



Audio Engineering Society Convention Paper 6485

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Reproduction of auditorium spatial impression with binaural and stereophonic sound systems

Paolo Martignon¹, Andrea Azzali¹, Densil Cabrera², Andrea Capra¹, and Angelo Farina¹

¹ Industrial Engineering Department, Università di Parma, Via delle Scienze, 43100 Parma, Italy
paolo.martignon@inwind.it

² School of Architecture, Design Science and Planning, University of Sydney
Sydney, NSW 2006, Australia
densil@arch.usyd.edu.au

ABSTRACT

Binaural room impulse responses convolved with anechoic recordings are commonly used in auditorium acoustics design and research. Binaural and stereophonic (O.R.T.F.) room impulse responses, which had been recorded in five concert auditoria, were used in this study to test the spatial audio quality of four reproduction systems: conventional stereophony, binaural headphones, stereo dipole, and double stereo dipole. Anechoic music, convolved with the impulse responses, was reproduced over these systems. The systems were matched as closely as possible to each other, and to the sound levels that would occur in the auditoria for the musical source. In a subjective test, subjects rated the room size, sound source distance and realism of the reproduction. The stereo dipole and O.R.T.F. stereophonic systems appear to work better than the headphone and double stereo dipole systems.

1. INTRODUCTION

Binaural audio recordings and binaural room impulse responses convolved with anechoic recordings are commonly used in auditorium and room acoustics design and research. Without individualization, such recordings and convolutions may be subject to substantial spatial distortions when listened to using headphones or other playback systems designed for

binaural signals. Since localization of sound around the aural axis depends largely on the highly individual acoustical filtering provided by pinnae, localization is a primary aspect of this spatial distortion. Nevertheless, non-individualized binaural recordings are very convenient, in terms of being easy to obtain through room acoustical measurement and computer simulation, as well as from existing databases. Despite their limitations, they can certainly be helpful in appreciating the acoustical qualities of auditoria, at least in relative

terms. This study examines three options for presenting audio recordings from concert auditoria in binaural format, as well as conventional stereophonic presentation. It investigates the ability of the audio reproduction formats to convey sound source distance and room size in the context of concert auditoria, and rates the subjectively assessed realism of the audio formats.

1.1. Two-channel audio formats

This section summarizes key characteristics of the audio formats considered in this research project.

1.1.1. Binaural techniques

Dummy head recordings and binaural simulations record or predict the sound at the ears, which can then be reproduced using headphones or other techniques including cross-talk canceling loudspeaker systems. A thorough review of binaural techniques, especially using headphone presentation, is given by Møller [1]. He summarizes the problems of binaural headphone techniques as including localization errors around the cones of confusion (and especially the difficulty in establishing a frontally localized source), and a lack of response of the system to head movements. While the former of these problems can be solved using individualization, and the latter using head-tracking, the present paper is concerned with systems with neither individualization nor head-tracking. Other authors cite inside-the-head localization as a problem, but Møller et al. [2] find no instances of this in test using a carefully calibrated non-individualized binaural headphone system. Headphone equalization is probably the most subtle key aspect of using a non-individualized binaural headphone system: simply reproducing a dummy head recording over unequalized headphones means that the sound is subject to the manufacturer's designed frequency response (which is unlikely to be optimized for binaural reproduction), and subject to effects of both the dummy head ear and listener's ear effects. One solution involves compensating for the non-flat transfer function between the headphones and the microphones of the original dummy head used to make the recordings. Møller et al. [2] find that the error in auditory distance perception increases when using non-individualized a headphone binaural system (compared to individualized headphone binaural, and to natural listening, for source distances of up to 5 m), but they did not find a systematic shift in perceived distance.

Cross-talk cancellation provides an alternative to headphones for presenting binaural recordings and simulations. Originally proposed in the 1960s [3, 4], this approach was famously used for auditorium acoustical assessment by Schroeder et al. in 1974 [5]. This technique reproduces the sound from the two ears of a head (or model or simulation thereof) at the two ears of a listener, using at least two loudspeakers. At a specified head position, the cross-talk from the right loudspeaker to left ear, and from the left loudspeaker to right ear, is cancelled by signals from the complementary loudspeaker. There are limits to this at low frequencies, because inter-aural level differences are naturally small or negligible. The short wavelengths at high frequencies can make the listener's head position critical for effective operation. Cross-talk cancellation also requires an absorbent acoustic environment to be effective.

More recently, a refinement of cross-talk cancellation known as the stereo dipole has been developed, investigated and applied. This is a type of cross-talk cancellation where the two loudspeakers are located close together, so as to approximate co-located monopole and dipole sources. Kirkeby et al. [6] find that this configuration (with a 10° interval between loudspeakers as seen by the listener) minimizes the ringing artifacts in the cross-talk canceling filters, and expands the area in which the cross-talk cancellation is effective (allowing greater listener head movement, [cf. 7]). The cost of closely located sound sources is that the low frequencies require a great boost, and so cross-talk cancellation at low frequencies becomes very inefficient. One solution to this problem is to have greater separation between low frequency drivers than high frequency drivers. Another solution is to institute a cut-off frequency below which cross-talk cancellation is abandoned, and the loudspeakers merely reproduce the binaural channels without additional processing. The present study, which uses stereo dipole, applies both of these solutions.

One clear advantage of the stereo dipole technique over binaural headphones is its ability to generate frontally located auditory images. Having the loudspeakers at what is probably the most important position for localization appears to solve this problem. Another related advantage is that, to the extent that the system tolerates head movements, the sound field is not locked to the head, and so localization may be able to benefit from at least small head movements.

The double stereo dipole is an extension of the simple stereo dipole system, with both front and rear stereo dipole loudspeaker pairs. This facilitates the impression of sound coming from behind the listener. However, the listener head position becomes critical for this loudspeaker arrangement, because the desired interference between front and rear stereo dipoles occurs over a quarter of a wavelength.

1.1.2. Conventional stereophony

Conventional two-channel stereophony is perhaps not used at all in auditorium acoustics research. However, it is very commonly used in music reproduction for entertainment purposes, and there are innumerable recordings of musical performances in auditoria made using various stereophonic microphone techniques. The present study uses the O.R.T.F. stereophonic microphone array, consisting of two cardioid microphones separated by 17 cm and by an angle of 110°. In a comparison of various stereophonic microphone arrays, Hugonnet and Jouhaneau [8] find that coincident techniques (such as XY and MS) yield the most accurate lateral localization, while closely spaced techniques (including O.R.T.F.) yield the finest distance discrimination. In another comparison, Ceoen [9] found a subjective preference for recordings made using the O.R.T.F. system (these were recordings of an orchestra in an auditorium), and this preference appears to be due to the configuration's ability to convey the spatial impression of the auditorium [10].

2. METHOD

2.1. Auditoria and impulse response measurements

This study exploits a collection of auditorium impulse responses previously made by Farina and colleagues [11]. The key characteristic of the selected impulse responses is that the same equipment and procedure was used in each case, with the signal gain structures fully documented. Measurements had been made using a dodecahedron loudspeaker plus a subwoofer as the sound source on stage. The test signal was an exponential swept sine wave. Equalization had been applied to this signal for a constant spatially averaged output power from the loudspeaker. A Neumann KU70 dummy head was used as the binaural microphone, with a pair of Neumann AK40 cardioid microphones in the O.R.T.F. configuration for two channel stereophonic recording. In addition, a Soundfield B-format

microphone, which includes an omnidirectional output channel, was on a boom 1 m ahead of the dummy head. This configuration and method is described in more detail by Farina and Ayalon [11].

The five auditoria used in this study were the large, medium and small halls in Rome's Parco della Musica, Parma's Auditorium Paganini, and Kirishima's Miyama Conseru in Japan. Two receiver positions were chosen for each auditorium. In every case, the receiver was on the longitudinal axis of symmetry of the auditorium, and the source 1 m off this axis, on the stage.

Room acoustical parameters were extracted from the selected impulse responses. These included reverberation time (T30), early decay time, clarity index (C80), speech transmission index, bass ratio, treble ratio, lateral fraction, and inter-aural cross correlation coefficient (IACC). Octave band values were transformed to single number values using the recommendations in ISO3382 [12]. Strength factor (G) was not determined, but the reproduced sound pressure level (L_{eq}) of each stimulus (see below) was.

2.2. Listening room and apparatus

The listening room floor was 4.5 m x 3.2 m, with a ceiling height of 4.2 m. Sound absorbing panels were attached to most of the wall space up to a height of 2 m. Absorbers were also suspended near the ceiling, and placed on the floor. Materials likely to absorb low frequency sound (such as cardboard panels and boxes) were included in the room acoustical absorption. The measured mid-frequency reverberation time (using the experiment loudspeakers as sources, and dummy head in the subject's position as receiver) was 0.2 s, with an increase in reverberation time the low frequency range. Background noise level, with the audio equipment operating, was measured at NCB 25 [13].

The axis of symmetry of the loudspeaker array was not aligned with the room, nor was the listening position in the room's center. Loudspeakers were at a distance of 1.5 m from the listening position. Prototype Audiolink AL105 loudspeakers were used for the conventional stereophonic pair, $\pm 30^\circ$ from the median line of symmetry. Genelec S30D reference studio monitors were used for the front stereo dipole, on their sides so that the tweeters were 22 cm apart, the mid-range drivers 43 cm apart, and the woofers 83 cm apart (measuring between driver centres). This corresponds to respective angles of 4° , 8° , and 16° from the median

line of symmetry (the angle seen by the subject between loudspeakers is double these values). The rear stereo dipole pair had QSC AD-S82H loudspeakers, with driver centers separated by 45 cm, corresponding to a 9° angle from the midline.

Although different loudspeaker models were used, the frequency responses of all systems were matched using 4096 tap inverse filters between 100 Hz and 20 kHz, developed using the algorithm of Kirkeby et al. [14]. One point in favour of this system matching was that the audio content of the experiment was undemanding on the loudspeakers, having little low frequency content and requiring only modest sound pressure levels at the listening position. Specifically, inverse filters were designed: (i) for the conventional stereophonic system to flatten the frequency response to an omnidirectional measurement microphone at the listener position; (ii) for the headphones to flatten the frequency response from the headphones to the dummy head; and (iii) for the stereo dipole systems, to provide cross-talk cancellation from 250 Hz and a flat frequency response between the binaural channels and dummy head (in the listening position) from 100 Hz.

Although the room had windows, they were almost entirely covered with opaque panels, so that the experiment was conducted in the light of the computer monitor, with just a little additional ambient light. Most of the surfaces in the room, at least below a height of 2 m, were dark grey or black, and little other than the experiment computer display was visible to a subject once their eyes had adapted to the computer monitor.

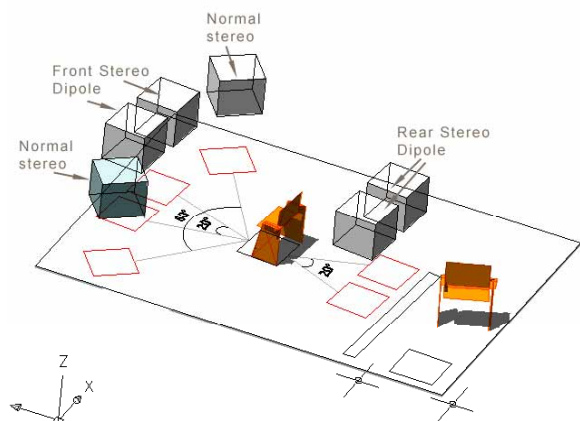


Figure 1 Sketch of the listening room configuration.

2.3. Stimulus generation

A calibrated anechoic recording was used in this project so that the reproduced sound pressure levels could be realistic. This was of a piano accordion, with a measurement microphone at a distance of 2.5 m directly in front of the performer. The music was “La ballata di Michè” (“Miky’s Ballad”), by Fabrizio de Andrè: a waltz, with a legato melody and articulated accompaniment. The octave band sound pressure levels of the source, normalised to 1 m, are shown in Figure 2. The A-weighted L_{eq} of the piano accordion normalized to 1 m is 80 dB(A). The recording was approximately 45 seconds in duration.

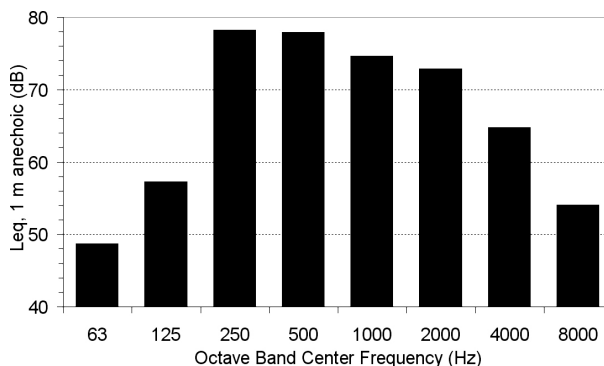


Figure 2 Octave band equivalent sound pressure level of the accordion, normalized to a microphone distance of 1 m.

Impulse responses created using a dodecahedron loudspeaker are not ideal for use in listening experiments (convolved with anechoic recordings). Typical sound sources, such as individual musical instruments or a human voice, are usually directional, rather than omnidirectional. An omnidirectional source will yield a lower direct-to-reverberant energy ratio than a source directed to the listener in an auditorium, resulting in reduced clarity for the listener. A second limitation of dodecahedral loudspeakers is their sensitivity as a function of frequency and radiation angle varies substantially due to interference between the twelve drivers. At high frequencies, the individual drivers also have their own directivity, resulting uneven sound radiation. The duration of an anechoic impulse response from a dodecahedral array is long, determined by the size of the dodecahedron. Although the room impulse responses used in this study were made with a dodecahedral loudspeaker (plus subwoofer), some attempt was made to address these problems. Firstly, the spatially averaged spectral irregularity of the

loudspeaker was compensated for by equalising the measurement signal (as mentioned previously). This is probably an adequate solution for all but the direct sound. Secondly, the direct sound was addressed by substituting the measured direct impulse with an ideal direct impulse. In the case of the O.R.T.F. impulse responses, this ideal signal was simply a single sample impulse, which has an almost flat frequency response up to the Nyquist frequency. For the dummy head the signal was the 0° anechoic impulse response for that dummy head. The direct sound of each room impulse response was measured, using a 256-sample fast Fourier transform (Blackmann-Harris window, sampling rate of 48 kHz) centered on the first major peak in the impulse response. The 256 sample ideal signals (with the impulse peak at the 129th sample) were substituted for the direct sound, scaled to have the same acoustic energy as the original 256 samples (measured at 500 Hz). The remaining part of the room impulse responses, consisting of early reflections and reverberant decay, was attenuated by 3 dB relative to the direct sound, thereby producing a simplistic approximation of a sound source with a directivity index of 3 dB facing the listening position.

Verification of the impulse response relative calibration was done by examining the relationship between the direct sound level and source-receiver distance. Notwithstanding effects of very early reflections, dissipation of acoustic energy in the air, and variation in loudspeaker directivity (depending on its orientation), the direct sound pressure level at the receiving position should follow the free field ideal of -6 dB per doubling of distance. Consistency with this principle was examined at 500 Hz (where air dissipation should be negligible, and the loudspeaker omnidirectional), as illustrated in Figure 3. There is general agreement between measurement and theory, with an rms error of less than 1 dB, but deviations of up to 2 dB.

The edited impulse responses (both ORTF and dummy head) were convolved with the anechoic recording of piano accordion, at a constant gain. In order to calibrate the gains of the playback systems in the listening room, a 500 Hz octave band noise signal was created with a known level difference to the anechoic recording microphone calibration tone. This was convolved with the direct impulse only of one of auditorium situations (O.R.T.F. format) using the same processing gain structure as for the music convolutions. The reproduced sound pressure level of the stereophonic loudspeaker system was adjusted to match that predicted by the

source-receiver distance in the auditorium (assuming direct sound only). This established the playback gain structure for the stereophonic system, such that the speech and accordion were reproduced in the listening room at approximately the same sound pressure levels as would have occurred in the auditoria.

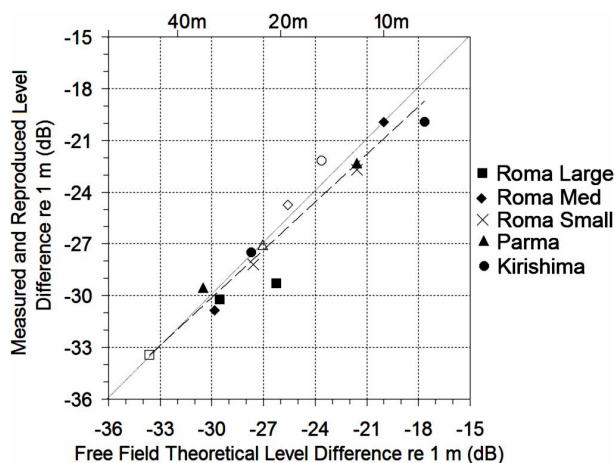


Figure 3 Comparison between theoretical free field and measured sound levels for various receiver positions in the five auditoria, at 500 Hz.

While the gains of the three binaural playback systems could be matched simply by dummy head measurements at the listening position, there is, to some extent, an arbitrary relationship between the stereophonic and binaural system gains, because their spatial sensitivity is different, and spatial sensitivity varies substantially with frequency in the case of the binaural system. It is certainly possible to match the microphone systems for free field sensitivity, or for diffuse field sensitivity – but these results are quite different, and in an auditorium the sound-field is at neither of these extremes. Therefore a simple approach to microphone system matching was taken in the playback system – such that the mean broadband sound pressure level difference of equivalent recordings (room impulse responses convolved with anechoic speech or accordion) was 0 dB (standard deviation of 1.2 dB). Having some stimuli with somewhat greater or lesser sound pressure levels over the binaural systems, relative to the stereo system) could influence the subjective parameters investigated, and as such was considered to be a useful component in the subjective comparison between these systems.

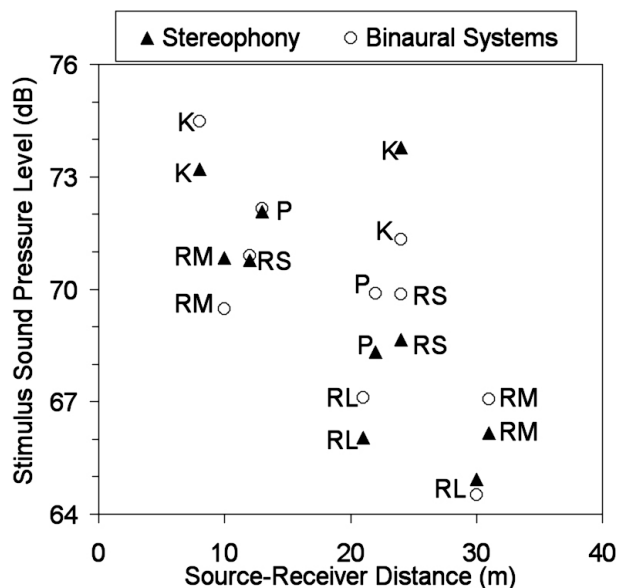


Figure 4 Unweighted Leq of the sound stimuli, measured with a dummy head microphone at the listener position in the listening room. Initials refer to the auditoria (Kirishima, Parma, Rome Large, Rome Medium and Rome Small).

2.4. Experiment Procedure

With ten auditorium situations, four audio playback systems and three response scales, presenting every stimulus to every subject was not considered to be feasible. Instead, each subject assessed five auditorium situations and two audio systems. The assignment of the auditorium situations and audio systems for each subject was done by counterbalancing between subjects.

The experiment was conducted using purpose-written software. The software presented the ten combinations of situation and audio system as randomly assigned buttons across the top of the visual interface (Fig 5). Pressing one of these buttons (using a wireless mouse) would cause the sound to play, and pressing another of them would switch the sound almost immediately to that of another stimulus, with approximately the same time in the musical performance. Hence, the subject could switch between stimuli whenever desired, listening to them in any order that they wished as many times as they wished. The three questions were displayed throughout the experiment, but only the first question was available for response until all stimuli received ratings (similarly, the third question was inactive until a full set of responses was received for the second

question). However, subjects could see and change their ratings for previous questions at any time. The question order was randomized between subjects.

The three questions were in Italian (Fig 5), and are roughly translated as “How large is the room that you are listening to?”, “How realistic is the sound?” and “How distant is the artist in meters?”

A computer screen was positioned directly in front of the subject (supported by the front stereo dipole loudspeakers). As well as presenting the response interface, the screen meant that the subject was almost always facing the front, which is an advantage for the loudspeaker based playback systems. The subject’s chair had a small integrated table, on which they operated a wireless computer mouse.

The subject was not given any information (other than the sound itself) on which loudspeaker system was being used for a stimulus. However, subjects were instructed by the computer program to put on the headphones when they switched to a headphone stimulus, and to remove the headphones when they switched to a loudspeaker stimulus. Clearly, this meant that subjects had a heightened awareness of the headphone technology, while the loudspeaker systems were differentiated merely by their sound.

Thirty subjects, all with musical backgrounds, participated in the experiment.

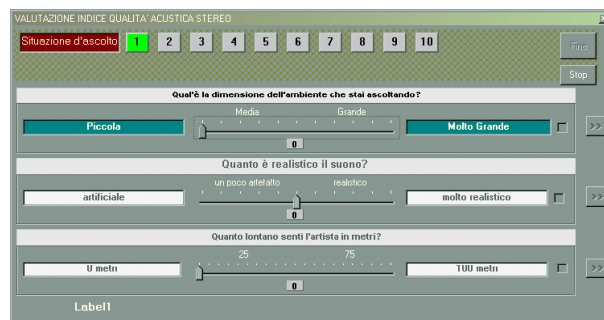


Figure 5 Control and response interface for the experiment with initial settings.

3. RESULTS

3.1. Auditory Distance Estimates

Analysis of variance (ANOVA) shows a significant effect for audio system ($f=3.86$, $p=0.0099$, $df=3$) and a stronger effect for situation ($f=11.45$, $p<0.0001$, $df=9$). A Scheffe test shows significant mean differences between binaural headphones and conventional stereophony ($p=0.015$), but not between any other pairs of audio systems. There are significant mean differences ($p<0.05$) between 16 of the 45 pairs of situations.

The results (Fig 6) show some match between physical and estimated distance for all four audio systems. While the best correlation is found for the double stereo dipole (Table 1), the smallest rms errors are found for O.R.T.F. stereophony and the single stereo dipole systems. Using logarithmic distance units, the stereo dipole system has smallest rms error. A correlation coefficient is not sensitive to absolute matches in values, but instead evaluates the goodness of fit of the data to a straight line. The rms error measures are sensitive to absolute deviations, and that using logarithmic units measures the error proportionate to distance (i.e. it tolerates larger errors at greater distances). The headphone system yields the weakest match of estimates to source-receiver distance, in all three evaluations. The authors favor the logarithmic unit rms evaluation.

	Correlation (r^2)	Rms Error (m)	Rms Error (log(m))
O.R.T.F.	0.39	9.3	0.23
Headphones	0.34	14.9	0.28
Stereo Dipole	0.59	9.6	0.19
Double Stereo Dipole	0.63	10.3	0.22

Table 1 Correlations and rms errors for auditory distance estimates, with respect to physical source-receiver distance.

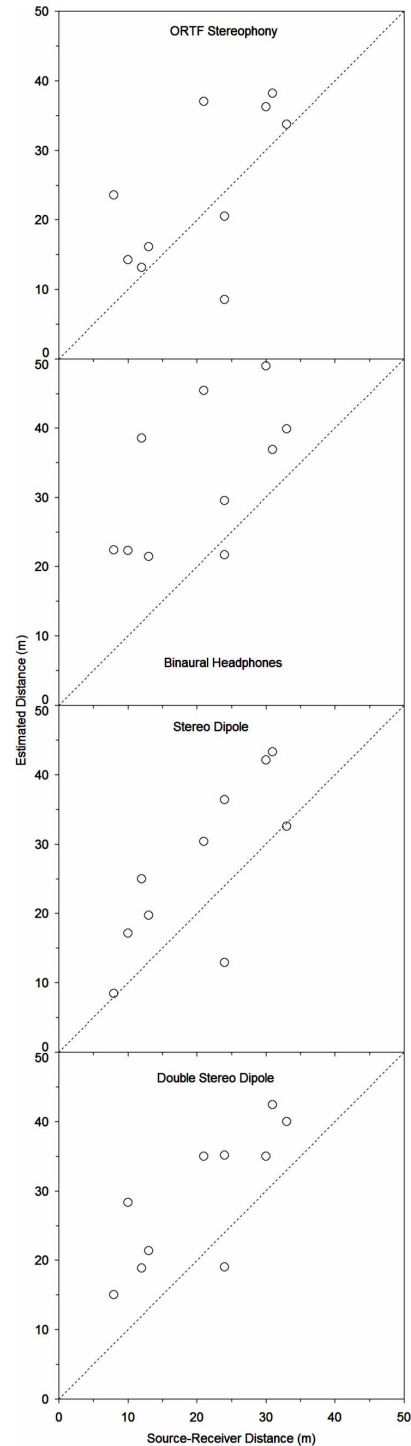


Figure 6 Mean auditory distance estimates for the four audio systems, shown in relation to the source-receiver distance of the impulse response measurements.

Generally the sound pressure level of the stimuli decreases with source-receiver distance (as shown in Fig 4). However, in the Kirishima concert hall, the 24 m distance received approximately the same sound pressure level as the 8 m distance using the O.R.T.F. microphone array. For the same pair of positions, the binaural microphone array sees a 3 dB reduction in level over distance. While these effects are explained by the unusual design of the auditorium (especially the ceiling reflection), and the different spatial sensitivity of the microphone arrays, they create a situation where auditory distance perception is likely to diverge from veridical, and also is likely to differ for the two audio recording systems. The correlations between stimulus sound pressure level and distance are $r=-0.74$ and $r=-0.69$ for the binaural and stereophonic systems respectively.

Distance estimates are related to the sound pressure level of the stimuli, most strongly for conventional stereophony and the stereo dipole systems. Mid frequency reverberation time (T30 – ranging from 1.8 s to 2.4 s) and inter-aural cross correlation coefficient (IACC – ranging from 0.12 to 0.48) also are significant correlates of auditory distance for some of the audio systems, as shown in Table 2.

	SPL	T30	IACC
O.R.T.F.	-0.86	0.67	-0.39
Headphones	-0.79	0.58	-0.59
Stereo Dipole	-0.82	0.34	-0.73
Double Stereo Dipole	-0.76	0.44	-0.67

Table 2 Correlation coefficients (r) between objective stimulus or room acoustical measurements and auditory distance estimates.

3.2. Auditory Room Size Ratings

ANOVA shows that room size ratings are significantly affected by situation ($f=6.89$, $p<0.0001$, $df=9$), but not significantly by audio system ($f=2.4$, $p=0.066$, $df=3$). Alternatively, an analysis considering auditorium instead of individual situations shows a significant effect for auditorium ($f=8.47$, $p<0.0001$, $df=4$), and a similarly non-significant effect of audio system. Results are shown in Figure 7.

Auditorium length provides some correlation with auditory room size ratings, at least for O.R.T.F. stereophony ($r=0.91$ for mean ratings of auditoria). There are no other correlations between room dimensions (length, width, footprint) and room size ratings for any audio system. The Rome small hall's size appears to be overestimated for the three binaural techniques. Single stereo dipole is not sensitive to the Rome large hall's greater physical size.

	Physical Distance	Estimated Distance
O.R.T.F.	0.46	0.95
Headphones	0.44	0.85
Stereo Dipole	0.33	0.58
Double Stereo Dipole	0.56	0.86

Table 3 Correlation coefficients (r) between auditory room size ratings and source-receiver distance (physical and estimated).

To some extent, there is an inherent relationship between room size and source-receiver distance, because large distances are impossible in small rooms. This helps to explain the high correlations between distance estimates and room size ratings, shown in Table 3, for three of the four audio systems. However, these correlations are higher than the respective correlations between room size ratings and actual source-receiver distance. In the case of the O.R.T.F. system there is little to distinguish room size ratings from distance estimates. The largest distinction between these subjective scales is found for the stereo dipole system. Figure 8 compares the ratings for these two systems.

Table 4 shows correlations between stimulus or room acoustical parameters and auditory room size ratings. For binaural headphones, early decay time (EDT) is the strongest correlate. For the two stereo dipole systems, IACC is the strongest correlate. For conventional stereophony, the strongest correlate is stimulus SPL – as would be expected considering the close relationship with auditory distance estimates for this audio system – but correlation with reverberation time is almost as strong.

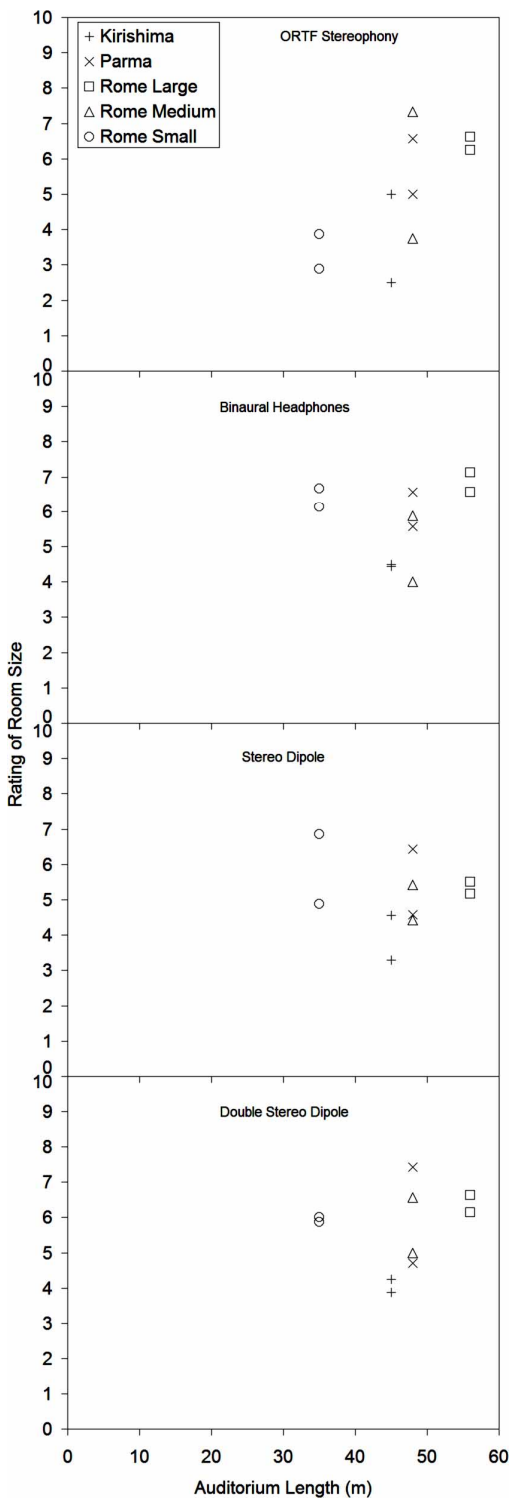


Figure 7 Mean auditory room size ratings for the four audio systems, shown in relation to the physical auditorium length.

	SPL	T30	EDT	IACC
O.R.T.F.	-0.74	0.72	0.57	-0.37
Headphones	-0.54	0.54	0.75	-0.69
Stereo Dipole	-0.43	0.18	0.32	-0.69
Double Stereo Dipole	-0.67	0.45	0.45	-0.79

Table 4 Correlation coefficients (*r*) between objective stimulus or room acoustical measurements and auditory room size ratings.

One striking difference between the room size ratings for the audio systems is in the results for the smallest auditorium (Rome Small). This auditorium receives larger room size ratings for the binaural systems than for O.R.T.F. stereophony. Kirishima, the second smallest auditorium, receives smaller room size ratings for the binaural systems. In terms of the acoustical parameters, IACC has a large contrast between these auditoria, with low values for Rome Small (0.14 and 0.15) and high values for Kirishima (0.48 and 0.45). The ability of the binaural systems to convey this contrast is inherently greater than the O.R.T.F. system, and this seems to be reflected in the correlations between room size ratings and IACC in Table 4.

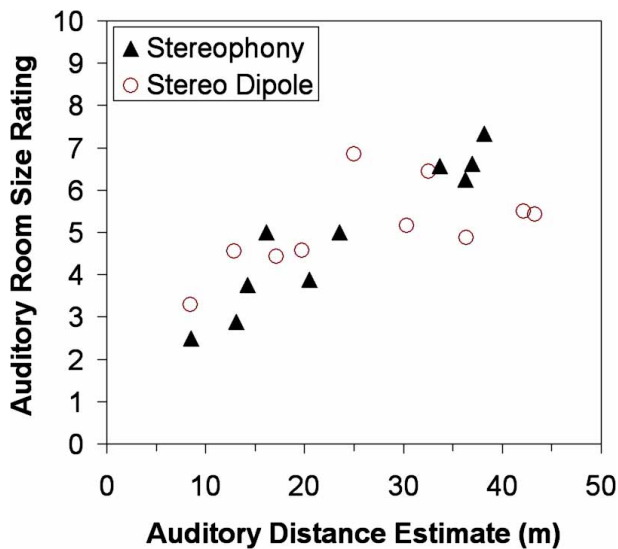


Figure 8 Comparison between auditory distance estimates and auditory room size ratings for the O.R.T.F. stereophonic system and the stereo dipole system.

3.3. Realism Ratings

ANOVA shows that situation does not significantly affect realism ratings ($p=0.3$), and that audio system significantly affects realism ($f=4.15$, $p=0.0068$, $df=3$). Binaural headphones were rated as the least realistic, and O.R.T.F. stereophony the most realistic (a Scheffe test shows that these two are significantly different). Single stereo dipole has a mean realism rating almost as great as O.R.T.F. stereophony, as shown in Fig 9.

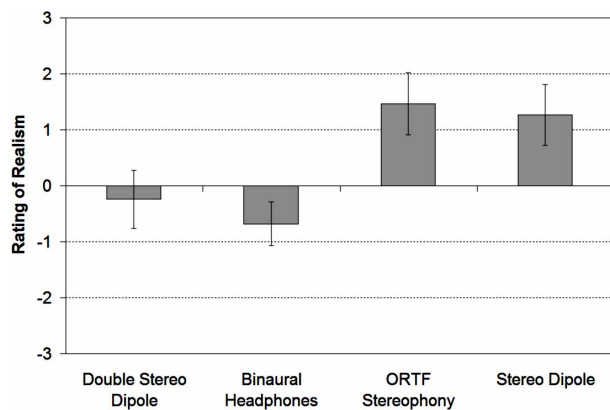


Figure 9 Mean auditory realism ratings for the four audio systems, ± 1 standard error.

It is not known how natural sound (in real concert halls) would be rated for realism. Nevertheless, we could assume that the subjects (who were experienced in music) were making judgments in reference to their memories of real concert auditorium sound. Subjects were asked to imagine themselves in an auditorium, rather than in a listening room with loudspeaker-reproduced sound. Assuming that these ratings do reflect experience of reality, then the O.R.T.F. stereophony and single stereo dipole system succeed best in conveying realistic sound to a listener.

4. DISCUSSION

As an assessment of four non-individualized two-channel audio systems for auditorium simulations, this study is limited by the fact that judgments of distance and room size have not been made in the actual auditoria. Hence, while it seems reasonable to rate systems based on the accuracy of subjective responses (e.g. accuracy of auditory distance estimates, in relation to source-receiver distances), it is not known whether auditory distance would be judged accurately were it possible to instantly transport blindfolded subjects

between the real auditoria. In the case of room size ratings, even though physical room length provides the best physical correlate for one audio system, it is not known whether such judgments in actual rooms would be similarly correlated to room length. The ratings of realism do not suffer this limitation, assuming that the actual auditoria would achieve full realism.

Previous studies of auditory perception of distance and room size show that the acoustical features of stimuli can have a strong effect, sometimes stronger than the effects of actual distance or room size. With respect to auditory distance perception in rooms, sound pressure level and aspects of reverberation (eg direct to reverberant ratio) can have strong effects. Unusually long reverberation times yield larger distance estimates [15, 16].

The weak or non-existent relationships between auditory room size ratings and actual room size in the present study are at odds with some previous study results, which showed that subjects can judge the physical size of rooms just by listening, at least in some circumstances [17, 18]. Nevertheless, previous studies also show that acoustical characteristics (especially reverberation time or reverberation level) can have a larger effect on perceived room size than the actual room size [17, 19, 20]. Since none of the rooms in the present study were small (all were large or very large), cues for discriminating room size were subtle, maybe too subtle for the actual room size to be conveyed when confounded with other differences between the auditorium situations. With regard to purely acoustic influences on room size perception, the four audio systems do not show the same tendencies – suggesting that further research is needed to understand this area.

There are natural correlations between the main acoustical cues for distance and room size. A small room is associated with high sound pressure levels (due to the reverberation level), and high sound pressure levels are also a cue to source proximity (due to the direct sound dispersion over distance). Reverberance is associated with large rooms (due to the long mean free path), and also with distant sources (due to the low direct to reverberant ratio). Hence, similarities between auditory distance estimates and room size ratings could be expected, although previous studies find some divergence between these [15, 20].

There are many other limitations to the study, including the use of a non-anechoic listening room (anechoic

conditions would be ideal for the cross-talk canceling systems), the use of different loudspeaker models (even though the frequency responses of these were matched), and the limited number of auditorium situations tested. Nevertheless, the study does yield apparently useful results such as:

- Binaural headphone systems are less effective than alternatives for auditorium simulations. Headphones yield low realism ratings and relatively poor estimates of distance. This result is striking because binaural headphone systems are widely used in auralization applications.
- The double stereo dipole system is relatively ineffective. However, the likely explanation of this is that the listener's head was not restrained, so that sound quality and image stability in the high frequency range could have been degraded by incidental movements.
- The single stereo dipole system is effective in terms of realism ratings and distance estimation. Of the three binaural systems tested, this appears to be the best. Not having the rear loudspeakers eliminates the front-back interference problem which degrades the double stereo dipole at high frequencies. While distance estimates and realism ratings are most distinct for single stereo dipole, the basis of these room size ratings is not clear (but appears to be partly influenced by IACC).
- The O.R.T.F. stereophonic system yields high ratings of realism, and appears to be the only system in which ratings of room size can be related to a physical variable (room length). However, distance estimations are less effective than for the stereo dipole system, and there is scarcely any distinction between distance estimates and room size ratings for the O.R.T.F. system.

An important distinction between the audio systems studied here and systems designed for entertainment is that the aim was realism, rather than listener enjoyment. The playback level of these systems was apparently less than typical playback levels for music entertainment [21, 22], but instead matched to the sound levels that would have been experienced for the instrument in the auditorium situations. Realism may or may not be a goal of entertainment systems, but it is a key attribute of any audio system to be used in the simulation of acoustic spaces for empirical research. While the O.R.T.F. and

stereo dipole systems both achieved good results in this study, the stereo dipole system has an inherent advantage over conventional stereophony in this respect, because it aims to convey the auditorium sound field experienced at the modeled head ears to the listener's ears. By contrast, conventional stereophony aims to reproduce the acoustic impression of the recorded space using a more approximate technique. Furthermore, it is not normally used at seat positions in an auditorium, but instead is used close to the stage, near the musical performance.

5. CONCLUSIONS

This study examined the reproduction sound quality of four non-individualized two-channel audio systems for a solo instrument in five concert auditoria. The main finding is that the stereo dipole appears to provide the most plausible reproduction. O.R.T.F. stereophony also yields a subjectively rated realistic reproduction, but fails to distinguish auditory distance from auditory room size perception. This may be related to the apparent influence of IACC on room size ratings in binaural systems. The problems with binaural headphone and double stereo dipole reproduction are well understood.

6. ACKNOWLEDGEMENTS

The authors are grateful for the assistance of Alberto Amendola, Paolo Bilzi, ASK Industries, Casa della Musica, and Tommaso Dradi (piano accordion) in this research project.

7. REFERENCES

- [1] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, Volume 36, Issue 3-4, pp. 171-218, 1992.
- [2] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: do we need individual recordings?" *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451-469, 1996.
- [3] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," *J. Audio Eng. Soc.*, vol. 9, pp. 148-151, 1961.
- [4] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Int. Conv. Rec.* 7, pp. 150-155, 1963.

- [5] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters," *Journal of the Acoustical Society of America*, vol. 56, no. 4, pp. 1195-1201, 1974.
- [6] O. Kirkeby, P. A. Nelson and H. Hamada, "The 'stereo dipole' – a virtual source imaging system using two closely spaced loudspeakers," *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387-395, 1998.
- [7] T. Takeuchi, P. A. Nelson, O Kirkeby and H. Hamada, "Robustness of the performance of the 'stereo dipole' to misalignment of head position," 102nd Audio Eng. Soc. Conv., Munich, Preprint 4464 (I7), 1997.
- [8] C. Hugonnet and J. Jouhaneau, "Comparative spatial transfer function of six different stereophonic systems," 82nd Audio Eng. Soc. Conv., London, Preprint 2465(H-5), 1987.
- [9] C. Ceoen, "Comparative stereophonic listening tests," *J. Audio Eng. Soc.*, vol. 20, no. 1, pp. 19-27, 1972.
- [10] M. Wöhr, G. Theile, H.-J. Goeres and A. Persterer, "Room related balancing technique method for optimizing recording quality," *J. Audio Eng. Soc.*, vol. 39, no. 9, pp. 623-631, 1991.
- [11] A. Farina and R. Ayalon, "Recording concert hall acoustics for posterity," 24th International Audio Eng. Soc. Conf. on Multichannel Audio, Banff, Canada, paper no. 38 (2003).
- [12] International Organization for Standardization, ISO 3382 (1997), *Acoustics—Measurement of reverberation time of rooms with reference to other acoustical parameters*
- [13] American National Standards Institute, ANSI S12.2-1995, *Criteria for Evaluating Room Noise*.
- [14] O. Kirkeby, P. A. Nelson, P. Rubak and A. Farina, "Design of cross-talk cancellation networks using fast deconvolution," *Audio Eng. Soc. 106th Conv.*, Munich, Germany, Preprint 4916 (J1).
- [15] D.H. Mershon, W.L. Ballenger, A.D. Little, P.L. McMurtry, and J.L. Buchanan, "Effects of room reflectance and background noise on perceived auditory distance," *Perception*, vol. 18, pp. 403-416, 1989.
- [16] D. Cabrera and D. Gilfillan, "Auditory distance perception of speech in the presence of noise," *Proc. Int. Conf. on Auditory Display*, Kyoto, Japan, pp. 431-439, 2002.
- [17] J. Sandvad, "Auditory perception of reverberant surroundings," *Journal of the Acoustical Society of America*, 105(2), Pt. 2, p. 1193 (paper 3pSP3), 1999.
- [18] R. McGrath, T. Waldmann, and M. Fernström, "Listening to rooms and objects," *Proceedings of the 16th Audio Eng. Soc. Int. Conf.*, Rovaniemi, Finland, pp512-522, 1996.
- [19] S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki, "Psychoacoustic cues in room size perception", *Proceedings of the 116th Audio Engineering Society Convention*, Berlin, 2004.
- [20] D. Cabrera, D. Jeong, H. J. Kwak and J.-Y. Kim, "Auditory room size perception for measured and modeled rooms," *Internoise*, Rio de Janeiro, 2005.
- [21] C. D. Mathers, K. F. L. Lansdowne, "Hearing risk to wearers of circumaural headphones: An investigation." *BBC Research Report RD 1979/3*.
- [22] Condamines, R., "Relation between the passband and the preferred listening level for music", *EBU Review*, no. 139, pp. 124 – 127, (June 1973).