Available online at www.sciencedirect.com

ScienceDirect

Journal of Sound and Vibration 317 (2008) 378–399

JOURNAL OF SOUND AND VIBRATION

www.elsevier.com/locate/jsvi

Subjective study of preferred listening conditions in Italian Catholic churches

Francesco Martellotta*

Dipartimento di Architettura e Urbanistica, Politecnico di Bari, via Orabona 4, I-70125, Bari, Italy

Received 20 March 2007; received in revised form 14 February 2008; accepted 10 March 2008
Handling Editor: C.L. Morfey
Available online 5 May 2008

Abstract

The paper describes the results of research aimed at investigating the preferred subjective listening conditions inside churches. The effect of different musical motifs (spanning Gregorian chants to symphonic music) was investigated and regression analysis was performed in order to point out the relationship between subjective ratings and acoustical parameters. In order to present realistic listening conditions to the subjects a small subset of nine churches was selected among a larger set of acoustic data collected in several Italian churches during a widespread on-site survey. The subset represented different architectural styles and shapes, and was characterized by average listening conditions. For each church a single source–receiver combination with fixed relative positions was chosen. Measured binaural impulse responses were cross-talk cancelled and then convolved with five anechoic motifs. Paired comparisons were finally performed, asking a trained panel of subjects their preference. Factor analysis pointed out a substantially common underlying pattern characterizing subjective responses. The results show that preferred listening conditions vary as a function of the musical motif, depending on early decay time for choral music and on a combination of initial time delay and lateral energy for instrumental music.

1. Introduction

Churches are acoustically complex places. They are complex from the geometric point of view, because their dimensions and shapes may be extremely varied, passing from simple auditorium-like churches to complicated baroque churches characterized by curved walls, side chapels, vaults and domes which often act as partially coupled volumes. The range of the possible sound messages is also extremely varied passing, during the same liturgical service, from the spoken word to organ music and congregational singing. In addition, churches are the natural places where sacred works, both choral and instrumental, should be performed. Finally, it should be pointed out that where adequate performance spaces are lacking (and in Italy this happens frequently), churches often host orchestral music concerts.

The Second Vatican Council [1] modified the Catholic liturgy, giving new importance to the teaching role of the celebration, emphasizing the need to understand the spoken word. This fact, combined with the

---

*Tel.: +39 0805963631; fax: +39 0805963701.
E-mail address: f.martellotta@poliba.it

0022-460X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2008.03.014

availability of new technical means and with the reverberant nature of most churches led to a widespread use of electro-acoustic systems to convey the word to the assembly of the faithful. Nowadays nobody would think of a church without a speech reinforcement system and this is an invaluable aid for the acoustician because in this way the natural acoustics may respect the "sacredness" of the place, optimizing at the same time the needs of music and singing. In fact, the reverberant acoustics of churches is well suited to organ music and Gregorian chants. Congregational chants also benefit from reverberation because it increases the sense of participation encouraging other singers to join in.

More critical listening conditions are found when orchestral music is played inside churches. In fact, according to well-established literature about concert hall acoustics [2,3], churches are more reverberant than the ideal. However, it is not unusual for both musicians and listeners to report positive comments, suggesting the investigation in greater detail of the preferred listening conditions inside this particular group of buildings.

Subjective tests, involving a statistically significant number of people, should be used in order to improve the knowledge of the complex relations between geometry, acoustics, and subjective preferences. Several examples can be found in the literature, mostly related to concert hall or opera house acoustics. Some studies are based on indirect comparisons made by collecting experts' judgments of given halls [3], others were carried out using an on site listening test [4,5], others using simulated sound fields [6], others recording live or reproduced music in real concert halls and later playing it back using earphones or loudspeakers with cross-talk cancellation [7,8]. For churches, the number of investigations is much lower. Carvalho [9] employed a group of subjects who attended live music performances in different churches. Desarnaulds [10] and Meyer [11] collected the judgments given by members of several parish communities on the acoustics of their churches, obtaining curves of optimal reverberation times as a function of room volume.

Comparisons between different places (even when they are made using the same group of listeners and the same musicians) are arduous because of the short-term nature of the acoustic memory which prevents listeners from transposing themselves from one place to another. In addition, live performances may be strongly influenced by the manner in which the musicians play. In fact, they may (even unconsciously) adapt to the particular acoustics of a room. On the other hand, laboratory methods based on loudspeaker reproduction of anechoic material recorded in real rooms, or convolved with measured impulse responses, allow nearly instantaneous comparisons of different acoustic conditions avoiding any dependence on the performers' style, but have some drawbacks in terms of the realism of the presentation. In fact, as will be better explained below, the simulation of the directivity and extent of real sound sources by means of electro acoustic sources and the accuracy of the playback system reproducing the actual sound field are the main concerns. However, careful arrangement of the experiment may minimize these problems, providing an acceptable compromise between scientific accuracy and realism.

This paper presents the results of research which used the laboratory method in order to understand the relations between subjective preference and the architectural and acoustic aspects observed in a group of Italian Catholic churches also taking into account the effect of different musical motifs.

2. The on site survey

A systematic survey [12] of the acoustic conditions of more than forty Italian Catholic churches was started in 2001. The churches differed in style (from Early-Christian to Modern), in size (from 2000 to 160 000 m^3), in typology (basilica plan, cross plan, central plan, and so on), and obviously in acoustics (reverberation time varying from 2 to 12 s).

All the measurements were carried out complying with the ISO 3382 standard [13]. An omni-directional sound source made of 12 120 mm loudspeakers (with a frequency response from 100 Hz to 16 kHz) mounted on a dodecahedron, together with an additional sub-woofer to cover the frequencies below 100 Hz, was used. High-quality impulse responses were collected by using a B-format microphone (Soundifield Mk-V) and a binaural head and torso (B&K 4100D). The signal used to excite the rooms was a constant envelope equalized sine sweep generated with MATLAB according to Müller and Massarani [14] so that the spectrum of the radiated sound was substantially flat from the 50 to 16 kHz third-octave bands. The room responses were recorded at a sampling rate of 48 kHz and 24 bit depth, to obtain, after deconvolution (performed again with MATLAB), impulse responses with very low noise (the signal to noise ratio was generally higher than 60 dB even at the lowest frequencies).

In each church several source–receiver combinations were analysed, with a minimum of two sources (placed in front and on the side of the altar) and an average of nine receivers placed in one half of the floor if the church was symmetrical (Fig. 1), otherwise they were spread out to cover the whole floor area uniformly. The source and the microphones were 1.5 and 1.2 m from the floor surface, respectively. The B-format microphone pointed with the *X*-axis toward the sound source, while the binaural head was placed on the seat facing the altar (with no head rotation).

The whole set of IRs collected in this way was later used to calculate the most important acoustical parameters according to the ISO 3382 standard [13], providing a large amount of data summarized in Ref. [12]. For the purposes of the present research only one source–receiver combination per church was chosen. This could be considered a rough approximation because of the point-to-point variation in the acoustic parameters inside a church and because a single sound source, although omni-directional, can hardly represent the actual directivity and extent of multiple real sources.

The first problem was addressed by taking into account a receiver located in the main nave at about one third of the distance between the source and the back wall (receiver 3 in Fig. 1). In this way all the energy-based acoustic parameters that, as shown in Refs. [12,15], are strongly dependent on source–receiver distance and on early to reverberant ratios, were referred to a position which varied according to the specific geometry, reasonably representing the ‘‘average’’ conditions to which most of the listeners are exposed. In fact, taking into account the just noticeable differences (JNDs) for each parameter (i.e. the smallest change that can be detected by a listener) and the logarithmic variation of the parameters as a function of the distance (with large variations close to the source and much smaller variations when the distance grows), the area in which a listener is substantially unable to detect acoustic differences covers a large part of the main listening area. Conversely, spaciousness parameters, such as LF (lateral fraction) and IACC, depend on both the source receiver distance and on the church width (and on the actual distance from reflecting walls), therefore in those cases where the main nave was wider (so that acoustic parameters varied significantly from the centre to the sides), an additional receiver was considered in a position near the side wall and at the same distance from the source (receiver 4 in Fig. 1).

The location of the sound source in front of the altar and on the symmetry axis was chosen in order to provide the best listening conditions for the audience (as in many cases choirs and organs are located in hidden positions). The use of a single point source was certainly a rough approximation for symphonic music because people expect to be able to locate different instruments, while for choir and organ this was a minor problem because their dimensions are generally smaller. However, it should be taken into account that currently available anechoic materials are, at most, stereophonic and, in this case, the spaciousness of the actual source might be correctly rendered only at a fixed position, which would hardly coincide with actual listening positions (that are generally far from the source and certainly not on the symmetry axis). Consequently, it was concluded that even using two sound sources the improvement in realism would have been marginal, so the simpler solution of a single sound source was finally preferred.

A representative sub-sample of the whole set of available churches was chosen in order to carry out listening tests. First, the places with too extreme acoustic conditions (i.e. very long reverberation time) were excluded

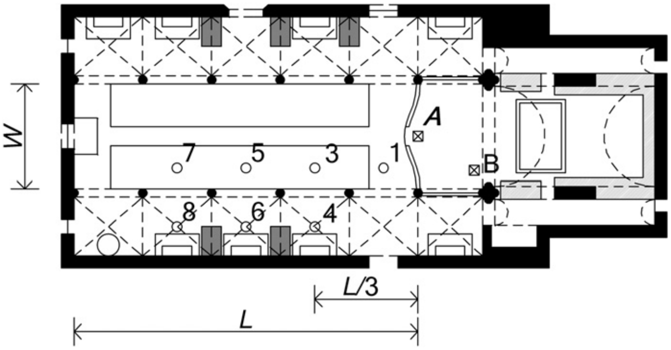

Fig. 1. Typical layout of source and receiver placement in churches.

from the study as being scarcely significant in a listening task. Then the similarities between churches built according to the same architectural style and having a comparable shape allowed to take into account only nine churches representative of different styles, typologies, and dimensions. In three cases an additional receiver was thought to be necessary to represent the variations in the lateral reflected energy, so 12 different IRs were originally considered. However, as the paired comparison procedure proved to be very time consuming it was decided to eliminate two of the additional receivers and only 10 IRs were finally considered. A summary of the main geometric and acoustic parameters corresponding to these IRs is reported in Table 1.

It should be underlined that all the measurements were carried out in unoccupied conditions which means that the acoustic conditions presented to the listeners might be more or less different from the actual conditions which would be observed during a service or a concert. In most cases pews cover only a small part of the floor (and consequently of the whole exposed surface), therefore small variations should be observed. However, this has a marginal influence on the results of the present study because the preferences expressed by subjects were correlated to the values of the actual acoustic parameters to which they were exposed. Consequently, any preference eventually found for given values of a parameter should be simply applied to occupied conditions.

3. The listening test

3.1. Playback method

During the survey both B-format and binaural IRs were collected, so different techniques of sound field rendering were available. However, at this stage of the research only transaural presentation of binaural signals was employed. This technique was a good compromise because, by only using two channels, it allowed a quite accurate reproduction of the sound field (avoiding most of the drawbacks of the headphone presentation such as in-head localization and front-back confusion, provided that cross-talk cancellation filters are correctly calculated). Two closely spaced loudspeakers, spanning an angle of 10° according to the so-called "stereo dipole" configuration [16], were used in order to obtain a virtual image of the surveyed churches. Although this arrangement is sometimes criticized for its less than optimal separation of channels [17,18], it is capable of ensuring a robust performance with respect to head movement (that cannot be avoided during the listening test if good naturalness is to be achieved) and of providing accurate localization for target azimuths ahead of the listener, even though back-to-front errors are sometimes observed [19,20]. However, the latter problem was considered a minor concern in the present research because for the selected source–receiver combinations the direct sound and the most important reflections always came from the front.

Table 1
Summary of the acoustical parameters measured in the selected sample of IRs

ID	S–R combin.	Style	Volume (m^3)	Width (m)	S–R distance (m)	Δt_i (ms)	T30 (500–1 k) (s)	EDT (500–1 k) (s)	BR*	C80 (500–2 k) (dB)	Ts (500–1 k) (ms)	LF (500–2 k) (%)	1-IACC$_E$ (500–2 k) (%)
A	A03	Rm	20 000	22	9.5	20	5.4	5.6	1.09	–4.2	342	31.1	79.0
B	A05	Rm	10 500	15	11.9	17	2.1	1.8	1.10	0.3	130	24.9	57.3
C	A04	Gt	33 100	26	18.1	16	5.7	6.3	0.96	–9.2	489	28.9	76.0
D	A02	Rn	19 000	32	10.8	9	8.9	8.4	1.26	–7.6	632	16.1	66.3
E	A04	Rn	39 000	36	20.6	25	5.1	5.4	1.07	–7.4	389	18.9	48.0
F	A03	Ba	8700	30	8.7	45	3.3	3.2	1.13	–4.2	240	33.3	67.0
G	A01	Ba	16 400	34	9.9	34	7.2	7.5	1.03	–5.4	500	15.2	38.3
H	A03	Mo	5500	16	11.2	15	6.3	6.3	1.07	–5.9	437	40.3	68.3
I1	A03	Mo	9000	25	12.5	27	4.4	4.5	1.12	–4.4	316	24.6	48.7
I2	A04				18	5	4.4	4.7	1.12	–5.4	352	30.9	82.3

Style abbreviations: Rm—romanesque, Gt—gothic, Rn—renaissance, Ba—baroque, Mo—modern.

*Bass ratio, defined as ($T30_{125–250}/T30_{500–1\,k}$).

The signals radiated by the loudspeakers were cross-talk cancelled in order to remove the part of the sound that reaches the right ear from the left loudspeaker and vice versa. The cross-talk cancellation was performed through convolution of the two binaural signals with a set of four inverse filters. The latter were obtained by inverting the IRs measured in the listening room using a frequency-domain deconvolution method with regularization to prevent excessive boost at the extremes of the frequency range [21].

In order to assess the performance of the playback system the amount of cross-talk cancellation was measured using a 10 s stereo signal whose left channel was white noise and whose right channel was silence. These signals were convolved with the four cross-talk cancellation filters and played-back through the loudspeakers and simultaneously recorded by the dummy head and torso. The amount of cross-talk cancellation was defined as the difference between the left and right power spectra. It can be observed (Fig. 2) that the spectrum at the left ear is relatively flat, while the spectrum at the right ear is far from perfect silence. However, the cancellation within the interval from 200 to 8 kHz averaged over five positions is about 17 dB with maximum values above 30 dB and minimum values which in a couple of cases equal 6 dB, these values appear in agreement with those reported in Refs. [17,18]. In particular, when the manikin is exactly in the position in which the filters were calculated the cancellation is 20.5 dB, when it is moved 50 mm sideward (independent of the direction) the cancellation degrades by 6 dB, while a fore-and-aft displacement of the same magnitude degrades the cancellation by only 3 dB. With respect to head rotations it was measured that a 7.5° rotation towards the left degraded the cancellation by 5 dB, while a 15° rotation in the same direction degraded the cancellation by 9 dB. Rotations in the opposite direction showed negligible attenuations as a consequence of the masking effect of the head.

A further check of the accuracy of the system consisted of the comparison of the reproduced acoustical parameters with those measured in the nine churches. The binaural impulse responses measured in the chosen position in each church were convolved with a logarithmic sweep, cross-talk cancelled, and finally played back through the system. The resulting signals measured at the two ears of the dummy head were processed to calculate monaural acoustic parameters (with reference to the left ear only) and binaural parameters. Monaural parameters included reverberation times (T30, T20, T10, and EDT) and energetic ratios (C80, C50, D50, and Ts), while IACC was used as a binaural measure. In addition, as the sound localization is strongly dependent on low frequency (below 1500 Hz) inter-aural time difference (ITD) [22] this parameter was included in the analysis in order to provide a measure of the accuracy of the system in reproducing localization cues. ITD values were calculated after low-pass filtering the binaural impulse responses with a cut-off frequency of 1500 Hz, while for the other parameters the differences between measured and reproduced values were calculated for the six octave bands from 125 to 4000 Hz and averaged over the octave bands. In all the cases the differences were expressed in terms of JNDs. The currently adopted values of JND are summarized in Table 2. For ITD the JND is about 10 μs [23] which corresponds to an angular resolution of 1.2°. However,

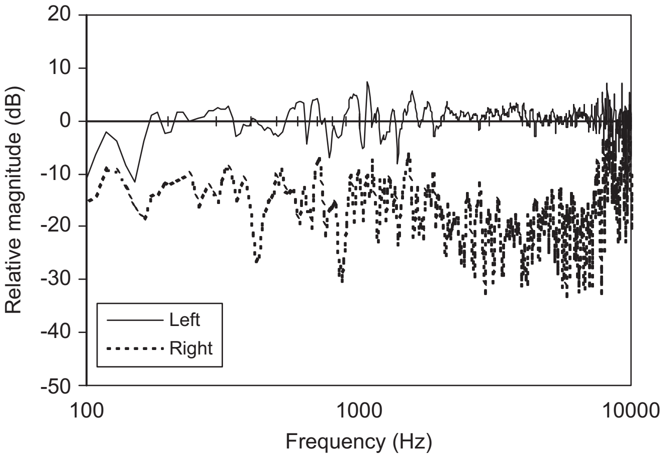

Fig. 2. Magnitude spectra of the signals at the microphones of the dummy head (averaged in five positions) when the playback system radiates white noise (10 kHz low-pass filtered) to the left ear and silence to the right ear.

Table 2
Summary of the JND values for the selected acoustic parameters

Parameters		JND
Reverberation	T30, T20, EDT	5%
Energy fractions	C80, C50	1 dB
	D50	5%
	Ts	10 ms
Spaciousness	IACC	0.08
Localization	ITD	0.08 ms

Table 3
Individual values of the mean differences expressed as a function of JND for the three groups of acoustic parameters

	Reverberation	Energy ratios	IACC	ITD
A	0.32	0.58	0.97	0.91
B	0.55	0.73	0.67	0.23
C	0.35	0.53	0.95	0.19
D	0.50	0.67	0.75	0.86
E	0.37	0.42	0.48	0.00
F	0.48	0.28	0.46	0.00
G	0.38	0.81	0.67	0.23
H	0.34	0.54	0.86	0.82
I1	0.50	0.80	0.73	0.00
I2	0.48	0.65	0.78	0.23
Mean	0.43	0.60	0.73	0.35
Std. dev	0.08	0.17	0.17	0.37

for the purposes of the present research an accuracy of 10° was considered acceptable, and consequently the JND for ITD was assumed to be 80 μs. The errors were finally averaged over the four groups (reverberation, energy ratios, IACC, ITD) and the results, reported in Table 3, suggest that all the parameters are reproduced quite accurately, with reverberation and energy ratios showing mean differences of 0.43 and 0.60 JNDs, respectively, while spaciousness shows slightly higher errors, with a mean value of 0.73 JND. The ITD shows an average error of 0.35 JND (corresponding to an angular error of about 4°), mostly due to the almost perfect match observed in seven cases out of ten. None of the individual values exceeded the JND limit, even though in two cases (namely churches A and C) the mean IACC error was close to unity.

3.2. The source signal

Five source signals were used during the listening test. Two were voice-only excerpts taken from a Gregorian chant and a modern polyphonic choral song. The instrumental pieces included an organ composition and two symphonic pieces taken from the Classical and Romantic music repertoire.

The Gregorian motif was a 22 s excerpt taken from the "Pange Lingua" hymn (attributed to St. Thomas Aquinas) sung in Phrygian mode and recorded in practically anechoic conditions. The excerpt is characterized by voice modulation and by the typical phrasing of the hymn. As it is a voice-only music motif, its spectrum (Fig. 3) is unevenly distributed and is mostly rich in mid frequencies with a notable contribution from low frequencies.

The polyphonic choral motif was a 27 s excerpt from Randall Thompson's "Alleluia" composed in 1940, taken from the DVD "Anechoic Choral Recordings" distributed by Wenger. This work was composed for a four part chorus of unaccompanied voices and the selected excerpt is characterized by a slow tempo and an initial contrapuntal part with a series of crescendos and diminuendos which terminates with a forte passage.

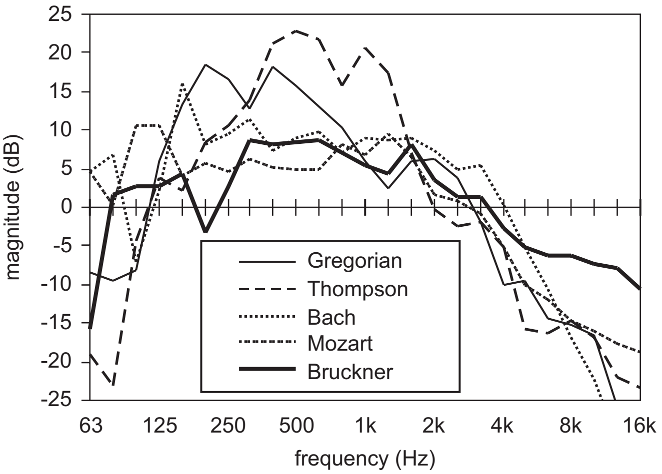

Fig. 3. One-third octave band spectra of the anechoic motifs used in the test.

The spectral content (Fig. 3) is quite different from the Gregorian piece, showing a lesser content of low frequencies.

The organ motif was a 21 s excerpt taken from J.S. Bach's "Fantasia in G minor" (BWV 542) played with the Hauptwerk pipe organ simulator without reverberation. The composition is characterized by the use of chromaticism and frequent modulation. The selected excerpt is characterized by a double sequence of a melodic part (typical of the fugue) followed by short chords covering several octaves and providing a significant contribution of sound energy from 125 to 4000 Hz (Fig. 3).

Both the symphonic motifs were taken from the CD "Anechoic orchestral music recording" distributed by Denon. The Classical motif was a 24 s excerpt of the 2nd theme in A+ of the Overture to "Le Nozze di Figaro" by W.A. Mozart. It is a fast passage (the tempo is *presto*) which dynamically evolves from *piano* to *forte* with a sequence of chords played *forte* by all the instruments and followed by a reply from the strings. The arrangement of the instruments covers a wide range of frequencies and provides an almost flat spectrum from 50 Hz to 5 kHz (Fig. 3).

The Romantic motif was a 28 s excerpt from the first movement of Symphony No. 4 in E-flat major "Romantic" by Anton Bruckner. The selected excerpt is well balanced in spectral terms (Fig. 3), it is characterized by a slow initial part that grows in crescendo from *pianissimo* to *fortissimo*, concluding with a fast and rhythmic musical passage played by the whole orchestra.

After the convolution with anechoic motifs and the cross-talk cancellation, all the recordings were adjusted to give an average level of about 75 dB(A) during the presentation. This decision was made because it is known from the literature [8] (and was confirmed during a preliminary test) that listening level may strongly influence subjective preference and this could prevent listeners from giving the proper emphasis to other aspects of the sound field.

The dependence of subjective preference on music motif has been extensively studied by Ando [24], who found a dependence of the preferred acoustic parameters on the long-time, τ_e, and the short-time, $(\tau_e)_{\min}$, duration of the auto-correlation function calculated for each music motif. Even though Ando's work is based on purely synthesized sound fields providing carefully studied combinations of acoustical attributes, its conclusions can be conveniently used as an interesting reference to better understand the results of the present research. Consequently, both τ_e and $(\tau_e)_{\min}$ were calculated for the selected motifs according to the procedure specified in Ref. [24]. The results, reported in Table 4, show significant differences between the two vocal motifs, with Thompson's excerpt characterized, as expected, by a short $(\tau_e)_{\min}$ and Gregorian chant having a much longer value, comparable with that of the Classical symphonic excerpt. The longest value is shown by organ music, while the Romantic symphony shows interesting differences between the long-time and the short-time duration which reflect the differences between the long initial crescendo (and legato) and the fast final passage.

Table 4
Long-time and short-time effective duration of the auto-correlation function calculated for the five music motifs

Music motif	τ_e	$(\tau_e)_{\min}$
Gregorian chant	91	59
Thomas' "Alleluia"	26	22
Bach's "Fantasia in G Minor"	–	164
Mozart's Overture from "Le nozze di Figaro"	70	69
Bruckner Romantic Symphony	134	80

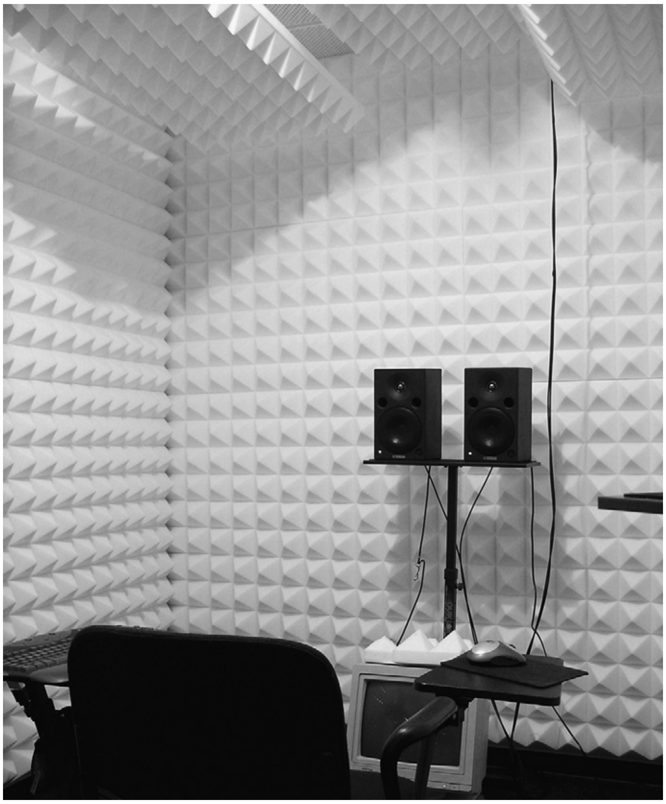

Fig. 4. The listening room.

3.3. The listening room

The listening room (Fig. 4) was designed in order 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 [25] was followed when possible, given that the scope of the present room is rather different from the "reference listening room" described in the norm. The room has a flat frequency response and a reverberation time of 0.09 s at medium frequencies, decreasing to 0.05 s at 8 kHz. Below 250 Hz the reverberation time gradually grows up to 0.35 s at 63 Hz. The room is acoustically insulated, using a floated construction, in order to minimize external noise.

The listening room has a nearly rectangular plan, with internal dimensions of 3.70 × 2.50 × 2.40 m. The floor area is about one third of the area suggested by ITU recommendations for multichannel reproduction, but the room is designed to receive one listener at a time, so the smaller dimensions are acceptable.

The two loudspeakers (Yamaha MSP5) have a flat frequency response from 60 Hz to 30 kHz and are placed on a stand in front of the wall with a span of 10° as seen by the listener. The level of the speakers was carefully aligned at the centre of the listener position.

3.4. The subjects

The listening tests were performed by a total of 143 persons aged between 20 and 60. Some of the subjects listened to more than one music motif, therefore the number of participants in each session is reported in Table 5. None of them had experience in hearing artificial head recordings, but they all had experience in critical listening of live or recorded musical performances. Most of the subjects (about 60 percent) have also played an instrument, sung in a choir, or studied music. None of the subjects reported any hearing problems.

This sample of listeners was the result of pre-screening which was carried out during the training phase. In fact, in agreement with ITU-R 1116-1 [25] all the listeners were trained before the test. The training phase consisted of a first check of the cross-talk system with pink noise sent first to the left ear and then to the right one and finally to both (giving a perfectly centred monophonic sound image). The listeners were invited to find their best listening position as the most comfortable spot where they were able to perfectly localize the sounds, alerting them to avoid head movements which could cause sudden changes of the sound localization. Later on they carried out a more accurate localization test in which they were exposed to six sounds coming from different directions on the horizontal plane located at the following angles from the median plane: $-105°$, $-45°$, $-15°$, $+30°$, $+60°$, $+120°$. Positions in the back were given less relevance because, as stated above, in the surveyed churches both the direct sound and the early reflections mostly come from the front. Each sound was a sequence made of an introductory male voice followed by notes of different pitch originally recorded through the binaural head and processed in the cross-talk cancellation system. The 58 subjects (out of 201) who failed to localize all the sound directions with an accuracy of $\pm 15°$ on the azimuthal plane completed the training procedure but were not included in the listening test. Most of the excluded subjects (Fig. 5) localized back sources on the front, pointing out, as reported in Ref. [19], a significant difference between the head related transfer function (HRTF) of the subject and that of the dummy head used to calculate cross-talk filters, which could consequently cause incorrect rendering of the sound field in spatial terms.

After this phase the listeners were exposed to all the material that they had to grade later in the test. This was done using a computer interface similar to that used during the test described in detail below. In this way they were able to interact with the interface, practicing using it and, at the same time, listen to all the sounds in a series of paired comparisons. Throughout this phase the subjects could freely interact with the trainer in order to convey their sensations and discuss the differences they were able to detect.

All the subjects were informed that they had to listen to sound fields reproducing the acoustic conditions of different churches, consequently they were strongly invited to focus their attention on the effect of the room rather than on purely musical details.

3.5. The listening test

The grading of the different listening conditions was obtained by means of paired comparisons of the recordings. In fact, the differences between stimuli were often small and a preference for one or another could not rely on long- and medium-term memory, so the listeners had to rely exclusively on short-term memory.

Table 5
Number of participants as a function of the music motif

Music motif	Initial participants	Consistent participants
Gregorian chant	55	50
Thomas' "Alleluia"	36	29
Bach's "Fantasia in G Minor"	34	29
Mozart's Overture from "Le nozze di Figaro"	40	36
Bruckner Romantic Symphony	65	52

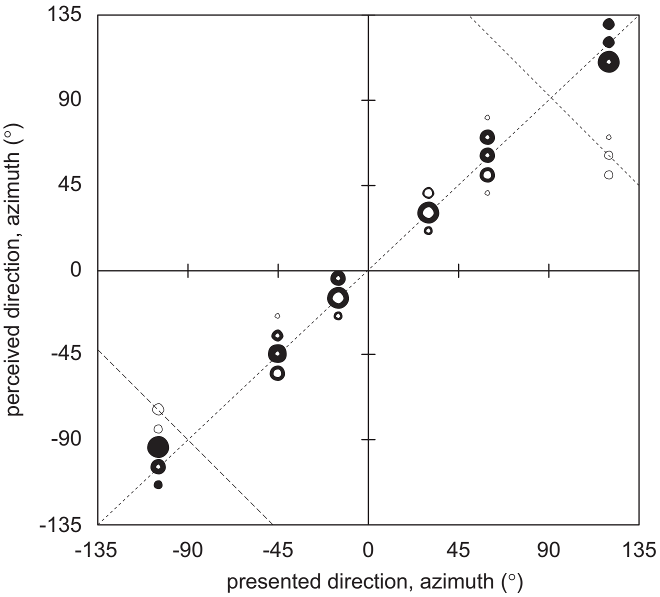

Fig. 5. Results of the localization test. Circle dimension is proportional to the number of responses. Black circles correspond to selected subjects. White circles correspond to excluded subjects.

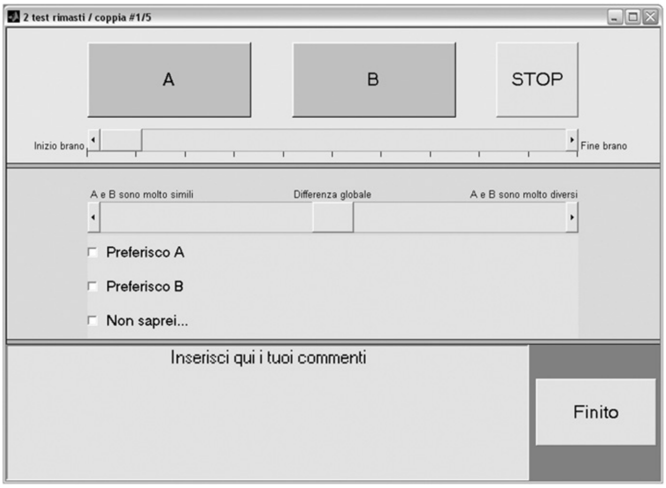

Fig. 6. Graphical interface used to carry out paired comparisons.

In order to make comparisons easier, and to make effective usage of the short-term memory a near-instantaneous switching between the stimuli was allowed with a short fade-down/change-over/fade-up to avoid possible artefacts due to the sudden change.

The whole listening test was controlled by the subject using a graphical interface developed in MATLAB (Fig. 6). The interface was based on the LISE environment developed by Rioux [26], modified in order to allow the listener to switch between signal A and signal B at any moment, to stop the sound, and to choose the exact point to listen to. The stimuli were presented according to a random sequence without repetitions. However, at the end of the listening session five comparisons were repeated without informing the subjects in order to check the consistency of their judgment criteria and, consequently, their reliability.

For each comparison the subjects had to express their preference for one of the signals. A third choice (“I don’t know…”) was given the subjects for those cases in which the stimuli were too similar or both annoying (or preferred). In this way the listeners were not forced to make a choice which might result in a random decision, considerably reducing circular errors which takes place when a listener expresses his preference for sound A over sound B, sound B over sound C, and also sound C over sound A [27]. Finally, the subjects were left free to type their comments (if any) about their sensations and about the criteria they used to make their decision. A control check prevented them from skipping a paired comparison without expressing their judgment.

After each comparison the preferred signal received a "+1" score, while for the other signal it was "−1". For no preference, both stimuli were given a score of zero. The preference scores for each signal were finally summed to obtain the ranking.

3.6. *Post screening of the subjects*

Once the subjects had completed the test, a post-screening was made in order to check their reliability. Temporal consistency of the judgment and circular error rate (CER) were assumed as criteria to discriminate unreliable listeners.

The temporal consistency of the judgment was evaluated by comparing the responses given by the subjects at the beginning of the test with those given to the five comparisons repeated at the end of the session. The five pairs were selected in order to compare cases significantly dissimilar and allow the listeners to make an easy choice, so that a change in preference could only be attributed to an unstable (or random) judgment criterion. The consistency of a listener was therefore given by the number of correct matches between the ordinary test and the corresponding five duplicates. Listeners with less than four correct matches were considered as inconsistent and their results discarded from the final analysis.

Circular errors may originate from an actual similarity between the stimuli which makes evaluation difficult for the listener, but the third choice was included in the test to prevent this. So when a circular error takes place it may point out a real inaccuracy of the listener or an alteration of the assessment criteria. Since circular errors cannot be completely eliminated even with the introduction of the third choice, their number was reasonably assumed as a measure of the subject's consistency. The CER was calculated as the ratio of the circular errors made by a given subject to the total number of triads given by the number of combinations of all the stimuli taken three at a time. Listeners with a CER greater than 20 percent were considered to be inconsistent and their results discarded.

The final number of consistent subjects (on average 84 percent of the total), counting only those satisfying both criteria, is reported for each music motif under analysis in Table 5.

4. Results

4.1. *Method of analysis*

The results of each listening test were assembled in a preference matrix whose entries indicated how many times each church was preferred by each listener. Possible rankings varied between +9 and −9. The total ratings for each music motif are reported in Fig. 7.

The matrix of the subjective ratings was subjected to a linear factor analysis based on the principal component analysis (PCA) extraction method [28]. Factor analysis is frequently used in psychology and in subjective preference studies [7,8,29] because it allows to explain whether observed variables (in this case subjective preference ratings), depending on several parameters, may be explained by a smaller number of hidden, or latent, factors common to all the variables. Without entering in the mathematical details of the method (well described in Ref. [7] and handled using SPSS software), once the N common factors are defined, each subject is identified in the space of factors by a set of N coordinates which are named "factor scores". The position of each subject in the factor space allows to understand in a simple and straightforward way the individual differences and the attitude towards the factors. In fact, if all the subjects are influenced in the same way by only one factor they will cluster towards the positive direction of that factor axis. Conversely, clustering of subjects on opposite direction of a given factor indicates two groups with contrasting attitude towards that factor. However, even though the analysis of factor scores may provide a quick understanding of

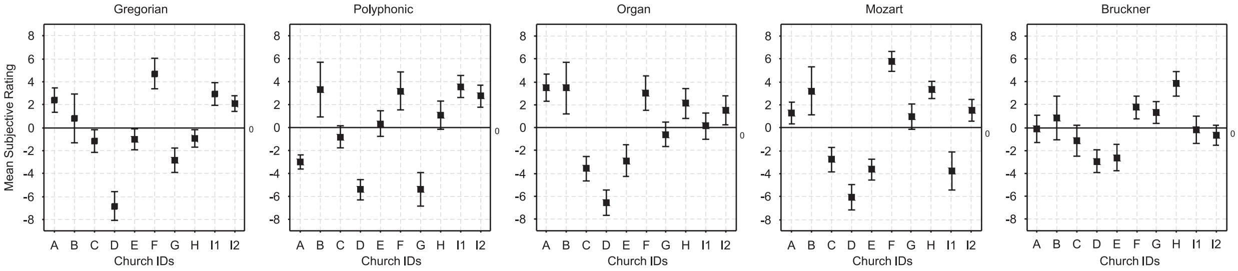

Fig. 7. Average subjective ratings for the five motifs under investigation. Error bars indicate 95 percent confidence intervals.

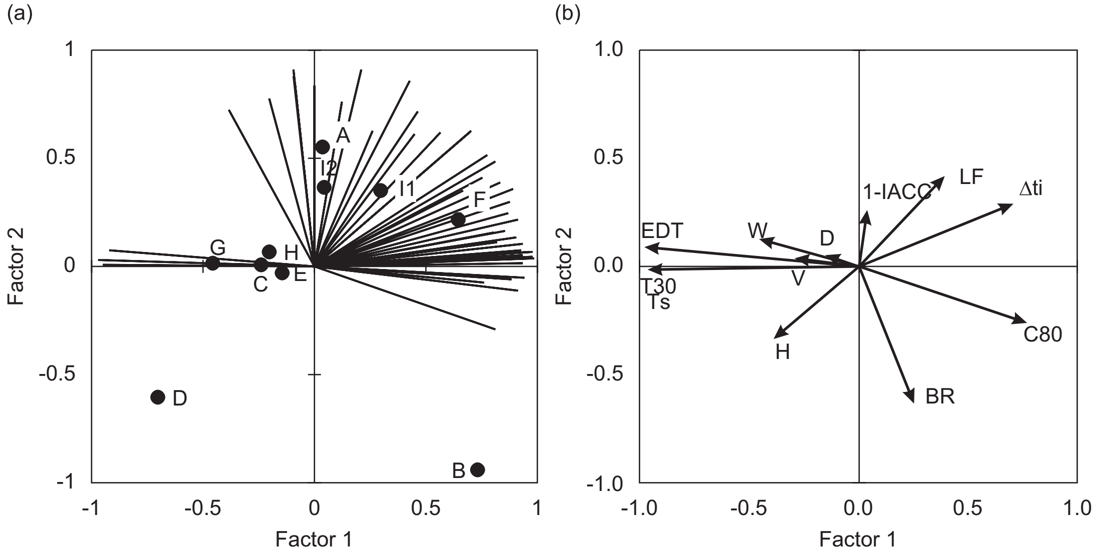

Fig. 8. (a) Preference plane for Gregorian chant. Capital letters indicate the listening conditions and vectors indicate the listeners. Projection of hall position on vectors approximate listeners individual preferences. (b) Correlations between objective parameters measured in the nine churches and subjective factors.

subjective behaviour, the main usefulness of factor analysis stays in the idea that each subjective preference may be expressed as the linear combination of the common factor scores. The combination coefficients (named "factor loadings") are calculated to reproduce best the observed variable from the factor scores and are the same for all the subjects. The factor loadings provide the basis for attempting to interpret the nature of the common factors. In fact, high loadings represent ratings that are influenced strongly by the factor, while low loadings represent ratings that are influenced weakly by the factor. The final step is the analysis of the correlations between factor loadings and objective acoustic measures corresponding to each case, which allows to define the actual dependence of subjective preference on acoustic parameters.

To aid interpretation, the results of factor analysis are represented in geometric form. In the present case, where the task is one of assigning subjective preferences, the factor space is also called a *preference space* (Fig. 8a). The full circles with capital letters represent the 10 different listening conditions included in the test and the vectors represent the different listeners. On the preference plane the projection of the circles on one of the axes represents the corresponding factor loading, and the projection of the listener's vector represents the factor score (i.e. the weight given to that factor in the judgment), while the projection of the point representing the different listening conditions on a listener's vector approximately represents the rating received by the respective subject. In some cases, in order to ease the interpretation of the results, a rotation of the axes may be applied. For the sake of clarity the correlations between objective parameters and extracted factors are graphically represented on a different factor plane (Fig. 8b) as vectors whose projection on the axis measures the respective correlation coefficient. So, the acoustic parameters with the highest correlation with one factor (and preferably low with the others) are the most suitable to quantify in objective terms (and hence predict) subjective preference.

Multioctave-band averages of the most important acoustic parameters calculated under the chosen listening conditions and a selection of geometrical parameters (Table 1) were used to investigate the correlation with the factors. The acoustical parameters included reverberation time T30, early decay time EDT, and centre time Ts calculated from 500 to 1000 Hz, clarity C80, LF, and inter-aural cross-correlation coefficient (1–IACC) calculated from 500 to 2000 Hz, and finally the bass ratio BR which is a single-number parameter defined as the ratio of the average T30 at 125 and 250 Hz to the average T30 at 500 and 1000 Hz, and the initial time delay Δt_i defined as the time delay between the strongest reflection and the direct sound. Other acoustic parameters, or other multioctave-band averages were also calculated but were not taken into account because they were correlated with the above-mentioned descriptors (i.e. they were not independent) and showed lower correlations with subjective responses. The geometric parameters taken into account were the room volume V, the average width of the main nave W, the average height of the nave H, and the source–receiver distance D. All the parameters were calculated from three-dimensional CAD models of the churches based on on-site measurements.

Finally, the average ratings (normalized in order to vary from -100 to $+100$) relative to each motif were subjected to linear regression analyses. Where linear models were not satisfactory and where the factor analysis pointed out a dependence on multiple parameters, general linear (and in one case quadratic) models were calculated. The procedure for general linear models was based on forward stepwise modelling with an α to enter equal to 0.10. The reliability of the models was assessed by means of the determination coefficient R^2 and of the result of the F-test at a 95 percent confidence level. The correlation was significant when the residual probability p associated to F was smaller than 0.05.

4.2. Gregorian chant

The analysis extracted two significant factors explaining 59 and 20 percent (56 and 23 percent after rotation) of the total variance. The results of the factor analysis reported on the *preference space* (Fig. 8a) showed that most of the listeners gave a positive weight to both factors, with only a couple of subjects giving a strong negative weight to factor 1. The churches receiving the highest ratings were F, I, and A. The worst were churches D and G (which have the longest reverberation times).

The analysis of the correlations between the factors and the objective parameters (Fig. 8b) shows that factor 1 was highly (and negatively) correlated with EDT ($R = -0.98$), Ts ($R = -0.97$), and T30 ($R = -0.96$). C80 and Δt_i were positively correlated, but their significance was lower (0.77 and 0.70, respectively). The second factor was negatively correlated to BR (with a quite low significance $R = -0.63$) and positively related to LF but with a poor significance ($R = 0.40$).

Combining the things together it can be concluded that the large majority of the subjects was influenced by reverberation attributes and preferred lower values. Only a couple of subjects showed an attitude exactly opposite to the majority (preferring longer reverberation), and a small group showed a better agreement with factor 2 which also showed a generally positive correlation with most subjects, indicating a preference for lower BR values.

The good correlation between early decay time and subjective ratings suggested to plot the total ratings achieved by each church as a function of EDT (Fig. 9), showing that the global preference increased as EDT decreased to a minimum of 3.2 s (corresponding to church F), below this value preference decreased again as church B got a low normalized rating. Consequently, a quadratic regression model was applied, providing the following best-fit curve:

$$\mathrm{GREG} = -4.47\,\mathrm{EDT}^2 + 31.9\,\mathrm{EDT} - 27.2, \tag{1}$$

having high statistical significance ($R^2 = 0.85$, $F = 19.85$, $p = 0.001$) and showing a maximum at 3.6 s. A comparison with the preferred reverberation time calculated according to Ando's theory (which is equal to $23\tau_e$ [24]) shows that even assuming the long-time value, equal to 91 ms, the resulting preferred T30 should be 2.1 s, considerably lower than the value observed, suggesting that in churches, given the "sacred" character of the music, listeners may accept (and prefer) longer reverberation times than in concert halls.

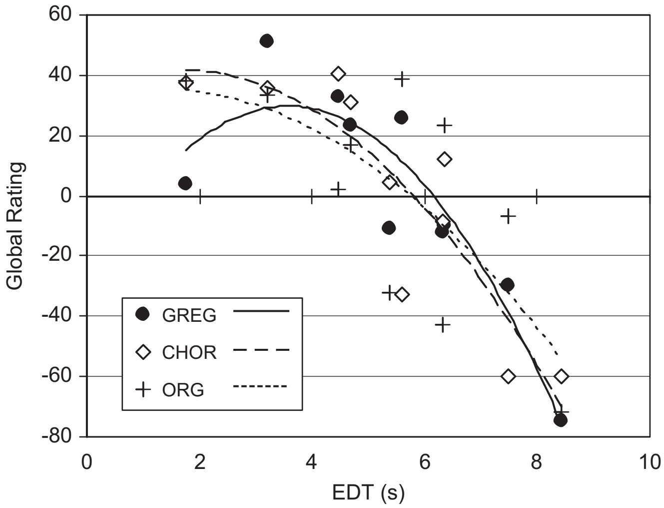

Fig. 9. Correlation between normalized global ratings for both vocal and organ motifs and EDT.

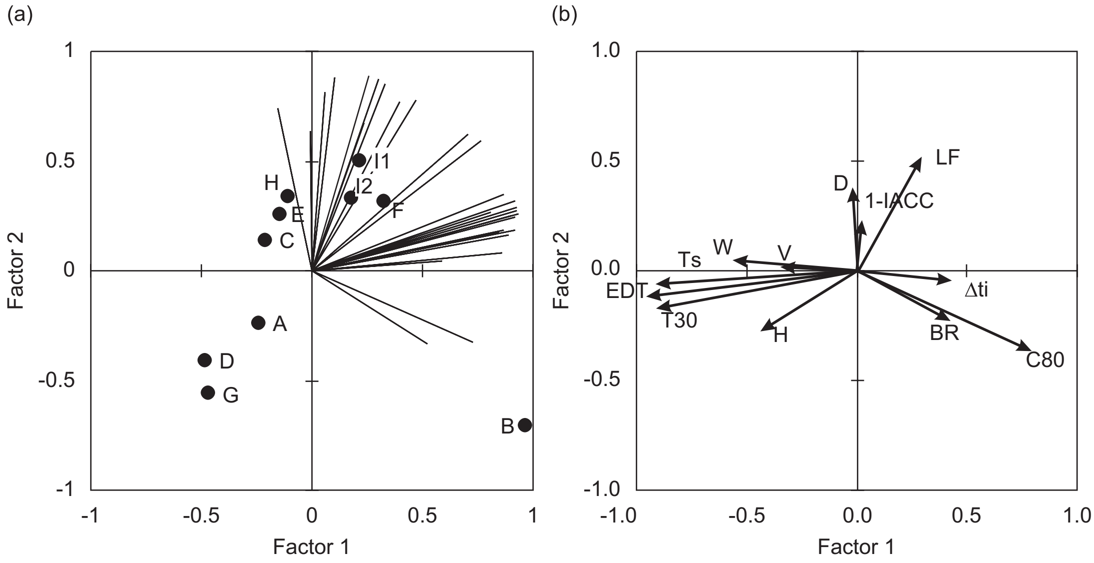

Fig. 10. (a) Preference plane for polyphonic choral music. (b) Correlations between objective parameters measured in the nine churches and subjective factors.

4.3. *Polyphonic choral music*

The analysis extracted two significant factors explaining 58 and 18 percent (48 and 29 percent after rotation) of the total variance. The *preference space* (Fig. 10a) shows that almost all the listeners gave a positive weight to both factors, with a few subjects giving a slightly negative weight to factor 1 or factor 2. The churches receiving the highest ratings were F, I, and B and the worst were again churches D and G.

The analysis of the correlations between the factors and the objective parameters (Fig. 10b) shows that factor 1 was highly (and negatively) correlated with EDT ($R = –0.95$), Ts ($R = –0.91$), and T30 ($R = –0.91$). Again, C80 was positively correlated, but its significance was lower ($R = 0.80$). All the other parameters showed poor correlations. The second factor was only related to LF although with relatively low significance

($R = 0.52$). BR performance was less significant in this case, possibly because the choral music excerpt used contained a lot less low frequencies and, consequently the preference for a flat response was less emphasized.

Combining the results it appears that a large part of the subjects preferred shorter reverberation, while a smaller group was indifferent to this attribute, showing a preference (although not clearly expressed) for higher values of LF. However, apart from BR there was a substantial agreement with the results obtained with the Gregorian excerpt, and even in this case the plot of the total ratings recorded for each church as a function of EDT shows (Fig. 9) that the global preference increased as EDT decreased up to a minimum of 4.5 s (corresponding to church I1), below this value preference remained mostly constant, indicating that for this kind of choral music (polyphonic and with varied dynamic) listeners accepted (and preferred) even the lowest reverberation time used in this study. The best fit curve was again quadratic with the following equation:

$$\mathrm{CHOR} = -2.45\ \mathrm{EDT}^2 + 8.18\ \mathrm{EDT} + 35, \tag{2}$$

having again high statistical significance ($R^2 = 0.79$, $F = 13.84$, $p = 0.004$), and showing a maximum at about 1.7 s. Even in this case this value is considerably higher than the preferred value calculated with Ando's theory which is 0.6 s.

4.4. Organ music

The analysis extracted two significant factors explaining 54 and 15 percent of the total variance. The effect of rotation was negligible, so the nonrotated solution was taken into account. The *preference space* (Fig. 11a) shows that a large majority of listeners gave a positive weight to factor 1, so that apart from three exceptions (showing a substantially indifferent attitude), it could be considered a consensus factor. Factor 2 played a less significant role and listeners gave contrasting ratings. The churches receiving the highest ratings were A, B, F, and H, whereas the worst were churches D and C.

The analysis of the correlations between the factors and the objective parameters (Fig. 11b) shows that factor 1 was highly correlated with Ts ($R = -0.88$), C80 ($R = 0.86$), while T30 and EDT had lower correlations (with $R = -0.78$) comparable with Δt_i ($R = 0.77$). LF was better correlated with factor 1 but its significance was low ($R = 0.56$). Among the geometrical parameters, church width showed a negative correlation with $R = -0.65$. The second factor was much less significant and was related only to BR ($R = -0.62$). This result is apparently comparable with that observed for Gregorian chant but in this case the subjective preferences are

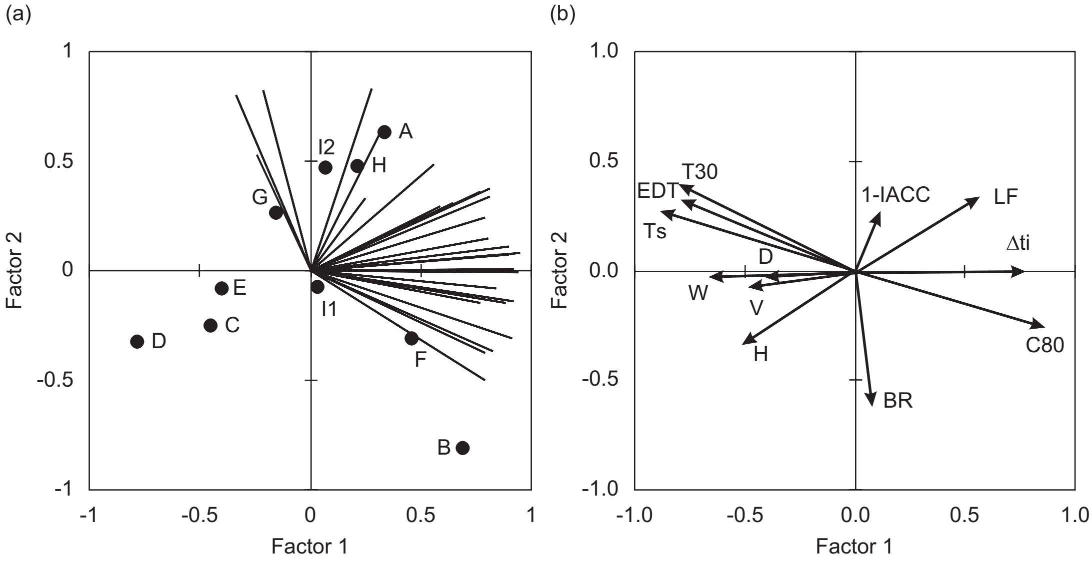

Fig. 11. (a) Preference plane for organ music. (b) Correlations between objective parameters measured in the nine churches and subjective factors.

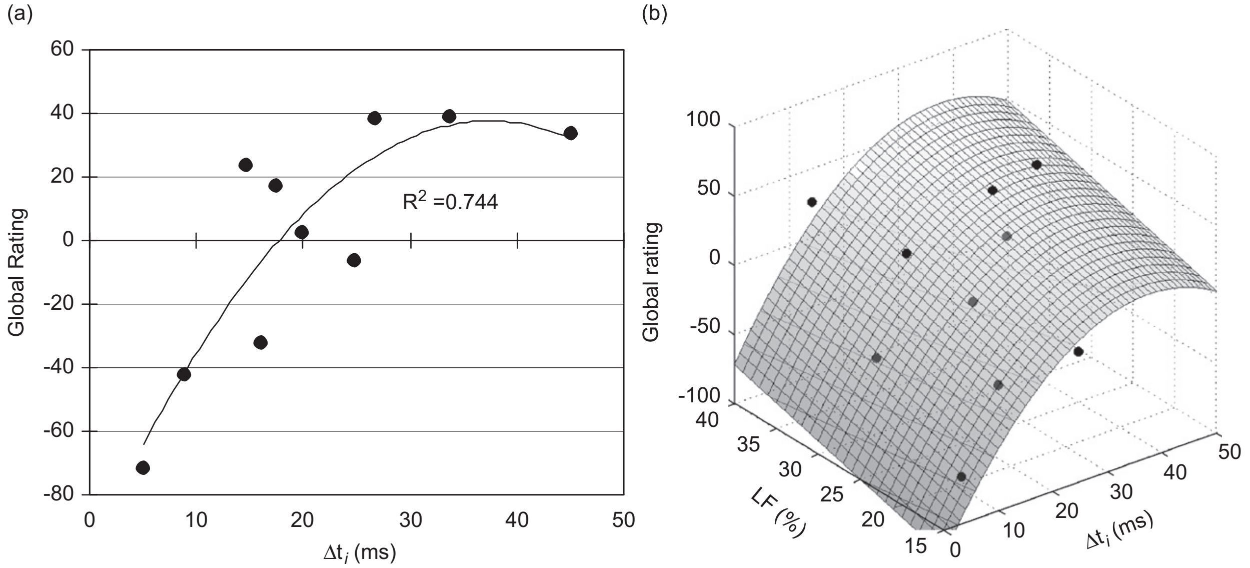

Fig. 12. (a) Correlation between normalized global ratings for organ and initial time delay (Δt_i). (b) Regression surface for normalized global ratings for organ music as a function of initial time delay (Δt_i) and lateral fraction (LF).

distributed across the factor with both positive and negative weights, indicating that it (and consequently BR) only represents individual differences.

The analysis of the correlations between factors and objective parameters shows a picture that differs from those observed before because on one hand the subjects were well clustered in preferring positive values of factor 1 but, on the other hand, none of the objective parameters showed a strong correlation with that factor. In fact, reverberation played a less significant role, possibly because of a wider range of preferred values (Fig. 9) which extends from 2 up to 6 s (including the optimal value of 3.8 s given by Ando's theory). Among the other parameters Ts and C80 showed the highest correlation with factor 1, but they were also weakly related to factor 2, while Δt_i was only related to factor 1 being a potentially good descriptor of subjective preferences.

In order to clarify which of these parameters could better represent subjective preference they were correlated with the total ratings, showing a clearly nonlinear behaviour because rating grows as Δt_i and C80 increase (and Ts and EDT decrease) and then stabilizes. As a consequence, a quadratic best-fit curve seemed appropriate and showed the best correlation with Δt_i ($R^2 = 0.74$), C80 ($R^2 = 0.70$) and Ts ($R^2 = 0.69$), followed by EDT ($R^2 = 0.53$). The latter correlation was less significant than the others but provided a regression equation very similar to that obtained for the choral music (Fig. 9). Among the other parameters Δt_i showed the best performance (Fig. 12a) indicating a preference for delays of about 40 ms, lower than the optimal value suggested by Ando who assumes $\Delta t_i \approx (\tau_e)_{\min}$ (in this case 164 ms). However, the longest delay observed in the present survey was 45 ms, so it cannot be stated whether longer values could have been preferred most.

Finally, the residual variance shown by the individual ratings proved to be well related to two independent parameters related to factor 1 (namely LF and W). When LF was used together with Δt_i the best performance was obtained, with a significant correlation with the residual variance ($R^2 = 0.63$). As a consequence, the total rating could be expressed as a combination of Δt_i and LF according to the following formula (Fig. 12b):

$$\mathrm{ORG} = -0.097\Delta t_i^2 + 7.258\Delta t_i + 1.84\mathrm{LF} - 145.9. \tag{3}$$

The correlation between predicted and actual ratings was high ($R^2 = 0.93$, $F = 45.8$, $p < 0.001$).

4.5. Classical symphonic music

For Classical symphonic music the analysis extracted two significant factors explaining 59 and 15 percent of the total variance. The *preference space* (Fig. 13a) shows that, as observed for organ music, the large majority

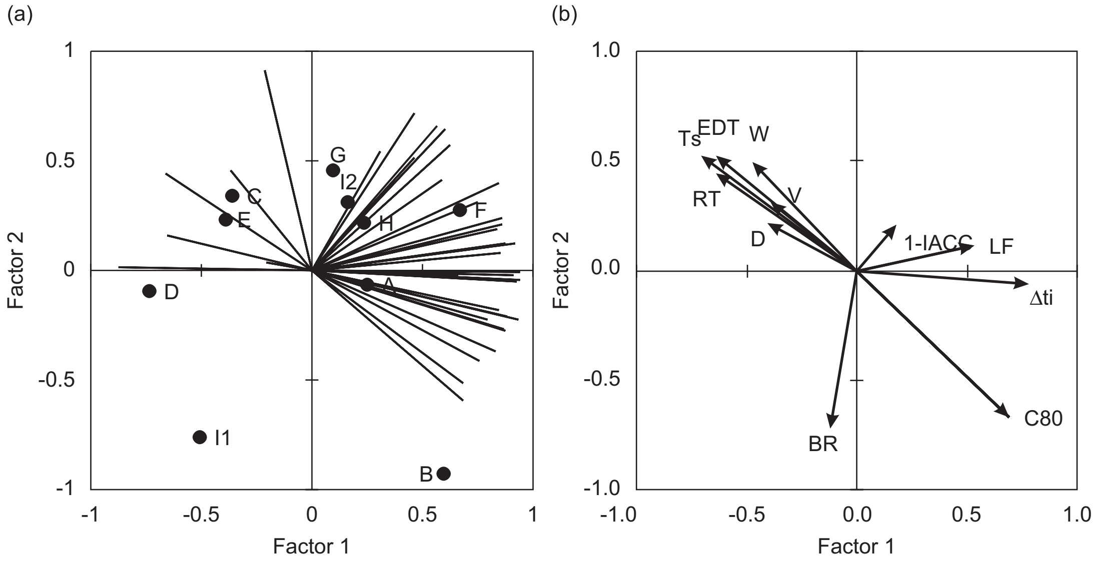

Fig. 13. (a) Preference plane for Classical symphonic music. (b) Correlations between objective parameters measured in the nine churches and subjective factors.

of the listeners was clustered around positive values of factor 1, with a minor number of subjects assigning negative weights to factor 1. As for organ music factor 2 plays a minor role with both positive and negative weights indicating that it only represents individual differences. The churches receiving the highest ratings were F, H, and B while the worst were churches D, I1, and E. It is particularly interesting to point out the difference between the ratings obtained by conditions I1 and I2, respectively, one of the worst and one of the best rated listening conditions. The two points mainly differ in the amount of lateral energy and inter-aural cross-correlation, confirming (although in a quite "extreme" way) the importance of spaciousness for symphonic music.

The analysis of the correlations between the factors and the objective parameters (Fig. 13b) shows that factor 1 was best correlated with Δt_i ($R = 0.78$), C80 ($R = 0.69$) and Ts ($R = –0.71$), while T30 and EDT had lower correlations, with R equal to -0.64. LF was related to factor 1 with a correlation of 0.53. Among the geometrical parameters the church width showed a negative correlation with $R = –0.48$. The second factor was related negatively to BR ($R = –0.72$) and to C80 ($R = –0.67$). These results are similar to those observed for organ music, but in this case the role of reverberation and clarity on the main factor is less significant, in particular because several listeners showed an indifferent attitude towards these parameters, possibly because for this kind of music (usually performed in different places) the attitude to accept longer reverberation times is reduced and, consequently, the subjects gave more importance to other acoustic parameters.

In fact, the analysis of the parameters related uniquely with factor 1 (namely Δt_i and LF) showed that taking one parameter at a time into account, Δt_i showed the highest linear correlation with $R^2 = 0.53$. Despite some fluctuations, the regression equation indicated a preference for increasing values of initial delay. Taking into account that the longest observed delay is 45 ms, this result is compatible with the predictions of Ando's theory which assumes a maximum towards 70 ms. As observed for organ music, the combination of multiple attributes of the sound field improved the accuracy of the prediction and the amount of variance explained. General linear models based on stepwise forward regression showed that a combination of Δt_i and LF provided better results, explaining an additional 19 percent of the variance, so that the final regression, characterized by $R^2 = 0.724$, $F = 9.17$, and $p = 0.011$, yields:

$$\mathrm{MOZ} = 2.27\mathrm{LF} + 2.03\Delta t_i - 103.2. \tag{4}$$

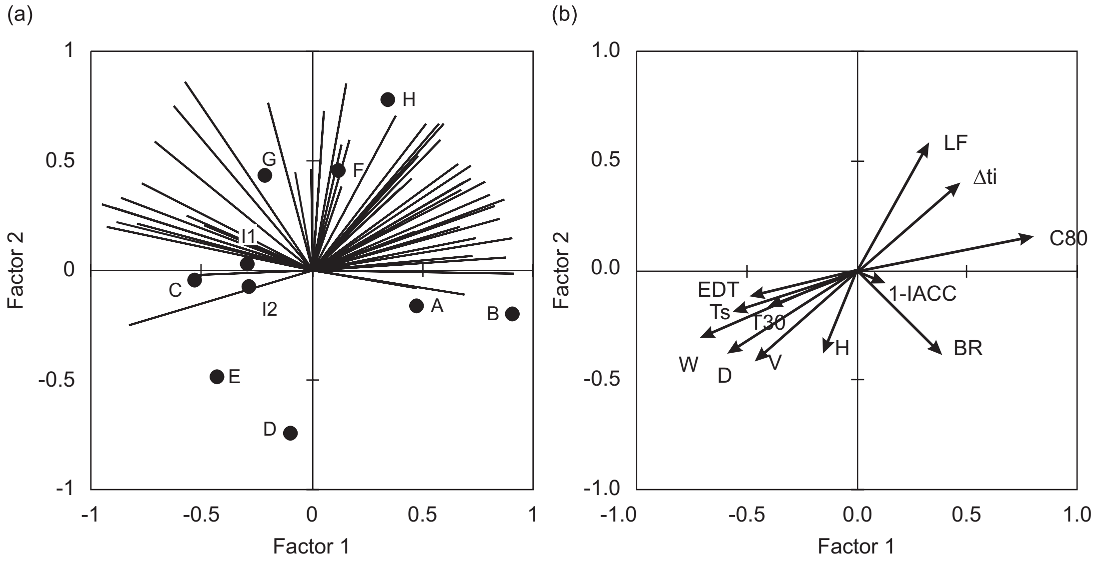

Fig. 14. (a) Preference plane for Romantic symphonic music. (b) Correlations between objective parameters measured in the nine churches and subjective factors.

4.6. Romantic symphonic music

The analysis extracted two significant factors explaining 46 and 17 percent (34 and 30 percent after the rotation) of the total variance. The *preference space* (Fig. 14a) shows that subjective responses were distributed over the first two quadrants, with two thirds of the listeners giving a positive weight to factor 1 and one third showing opposite preferences, indicating the presence of two distinct groups with different tastes. Conversely, factor 2, which practically explains the same variance as factor 1, nearly always received positive weights behaving as a consensus factor. The churches receiving the highest ratings were H, F, and B whereas the worst were churches D and E.

The analysis of the correlations between the factors and the objective parameters (Fig. 14b) shows that factor 1 was best correlated with C80 ($R = 0.80$) and W ($R = -0.72$), while T30 and EDT had lower correlations, with R below -0.50. The second factor was significantly related only to LF ($R = 0.59$). In addition to the reasons given for Classical music, the poor performance of the reverberation parameters, including Ts, might be explained as a result of the long crescendo, which probably masked the actual reverberation, and of the complex nature of the music motif which, being slow at the beginning and fast at the end, might have induced the listener to prefer different conditions according to the part they gave more importance.

The distribution of the subjective preferences across both sides of factor 1 indicate that the latter only represents individual differences and consequently no significant correlation between total ratings and objective parameters related with it might be found. Conversely, LF, the best related parameter with factor 2 (which behaves as consensus factor) also shows the best correlation with total ratings even though the correlation is quite low ($R^2 = 0.40$, $p = 0.051$), mostly because of an outlier (church G), which got a strangely high rating despite the low LF. If this value is excluded from the analysis the correlation improves considerably ($R^2 = 0.767$, $F = 23.01$, $p = 0.002$) yielding the following equation:

$$\mathrm{BRK} = 2.77\mathrm{LF} - 78.1. \tag{5}$$

It is interesting to observe that despite the large variation in terms of correlation the resulting equations were quite similar, with a limiting value between negative and positive ratings of 26.4 percent in the first case, and of 28.2 percent in the second.

4.7. Discussion

The underlying pattern of objective parameters influencing the two significant factors was quite similar for all the music motifs. The reverberation parameters and centre time were negatively related to factor 1, while clarity and initial time delay gap were related positively. The LF was positively related to both factors but, more frequently, the highest correlation was with factor 2. The only other parameter showing a correlation with factor 2 was the bass ratio but in most cases the listeners were not in agreement on positive or negative values so it can be considered a descriptor of individual differences. The poor performance of the IACC can be explained by the particular combination of acoustic parameters used in this research (so that good IACC values were often associated to bad values of other acoustic parameters). The low correlation with LF ($R^2 = 0.42$), which is not unusual in churches where columns, pulpits, and other elements may contribute to decrease IACC without increasing LF, might explain the different reaction of the listeners to these parameters which should represent the same subjective attribute of the sound field. Among the geometric parameters only church width showed a negative correlation with factor 1 which was particularly significant when instrumental music was used, confirming the importance of having narrower rooms for this kind of music [2,4,8].

Despite the common pattern of objective parameters, subjective preferences changed significantly as a function of the music motif. Three different behaviours were identified.

For vocal music the listeners almost unanimously gave positive weight to the main factor, indicating that the preference was influenced by attributes related to reverberation, and in particular by early decay time, to which it was related by means of a quadratic regression equation with very high statistical significance.

For organ and Classical symphonic music the subjects were equally coherent in preferring the main factor which, in this case, was related to early decay time, clarity, initial time delay gap, and the LF. However, the reverberation parameters were correlated also to the second factor and consequently their correlation with total subjective ratings was lower. For organ music this depended on the wider range of preferred reverberation times (varying from 2 to 6 s), while for Classical symphonic music this probably depended on a reduced attitude to tolerate longer reverberation for this kind of music. A combination of Δt_i and LF, the parameters related only to factor 1, provided a best-fit surface that predicted subjective ratings quite accurately.

For Romantic symphonic music the largest individual differences appeared with a scattering of the preferences across factor 1, compensated by a substantial agreement on positive values of factor 2 (i.e. for largest LF values). The significant number of subjects showing indifference or a negative preference for factor 1 could be explained by the double character of the music motif that has a long crescendo that probably masked the reverberation, while the fast and rhythmic conclusion required higher clarity and shorter reverberation. In addition, some of the subjects reported that since they could not find ideal listening conditions for symphonic music they tended to prefer those aspects that emphasized the "character" of the place. This might explain the presence of two distinct groups of preference with respect to factor 1, and the consequently lower amount of variance explained. Conversely, the substantial agreement on factor 2 explained the good performance of LF in correlating with global ratings.

Although with different degrees of significance, the results showed that optimal listening conditions in churches depended, as a function of different music motifs, on reverberation, LF, and initial time delay. The three independent acoustic parameters are in good agreement with those generally accepted for concert halls [3,4,8,24] where LF is sometimes replaced with IACC and the loudness level (which was deliberately neglected in this paper) also plays an important part. However, it is worth noticing that in none of the observed cases the three parameters appeared combined together. For vocal music the reverberation explained nearly all the variance of the subjective preferences, but it had no influence on preferences for instrumental music which followed a sort of adaptation process deriving from the impossibility of finding optimal conditions. The result of this adaptation was an indifference towards reverberation in favour of acoustic parameters that had values closer to those observed in concert halls. A similar behaviour was observed also in Ref. [8] where reverberation was well related to subjective ratings only when halls with T30 below 2 s were considered.

A similar adaptation process was also observed in terms of preferred values of the selected parameters, and in particular of EDT. In fact, even though optimal values determined for vocal and organ music according to Ando's theory as a function of $(\tau_e)_{\min}$ are based on different premise, and consequently any comparison

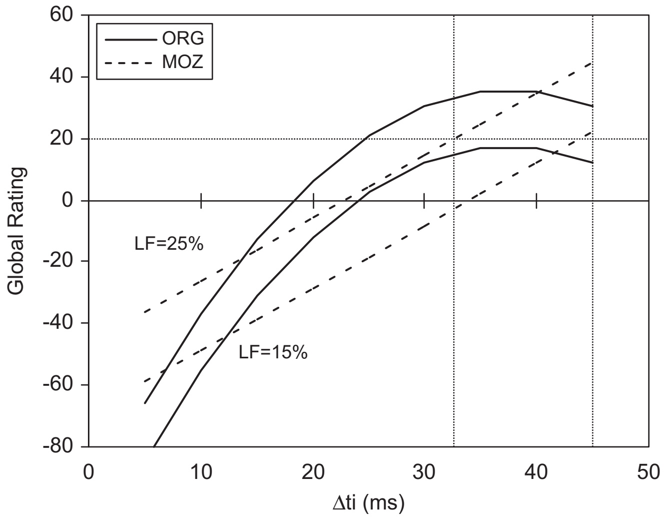

Fig. 15. Regression curves for normalized global ratings for instrumental music as a function of initial time delay Δt_i for given LF values.

should be considered carefully, the results showed that in churches listeners preferred longer values varying from 1.7 s for the polyphonic excerpt to 3.6 s for Gregorian chant. Similarly, preferred values of initial time delay are longer than those suggested by Beranek [3] for concert-halls (varying between 15 and 25 ms) and lower than those suggested by Ando as a function of $(\tau_e)_{\min}$. LF is the only parameter which shows preferred values close to 30 percent, in good agreement with those determined for concert-halls.

The dependence of optimal listening conditions on EDT for vocal music, and on Δt_i and LF for instrumental music suggests that ideal acoustics for a wide range of purposes might be obtained in churches by properly adjusting the above mentioned parameters. As can be seen in Fig. 9, if a minimum rating of +20 is assumed as satisfactory, EDT values varying between 2.1 and 4.6 s provide good listening conditions for both Gregorian and polyphonic music. The regression curve drawn for organ music, despite its lower significance, shows that limiting the interval to 4.2 s can also provide satisfactory values for this kind of music. This interval appears in good agreement with the lower limits of optimal reverberation times found by Desarnaulds [10] and Meyer [11] as a function of the volume. However, those values are referred to unoccupied churches, while in the present work the subjects judged the presented listening conditions, therefore the optimal interval found should be better referred to occupied conditions. Consequently, the agreement with the lower interval found in Refs. [10,11] appears reasonable as longer reverberation may be found in unoccupied conditions. For organ and Classical symphonic music the projection of the different regression surfaces (Fig. 15) shows that when LF is below 15 percent no Δt_i value can provide a satisfactory rating (above +20). Conversely, when LF equals 25 percent any Δt_i varying between 32 and 45 ms ensures a rating above +20 independent of the music program. For Romantic music the requirements for LF are a bit stricter because values above 28 percent only ensure a positive rating, while LF must be at least 35 percent in order to have ratings above +20.

5. Conclusions

This paper presents the results of research aimed at investigating the subjective evaluation of church acoustics by means of laboratory listening tests based on paired comparisons and carried out in a specifically designed room using transaural presentation of binaural recordings, with loudspeakers arranged according to the stereo dipole configuration. Five different music motifs, spanning Gregorian chants to Romantic symphonic music, were used in order to test the churches under different conditions. The preference matrix obtained by collecting subjects' rankings was subjected to factor analysis.

All the analysed musical motifs showed a very similar underlying pattern of objective parameters influencing the two significant factors. Reverberation parameters and centre time were negatively related to the main factor, while clarity was related negatively. The LF was positively related to both factors but, in general, the highest correlation was with the secondary factor. Among the geometric parameters only church width showed an acceptable (negative) correlation with the main factor.

Despite the common pattern of objective parameters, subjective preferences changed significantly as a function of the musical motif. For vocal music the preference was influenced almost exclusively by reverberation and by EDT in particular. For organ and for Classical symphonic music the preference depended on a combination of Δt_i and LF, even though EDT still played an important role for organ. For Romantic symphonic music the greatest individual differences appeared, with a scattering of the preferences across the main factor mostly due to the impossibility of finding ideal listening conditions which induced the listeners to prefer those cases that emphasized the "sacred character" of the place. The analysis of the correlations between total ratings and acoustical parameters allowed the definition of an optimal range of EDT from 2.1 to 4.2 s for vocal and organ music. For instrumental music the optimal listening conditions depended on the combination of initial time delay and LF.

Further investigations are required in order to obtain even more realistic listening conditions, using more than two channels and searching for a more accurate simulation of real sound sources. In addition a proper choice of different listening conditions should be made in order to clarify the relationship between subjective preference and acoustic and geometric parameters. The findings of the present research and the large amount of available data might be the starting point for more detailed research, possibly involving speech intelligibility as well as music and singing.

Acknowledgements

This study has been carried out within the national interest program of scientific research "*The acoustics of worship places*", funded by the Italian Ministry of Universities and Research. Special thanks are due to Prof. Ettore Cirillo for his continuous support and interest in the present research and to Bengt-Inge Dalenback for his comments and his kind collaboration in reviewing this paper. The author is finally indebted to Michele D'Alba who untiringly organized and carried out the listening tests and to Christian Skaug for his help with the statistical analysis of the results. A final thank is due to the anonymous reviewers that with their valuable comments helped to improve this paper.

References

[1] Constitution on the Sacred Liturgy, Promulgated by Pope Paul VI, 1963.
[2] M. Barron, *Architectural Acoustics and Auditorium Design*, Chapman & Hall, London, 1993.
[3] L.L. Beranek, *Concert and Opera Halls—How They Sound*, AIP, Woodbury, 1997.
[4] M. Barron, Subjective study of British symphony concert halls, *Acustica* 66 (1988) 1–14.
[5] A. Farina, Acoustic quality of theatres: correlations between experimental measures and subjective evaluations, *Applied Acoustics* 62 (2001) 889–916.
[6] Y. Ando, Calculation of subjective preference at each seat in a concert hall, *Journal of Acoustical Society of America* 74 (1983) 873–887.
[7] L. Cremer, H.A. Muller, *Principles and Applications of Room Acoustics*, Vol. 1, Applied Science, London, 1982.
[8] M.R. Schroeder, D. Gottlob, K.F. Siebrasse, Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters, *Journal of Acoustical Society of America* 56 (4) (1974) 1195–1974.
[9] A.P.O. Carvalho, A.E.J. Morgado, L. Henrique, Relationship between subjective and objective acoustical measures in churches, *Building Acoustics* 4 (1) (1997) 1–20.
[10] V. Desarnaulds, De l'Acoustique des églises en Suisse—Une Approche Pluridisciplinaire (The Acoustics of Swiss Churches—A Multidisciplinary Approach), Doctoral Thesis, Ecole Polytechnique Federale de Lausanne, 2002.
[11] J. Meyer, *Kirchenakustik (Church Acoustics)*, Verlag Erwin Bochinsky, Frankfurt am Main, 2003.
[12] E. Cirillo, F. Martellotta, *Worship Acoustics and Architecture*, Multi-Science Publishing, Brentwood, 2006.
[13] ISO-3382, *Acoustics—Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, ISO, Geneva, Switzerland, 1997.
[14] S. Müller, P. Massarani, Transfer-function measurement with sweeps, *Journal of the Audio Engineering Society* 49 (2001) 443–471.
[15] E. Cirillo, F. Martellotta, Sound propagation and energy relations in churches, *Journal of the Acoustical Society of America* 118 (1) (2005) 232–248.
[16] O. Kirkeby, P.A. Nelson, H. Hamada, The stereo dipole—a virtual source imaging system using two closely spaced loudspeakers, *Journal of the Audio Engineering Society* 46 (5) (1998) 387–395.

[17] M.R. Bai, C.C. Lee, Objective and subjective analysis of effects of listening angle on cross-talk cancellation in spatial sound reproduction, *Journal of Acoustical Society of America* 120 (4) (2006) 1976–1989.

[18] N. Prodi, S. Velecka, The evaluation of binaural playback systems for virtual sound fields, *Applied Acoustics* 64 (2003) 147–161.

[19] T. Takeuchi, P.A. Nelson, Robustness to head misalignment of virtual sound imaging systems, *Journal of Acoustical Society of America* 109 (2001) 958–971.

[20] M.R. Bai, C.W. Tung, C.C. Lee, Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm, *Journal of Acoustical Society of America* 117 (2005) 2802–2813.

[21] O. Kirkeby, P.A. Nelson, P. Rubak, A. Farina, Design of cross-talk cancellation networks by using fast deconvolution, *Proceedings of 106th AES Convention*, Munich, 8–11 May 1999.

[22] F.L. Wightman, D.J. Kistler, The dominant role of low-frequency interaural time differences in sound localization, *Journal of the Acoustical Society of America* 91 (3) (1992) 1648–1661.

[23] R.G. Klumpp, H.R. Eady, Some measurements of interaural time difference thresholds, *Journal of the Acoustical Society of America* 28 (5) (1956) 859–860.

[24] Y. Ando, *Architectural Acoustics*, Springer, New York, 1998.

[25] ITU-R BS.1116-1. Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems, Geneva, Switzerland, 1994.

[26] V. Rioux, Sound Quality of Flue Organ Pipes—An Interdisciplinary Study on the Art of Voicing, Doctoral Thesis, Chalmers University Gothemburg, 2001.

[27] E. Parizet, Paired comparison listening tests and circular error rates, *Acustica—Acta Acustica* 88 (4) (2002) 594–598.

[28] J.O. Kim, C.W. Mueller, *Introduction to Factor Analysis*, Sage Publications, Newbury Park, CA, 1978.

[29] G. Plenge, P. Lehmann, R. Wettschureck, H. Wilkens, New methods in architectural investigations to evaluate the acoustic qualities of concert halls, *Journal of the Acoustical Society of America* 57 (6) (1975) 1292–1299.