools used today. These issues have been further assessed in a recent independent study carried out by Cabrera *et al.* [23], in which the same four software tools tasted here are included. Their findings on the effects of both the background–noise treatment and the frequency selectivity on the calculation of parameter values, complement the results obtained in this study. An improvement in this regard could ensure an almost perfect agreement of parameter values at all frequency bands.

This research provides an insight into the reproducibility and repeatability of room acoustic characterisation results, and the assessment of the differences between these results when using different software tools in terms of the subjective discrimination threshold for each objective quantity, and hence their reliability can be assessed and the comparability of data between different research groups can be ensured. Since, according to the ISO 3382-1, no acoustic software certification is required, this paper may serve as useful supplementary material for software users if the details pointed out here are taken into account when analysing the results, in order to ensure that the acoustic software used exerts no influence on the acoustic characterisation of a room.

Acknowledgement

This research has been financially supported by FEDER funds and the Spanish Ministry of Economy and Competitiveness, with reference BIA2014-56755-P.

References

[1] C. Hak, R. Wenmaekers, L. van Luxemburg: Measuring room impulse responses: Impact of the decay range on derived room acoustic parameters. Acta Acustica united with Acustica **98** (2012) 907–915.

[2] M. Guski, M. Vorländer: Comparison of noise compensation methods for room acoustic impulse response evaluations. Acta Acustica united with Acustica **100** (2014) 320–327.

[3] R. Kürer, U. Kurze: Integrationsverfahren zur Nachhallauswertung (Integration procedure for evaluating reverberation). Acustica **19** (1967) 313–322.

[4] A. Lundeby, T. E. Vigran, H. Bietz, M. Vorländer: Uncertainties of measurements in room acoustics. Acustica **81** (1995) 344–355.

[5] W. T. Chu: Comparison of reverberation measurements using Schroeder's impulse method and decay-curve averaging method. Journal of the Acoustical Society of America **63** (1978) 1444–1450.

[6] D. R. Morgan: A parametric error analysis of the backward integration method for reverberation time estimation. Journal of the Acoustical Society of America **101** (1997) 2686–2693.

[7] N. Xiang: Evaluation of reverberation times using a nonlinear regression approach. Journal of the Acoustical Society of America **98** (1995) 2112–2121.

[8] X. Pelorson, J.-P. Vian: On the variability of room acoustical parameters: reproducibility and statistical validity. Applied Acoustics **37** (1992) 175–198.

[9] D. De Vries, E. M. Hulsebos, J. Baan: Spatial fluctuations in measures for spaciousness. Journal of the Acoustical Society of America **110** (2001) 947–954.

[10] R. San Martín, I. B. Witew, M. Arana, M. Vorländer: Influence of the source orientation on the measurement of acoustic parameters. Acta Acustica united with Acustica **93** (2007) 387–397.

[11] S. Bradley: An international comparison of room acoustics measurement systems. Internal Report, IRC-IR-714 National Research Council, Canada, 1996.

[12] B. F. G. Katz: International round robin on room acoustical impulse response analysis software 2004. Acoustic Research Letters Online **5** (2004) 158–164.

[13] P. Fausti, A. Farina: Acoustic measurements in opera houses: Comparison between different techniques and equipment. Journal of Sound and Vibration **232** (2000) 213–229.

[14] F. Policardi: MLS and Sine-Sweep technique comparison in room-acoustic measurements. Elektrotehniški Vestnik **78** (2011) 91–95.

[15] J. Segura, A. Giménez, J. Romero, S. Cerdá: A comparison of different techniques for simulating and measuring acoustic parameters in a place of worship: Sant Jaume Basílica in Valencia, Spain. Acta Acustica united with Acustica **97** (2011) 155–170.

[16] ISO 3382-1:2009(E): Acoustics - Measurement of room acoustic parameters. Performance spaces. International Organization for Standardization, Geneva, Switzerland, 2009.

[17] C. Hak, R. Wenmaekers, J. Hak, L. Van Luxemburg, A. Gade: Sound strength calibration methods. Proceeding of the International Congress on Acoustics, Sydney, Australia, 2010, 1–6.

[18] B. F. G. Katz: In situ calibration of the sound strength parameter G. Journal of the Acoustical Society of America **138** (2015) 167–173.

[19] C. L. Christensen, G. Koutsouris, J. Gil, J. H. Rindel: Applying in-situ recalibration for sound strength measurements in auditoria. Proceedings of the Institute of Acoustics **37** (2015) Pt.3.

[20] Patent No. 7,023,453: IBM SPSS Statistics 19 Core System, User's guide licensed materials. Property of SPSS Inc., an IBM Company. Copyright SPSS Inc. 1989, 2010.

[21] L. Prieto, R. Lamarca, A. Casado: Assessment of the reliability of clinical findings: the intraclass correlation coefficient. Medicina Clinica (Barcelona) **110** (1998) 142–145.

[22] J. Blauert, N. Xiang: Acoustics for engineers. Second edition. Springer-Verlag, Berlin, 2009.

[23] D. Cabrera, J. Xun, M. Guski: Calculating reverberation time from impulse responses: a comparison of software implementations. Acoustics Australia (2016) in press.

[24] T. Okano, L. L. Beranek, T. Hidaka: Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls. Journal of the Acoustical Society of America **104** (1998) 255–265.

[25] M. Vorländer: Auralization: Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. First edition. Springer-Verlag, Berlin, 2008.

[26] J. S. Bradley, R. Reich, S. G. Norcross: A just noticeable difference in C50 for speech. Applied Acoustics **58** (1999) 99–108.

[27] D. Cabrera, D. Lee, G. Leembruggen, D. Jimenez: Increasing robustness in the calculation of the speech transmission index from impulse responses. Building Acoustics **21** (2014) 181–198.