The just noticeable difference of center time and clarity index in large reverberant spaces

F. Martellotta[a)]

Dipartimento di Architettura e Urbanistica, Politecnico di Bari, via Orabona 4, I70125 Bari, Italy

(Received 22 February 2010; revised 26 May 2010; accepted 27 May 2010)

Just noticeable difference (JND) values are available for most acoustical parameters currently used in practice. However, they have been determined with reference to conditions typically encountered in concert halls and in rooms for speech, covering a range of reverberation times (T) spanning from 0.5 s to 2 s. When reverberation gets longer, the relationship between measured parameters describing acoustic clarity may change significantly and subjective perception might also be different. The proposed research investigates the influence of reverberation time on JND for clarity measures taking into account three reference cases having T values varying from 2 s to 6 s. Measured B-format impulse responses were properly modified to introduce the desired changes and then auralized with two music motifs for presentation on a 4-channel playback system. Listening tests based on paired comparisons were carried out to determine subjective limens. The results proved to be independent of music motifs and showed that JND in the clarity index is almost independent of T, while JND in the center time is significantly related to T and can be assumed as the 8.5% of the reference T_S value.

 [DOI: 10.1121/1.3455837]

PACS number(s): 43.55.Hy [NX] Pages: 654–663

I. INTRODUCTION

Just noticeable difference (JND) is defined as the smallest perceivable change in a given measurable parameter, and is an important factor which correlates the subjective dimension of sound perception to its objective measures. JND measures the sensitivity of the listeners to a change in a given parameter. Consequently, a precise estimate of the JND is of fundamental importance for a number of applications, such as determining the accuracy of a measurement, assessing the reliability of prediction software,[1] or finding the required spatial resolution of a measurement grid capable of describing the subjective variations of the sound field in a room.[2] Given the large number of acoustic parameters which can be measured nowadays and their widespread usage, it is particularly useful to have well known JND values for each one of them.

The first studies into this field were carried out with reference to reverberation time (T), as reported by Cremer and Muller,[3] showing that for T values above 0.6 s the JND was about 4% of the observed value, while below 0.6 s the subjective limen became "absolute" and approximately equal to 0.024 s. Above 0.6 s the relative variation JND was strictly dependent on T values, showing a U-shaped trend, with high values at short T, reaching a minimum (of about 4%) between 1 s and 2 s, and then increasing again up to 5% when T approached 6 s. More recent studies,[4] based on realistic binaural reproduction of sound fields, showed that below 0.6 s the difference limen was 0.042 ± 0.015 s, twice the value resulting from previous studies. However, significant variations appeared as a function of the program material used, suggesting that the transient nature of the signals may explain such variance.

Reichardt *et al.*,[5] using synthetic sound fields, studied difference limens for the delay and the level of individual reflections in an impulse response. These results could be converted into JND for clarity, but as they refer to an impulse response not properly simulating concert-halls conditions they were not considered as significant.

Later on, Cox *et al.*[6] carried out a newer set of experiments, still based on synthetic sound fields, but using a sequence of early reflections and reference acoustic conditions more similar to those typically observed in concert-halls (T $=2.1$ s). They provided a more detailed discussion of JND for clarity parameters and for spatial impression, also taking into account the effect due to different music motifs. The resulting difference limen for center time (T_S) was, on average, 8.6 ± 1.6 ms for a reference value of about 80 ms. They also found a significant difference as a function of music motif, with a larger value (11.4 ± 2.7 ms) corresponding to a more legato and slow moving motif, and a nearly halved value (5.7 ± 0.9 ms) when the motif was faster. From this measurement, taking into account a theoretical relationship, they also derived the corresponding JND of clarity index (C_{80}) which for the slow moving motif was 0.92 ± 0.22 dB and for the faster motif was 0.44 ± 0.07 dB, giving an average of 0.67 ± 0.13 dB.

Further investigations were carried out by Bradley *et al.*[7] mostly focusing on clarity of speech (C_{50}). They considered three different reverberation times, varying from 0.5 s to 2.0 s, showing that the corresponding variations in JND were not statistically significant. Thus they concluded that under the observed conditions, JND for C_{50} was independent of reverberation time and was 1.1 dB. The corresponding JND for

[a)]Author to whom correspondence should be addressed. Electronic mail: f.martellotta@poliba.it

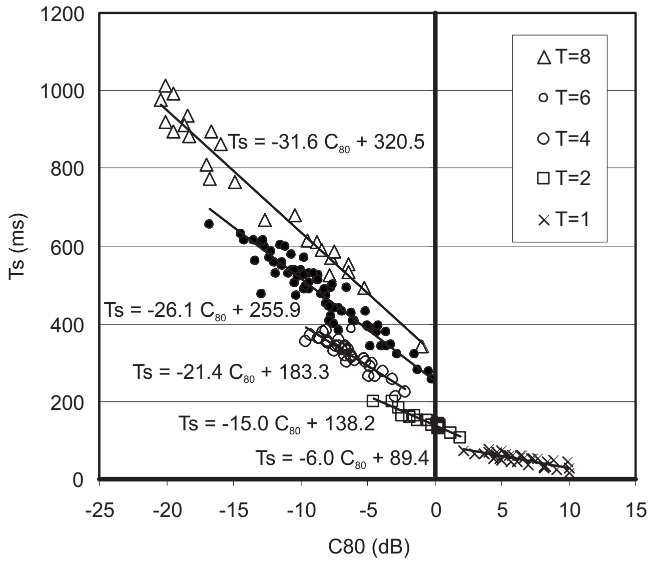

FIG. 1. Different relationship between experimental values of C_{80} and T_S (averaged over 500 and 1000 Hz octave bands) as a function of reverberation time. Pairs of C_{80} and T_S values were measured in different churches (Ref. 11) and grouped according to the corresponding reverberation times.

related parameters, such as C_{80}, was equal to 0.9 dB, showing substantial agreement with the upper value found by Cox *et al.*,[6] even though the temporal structure of speech is more likely to be comparable with a fast moving music motif, suggesting that the similarity was just apparent.

Okano[8] tried to shed more light on the perception of differences in relevant attributes of sound fields in relation to intensity variations of early reflections, considering different sound fields typical of concert-halls. Even though this research was based on the responses of just three listeners, the results showed, among the other things, that larger clarity variations (equal to 1.5 ± 1.0 dB) are required in order to be subjectively perceived compared to previous studies. In addition, the same study showed that varying the intensity of a given reflection may be more easily detected subjectively in terms of spaciousness or loudness rather than clarity. This suggests that varying a single reflection may be counterproductive when investigating JND for clarity measurements.

Recently, Ahearn *et al.*[9] carried out a preliminary study on JND for C_{80} which investigated different reverberation times (1.6 and 2.1 s), different music motifs, and a large number of subjects. The results proved to be significant only when a selection of the most reliable participants was considered and gave an average JND of about 1.6 dB, with significant variations as a function of music motifs and reverberation time.

All the above mentioned studies used synthetic sound fields and were focused on acoustic conditions typically found in concert halls or auditoria, that is taking into account rooms with reverberation times generally below 2.1 s. Bradley's study suggested that below that limit there should be no dependence of JND as a function of the reverberation time of the room. However, in large reverberant rooms such as churches, reverberation time is likely to be above that limit and, when T changes significantly, the relationship between C_{80} and T_S also changes (Fig. 1) leading to possible discrepancies between JND for different parameters. In fact, a 1 dB variation in C_{80} corresponds to a 6 ms change when $T=1$ s, which gradually increases up to 32 ms when $T=8$ s. This means that when reverberation time grows, even though JND remains the same for C_{80}, it is likely to change for T_S, or vice versa.

Given this premise, the present research aims at investigating possible dependence of JND for T_S (and C_{80}) as a function of reverberation time, taking into account the particular case of large reverberant enclosures. Measured B-format impulse responses were used as a reference (instead of purely synthetic sound fields), in order to provide the most realistic conditions.[10] Small modifications were made in order to produce a variation in clarity parameters without affecting other acoustic aspects. The resulting impulse responses were convolved with anechoic material and played back on a 4 channel system. A varied set of listeners was then exposed to pairs of sound fields in order to decide whether they were identical or different (and in the latter case they had to identify which one between A and B was clearer). Three different reverberation times were used as a reference, respectively equal to 2 s, 4 s, and 6 s, together with two different music motifs.

II. EXPERIMENTAL SET-UP

A. Sound field characteristics

Three different sound fields with increasing reverberation times were taken as a reference in order to investigate possible dependence on this parameter. A lower limit of $T=2$ s was assumed for the reference signal, which corresponded to conditions observed in concert-halls. Then, in order to take into account acoustic conditions characterized by longer reverberation times and, at the same time, minimize the number of listening tests to be performed, two additional reference conditions were used, respectively having a mid-frequency reverberation time of about 4 s and 6 s.

In order to improve the realism of the listening test, measured impulse responses (IRs) recorded with a B-format microphone (Soundifield Mk-V) were considered as a valid alternative to purely simulated responses given the good performance this method showed.[10] The three reference IRs were chosen from a large set of experimental measurements collected in churches,[11] according to the desired acoustic characteristics. All the measurements were carried out complying with the ISO 3382 standard,[12] using an omnidirectional sound source, together with an additional subwoofer to cover the frequencies below 100 Hz. The rooms were excited with a constant envelope equalized sine sweep generated using MATLAB according to Müller and Massarani[13] so that the spectrum of the radiated sound was substantially flat from the 50 Hz to 16 kHz third-octave bands. The resulting impulse responses, obtained after deconvolution, had very low noise (the signal to noise ratio was generally higher than 60 dB even at the lowest frequencies) in order to be safely used for auralization purposes. Consequently, all the measured reverberation times were calculated over a 30 dB decay.

TABLE I. Summary of the measured acoustic parameters in the three churches. Multi octave averages according to ISO 3382.

Church	San Cataldo	San Lorenzo	Sant'Ambrogio
T (s) (500–1000 Hz)	2.1	4.1	6.0
EDT (s) (500–1000 Hz)	1.8	4.2	5.8
BR	1.10	1.0	1.10
T_S (ms) (500–1000 Hz)	130	286	378
C_{80} (dB) (500–1000 Hz)	−0.8	−4.2	−4.5
C_{50} (dB) (500–1000 Hz)	−3.8	−6.4	−5.4
D_{50} (%)(500–1000 Hz)	31	19	22
LF (125–1000 Hz)	0.25	0.33	0.24
IACC (500–2000 Hz)	0.43	0.29	0.36

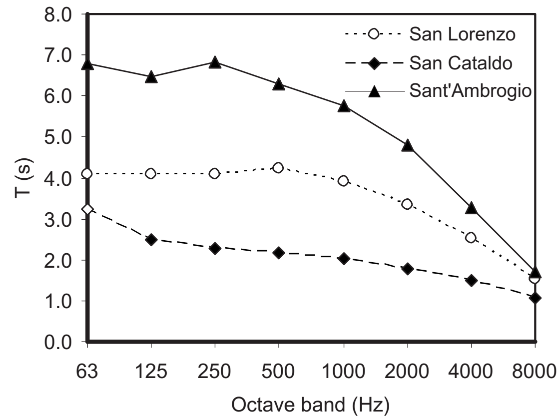

FIG. 2. Measured reverberation time (T) as a function of octave bands for the three selected churches.

The selected churches were the cathedral of San Cataldo in Taranto ($V=10\,500\ \mathrm{m}^3$), having a mid-frequency T of 2.1 s, the church of San Lorenzo in Turin ($V=12\,000\ \mathrm{m}^3$) having a mid-frequency T of 4.1 s, and the Basilica of Sant'Ambrogio in Milan ($V=23\,000\ \mathrm{m}^3$), having a mid-frequency T of 6.0 s. For each church the sound source was located in front of the altar and the receiver was chosen among those located in the main nave at a distance between 2.5 and 2.9 times the critical distance to prevent direct sound from being too strong and ensure a comparable balance between early and late reflections. Details of the acoustic parameters measured in the three churches at the selected positions are given in Table I. In all the cases the reverberation time varied as a function of frequency, according to the typical behavior observed in churches, with low-frequency values about 10% longer than mid-frequencies, and high-frequency values rapidly decreasing as a result of air absorption (Fig. 2).

B. Sound field simulation

The listening room had a nearly rectangular plan, with internal dimensions of $3.70\times2.50\times2.40$ m and was designed to be as dry as possible, given architectural limitations, in order to allow for the reproduction of virtual sound fields. Recommendation ITU-R BS.1116-1[14] was followed when possible, given that the scope of the room used was rather different from the "reference listening room" described in the norm. The room used had a flat frequency response and a reverberation time of 0.07 s at medium frequencies, decreasing to 0.03 s at 8 kHz. Below 250 Hz the reverberation time gradually grew up to 0.30 s at 63 Hz (the reverberation time of the octave band filter being approximately 0.15 s). The room was acoustically insulated, using a floating construction, in order to minimize external noise to levels below 20 dB (A weighted).

The four loudspeakers (Yamaha MSP5) have a flat frequency response from 60 Hz to 30 kHz and were placed on stands and arranged in a square rig with the listener located at the center at a distance of 1.3 m from each one of them. The level of the speakers was carefully aligned at the listening position.

In order to determine the acoustic parameters characterizing the reproduced sound fields, the selected B-format IRs were convolved with a logarithmic sine sweep and then Ambisonic decoded in order to be played on the square rig loudspeaker arrangement. The resulting simulated room responses were picked up at the listening position using the same microphones used during the on-site survey and finally deconvolved. The resulting acoustic parameters are reported in Table II. A comparison with the corresponding parameters measured on-site shows remarkable agreement. In particular, differences in monaural parameters were negligible and largely within conventionally accepted JND values. Conversely, LF shows differences which, in some cases, are close to or slightly above JND values. Possible causes of this discrepancy may be found in the non perfectly anechoic conditions of the listening room in the low frequency range, which might have added some additional lateral reflections.

TABLE II. Summary of the reproduced acoustic parameters in the three churches. Multi octave averages according to ISO 3382. Values reported in parentheses correspond to relative (for T and EDT) and absolute differences between measured and simulated acoustic parameters, compared with conventionally accepted JND values reported in the final column.

Church	San Cataldo	San Lorenzo	Sant'Ambrogio	JND
T (s) (500–1000 Hz)	2.1 (0.0%)	4.1 (+0.2%)	6.0 (−0.1%)	5%
EDT (s) (500–1000 Hz)	1.8 (−3.9%)	4.1 (+2.0%)	5.8 (+1.1%)	5%
T_S (ms) (500–1000 Hz)	130 (−0.3)	276 (+6.8)	372 (+5.6)	10 ms
C_{80} (dB) (500–1000 Hz)	−0.7 (−0.1)	−4.1 (−0.1)	−4.4 (−0.2)	1 dB
C_{50} (dB) (500–1000 Hz)	−3.8 (0.0)	−6.7 (−0.3)	−5.2 (−0.2)	1 dB
LF (125–1000 Hz)	0.32 (−0.07)	0.28 (+0.04)	0.28 (−0.05)	0.05
IACC (500–2000 Hz)	0.42	0.42	0.48	0.075

IACC values measured at the listening position were also given, even though differences were not considered as during the on-site measurements the B-format and binaural microphones were not coinciding (the average distance between them being about 0.5 m) and misalignments might sometimes take place, consequently any direct comparison would not be significant.

However, such small discrepancies were not considered as a major problem. In fact, the simulated values are well within the range of values typically observed in churches and, above all, the aim of the research was to investigate the sensitivity of listeners to sound field changes, so the main concern was to ensure that the desired modifications to the reference sound fields were correctly rendered by the playback system.

C. Sound field modification

In order to investigate subjective sensitivity to center time (and clarity) variations, the selected B-format IRs had to be carefully modified. Previous research[6,7] used simulated sound fields, and variations were obtained by simply increasing the level of the reflections or changing their arrival time. When dealing with measured IRs, things are more complex because of the presence of diffuse reflections. In addition, varying the intensity and the arrival time of a single reflection may easily cause changes in the spatial perception of the given sound,[8] thus, in order to keep such variations to a minimum, the amplitude of the early part of each IR (including direct sound and the reflections arriving within the first 80 ms) was gradually reduced in order to determine, over the octave bands from 125 Hz to 4 kHz, a mean increase in T_S of 10 ms (Fig. 3). As the modified part was at the very beginning of the IR the reverberation time was barely affected and the *EDT* showed only slightly larger variations. In addition, as the reduction involved both direct sound and early reflections and was equally applied to all the B-format components of the IR (W, X, Y, and Z) no significant variation was expected either in terms of *LF* or *IACC*. The procedure was then repeated on the new IR until the desired range of T_S variations was fully covered. For the 2 s case, four IRs were obtained, corresponding to a 40 ms difference in T_S. For the 4 s case, five IRs were obtained, corresponding to a 50 ms difference in T_S. Finally, for the 6 s case, seven IRs were obtained, corresponding to a 70 ms difference in T_S.

In order to check whether the proposed modifications determined the desired effects, the B-format IRs were convolved again with a sine sweep, decoded, and played back in the listening room. Results are reported in Table III and show that the proposed method was actually successful in producing the desired differences in T_S (and consequently in other energy ratios), while keeping the other acoustic parameters well within JNDs. It is interesting to observe that, as stated above, when the reverberation time increases, the same 10 ms variation in T_S corresponded to a decreasing variation in C_{80} and C_{50}. In fact the resulting total difference for C_{80} was about 4 dB independent of the reference T value.

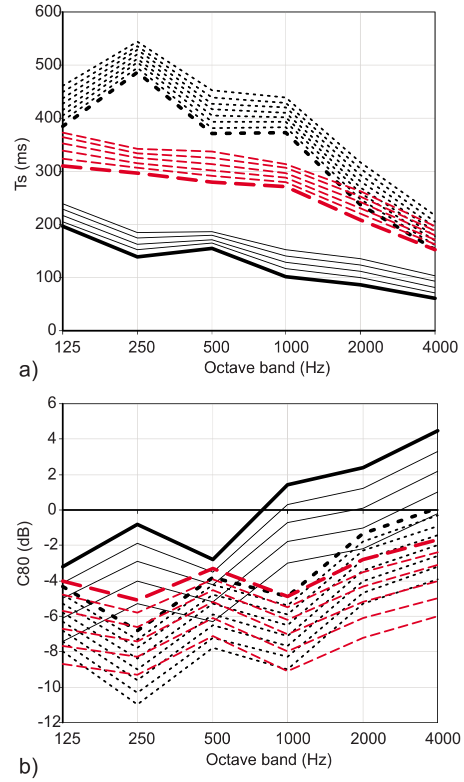

FIG. 3. (Color online) Plot of the frequency distribution of changes in T_S (a) and C_{80} (b) following the applied modifications to reference IRs (thicker lines). (—) $T=2$ s; (– –) $T=4$ s; (- - -) $T=6$ s.

D. The source signal

As shown by other studies,[4,6,8] JND may also depend on the characteristics of the motif used to present different sound fields. Consequently, even though the number of combinations of IRs was large and the procedure relatively time consuming, two different source signals were finally used.

The first was a 5 s excerpt taken from the “Pange Lingua” hymn (attributed to St. Thomas Aquinas) sung in Phrygian mode and recorded in practically anechoic conditions. This kind of liturgical singing was considered particularly suitable to present sound fields recorded in churches and, taking into account the results reported in Ref. 15, it should perfectly fit the acoustic conditions under investigation. As it is a voice-only music motif, its spectrum (Fig. 4) is unevenly distributed and is mostly rich in mid frequencies with a notable contribution from low frequencies.

The second was a 9 s excerpt from a Weber theme for cello (taken from the CD “Music for Archimedes” distributed by Bang&Olufsen) recorded in anechoic conditions. This solo piece of music is well suited to being convolved with IRs obtained from a single-point sound source, it is characterized by faster tempo and staccato notes compared to the Gregorian chant. Its spectrum is characterized by a notable content of medium and high frequencies (from 200 Hz to 4 kHz), and the attack/decay transients of the cello are shorter than those of the human voice, providing substantially different listening conditions.

In order to compare the sound signals according to Ando’s theory,[16] the long-time, τ_e, and the short-time, $(\tau_e)_{min}$, duration of the auto-correlation function were calculated for

TABLE III. Summary of the relative (for T and EDT) and absolute differences (over octave bands from 125 Hz to 4 kHz) in acoustic parameters due to manipulation of early part of selected IRs. "Individual" values represent average differences between each modified IR. "Total" values represent global differences between original (reference) IR and the one with the largest modifications.

	T=2 s		T=4 s		T=6 s	
	Individual	Total	Individual	Total	Individual	Total
T (%)	0.3%	1.2%	0.1%	0.4%	0.1%	0.6%
EDT (%)	1.1%	4.6%	0.5%	2.4%	0.3%	2.3%
T_S (ms)	11	43	10	51	10	70
C_{80} (dB)	1.1	4.3	0.9	4.3	0.6	4.0
C_{50} (dB)	1.3	5.0	0.9	4.7	0.6	3.9
LF	0.010	0.039	0.008	0.042	0.009	0.065
IACC	0.007	0.028	0.006	0.032	0.004	0.025

each music motif. In agreement with the above considerations about the structure of the motifs, the Gregorian chant shows longer values $[(\tau_e)_{\min}=84\ \text{ms},\ \tau_e=105\ \text{ms}]$ compared to Weber's theme $[(\tau_e)_{\min}=57\ \text{ms},\ \tau_e=60\ \text{ms}]$, suggesting that the first is most suitable for more reverberant spaces.

Both the selected signals were convolved with the previously arranged B-format IRs and finally decoded for Ambisonic presentation. As any level difference may represent an important clue to detect possible differences all the convolved signal were carefully adjusted in order to give for each signal a mean level of about 75 dB (A-weighted) during the presentation.

E. The listening test

The most delicate part in JND research is certainly represented by the listening test. The *method of minimal changes*, or *method of limits*,[17] is one of the most reliable and efficient tools for getting the difference limens of a quantity. According to this method the sounds are presented in pairs (sound A and sound B, differing in the selected characteristic) and all the subjects have to judge is whether or not there is a difference in the sounds.

The pairs start with a large difference in clarity, and subsequent pairs have a decreasing clarity difference between them. Eventually audible parity is met and the pairs sound the same. Then the clarity difference starts to increase until once again listeners can hear a difference in the pairs.

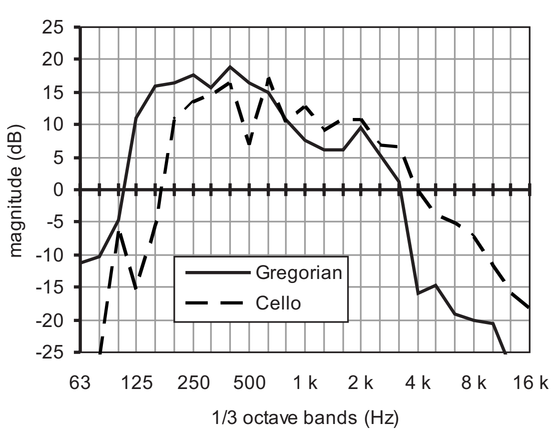

FIG. 4. One-third octave band spectra of the anechoic motifs used in the test

When going from noticeable differences to not noticeable differences (descending series) subjects are expected to perceive the differences up to a certain change in the parameter value. This first difference limen (DL^-), named "just not noticeable difference" is defined as the average of the last noticeable and of the first not noticeable difference. Then the listeners should keep on listening without being able to perceive differences until, when the difference grows again (ascending series), they again become able to perceive differences. The difference limen when a listener begins to perceive differences (DL^+) represents the "just noticeable difference" and is given, again, by the average of the last not noticeable and of the first noticeable difference. Consequently, for each listener the resulting JND is the average of the two limens, provided that statistical analysis (performed with the Student t-test) confirms that the two sets of responses given by the listeners are not significantly different.

Ideally, the pattern of subject's responses should be a sequence of "different," followed by a sequence of "same," and finally a new sequence of "different." If the responses do not fit this pattern the listener is likely to be guessing and is not included in the analysis.

In order to introduce a further element to investigate subjects' reliability in discriminating between presented sounds, the listeners were asked to assess (if able to perceive a difference) which of the two sounds was clearer than the other and how much. This proved useful to focus listeners' attention on the only subjective aspect where differences could be found and allowed us to check the reliability of their answers. In fact, when subjects stated that sounds were different but they failed to identify the clearer sound, it meant that they were guessing or that the sounds differed by an amount which was below their subjective limen but they were not recognizing it. The latter effect might derive from the knowledge of the order in which the pairs are presented, which introduces less uncertainty in the subject's judgment, but makes the measurement prone to habituation and expectation. Habituation consists of waiting until sometime after the true limen before switching the judgment from "different" to "same" (or vice versa). Expectation consists in switching the judgment from "different" to "same" (or vice versa) before the true limen is reached. In order to interpret the results in the most rigorous way, when a listener rated the sound as "different" but failed to identify the clearer one, the response was actually assumed to be a "same." Then the sequence was analyzed and if it respected the pattern described above (i.e., if the "error" happened immediately before or after the sequence of "same") the listener was still considered reliable, otherwise he/she was excluded from the analysis.

The whole test was controlled by the listeners by means of a graphical interface developed in MATLAB, originally based on the LISE environment developed by Rioux[18] and subsequently modified in order to support multi-channel audio playback, optional near-instantaneous switching between signals, choice of starting and ending point, and loop between signals under test. For each comparison the subjects had to state whether or not they were able to perceive differences, and, in the first case, had to express (using a slider) which signal was clearer and by how much.

For a given music motif the listening test consisted of a training session in which, for each of the three reference cases (differing in T), the signals with the largest differences were presented twice in order to let the listener become familiar with the interface and, above all, with the sounds. The purpose of the research was clearly explained during individual sessions and listeners were assisted during their first listening experiences in order to let them focus on acoustic clarity and learn to make distinctions between different signals. At the end of this step (which was presented independently of the listener's experience and which could be repeated in case the listener showed significant problems identifying the clearest sounds), subjects performed the actual listening test made up of three different series one after the other in succession (i.e., one for each of the three base cases). The base cases were presented in random order to prevent fatigue effects. In addition, in order to prevent habituation and expectation effects, the number of comparisons changed for each base case.

F. The subjects

About 40 subjects participated the test, they varied in age from 20 to 55 years, they had different musical and listening background (from completely inexperienced listeners to professional musicians and audio engineers), in order to have a sample representing a broad range of typical listeners. Their hearing was not tested but they were only accepted if they had no known hearing problem.

The subjects started by listening to one of the motifs (selected randomly) and, after a variable period of time (generally three to four weeks) they were asked to repeat the test using the other motif. In this way any training or memory effect was excluded.

A careful post screening of the subjects was carried out and all the listeners whose responses did not comply with the pattern explained above were excluded from the analysis. The imposed conditions proved to be very demanding as the resulting number of "reliable" listeners was reduced to 13, with a minimum of 10 for the 2 s base case (Table IV). Most of the subjects reported that the test was "difficult," some had serious problems in identifying as different even the signals with the largest differences, others had problems in actually understanding the concept of clarity. Finally, a few subjects reported problems with the graphical interface or failed to find an effective listening method. It is interesting to observe that, apart from one of the subjects who was unable to perform the second session, all those who correctly performed the first test, also performed well in the second one. A selection of five of the listeners who performed badly in the first test, but had shown only minor problems, were also invited to perform the second one. An improvement in their performance was expected following their previous experience. They were further instructed about the importance of reporting their auditory sensations and were carefully assisted during the training session in order to help them understand the concept of clarity and how to discriminate it. Nonetheless, only one of them provided reliable responses over all the base cases, suggesting that the others were probably unable to keep the same judgment criterion for all the cases, or, more simply, they were unable maintain concentration for a long time. The final sample remained well assorted, including four musicians, one audio engineer, and eight normal listeners.

TABLE IV. Summary of the subjects participating the tests complying with the selection criteria.

	$T=2$ s	$T=4$ s	$T=6$ s
Gregorian	12	13	13
Cello	10	13	13

III. RESULTS

The JND of a given parameter is defined as the smallest perceivable change detected by 50% of the subjects. In statistical terms this value is the average of individual JNDs, provided that they are available (as when the "method of minimal changes is used[6]). When subjective responses are collected by means of different methods, such as the "method of constant stimuli,"[7] the percentage of subjects perceiving a difference is plotted as a function of the change in the measured parameter and the JND is defined by means of regression analysis according to its definition as the change corresponding to 50% of detections. When using the method of minimal changes this plot may be useful to show the distribution of the results, but the conventional approach is to calculate JND as the mean (with the corresponding confidence interval) of the individual difference limens, obtained, according to the procedure described above, as the average of DL^- and DL^+, provided that the corresponding sets of values referred to the whole sample of listeners are not significantly different. In statistical terms the similarity between two different samples may be assessed by means of a Student t-test, which (through the calculation of the test statistic t, together with the number of degrees of freedom) provides the residual probability p that the samples have the same mean and standard deviation. Generally $p \leq 0.05$ is assumed as proof that the samples have different statistical descriptors. This test was also applied to investigate the JND

TABLE V. Summary of the measure JND with reference to Ts.

	$T=2$ s (ms)	$T=4$ s (ms)	$T=6$ s (ms)
Gregorian	14.3 ± 3.7	21.7 ± 4.3	33.9 ± 4.8
Cello	12.8 ± 2.4	19.2 ± 4.8	30.2 ± 4.7
Average	13.6 ± 2.3	20.4 ± 3.2	32.1 ± 3.4

dependence on different motifs and on different reverberation times by comparing the similarities of the different results.

A. JND for center time T_S

With reference to the Gregorian motif, no significant difference appeared between DL^- and DL^+ for each of the base cases. The limens differed by a maximum of 4 ms (when $T=4$ s) and the *t*-test proved that this difference was not significant ($p=0.279$). Therefore their average was assumed as the actual JND for each base case (see Table V), showing that increasing the reverberation time determined an increase in the JND of T_S. The confidence intervals were quite large compared to previous studies[6] and they were actually close to each other. However, the *t*-test confirmed that the mean values were different. In fact, the residual probability values calculated between the 2 s case and the others were respectively 0.018 and less than 0.001, while the *p* value between the 4 s case and the 6 s case was 0.001, well below the 0.05 limit.

With reference to the cello theme by Weber, the results for center time showed some interesting differences (Table V). First of all, the analysis of DL^- and DL^+ for each base case pointed out a different trend (Fig. 5). In fact, even though the *t*-test confirmed that their average values were statistically the same ($p=0.088$ for $T=6$ s), the two values differed by a larger amount following a systematic trend for higher DL^- (i.e., the jnnd) and lower DL^+ (i.e., the jnd), suggesting that listeners generally found it easier to detect the differences after reaching subjective parity. This behavior is frequently encountered in this kind of tests[4] and results from "expectation" errors which might appear more clearly after reaching parity because the cello theme has shorter attack/decay transients than the slow moving Gregorian excerpt which make the detection of the differences easier.

This may also explain the generally smaller JND compared to the corresponding values found for the Gregorian motif, with differences varying from 2 ms to about 4 ms. However, the *t*-test applied to samples corresponding to different motifs but the same base case proved that the observed differences were not statistically significant, at 95% confidence level. In fact the residual probabilities *p* for the three base cases were respectively 0.52, 0.46, and 0.29. So, despite the observed differences, JND values were not motif-dependent. The relationship with reverberation time appeared again, even though the *t*-test applied to the 2 s and 4 s base cases gave a $p=0.044$, suggesting that the difference was on the very limit of significance, while in the other cases, between 2 s and 6 s *p* was less than 0.001, and between 4 s and 6 s *p* was 0.004.

However, taking into account that JND values were not motif-dependent it seemed more appropriate to calculate the resulting averages for each base case (Table V) which (thanks to the larger number of cases) were all statistically different ($p<0.005$ for all the combinations), clearly showing the dependence as a function of reverberation time.

The analysis of the plot of the percentage of subjects who perceived differences as a function of changes in T_S (Fig. 6) showed a behavior in good agreement with previous

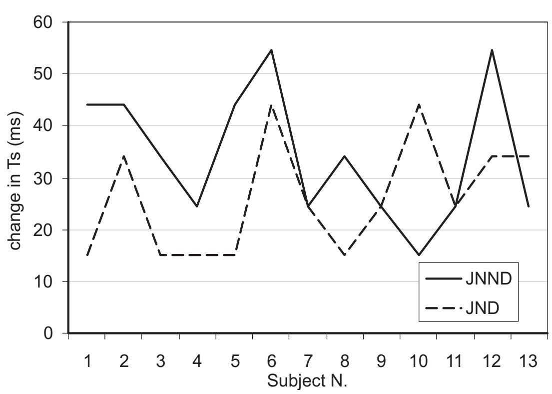

FIG. 5. Variation of JND and JNND as a function of different subjects. Base case $T=6$ s, in combination with Cello theme.

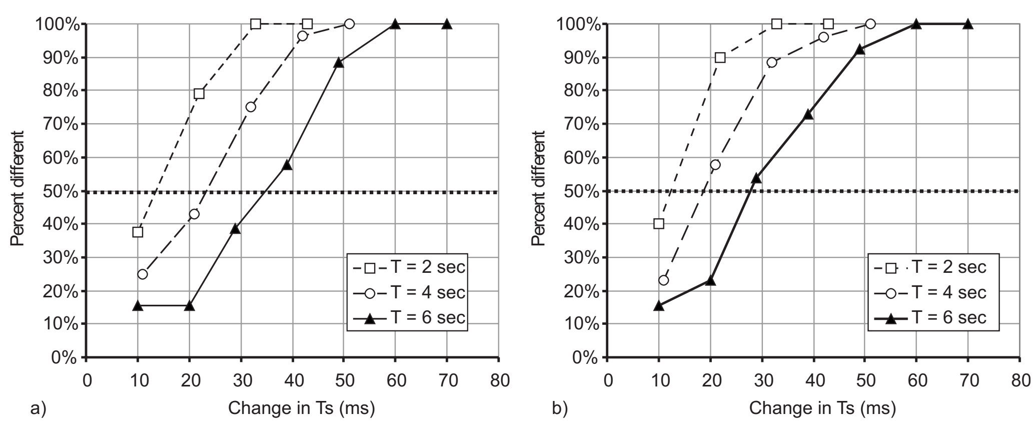

FIG. 6. Plot of the psychometric function showing the percentage of subjects detecting a change in T_S as a function of the parameter variation. (a) Gregorian chant; (b) Cello theme

TABLE VI. Summary of the measured JND with reference to C80.

	$T=2$ s (dB)	$T=4$ s (dB)	$T=6$ s (dB)	Mean (dB)
Gregorian	1.34 ± 0.34	1.52 ± 0.32	1.78 ± 0.27	1.55 ± 0.18
Cello	1.21 ± 0.22	1.37 ± 0.40	1.56 ± 0.30	1.39 ± 0.19
Average	1.28 ± 0.21	1.45 ± 0.25	1.67 ± 0.20	1.48 ± 0.13

observations. In particular, for a given change in T_S subjects listening to the cello theme generally showed higher percentage of detection compared to Gregorian chant. The plots also showed that excluding the largest differences, for which 100% of subjects detected a variation, the dependence between the two variables was substantially linear. The correlation coefficients R^2 were very large (>0.95) and allowed the definition of a JND corresponding to 50% of detections. For the Gregorian chant the resulting values for the three base cases were the following: 13.5 ms, 22.3 ms, 33.0 ms. So they were in good agreement with the results of the previous analysis. For the cello theme the resulting values for the three base cases were the following: 12.4 ms, 19.3 ms, 28.7 ms. Again the agreement with the previous values was good, only the 6 s value was slightly shorter.

B. JND for clarity index C_{80}

The same subjective ratings were then used to determine the JND for clarity index (C_{80}). For the Gregorian chant the differences between DL^- and DL^+ for each base case were negligible (the smallest p-value was 0.29) and their average was assumed as the JND (Table VI), showing a slight increase of JND as T grows. However, for this parameter the between-case differences were not significant, as the residual probability p was much greater than 0.05 for all the combinations (apart from the 2 s and 6 s for which $p=0.06$), suggesting that a single average value might be more appropriate than three distinct. This was also confirmed by the analysis of the plot (Fig. 7) which showed that the points representing the three base cases were more clustered. The resulting average value of the JND was 1.55 ± 0.18 dB, greater than the normally accepted value which is similarly defined with reference to a 2 s reverberation time (even though in that case the reference C_{80} was much higher, being 3 dB). This might depend on the slow moving musical structure of the Gregorian chant which tends to mask clarity variations.

In fact when the results referred to the cello theme were considered some differences appeared. As already observed for T_S DL^- and DL^+ differed by about 0.4 dB for the first two cases, and by 0.5 dB for the 6 s case. In the latter case, according to the t-test, the difference was on the very limit of significance (as $p=0.068$). However, despite such boundary conditions, the JND could still be calculated as the average of DL^- and DL^+ for each base case (Table VI). The t-test demonstrated that the variations between base cases were not significant (the smallest p-value being 0.09 between $T=2$ s and $T=6$) and, consequently, a unique mean value could be conveniently used for the cello theme, equal to 1.39 ± 0.19 dB. The plot reported in Fig. 6 shows that the measured points were nearly coinciding, defining a uniform relationship between the percent different and the actual change in parameter value.

In comparison with the results for the Gregorian chant the JND values for clarity were generally smaller by about 0.2 dB. However, the t-test confirmed that the change was not statistically significant (p-values were respectively 0.52, 0.67, and 0.30), so the values were not motif-dependent. This suggested that an average value could be determined for each base case (Table VI). However, in this case the t-test showed that (following the increased number of observations) the differences between the extreme base cases (2 s and 6 s) were statistically significant ($p=0.012$), while the differences between adjacent cases were not statistically significant (lowest $p=0.18$). This suggests that there is a mild dependence on T, but, given the relatively small global variation, a unique JND value, independent of music motif and of T, still seemed a better choice for C_{80}. Averaging all the previous data the final value was 1.48 ± 0.13 dB.

C. Discussion and comparison with previous studies

In summary, the results discussed above showed that the perceptual differences between ascending and descending series were mostly negligible. The cello theme seemed to allow the listeners to detect jnd (DL^+) more easily than jnnd (DL^-),

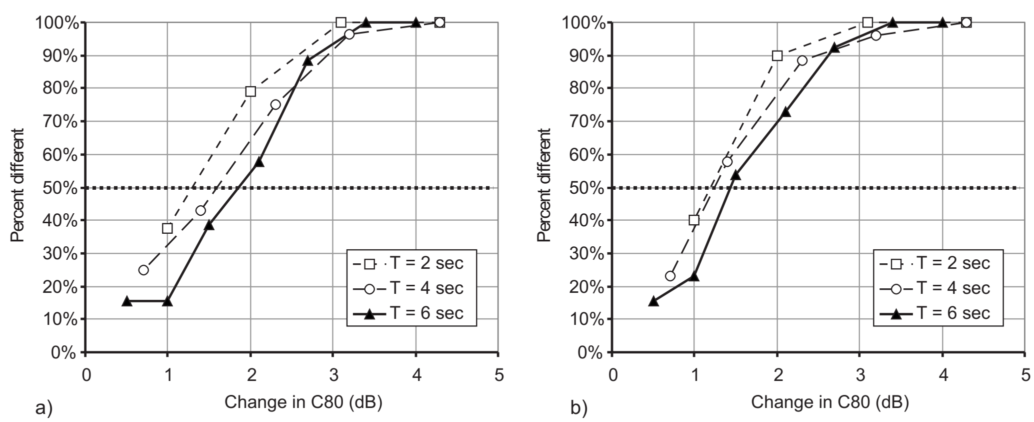

FIG. 7. Plot of the psychometric function showing the percentage of subjects detecting a change in C_{80} as a function of the parameter variation. (a) Gregorian chant; (b) Cello theme

but statistical tests showed that such differences were not significant, so the mean of both values was assumed for each motif and each base case.

Similarly, even though the faster music motif generally determined a smaller JND for both T_S and C_{80}, such differences were not statistically significant (at least with reference to the motifs used in this paper), suggesting that the mean value of the JNDs corresponding to different motifs was a better descriptor.

The above considerations applied to both JND of T_S and C_{80}. However, when the relationship with the reverberation time was investigated (i.e., when the results of the three base cases were taken into account), significant differences appeared between the two acoustic parameters.

In fact, JND of C_{80} showed only mild variations, with adjacent confidence intervals generally superimposed, which proved to be statistically significant only when the extreme base cases are considered ($T=2$ s and $T=6$ s) and when results from both music motifs are taken together. This suggested that assuming JND to be independent of T might be a safer practical assumption, in agreement with the results found by Bradley[7] for C_{50} over the interval from 0.5 s to 2 s, and by Ahearn *et al.*[8] However, it must be underlined that the JND value found by Bradley was 0.9 dB, while the average JND found in the present research and in ref. 8 was about 50% larger. This might depend on the decreasing trend shown by JND as a function of T, combined with the fact that speech signals are very different from music (being made of "shorter" sound units), which may ease the identification of smaller differences. The differences are even larger if results are compared with those by Cox *et al.*,[6] who found JND to be 0.92 ± 0.22 dB for the slow moving motif, and 0.44 ± 0.07 dB for the faster motif. This might also depend on the simulated impulse responses used in Refs. 6 and 7 which were based on a discrete number of early reflections (8 to 9) which is small compared to the much larger number (31) of early reflections used in Ref. 8 and to the continuous sequence of early diffuse reflections found in the measured IRs used to auralize the motifs in this study. As reported by Cox *et al.*[6] to justify the differences with previous studies by Reichardt,[5] the larger number of early reflections is likely to mask the changes in clarity, leading to an increase in JND. Finally, a further explanation of the differences might be the use in Ref. 6 of a theoretical formulation to calculate JND for clarity from values found for T_S, which might introduce some inaccuracies, at least in some cases.

In fact, according to the equation [Eq. (7) in Ref. 6], a constant JND for clarity would correspond to a variable JND for T_S, which is in agreement with the findings of the present paper. When T changes from 2 s to 6 s (and T_S proportionally changes from 130 ms to 378 ms) the resulting JND for T_S varies from 14 ms to 32 ms. However, according to the mentioned theoretical relationship the corresponding JND of C_{80} for the given C_{80} values should vary between 0.17 dB and 0.32 dB, which is clearly below the actual values. Similarly, a constant JND of 1.5 dB for C_{80} would correspond to a JND for T_S varying between 45 ms (for $T=2$ s) and 128 ms (for $T=6$ s). Such results suggest that crude application of the formula may not be recommended in rooms where the early reflection distribution is far from being ideal.[19] In fact, the application of a model specifically designed for churches[20] shows that the relationship between T_S and C_{80} is almost perfectly linear within each church (Fig. 8), with a slope which increases as the reverberation time grows, so that assuming a constant JND equal to 1.5 dB for C_{80} would give a theoretical JND for T_S varying from 20 ms (at $T=2$ s), to 27 ms (at $T=4$ s), and finally to 33 ms (at $T=6$ s). The resulting values do not perfectly coincide with the measured values, but a better agreement may be obtained assuming variable JND values for C_{80}.

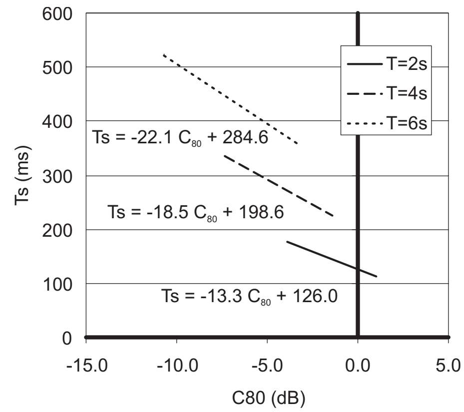

FIG. 8. Theoretical plots of T_S vs. C_{80} determined, according to the double slope model presented in Ref. 20, as a function of different reverberation times. Parameters for the theoretical model are chosen in agreement with the characteristics of the three churches used as a reference

The dependence between JND of T_S and the reverberation time was better explained by taking into account the reference T_S value for each base case (Fig. 9). In fact, a regression analysis between JND and T_S showed that the intercept of the regression equation was almost negligible, so that even when forcing the intercept to zero the variation in the statistical significance was small ($R^2=0.893$), leading to the conclusion that the ratio between the difference limen and the reference T_S (i.e., the Weber ratio in psychophysical terms) might reasonably be considered constant and equal to 0.084. This suggested that the JND for T_S should not be an absolute limen but it should be relative to the reference value. A simple validation of this approach was carried out

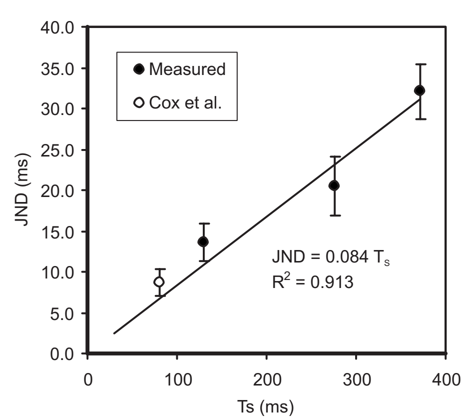

FIG. 9. Plot of JND of Ts vs corresponding reference values of the same parameter compared to the value resulting from Ref. 6. Error bars correspond to confidence intervals.

by considering that in Cox *et al.*[6] T_S was 80 ms and, according to the relative limen, this would give a JND of 7.2 ms, in reasonable agreement with the mean value they found (8.6 ± 1.6 ms). Unfortunately, further comparisons were not possible because T_S values are not given.

The relative variation of JND as a function of the reference T_S value introduces a further element of discussion related to the frequency dependence of this parameter. All the above considerations have been referred to multi-octave averages of reference parameters (mostly focused on mid-frequencies) and on wide band averages of modified acoustic parameters, but at higher frequencies the T_S value is generally lower and hence even the JND is likely to be more easily detectable, provided that the sound signal has a significant contribution in that spectral region. The observed (although not statistically significant) difference between the Gregorian chant and the cello theme might then also depend on the larger amount of high frequencies characterizing the second motif. In fact the listeners might have perceived the differences by focusing on the high frequency sounds of the staccato notes.

IV. CONCLUSIONS

The dependence of just noticeable difference for clarity measures (C_{80} and T_S) on reverberation time was investigated by means of listening tests based on measured B-format IRs auralized with two anechoic motifs (a Gregorian chant and a cello theme by Weber) and then presented on a 4-channel system. Three reference cases were investigated with reverberation times varying between 2 s, 4 s, and 6 s. The results were based on the responses of 13 non-expert listeners and showed that JND for C_{80} was substantially independent of music motif and of T, with an average value of 1.5 ± 0.1 dB. Conversely, the JND in T_S showed a clear dependence on T, while still being independent of the music motif. This behavior followed the complex relationship between T_S and C_{80} according to which when reverberation time increases the same variation in C_{80} determines an increasing change in T_S. The study of the correlation between JND in T_S and the corresponding parameter values for each base-case suggested that JND should be expressed in relative terms as 8.5% of the reference value. Even though the present research was mostly focused on long reverberant enclosures the comparison with previous studies dedicated to typical concert-hall conditions showed that despite some differences, possibly depending on the more realistic conditions (based on measured IRs) and on different material used to present the cases, the results may well complement each other. In fact, when dealing with T_S values typically found in concert halls (generally below 140 ms), the resulting JND should be about 10 ms, the value conventionally used in practice. However, when reverberation time is longer and T_S increases a 10 ms JND may be too restrictive and the relative difference limen may provide more accurate results. In addition, the relatively low number of consistent subjects (13 out of 40) suggests that the results found should be assumed as lower limits, as "ordinary people" would require larger differences both in C_{80} and T_S to perceive differences. Further investigations are nonetheless required in order to better investigate the influence of reverberation time on JND for clarity and to better clarify the influence of music motifs both in terms of frequency content and tempo, possibly involving larger samples of listeners.

[1] I. Bork, "A comparison of room simulation software—The 2nd round robin on room acoustical computer simulation," Acust. Acta Acust. **86**, 943–956 (2000).

[2] F. Martellotta, E. Cirillo, A. Carbonari, and P. Ricciardi, "Guidelines for acoustical measurements in churches," Appl. Acoust. **70**, 378–388 (2009).

[3] L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics* (Applied Science, London, 1982), Vol. **1**, pp. 503–509.

[4] T. I. Niaounakis and W. J. Davies, "Perception of reverberation time in small listening rooms," J. Audio Eng. Soc. **50**, 343–350 (2002).

[5] W. Reichardt and W. Schmidt, "The detectability of changes in sound field parameters for music," Acustica **18**, 275–282 (1967).

[6] T. J. Cox, W. J. Davies, and Y. W. Lam, "The sensitivity of listeners to early sound field changes in auditoria," Acustica **79**, 27–41 (1993).

[7] J. S. Bradley, R. Reich, and S. G. Norcross, "A just noticeable difference in C50 for speech," Appl. Acoust. **58**, 99–108 (1999).

[8] T. Okano, "Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections," J. Acoust. Soc. Am. **111**, 217–229 (2002).

[9] M. J. Ahearn, M. J. Schaeffler, R. D. Celmer, and M. C. Vigeant, "The just noticeable difference of the clarity index of music, C80," J. Acoust. Soc. Am. **126**, 2288 (2009).

[10] C. Guastavino and B. Katz, "Perceptual evaluation of multidimensional spatial audio reproduction," J. Acoust. Soc. Am. **116**, 1105–1115 (2004).

[11] E. Cirillo and F. Martellotta, *Worship, Acoustics, and Architecture* (Multi-science, Brentwood, 2006), pp. 51–200.

[12] ISO 3382, Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters. (ISO, Geneva, Switzerland, 1997).

[13] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," J. Audio Eng. Soc. **49**, 443–471 (2001).

[14] ITU-R BS 1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Geneva, Switzerland, 1994.

[15] F. Martellotta, "Subjective study of preferred listening conditions in Italian Catholic churches," J. Sound Vib. **317**, 378–399 (2008).

[16] Y. Ando, *Architectural Acoustics* (Springer, New York, 1998), pp. 7–19.

[17] W. H. Ehrenstein and A. Ehrenstein, "Psychophhysical methods," in *Modern Techniques in Neuroscience Research*, edited by U. Windhorst and H. Johansson, (Springer, Berlin, New York, 1999), pp. 1214–1229.

[18] http://lab5.ta.chalmers.se/~vincent/doc/lise/lise.html (Last viewed 5/26/2010).

[19] E. Cirillo and F. Martellotta, "Sound propagation and energy relations in churches," J. Acoust. Soc. Am. **118**, 232–248 (2005).

[20] F. Martellotta, "A multi-rate decay model to predict energy-based acoustic parameters in churches," J. Acoust. Soc. Am. **125**, 1281–1284 (2009).